Dagobert Soergel

UBLIS 571
Information Organization

Spring 2016

Lecture Notes

red

Weeks 1 - 14.
Reminder to look at the model catalog

Model catalog

Required. Refer to this throughout the course.

1 Soergel, Dagobert with Miller, Amy
Model Catalog for UBLIS 571. **Including a summary of the MARC Format.**
2016 January. 46 p.

The model catalog gives many examples of cataloging documents, including a Web site, using AACR2R (**A**nglo-**A**merican **C**ataloguing **R**ules. 2. edition, revised) and the MARC (**MA**chine **R**eadable **C**ataloging) format. It also gives for the same documents a comparison between cataloging using AACR2, (the old cataloging rules that are still important to know about) and RDA (Resource Description and Access, the new cataloging rules taking over gradually)..

In addition, it includes an **outline of the MARC format** for study

The **examples** are useful for

Lecture 4.2. Data schemas and formats,

Lectures 7.1-7.2. Cataloging and metadata. Bibliographic control: description, entries, and access and
Assignment 8. Descriptive cataloging practice

Lectures 10.2 - 11.2 Analysis and use of DDC, Yahoo, and LCC
and
Assignments 13.1-4 Subject cataloging and searching practice.

red

red

Weeks 1 - 2. M January 25 - W February 4

Part 1.
Foundations. Knowledge and knowledge representation.

red

Things to do in Week 1, M January 25 - W January 27

	Assignments due W Jan. 27. ☐ required ○ optional	√
Review answer key(s)	None	
Assignment(s)	Personal introduction: Introduce yourself to the class	☐

	New topics this week

	1.1a General introduction to the course	
Readings	1 Objectives etc. (pink) Also have a look at Assignments 1-3.	☐
	2 Textbook Ch. 1 **Information Systems for Problem Solving**	☐
	3 Berners-Lee, Tim; Hendler, James; Lassila, Ora 2001 **The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities**. Scientific American. 2001 May. Famous, still visionary. How course concepts apply in the future.	☐
	4 Coyle, Karen. 2012 **A New World of Data. Populating the Semantic Web**. Am. Libraries; 2012 July, p. 21. Library cataloging data as an integral part of linked data on the Semantic Web. Importance of entity-relationship modeling for linked data.	☐
	5 Gardner, Eileen 2013 **Application of UBLIS571 Course Concepts to School Librarianship**. UBLIS 571 Spring 2013. Optional, **For LMS Required**	○ ☐
	6 Soergel, D. 2004. **Information retrieval**. Berkshire Encycl.on Human-Computer Interaction.	○
	7 Soergel, Dagobert. **Information organization**. Berkshire Encycl.on HCI. 2004. 6 and 7 are **optional** overview articles directed at a general audience; first introduction to many course concepts. If you have trouble early in the course, come back to these.	○
Lecture	Read the general introduction materials and the syllabus (course cover page, Quick Start guide, etc. at the very beginning of the course packet). Includes **Lecture 1.1a read text and/or listen to the audio from the slides** (10 min)	☐

	1.1b Overview of the course and course materials	
Lecture	**Lecture 1.1b slides** (25 min) (Optional, done as part of the General Introduction)	○

	1.1c Information Professionals in the 21st century	
Readings	1 Special Libraries Association. **Special Librarians Putting Knowledge to Work. Competencies for Information Professionals of the 21st Century** (2014 draft).	☐
	2 U.S. Department of Labor. Bureau of Labor Statistics **Occupational Outlook Handbook. Librarians** (a quite traditional view, limited).	○
Lecture	**Lecture 1.1c slides** (25 min)	☐

over

	1.2 Information systems and information structure.	
Readings	Look over Assignments 1 - 3 so you can ask any questions you may have. 1 Lecture 1.2 Objectives etc. (pink sheet). .	☐
Lecture	**Lecture slides** (75 min)	☐

Learning blog	Learning blog Week 1 due Jan. 27	○

	Assignments assigned W Jan. 27
Assignments assigned	‣**Assignment 1, Hypermedia exploration: Perseus and Freebase** (2.5 hours) (due `Feb. 10) ‣**Assignment 2, Bibliographic retrieval system exploration: MEDLINE** (3 hrs) (due `Feb. 10) ‣**Assignment 3, Online catalog search exercise** (1.5 hours) (due `Feb. 10)

pink

Part 1.

January 27 - February 3

Foundations. Knowledge and knowledge representation

Lectures 1.1a-c

January 27

1.1a. General introduction to the course done in introduction

1.1b. Overview of the course and course materials done introduction

1.1c. Information professionals in the 21st century

Learning objectives	1 Have a sense of the pervasive applications of organization of knowledge concepts and principles. (P2.3.0,1) 2 Have an overview of the course and know what to expect and what is expected of you. 3 Have an **appreciation for** 3.1 the **wide variety of information tasks** that an education in information studies enables you to undertake and 3.2 the **wide variety of information environments** you can work in. Put differently, you should gain a sense of the breadth of functions and the breadth of environments you can work in, the breadth of careers. (P4.1) 4 Have an **appreciation for the wide variety of information systems** that exist, including expert systems. (P2.0.2) 4.1 Be able to characterize an information system or a job in an information system using the facets given in Lecture 1.1b. (P2.0.2,1) 5 Have an idea of Organization of Knowledge concepts and skills that are needed in practice. (P2.3.0,2)
Practical significance	For career planning it is important to understand the wide array of systems and jobs in which knowledge organization and skills can be used

pink

Lecture 1.1c. Information professionals in the 21st century (25 min)

What do information professionals do? p. ~13

See sample CVs in Supplement SLecture 1.1c

Types of information systems p. ~14 - 15 see also supplement SLecture 1.2.2

On learning, teaching, and the organization of information p. ~16

Salaries of information professionals See supplement SLecture 1.1c

Lecture 1.1c. Information professionals in the 21st century (25 min)

What do information professionals do?

<table>
<tr><td>Answer questions, find things</td><td>

Explore the information need with the user:
Understand the user’s problem
understand what the user knows already,
understand how the user thinks
Find answers in external and internal sources, such as
Library catalogs, bookstore catalogs (mostly online now),
Reference tools (bibliographies, biographical tools, almanacs, encyclopedias, etc.), print or online
Substantive databases (numeric, such as census, stocks, genomic db)
Maps
The Web at large, intranets
Archives (find records on a subject even though they are not indexed by subject)
A repository of instructional materials
Make a report that draws on several sources (extensive example: Congressional Research Service reports for Congress)
Organize the answer for quick perusal
</td></tr>
<tr><td>Organize things so they can be found</td><td>

Catalog books using the MAchine-Readable Cataloging (MARC) format and the Anglo-American Cataloging Rules (AARC2-2002) or soon RDA
Catalog Web pages using Dublin Core
Catalog learning materials using educational metadata standards
Format documents using XML
Write abstracts for and index journal articles
</td></tr>
<tr><td>Help people produce information</td><td>

Assist in editing and formatting documents
Help teachers in creating lesson plans (find instructional materials or learning objects, help format the lesson plan, help format materials for students, for example using graphic organizers)
Create Web pages (for the library or school media center on the organization)
</td></tr>
<tr><td>Teach</td><td>Teach people

how to find information. Requires teaching them about information organization
how to assess and evaluate information
how to use and integrate information
how to present information
</td></tr>
<tr><td>Develop and set up systems for all of the above</td><td>

Set up bibliographic and other databases, including library catalogs
Set up an intranet or a more ambitious enterprise portal that supports the work of all people in the organization
Set up document templates for easy creation of documents
Develop classification schemes, thesauri, taxonomies for special user groups (Each US agency must have a taxonomy to present its material to the public)
Help users with setting up their own personal information systems
</td></tr>
</table>

Types of information systems and information environments

Information systems can be classified along many dimensions or facets. Any specific information system can be characterized by a combination, one concept from each dimension, for example

A system

- dealing with documents (text and images), loosely structured information
- using plain retrieval
- dealing with published or semi-published information
- serving a government agency
- information used for research and patient care
- dealing with the medical domain
- using paper technology for storage and accessing digital information

= a traditional medical government library

Sample dimensions (facets) for characterizing information systems
Types of information (such as bibliographic data, text and images, multimedia, numerical and other primary data, organization data and records); **Degree of structure of the information** (unstructured or loosely structured information as in text vs. tightly structured information as in numeric databases) **Processing to create answers: plain retrieval vs. drawing conclusions** **Origin of information** (such as generally published information - paper or online, government information, organizational information, information. about customers or patients); **internal vs. external information** **Users of the information, audience or organization served** (groups - such as children, farmers, scholars, urban communities - or organizations - such as schools; universities or colleges; government agencies; businesses); **Uses of information** (such as research, learning, problem solving, decision making, patient care, collaboration, day-to-day transactions); **Subject field** (such as physics, medicine, or anthropology); **Technical means of providing access** (such as paper vs digital). The combinations are many, illustrating the flexibility and diversity of the information field

The table on the next page gives some examples for these dimensions. The information systems listed have characteristics for the other dimensions too but we do not list these.

Sample systems illustrating selected dimensions

Information & processing	**Expert systems** (medical diagnosis, computer configuration, detecting mineral deposits from satellite images, loan approval, etc.) **Software libraries** / databases of computer programs / modules for ease of access and reuse **Employment service databases** **Personnel information system** (usual personnel data plus skills and assignments to manage an organization's workforce) **Geographical information system (GIS)**
Users	**Information systems in organizations** Knowledge management: Make sure all applicable information is used to best advantage by organizing all types of internal and external sources of information – paper files and computer files no matter who keeps them, people and the knowledge they have in their heads – for access and usability. Information resources management Management information systems (MIS), Decision support systems Day-to-day transaction systems (order, inventory, etc.) Records management, archives (especially electronic records) A **personal information system** managing Web bookmarks, bibliographic references, downloaded Web pages, computer files, paper documents in personal collection, all kinds of notes, addresses, appointments
Use	**Instructional information systems** matching learner's needs with instructional materials In formal educational institutions In organizations for training (this is big business! Coordinate with personnel information system) For both: long-distance learning
Technology	**Paper libraries** of all kinds (public, academic and school, special, personal) **Online information systems** Digital libraries Intranets An organizations Web site Any kind of **computer database** Bibliographic databases (e.g., Medline or OCLC's WorldCat) OCLC = Online Computer Library Center, the world's largest cooperative cataloging agency) Full-text databases (e.g., Westlaw or Lexis for law) Multimedia databases. Problem of retrieving still and moving images Substantive databases (directories, statistical data, material properties data, genomic data, data on chemical substances, electronic health records (EHR))

On learning, teaching, and the organization of information
Helping students to learn how to organize information, to learn how to learn

Implications for curriculum and instruction

"The findings from this study suggest that today's students need to learn – in a way that transcends their learning of specific content – a good deal about the structure of knowledge and about the importance of that structure. In order to learn in an information-rich environment, they need to learn

(1) that knowledge is indeed structured in meaningful ways;
(2) that various structures can be applied to various kinds of knowledge; and
(3) that a key part of learning is learning how to create personal structures that organize their own learning accurately and coherently.

They must learn that knowledge is an organized, systematically related set of ideas and that they need to work at building an understanding of that organization as well as learning the individual ideas. They must learn the nature and uses of various kinds of structures–for example, time lines, maps, and hyperlinks as well as traditional narrative structures–that they can use as tools for building their own knowledge. They must learn criteria and procedures for building appropriate and coherent structures that will allow them to integrate and communicate their thoughts. A curricular emphasis on teaching students how to structure information is, I believe, the most important implication for learning and teaching that stems from the presence of the information-rich environment in which we and our students live.

Learning theorists tell us that learning consists of constructing mental models or schemas, structures that are comprised of ideas and patterns or frameworks that organize and link those ideas. At some basic level, then, learning is the equivalent of organizing information. And no one in a school knows more about organizing information than the library media specialist best. Helping both teachers and students understand and learn to create a variety of ways to structure information is the key task for our profession in these best and worst of times."

From

Delia Neuman
Learning in an information-rich environment: Preliminary results
Treasure Mountain/Elms Research Retreat
Elms Resort and Spa
Excelsior Springs, MO
May 31, 2002

Book
Neuman, Delia
Learning in Information-Rich Environments: I-LEARN and the Construction of Knowledge in the 21st Century.
Springer; 2011. 156 p. ISBN-10: 1441905782 ISBN-13: 978-1441905789

Delia Neuman is director of the LMS program at Drexel University.

pink Part 1. Foundations. Knowledge and knowledge representation, cont.

Lecture 1.2 ***January 27***

Information systems and information structure. Expert systems (75 min)

Learning objectives	1 Gain an **appreciation for the variety of information systems** that exist, including knowledge-based systems / expert systems. (P2.0.2) 2 Appreciate how a **system that can draw logical inferences (reason) can save a user considerable time** and improve problem solving and decision making. (P2.3.0,3) 3 **Understand the importance of information structure** / knowledge representation as the heart of an information system. (P2.3.1.0) 3.1 Understand how **data need to be structured to allow automatic reasoning.** (P2.3.1,0.1) 3.2 **Be able to combine different kinds of data (conduct chained searches) to find an answer**. (P2.3.2,1) 4 Have a **first idea of the entity-relationship approach** to knowledge representation. (P2.3.1,1.1)
Practical significance	Being knowledgeable about databases is a requirement for every executive assistant, let alone information professionals. Databases are the key to dealing efficiently with many types of information. Knowing about many types of information systems makes your skills more widely applicable and thus increases career opportunities. Expert systems are now widely used in many subject areas, such as medicine, computer system configuration, and processing of loan applications; see the list of examples at the end of Lecture 1.2. Designing or understanding the information structure of a system is key to building or using the system. The *entity-relationship (E-R) approach* is the most natural and at the same time most general way for representing information. The E-R approach is used in structuring and querying *linked data* that make up the *Semantic Web*.

Note on terminology: The Artificial Intelligence (AI) community speaks about *knowledge representation*, the database community speaks about *data modeling*.

Key idea:

Combining different kinds of data to find an answer. (Inference. Chaining)

Done by people - reference librarian or user consults different databases as needed to find all the facts needed to construct an answer.

Done by systems - all facts must be accessible to the system

Relates to ➜UBLIS 518 Reference Sources and Services (concepts important for how to search)

pink

Introduction

Purpose of an information system generally: **Answer questions** by either finding an answer that exists ready-made in the database or deducing an answer from multiple statements in the database. Answering a question means **going from something known to something unknown**.

The lecture will show through examples **how information structure is used to find answers**.

We will look at three examples of information systems:

Lect. 1.2a	An expert system for medical prescriptions
Lect. 1.2b	A database that supports the operation of a university (Organizing Info., Chapter 3)
Lect. 1.2c	Medline, a bibliographic information system (Supplement SLecture 1.2c)

****

Lecture 1.2a. An expert system for medical prescriptions

Purpose	From the data in the patient record, including new diagnoses, and medical knowledge, find drugs the patient should take.
Questions	1 What new disease does patient Fred have? *Known:* Patient Fred ➜ *unknown:* ?Disease 2 What drugs are used to treat asthma? *Known:* Disease asthma ➜ *unknown:* ?Drug 3 What drugs should patient Fred take? *Known:* Patient Fred ➜ *unknown:* ?Drug

P. ~22 shows examples of data that are needed to come up with a prescription for a patient.
A physician deciding on a prescription combines these data through a reasoning process.

To prescribe a drug, a physician proceeds as follows

A Find the disease newly added the patient record (patient data, specific to each patient).

B Find out what drugs are available to treat that disease (drug treatment data, med. textbook)

C Check for each of these drugs what, if any, harmful effects they may have on the patient.

In most medical environments, there are three simple databases each of which
• contains one kind of facts and • can be queried to retrieve facts as they are stored in the database.

The physician is left with the task of combining these facts to decide what drug to prescribe.

We will show these databases, and then develop an **expert system** that has access to all three kinds of facts and can combine them to deduce a suggestion for a drug to be prescribed, saving the physician time.

Note: Customarily, the term "fact" is used in this context, but "statement" or "assertion" would be better. "Fact" implies that an assertion is true, but not all so-called "facts" in a database are true.

Read p. ~23-24 (but do not get stuck), then look at / listen to the slides for Lecture 1.2a.

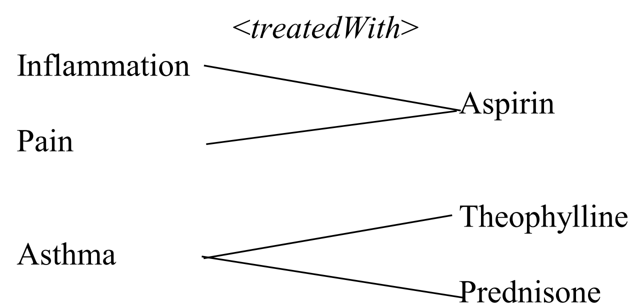
A Patient facts

Patient facts	A1 Fred <*hasDisease*> LiverDisease A2 Fred <*hasDisease*> Asthma (New disease, basis for question) A3 Phil <*hasDisease*> Inflammation

B Treatment facts

Question	What drugs are used to treat asthma? Asthma <*treatedWith*> ?DrugX
Treatment facts	B1 Inflammation <*treatedWith*> Aspirin B2 Pain <*treatedWith*> Aspirin B3 Asthma <*treatedWith*> Prednisone B4 Asthma <*treatedWith*> Theophylline
Answer	Theophylline, Prednisone

Graphical representation of facts (same as an RDF graph)



C Contraindication facts (one consideration in harmful side effects)

Contra-indication facts	C1 Aspirin <*contraIndicatedWith*> PepticUlcers C2 Theophylline <*contraIndicatedWith*> PepticUlcers C3 Theophylline <*contraIndicatedWith*> Arrhythmia C5 Prednisone <*contraIndicatedWith*> LiverDisease

The physician uses contraindication facts to filter out drugs that treat the disease in question but would do harm to the patient and retain only drugs the patient tolerates.

The general rules for the drug prescription reasoning process are given on p. ~21 -22. These rules must be applied to a specific patient, in the example Fred. How this is done is explained step by step in class and in the slides for Lecture 1.2a. Read p. ~23-24

Developing a system that can deduce answers by combining facts

Saves the physicians time. We need to formulate rules on how to combine facts in a language a computer can understand). Here we use *Prolog* (Programming in logic). Lecture 5.1 introduces SPARQL, a query language for linked data in the Semantic Web

Overall strategy: Divide the problem into two sub-problems (as the physician does)

Question	What drug(s) should Fred take? Fred <*shouldTake*> ?DrugZ
Rule	?PersonX <*shouldTake*> ?DrugZ IF ?PersonX <*hasCandidateDrug*> ?DrugZ AND ?PersonX <*tolerates*> ?DrugZ /* filter for: drug does no harm */

First sub-problem. Combine patient facts with treatment facts to find candidate drugs

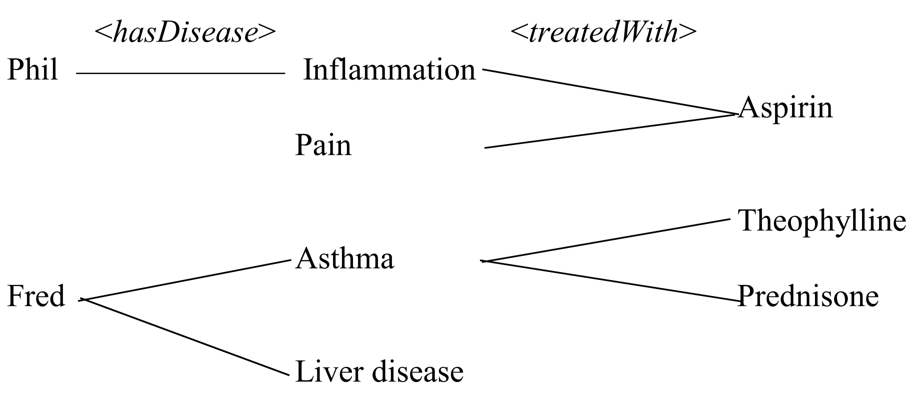
Question	What are candidate drugs for Fred Fred <*hasCandidateDrug*> ?DrugZ
Rule	?PersonX <*hasCandidateDrug*> ?DrugZ IF ?PersonX <*hasDisease*> ?DiseaseY AND /* from patient facts, A */ ?DiseaseY <*treatedWith*> ?DrugZ /* from treatment facts, B */
Answer	Theophylline, Prednisone

This rule does a chained search, as further illustrated in the graph:

1. The first condition starts from a person and finds a disease.
2. The second condition starts from the disease found and finds candidate drugs

.

Graphical representation of facts. Chained search



Note: We could build a drug prescription expert system that considers just the newest patient disease to be treated and treatment facts. But such a system would not be good for patients; **more knowledge is needed to avoid drugs that would harm the patient**.

Second subproblem: Combine patient facts with contraindication facts to retain only drugs the patient tolerates.

For example, Prednisone must not be taken by a person with liver disease, that is:

Prednisone <*contraIndicatedWith*> Liver disease

Question	What drugs does Fred tolerate. Fred <*tolerates*> Drug Z In the example, we need to apply this check to Theophylline and Prednisone Answer is Yes / No (True / False)
Rules	?PersonX <*tolerates*> ?DrugZ IF ?DrugZ <*contraIndicatedWith*> ?DiseaseW AND ?PersonX <*freeOf*> ?DiseaseW ?PersonX <*freeOf*> ?DiseaseW IF NOT (?PersonX <*hasDisease*> ?DiseaseW) ?PersonX <*tolerates*> ?DrugZ IF NOT (?DrugZ <*contraIndicatedWith*> any ?DiseaseW) /* If a drug has no contraindications at all, it can be prescribed without checking the patients diseases. */
Answer	Theophylline

Note: If a drug is not ruled out based on the diseases seen from the patient record, a good system would **alert the physician to all contraindications** so that the patient can be checked out for these conditions. Omitted in our simple example.

*** Look at the slides for Lecture 1.2a now.

Further refinements	• drug effectiveness • drug side effects and their severity • drug interactions and incompatibilities • drug cost A system containing all these data for a large number of drugs can prescribe as well as a human expert and would be called an **expert system.**

Some additional types of data used:

Drug effectiveness data (effectiveness may depend on several factors):

Disease <*treatedWith*> (Drug, Effectiveness, Ethnicity, Gender, Age)

Drug incompatibility data (bad effects that can happen if a person takes two drugs)

Drug <*incompatibleWith*> (Drug, Effect)

Note that relationships are more complex (they relate more entities) to accurately reflect reality.

Lecture 1.2b. **A database that supports the operation of a university**

Discussion of the example presented in Textbook, Chapter 3

Partial conceptual schema and some illustrative data for a university database

Figures 3.2a and 3.3a are from the Textbook, Chapter 3, and embellished here.

See next page

We will discuss in class how these interconnected data can be used to find answers by combining different kinds of data ("facts"); in the graphical representation this can be visualized as "chaining".
For online / just reading : Use p. ~26 - 27. Take these pages out of the binder so you can look at the figures at the same time.

Key idea

Chained search

Combining different kinds of data to find an answer. (Inference. Chaining)

Here you will learn about conceptual data schemas. A conceptual data schema is, as the name says, a definition of the data in a database at the conceptual or logical level. To implement a database following a conceptual data schema one uses a database management system (DBMS), such as Microsoft Access, FileMaker Pro, ORACLE, or an RDF triple store. This is a topic in UBLIS 506 *Information Technology*.

See supplement SAssignment 6a for an explanation and exercise of how to implement the University Database in MS Access.

In supplement, SLecture 1.2d

Types of information systems from simple to complex (and more useful)

Characteristics of a good information system

Advanced ideas to ponder

Expert system examples

Who were the students in the Fall 1979 offering of the course Vegetable Pickling?
Entity types involved: *Course offering*
Person
Relationship: ⟨has student⟩

Who taught the Fall 1979 offering of the course Vegetable Pickling?
Entity types involved: *Course offering*
Person
Relationship: ⟨has instructor⟩

What are the prerequisites of the course Vegetable Pickling?
Entity type involved: *Course*
Relationship: ⟨has prerequisite⟩

What grade did J. Doe get in Vegetable Pickling in Fall 1979?
Entity types involved: *Course offering*
Person
Grade

However, the situation in this example is more complex: a grade does not pertain to a person alone nor to a course offering alone but rather to a person–course offering combination. We consider such a combination as a new, composite entity. Thus, we revise the entity types as follows:

Revised entity types: (*Course offering* ⟨has student⟩ *Person*)
Grade
Relationship: ⟨has attribute⟩
Answer contains: A list of grade values (one value in the example)
Selection criteria: Grade value is assigned to the combination entity (Fall 1979 offering of Vegetable Pickling ⟨has student⟩ J. Doe)

Fig. 3.1 Sample search topics and their analysis

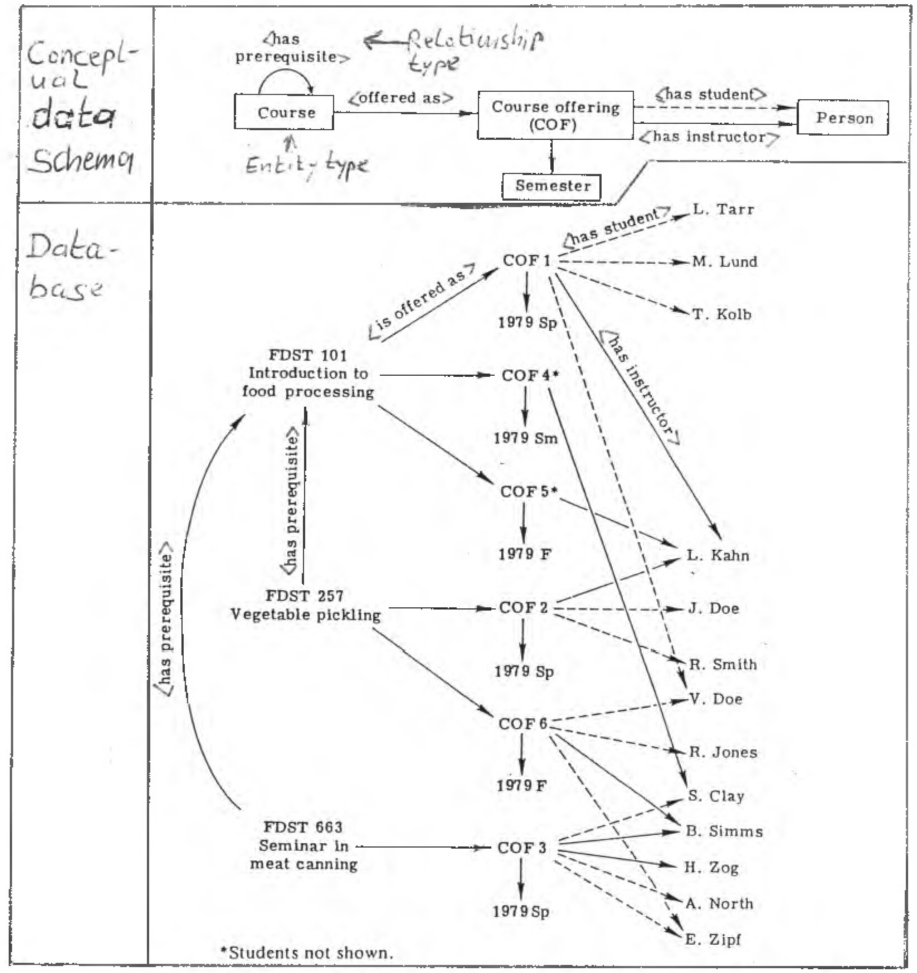


Fig. 3.2a Structure of a University Data Base. Basic example. Graphic representation

Fig. 3.2b Structure of a University Data Base. Basic example. Tabular representation

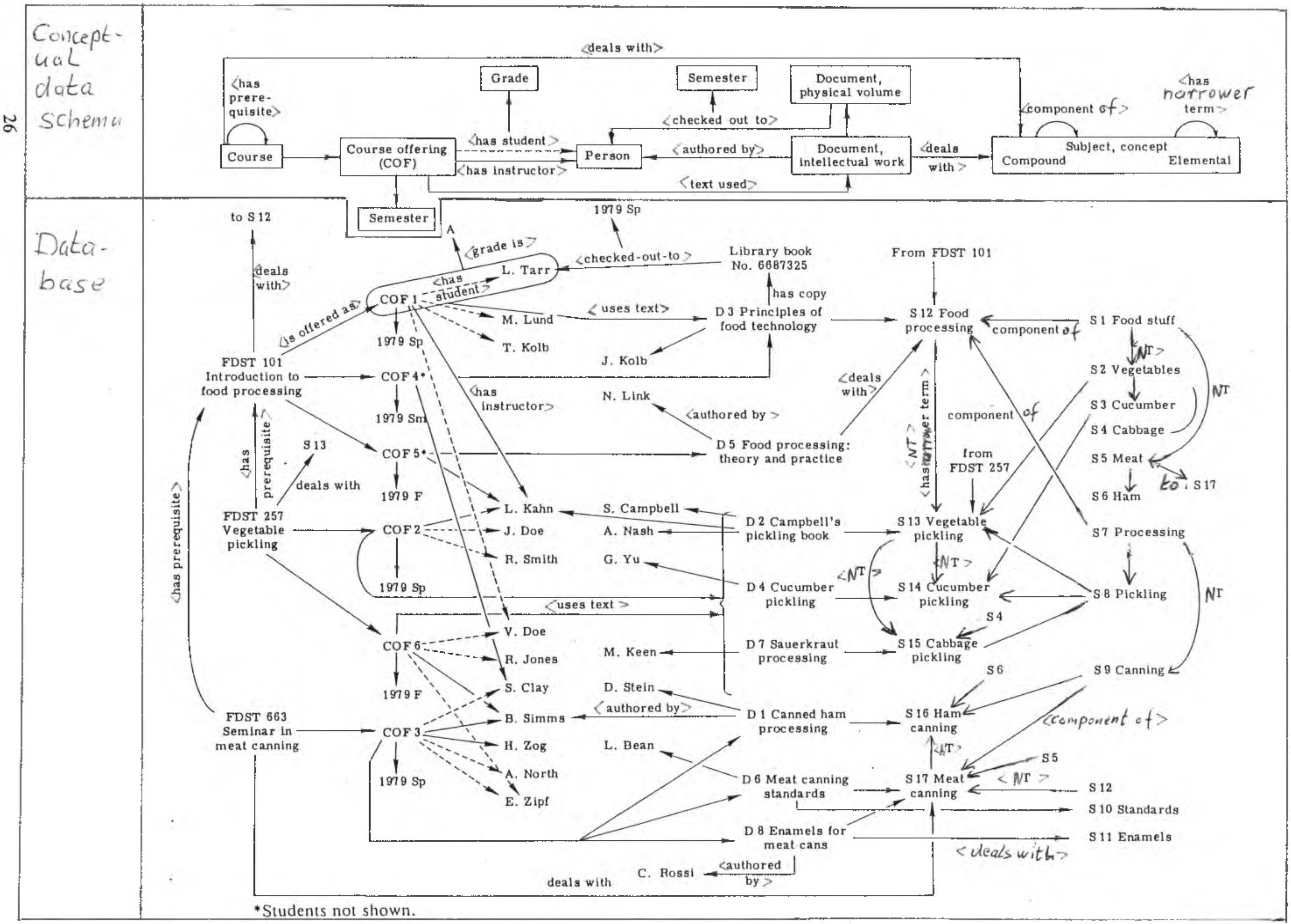


Fig. 3.3a Structure of a University Data Base. Expanded example. Graphic representation

Text to go with Figures 3.2a and 3.3a.

Figure 3.2a (p ~26) shows how the structure of a database is derived from expected questions. This is the key principle of user orientation that goes through the course as a red thread.

Read Textbook Sections 3.1 and 3.2 for an explanation of entity types and relationship types; together they define what kinds of data can be stored in a database, they constitute the conceptual data schema.

The database structure needs to support searching, so we will discuss some queries, from very simple to more complex. First try to find the answers yourself, then look at the answers starting on p. ~31.

In the small network diagram (Fig. 3.2a)

Q1 Find all persons who are students in course offering COF1.

Q2 Find all persons who at one time or other were students in some offering of FDST 101.
Hints: You need to follow a two-step chain from FDST 101 to the persons.
Write out this chain (1) generally (using entity types) (2) with specific entity values.

In the large network diagram (Fig. 3.3a, p. ~27**)**

Q3 Find all documents on S12 Food processing.
Note: All arrows can be followed in both directions. The direction of the arrow in the diagram depends on the label. If the label was *coveredIn*, the arrow would point from the Subject to the Document.

Q3a **Find all books on food processing and any of its narrower concepts**, such as vegetable pickling.

Explore the network structure for searching (chained searching)

Q4 Assume that there are no relationships Document <*dealsWith*> Subject; in other words, this university library has no subject access in its catalog.
Query: Find all documents on S12 Food processing.
Hint: Find a path (chain) trough the network that starts with a Subject and ends with a Document.

Q5 Find all course offerings where an instructor is also an author.
Hint: Starting from a course offering, find the instructor(s).
Starting from the same course offering, find the author(s) of its textbook(s).
If you come to the same person both ways, the course offering meets the condition.

Answers to the queries

In the small network diagram (Fig. 3.2a)

Q1 Find all persons who are students in course offering COF1.

Answer. Start from COF1, follow the <*hasStudent*> link and find
L. Tarr, M. Lund, and T. Kolb
This is a one-step search using only one type of data, namely which persons are enrolled in a given course offering.

Q2 Find all persons who at one time or other were students in some offering of FDST 101.
Hints: You need to follow a two-step chain from FDST 101 to the persons.
Write out this chain (1) generally (using entity types) (2) with specific entity values.

Answer. Start from FDST 101,follow the <*isOfferedAs*> link to COF01, from there follow the <*hasStudent*> link and find
L. Tarr, M. Lund, and T. Kolb.
Repeat from FDST101 to COF04 and then CO05 (students are omitted from graph). This is a two-step search (chained search) using two types of data, namely
(1) what are the course offerings for a given course and
(2) which persons are enrolled in a given course offering.
This is a very simple inference.

In the large network diagram (Fig. 3.3a)

Q3 Find all documents on S12 Food processing
Note: All arrows can be followed in both directions. The direction of the arrow in the diagram depends on the label. If the label was *coveredIn*, the arrow would point from the Subject to the Document.

Answer. Start from S12, follow <*dealsWith*> backwards, find books D3 and D5; one-step search using only subject indexing data for books.

Question: Just starting from S12, do you find all documents on Food processing? Or should D2 Campbell's Pickling Book be found as well?
If the user needs documents that deal comprehensively with all aspects of Food processing, such as textbooks or handbooks, then starting just from S12 is fine; this is called a "general references" search. But if the user needs every document on every subtopic of Food processing, then the system needs to (1) find these subtopics using information about relationships between subjects/concepts as found in a classification scheme or thesaurus and (2) from each subtopic find the applicable documents using information about relationships between subjects and documents as found in a catalog; this is called an "inclusive search".
Ass. 2 Medline has another example of general references search and inclusive search.

Q3a How can the system find all books on food processing and any of its narrower concepts, such as vegetable pickling?

Answer. Start from S12, follow <**dealsWith**> backwards, find books D3 and D5; one-step search using only subject indexing data for books. Then, to find more books, from S12

follow the <*hasNarrowerTerm*> link, find S13, and from there, following <*dealsWith*> backwards, find D2. Repeat until no more narrower terms are found.
As pointed out above, this is a two-step search (chained search) using two types of data: data about concept relationships (found in a thesaurus) and subject indexing data for books (found in the library catalog).

Explore the network structure for searching (chained searching)

Q4 Assume that there are no relationships Document <*dealsWith*> Subject; in other words, this university library has no subject access in its catalog.
Query: Find all documents on S12 Food processing
Hint: Find a path (chain) trough the network that starts with a Subject and ends with a Document.

Answer. Start from S12, follow <*dealsWith*> backwards to FDST 101, from there follow <*isOfferedAs*> to COF01 and then <*usesText*> to document D3. Repeat through COF04, leads again to D3. Repeat through COF05, leads to D5.
This is a three-step chained search using three different types of data.

Q5 Find all course offerings where an instructor is also an author.
Hint: Starting from a course offering, find the instructor(s).
Starting from the same course offering, find the author(s) of its textbook(s).
If you come to the same person both ways, the course offering meets the condition.

Answer.
Start from COF01 and find the instructor, L. Kahn.

Start from COF01 again and find the textbook, D3.

From D3 find the author J. Kolb.

The instructor and the textbook author different, COF01 is not an answer to the query.

Try the next course offering:

Start from COF02 and find the instructor, L. Kahn.

Start from COF02 again and find the textbook, D2.

From D2 find the authors S. Campbell, L. Kahn (could stop here), and A. Nash.

Found a textbook author who is the same person as the instructor, COF02 is an answer.

Note: If Kahn had not been an author of D2, you could still try to go through the other two textbooks,D4 and D7. Once you find a match you can stop. If you worked through all possibilities and did not find a match, the course offering is not an answer.

This is an example of how flexible this database structure is. Nobody thought about this question when designing the database. But the "atomic" facts in the database can be combined in many different ways to answer even unexpected questions (not all of them, of course).

red

Things to do in Week 2, January 27 - February 3

	Assignments due W Feb. 3 ☐ **required** ○ **optional**	√
Review answer key(s)	None	
Assignment(s)	None	

	New topics this week	

	2.1 Nature of Knowledge	
Readings	1 **Lecture 2.1 Objectives etc**. (pink). Also have a look at Assignment 4	☐
	2 Textbook Ch. 2 **The Nature of Information**	☐
	3 Skemp, Richard R. *The psychology of learning mathematics*. Expanded Amer. Ed. (Also: *2. ed. 1986,* 1.ed. 1971, page numbers vary) Chapter 2 **The formation of mathematical concepts**, p. 9 -21 **Required** Chapter 3 **The idea of a schema**, only p. 22-29 **Required** Chapter 5 **Symbols**, p. 46-55 **Required** Excellent general introduction to the structure of knowledge and its representation and the nature of concepts, applicable to any subject, not just math. Read carefully.	☐
Lecture	**Read the lecture notes** like a book chapter - no slides	☐

	2.2 Knowledge representation		
Readings	1	**Lecture2.2 Objectives etc**. (pink)	☐
	2	Textbook Section 9.3 **Criteria for the design and evaluation of data schemas** (p. 150-152)	☐
	3	Lindsay & Norman. *Human information processing. Intro to psych.* NY: Ac. Pr., 1972. Chapter 10. **The structure of memory (semantic networks, DS),** p. 374-401 Chapter 11. **Memory processes [restructuring semantic networks]**, p. 402-434	○
	4	Jonassen, David H.; Beissner, Katherine; Yacci, Michael. Structural knowledge: Techniques for representing, conveying and acquiring structural knowledge. Hillsdale, NJ: Lawrence Erlbaum, 1993. Ch. 12. **Implicit methods for conveying structural knowledge through frames and slots,** p. 125-133.	☐

over

	5	Parsaye, Kamran; Chignall, Mark. *Expert systems for experts*. New York: John Wiley and Sons, 1988. Section 2.2.3. **Frames: Packaged Structures**, p. 48-57	○
	6	Fikes, Richard and Kehler, Tom. **The role of frame-based representation in reasoning**. Communications of the ACM. 1985; 28(9): 904-920.	○
Lecture	**Read the lecture notes** like a book chapter - no slides		□
Learning blog	Learning blog Week 2 due February 3		○

	Assignments assigned W. February 3	
Assignments assigned	‣**Assignment 4,**	**Restructuring two sets of data using hierarchical inheritance** (1.5 hours) (Due `Feb. 10)

pink **Lecture 2.1** ***February 3***

The nature of knowledge

Learning objectives	1 **Understand the characteristics and facets of different types of knowledge** (P1.1.1) Be able to apply this understanding to 1.1 the understanding and analysis of information needs (P2.1+) 1.2 the organization of information (P2.3) 1.3 the assessment and evaluation of information in all forms. (P2.2.2,2.3; P2.5.2,3) 1.4 the design of learning experiences (P2.6.1,2.1) 2 **Understand findings from cognitive psychology on the way people form and deal with concepts** (P1.1.2) Be able to apply these findings to 2.1 the understanding and analysis of information needs (P2.1+), 2.2 the design of classifications and search support (P2.3), 2.3 information presentation (P2.3.1,3; P2.5.2,5), 2.4 designing learning experiences (P2.6.1,2.1) 3 Understand the pervasive role of classification throughout the human endeavor. (P2.3.9,1)
Practical significance	**Knowing about types of knowledge is important for** • understanding and analyzing information needs (as in interviewing a library patron before doing a search – reference interview); • analyzing and assessing information found; • determining how to organize and process information /knowledge in accordance with its type; • matching documents to the needs of the patron according to the type of information they contain. **Knowing about types of concepts is important for** • understanding how people think and, therefore, how they ask questions, how they determine relevance, and how they process information. • The answers, in turn, determine how information should be retrieved (retrieval should approximate human relevance judgment or improve on it) and what information should be presented to a user in what form.

pink

Outline

1 Types of knowledge: characteristics/facets/dimensions

1.1 Types of knowledge by content

1.2 Types of knowledge by scope of applicability

1.2.1 Knowledge about regularities (laws, rules) vs.
knowledge about individual detail from which the regularities can be derived

1.2.2 Scope of applicability to natural or social phenomena.
Scope of validity of a statement in space and time.

1.3 Types of knowledge by degree of "vagueness" of knowledge

1.4 Types of knowledge, other aspects

2 The nature of concepts / categories

2.1 Types of concepts. Individual concepts and class concepts

2.1.1 Individual concepts – individual entities. Persistence over time

2.1.2 Class concepts / categories. Simplified account

2.1.3 Mass concepts (oil, flour, sugar) (how much?)
vs. count concepts (sugar cubes, books) (how many?)

2.1.4 Abstract concepts (freedom, justice). Can define the concrete class of all countries in which freedom prevails.

2.2 Objectivist vs. organism-centered view of categories

2.3 Explicit definition of categories vs. prototypes and fuzzy membership.
Radial categories

2.4 Basic level categories (Eleanor Rosch)

Preliminaries

The notes for Lecture 2.1 are comprehensive, including elaboration and examples previously discussed only in the seated class. Thus the notes are like a book chapter, so there is no need for an audio recording.

The general concepts are illustrated with examples from various subject domains. A certain amount of general education is assumed. If you do not know about Kepler's laws of planetary motion (in general, no need to know the actual formulas) or the gravitational constant, look it up on Google. After all, you are studying to be information specialists who should follow the motto "if you do not know it, look it up".

The questions are often rhetorical in that the answer is obvious and does not need to be given. Nevertheless, they encourage you to think.

1 Types of knowledge: characteristics/facets/dimensions

➜UBLIS 518 Reference Sources and Services: Selection of reference tools

Think about why we talk about different types of knowledge. How can understanding types of knowledge help you

- design databases?
- answer a reference question?

Think about this for all the different types of knowledge discussed

1.1 Types of knowledge by content

1.1.1 Definitional knowledge (dictionary) vs. **assertive knowledge** (encyclopedia, world almanac). Essential vs accidental attributes

- Essential attribute: An action must involve forcible entry to fit the definition of burglary.
- Accidental attribute: The action took place at 2 am.

(These are relative distinctions, see discussion of concepts in Section 2.1 below.)

In principle, **definitions are not true or false**, right or wrong; they may be useful or not useful. Information about the use of a word or phrase or other sign or symbol in a given community and context can be true or false.

Assertions (statements about the real world or an imaginary world), on the other hand, **can be true or false**.

1.1.2 Knowledge about **static relationships** ("what is the area of Nigeria") vs. knowledge about **events and actions** ("what has happened in the religious strife in Nigeria")
More generally: **declarative** (what is) **vs. procedural** (how to) **knowledge**

Information about static relationships is easier to organize and to find than information about events or actions:

- To find the area of Nigeria, just look in the World Almanac or a similar source.
- Where would you find information about *the religious strife in Nigeria over the past 10 years*? How likely is it that you could all information needed in one place?

1.1.3 Knowledge by **subject area**, such as *physics* or *environmental safety* or *healthy behavior*.

1.1.4 Knowledge by **relationship type used in a statement** (by type of data), such as Disease <*treatedWith*> Drug or Person <*memberOf*> Organization

1.2 Types of knowledge by scope of applicability

The more widely applicable an item of knowledge, the more important to obtain it, validate it, and store it in an easily accessible form. There are several aspects or facets of scope.

1.2.1 Knowledge about regularities (general laws, rules) vs. knowledge about individual detail from which the regularities can be derived

Knowledge about regularities that apply to many cases or throughout a system (e.g. medical text-book knowledge) vs. knowledge about individual detail that applies only to the individual case (e.g., an individual patient). (Go back to Lecture 1.2 and see what data are general and what data apply only to individual persons.)

You should be able to get a sense of the distinction between **Regularity** and **Individual detail** from the many examples given below.

Examples:

Regularity:	Asthma <*treatedWith*> Theophylline	[Applies to all asthma patients]
Individual detail:	Fred <*hasDisease*> Asthma	[Applies just to one patient.]
Regularity:	FDST 257 <*hasPrerequisite*> FDST 101 FDST 257 <*isOfferedAs*> COF02	[Applies to all students in any section] [Applies to all students in this section]
Individual detail:	COF02 <*hasStudent*> R. Smith	[Applies to this student]
Regularity:	Hearing tests <*hasNarrowerTerm*> Audiometry	[Applies to all searches for these concepts.]
Individual detail:	Document 4 <*dealsWith*> Audiometry	[Applies to retrieval of just one document.]
Regularity:	Kepler's laws of planetary motion	[Applies to all planets at all times]
Individual detail:	The observational data about planet positions	[Each observation applies to the position of one planet at one time]
Regularity:	Burglary is punishable with 3 - 10 years of prison.	[Applies to all burglaries]
Individual detail:.	Weaver broke a large Window and entered the house. He took . . .	[Applies to this particular burglary]

To generalize from two of these examples.

Domain	Type of fact	Examples
Medical	**Regularities, general facts**	Facts about symptoms and diseases and treatments are broad; they apply to many persons (patients, cases).
	Individual detail facts	Facts about an individual patient are narrow; they apply only to one person (patient, case).
Subject access to documents	**Regularities, general facts**	Facts about concept relationships are broad; they apply to all searches for the concepts involved and affect the retrieval of all documents indexed by one of the concepts.
	Individual detail facts	Facts linking a document to a concept (indexing facts) are narrow; they affect only the retrieval of this one document.

Science is all about deriving regularities (scientific laws) that can be used to "recreate" individual detail by deduction. For example, Kepler's laws of planetary motion can be used to compute the position of a planet at any given time from the position of the planet at one time and two parameters that characterize the orbit of the planet around the sun. From three data values we can compute with the aid of Kepler's law thousands of observations about the position of the planet made by astronomers and reported in huge books (for example by the astronomer Tycho Brahe (1546 – 1601), an enormous savings of storage space. More importantly we can compute the position of the planet at any time, past or future. **Ethics** is about establishing and justifying general rules that can guide behavior in many situations. Likewise, **statutory law** and regulations state general rules to be applied/followed in many individual situations.

Ways to reason from past experience	
Regularities or laws are known	**Reasoning from general laws (deductive):** Draw conclusions on specific cases to which the laws can be matched.
Regularities or laws are <u>not</u> known	**Case-based reasoning (inductive)**: Find similar past cases and assume the new case will have similar outcomes. Examples: Weather forecast Decide a legal case where the law is inconclusive (precedents) Decide on patient treatment based on past experience with similar cases

Examples of **deductive reasoning**:

- A physician interpreting the patient's symptoms diagnoses the disease Asthma (as opposed to some other *disease* that causes difficulty in breathing, then consults a textbook to find that Asthma can be treated with Theophylline and deduces that the patient should take Theophylline.

- A judge determines that the facts of a case meet the definition of *burglary without aggravating circumstances* given in the law, finds that the law specifies a punishment of three years in prison and deduces that the accused should be so sentenced.

Examples of **case-based reasoning:**

- A physician remembers that a patient she encountered five years ago had a similar combination of symptoms as the patient before her responded well to a lactose-free diet and suggests a lactose-free diet to the patient before her
- A merchant brings a case against members of a flash mob that impeded business and caused a loss of $10,000. The judge cannot find a specific law but finds a similar case decided by another court that ruled that the flash map participants exercised their right to free expression and cannot be forced to make up the merchant's loss. So the judge issues the same ruling.
- Weather forecasters look through large amounts of weather data to find a configuration (temperature in different places and altitudes, wind strength and direction in different places, precipitation, etc.) to the configuration that exists now and predict the weather based on what happened in a similar situation in the past.

Questions:

Which kind of reasoning requires more data, deductive reasoning or case-based reasoning?

What kind of retrieval is needed for case-based reasoning?

Two important specific kinds of knowledge about regularities	
Type of knowledge	**Examples**
Knowledge about restrictions on data values	*A male individual of a mammal species cannot be pregnant.* *A two-year-old human cannot weigh more than 30 pounds.* Used for checking data for errors
Default knowledge	**Default** knowledge: *A car has **four** wheels.* Specific knowledge about an **individual case** / knowledge about exceptions: *The Runabout has **three** wheels.* Default knowledge used all the time in daily life. Default values in data entry forms

1.2.2 Scope of applicability to natural or social phenomena. Scope of validity of a statement in space and time

Regularities can differ in the scope in which they apply.

Examples

Narrow scope	Broad scope
The **natural law** on the **free fall of objects towards the earth** applies only on the earth. **Kepler's laws** apply only to the movement of objects moving in an orbit around another object (originally they were conceived as applying only to the movement of the planets around the sun).	The **general law of gravity** applies to many phenomena throughout the universe; many more specific laws, like the two mentioned, can be derived form it.
A **property value for a specific material** (such as the electrical conductivity of copper) applies only to phenomena involving that material.	The **gravitational constant** holds through the entire universe (or so physicists think) and is involved in many phenomena.
A **social rule**, custom, or etiquette rule **that applies only in one country**	**A social rule**, custom, or etiquette rule **that applies world-wide**
Many **rules of grammar apply to only one language family** (such as all Indo-European languages) or to only a single language.	Some linguists believe that some principles apply to all languages (**language universals**).

- From the standpoint of science, which is more important to measure precisely, the electrical conductivity of copper or the gravitational constant?
- A business traveler needs to know whether a social rule, custom, or etiquette rule that applies world-wide or only in his own country. Otherwise he may offend potential business partners in a foreign country and loose a deal

Related distinction	
Domain-specific knowledge Examples: Effects of a drug; How to teach math to fourth-graders	**Common sense knowledge** Examples: Use cost-benefit analysis General principles of management

Examples:

> A World Bank project aimed at improving schools by training education ministry staff in
>
> (1) general management procedures, especially procurement,
> (2) domain-specific (in World Bank speak: sector-specific) knowledge in education.
>
> The country staff found (1) just as useful or more useful because it could be transferred to other domains.

UBLIS program objectives make the following distinction:

2 **Graduates apply domain knowledge and skills required in diverse information environments**
Including information needs assessment, collection management, knowledge organization, information technology, user services, and pedagogy and information literacy instruction.

3 **Graduates demonstrate professional competencies, including leadership, critical thinking, inquiry, communication, collaboration, reflective practice, and ethical adherence. [In other words, the possess general knowledge and skills needed across professions.]**
Skills in management, communication and collaboration, research, and critical thinking.

What you learn in UBLIS 571 is applicable primarily in your future jobs as information specialists; it is domain-specific.

What you learn in UBLIS 581Management of Libraries & Info Agencies you can use to manage a library, a museum, or a restaurant; it is common-sense knowledge. What you learn in a course on research methods you can use for research in many domains.

Question: How are scope and usefulness of knowledge related?

1.3 Types of knowledge by degree of "vagueness" of knowledge

"Vagueness" is a vague umbrella term for the more well-defined distinctions listed below.

Concepts of "vagueness" of knowledge can be applied, for example, to knowledge of document relevance to a user's request (see below).
Rarely is knowledge "hard and fast"; it is important to assess a given knowledge item along the vagueness dimension to see how sound a basis it is for decision making.

You should be able to figure out for each subtype how important the distinction it makes is for different users and uses.

1.3.0 Introductory examples

- When giving a user data on measurements of natural or social phenomena, you must give an error range unless the user needs just a general ball park figure. In the literature values are often given with more precision than is warranted by the method by which the values are obtained (such as giving a poll number such as "approval rating 51.5%" when the margin of error of the poll is ±5%). As an information professional you must spot such cases (that border at the fraudulent) to protect the unsuspecting user.
- For making decisions about economic matters (for a company or whether to refinance your mortgage) it is very important to know the *certainty* of a piece of data, especially of economic forecasts. As a reference librarian you must give the user such metadata (data about data) describing the certainty with which the data in your answer hold.
- Do give data with exactly the precision the user needs, neither more nor less. This is a special case of Grice's Maxim of Quantity (see below)

Grice's Conversational Maxims

http://en.wikipedia.org/wiki/Paul_Grice, http://en.wikipedia.org/wiki/Gricean_maxims#Grice.27s_Maxims

Maxim of Quality: Truth

- Do not say what you believe to be false.
- Do not say that for which you lack adequate evidence.

Maxim of Quantity: Information

- Make your contribution as informative as is required for the current purposes of the exchange.
- Do not make your contribution more informative than is required.

Maxim of Relation: Relevance

- Be relevant.

Maxim of Manner: Clarity

- Avoid obscurity of expression. ("Eschew obfuscation")
- Avoid ambiguity.
- Be brief ("avoid unnecessary prolixity").
- Be orderly.

1.3.1 Precise vs. imprecise knowledge

One aspect of precision is numerical precision with which data can or should be given. Most measurements have an error range; often values should be rounded.

Precision is used here as generally used in the language. Do not confuse with precision as a measure of retrieval performance ("recall and precision"); that usage is specific to the information retrieval domain.

Examples

Knowing that a quote is from Shakespeare is hardly precise. A little more precise: knowing the quote is from Macbeth. Very precise: knowing it is from Macbeth, Act 2, Scene 3.

Persons' names with increasing precision: last name with initial; full first name; plus middle initial; full middle name; plus date of birth (important in cataloging).

Numeric values derived from empirical data, such as physical measurements or poll results, are subject to error; they have an error range. In reporting such numbers, give only significant digits and preferably indicate the error range to avoid conveying unwarranted precision.

- **Do not give data with more precision than is known (warranted by the measurement)** (e.g., polling data).
- **Do not give data with more precision than is required** even if more precision is known. For example, in a financial report to the board of a $10 million organization, round numbers to the nearest thousand, but for balancing the books use numbers to the penny.

1.3.2 Certain vs. uncertain knowledge

Linked with risk. Combined with precision: Confidence intervals.

A brief lesson in numerical literacy (numeracy)

A **confidence interval** gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. (Definition taken from Valerie J. Easton and John H. McColl's Statistics Glossary v1.1) (from/www.stat.yale.edu/Courses/1997-98/101/confint.htm).

Example: From a poll of 1,000 eligible voters asking about their approval of the US President, we may find that 450 or 45% approve and 55% disapprove; that is a *sample parameter*. But what we really want to know is what percentage of all eligible voters approve; that is the *population parameter*. To find out the actual value, we would have to ask all eligible voters. Since that is too expensive, we use the sample parameter from the poll (45%) as an estimate of the population parameter. But how confident are we that this estimate is close to the population parameter. Statisticians can tell us the probability (a quantitative expression of confidence) that the population parameter falls into a given range, for example as follows:

Range of approval % in the population (confidence interval)	Probability (level of confidence)
44% - 46%	90%
43% - 47%	95%
41% - 49%	99%

The larger the range, the more certain can we be that the true population parameter falls into the range.

Question: How does knowing all this make you a better reference librarian?

Yes/no statements (such as "facts" or rules in an experts system) vs. **probabilistic statements**.

- We can pretend that a document is either relevant to a user’s question or it isn’t, with no shades of gray in between (see next point for a different stance), yet still say that document X has a probability of 0.7 of being relevant. Our knowledge about relevance is uncertain.
- A medical condition, observation, or measurement may point to (be a symptom of) an underlying disease usually not with certainty but with some probability.

1.3.3 Graded assertions

For example, a document can be highly relevant or somewhat relevant. This can be expressed by a numerical score between 0 and 1. In other words, the set of relevant documents has no sharp boundary but rather is a fuzzy set. We cannot say that a document is a member of the set or that it is not a member of the set; rather, membership in a fuzzy set is a matter of degree. Shades of gray rather than black-and-white.

Example: In medical diagnosis, conditions that are used as symptoms (such as a headache) are determined at a given level of severity. The probability of a headache pointing to a given disease may well depend on the severity of the headache.

1.3.4 Unambiguous vs. ambiguous statements (including intentional ambiguity)

In the negotiations after the end of World War 2 the Western Allies and the Soviet Union agreed on the Oder-Neisse line as the border between Poland and Germany That agreement was ambiguous since the river Oder has two tributaries named *Neisse*. After the agreement was signed, the Western Allies wanted the border implemented on Neisse 1, but the Soviet Union insisted that the agreement meant Neisse 2, which moved a much larger chunk of land from Germany to Poland.

1.3.5 Facts (“true”, “objective”) or statements asserted as facts vs. opinion

"Hard" statements vs. judgment statements. News page vs. editorial page. Whose judgment or opinion?

1.3.6 Beliefs. Stronger than opinion. In a relativist view, nothing can be known for sure, there are only beliefs held by various parties. A database needs to indicate whose belief.

1.3.7 Knowledge about the accuracy, certainty, or trustworthiness of facts or rules

1.4 Types of knowledge, other aspects

1.4.1 Modality of knowledge items: descriptive, prescriptive, statement of possibility

Descriptive statement:	The car is going 50 miles per hour (what is)
Prescriptive statement:	The speed limit is 45 miles per hour (what should be) (a prescription for drivers as to speed of their cars)
Possibility statement:	The car can go 100 miles per hour (what could be)

More examples

Descriptive knowledge:	Knowledge about the effects of calorie intake, specific nutrients (such as vitamin E), and exercise.
Prescriptive knowledge:	Guidelines on nutrition and exercise.

People writing to an advice columnist report the facts of the case as they see them – descriptive knowledge. The advice columnist tells them what to do – prescriptive knowledge.

The law is prescriptive knowledge

Politicians and planners deal with the art of the possible; they need knowledge of what is possible. For example, some people claim to know that it is not possible to change Social Security because public opinion is against it and the votes to pass legislation are not there. On the other hand, visionary politicians may defy conventional knowledge of what is possible and make things possible. Time horizon of statements about possibility.

On possibility statements

A famous quote from Robert F. Kennedy (used in different variations in many of his speeches

"There are those that look at things the way they are, and ask why? I dream of things that never were, and ask why not?"
Robert F. Kennedy. Robert Kennedy in His Own Words: The Unpublished Recollections of the Kennedy Years   www.goodreads.com/work/quotes/332487

A paraphrase of a quote from George Bernhard Shaw

"You see things; and you say 'Why?' But I dream things that never were; and I say 'Why not?'"
From the play Back to Methuselah (1921), Part 1, Act 1, "In the Beginning"; said by The Serpent to Eve in the Garden of Eden.
www.quotecounterquote.com/2011/07/i-dream-things-that-never-were-and-say.html
(I may have violated Grice's maxim of quantity in giving so much information about the quote.)

1.4.3 Knowledge about what kinds of knowledge are important: Conceptual data schema (introduced in Lecture 1.2)

2 The nature of concepts / categories / classes

Importance: The nature of concepts is fundamental to information processing in people and in machines (see readings, particularly Skemp). Another way of looking at types of knowledge.

2.1 Types of concepts. Individual concepts and class concepts

2.1.1 **Individual concepts – individual entities**. Persistence over time.

2.1.2 **Class concepts / categories**. Simplified account

See Sections 2.2 - 2.4 for a discussion of the complexities of the structure of categories.

Concepts have	
Intension, **intensional definition,** "meaning"	**Definitional knowledge** as opposed to world knowledge (empirical knowledge) A concept or class defined in terms of attributes or characteristics all entities must possess in order to be members of the class. These are called **essential attributes or characteristics**. A characteristic of an individual entity can expressed in several ways: (1) the entity possesses an attribute (2) the entity is capable of entering a given relationship (occupy a comparable place in a network of relationships) A **query formulation is a definition**. It defines what it means for a document (or a person, or a computer program) to be relevant for the user. It encapsulates the user's intention. **Description logic** provides a formal way for defining concepts in a classification or ontology used in the semantic Web.
Extension	The set of individual entities belonging to the category For example, the set of all relevant documents In some cases it is possible to define a category by exhaustively listing all its members. This is called an **extensional definition**. Some defined concepts have empty extensions in reality, see below.

Example definitions (isa = is a type of):

Class	Definition
red balls	All objects that meet two conditions: Object <*hasType*> Ball **and** Object <*hasColor*> Red
pews	All objects that meet two conditions: Object <*isa*> Bench **and** [Object <*locatedIn*> Building, Building <*isa*> Church]
government documents	All documents that meet the conditions: Document <*publishedBy*> Organization, Organization <*hasType*> Government agency.
water-soluble substances	All chemical substances that meet the condition Substance <*solubleIn*> Water.
?	All English words that meet the condition Word <*canServeAsObjectFor*> Refine.

Essential attributes	Attributes that are used in the definition of a class
Accidental attributes	Any other attribute that one or more members of a class may possess

It often happens that all members of a class share an accidental attribute, that is an attributes that is not defining but nevertheless present in each member of the class. Such a general law can be determined by observation.

Example: Assume it is true that all government documents are authoritative.
Thus, if we have ascertained that a given document meets the definition for government document, we can conclude that the document is authoritative (knowledge of a regularity).

It is this ability to predict the behavior of an entity once it has been identified as belonging to a concept/category that makes for the usefulness of concepts; concepts are essential for economy of mental operations.

In law:

Fit facts of the case to a legal concept, for example determine that the facts of a case meet the definition of burglary.
Then apply the legal rule applicable to that concept.

Erroneous generalization: Stereotypes.

Relationship of definition to empirical data. Examples: One can define an animal species through a list of attributes such that no animal existing in nature fits the definition, for example, "an animal that looks like a horse and has a horn in the middle of the forehead". There are no unicorns in the real physical world, but there are plenty in fiction. Such fictitious animals are written about and depicted and become objects of searches. *Imaginary animals* is a very popular search topic in the International Children's Digital Library (ICDL).

2.1.3 **Mass concepts** (oil, flour, sugar) (how much?)
vs. **count concepts** (sugar cubes, books) (how many?)

2.1.4 **Abstract concepts** (freedom, justice).

Can define the concrete class of all countries in which freedom prevails.

2.2 Objectivist vs. organism-centered view of categories

(Quotes from Lakoff, *Women, fire, and dangerous things*. U. of Chicago Pr.; 1987)

Next page

2.2 Objectivist vs. organism-centered view of categories

Important: Information is not just transmitted but needs to be actively processed and assimilated by the receiver or learner (see the last paragraph of this section).

Objectivist view of categories (as characterized by George Lakoff)
• Symbols that correspond to the external world are *internal representations of external reality*. • Abstract symbols may stand in correspondence to things in the world independently of the particular properties of the organism that holds the symbols. • Since the human mind makes use of internal representations of external reality, the mind is *a mirror of nature*, and correct reason mirrors the logic of the external world. (p. XIII)

Organism-centered view (DS term) of categories (George Lakoff)
• *Summary of the objectivist view for contrast:* Do meaningful thought and reason concern merely the manipulation of abstract symbols and their correspondence to an objective reality, independent of any embodiment (except, perhaps, for limitations of the organism, such as being color-blind)? • *Organism-centered view*: Or do meaningful thought and reason essentially concern the nature of the organism doing the thinking – including the nature of its body, its interactions in its environment, its social character, and so on? (p. XV - XVI) **Embodied cognition** versus symbolic representation

A balanced view (D. Soergel)
• Interacting with the physical, social, and intellectual world around us and with our own selves, we form complex mental models which allow us to better understand the world and better understand ourselves and thus help us to take actions in the world towards achieving our objectives. This formation of mental models has a social dimension; it is often done in interaction with or building on the models of others – as in group learning. • These mental models, which include concepts / categories, reflect a structuring of experience in ways useful to the person. A person's experience is shaped by perceptions of the world (within the limitations of the person's faculties for perception and thought) and by the modes of interaction with objects in the world. Thus, **a mental model is not simply a mirror image of the outside world but rather an actively shaped image, adapted to the person's needs**, often distorted, often enriched (or contaminated, depending on one's point of view) with elements that have no counterpart in the "real" world (but might well be realized as the person shapes the world). *See Supplement* SLecture 2.1

The sense-making approach to information service: a person **must make personal sense of information given**
The view of mental models, concepts, and categories presented above is important for an understanding of how people use information and what information should be provided to people. In this view, a person must assimilate information into her mental model; a person **must make personal sense of the information**. Different people may get different things out of the same document. In that sense, one might say that information does not exist objectively, but only as it gives rise to a change in a person's mind. Or that a book does not convey information as much as it is a stimulus for the reader to create and elaborate her own information in her own mind.

The sense-making approach in related disciplines	
In literary theory	This is the position taken by **reader response theory**. The expression "I did not get much out of it" is in tune with this "active reader" position. The art of giving the reader a "relevant" book, then, is to find a book that allows this reader with his background and mental models to "get something out of" the book, to construct his own knowledge, updating his mental models in a way that will help him to find better solutions to the problems he faces.
In education	The **constructivist theory of learning** holds that we learn best by constructing or reconstructing knowledge for ourselves. **Discovery learning** is a closely related approach. It holds that students learn best when they explore a subject and discover facts and relationships for themselves. In science this means that students discover scientific laws through their own experiments. The **job of the teacher or information specialist** then is to create an environment, including access to information, that enables students or users to do their own discovery and knowledge construction with guidance provided only to the extent necessary ("scaffolding").

2.3 Explicit definition of categories vs. prototypes and fuzzy membership. Radial categories

Prototype. Example *Chair*:

> *Chair*, living room chair, kitchen chair, lawn chair, easy chair, rocker, armchair, chaise longue, bar stool, stool?, footstool??, ottoman??

Necessary attributes vs. sharing a sufficient number of attributes.

Knowledge of concepts stored in memory as explicit definitions or prototypes or sets of examples? In reality a mixture of all three?

Importance of examples in thesaurus scope notes

Radial category. Example**:** *Mother*
(a category that has a central case but then many cases deviating more or less in different directions)

> There are many "models" of what a *mother* is (Lakoff 1987, p. 83).
>
> > "The central case, where all the models converge, includes a mother who is and always has been female, and who gave birth to the child, supplied her half of the child's genes, nurtured the child, is married to the father, is one generation older than the child, and is the child's legal guardian."
>
> The following cases share some, but not all, of these features. The first four emphasize a biological perspective, the others a social perspective.
>
> - Biological mother (also called natural mother, but that term was abandoned)
> - Birth mother (term for biological mother in the context of adoption)
> - Surrogate mother
> - Genetic mother
> - Rearing mother
> - Stepmother
> - Adoptive mother
> - Foster mother
> - Unwed mother
>
> **Importance for reference interview and searching.**
>
> What would you search for if the user says he wants documents about *mother*?
> Understanding radial categories helps you ask the user questions to better pinpoint the information need.

Application to retrieval:

The documents (or statements, such as statements of fact or hypotheses, etc.) relevant to a query constitute a category (a subset of all documents). We can define such a category in two ways:

- through a query formulation that explicitly specifies the features that make a document relevant (expressing the intent of the user, intensional definition). (This query formulation could be applied in a Boolean search (to be retrieved, a document needs to meet all conditions) or in a ranked retrieval search (documents are retrieved even if they do not meet all conditions exactly, and are ranked by how closely they meet the conditions);
- through a sample document that serves as a prototype of relevant documents or several documents that serve as examples(“find more like this”). Related to extensional definition. For example, in a set of documents found the user may mark those she considers relevant without specifying a formal query (formal definition of what is relevant). A system can take the set of documents marked relevant and find more documents that fit into the set.

The category “relevant documents” can be a radial category when there are different ways in which a document can be relevant to the query. Needs several query formulations.

2.4 Basic level categories: Application of categories or concepts to action (Eleanor Rosch)

From this perspective, what categories are most useful and worth the effort to learn?

Example. Gain in knowledge of what to do as concepts get more specific

- If somebody tells you that there is a piece of *furniture* in a room you have been assigned, that does not tell you much. You still do not know what you can do in the room.
- If somebody tells you there is a *chair* in the room, that tells you a lot more; you know you can sit down.
- If somebody tells you there is a *easy chair* in the room, you know a little more, but not much more; you still know only that you can sit down (perhaps a bit more comfortably).

There is a **big information gain from *furniture* to *chair***, but a **small gain from *chair* to *easy chair***. So it is worthwhile to learn about the category of *chair*, but the added benefit of knowing all the many specific types of chair would be low and the learning effort would be very high.

chair is at the optimal level in the hierarchy, it is a **basic level category**.

More examples

Superordinate	Basic level	Subordinates	
Furniture	Chair	Kitchen chair	Living room chair
		Lawn chair	Bar stool?
		Easy chair	Footstool??
	Table	Dining room table	Kitchen table
		Card table	Pool table

The idea of basic level categories is important for information services and learning and instruction in many ways, for example:

- Users tend to ask questions using basic level categories even if their information need is more specific. The reference librarian (or a computer system designed to replace the reference librarian) needs to recognize when it is necessary to probe for the real specific information need.
- Basic level categories are learned first. A classification scheme for children must be built from basic level categories.

Empirical results of studies in cognitive psychology

- **Number of attributes.** Subjects were given words that name categories of objects, such as *furniture*, *chair*, *lawn chair*, and were asked to list **attributes** of that category.

  For **superordinate categories**, such as *furniture*, subjects listed **few attributes**.

  For **intermediate categories**, such as *chair*, subjects listed **many attributes**. **Basic level**

  For **subordinate categories**, such as *easy chair*, subjects listed **a few additional attributes** beyond those for *chair*.

  "Basic level categories are the most inclusive level of classification at which **objects have a significant number of attributes in common**." (p. 214)

- **Functionality:** Basic level categories are the most inclusive level of classification at which **objects share substantive functionality**. (p. 217). Example:

  furniture (table, chair, cabinet) (**superordinate category**):
  Few, if any movements or other things you do are **in common** to all types of furniture.

  chair (any type (**basic level**):
  People make the **same kind of movements** (sitting down) for all types of chair

  easy chair (**subordinate category**):
  movements are hardly different from any other type of chair

- **When learned?** Basic level categories are **learned first**.

Level	Number of attributes	Number of instances	Number of categories	Usefulness for action
Superordinate	Few attributes	Gazillion instances	Few categories	Low
Basic level	Many attributes (high information gain)	Many instances	Medium number of categories	High
Subordinate	Only a few more attributes (low information gain)	Few instances	Very large number of categories	Only slightly higher

Note: Basic level may depend on group - culture and subculture.

Elaboration in supplement SLecture 2.1

pink **Lecture 2.2** ***February 3***

Knowledge representation

➜ UBLIS 506 IT, UBLIS 569 Data Management

Learning objectives	1 Graduates understand and are able to apply **specific approaches to the organization / representation/ modeling of data / information / knowledge** (P2.3.1,1#, especially the entity-relationship approach (P2.3.1,1.1) and frames (P2.3.1,1.5) 1.1 See how frames are useful to help learners acquire and organize knowledge. (P2.3.1,1.5.1) 2 **Understand different mechanisms of knowledge processing.** (P2.3.1,2) 2.1 Understand the principle of **hierarchical inheritance** and be able to apply it to achieve more compact internal storage in the computer and in the mind and more compact and more easily grasped external representation. (P2.3.1,2.1) 3 **Be able to apply this understanding to finding answers to a user's question by combining knowledge from one or more systems**. (P2.3.2,1) 4 Understand the pervasive role of classification throughout the human endeavor. (P2.3.9,1)
Practical significance	Knowing about system-internal knowledge representation and conceptual data schemas is important for organizing a body of knowledge for retrieval and beyond that, for inference, that is, for a system that can draw conclusions from the knowledge stored (and thus create new knowledge), rather than simply retrieving what is there.

Key idea:

Hierarchical inheritance

Using knowledge about a class to answer questions about any member of the class.

Using knowledge about a broad class to answer questions about any subordinate class or any member of such a subordinate class.

Reorganize a set of data so it requires less storage space and can be communicated and assimilated more efficiently.

pink

Outline

0	Forming categories in a set of entities to create a more efficient data structure using hierarchical inheritance – Introductory exercise with two examples
1	Definition of knowledge representation (in the mind, on paper, for computers) Examples
2	Approaches to knowledge representation
3	Some mechanisms in knowledge representation
4	Criteria for describing and evaluating knowledge representations (Suppl. SLecture 2.2)

For this lecture, the notes stand on their own. Just read, no audio.

0 Forming categories in a set of entities to create a more efficient data structure using hierarchical inheritance.

Introductory exercise with three examples.

*S*eated: In-lecture exercise. *O*nline: Do on your own, then go to answers on p. ~68

Example 1. Restaurant menu

Consider the menu listing below (the actual menu as it appears on the restaurant's website is on the back of this page). How easy is it for a guest to figure out what the differences between the options are? How many things are different, how many the same? Present the information given in the three options so that the guest can see quickly what is the same and what is different, using less text to boot.

Third Course		
Fondue Feast	**Fondue Fusion**	**Lobster Indulgence**
Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.	Lobster Tail, Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.	Lobster Tail(s), Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.

Revised Presentation
After you did your own, see p. ~68 for an answer

Example 2. Bibliographic records on facing page, in the MARC format

Reorganize these bibliographic records, using hierarchical inheritance for efficient storage

Elaboration: You can guess what the MARC tags 100, 245, etc. mean or you can look at the beginning of the Model catalog where the MARC tags are defined. Look at these bibliographic records; two of them "do not belong". What about the other four? How hard would it be for a user who gets these records returned in a search to figure out what the differences are? **What information is the same for all four records**, what information is different for individual records? Present the information given in these four records so that the user can see quickly what is the same and what is different, using less text to boot. The idea is exactly the same as in example 1.

****

Belongs to previous page: Image of the actual restaurant menu (from their website)

Example 2. Bibliographic records. See instructions on facing p. ~64.

Document 1

100 1 Mager, Robert Frank, $d 1923-
245 10 Developing attitude toward learning / $c Robert F. Mager.
260 Belmont, Calif. : $b Fearon/Pitman Publishers, $c c1968.
300 vii, 104 p. ; $c 22 cm.
650 0 Interaction analysis in education.
650 0 Learning, Psychology of.
650 0 Group work in education.
650 0 Classroom management.

Document 2

100 1 Candelora, D[eborah] M.
245 10 Hands-on technology program $h [computer file]
246 HOT program
260 [Ramsey, NJ]: $b [Galaxy Networks], $c 1996-
500 Title from the home page HTML title
500 Material copyrighted by D. M. Candelora
500 Accessed 1998 Feb. 2
650 0 Science - Study and teaching (Elementary) - Aids and devices
650 0 Science - Experiments
650 0 Computers - Study and teaching (Elementary) - Aids and devices
650 0 Mathematics - Study and teaching (Elementary) - Aids and devices
650 0 Learning by discovery
650 0 Active learning
856 4 $u www.galaxy.net/~k12/ $n Ramsey, NJ

Document 3

100 1 Mager, Robert Frank, $d 1923-
245 10 Developing attitude toward learning : $b or SMATs 'n' SMUTS / $c Robert F. Mager.
250 2nd ed.
260 London : $b Kogan Page, $c 1991, c1990.
300 116 p. ; $c 23 cm.
650 0 Interaction analysis in education.
650 0 Learning, Psychology of.
650 0 Group work in education.
650 0 Classroom management.
650 0 Students $a Motivation

Document 4

100 1 Conant, James Bryant, $d 1893-1978
245 10 The comprehensive high school; $b a second report to interested citizens $c by James B. Conant.
260 New York, $b McGraw-Hill $c [1967]
300 vi, 95 p. $c 21 cm.
650 0 Education, Secondary
650 0 Education $z U.S. $y 1945-
650 0 Comprehensive High Schools $z U.S. $y 1945

Document 5

100 1 Mager, Robert Frank, $d 1923-
245 10 Developing attitude toward learning, $b or, SMATS "n" SMUTS / $c Robert F. Mager
250 2nd ed.
260 Belmont, Calif. : $b David S. Lake, $c c1984.
300 x, 116 p. : $b ill. ; $c 24 cm.
490 1 The Mager library
500 Rev. ed. of: Developing attitude toward learning. 1968.
650 0 Interaction analysis in education.
650 0 Learning, Psychology of.
650 0 Group work in education.
650 0 Classroom management.

Document 6

100 1 Mager, Robert Frank, $d 1923-
240 10 Developing attitude toward learning. $l Spanish
245 10 Desarrollo de actitudes hacia la ensenanza / $c Robert F. Mager.
260 Barcelona : $b Martacinez Roca, $c c1985.
300 158 p. : $b ill. ; $c 19 cm.
650 0 Interaction analysis in education.
650 0 Learning, Psychology of.
650 0 Group work in education.
650 0 Classroom management.

Example 3. Food products
(Data on facing page and the table to the right, modeled after Consumer Reports, easy to see cols with same ●)

FP Features	FP1	FP2	FP3	FP4
Spinach	●	●		
Beans			●	●
Frozen	●	●		
Canned			●	●
Libby	●		●	
Campbell		●		●

The principle and process you should have discovered from Examples 1 and 2 is this:

Step 1: In the set of all items, find groups of items that have information in common.

In Example 1, the menu, there is only one group of three menu choices, so this step is trivial.
In Example 2, bibliographic records, there are four records that form a group. For now, we ignore the other two (each forms a group of 1).

Step 2: Determine what information is in common to all items.

Example 1: What is in common to the three menu choices?
Example 2: What is in common to the four bibliographic records in the group?

Step3: Reorganize the information for more efficient storage and display.

- Create a new *common record* that includes this common information.
- For each of the original items include a reference to the common record and then list only the item-specific information that is not in the common record.

You can find examples in your own answers for Examples 1 and 2 or in the answers provided (p. ~71 - 72). Each member of a group is said to *inherit* all information that is in the common record for the group; this is called **hierarchical inheritance**.

In Examples 1 and 2 there is only one group. Example 3 has two groups.

Step 1: Find groups of items that have information in common. Already done. In real life and in Assignment 4 you must do this step yourself.

Now do Steps 2 and 3 for the first group:

Step 2: Determine what information is in common to food products 1 and 2

Step 3: Reorganize the information

- Create a new *common record* for the first group; include the information in common to food products 1 and 2.
- Create new records for food products 1 and 2: Include a reference to the common record and then list only the information specific to food product 1 and food product 2, respectively that is not in the common record.

Repeat for the second group, food products 3 and 4.

Grouping items by their common characteristics and creating a common record is a process of abstraction, an important element in the process of concept formation.

Original database	Restructured database
Food product 1. Libby's frozen spinach *Food*: spinach *Preservation*: frozen *Manufacturer*: Libby **Food product 2. Campbell's frozen spinach** *Food*: spinach *Preservation*: frozen *Manufacturer*: Campbell	
Food product 3. Libby's canned beans *Food*: beans *Preservation*: canned *Manufacturer*: Libby **Food product 4. Campbell's canned beans** *Food*: beans *Preservation*: canned *Manufacturer*: Campbell	

Note: The restructured database has more records, but they are much shorter

A good **example for hierarchical inheritance** is a cookbook which may have a basic recipe for **potato soup** with a lengthy description and then many variations that say, for example

Rosemary potato soup

"Use the recipe for potato soup but add 1 ½ teaspoons minced fresh rosemary ."

Answers to exercises Lecture 2.2, Section 0 (exercises starting on p. ~63)

Restaurant menu

Consider the menu listing below (the actual menu as it appears on the restaurant's website is on the back of this page).

How could this menu be presented so it takes less space and is easier to read

Hint: Try to figure out what is actually different for each choice

Third Course		
Fondue Feast	**Fondue Fusion**	**Lobster Indulgence**
Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.	Lobster Tail, Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.	Lobster Tail(s), Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.

Revised Presentation		
Third Course		
All Sauerbraten NY Strip Nueske's Applewood Smoked Bratwurst Hefeweizen Marinated Shrimp Roasted Garlic Crusted Chicken Sun Dried Tomato Ravioli Fresh Vegetables		
Fondue Feast	**Fondue Fusion**	**Lobster Indulgence**
	Lobster Tail	Lobster Tail(s)
Filet Mignon	Filet Mignon	

Hierarchical inheritance for efficient storage of bibliographic data

This example uses hierarchical inheritance "with override": If a field from the parent record is repeated in the child record with different information, the child information overrides what would otherwise be inherited from the parent. (Documents 2 and 4 are not related to any other document and thus are omitted here.)

Generic record A
100 1 Mager, Robert Frank, $d 1923-
245 10 Developing attitude toward learning / $c Robert F. Mager.
650 0 Interaction analysis in education.
650 0 Learning, Psychology of.
650 0 Group work in education.
650 0 Classroom management.

Document 1 Inherits from: A
260 Belmont, Calif. : $b Fearon/Pitman Publishers, $c c1968.
300 vii, 104 p. ; $c 22 cm.

Document 6 Inherits from: A
240 10 Developing attitude toward learning. $l Spanish
245 10 Desarrollo de actitudes hacia la ensenanza
260 Barcelona : $b Martacinez Roca, $c c1985.
300 158 p. : $b ill. ; $c 19 cm.

Generic record B Inherits from: A
245 10 $b or SMATs 'n' SMUTS
250 2nd ed.

Document 5 Inherits from: B
260 Belmont, Calif. : $b David S. Lake, $c c1984.
300 x, 116 p. : $b ill. ; $c 24 cm.
490 1 The Mager library
500 Rev. ed. of: Developing attitude toward learning. 1968.

Document 3 Inherits from: B
260 London : $b Kogan Page, $c 1991, c1990.
300 116 p. ; $c 23 cm.
650 0 Students $a Motivation

Another document example

Three-part journal article, three separate entries

Solomon, P. (1997a). **Discovering information behavior in sense making. I. Time and timing**.
Journal of the American Society for Information Science, 48(12), 1097-1108.

Solomon, P. (1997b). **Discovering information behavior in sense making. II. The social**.
Journal of the American Society for Information Science, 48(12), 1109-1126.

Solomon, P. (1997c). **Discovering information behavior in sense making: III. The person.**
Journal of the American Society for Information Science, 48, 1127-1138.

Three-part journal article, one consolidated entry

Solomon, P. (1997). **Discovering information behavior in sense making**.
Journal of the American Society for Information Science.
I. Time and timing. 48(12), 1097-1108.
II. The social. 48(12), 1109-1126.
III. The person. 48(12), 1127-1138.

Note: Why people use such poorly conceived symbols as Roman numerals when the far better alternative of Arabic numerals is available is beyond me (perhaps they think it looks more erudite and impressive), but that is how it is done in the journal so that it is how it must be recorded.

Example 3a. Food products – database answer partial
See whether you can now complete the answer for the second group

<table>
<tr><th>Original database</th><th>Restructured database</th></tr>
<tr><td>

Food product 1. Libby's frozen spinach
Food: spinach
Preservation: frozen
Manufacturer: Libby

Food product 2. Campbell's frozen spinach
Food: spinach
Preservation: frozen
Manufacturer: Campbell</td><td>Common record A
Food product A. Frozen spinach
Food: spinach
Preservation: frozen

Food product 1. Libby's frozen spinach
inherits from: Food product A
Manufacturer: Libby

Food product 2. Campbell's frozen spinach
inherits from: Food product A
Manufacturer: Campbell</td></tr>
<tr><td>

Food product 3. Libby's canned beans
Food: beans
Preservation: canned
Manufacturer: Libby

Food product 4. Campbell's canned beans
Food: beans
Preservation: canned
Manufacturer: Campbell</td><td></td></tr>
</table>

Note: The restructured database has more records, but they are much shorter

Example 3a. Food products – database answer complete

Original database	Restructured database
	Common record A **Food product A. Frozen spinach** *Food*: spinach *Preservation*: frozen
Food product 1. Libby's frozen spinach *Food*: spinach *Preservation*: frozen *Manufacturer*: Libby	**Food product 1. Libby's frozen spinach** *inherits from*: Food product A *Manufacturer*: Libby
Food product 2. Campbell's frozen spinach *Food*: spinach *Preservation*: frozen *Manufacturer*: Campbell	**Food product 2. Campbell's frozen spinach** *inherits from*: Food product A *Manufacturer*: Campbell
	Common record B **Food product B. Canned beans** *Food*: beans *Preservation*: canned
Food product 3. Libby's canned beans *Food*: beans *Preservation*: canned *Manufacturer*: Libby	**Food product 3. Libby's canned beans** *inherits from*: Food product B *Manufacturer*: Libby
Food product 4. Campbell's canned beans *Food*: beans *Preservation*: canned *Manufacturer*: Campbell	**Food product 4. Campbell's canned beans** *inherits from*: Food product B *Manufacturer*: Campbell

Note: The restructured database has more records, but they are much shorter

1 Definition of knowledge representation

(in the mind, on paper, for computers)

Knowledge representation is the expression of knowledge through a system of symbols or signs, such as words, Dewey numbers, or icons. A knowledge representation scheme must provide

- symbols that refer to objects in the world or objects in the mind (put differently, symbols that refer to entity values, roughly, nouns);
- symbols that refer to relationship types (roughly, verbs);
- a syntax that allows for the expression of statements consisting of entity identifiers linked through relationship symbols.

Natural language is a very expressive knowledge representation system, but it is hard for a computer system to figure out what a natural language text means and then act on this knowledge. Need simpler KR systems for useful computer systems.

Approaches	Entity-relationship representation (very common in the database field)
	Semantic network representation
	Frame representation (artificial intelligence & object-oriented programming)

Note: In 571 we talk about knowledge representation in the abstract. Implementation in databases is treated briefly in 506 Information Technology and extensively in more advanced IT courses. In the supplement SAssignment 6 there is an assignment that takes you through creating and querying a simple Microsoft Access implementation of the University Database.

Knowledge representation examples (for computer systems)

A small example with data on birds

A large example with data on food products

The small example.

A small set of data represented in three different ways for you to have a good look to get a feel for different representations. The entity-relationship representation is the most basic. It is the representation used for Linked Data. The other two representations show exactly the same data, each in a way that is easier to understand for the human reader.

Semantic network.
Data common to all birds (ignoring a few exceptions) are listed only for Bird (the common node for the group of all birds; these data are understood to be true for every node under Bird. The subordinate nodes **inherit** everything that is said about Bird. Based on the information in the semantic network:

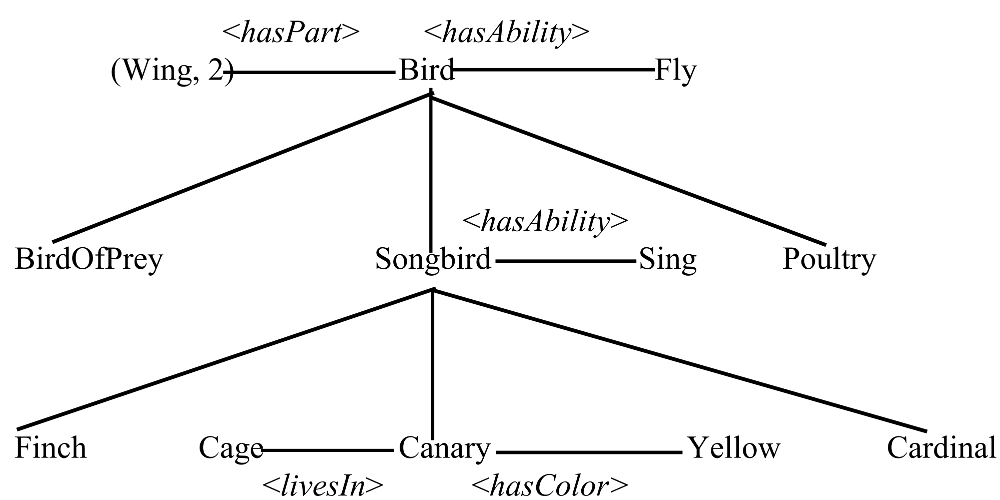
Write down all statements you can make about Song Bird.

Write down all statements you can make about canary?

Small example: Representation of data on birds

Entity-relationship representation of data on birds (nicely ordered)					
Bird	*<isa>*	Animal	Poultry	*<isa>*	Bird
Bird	*<hasAbility>*	Fly	Canary	*<isa>*	Songbird
Bird	*<hasPart>*	(Wing, 2)	Canary	*<hasColor>*	Yellow
BirdOfPrey	*<isa>*	Bird	Canary	*<livesIn>*	Cage
Songbird	*<isa>*	Bird	Finch	*<isa>*	Songbird
Songbird	*<hasAbility>*	Sing	Cardinal	*<isa>*	Songbird

Semantic network representation of data on birds



Key idea: **Hierarchical inheritance**	Data common to all birds (ignoring a few exceptions) are listed only for Bird (the common node for the group of all birds); these data are understood to be true for every node under Bird. The subordinate nodes **inherit** everything that is said about Bird. Based on the information in the semantic network, we can say Songbird *<hasAbility>* Fly and Songbird *<has part>* (Wing, 2) Songbird **inherits** these characteristics from Bird. ▸**Now write down everything you can say about Canary** based on the semantic network Note: Just having a hierarchy alone does not make hierarchical inheritance. **Only when the hierarchy is used to pass down characteristics from higher nodes to lower nodes is there hierarchical inheritance.**
Spreading activation	Activation (or attention) may spread from a node to a neighboring node: A person thinking about yellow (yellow is activated) may be reminded of a canary (canary is activated), and then of songbird and then of a bird singing (sing is activated) and then, again starting from songbird, of any bird and thus of a bird flying)

Small example. Frame representation of data on birds with hierarchical inheritance
See next page for more explanation

Frame for:	**Bird**	
inheritsFrom:	Animal	/* *inheritsFrom* and *isa* are the same. */
includesSpecific:	Songbird; BirdOfPrey, Poultry	/* This slot does not inherit down. */
hasColor:		
hasPart:	(Wing, 2)	
hasAbility:	Fly	
livesIn:		

Frame for:	**Songbird**	[ ] = [inherited]
inheritsFrom:	Bird; [Animal]	
includesSpecific:	Canary; Finch; Cardinal	
hasColor:		
hasPart:	[(Wing, 2)]	Anything that is true for a bird is true for a songbird through inheritance
hasAbility:	Sing; [Fly]	
livesIn:		

Frame for:	**Canary**	[[inherited from two levels above]]
inheritsFrom:	Songbird; [Bird]; [[Animal]]	Anything that is true for a bird or a song bird is true for a canary through inheritance
includesSpecific:		
hasColor:	Yellow	
hasPart:	[[(Wing, 2)]]	
hasAbility:	[\Sing]; [[Fly]]	Anything that is true for a bird or a song bird is true for a canary through inheritance
livesIn:	Cage	

Frame for:	**Penguin** /* added to illustrate inheritance override */
inheritsFrom:	Bird; [Animal]
includesSpecific:	
hasColor:	White; Black
hasPart:	[(Wing, 2)]
hasAbility:	Swim; NOT Fly; [Fly](example of **overriding** an inherited piece of data)
livesIn:	Antarctica

Optional. Look at the Web pages for the following two birds (from www.allaboutbirds.org)

www.allaboutbirds.org/guide/Black-bellied_Whistling-Duck/id
www.allaboutbirds.org/guide/Wood_Duck/id

Observe the common template used to organize the information. Not all birds in the database have such complete data.

Explanation of frames

A frame is collection of data item. Each data item is a slot. Here we have a collection of frames that all have the same slot structure. Each data item is from the entity-relationship representation, so the slot names (labels, tags) are the relationship names.

A frame is a record with additional capabilities. The only additional capability used in the example is hierarchical inheritance.

The data in [ ] are inherited. They could be omitted because a frame system would know they are there by inheritance. They are included for the convenience of the reader. A plain record system as used in most library catalogs does not implement hierarchical inheritance so all information about a work must be included in a record, whether on a conceptual level it is inherited or not

The override example illustrates how we can have Penguin under Bird without saying that a Penguin can fly, as would be implied by inheritance.

Frame representation of data on birds with hierarchical inheritance. **Repeated**
See previous page for more explanation

Frame for:	**Bird**	
inheritsFrom:	Animal	/* *inheritsFrom* and *isa* are the same. */
includesSpecific:	Songbird; BirdOfPrey, Poultry	/* This slot does not inherit down. */
hasColor:		
hasPart:	(Wing, 2)	
hasAbility:	Fly	
livesIn:		

Frame for:	**Songbird**	[ ] = [inherited]
inheritsFrom:	Bird; [Animal]	
includesSpecific:	Canary; Finch; Cardinal	
hasColor:		
hasPart:	[(Wing, 2)]	Anything that is true for a bird is true for a songbird through inheritance
hasAbility:	Sing; [Fly]	
livesIn:		

Frame for:	**Canary**	[[inherited from two levels above]]
inheritsFrom:	Songbird; [Bird]; [[Animal]]	Anything that is true for a bird or a song bird is true for a canary through inheritance
includesSpecific:		
hasColor:	Yellow	
hasPart:	[[(Wing, 2)]]	
hasAbility:	[\Sing]; [[Fly]]	Anything that is true for a bird or a song bird is true for a canary through inheritance
livesIn:	Cage	

Frame for:	**Penguin** /* added to illustrate inheritance override */
inheritsFrom:	Bird; [Animal]
includesSpecific:	
hasColor:	White; Black
hasPart:	[(Wing, 2)]
hasAbility:	Swim; NOT Fly; [Fly](example of **overriding** an inherited piece of data)
livesIn:	Antarctica

Optional. Look at the Web pages for the following two birds (from www.allaboutbirds.org)

www.allaboutbirds.org/guide/Black-bellied_Whistling-Duck/id
www.allaboutbirds.org/guide/Wood_Duck/id

Observe the common template used to organize the information. Not all birds in the database have such complete data.

Large example: Representation of data on food products

Purposes of the food information system	**Government agency**: Determine the safety of a food product **Consumer:** Find food products to be avoided with a given allergy **Cook:** Prepare a food product **Food manufacturer:** Produce the ingredient label
Sample questions	Find all products to be avoided by people allergic to eggs. I have cauliflower, onions, and tomatoes I need to use up. Find a good recipe.

The complete data on food products are shown only in the entity-relationship representation, as in the following examples:

FP0 *<hasName>* "Food product"

FP1 *<hasName>* "Vegetable product"
FP1 *<isa>* FP0 # "Food product"
FP1 *<comesFromSource>* Plant

FP2 *<hasName>* "Meat product"
FP2 *<isa>* FP0 # "Food product"
FP2 *<comesFromSource>* Animal
FP2 *<comesFromPart>* Carcass

FP3 *<hasName>* "Egg product"
FP3 *<isa>* FP0 # "Food product"
FP3 *<comesFromSource>* Animal
FP3 *<comesFromPart>* Egg

FP4 *<hasName>* "Prepared food"
FP4 *<isa>* FP0 # "Food product"
FP4 *<processedBy>* *

The statements are arranged so that one can easily see what a frame representation (a frame for each food product) would look like. In this representation I used the short *isa* (A is a kind of B) instead of the long *inheritsFrom*.

Notes (can omit in first reading)

On the notation for entity-relationship representation. Lectures 1.2 - 4.2 use an easily readable notation to represent statements that link entities through relationships. The names for relationship types are included in < >. If the relationship links only two entities (a very frequent case). these statements are triples:

Entity1 <relationship> Entity2, for example, FP15 <comesFromSource> Chicken.

Lecture 5.1 on RDF introduces the notation predominantly used in machine-readable files of triples; it is still readable, but less so.

On entity identifiers. Entity values (such as specific food products, specific plants and animals, chemical substances) must be identifiable; they need an identifier. (Think of license plates or product serial numbers or Universal Product Codes.) The identifier can be a number, such as FP3 (preferred, on the Web: Universal Resource Identifiers) or a natural language term, such as Carrot (more readable). In the example I used numbers for food products to illustrate the principle and terms for everything else to improve readability. Food products must have human-readable names. The name of a food product is given just another piece of data; the name itself is a string, indicated through " ". For readability, I added the food product name (in " ") after each mention of a food product ID, even though this is redundant.

Entity-relationship (E-R) representation

Conceptual data schema (entity types and relationship types covered; *isa* is short for *is a*)

Entity types

Name
FoodProduct (**FP**)
Organism
  Person
OrganismPart
Substance
Form
Process
Purpose
Container
GoodCommodity
  Portion
LegalEntity
  Person
MoneyNumber

Relationship types

FoodProduct	<*hasName*>	Name
FoodProduct	<*isa*>	FoodProduct
FoodProduct	<*comesFromSource*>	Organism
FoodProduct	<*comesFromPart*>	OrganismPart
FoodProduct	<*isExtractedSubstance*>	Substance
FoodProduct	<*isMadeFrom*>	FoodProduct[1]
FoodProduct	<*hasIngredient, withPurpose*>	FoodProduct, Purpose
FoodProduct	<*contains*>	Substance (omitted in the examples)
FoodProdcuct	<*processedBy*, *withIntensity*, *withPurpose*>	Process , (e.g. Cooking) Intensity,(e.g. FullyCooked) Purpose (e.g. MakeEdible)
FoodProdcuct	<*hasForm*>	Form
FoodProdcuct	<*packedIn*>	(Container, Volume)
Organism	<*eat*>	Portion Or Substance OR FP
<*buy/sell*> (LegalEntity [seller], LegalEntity [buyer], Good, MoneyNo.)		

[1]Not used in the lecture examples, but in the reading on a Food Description Language

Entity values can be seen from the E-R statements (FP0, FP1, etc, Plant, . . .) FP = FoodProduct

E-R statements [] inherited from one level above [[]] inherited from two levels above # Comment

FP0	***<hasName>***	**"Food product"**
FP1	***<hasName>***	**"Vegetable product"**
FP1	*<isa>*	FP0 # "Food product"
FP1	*<comesFromSource>*	Plant
FP2	***<hasName>***	**"Meat product"**
FP2	*<isa>*	FP0 # "Food product"
FP2	*<comesFromSource>*	Animal
FP2	*<comesFromPart>*	Carcass
FP3	***<hasName>***	**"Egg product"**
FP3	*<isa>*	FP0 # "Food product"
FP3	*<comesFromSource>*	Animal
FP3	*<comesFromPart>*	Egg
FP4	***<hasName>***	**"Prepared food"**
FP4	*<isa>*	FP0 # "Food product"
FP4	*<processedBy>*	*
FP5	***<hasName>***	**"Soup"**
FP5	*<isa>*	FP4 # "Prepared food"
FP5	*<processedBy>*	*
FP5	*<hasForm>*	LiquidOrSemiliquid
FP11	**<hasName>**	**"Diced carrots"**
FP11	*<isa>*	FP1 # "Vegetable pr."
FP11	*<comesFromSource>*	CarrotPlant
FP11	*<comesFromPart>*	Root
FP11	*<hasForm>*	Diced
FP12	***<hasName>***	**"Cut green beans"**
FP12	*<isa>*	FP1 # "Vegetable pr."
FP12	*<comesFromSource>*	BeanPlant
FP12	*<comesFromPart>*	ImmatureFruit
FP12	*<hasForm>*	Cut
FP13	***<hasName>***	**"Chicken broth"**
FP13	*<isa>*	FP2 # "Meat product"
FP13	*<comesFromSource>*	Chicken
FP13	*<comesFromPart>*	Meat
FP13	*<comesFromPart>*	Bones
FP13	*<isExtractedSubstance>*	{Fat, Protein, Flavor}
FP13	*<processedBy>*	Cooking
FP13	*<hasForm>*	Liquid
FP14	***<hasName>***	**"Cubed cooked chicken"**
FP14	*<isa>*	FP2 # "Meat product"
FP14	*<comesFromSource>*	Chicken
FP14	*<comesFromPart>*	SkeletalMeat
FP14	*<processedBy>*	Cooking
FP14	*<hasForm>*	Cubed
FP15	***<hasName>***	**"Eggs"**
FP15	*<isa>*	FP3 # "Egg product"
FP15	*<comesFromSource>*	Chicken
[FP15	*<comesFromPart>*	Egg
FP16	**<hasName>**	**"Durum wheat flour"**
FP16	*<isa>*	FP1 # #Vegetable pr."
FP16	*<comesFromSource>*	DurumWheat
FP16	*<comesFromPart>*	SeedKernel
FP16	*<hasForm>*	Ground
FP17	***<hasName>***	**"Noodles"**
FP17	*<isa>*	FP4 # "Prepared food"
FP17	*<hasIngredient>*	FP16 #Durum wheat flour
FP17	*<hasIngredient>*	FP15 # "Eggs"
FP17	*<processedBy>*	Mixing
FP17	*<processedBy>*	Extruding
FP17	*<processedBy>*	Drying
FP18	***<hasName>***	**"Flavoring"** # detail omitted
FP19	***<hasName>***	**"BHT"** # detail omitted
FP20	***<hasName>***	**"Chicken noodle soup"**
FP20	*<isa>*	FP5 # "Soup"
FP20	*<hasIngredient>*	FP13 # "Chicken broth"
FP20	*<hasIngredient>*	FP14 # "Cubed cooked chicken"
FP20	*<hasIngredient>*	FP11 # "Diced carrots"
FP20	*<hasIngredient>*	FP12 # "Cut green beans"
FP20	*<hasIngredient>*	FP17 # "Noodles"
FP20	*<hasIngredient>*	FP18 # "Flavoring"
FP20	*<hasIngredient, w/ purpose>*	FP19 # "BHT", Preservation
FP20	*<processedBy, w/intensity, w purpose>*	Boiling, Fully cooked, {Make edible, Preservation}
FP20	*<hasForm>*	LiquidWithSolidChunks

FP21	***<hasName>***	**"Diced parsley"** # statements not shown
FP22	***<hasName>***	**"Campbell's Chicken Noodle Soup"**
FP22	*<isa>*	FP20 # "Chicken noodle soup
[FP22	*<hasIngredient>*	FP13 # "Chicken broth"**]**
[FP22	*<hasIngredient>*	FP14 # "Cubed cooked chicken meat"**]**
[FP22	*<hasIngredient>*	FP11 # "Diced carrots"**]**
[FP22	*<hasIngredient>*	FP12# " Cut green beans"**]**
FP22	*<hasIngredient>*	FP21 # "Diced parsley" # specific to "Campbell's Chicken Noodle Soup"
[FP22	*<hasIngredient>*	FP17 # "Noodles"**]**
[FP22	*<hasIngredient>*	FP 18 # "Flavoring"**]**
[FP22	*<hasIngredient,*	FP 19 # "BHT",
	w/purpose>	Preservation**]**
[FP22	*<processedBy,*	Boiling,
	w/intensity,	Fully cooked,
	w/*purpose>*	{Make edible, Preservation}**]**
FP22	*<packedIn>*	Steel can # specific to "Campbell's Chicken Noodle Soup"

[ ] inherited from one level above, **[[ ]]** inherited from two levels above

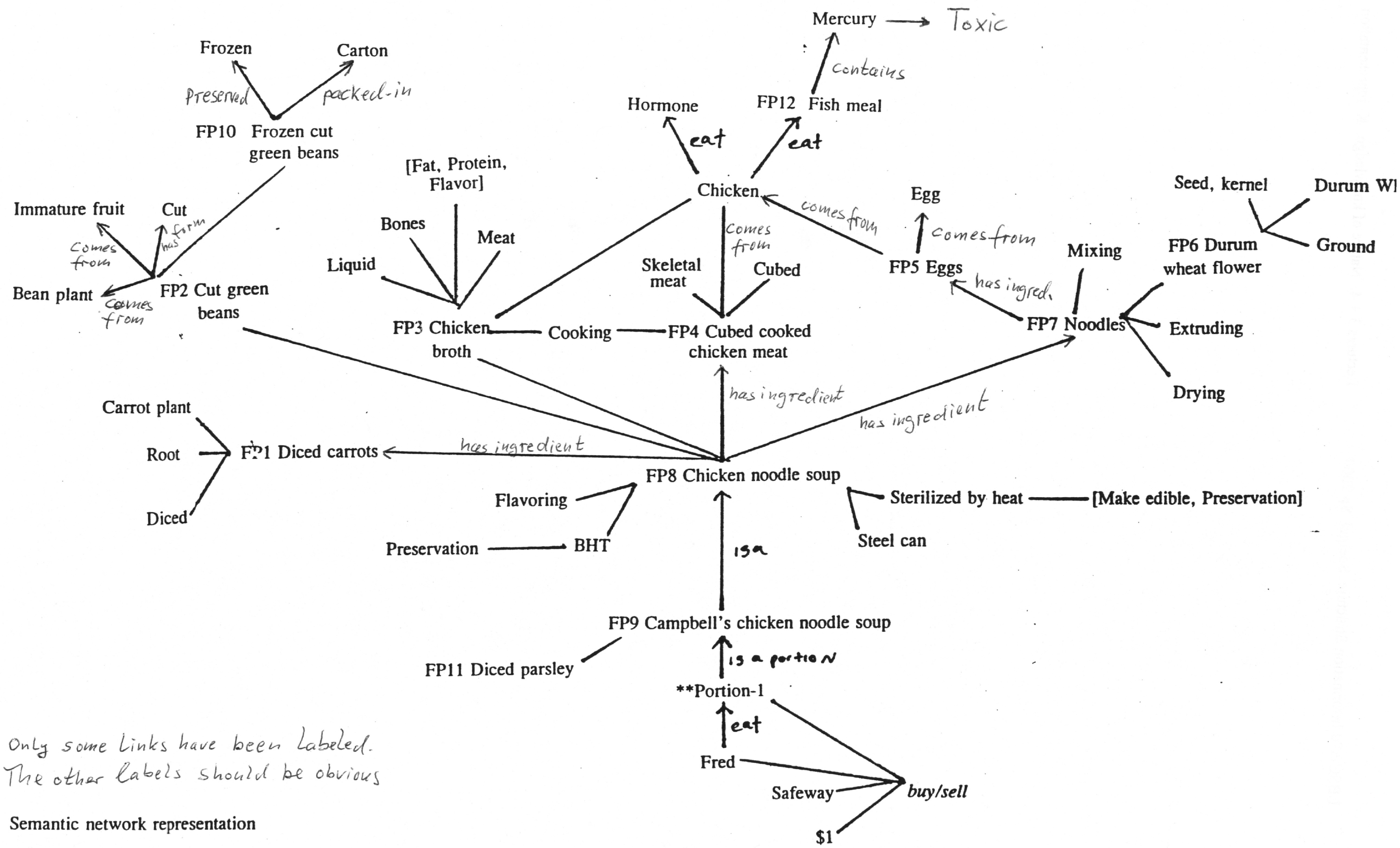
Portion-1	***<isaPortionOf>***	**FP22 # "Campbell's chicken noodle soup"**
[[Portion-1	*<hasIngredient>*	FP13 # "Chicken broth"**]]**
[[Portion-1	*<hasIngredient>*	FP14 # "Cubed cooked chicken meat"**]]**
[[Portion-1	*<hasIngredient>*	FP11 # "Diced carrots"**]]**
[[Portion-1	*<hasIngredient>*	FP12 # "Cut green beans"**]]**
[Portion-1	*<hasIngredient>*	FP21 # "Diced parsley"**]**
[[Portion-1	*<hasIngredient>*	FP17 # "Noodles"**]]**
[[Portion-1	*<hasIngredient>*	FP18 # "Flavoring"]]
[[Portion-1	*<hasIngredient,*	FP19 # "BHT" *purpose* Preservation**]]**
[[Portion-1	*<processedBy,*	Boiling
	w/purpose >	{Make edible, Preservation}**]]**
[Portion-1	*<packedIn>*	(Steel can, 10 fl oz)**]**

FP23	***<hasName>***	**"Frozen cut green beans"**
FP23	*<isa>*	FP12 # "Cut green beans"
[FP23	*<comesFromSource>*	Bean plant**]**
[FP23	*<comesFromPart>*	Immature fruit**]**
[FP23	*<hasForm>*	Cut**]**
FP23	*<processedBy,*	Freezing
	w/ purpose >	Preservation
FP23	*<packedIn>*	Carton

***<buy/sell>* (Safeway, Fred, Portion-1, $1)**

Fred	*<eats>*	Portion-1

FP24	***<hasName>***	**"Fish meal"**
FP24	*<contains>*	Mercury
Chicken	*<eats>*	Hormone
Chicken	*<eats>*	FP24 # "Fish meal"



Only some links have been labeled.
The other labels should be obvious

Semantic network representation

Large example. Food product data in frame representation

Just a few examples are shown. Continue on next page

Large example. Food product data in frame representation

A frame is a packet of information consisting of many "compartments" called *slots*; each slot specifies a type of information. Each compartment is filled with the appropriate piece of information, called the *slot filler*.

There are two types of frames: **Extended frames** and **minimal frames**.

An **extended frame** collects many pieces on information about one entity; it is entity-focused. You have seen examples of extended frames in the bird example. Two examples are given on the next page:

a full frame instance (including all information, including what is inherited) for FP20
a short frame instance excluding inherited information, for FP22

An extended frame packages many statements about the same entity, the head of the frame. Each statement is based on a relationship type, usually two-way relationships. The left argument of the relationship is the head of the frame, the frame slot specifies the relationship type, and the slot filler is is the right argument is the slot filler:

Head of the frame	Slot	Slot filler
FP 20	*<hasForm>*	Liquid with solid particles

The pieces of information in an extended frame could be stored as separate statements.

The frame instance for FP22 takes very little space yet gives a lot of information because it takes advantage of hierarchical inheritance. The *<isa>* (*<inheritsFrom>*) FP20 packs a lot of punch: It pulls in all the information given in the frame for FP20.

Looking ahead: A frame is like a record (or a row in a relational table or an object in a object-oriented database), the frame slots correspond to the data fields in a record (or the columns in a relational table). But frames have more power than plain records. Frame systems (like object oriented databases) implement hierarchical inheritance

Slots in a frame often correspond to **facets** in a faceted classification. The frame for FP22 shows correspondence to the facets from the paper on a food description language.

Another example of hierarchical inheritance: Recipes in a cookbook

The cookbook may have a basic recipe for *basic potato soup*. That recipe may take up two pages, explaining how to select the best potatoes, how long to cook them, how to mash them up, how much water to add, etc. Then there may be 20 recipes for specialty potato soups. Each of these takes just a few lines. For example, the recipe for *leek potato soup* might say simply "Use recipe for *basic potato soup*. Cut four medium leeks into small pieces and add to liquid potato soup before cooking." This saves printing the two pages on how to make basic potato soup twenty more times.

Head of frame: FP20		**Facet**
Slot	**Slot filler**	
<hasName>	**"Chicken noodle soup"**	
<isa> (<inheritsFrom>):	FP5 # "Soup"	*A Product type*
# Slots dealing with food origin	N/A	
<comesFromSource>	N/A	*B1 Food source*
<comesFromPart>	N/A	*B2 Part*
<isExtractedSubstance>	N/A	
<isMadeFrom>		
<hasIngredient>	FP13 # "Chicken broth"	*B3 Ingredient*
. . .		
<hasIngredient, w/*purpose> Preservation*	FP19 # " BHT"	*D4 Method of preservation*
# End food origin		
<processedBy,	Boiling	*D2 Cooking method*
withIntensity>	Fully cooked	*D1 Degree of preparation*
<processedBy, *w/purpose> Preservation*	Boiling # sterilizing by heat	*D4 Method of preservation*
<hasForm>:	Liquid with solid particles	*C Phys. state, shape, form*
<packedIn>	N/A	*E2 Container, wrapping*

Head of frame: FP22 (inherits most of its information from FP20; inherited slots are not repeated, saving lots of space)	
Slot	**Value**
<hasName> *<isa>* *<hasIngredient>* *<packedIn>*	**Campbell's Chicken Noodle Soup** FP20 Chicken noodle soup FP21 Diced parsley Steel can

Think of this type of inheritance in the context of recipes in a cookbook.

A minimal frame: **Instance of a frame for buy/sell** (a relationship with four arguments)
A minimal frame gives information for one relationship instance; it focuses on the relationship (whereas an extended frame focuses on the head entity).

Note that buy and sell are different names for the same underlying event:

> **buy** is used as the name of the event if the emphasis is on the buyer (the receiver of goods or services)
> **sell** is used as the name of the event if the emphasis is on the seller (the source of goods or services)

buy/sell	
*Source*OfGoodOrService / *Receiver*OfMoney*:*	Safeway
*Receiver*OfGoodOrService / *Source*OfMoney*:*	Fred
GoodOrService:	Portion-1 (a particular can of Campbell's chicken noodle soup)
MoneyAmount:	$1

Relationships can relate two, three, four, or more pieces of information (called arguments) needed to make a complete statement. In the entity-relationship version, we wrote the same information as:

<*buy/sell*> (Safeway, Fred, Portion-1, $1)

The frame is just a different way of writing this, specifying the role each piece of information plays. All slots are essential; each value depends on all the others. The same information cannot be expressed in separate statements (as is the case for extended frames). That is why the frame is called **minimal**.

Linguists specify for each verb or group of verbs the slots that must be filled in order to make a complete statement with the verb; they call this specification a **case frame**. So the above is a case frame for the verb buy and equally for the verb sell.

New topic: Example for activation of all frame elements when one slot value is activated.

Assume you stand at a street corner, opposite is a store with a little clock tower, a lady in fancy hat walks by. All of a sudden a red sports car appears and hits a pedestrian crossing the street. You store the whole scene in your mind as frame with slots:

Frame: Street scene	
General scenery:	Street corner
Building characteristics	Store with small clock tower
People:	Lady in crazy hat, Person crossing street, many people walking by
Cars	red sports car
Events	red sport car hits pedestrian
. . .	

Two years later you see a red sports car, activating one element in your street scene. The scene you saw two years ago flashes before your eyes; activation of one frame element activates the whole frame, very similar to spreading activation.

2 Approaches to knowledge representation

Summary of concepts covered in examples

Entity-relationship approach

Semantic networks

Frames

Role of frames

Grid for data acquisition

Template for data output (for example, city data frame in Wikipedia)

Activation of all frame elements when one element is activated
(Seeing *parsley* may activate in a person's mind the whole frame for *Campbell's Chicken Noodle Soup*)

Types of frames

Extended frame (DS term)

An extended frame includes additional slots that represent further relationships, usually binary relationships from the focal entity to other entities.

Minimal frames (DS term)

A minimal frame represents a multi-way (n-ary) relationship – each slot corresponds to one argument of the relation. No slot could be omitted without making the frame incomplete, that is, making at least one other slot value indeterminate.

3 Mechanisms in knowledge representation

Spreading activation

Hierarchical inheritance

Restrictions on values

Default values (for example, telephone area code in the database of a local charity)

Procedural attachments (procedures to be called when data are entered in the slot)

4 Criteria for describing and evaluating knowledge representations

(advanced) See supplement SLecture 2.2

red

Week 3. February 3 - February 10

Part 2.
Structure and evaluation of information systems

red

Things to do in Week 3, W February 3 - February 10

	Assignments due W. Feb. 10 ☐ required ○ optional	√
Review answer key(s)	None	
Assignment(s)	Assignment 1: Hypermedia exploration: Perseus and Freebase (2.5 h) (ass. `Jan. 27, L1.2)	☐
	Assignment 2: Bibliogr. retrieval system exploration.: MEDLINE (3 h)(ass. `Jan. 27, L1.2)	☐
	Assignment 3: Online catalog search exercise (1.5 hours) (ass. `Jan. 27, L1.2)	☐
	Assignment 4: Restructure set of data. Restructured DB. ☐ network ☐ (1.5h) (ass. `Feb. 3, L2.2)	☐

New topics this week

	3.1 The structure of information systems	
Readings	1 **Lecture 3.1 Objectives etc**. (pink). Also have a look at Assignment 5	☐
	2 Textbook **Ch. 5 The Structure of Information Systems**	☐
Lecture	**Lecture 3.1 slides**	☐

	3.2 Objectives and performance measures for info. systems	
Readings	1 Lecture 3.2 Objectives etc. (pink)	☐
	2 Textbook Ch. 6 **Systems Analysis**	☐
	3 Textbook Ch. 7 **Assessment of users' problems and needs**	☐
	4 Textbook Ch. 8 **Objectives and performance measures for ISAR systems** (for discussion)	☐
Lecture	**Lecture 3.2 slides** (80 min)	☐

Learning blog	Learning blog Week 3 due February 10	○

Assignments assigned February 10

Assignments assigned	‣**Assignment 5, Analytical description of an information system** (3 hours) (due `Feb. 17)

red

pink

Part 2.

February 10

Structure and evaluation of information systems

Lecture 3.1

February 10

The structure of information systems (Textbook, Section 5.1)

Learning objectives	1 **Know and understand the functional components of information systems** and be able to use this framework (P2.0.1) 1.1 to analyze and design information systems (P2.0.1,1) 1.2 to integrate the subject matter from this and other courses. (P2.0.1,2) 2 **Draw specific implications on the design and operation of specific functions in an information system.** Be inspired to be proactive in finding out about information needs and take an active role in discerning information needs 2.1 Be inspired to be proactive in finding out about information needs. (P2.1.0,1) 2.2 Understand the importance of determining requirements without being hemmed in by constraints of present systems or present technology. (P2.1.0,2) 2.3 Be inspired to play an assertive professional role in helping users determine their true information needs. (P2.1.0,3) 2.4 Appreciate the need for understanding the relevance criteria of a specific user in a particular situation for conducting a good search. (P2.1.0,4). 2.5 Understand the added value of post-search information analysis by software and/or an information professional. (P2.5.2,5.1) 3 **Understand the wide variety of information systems** (P2.0.2)
Practical significance	• To **design, operate, or use an information system** or a specific function in it, you must understand the information system components, their inputs, output, and functioning. • To **take advantage of all available career opportunities**, you must understand the multitude of information systems and information environments in which the knowledge and skills acquired in UBLIS can be applied. • The information system diagram provides a framework for organizing information from many courses.

Additional examples of system descriptions in supplement, SLecture 3.1. Also useful for Assignment 5 System description.

pink

In-lecture exercises (in preparation for Assignment 5, where you analyze an information system)

See Assignment 5, Assignments p. ~61, use template on Assignments p. - **63**

Additional question (used on p. ~92, put here to save space)

Determine the percentage of resources allocated to each of the following functions (Refer to Section 2.6 of the Textbook).

(1a) Assist the user in identifying relevant documents (intellectual access). (A user is given a list of references to documents relevant for her problem or topic. She must then consult these documents and extract the information needed.)

(1b) Make available known documents (physical access). (A user requests specific documents, often documents found through 1a, and is loaned or given copies.)

(2) Provide tailor-made packages of substantive data. (A user needs information on a certain topic and is given a report that contains just the information she needs, neither more nor less. This report may be prepared by information center staff or a computer program by extracting information from documents or it may be the result of a search of a substantive database.) (See Section 2.5 of the Textbook for an explanation of "substantive data".)

What kind of data do you need to answer this question?

Refer to the figure on the following (facing) page (which conveniently combines Textbook Figures 5.1c, Textbook p. 47 and 5.6, Textbook p. 58). Using this figure, analyze the following:

1 A **Web search system** (like Google or Yahoo) as an information system.

2 A **special library** as an information system.
Question: How does interlibrary loan fit into the information system framework?

3 The **production and use of a textbook** in the information system framework.

To analyze Google as an information system using the structure on p. ~97, look at each box and determine what Google does using what files (data) and rules

Same for Yahoo Directory, http://dir.yahoo.com/
Short description: http://help.yahoo.com/l/us/yahoo/directory/basics/basics-03.html

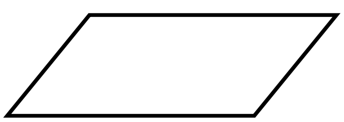
For online: Post your results to the discussion board for Week 3. The recording will take you through some of this analysis

Additional question: Determine the percentage of resources allocated to each of the following functions See Section 2.6 of the Textbook and see elaboration on p. ~97, put there to save space).

Supplement SLecture 3.1 gives three examples of system descriptions, also useful for the system description assignment.

Note on diagram conventions

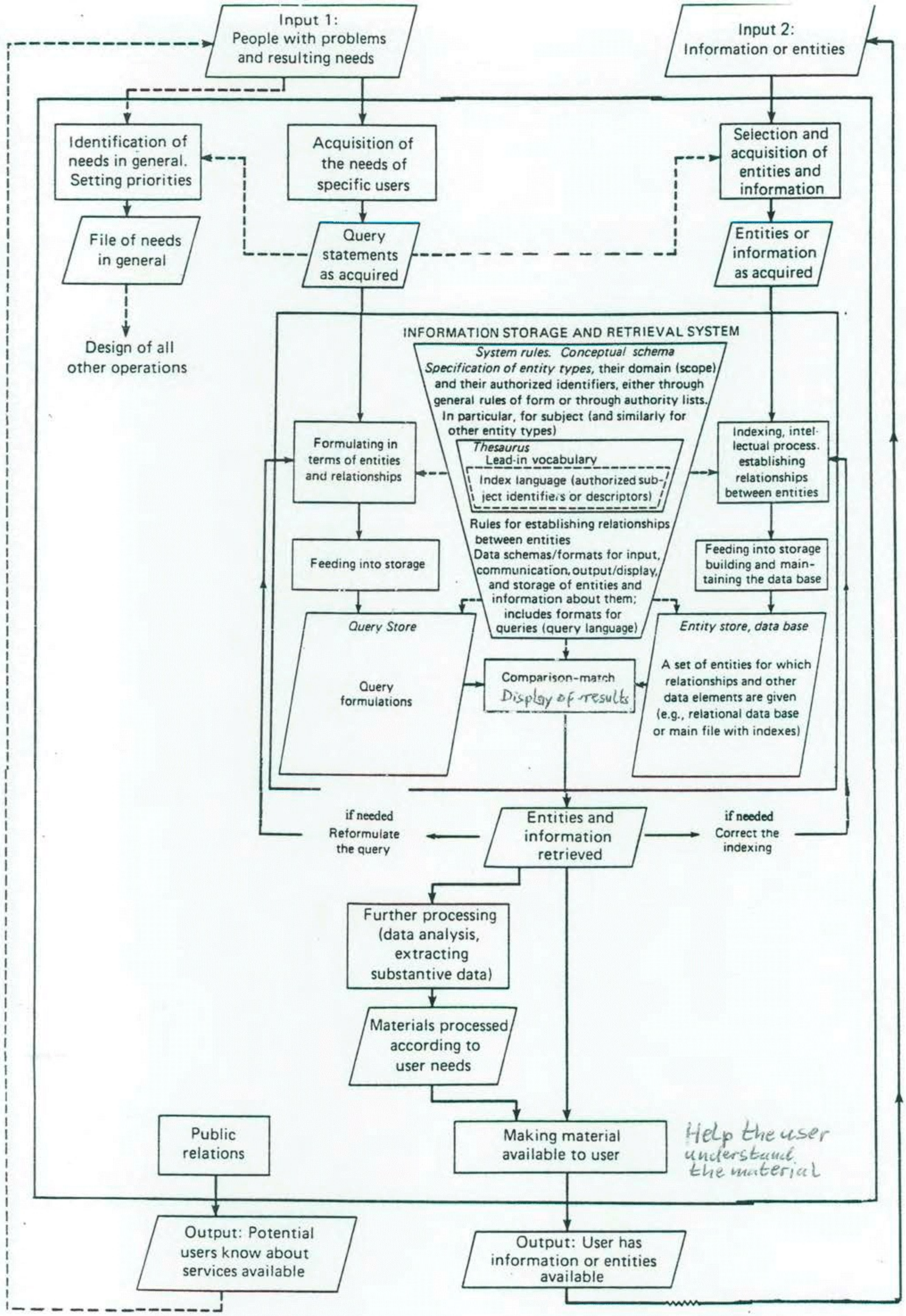
Denotes a file, data, inputs and outputs. Could be a group of people and their problems as an input



Denotes a process that transforms inputs into outputs

Sequence of processes and files, flow of data

Control of processes or file organization



Text Figures 5.1c and 5.6 combined
Repeated from page 58 Lecture 3.1

pink **Lecture 3.2** ***February 10***

Objectives and performance measures for information systems

(Organizing. Information, Chapter 8)

<table>
<tr><td>Learning objectives</td><td>
1 Understand the objectives of information systems and associated performance measures. Be aware both of the difficulties of defining objectives and measures and of the importance of having clearly defined objectives and measures. (P2.0.4)

2 Be able to apply objectives and measures to (P2.0.4,1)
• design, evaluation, planning, and monitoring operations;
• communicating requirements to systems analysts
• the evaluation of an information system as a whole and the analysis of the contributions of individual information system components;
• the selection of information systems (reference tools, databases, search engines, library software etc.) for acquisition (including training for use);
• the conduct of individual searches, including
    • the specification of individual user/search requirements;
    • the selection of an information system (database and search system) that can be expected to meet these requirements;
    • the determination of optimal search effort (resources to be allocated to the search);
    • the evaluation of search results and determining when to stop searching.

3 Specifically, understand retrieval performance measures and be able to apply these to the tasks specified in Objective 1. (P2.0.4; P2.5.2,2)
</td></tr>
<tr><td>Practical significance</td><td>Incorporated into the statement of objectives.

See Textbook Chapter 8, Introduction, and Section 8.5 for elaboration.</td></tr>
</table>

Discussion questions: see next page

pink

Discussion questions

1 Consider the definition of relevance and of performance measures in general in the context of an information system with data on the structure of a nuclear power plant to be used in case of malfunctions. The system gives detailed information about all components down to the last pipe and valve, their functions and interrelationships.

2 Consider performance measures for the following information system. The purpose of the information system is to assist in solving crimes. The system stores reports of past crimes — both solved and unsolved — and indexes them by various features of the modus operandi. To use the system, the detective formulates a query based on the features of the unsolved crime. The system provides reports of similar crimes; these might assist in solving the crime in question.

3 Consider a Web search service (such as AltaVista or Lycos) that produces ranked retrieval output. Picture two users. User 1 needs a quick answer to a question, and user 2 needs a comprehensive list of materials (for example, a listing of all classifications schemes and thesauri available on the Web). What performance measures would be appropriate for each type of user?

Textbook, Chapter 8 review, especially

The need for performance measures for guiding system design, Figure 8.1, Textbook p. 111

Deriving performance measures **for bibliographic information systems**,
Figures 8.2 and 8.3.
Figure 8.3 updated on Lecture Notes p. ~102 (next page)

The concept of relevance, Textbook Section 8.4

Relates to ➔UBLIS 518 Reference Sources and Services, important for how to search

Have the Textbook open to p. 111

Have the lecture notes open to Figure 8.3, p. ~98

The discussion of Figure 8.3 is useful for understanding how a user processes information and therefore has implications for pre-processing and arranging search results

For online, listen to audio

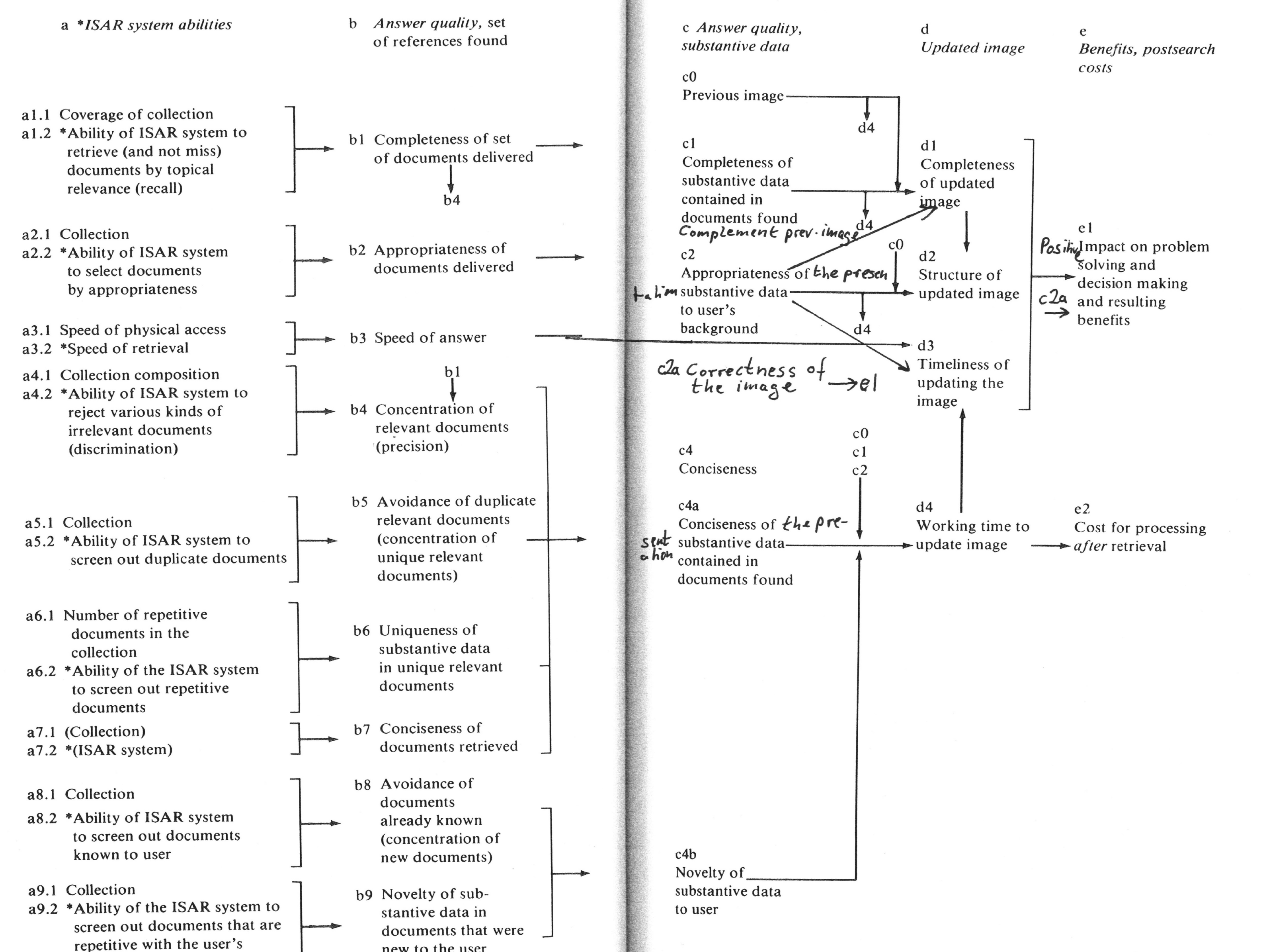


Fig. 8.3 Derivation of a local performance measure for a bibliographic ISAR system.

Relevance criteria for different purposes

Relevance criteria of teachers selecting oral history materials for a lesson

Based on a study of teachers using materials from the Shoah Foundations archive of videotaped interviews with Holocaust survivors. http://sfi.usc.edu/

- **Relevant to teaching content and method**
  - **Relationship to lesson theme**
  - **As part of broader curriculum**
    - Relates to other schoolwork
    - Variety for the classroom
    - Vocabulary
  - **Characteristics of oral history interviews**
    - Flow of interview
    - Expressive power
    - Language & verbal expression
    - Diction
    - Nonverbal communication
  - **Characteristics of the story**
    - Positive message for students
    - Role of interviewee in Holocaust events
  - **Relationship of story to student**
    - Students connect with passage
    - Students identify with interviewee (e.g., based on demographic characteristics)
    - Radical difference from students' reality
- **Represents different populations**
  - Race
  - Age of interviewee during Holocaust events
- **Appropriateness**
  - Developmental appropriateness
  - Acceptability to stakeholders
- Technical production quality
- Length-to-contribution ratio

Topical relevance for scholars

Types of topical relevance	
Topic: Food in Auschwitz	
Relevance type	Example
"Classical relevance" (TREC definition)	
• Provides direct evidence	Describes types of food and portions served
• Provides indirect/circumstantial evidence	Describes undernourished people
Additional relevance types	
• Provides context	(1) Reports on intensity of manual labor (2) Availability of food in the area around the camp
• Useful as a basis for comparison	Food situation in a different camp
Pointer relevance	
• Provides pointer to a source of information (The information pointed to can be relevant in any of the ways listed above)	Mention of a study on the subject

TREC = **T**ext **RE**trieval **C**onference
A yearly competition of information retrieval systems performing specified retrieval tasks on a given test collection held at NIST (National Institute of Standards and Technology)

CLEF = **C**ross-**L**anguage **E**valuation **F**orum
The corresponding activity in Europe. Uses the MALACH speech retrieval test collection for one task.

red

Weeks 4 - 5. February 10 - February 24

Part 3.
Information Retrieval
General principles and methods

red

Things to do in Week 4, W February 10 - February 17

	Assignments due W February 17 ☐ **required** ○ **optional**	√
Review answer key(s)	Assignment 1: Hypermedia exploration: Perseus and Freebase (assigned L1.2, `Jan. 27)	☐
	Assignment 2: Bibliographic retrieval system exploration: MEDLINE (assgnd L1.2, `Jan. 27)	☐
	Assignment 3: Online catalog search exercise (assigned. L1.2, `Jan. 27)	☐
	Assignment 4: Restructuring a set of data in db and sem. net representation (ass.. L2.2, `Feb. 3)	☐
Assignment(s)	Assignment 5: Analytical description of an information system (3 hrs) (assgnd. L3.1, `Feb. 10)	☐

	New topics this week

	4.1 Searching linked data. Integrated information structure model	
Readings	1 **Lecture 4.1 Objectives etc**. (pink). Also have a look at Assignment 6.	☐
	2 Soergel **Searching Linked Data** Prologue and p. 1 - 17, **have at hand**	☐
	3 Soergel **Language of Foods**	○
Lecture	**Lecture 4.1 slides** (65 min)	☐

	4.2a Data schemas and formats. Review	
Readings	1 **Lecture 4.2a Objectives etc**. (pink)	☐
	2 Textbook Ch. 3 **Structure Of Information**	☐
	3 Textbook Ch. 9 **Data Schemas and Formats**	☐
	4 **Model Catalo**g (very first Reading in packet) (useful to look at)	☐
Lecture	**Lecture 4.2a slides** (30 min)	☐

	4.2b Data schemas and formats. In-lecture exercise	
Lecture	**Lecture 4.2b slides** (30 min)	☐

over

Learning blog	Learning blog Week 4 due February 17	○

	Assignments assigned W Feb. 17	
Assignments assigned	‣**Assignment 6, Developing a conceptual data schema** (1.5 hours)(due `Feb. 24)  ‣**Feedback quiz** (30 min.)(due `Feb. 24)	

pink

Part 3

February 17 - 24

Information retrieval. General principles and methods

Lecture 4.1 *February 17*

Searching linked data. An integrated information structure model

Note: Linked data are data stored on the Web as "triples", statements that link two entities through a relationship as illustrated in Lecture 2.2.

Learning objectives	1 **Understand a general model of information retrieval** based on the entity-relationship approach. (P2.3.1,1.2) 2 **Be able to use this general model to** (P2.3.1,1.2) 2.1 recognize common principles across types of information systems and develop an overall vision of retrieval possibilities 2.2 analyze specific systems and information retrieval operations in terms of this general model. 2.3 use existing systems in new and imaginative ways, in particular, to use several different systems in synergistic ways for more sophisticated searches. (P2.3.2,1) 2.4 design new systems with increased power, for example to search for Linked Open Data on the Web, see Lecture 5.1.
Practical significance	Understanding the principles discussed in this lecture will make you a better searcher. You may never be able to use a unified integrated system of the kind described. **But you can use existing systems in combination to achieve improved search results enabled by the way of thinking presented here**. In other words, you can build your own “virtual” integrated information structure whenever a search requires it.

Key idea:

Combining different kinds of facts to find an answer. (Inference. Chaining)

Done by people - reference librarian or user consults different databases as needed to find all the facts needed to construct an answer, often chaining data from different databases together.

Done by systems - all facts must be accessible to the system. This could be a system that accesses multiple databases.

Relates to and elaborates on **Lecture 1.2 Information systems and information structure**

Also read beforehand p. ~110 and also p. ~111-112 (that is why they are pink)

The reading *A general model for searching linked data* serves as lecture notes for this lecture.

The reading presents a powerful conceptual framework for searching linked data. Lecture 5.1 introduces SPARQL, an actual but less powerful query language that is a W3C standard.

pink

A general model for searching linked data.

Part 1. Basic structure and search commands

pink This recorded lecture will present the material from the reading

A general model for searching linked data
OR Design of an integrated information structure interface

as if it was a presentation at a conference. **Please read the Prologue and p. 1 - 17 of this reading beforehand**; these sections give examples. The lecture does not assume that you have read the remainder of this reading; rather, the reading is a back-up reference.

Restatement of the objectives: Through this lecture you should

- Get a better understanding of entity-relationship data modeling
- Improve your skills as a reference librarian through better understanding of chained searching, often using multiple data sources
- Acquire a general framework for understanding retrieval system features such as inclusive (hierarchically expanded) searching.
- Get a forward-looking sense what retrieval systems could do in the new world of Linked Open Data (LOD)

Recommended exercise before listening to the lecture

As an example for the dimensions given in the reading Figure 7, p. 17 (and also examples of the kinds on the searches discussed in the reading), you could do search in the UB Libraries catalog:

Type *Library instruction* in the search box
Starting point: a search key, entered by the user

Scroll down to
Bibliographic instruction : a handbook /
by Renford, Beverly. Published 1980

Click on the title to see the full record. Many of the elements of the records can be used as starting points for a new search. Rather than having to enter a person's name to find documents he or she authored, you can just click on the name displayed on the screen to start an author search. The result is no different from doing the same search by entering the person's name in the search box as author (Dimension 2). Try it.

You can start a search for the subject *Library orientation > Handbooks, manuals, etc.*
By clicking on the second element, *Handbooks, manuals, etc*

See what happens if you click on *Library orientation*

pink

Note: When you click on *Library orientation*, the UB catalog does a search

Subject Library orientation <*isSubjectHeadingFor*> Document X

(to be precise: the system finds all documents for which the words library and orientation occur among all assigned subject headings)

That is all a user can do. A user may want to find subjects that are in some way related to Library orientation, which would be the search

Subject Library orientation < *hasST OR hasBT OR hasNT OR hasRT*> Subject X

ST	*hasSynonymousTerm*
BT	*hasBroaderTerm*
NT	*hasNarrowerTerm*
RT	*hasRelatedTerm*

There is no provision in the UB catalog to do that (but you will later learn how to use Library of Congress authority databases to do that kind of search). The UB catalog could, upon the user clicking on Library orientation, ask the user

Do you want to find

Books that have the words library and orientation in their subject headings
Books that have the words library and orientation in their subject headings or title
Subject headings related to Library orientation
Courses on Library orientation

Now you can acquire an understanding of the general system proposed by listening or reading or both.

- When listening to the lecture, be sure to have the reading in front of you.
- You can also just read p. 18 - 50 of the reading

pink **Lecture 4.2a+b** ***February 17***

Conceptual data schemas and input, storage, and output/presentation formats

(Textbook, Sections 9.1, 9.2, 9.4, and 9.5)

Learning objectives	1 **Be able to analyze or design the conceptual data schema** of an information system, specifically using the entity-relationship approach. (P2.3.1,1.4#; P2.3.4) 1.1 analyze the conceptual data schema underlying an info system (P2.3.4,1); 1.2 judge the adequacy of this schema with respect to the queries to be answered (P2.3.4,1); 1.3 use the knowledge of the schema to exploit fully the possibilities of obtaining answers from the information system (P2.3.4,1); 1.4 select or design a conceptual data schema for an information system based on user requirements. (P2.3.4,2) 2 **Be able to analyze and select or design the input formats and output formats** used to interact with an information system: (P2.3.1,4.12) • input formats that make data entry complete, error-free, and easy; • output formats (for reports, such as recurring bibliographies, or the display of search results) that contain all the information needed (and no more) in an easy-to-read form. **3 Understand the structure of a MARC record** (P2.3.6,2)
Practical significance	1 **For designing information systems:** The success of any information system depend(s vitally on the complete-ness of the information included. The conceptual data schema determines what information can be included in the system and what information is elicited from the people that enter data into the system. Input and output formats determine how easy it is to interact with the system. 2 **For using information systems (including reference tools):** **To get the most out of an information system** in terms of being able to do different types of searches, you need to know its conceptual data schema. **To select the appropriate information system**, you need to be familiar with the conceptual data schemas of many information systems. **To do the kind of power search** that draws on multiple information systems simultaneously requires even more knowledge of conceptual data schemas.

Schema Arrangement of parts in some order, showing interrelationships.

This topic is closely related to document structure&design, to be discussed in Lectures 5.2-6.1.

The conceptual data schema is, as the name says, at the conceptual or logical level. To implement a database following a conceptual data schema one uses a database management system (DBMS), such as Microsoft Access, FileMaker Pro, or ORACLE. This is a topic in UBLIS 506 Information technolog

See SAssignment 6a for an explanation and exercise of how to implement in the University Database in MS Access.

Over

pink

Lecture 4.2a. Data schemas and formats. Review (30 minutes)

Starting on the next page

Listen to the audio for Lecture 4.2a.
Note: The audio gives the starting page as 78, but it should be p. ~114. Add 36 to any other page number mentioned.

The audio also says the exercise is done first; ignore that. Do Lecture 4.2a (Notes + audio) first, then proceed to the exercise Lecture 4.2b.

Uses of the different types of information in an information system

A type of information (a fact type, see Lecture 1.2), as defined by a relationship type, may be used for one or more of the following functions.

- **Retrieval, drawing inferences, statistical analysis**

  Example: From drug prescription expert system

  Disease <*treated with*> Drug

  This piece of data is used for
  - plain retrieval of medical knowledge;
  - inference in conjunction with patient data.

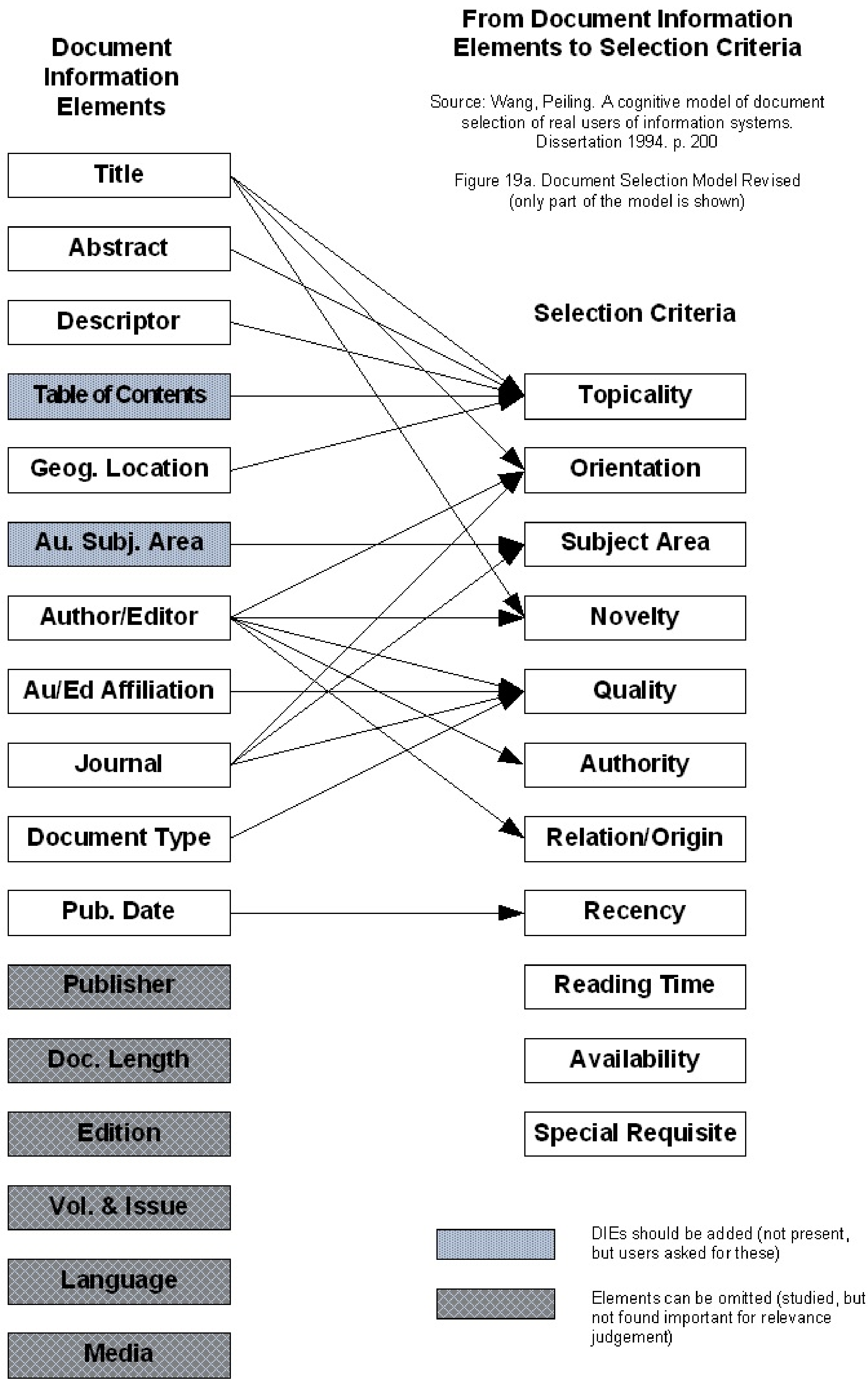
- **Arranging retrieval output**
  - Example: Arranging a long list of document records retrieved from an OPAC (Online Public Access Catalog) by call number
  - Example: Arranging output from a Web search by URL (Uniform Resource Locator), which would bring pieces of one Web document that consists of several pages together in the output list

- **Providing information to the user**, either the substantive data sought or information about a document that enables the user to judge the relevance of the document.

The conceptual data schema designer must weigh the cost for including a type of information against the benefit in terms of these three functions.

As an example, look at Wang's list of "Document Information Elements" users considered in document selection or wished to have available (next page). These results should be used as a guideline in systems design when deciding what information to include in the system and what information to present to users in the output format. Users did consider information elements that are linked to the document indirectly, such as the subject area of the document author. As will be further elaborated below, the information system must assemble all this information about a document, possibly obtaining information from other databases, such as a database about persons.

Facing page:
Use of "Document Information Elements" in decisions on relevance and document selection


Document Information Elements
From Document Information Elements to Selection Criteria
Source: Wang, Peiling. A cognitive model of document selection of real users of information systems. Dissertation 1994. p. 200
Figure 19a. Document Selection Model Revised (only part of the model is shown)
Title
Abstract
Descriptor
Table of Contents
Geog. Location
Au. Subj. Area
Author/Editor
Au/Ed Affiliation
Journal
Document Type
Pub. Date
Publisher
Doc. Length
Edition
Vol. & Issue
Language
Media
Selection Criteria
Topicality
Orientation
Subject Area
Novelty
Quality
Authority
Relation/Origin
Recency
Reading Time
Availability
Special Requisite
DIEs should be added (not present, but users asked for these)
Elements can be omitted (studied, but not found important for relevance judgement)

From conceptual data schema to records (Textbook, Section 9.2)

Record format	• A record is an assembly of information about a given entity, such as an event, a person, or a document for input, storage, communication, or display. • The **record format** determines how the different pieces of information are arranged in the record. • A record is a simple frame. Slot in a frame = data field in a record. • The evaluation criteria for schemes of knowledge representation (Lecture 2.2, Section 3.4) apply to records; see also Textbook, Section 9.3. • Many records are extended frames: they incorporate many binary statements that link the focal entity of the record to another entity. Each statement could stand on its own. The presentation of these data in a record is more concise and may be more intuitive and more easily grasped than a series of statements. See the MARC record format (facing page) and the examples in the Textbook Ch. 9.
Input record. Input format	The best way for eliciting input from the system operator (for example, a cataloger) is often an **input record** , a form with blanks (slots) to fill in.
Storage record. Storage format	Some systems store data internally in tables (relational database), where information about a given entity may be distributed over several tables. Other systems store data internally in records or frames, assembling all information about one entity, e.g., a book, in one place (but making it more difficult to assemble information about entities of other types, e.g., persons).
Communica-tion record	A record in a common **communication format** to transfer data from one system to another. Each system may use its own internal format. Examples: MARC, Z39.50
Output/ display record. Display format	To present information about an entity in a format easily understood by the user, the information system must assemble the desired types of information into an output record. The information elements in the output record may be linked to the entity directly (for example, the direct link between a document and the person who authored it) or indirectly (for example, the indirect link between a document and the organization with which the author is affiliated).

MARC format	Opposite is a sample record format, the MARC (MAchine-Readable Catalog) format. (MARC was developed by the Library of Congress starting in 1962 for the interchange of bibliographic data and has become a widely used standard). For each data field, the corresponding statement template (relationship type plus entity types related) is given. A more complete list of MARC fields is found with the description of the model catalog (first in the general package). When a record format, such as MARC, exists, one can "re-engineer" the E-R conceptual data schema. In the design of a new system, one should start with the E-R schema and design record formats for input and display.

Correspondence between data fields in a MARC record and relationship types (simplified)

MARC field		Rela tionship type
100	Main Entry-Personal Name	Document <*hasCreator*> **Person** (who is main entry)
110	Main Entry-Corporate Name	Document <*hasCreator*> **Organization**
111	Main Entry-Meeting Name	Document <*hasCreator*> **Meeting**
130	Main Entry-Uniform Title	Document <*hasTitle*> **Text** (if Title is main entry)
245	Title Statement	Document <*hasTitle*> **Text**
260	Publication, Distribution, etc. (Imprint)	Document <*hasPublishery*> **LegalEntity** (Person or Org.) this Organization <*locatedIn* **Place** (chain) Document <*hasPublicationDate*> **Date**, ... (distinguished by using subfields)
490	Series Statement	Document <*partOf*> **Document** (which is a Series)
600	Subject Added Entry-Personal Name	Document <*hasSubject*> **Person**
610	Subject Added Entry-Corporate Name	Document <*hasSubject*> **Organization**
611	Subject Added Entry-Meeting Name	Document <*hasSubject*> **Meeting**
630	Subject Added Entry-Uniform Title	Document <*hasSubject*> **Document**
650	Subject Added Entry-Topical Term	Document <*hasSubject*> **Topic**
651	Subject Added Entry-Geographic Name	Document <*hasCoverage*> **Place**
700	Added Entry--Personal Name	Document <*authoredBy*> **Person**
710	Added Entry-Corporate Name	Document <*emanatedFrom*> **Organization**
711	Added Entry-Meeting Name	Document <*emanatedFrom*> **Meeting**
730	Added Entry-Uniform Title	Document <*hasTitle*> **Text**
850	Holding Institution	Document <*heldBy*> **Organization**

The MARC record is an extended frame with information about a document; it incorporates many statements that link the document to some other entity (a Person, a Date, a Subject, etc.). Each statement could stand on its own; the data in the MARC record could be represented as a series of statements. In the example, the names (or labels) for the relationship types follow largely the Dublin Core standard (see Lecture 7.2b). Each statement is a triple. In the Linked Open Data (LOD) initiative on the Web data are triples See.

http://en.wikipedia.org/wiki/Linked_data LOD in libraries: http://kcoyle.net/presentations/links.html

www.oclc.org/multimedia/2011/files/globalcouncil/Buzash_Calhoun_Dunsire_Linked_Data.pdf, to slide 14

http://www.w3.org/2005/Incubator/lld/wiki/Main_Page

British National Bibliography: www.bl.uk/bibliographic/datafree.html, search at http://bnb.data.bl.uk/

Library of Congress: http://id.loc.gov/ **OCLC: www.oclc.org/data.en.html**.

MARC Record (simplified)	**E-R statements** (in this case triples, linked data on the Web)
010 2009038344	0838985130 <*hasLCControlNo*> 2009038344 # Comment: 0838985130 is the identifier for the book
040 DLC $c DLC $d YDX $d MTG $d CDX $d BWX	
020 9780838985137 (pbk. : alk. paper) [ISBN 13 digits]	
020 0838985130 (pbk. : alk. paper [ISBN 10digits]	0838985130 <*hasISBN13*> 9780838985137
050 00 ZA3075 $b .B83 2009	
050 4 Z674 $b .A75 no.61	0838985130 <*hasSubject*> LCC#ZA3075, 0838985130 <*has*LCCallNo> ZA3075 $b .B83 2009
082 00 028.7071/1 $2 22	0838985130 <*hasSubject*> LCC#Z674, 0838985130 <*has*LCCallNo> Z674 $b .A75 no.61
092 $b	0838985130 <*hasSubject*> DDC#028.7071/1
049 SBSM	
100 1 Budd, John, $d 1953-	0838985130 <>
	0838985130 <*hasCreator*> Budd, John, $d 1953-
245 10 Framing library instruction / $c by John Budd.	Budd, John <*hasBirthDate*> 1953
260 Chicago : $b Association of College and Research Libraries,	0838985130 <*hasTitle*> Framing library instruction
	0838985130 <*hasPublisher*> Association of College and Research Libraries
$c 2009.	Association of College and Research Libraries <*locatedIn*> Chicago
300 v, 197 p. ; $c 23 cm.	0838985130 <*hasPublicationDate*> 2009 # Comment: In MARC, date is a subfield
490 1 ACRL publications in librarianship ; $v no. 61	0838985130 <*hasNoOfPages*> (v, 197), 0838985130 <*hasHeight*> 23 cm
	0838985130 <*isPartOf*> (ISSN-0193-1784, 61),
504 Includes bibliographical references (p. 186-194) and index.	ISSN-0193-1784 <*hasTitle*> ACRL publications in librarianship
	0838985130 <*includesTypeOfMaterial*> BibliographicalReferences
505 0 A framework for instruction	0838985130 <*includesTypeOfMaterial*> Index
-- Beyond information literacy -- In the classroom -- Cognition and clear thinking -- A vision for learning -- Putting it all together.	0838985130 <*hasTOC*> A framework for instruction -- Beyond information literacy -- In the classroom -- Cognition and clear thinking -- A vision for learning -- Putting it all together.
650 0 Information literacy $x Study and teaching (Higher)	0838985130 <*hasSubject*> LCSH#Information literacy $x Study and teaching (Higher)
650 0 Information resources $x Evaluation $x Study and teaching (Higher)	0838985130 <*hasSubject*> LCSH#Information resources $x Evaluation $x Study and teaching (Higher)
650 0 Critical thinking $x Study and teaching (Higher)	0838985130 <*hasSubject*> LCSH#Critical thinking $x Study and teaching (Higher)
650 0 Library orientation for college students.	0838985130 <*hasSubject*> LCSH#Library orientation for college students.
650 0 Academic libraries $x Relations with faculty and curriculum	0838985130 <*hasSubject*> LCSH#Academic libraries $x Relations with faculty and curriculum

Lecture 4.2b. Data schemas and formats. In-lecture exercise (40 minutes)

Note: URLs for all documents connected with this exercise are visible in UBlearns and on the course website.

To get the most out of this, do the following:

(1) write down some questions. For each question, think what kind of data are required to answer it and based on that note the entity types and relationship types needed to represent such data.

(2) Study Example 1 (get the print.pdf and the audio) with the provided audio and, if you still need more practice or if you aspire to be certified as a Library Media Specialist, study Example 2 (get the print.pdf)

After the seated in-lecture exercise or after reviewing the example(s) you can work on Assignment 6.

Two examples are presented on the following pages

1 **The information system for a computer club** (example normally used)

2 **A totally integrated information system for a school**
(Main example if the majority of the students in the class are planning to get LMS certification)

There is a "print" document for this example on the course website.

Note: The purpose of this exercise is to gain a deeper understanding of the entity-relationship approach to modeling data. This purpose is achieved best by thinking from scratch about the needed entity types and relationship types. In real life, on the other hand, system builders should take advantage of existing conceptual data schemas, particularly those in wide use, and use accepted entity types and relationship types as much as possible. This saves time and makes systems more interoperable.

In-lecture exercise developing a conceptual data schema

Example 1. Information system of a large computer users' group

such as the Washington Apple Pi, www.wap.org or the
Western New York Computer Society, www.wnycomsoc.org

A computer users' group has the purpose of helping members to better use their computers.

Some functions of a computer users group

- a library for members to use
- a newsletter with articles and product reviews
- special interest groups (hold meetings, have a chair)
- a group purchase program
- a list of experts on specific subjects that have agreed to be on call to answer member questions

Sample questions with entity types and relationship types

Who knows about printers?
Entity types: Person, Subject
Relationship type: Person <*knowsAbout*> Subject

I am looking for a review of Microsoft Word 12
Entity types: Document, SoftwareMakeAndModel
Relationship type: Document <*reviews*> SoftwareMakeAndModel

What is a good word processor for Red Hat Linux
Entity types: New: Function, Quality; already noted: SoftwareMakeAndModel
Relationship types: SoftwareMakeAndModel <*servesFunction*> Function
SoftwareMakeAndModel <*worksWith*> SoftwareMakeAndModel
SoftwareMakeAndModel <*hasQuality*> Quality

Your sample questions / reports from the database

Layout for recording entity types and relationship types that emerge from the sample questions. Computer users group

Entity types	Relationship types

In-lecture exercise developing a conceptual data schema

Example 2. A totally integrated information system for a school

Functions of a totally integrated information system for a school

This is about envisioning a system beyond what few, if any, systems can do today. The system is to support all members of the school community with finding all information they need, producing reports or answering ad-hoc questions. Such a system is not pure fantasy; below is some information about a system called renweb (no endorsement, just a system I could find some information about).

www.renweb.com/Wheel/Integrated.aspx

www.blytheco.com/mip_fund_accounting/renweb.asp (look at this to get an idea of functions)

www.blytheco.com/attachments/products_and_services/mip_fund_accounting/misc/RenWebBrochure.pdf

www.renweb.com/Blog/EntryId/118/RenWeb-Launches-NEW-Fully-Integrated-Library-Management-Solution.aspx

I also posted a draft paper of mine on a lesson planning system that presents some relevant ideas (www.dsoergel.com/SoergelCLASSProposalStart.pdf, strictly optional)

Now to **some additional illustrative functions**

- Track individual students' learning and alert students, teachers, and parents of problems, possibly suggesting materials useful in addressing deficiencies adapted to the student's learning style.
- Assist teachers in lesson planning (considering curriculum standards) and finding learning objects (educational materials) that will support achievement of lesson learning objectives. (A learning object can be a plan for an entire year for a given class, a plan for a three-week unit, an individual lesson plan, description of / materials for a class activity, an assignment, a quiz or test, …)
- Find books and other library materials as well as websites that will be useful for students in completing an assignment and be prepared for many students coming into the library to use these materials at the time the assignment is given.
- Find out which of these materials are available in nearby public library branches and automatically alert the public library branch(es) when high demand for these materials is expected.
- Keep track of materials students are using and correlate with achievement of learning objectives.
- Find parents who could be asked to participate in career day.

Here are two questions to get your thinking started. They are deliberately a bit more complex and out of the ordinary than you might expect. See what you can come up with; simpler and more obvious questions are quite OK for now.

Question 1

I am concerned about this girl who is the victim of bullying. Could you recommend a fiction book she could read to help her to stand up for herself.

Entity types:

- Person
- Condition (Being bullied)
- Document
- DocumentGenre

Relationship types:

Person	*<hasCondition>*	Condition
Document	*<usefulFor>*	Condition
Document	*<belongsToGenre>*	Fiction

Question 2

I am looking for a group activity in a lesson on the dangers and consequences of stereotyping / labeling / name calling in sixth grade.

Entity types:

- LearningObject
- LearningObjectType (GroupActivity)
- CurriculumSubject
- GradeLevel

Relationship types:

LearningObject	*<dealsWithSubject>*	CurriculumSubject
LearningObject	*<suitableFor>*	GradeLevel
LearningObject	*<belongsToType>*	LearningObjectType
LearningObject	*<isPartOf>*	LearningObject
LearningObject	*<dealsWithStandard>*	CurriculumStandard
LearningObject	*<hasQuality>*	Quality [could set values as 1 – 5 stars)

Your sample questions / reports from the database

Layout for recording entity types and relationship types that emerge from the sample questions. Integrated information system for a school

Entity types	Relationship types

red

Things to do in Week 5, W February 17 - February 24

	Review answer keys ☐ **required** ○ **optional**	√
Review answer key(s)	Assignment 5: Analytical description of an information system (assigned L3.1, Feb. 10)	☐
	Assignments due W Feb. 24 ☐ **required** ○ **optional**	√
Assignment(s)	Assignment 6: Developing a conceptual data schema (assigned. L4.2, Feb. 17)	.☐

	New topics this week

	5.1 RDF, linked data, SPARQL query language (40 min)	
Readings	*1 Lecture 5.1 objectives etc. (pink sheet).*	☐
Lecture	**Lecture 5.1 slides** (40 min)	☐

	5.2 Access to information: data structure and search modes (40 min)	
Readings	1 Lecture 5.2 objectives etc. (pink sheet). Also look over Term paper description	☐
	2 Textbook Chapter 10. **Elementary query formulation**	☐
	3 Textbook Chapter 11. **Data structures and access**	☐
	4 Lecture 5.2 Notes, Sections 1-3: Retrieval as Prediction, Boolean retrieval, Ranking	☐
Lecture	**Lecture 5.2 slides** (40 min)	☐

Learning blog	Learning blog Week 5 due February 24	○

	Assignments assigned W Feb. 24	
Assignments assigned	▸**Prepare description of term paper** using the form found after Assignment 6 (due `Mar. 2)	

red

pink **Lecture 5.1** (40 min) ***February 24***

RDF, linked data, SPARQL query language

(No Textbook Chapter)

➜UBLIS 518 Reference Sources and Services

Learning objectives	1 Understand the basics of RDF and its use for storing data as RDF triples (E-R statements with binary relationships) on the Web. (P2.3.1,1.1.1) 1.1 Be able to write RDF triples in the Turtle serialization format. (P2.3.1,1.1.1,1) 1.2 Be able to define classes (entity types) and properties (relationship types) in RDF following an example. (P2.3.1,1.1.1,2) 2 Understand name spaces on the Web and their significance and are able to use name space declarations. (P2.3.1,1.1.2) 3 Understand the principles and significance of linked data. (P2.3.1,1.1.3) 3.1 Be able to work with a systems person to put linked data sets on the Web. (P2.3.1,1.1.3, 1) 4 Be aware of the possibilities of using RDF data sets and linked data to answer users questions and be able learn how to search linked data. (P2.3.2,2) 4.1 Be aware of and be able use linked data browsers. (P2.3.2,2.1) 4.2 Be aware of the query language SPARQL and of the concept of a SPARQL endpoint. (P2.3.2,2.2)
Practical significance, examples	• Linked data are becoming an increasingly useful source for answering all kinds of user questions. Being able to search linked data will make you a better reference librarian. • Libraries can be providers of linked data. Catalog data become more useful when they can be used in conjunction with other data; many libraries post their catalog as linked data. Libraries can help community groups to post data sets as linked data.

Outline

1 Resource Description Framework (RDF). RDF Schema (RDFS)
2 Linked Data
3 SPARQL query language

pink

1 Resource Description Framework (RDF). RDF Schema (RDFS)

Definition	A general abstract data modeling method based on the entity-relationship (E-R) approach. Originally developed with metadata in mind, but is now used widely for any kind of data. (Soergel *Searching linked data* elaborates on this.) Entities / objects are called *resources*. Entity types are called *classes.* Relationship types are called *properties*. A statement is made *about* a resource (the subject of the statement), using a property and giving another resource as the *value.* Example (example 2 is better for using a URI identifier, not a string): www.dsoergel.com/ublis571 <*hasCreator*> “Dagobert Soergel” . www.dsoergel.com/ublis571 <*hasCreator*> www.dsoergel.com/ds . Such statements are called *triples* since they consist of three elements. RDF and RDFS both specify a format (through giving properties (relationship types)); these formats are used widely in data sets. Roughly speaking, RDFS specifies a format for defining a conceptual data schema: define classes (entity types) and properties (relationship types); RDF specifies a format for writing actual data statements.
Implementation	An RDF data set is a set of entity-relationship statements. Any subset of E-R statements (including the whole data set itself) can be represented as a graph. Conceptually, RDF is often thought of in terms of such RDF graphs. Representing RDF data sets in textual form is called *serialization* because the data represented in a 2-D RDF graph are now represented as a series of E-R statements arranged one after the other. Several serialization formats exist. The *Turtle* serialization format is the most popular, more so than the XML-based format. We will use the Turtle serialization format. www.w3.org/TR/turtle/
Name spaces	Classes, properties, and individual entity values all have names / identifiers. Names are given by many different systems; they may use the same name for different things. To avoid confusion, every system has its own name space identified by a URI (Universal Resource Identifier). The full name consists of the name space URI and the name given within the name space.

For a good overview with more detail see http://en.wikipedia.org/wiki/Resource_Description_Framework

Examples of serialized RDF graphs in the Turtle serialization format

Here is a simple example of an RDF graph. http://www.country.foodanddrugadministration.gov/data is a fictitious website; it establishes the names prefixed with fda: and defines a property *label* in the meaning used in the food domain. So fda:*label* is quite different from the property rdfs:*label* which means any type of name or designation. This illustrates the importance of name spaces. Name spaces ensure that there is no collision between names (of properties, classes, or entity values) defined in different systems using the same name for different things. Without the use of distinct name spaces, things would get horribly confusing, both to document creators and the automated systems processing RDF data.

To increase readability, properties are written in italics and white space is inserted.

Example 1. Food product data

```
@prefix fda: <http://www.country.foodanddrugadministration.gov/data>           # Name space declaration

http://www.someccountry.foodanddrugadministration.gov/data#FP0        # Full name space URL
   http://www.someccountry.foodanddrugadministration.gov/data#hasName      "Food product" .

fda:FP1      fda:hasName            "Vegetable product"          # Short name space prefix
fda:FP1      fda:isa                fda:FP0 .
fda:FP1      fda:comesFromSource    fda:Plant .

fda:FP2      fda:hasName            "Meat  product" ;            # ; next triple has same left side
             fda:isa                fda:FP0 ;
             fda:comesFromSource>   fda:Animal ;
             fda:comesFromPart>     fda:Carcass .
```

Example 2. Document data

```
@prefix dc: <http://purl.org/dc/terms/#>
@prefix ukoln: <www.ukoln.ac.uk/metadata/resources/#>

ukoln:document337  dc:title         "Guidance on expressing the Dublin Core within RDF" ;
                   dc:creator       <http://www.w3.org/People/EM/contact#me> ,  # URI for Eric Miller
                                    <www.linkedin.com/in/pau1mil1er>
                                    <http://danbri.org/foaf#danbri> ;            # Dan Brickley
                   dc:subject       "Dublin Core" ,
                                    "RDF" ,
                                    "XML" ;
                   dc:publisher     <http://dublincore.org/about-us#DCMI>
                                              # ficitious for the DublinCore Metadata Initiative
                   dc:contributor   <http://dublincore.org/about-us#DCMI>
                                    # ficitious for the"Dublin Core Data Model Working Group"
                   dc:date          "1999-07-01"
                   dc:format        dc:text/html                  # This is defined in dc:
                   dc:language      en                            # From RFC 5646, no name space

ukoln:document337  ...
```

The next few pages show
a worked-out example of representing some bibliographic data in RDF

For another example (the food database from Lecture 2.2) see supplement, SLecture 5.1.
The same database is shown in an XML representation in SLecture 6.1a

RDF Schema	It is useful, but not required, to define a conceptual data schema before entering data using the RDF data model, for example in the Turtle serialization format. Defining a conceptual data schema involves defining entity types (in RDF: classes) and relationship types (in RDF: properties). This is what is done here through an XML document that uses tags that are predefined in RDF Schema (prefix rdfs:). Please get the gist of this. I will not ask about details of syntax in an exam, but I may present a practical problem were using RDF like that might factor into the solution. And why would I ask a question like that? Because that might happen to you in real life. If it does, you should recognize that RDF representation might be helpful, find out about data sets you could draw on, and find somebody who can implement the idea. Again, some of you might become interested in this and learn enough to be experts at handling data in RDF Below is a very simple piece of an RDF schema definition for bibliographic data.
Classes (entity types)	Note the specification of **hierarchical relationships among classes**, for example the entity type (class) Person is defined as a subclass of LegalEntity using the tag rdfs: subClassOf. Also, there are **hierarchical relationships among properties (relationship types)** Also note that every class and every property has an internal ID and an external label. They can be the same but do not have to be. XML and RDF schemas for the food database from Lecture 2.2,is given in www.dsoergel.com/SYL2003FaLecturesAppendixNew.pdf (see Lecture 6.2b)

Example: defining an entity-relationship conceptual data schema using RDF and RDFS

The definition uses classes and properties in the RDF and RDFS specification (see http://www.w3.org/TR/rdf-schema/, to UFMG students also sent as attachment for convenience)
"The rdf:type property may be used to state that a resource is an instance of a class." The other properties defined in RDF or RDFS that are used her are self-explanatory due to their meaningful identifiers.

The possible uses of this conceptual data schema will be discusses after the entire example (schema and data).

Schema and data are defined or stated using the same format.

```
# File found at www.dsoergel.com/bibschema

@prefix rdf:       <www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:      <www.w3.org/2000/01/rdf-schema#> .
@prefix dsbibs:    <www.dsoergel.com/bibschema#> .  # Fictitious, refers to this file for consistency
@prefix dsbibd:    <www.dsoergel.com/bibdata#> .         # Fictitious
```

Definition of classes (entity types)

```
dsbibs:Book
    rdf:type             rdfs:Class ;
    rdfs:label           "Book" ;
    rdfs:comment         "The class of books" ;
    rdfs:subClassOf      rdfs:Resource .

dsbibs:LegalEntity
    rdf:type             rdfs:Class ;
    rdfs:label           "Legal entity" ;
    rdfs:comment         "The class of person or organizations" ;
    rdfs:subClassOf      rdfs:Resource .

dsbibs:Person
    rdf:type             rdfs:Class ;
    rdfs:label           "Person" ;
    rdfs:comment         "The class of persons" ;
    rdfs:subClassOf      dsbibs:LegalEntity .

dsbibs:Organization
    rdf:type             rdfs:Class ;
    rdfs:label           "Organization" ;
    rdfs:comment         "The class of organizations" ;
    rdfs:subClassOf      dsbibs:LegalEntity .
```

RDF schema definition example continued

Next we define properties (relationship types). Here is where the real power of RDF comes in. For example, the property *creator* (as defined below in the "name space" dsbibs: has an entity of class Book at its left side (domain) and an entity of class LegalEntity (a Person or Organization) at its right side (domain).

Definition of properties (relationship types)

dsbibs:title
rdf:*type* rdf:Property ;
rdfs:*label* "Title" ;
rdfs*:comment* "The name given to the resource" ;
rdfs:*domain* dsbibs:Book **;**
rdfs:*range* rdfs:Literal **.**

dsbibs:creator
rdf:*type* rdf:Property ;
rdfs:*label* "Creator" ;
rdfs*:comment* "A person or organization responsible for the content of a book" ;
rdfs:*domain* dsbibs:Book **;**
rdfs:*range* dsbibs:LegalEntity **.**

dsbibs:editor
rdf:*type* rdf:Property ;
rdfs:*label* "Editor" ;
rdfs*:comment* "A person or organization responsible for selecting, arranging, and formatting the content of a book" ;
rdfs:*subPropertyOf* dsbibs:creator **.**

dsbibs:affiliatedWith
rdf:*type* rdf:Property ;
rdfs:*label* "person affiliation" ;
rdfs*:comment* "The organization a person is affiliated with" ;
rdfs:*domain* dsbibs:Person **;**
rdfs:*range* dsbibs:Organization **.**

Data on some books structured according to the rdf schema just given see next page

You may want to take the next page out so you can compare with the definitions of the properties (relationship types)

Bibliographic data structured using the classes and properties just defined

File found at www.dsoergel.com/bibdata

@prefix *rdf*: <www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <www.w3.org/2000/01/rdf-schema#> .
@prefix dsbibs: <www.dsoergel.com/bibschema#> . # Fictitious, refers to this file for consistency.
@prefix dsbibd: <www.dsoergel.com/bibdata#> . # Fictitious

\# **Class membership definitions: Entity values assigned to classes (entity types)**

dsbibd:ISBN0126542619 rdf:*type* dsbibs:Book .# The Soergel bibliographic database uses ISBNs to identify books.

dsbibd:ISBN0081086007 rdf:*type* dsbibs:Book .

dsbibd:ISBN9780838912102 rdf:*type* dsbibs:Book .

dsbibd:ISBN0471046736 rdf:*type* dsbibs:Book .

dsbibd:Soergel rdf:*type* dsbibs:Person . # dsbibd uses last name as Person identifier. Not best

<www.simmons.edu/faculty#Chan> rdf:*type* dsbibs:person . # external identifier

dsbibd:ALA rdf:*type* dsbibs:Organization .

\# **Data on books (relating books to persons and organizations)**

dsbibd:ISBN0126542619
dsbibs:*title* "Organizing information" ; # The range of dsbibs:*title* is defined as literal.
dsbibs:*creator* dsbibd:Soergel . # The range of dsbibs:*creator* is defined as LegalEntity, which includes Person.

dsbibd:ISBN 081086007
dsbibs:*title* "Cataloging and classification" ;
dsbibs:*creator* <www.simmons.edu/faculty#Chan> .

dsbibd:ISBN 9780838912102
dsbibs:*title* "RDA: Resource Description and Access Print--2013 Revision (Includes July 2013 Updates) " ;
dsbibs:*creator* dsbibd:ALA .

dsbibd:ISBN0471046736
dsbibs:*title* "Information retrieval systems. 2. ed." ;
** dsbibs:*creator* dsbibd:Lancaster .

** See next page, Using the RDF conceptual data schema definition. If the conceptual data schema is used for integrity checking, this line will be rejected since Lancaster has not been defined as a person. If the conceptual data schema is used to infer the entity type of an unknown value, Lancaster will be assigned the entity type Person.

Using the RDF conceptual data schema definition	
Standard RDF use: No integrity checks	The RDF specification does not ask for integrity checks. The specification assumes that input data are correct and that missing class membership information can be derived from the input data. For example, when an RDF processor encounters dsbibd:ISBN0471046736 dsbibs:*creator* dsbibd:Lancaster . and dsbibd:Lancaster does not already belong to class dsbibs:LegalEntity (or one of its subclasses) as required in the definition of dsbibs:*creator*, then the RDF processor assigns dsbibd:Lancaster to the class dsbibs:LegalEntity even if dsbibd:Lancaster already belongs to another class. (In RDF an individual entity can belong to multiple classes.) This example is a bit more complex than most cases: The RDF processor can infer dsbibs:LegalEntity but not more specifically dsbibs:Person or dsbibs:Organization. In this use, the rdf:*type* statements in the beginning of the data file are not needed. A standard processor of RDF data even accepts statements that use a property that has not been defined. The processor can assign the property to the class rdf:Property but it cannot assign the entity values used to classes since domain and range are not known for an undefined property. There are clearly other advantages of defining properties explicitly, such as clear textual definition and consistent use of a property within a data set and across data sets. This is just like any other case of vocabulary control. RDF practice uses mostly well-defined properties, preferably from standard property vocabularies.
Traditional database use: Integrity checks	In traditional databases, it is assumed that input data contain errors. The conceptual data schema is used to do integrity checks to detect such errors. For example, if such a system encounters dsbibd:ISBN0471046736 dsbibs:*creator* dsbibd:Lancaster . and dsbibd:Lancaster does not belong to class dsbibs:LegalEntity (or one of its subclasses) as required in the definition of dsbibs:*creator*, then the processor rejects the statement. dsbibd:Lancaster needs to be explicitly assigned to dsbibs:Person before the statement will be accepted. That way errors (including misspelled names) will be detected and can be corrected. A system processing RDF data could be programmed to do integrity checks; that would go beyond the RDF specifications.
Tradeoffs	There is clearly a tradeoff: The RDF approach minimizes effort but allows for a higher level of inconsistency and errors. Note: If all the name authority data of the Library of Congress were put on the Web as Linked Open Data in a format that would declare every person as belonging to class Person and every organization to class Organization, the result would be that in almost all cases a **correctly spelled** name would pass muster.

2 Linked Data

The idea of Linked Data is that anybody can put data sets on the Web and that search engines and applications can use any combination of these data sets to find answers, draw inferences, compute numbers, or do any other kind of processing. Data sets are most commonly represented in one of the RDF serialization formats, often Turtle. This is an ideal, realization is slow. The following principles need to be followed to come closer to the ideal.

From Ducharme, Bob 2013.*Learning SPARQL*, p. 41-42:

"Linked Data is not a specification, but a set of best practices for providing a data infrastructure that makes it easier to share data across the Web. You can then use semantic web technologies such as RDFS, OWL, and SPARQL to build applications around that data.

Four principles of Linked Data (Tim Berners-Lee 2006) (I've bolded his wording and added my own commentary):

1. **Use URIs as names for things**. URIs are the best way available to uniquely identify things, and therefore to identify connections between things.
2. **Use HTTP URIs so that people can look up those names**. You may have seen URIs that begin with ftp:, mailto:, or prefixes made up by a particular community, but using these other ones reduces interoperability, and interoperability is what it's all about.
3. **When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).** While a URI can just be a name and not actually the address of a web page, this principle says that you may as well put something there. It can be an HTML web page, or something else; whatever it is, it should use a recognized standard. RDFS and OWL let you spell out a list of terms and information about those terms and their relationships in a machine-readable way- readable, for example, by SPARQL queries. Because of this, if a URI that identifies a resource leads to RDFS or OWL declarations about that resource, this is a big help to applications. (The asterisk in "RDF*" means that it refers to both RDF and RDFS as standards.)
4. I**nclude links to other URIs so that they can discover more things**. Imagine if none of the original HTML pages had a elements to link them to other pages. It wouldn't have been much of a web. Going beyond this HTML linking element, various RDF vocabularies provide other properties that let you say "this data (or this element of data) has a specific relationship to another resource on the Web." When applications can follow these links, they can do interesting new things."

An increasing amount of data is made available as linked data (see http://en.wikipedia.org/wiki/Linked_data, http://thedatahub.org/), **including library catalogs** (for example, http://thedatahub.org/dataset/bluk-bnb) and the LC Subject Headings (www.oclc.org/research/news/2011-12-14.htm). The BBC makes much of its data available this way. Many libraries now are working on putting their catalog data available as linked data, thus letting users interrelate the catalog with many other kind of data. You might be asked to participate in such a project.

****

The following pages give two simple queries using the SPARQL query language. (SPARQL stands for **S**PARQL **P**rotocol **A**nd **R**DF **Q**uery **L**anguage). Several large data providers have a *SPARQL Endpoint*, a website where one can enter SPARQL queries to search one or more data sets provided; no special software needed at the user end. The examples use the Wikipedia SPARQL endpoint at **http://dbpedia.org/sparql**

You can try these queries (or modifications) yourself (may need to try different browsers) .

3 SPARQL query language (using http://dbpedia.org/sparql)

SPARQL Explorer for Wikipedia (SPARQL = **S**parql **P**rotocol **A**nd **R**df **Q**uery **L**anguage)

```
PREFIX : <http://dbpedia.org/resource/>        # Note: the prefix is just :
PREFIX d: <http://dbpedia.org/ontology/>        # This part of the query
                                                # entered by user
SELECT ?artist ?album
WHERE
{                                               # Remember Prolog example,
  ?album d:producer :Timbaland .                # Lecture 1.2
  ?album d:musicalArtist ?artist .              # ? signifies a variable
}
```

SPARQL Results: (On the website, these are all hyperlinks; I converted to text for readability)

These are just the first results.

artist	album
:Missy_Elliott	:All_n_My_Grill
:Nicole_Wray	:All_n_My_Grill
:Keri_Hilson	:Breaking_Point_(song)
:Missy_Elliott	:Gossip_Folks
:Missy_Elliott	:Hit_Em_wit_da_Hee
:Missy_Elliott	:I'm_Really_Hot
:Missy_Elliott	:Sock_It_2_Me
:Missy_Elliott	:The_Rain_(Supa_Dupa_Fly)
:Hikaru_Utada	:Exodus_'04
:Tweet_(singer)	:Oops_(Oh_My)
:Nicole_Scherzinger	:Scream_(Timbaland_song)
:Timbaland	:Scream_(Timbaland_song)
:Madonna_(entertainer)	:4_Minutes_(Madonna_song)
:Brandy_Norwood	:Afrodisiac_(song)
:Nelly_Furtado	:All_Good_Things_(Come_to_an_End)

On the website, clicking on the link at Breaking Point results in a list of RDF triples giving many different types of information about that song. See facing page.

Some common prefixes for property (relationship type) vocabularies (aka ontologies)

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

About: Breaking Point (song)

An Entity of Type : work, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

"Breaking Point" is a song performed by American recording artist Keri Hilson from her second studio album, No Boys Allowed (2010). Mosley Music Group and Interscope Records released it as the album's lead single on September 7, 2010. The song was written by Timbaland, Jerome "J-Roc" Harmon, Hilson and Attitude, and produced by Timbaland with assistance from Harmon.

All are URIs (Universal Resource Identifiers, note the prefix),
unless in " " (indicates xsd:string) or (xsd:integer), (xsd:date), ... (note that there is no prefix)

Property	Value
dbpedia-owl:abstract	"Breaking Point" is a song performed by American recording artist Keri Hilson from her second studio album, No Boys Allowed (2010). . . ."
dbpedia-owl:album	dbpedia:No_Boys_Allowed
dbpedia-owl:format	dbpedia:Music_download
dbpedia-owl:genre	dbpedia:Contemporary_R&B
dbpedia-owl:musicalArtist	dbpedia:Keri_Hilson
dbpedia-owl:musicalBand	dbpedia:Keri_Hilson
dbpedia-owl:previousWork	dbpedia:Got_Your_Back
dbpedia-owl:producer	dbpedia:Timbaland dbpedia:Jerome_"J-Roc"_Harmon
dbpedia-owl:recordLabel	dbpedia:Interscope_Records dbpedia:Mosley_Music_Group
dbpedia-owl:recordedIn	dbpedia:Record_Plant
dbpedia-owl:releaseDate	2010-09-07 (xsd:date)
dbpedia-owl:subsequentWork	dbpedia:Pretty_Girl_Rock
dbpedia-owl:thumbnail	http://upload.wikimedia.org/wikipedia/commons/thumb/0/06/Breakingpoint.JPG/200px-Breakingpoint.JPG
dbpedia-owl:wikiPageID	28750325 (xsd:integer)
dbpedia-owl:wikiPageInLinkCount	19 (xsd:integer)
dbpedia-owl:wikiPageOutLinkCount	139 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	547235411 (xsd:integer)
dbpedia-owl:writer	dbpedia:Keri_Hilson dbpedia:Attitude_(rapper) dbpedia:Timbaland dbpedia:Jerome_"J-Roc"_Harmon
dbpprop:album	dbpedia:No_Boys_Allowed
dbpprop:artist	dbpedia:Keri_Hilson

Another SPARQL query example. Optional

To pose a SPARQL query one needs to have/know the following:

- a SPARQL processor on one's computer or a website where one can enter SPARQL queries;
- one or more data sets to search;
- the conceptual data schema for each data sets, especially the properties (relationship types) it uses and their identifiers (at least the properties to be used in the query);
- the identifier(s) for the specific entities one wants to use as starting points in the query.

For the first query example,

- I used a website,
- I wanted to search data sets provided by Wikipedia.
- This is an example from *Learning SPARQL*; the book told me that I can use the properties *producer* and *musicalArtist* defined in <http://dbpedia.org/ontology/>
- The book also told me that Timbaland is an entity identifier used in the data set <http://dbpedia.org/resource/>

From the listing of RDF triples on p. ~137 we can see many properties and entity identifiers. So we can issue a query to find all Rhythm and blues ballads performed by Keri Hilson and the albums in which they appeared:

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT ?song ?album
WHERE
{
 ?song        dcterms:subject        category:Rhythm_and_blues_ballads .
 ?song        dbpprop:artist         :Keri_Hilson .
 OPTIONAL { ?song    dbpedia-owl:album    ?album . }
}
```

SPARQL results:

song	album
:Breaking_Point_(song)	:No_Boys_Allowed
:Energy_(Keri_Hilson_song)	:In_a_Perfect_World...
:Slow_Dance_(song)	:In_a_Perfect_World...
:Make_Love	-

The last song has no album data. Without the keyword OPTIONAL it would not have been found since the third condition is false if a song does not have an album. But the third condition is needed to find the album when the information is there. OPTIONAL tells the query processor to find the song even if the third condition is false

To learn more about linked data and SPARQL: Every semester I offer a UBLIS 598 Directed Study.

pink **Lecture 5.2** ***February 24***

Access to information: Data structure and search modes

(Textbook, Chapters 10 and 11)

➜UBLIS 518 Reference Sources and Services

Learning objectives	1 Understand and be able to apply the basic principle of searching: use all available evidence to predict the degree of relevance of some entity. (P2.5.2,1.1) 2 Understand search algorithms: Boolean retrieval and ranked retrieval and the role of synonym expansion and hierarchic expansion in both and are able to apply this knowledge in systems that support inclusive searching as well as in systems that do not, including free-text searching. (P2.5.2,1.3) 2.1 Be able to formulate Boolean queries. (P2.5.2,1.3.1) 2.2 Understand the role of hierarchy in searching and be able to use hierarchy in searching. (P2.5.2,1.3.1) 2.3 Be able to formulate free-text (such as Google) queries enhanced through synonym and hierarchic expansion (P2.5.2,1.3.3) 3 Understand how a given data structure supports answering questions through retrieval and inference, especially how the tables in a database can be used together to answer questions, and how indexes help. (P2.4.2,1.1) 4 Graduates are able to analyze the storage structures (tables, record formats) of an information system and design simple storage structures, including the use of hierarchical inheritance. (P2.4.2,1.2) 5 Be able to design simple data access structures (for example, indexes). (P2.4.2,1.3)
Practical significance, examples	• When searching for X, use a reference tool where X can be found in the index. • Store data with minimal redundancy by using hierarchical inheritance. (OCLC and other bibliographic databases are tremendously redundant since they do not use a data structure that exploits hierarchical inheritance.) • Design simple databases, for example, paying attention to creating indexes in MS Access or MySQL (tradeoff between time for creating a database and searching it).

Over

pink

Outline

1 Retrieval as prediction. *Read this section, not discussed*

Query formulation: Find good clues that predict the relevance of an entity (document, person, computer program, etc.) for a given user with a given purpose
Retrieving a document or other entity is predicting that it is relevant to the problem to be solved at least to a certain degree. The challenge in formulating a query is to find the clues that predict that a document or other entity is relevant:

For documents:	What clues can predict that the document will be helpful in solving the problem at hand?
For persons to fill a job:	What clues can predict that a person will do well in the job?

Finding the right clues requires knowledge and may involve some guesswork:

- When searching for documents using free-text retrieval, the searcher must determine what words and expressions the author of a relevant document may have used in the title, the abstract, and the full text; this requires knowledge of how language is being used both in general and by specific scientific schools and even individual authors.
- When using descriptors assigned by an indexer, the searcher must determine what descriptors an indexer would have assigned to relevant documents; this requires knowledge of the index language, the indexing instructions, and the actual indexing practices. (Request-oriented indexing, to be discussed later, seeks to increase the probability that descriptors corresponding to user needs are included in the indexing language and assigned correctly in indexing.)

Of course, free-text terms and assigned subject descriptors are only two kinds of evidence. Many other clues can be considered, such as

- publication date,
- topical focus of the journal in which an article appears,
- reputation of the journal or publisher,
- reputation of the publisher,
- reputation of the author or the author's institution,
- etc.

Again. Think about a user's question where clues other that words expressing the search concept might be useful. To be a good reference librarian, you need to be imaginative in thinking of different clues that could lead to relevant items.

Example:
You want to find documents that list several programs for managing your personal bibliographic database. There are several terms for this concept, such as *bibliographic software*, *bibliography manager*, *reference manager*, *text database management system*. Rather than trying to come up with all possible terms to search, think of two or three programs you know, such as *Endnote, Zotero, and RefWorks*, and put them in the Google search box. Documents that mention all three of these might well mention others as well, and some of these documents will be comparative reviews.

2 Brief review of Boolean retrieval (Textbook, Chapter 10, p. 165 - 171)

Boolean operators, including the use of parentheses, are reviewed in textbook Sections 10,1 - 10.3. You are assumed to know simple Boolean searching, as in the following Google search:

ranking AND methods AND (review OR "state of the art" OR overview)
(AND is not needed since it is the default; added for clarity. "" searches for the phrase)

Natural language *and* is ambiguous, it means AND in some contexts and OR in others. Textbook Section 10.4 p. 168 explains how to tell from the context. Be sure you understand how to tell when *and* means AND and when *and* means OR

NOT is tricky, use judiciously. Many people do not understand how NOT works, you should. In Section 10.6, Figure on Textbook p. 170, be clear about the **relevant documents missed by the AND NOT query,** why they are relevant and why they are missed.
Section 10.6 gives an example of how to get around NOT, but in some cases, NOT (in some systems AND NOT) is the only way to narrow the query formulation.

3 Ranking documents by expected relevance (as in Google)

3.1 Introduction

Boolean retrieval: YES or NO – division of all documents in the system into **two classes**
A document either scores 1 and is retrieved or it scores 0 and is rejected: class 1: retrieved - expected to be relevant class 2: rejected - expected not to be relevant

Using two queries to get YES, MAYBE, NO – division of all documents into **three classes** This method is often useful in practice
Problems with formulation of queries. Consider the following query formulation consisting of four descriptors *A* Simulation AND *B* Traffic flow AND *C* Passengers AND *D* Rail transport Perhaps documents that contain any three of the four descriptors might be of some interest; could run the broader query (ABC OR ABD OR ACD OR BCD) in parallel: **Class 1**: retrieved in narrow search ABCD – expected to be clearly relevant **Class 2**: retrieved in broad search – expected to be somewhat relevant **Class 3**: not retrieved even in the broad search – expected to be not relevant In a system that uses Boolean retrieval, running a focused narrow query and a broader query is often a good idea to give the user a result list that has first the most important items and then more items to explore **–** a very coarse form of ranking

More refined ranking by expected relevance – continuous scale

Compute for each document a **quantitative measure of expected relevance** to the given search request. Instead of 3 classes of documents, we then get a **list of documents ranked according to expected relevance**. (In many systems the ranking is poor and does not approximate the user's intuition.) The measure is usually computed as the nearness or **similarity** between query formulation and document representation (not best), based on the number of descriptors they have in common. Different formulas are possible.
Very simple formula: percentage of query descriptors found in the document record. A document would get a score as follows

all four query descriptors	1.00
three query descriptors	.75
two query descriptors	.50
one query descriptors	.25
no query descriptors	.00

For each formula:

(a) crude form: only the exact term matches, uses no knowledge of concept term relationships and

(b) knowledge-based form: uses knowledge of concept and term relationships, for example to match a query term with a synonym or narrower term in the document.

Using OR descriptor combinations, as in the following query formulation:

Search topic
Traffic congestion in airports [terminals for air traffic] in NY or Boston or Washington

Assume two documents with index terms as follows:

Document 1:	*Traffic congestion*	*Terminals*	*Air traffic*	*Boston*
Document 2:		*Terminals,*		*New York*, *Boston, Washington*

The Google query (AND is implied) ("" instructs Google to search the phrase.)

Q1 "Traffic congestion" Terminals "Air Traffic" "New York" Boston Washington

treats all terms the same (six components) and would give the same score (simplified: 4/6 or .66) to both documents, even though document 1 is clearly more relevant.

Q2 "Traffic congestion" Terminals "Air Traffic" ("New York" OR Boston OR Washington)

considers the conceptual structure and expresses four conceptual components: the three subjects and the place component expressed by ("New York" OR Boston OR Washington). It would give the following scores:

Document 1 gets 1.0 (it matches all four conceptual components of the query)
Document 2 gets .5 (it matches only two conceptual components, *Terminal* and place)

Using OR in Google can give much better results

3.2 In-lecture exercise: **Ranking of retrieved documents**

All: Read Purpose and Task and p. ~146 - 147 as indicated

Purpose	1. To give you a better "feel" of how formulas for the computation of expected relevance work and what a rank list of documents looks like. 2. To have you compare the effectiveness of four formulas (two base formulas, each applied with two matching rules).
Task	Given are a query formulation and four document representations (descriptors assigned to the documents) and four formulas for the computation of expected relevance. The formulas are deliberately very simple; many more complex formulas are being used. 1. Using your own judgement, rank the documents 1-4 by their relevance to the query. 2. Compute for each document the coefficient of expected relevance according to four different formulas and list the documents in rank order by decreasing expected relevance. 3. Compare the four rankings with your intuitive ranking. State which is better. Briefly state why one formula works better than the other.

In this exercise we will conduct a thought experiment to test the performance of four ranking formulas. The next two pages give

1 A "test collection" consisting of four documents with descriptors and one query. To complete this test collection, you will supply a ranking of the documents by relevance to the test query (most relevant document is ranked first).

2 Explanation of four formulas to compute expected relevance for each document and thereby determine a ranking; the ranking is the final result.

There will be instructions for the seated class and for online.

"Test collection": Query formulation Q and descriptors for documents D1 - D4

Q	**B1.2** Rail transport		**AND** **E1.2** Traffic stations	**AND** **J1** Passenger transp.	**AND** **U15** US		
D1	**B1.2** Rail transport	**B3** Air t.	**E1.2** Traffic stations	**J1** Passenger transp.	**U15** US	**U20** Europe	**Q24** Traffic simulation
D2	**B1.2.1** Local rail transport		**E1** Traffic facilities	**J1** Passenger transp.	**U15** US		
D3	**B1** Ground transp		**E2.1** Vehicles	**J1** Passenger transp.	**U15** US		
D4	**B1.2** Rail transport		**E1** traffic stations	**J1** Passenger transp.	**U15** US		

The concept numbers or notations (like Dewey numbers) B1, B1.2, B1.2.1, J1, etc. come from an imaginary classification; they do express a hierarchy

Formulas for computing expected relevance

Base formulas 1 and 2

Base formula 1:	R = # of matching descriptors M / # of query descriptors Q
Base formula 2:	R = # of matching descriptors M / (# of query descriptors Q + # of doc. descr. D)

Matching rules. M, the number of matching descriptors depends on how we define that a query descriptor is matched by a document descriptor. The following table defines two matching rules.

Matching rule a: (crude)	**Exact descriptor match**: A query descriptor produces a match only if the document has exactly the same descriptor			
Matching rule b: (knowledge-based)	**Hierarchically expanded match**: A query descriptor produces matches as shown in the following examples:			
	Query descriptor	**Document descriptor**		**Match value**
	B1.2 Rail transport	**B1.2** Rail transport	Same	1
		B1.2.1 Local rail transport	Narrower	1
		B1 Ground transport	Broader	0.5

Note: The numbers in the column "Match value" are set arbitrarily for this exercise. One might count a narrower descriptor as 0.75 of a match, for example.

This results in four formulas for computing expected relevance:

How query descriptor to document descriptor match is determined	**How the score is computed: Base formula**			
	1	**M / Q**	**2**	**M / (Q + D)**
Ma Matching rule a: exact	1a	Ma / Q	2a	Ma / (Q + D)
Mb Matching rule b: knowledge-based	1b	Mb / Q	2b	Mb / (Q + D)

M Number of descriptors in common

Q Number of descriptors in Query

D Number of descriptors in Document

Seated: Work on the board to compute the relevance scores and derive the rankings

Online*:* See next page

Results

Expected relevance score for the query Q

Document	**Formula**			
	1a	**1b**	**2a**	**2b**
D1				
D2				
D3				
D4				

Ranking. Put document no in cell

Rank	**Intuitive**	**Formula**			
		1a	**1b**	**2a**	**2b**
1					
2					
3					
4					

Online see next two pages

For all: Then continue on p. ~150

Online version of the in-lecture exercise. **Instructions**

The exercise uses the following spreadsheet
http://www.dsoergel.com/UBLIS571DS-05.2-2Lecture5.2RankingSpreadsheet.xlsx
Open it and get familiar with it following the explanation

Explanation of the spreadsheet.

On the top are three tables for summary results:

Table 1	The query formulation and document representations for reference
Table 2	A table where the spreadsheet computes for each document (row) the expected relevance score under each formula (column)
Table 3	A table to record the rank ordering, first your intuitive ranking, then the ranking by each formula

Then there are **four tables, D1-D4, one for each document, to enter the match values from which the spreadsheets computes the expected relevance scores for each document in Table 2.**

Instructions for what you do

If you have trouble with the instructions, take a peek at Sheet 2, it has all the answers filled in.

Step 1	Rank the four documents D1 – D4 intuitively and put the document numbers in the appropriate cell of the ranking Table 3**.**
Step 2	**Do one document at a time**: Documents D1-D4, Tables D1-D4, respectively For each query descriptor tenter the match value under matching rule a and under matching rule b
	For **document D1** this easy: For each query descriptor there is a document descriptor, so all match values are the same: 1 **Row a (exact match)**: all match values are 1 **Row b (knowledge-based)**: all match values are 1
	For **document D2** this is a bit more difficult: **Row a (exact match)**: query descriptors *B1.2* and *E1.2*: no exact match in the document, 0 query descriptors *J1* and *U15:* exact match in document, 1 **Row b (knowledge-based)**: query descriptor *B1.2*: narrower document descriptor B1.2.1, by matching rule b the match value is 1 query descriptor *E1.2*: broader document descriptor E1, by matching rule b the match value is .5 query descriptor *J1*: Document descriptor same, 1 query descriptors *U15*: Document descriptor same, 1
	Do Documents 3 and 4. The spreadsheet will compute Table 4

Step 3	**Arrange the documents by rank** according to each formula (highest score is Rank 1) in the Ranking Table. For formula 1a, D1 and D4 have the same top score, so they both go into Rank 1 , Rank 2 is free, D2 and D3 go into Rank 3.
Step 4	**Compare the ranking with your intuitive ranking.** If you put D1 and D4 both into Rank 1 because they have all four query descriptors, followed by D2 and D3 in Ranks 3 and 4, respectively, then you will like formula 1b best. Formula 1a does not work as well because it does not consider that D2 has two descriptors that, while not the same, are related to query descriptors while D3 has only one. If you put D4 in Rank 1 and D1 in Rank 1 because it has extra descriptors while D4 matches the query exactly, then you might like formula 2b best: even though it places D2 above D1, it separates D2 and D3 and formula 2a does not. In either case, the knowledge-based formula works better than the exact-match (or dumb) formula. Few search engines use knowledge-based formulas because it is hard to assemble the required knowledge for a universal system. A plain search in Medline uses the dumb matching, but an inclusive (hierarchically expanded) search uses knowledge-based match using the knowledge about hierarchy available in MeSH (Medical Subject Headings). Whether a document should be penalized for having extra descriptors (D1 vs. D4) is a question answered differently by different people. The pro argument is that a document with extra descriptors does not have as much space left to deal with the query descriptors. But what if D1 is 300 pages and D4 10 pages? We would have to run retrieval tests to answer the question, but they might well be inconclusive, or the results might vary from question to question.

There are many ranking methods used by search engines. Often used ranking criteria:

1 **The frequency of query words in the document**: How often does the query word occur in the document. This is called TF (Term Frequency. The more frequent the word, the more emphasis on that word in the document. Some ranking formulas distinguish just *1* and *more than 1* (2, 3, 4, etc. are all treated equal) to counteract spamming.

2 **The rarity of a query word**: A document gets more weight for a rare word than for a word occurring in many documents. Extremely common words are ignored for ranking. Rarity is computed as 1/DF, where DF (Document Frequency) is the number of documents in the collection that contain the term one or more times; this is also known as IDF (inverse document frequency)

3 **The absolute and relative location of query words**: A document gets more weight for a query word in the HTML title, in an HTML META tag, or in the top portion of the document than for a query word someplace down in the body of the document. Also, a document gets more weight if two query words occur close together.

1 and 2 are often used together. The contribution of a word in a document to the document's relevance score is computed as

TF * IDF (Term Frequency in the document * Inverse Document Frequency)

4 Search modes and data structures

To execute a search, a retrieval system must operate on stored data. The problem is to devise a data structure and a search process that makes retrieval fast. We will discuss data structures for Boolean retrieval and data structure in semantic networks.

4.1 Review the data structures described in Chapter 11

Textbook, Section 11.1. Key messages:

- Need an index for fast searching
- How the index is built. Takes time to build an index
- Index takes up space (not a big problem anymore)
- In database implementation: Build indexes for data fields that are searched

Note: Data structures in semantic networks were discussed in Lecture 2.2.

4.2 Further elaboration of data structures (Advanced, ➜UBLIS 506 Information Technology)

See Supplement SLecture 5.2

red

Weeks 6 - 7. February 24 March 9

Part 4
The nature, design, and management of documents and records

red

Things to do in Week 6, W February 24 - March 2

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	Assignment 6: Developing a conceptual data schema (assigned L4.2, `Feb. 17)	☐
	Four-week quiz (assigned `Feb. 17)	☐
	Assignments due W March 2 ☐ required ○ optional	√
Assignment(s)	Short description of term paper (assigned. L5.2, `Feb. 24)	☐

	New topics this week

	6.1a Document functions and document design (information design) for people. Knowledge (re)presentation in text, images, and multimedia . (30 min)	
Readings	1 Lectures6.1a-6.2b Objectives etc. (pink) and *Lecture6.1a* Objectives etc. (pink).	☐
	2 Lecture 6.1a Notes Section 1 and Section 2	☐
	3 Soergel, Dagobert. **The nature of texts**. 1999. 12 p.	○
	4 Mayer, Richard E. **The Balloons Passage: Understanding requires a schema**. From Mayer, R. E. Thinking, Problem-Solving, Cognition. NY: Freeman, 1983, p. 207- 208.)	☐
	5 Novak, J. D.; Cañas, A. J. **The Theory underlying concept maps and how to construct them**. Tech. Report IHMC Cmap Tools, Florida Institute for Human and Machine Cognition, 2006-01, rev. 2008-01, 33 p. Retrieved 2012-04-09**. Required p. 1-12, LMS all**. http://cmap.ihmc.us/publications/researchpapers/theorycmaps/theoryunderlyingconceptmaps.htm.	☐
	6 Keyes, E**. Information design: Maximizing the power and potential of electronic publishing equipment**. IEEE Transact. on Professional Communic. 30(1) (1987): 32-37.	☐
	7 Soergel, D., comp. **Useful document design guidelines**. 2011 7a Soergel, D. **Supplemental style rules.** 2011	☐
	8 Lynch, P. J.; Horton, S. **Web style guide. Basic design principles for creating Web sites**. 3. ed. New Haven, CT: Yale U. Press; 2009. 352 p . At www.webstyleguide.com Ch. 3. Site design.. Ch. 4. Page design. Ch.. Typography.. Find on Web,	○
	9 Meyer, Bonnie J. F. **Following the author's top-level organization: an important skill for Reading comprehension**. In: Understanding Readers' Understanding: Theory and Practice. Tierney, R. J.; Anders, P. L.; Mitchell, J. M., eds., p. Hillsdale, NJ: Erlbaum, 1986. p. 59 - 76. In print readings. **Recommended. LMS**	○
	10 Rumelhart, David E. & Norman, Donald H. **Accretion, tuning, and restructuring**. Semantic Factors in Cognition. Cotton, J. W. & Klatzky, R.L., eds. Hillsdale, N.J.: Erlbaum, 1978, p. 37 - 53. In print readings **Recommended for LMS**	○
	11 Killeen. GEO Directors Review (optional) (an example of well-designed slides)	○
Lecture	**Lecture 6.1a slides (40 min)**	☐

over

	6.1b Document macrostructure and inter-document relationships (40 m)	
Readings	*1 Lecture 6.1b Objectives etc. (pink)*	☐
	2 ***Reading after.*** *Document template system example. To be announced*	
Lecture	**Lecture 6.1b slides (30 min)**	☐

	6.2a Formatting documents for interpretation by computer programs. Document markup languages (25 min)	
Readings	1 Lecture 6.2a Objectives etc. (pink).	☐
	2 Lecture notes 6.2a Brief intro. and basic principles, Lecture Notes p. ~201 - 204	☐
Lecture	**Lecture 6.2a slides**	☐

	6.2b Document analysis for retrieval and information extraction (45 m) **[Syntactic and semantic parsing in Supplement]**	
Readings	1 Lecture 6.2b Objectives etc. (pink). Also have a look at Assignment 7.	☐
	2 Lecture notes 6.2b Sections 1, 2, and beginning of 3 p. ~219 - 223	☐
	3 **Practical applications of linguistic software**. Short version Skim, peruse examples. Illustrates practical importance of text analysis and NLP.	☐
	3a **Practical applications of linguistic software. A lot more detail.**	○
	4 Feldman, Susan **NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval**. ONLINE, May 1999. 23(3): 62-64,66-68,70-72. www.scism.sbu.ac.uk/inmandw/ir/jaberwocky.htm. In printed readings. **Recommended**	○
	5 Etzioni, O.; Banko, M.; Soderland, St.; Weld, D. S. **Open Information Extraction from the Web**. Comm.of the ACM December 2008 51(12): p. 68-74.	☐
	6 Crombie, Winifred, original author ; Soergel, Dagobert, adapter. **Semantic relations between propositions**. Original article 1985, adaptation 1998. 7 p. Shows how universal the entity-relationship approach is.	☐
	7 Shuldberg, K. H.; MacPherson, M.; Humphrey, P.; Corley, J. **Distilling information from text: The EDS TemplateFiller system.** JASIST. 1993.10; 44(9): 493-507.	○
	8 Allen, James. **Natural language understanding**. 2nd ed. Redwood City, Calif.: Benjamin/Cummings, 1995. TOC and Ch. 1 in printed readings Table of contents and Chapter 1. Intro. to natural language understanding, p. 1-17.	○
Lecture	**Lecture 6.2b slides (40 min)**	☐

Learning blog	Learning blog Week 2 due February 3	○

	Assignments assigned W March 2	
Assignments	Assignment 7. Design a document template (2 hours) (due `Mar. 9)	
Concl. readings	Readings that integrate material from Week 6.See Things to do in Week 7, at the top	

pink **Part 4** ***March 2- 9***

The nature, design, analysis, and management of documents and records

➜UBLIS 506 Information Technology

Lectures 6.1a - 6.2b ***March 2 - March 9***

Document function, structure, design, and analysis (No Textbook chapter)

Scope	This part of the course requires a clarification of the scope, particularly what is meant by "document," and how this topic is approached by many disciplines from many angles.
Broad definition of "document"	Text has been defined as: "Any passage, spoken or written, of whatever length, that [forms] a unified whole." (Halliday) Document: any presentation of information in any form: • written or spoken text, still or moving images, or music and sound (a multimedia document combines all of these) • in any medium – print, computer screen, TV, radio, etc. "Document" includes learning objects, websites, human-computer interfaces, etc. The boundary between document and information system is not sharp.
Disciplines/ fields dealing with information presentation/ document design	• Text linguistics, discourse analysis • Rhetoric, English composition • Document design, including Web design • Information architecture • User interface design • Instructional design • Advertising design • Graphics design, including, for example, guidelines for slides • Formatting documents for interpretation by computer programs
Disciplines/ fields dealing w/ document analysis	• Linguistics. Natural language processing. • Artificial intelligence, ontologies • Machine learning

Learning objectives and practical significance for Lectures 6.1a - 6.2b are given on the following pages.

Learning objectives, Lectures 6.1-6.2b (Learning objectives for all lectures are integrated here)	**1 Understand the importance of knowledge of document functions and document types for providing good and appropriate information service to users.** (P2.2.2,2.1) 1.1 Understand the functions of documents and be able to discern which function a user requires. (P2.2.2,2.1.1) 1.2 Be aware of the different types of text and the communication purposes they serve. (P2.2.2,2.1.2)	L6.1a
	2 Understand principles of and gain a feel for good document design – document structure and presentation where the internal logical/conceptual structure drives the external form / is conveyed well through the external form. **Understand the importance of good document design** for the efficient transmission of information, both for human understanding and for machine processing. (Lectures 1.1-2.2 and 4.2 are also relevant) (all, foundation in L6.1a) . (P2.2.2,2.2)	L6.1a
	3 Be able to apply understanding of information structure and principles of document design and text structure to the **analysis, critique, assessment and evaluation of information and its design in all forms**. (P2.2.2,2.3) 3.1 Be able to assess quality and usability of documents f. collection development. 3.2 Be able to match documents to user characteristics and needs.	L6.1 a- b
	4 Be able to create good design for new documents. (all, foundation in L6.1a) Includes 6.3 (P2.3.1,4)	L6.1a - L6.2b
	5 Understand specific issues in document structure and be able to apply this understanding to document analysis and design, including text processing techniques. (P2.3.1,5) Note: Document structure can and must be considered at different scales, from the macro-level of major document parts to the micro-level of sentence structure. Lectures 6.1b and 6.1a focus on macro- to midi-scale, Lecture 6.2b on micro-scale (but not exclusively)	L6.1b - L6.2b
	5.1 Understand the importance and usefulness of document templates that codify document structure. (P2.3.1,5.1)	L6.1b
	5.2 Understand document type / document template systems with hierarchy and hierarchical inheritance. (P2.3.1,5.2)	L6.1b
	5.3 Be able to design a document template or template system at the conceptual level so it can be implemented by the IT department. (P2.3.1,5.3)	L6.1b
	5.4 Understand the principles of markup languages (HTML, specialized markup languages defined using XML) and markup / template definition languages (XML) and their importance for the implementation of good document design. (P2.3.1,5.4)	L6.2a
	5.4.1 Understand the "levels" of markup physical, formal logical and content logical.	L6.2a

	5.4.2 Understand the significance of XML.	L6.2a
	5.4.3 Have a general idea how document templates can be defined using XML.	L6.2a
	5.5 Understand information reuse and repurposing: a single properly structured internal document can, through transformation, give rise to many external and internal representations of all or selected content. (P2.3.1,5.5)	L6.2a
	5.6 Be able to recognize a situation where information reuse and repurposing would be useful and suggest a solution in general terms so it can be implemented by the IT department. (P2.3.1,5.6)	L6.2a
	5.7 Understand the importance of noun phrases as carriers of meaning in English. Be able to give examples. (P2.3.1,5.7)	L6.2b
	5.8 Understand the characteristics of context that determine the meaning of an ambiguous word and how these characteristics can be used for automatic word sense disambiguation (WSD). (P2.3.1,5.8)	L6.2b
	5.9 Understand text coherence and cohesion and their role in text understanding by people and by computer programs. (P2.3.1,5.9)	L6.2b
	5.9.1 Understand how the idea of structuring information in frames or in a semantic network applies to discerning the structure of text. (P2.3.1,5.9.1)	L6.2b
	5.9.2 Understand the problems of anaphoric reference and co-reference and their impact on information retrieval, esp. proximity searching, and information extraction. Understand why resolving anaphoric references is important for extracting assertions from text. (P2.3.1,5.9.2)	L6.2b
	5.10 Understand patterns used in text to represent relationships and how such patterns can be used for information extraction. (P2.3.1,5.10)	L6.2b
	5.11 Understand information extraction into frames. (P2.3.1,5.11)	L6.2b
	6 Have a general understanding of prevalent document analysis methods and their uses. **6**.1 **Be aware of the many useful applications of document analysis in dealing with vast quantities of tex**t, including automatic indexing for information retrieval, automatic abstracting and summarization, information extraction, and automated translation. 6.2 Have a general understanding of the potential of natural language processing for providing better, more pinpointed answers to users. 6.3 **Be able to recognize situations where information extraction and other methods of document analysis would be useful**, to suggest the use of appropriate software, and to participate in the selection of such software. (P2.3.3,1) 6.4 Have a general idea of how a computer program could do word sense disambiguation (WSD). [6.5 [For advanced study: Have some idea how automatic syntactic parsing (sentence diagraming) and semantic parsing work] (Suppl.SL6.2b) (P2.3.3,1.1).	L6.2b

pink **Practical significance**	**Lectures 6.1a - 6.2b** especially in the context of the Web. Inherit down to each lecture
1 General	1.1 Well-designed information presentation helps people assimilate and understand information and thus to cope with the ever-increasing amounts of information needed to function in a modern society.
2 Document production	2.1 Assisting in the **manual production of documents** is a very important problem particularly in large organizations such as the World Bank. Also important in helping students studying English composition. Note: On a day-to-day level, **most information specialists** (including librarians) must **produce documents** all the time (seen the list below). 2.2 **Automatic** or computer-assisted **generation of text** and documents 2.3 Devising **guidelines for document design** (as in a text on English composition)
3 Document retrieval	3.1 **Structure for storing information and selecting specific document parts for retrieval and display**. Many organizations now organize their documents into large text databases searchable on an intranet. 3.2 Devising systems that help users to find just the right documents or portions of documents for a given purpose 3.3 Devising computer systems that can assess relevance as a user would, assimilate information from a document, abstract or index a document
4 Document analysis and assimilation	4.1 **Understanding how people process documents** (assess relevance, assimilate information from a document, write an abstract of a document, or index a document) 4.2 Serving as the user's agent in judging the relevance and appropriateness of a document to the user's situation (background and purpose) 4.3 **Guidance in the analysis of documents**. Reading and evaluating documents, for example, scientific articles or news stories, is much easier if one understands their structure. Document processing by human indexers or machine indexing systems is based on document structure.

Discussion question: Design of multimedia documents

What combination is best for given communicative task, for example, teaching a concept, persuading people to do something or quit doing something, or giving instructions on how to put a piece of furniture together or how to use of a device? Generalization of text, which refers only to language.

pink **Documents produced by information specialists**

1 Presentation of search results (bibliography or substantive data)
2 New acquisitions list
3 Guide to the library, instructional materials
4 Guide to information on . . .
6 Promotional materials
7 Library newsletter
8 Meeting notes
9 Classification schemes and thesauri

Library Web site

Increasingly, libraries set up Web sites for use by their patrons; these Web sites include (but are not limited to) the kind of information listed above. (Hint: other libraries' Web servers are very useful information sources, for example www.lib.uchicago.edu/LibInfo/)

Outline for Lectures 6.1a - 6.2b

Macro Overall structure of documents

Lecture 6.1a Document functions and document design (information design) for people
Knowledge (re)presentation in text, images, and multimedia (30 min)

Lecture 6.1b Document macrostructure and inter-document relationships (40 min.)
Document macrostructure. Document templates
Hypermedia
Inter-document (inter-textual) relationships

Macro/micro

Lecture 6.2a. Formatting documents for interpretation by computer programs.
Document markup languages (20 min, more fully covered in ➜ 506)

Micro Detail within a document

Lecture 6.2b. Document analysis for retrieval and information extraction
[Syntactic and semantic parsing in Suppl](40 min)

pink

pink **Lecture 6.1a** (30 min) ***March 2***

Document functions and document design (information design) for people
Knowledge (re)presentation in text, images, and multimedia

Learning objectives From Lectures 6.1a - 6.2b, p. ~156 - 157	**1 Understand document functions and text types.** (P2.2.2,2.1) 1.1 Understand the functions of documents and be able to discern which function a user requires. (P2.2.2,2.1.1) 1.2 Be aware of the different types of text and the communication purposes they serve. (P2.2.2,2.1.2)
	2 Understand principles of and gain a feel for good document design – document structure and presentation where the internal logical/conceptual structure drives the external form / is conveyed well through the external form. **Understand the importance of good document design** for the efficient transmission of information, both for human understanding and for machine processing. (Lectures 1.1-2.2 and 4.2 are also relevant) (all, foundation in L6.1a) . (P2.2.2,2.2)
	3 Be able to apply understanding of information structure and principles of document design and text structure to the **analysis, critique, assessment and evaluation of information and its design in all forms**. (P2.2.2,2.3) 3.1 Be able to assess quality and usability of documents for collection development. 3.2 Be able to match documents to user characteristics and needs.
	4 Be able to create good design for new documents. (all, foundation in L6.1a) Includes 6.3 (P2.3.1,4)
Practical significance Inherited from Lectures 6.1a - 6.2b **plus**	To provide users with documents, Web sites, screens, and other information representations that optimally support understanding, you must be able to select or create such documents or other representations. Understanding document design is important for 1 creating documents 2 selecting documents for a collection 3 matching documents with users, in particular 4 Matching text type with user needs is important in answering questions. This has implications for cataloging. ➜ UBLIS 518 Reference Sources and Services: Judging appropriateness

Context	Lecture 2.2 focused on knowledge representation in computer systems and in the brain, so did Lecture 4.2 on data schemas and formats. This sequence of lectures focuses on external knowledge representation with the purpose of helping people to assimilate knowledge. But there is much overlap and interrelationship: • Semantic networks are used as external knowledge representation in the form of concept maps. • Record formats can be useful for external representation. • Internal representation often serves the main purpose of letting a computer program create multiple external knowledge representations (Lecture 6.2a on XML).

Discussion question (for the whole section)

How could the principles of text structure be applied to the structure of images?

Outline

1 Document analysis. Document functions *Sections 1 and 2 are just for reading*

2 Document or text types (summarized from Beaugrande)

3 Some principles for good document design

4 Means of expression, media modalities

- 4.1 High-level means of expression
- 4.2 Low-level means of expression
- 4.3 Methods for indicating parts of a document (large or small)

5 Further elaboration of these principles through document design examples

Read Sections 1 and 2 beforehand

1 Document analysis. Document functions

Sections 1 and 2 are just for reading.

Perspectives in document analysis	
	1 Internal document structure 2 **External context or communicative situation of the document**: 2.1 the creator (writer/speaker/designer) and the peruser (reader/listener/viewer) and their relationship 2.2 the functions the document fulfills

Document functions	
1 Informing, educating	Enable the reader/viewer to construct, reconstruct, or otherwise update his or her own mental image, to make sense out of the message presented. 1.1 Reporting results of research or scholarly endeavors. Describing objects or events. 1.2 Educating: helping the reader understand a new field or topic. 1.3 Providing small pieces of information quickly as needed. 1.4 Reporting the discourse at a meeting and its results (minutes). 1.5 Laying out a plan of action. 1.6 Giving instructions/prescriptions/orders (transmitting norms) (also below). Informing documents may be designed for reading in context or to enable the reader/viewer to quickly locate a piece of information she needs.
2 Instructing, persuading	Creating or changing beliefs, attitudes, or behaviors. Persuasion. 2.1 Giving instructions/prescriptions/orders (transmitting norms) (repeated here). 2.2 Persuading somebody to do something (vote for a person or issue, buy something). 2.3 Assisting in the treatment of mental or behavioral disorders by enabling the reader/viewer/listener to construct or reconstruct emotional/attitudinal structures, including self-image (bibliotherapy).
3 Entertaining	Providing entertainment or enjoyment.
	A document can serve multiple functions, especially it can entertain while it educates (and educate the better for it, “edutainment”).

2 **Document or text types** (summarized from Beaugrande)

The type of a text is determined by its internal structure and the communicative situation, especially the function of text. Major text types are listed and defined briefly below. The viewer/reader/listener can process a document more efficiently if s/he knows the its type
➜UBLIS 518 Reference Sources and Services: Matching text type with user's purpose.

Major text types
(There are many, many text types; this list is just the tip of the iceberg.)

Type	**Examples**
Facet 1	
Descriptive:	Review article; newspaper article; dictionary definition
Argumentative:	Logical proof; legal argument
Didactic:	Textbook
Narrative:	Fairy tale; letter
Conversational:	Reference interview
Facet 2	
Literary:	Prose (e.g., Novel); Poetry (e.g., Limerick)
Scientific:	Research study report

The classification of text types parallels roughly the classification of functions, but there is not a perfect one-to-one correspondence; for example, a poem may educate or persuade or entertain, or all three at once.

Elaboration of text types adapted from Beaugrande *Text, discourse, and process*, VII.1.8
See supplement SLecture 6.1a.

Question

Why should a reference librarian understand document functions and text types?
How does this help with ascertaining the user's information need?
How does this help with selecting the right documents for the user?

3 Some principles for good document design

Know the reader	• Problem to be solved / task to be accomplished • Information need • Background knowledge
Content	Select the information carefully - only what the reader needs to know. Avoid redundancy or use it purposefully.
Structure	**Elaborate in your own mind the intrinsic structure of the topic** / the phenomena to be presented - good document design is grounded in a thorough understanding and structuring of the topic. **Choose the external representation structure that best facilitates the assimilation of the intrinsic structure — form follows function**. Examples of external representation structures: Plain text, typographically structured text (such as a list or a linear arrangement of a hierarchy), table, diagram, picture. **A general structure that is often useful** Make schemas explicit. Provide advance orienters ("*Tell them what you are going to tell them*"). Give the detail ("*Tell them*"). Provide opportunity for rehearsal or application to fix the new information in the reader's mind. ("*Tell them what you told them.*")
Layout	Provide guide posts that indicate the overall context . (for example, running heads, navigation chains as in Yahoo). Point out relationships (cross-references, links). In documents intended for looking things up, such as a dictionary: Provide guiding headings at the top of each page. (Counterexamples: Library of Congress Subject Headings, Dewey Decimal Classification index, MeSH.)

4 Means of expression, media modalities

4.1 High-level means of expression

	Non-linguistic (depicting)	**Linguistic** (text, verbal, convention)
Auditory Hearing/audio/ sound	Sound Music	Spoken language Speech
	"Audons"	
Visual Sight/vision/ graphical	Images, pictures (photos, paintings, drawings, charts, diagrams) "Visuals" (including real objects and models) Still Moving	Written (printed) language "Text" in Information Retrieval
	Icons Pictograms Sign language	
Touch/tactile	Tactile representations, for example a three-dimensional map	Braille
Other senses	Smell is of little practical significance here. The kinesthetic senses do not apply (except perhaps in virtual reality applications)	
Audiovisual	Sound and images simultaneously (as experienced in real life), often includes speech, may include written text.	
Multimedia, hypermedia	Combination of pieces of presentations in different modalities: Display of written text may be followed by still image, or a series of still images with explanation in speech, which in turn may be followed by an audiovisual segment. Multimedia kit. Hypermedia — interlinked segments in several modalities	

Consider	Medium Arsenal of artistic expression, visual vocabulary (icons, symbols in comic strips) Image structure Style	Language (Chinese, English, etc.) Vocabulary Text structure Style

4.2 Low-level means of expression

1 Typography: Type face, type size
2 Highlight or lead symbol (triangle, bullet, square, pointing hand, etc.)
3 Graphical means for highlighting or de-emphasizing (often used to distinguish between options that are available at the moment and those that are not)
    Bold, blinking (use sparingly), reverse on different background, black vs. gray
4 Color (but 8% of the population are color blind) (also for emphasis)
5 Boxes and other means of grouping

4.3 Methods for indicating parts of a document (large or small)

1 Explicit labels 2 Arrangement 3 Type face	See the examples for different methods of displaying a catalog record in the Textbook, p. 160 - 161.

****

5 Further elaboration of these principles through a series of

Document design examples

List of examples on next page.

Take this sheet out of your binder so you can look at the list of examples as we go through them

5 Further elaboration of these principles through a series of

Document design examples

Example 1	Two formats for salary data
Example 2a	Alphabetical vs. meaningful display (Art and Architecture Thesaurus)
Example 2b	Alphabetical vs. meaningful display (Art and Architecture Thesaurus)
Example 3	Examples from the Longman Lexicon of the English Language
Example 4	Display of information on buildings on a site in Perseus
Example 5	Two displays of the same hierarchy
Example 6	Two displays of a catalog record in a public library OPAC (Online Public Access Catalog)

In the Supplement, SLecture 6.1a

Example 7	Winners and losers in the forecasting game (from Tufte 1983)
Example 8	Thermal conductivity of tungsten: Arrangement of labels to facilitate interpretation (from Tufte 1983)
Example 9	Napoleon's campaign to Russia (from Tufte 1983)
Example 10	Classified arrangement of descriptors in a document record for indexing test (Alcohol and Other Drug Thesaurus)
Example 11	Contents page from *Alcohol Research*

The syllabus and lecture notes are an example of document design, using boxes, labels, comparative columns, tables showing a concept space that has two dimensions (such as the table in this lecture) and color and striving for consistent format. For example, first pages of lectures follow a common format, so do first pages of assignments. Also running heads as guide posts for orientation in the document.

Example 1. Minimize labeling, enable comparison

Poor example: Library Jobs by Level, ALA survey 2009

2009 *ALA-APA Salary Survey: Librarian – Public and Academic* (*Librarian Salary Survey*)

Job title	Average salary
Director/Dean/Chief Officer	
Public Libraries	83K
Academic Libraries	92K
Deputy/Associative/Assistant Director	
Public Libraries	73K
Academic Libraries	80K
Dept Head/Branch Mgr/Coordinator/Senior Mgr	
Public Libraries	63K
Academic Libraries	62K
Manager/Supervisor of Support Staff	
Public Libraries	53K
Academic Libraries	55K
Librarian Who Does Not Supervise	
Public Libraries	51K
Academic Libraries	55K
Beginning Librarian	
Public Libraries	46K
Academic Libraries	46K

http://www.ala-apa.org/salaries/SalarySummary2008.pdf (Data extracted from Tables 4 and 6)

Improved example: Same data, different arrangement

Job title	Public	Academic
Director/Dean/Chief Officer	83K	92K
Deputy/Associative/Assistant Director	73K	80K
Dept Head/Branch Mgr/Coordinator/Senior Mgr	63K	62K
Manager/Supervisor of Support Staff	53K	55K
Librarian Who Does Not Supervise	52K	55K
Beginning Librarian	46K	46K

Examples 2a and 2b. Meaningful arrangement

From the Art and Architecture Thesaurus (AAT)

Poor example: Alphabetical arrangement	Improved example: Meaningful arrangement
<size: photograph formats> double whole plate half plate mammoth plate ninth plate quarter plate sixteenth plate sixth plate whole plate	**size: photograph formats** sixteenth plate ninth plate sixth plate quarter plate half plate whole plate double whole plate mammoth plate
Art and Architecture Thesaurus sequence	**Suggested meaningful sequence**

Figure 1. Alphabetical vs. meaningful sequence on same hierarchical level
(Art and Architecture Thesaurus)

In the **art genres** example on the next page, notice the advantage of having definitions / scope notes for related terms right next to each other.

Explanation of **scope note**
In a thesaurus or classification there are often notes that explain the usage of a descriptor (or give a definition); these are called scope notes. Advanced: It is better to distinguish between definition of a descriptor and a usage note. Most may be needed to help indexers and searcher understand what a descriptor means and how it should be used in indexing and searching.

<*art genres*> academic art amateur art apocalyptic art art brut children's art commercial art community art SN Includes art undertaken in conjunction with particular communities, often socially deprived, usually with the idea of producing an effect or inspiring response specifically within those communities, with no reference to widely established standards. For art intended to beautify or enrich public places, use **public art**. computer art court art crafts cybernetic art didactic art dissident art ethnic art fantastic art figurative art folk art funerary art naive art nonrepresentational art primitive art public art SN Use for art whose purpose is to beautify and enrich public places. For art undertaken in conjunction with particular communities, usually to produce an effect or inspire response specifically within those communities, use **community art**. rock art cave art serial art sofa art street art	**art genres** . **art genres by content or other intrinsic characteristics** . . figurative art . . . fantastic art . . . apocalyptic art . . nonrepresentational art . . cybernetic art . . serial art . . crafts . **art genres by standard** . . academic art . . folk art . . dissident art . **art genres by type of artist or origin** . . amateur art . . naive art . . art brut . . children's art . . computer art . . ethnic art . . primitive art . **art genres by audience, purpose, or display context** . . sofa art . . court art . . public art . SN Art whose purpose is to beautify and enrich public places. . . . community art SN Public art undertaken in conjunction with particular communities, often socially deprived, usually with the idea of producing an effect or inspiring response specifically within those communities, with no reference to widely established standards. . . . street art . . rock art . . . cave art [prehistoric, esp. paleolithic] . . didactic art . . commercial art . . funerary art
a. **AAT sequence**	b. **Suggested meaningful sequence**

Figure 2. Alphabetical vs. meaningful sequence.

Example from the Art and Architecture Thesaurus (AAT)

Example 4. Tabular representation of data with same structure

Longman Lexicon

12 nouns & adjectives : family relations

relationship \ sex	*male*	*female*
parent	father	mother
child in relation to parents	son	daughter
child in relation to other children	brother	sister

father [C] the male parent: *His father is a shopkeeper. She doesn't get on well with her father.* **fatherly** [B] *usu apprec* like a (good) father: *He has a pleasant fatherly manner.* **fatherhood** [U] the state of being a father: *Fatherhood seems to suit him; he loves the baby.*

sire [C] **1** *old use* a father or male ancestor [⇨ C17] **2** the male parent of an animal: *These racehorses all have famous sires.*

mother [C] the female parent: *Her mother used to run a shop. He doesn't get on well with his mother.* **motherly** [B] *usu apprec* like a (good) mother: *She is a very motherly woman.* **motherhood** [U] the state of being a mother: *Motherhood seems to suit her; she looks happy.*

dam [C] **1** *old use* a mother or female ancestor **2** the female parent of an animal: *The dog's dam won many prizes.*

son 1 [C] a male child: *Mr and Mrs Brown have two sons. He was carrying his baby son.* (*fig*) *Robert Burns is one of Scotland's greatest sons.* **2** [N] *affec* (form of address to a boy or younger man): *Come on, son; let's go now.* **sonny** [N] (used in speaking to a young boy): *Better go home to your mother, sonny.*

daughter [C] **1** a female child: *Mr and Mrs Smith have three daughters. She had brought her teenage* [⇨ C9] *daughter with her.* (*fig*) *Joan of Arc was one of France's greatest daughters.* **2** (*fig*) something thought of as a daughter: *French is a daughter language of Latin.* **daughterly** [B] *usu apprec* like a (good) daughter: *She has a daughterly love for the old man.*

brother [C] a male relative with the same parents: *My older brother works in New York. His brothers are all younger than him/than he is.* (*fig*) *All the men in this association are brothers with the same aims. He is not accepted by his brother lawyers in town.* **brotherly** [B] *usu apprec* like a (good) brother: *He has a very brotherly manner towards the little boys.*

sister [C] a female relative with the same parents: *His older sister started a new job last week. She's prettier than both her sisters.* (*fig*) *French is a sister language of Spanish.* **sisterly** [B] *usu apprec* like a (good) sister: *She has a sisterly love for him.*

sibling [C] *tech* a brother or sister: *Siblings do not always like each other.*

paternal 1 [B] *sometimes deprec* of, like, or received from a father: *He has a very paternal attitude towards his students. His uncle gave him some paternal advice.* **2** [Wa5;A] on one's father's side of the family: *She is his paternal grandmother.* **paternity** [U] **1** fatherhood: *Paternity is a state which most men reach.* **2** origin on the father's side: *The paternity of the child is unknown.*

maternal 1 [B] of, like, or natural to a mother; (kind) like a mother: *She has a warm, maternal nature.* **2** [Wa5;A] on one's mother's side of the family: *He is her maternal grandfather.* **maternity 1** [U] motherhood: *Women are prepared by nature for maternity.* **2** [A] for the purposes of or relates to giving birth: *She is going to have the baby in a maternity hospital.*

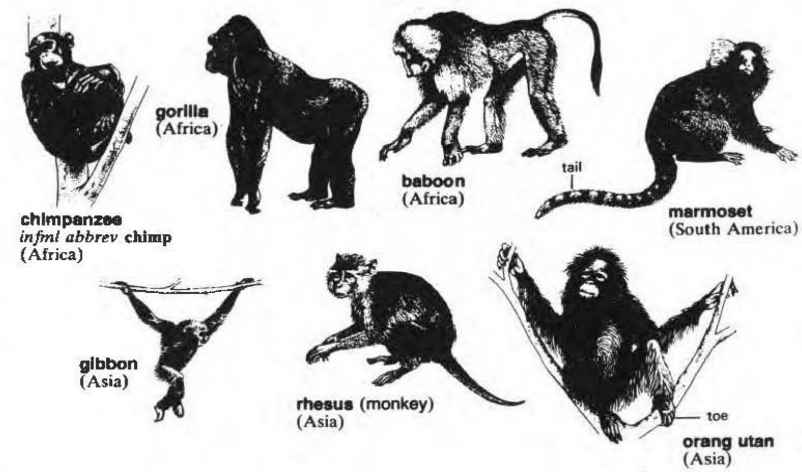
fraternal [B] *fml* of, like, or suitable to a brother: *They sent their fraternal greetings to him, as a member of the same group.*

filial [B] *fml* in the manner of a (good) son or daughter: *It is your filial duty to obey your father.*

C13 nouns : names for one's father and mother [N;C]

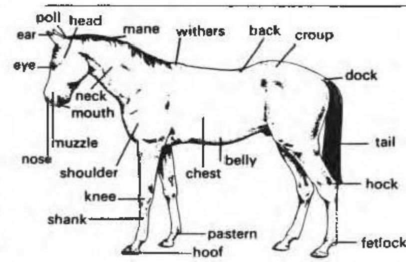
father

dad	*infml also with any much older man*	*Come on, Dad! The dads and mums were invited to see their children's work at the school*
daddy	*infml esp by & to children*	*Help me, Daddy! 'Is your daddy at home, dear?' he asked the little girl.*
pa	*infml esp AmE & BrE working class*	*Come on, Pa; time to go.*
papa	*old use esp by & to children, esp upper class*	*Papa, may we leave the room, please? Ask your papa about it, darling.*
poppa	*infml esp AmE*	*Come on, Poppa; let's go now. My poppa says I should work hard at school.*
pop	*infml esp AmE also with any much older man*	*Come on, Pop; let's have a drink.*



A51 nouns : the horse and similar animals [C]

[⇨ K199 HORSE RIDING, K200 HORSE RACING]



horse

Names for the horse according to age and sex		
age \ *sex*	*male*	*female*
full-grown	stallion	mare
young	colt	filly
very young newborn	foal	

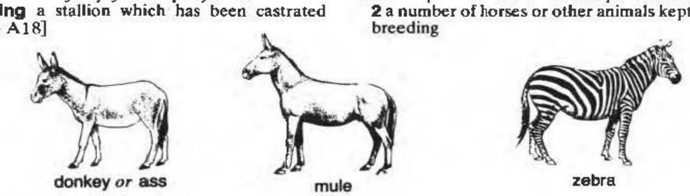
Other varieties and breeds of horses and similar animals

pony a horse of a small breed, esp as used by children: *The children enjoyed riding (on) the ponies. They enjoyed the pony rides.*

gelding a stallion which has been castrated [⇨ A18]

mustang a small wild horse of the North American plains

stud 1 *esp AmE* a male horse kept for breeding **2** a number of horses or other animals kept for breeding



donkey, ass (*fig*) a foolish person: *Don't be such an ass—It's a silly thing to do! He's a bit of a donkey; he does silly things.*

jackass 1 a male ass **2** (*fig*) a foolish person

mule a cross between a horse and a donkey: *The cart was pulled by two mules.*

zebra a striped animal from Africa, related to the horse

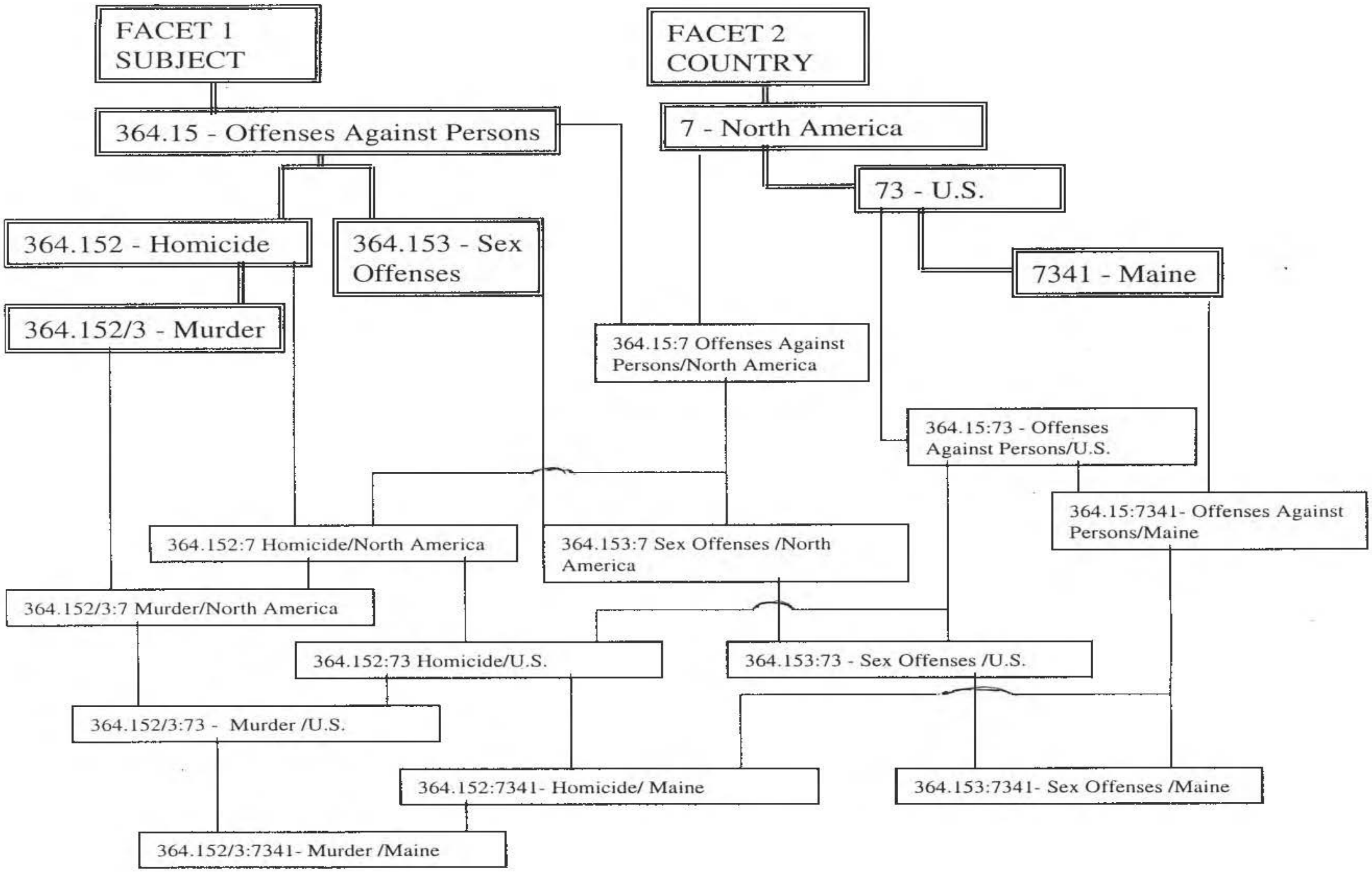
Results of a search for architecture (buildings) whose site is "Amphiaraion" in the region of Attica (from an old version of Perseus)

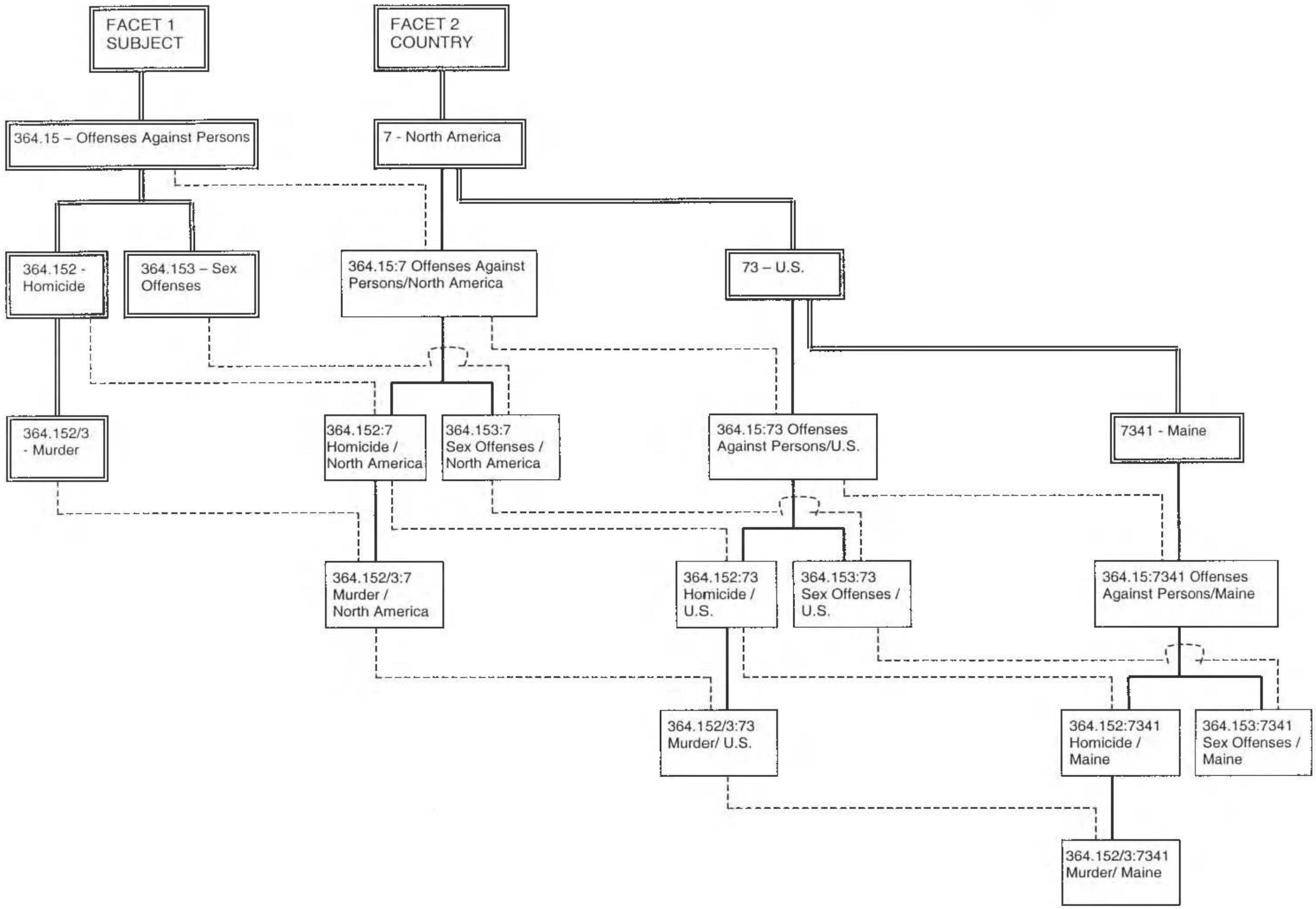
All buildings shown are in the region of Attica on the site Amphiaraion.

Name **All "Amphiaraion, ..."**	**Summary**	**Period**	**Type**
Earlier Temple of Amphiaraios	Small temple; on the western end of the Terrace of Dedications in the Sanctuary of Amphiaraios	Late Clas./Hell.	Temple
Klepsydra	Water clock and small annex; southeast of the Sanctuary of Amphiaraios, across the stream and east of the temple of Amphiaraios	Hellenistic	Klepsydra
Stoa	Stoa; on the east side of the Sanctuary of Amphiaraios, southeast of Theatre	Late Classical	Stoa
Temple of Amphiaraios	Temple; at the Western end of the Sanctuary of Amphiaraios	Hellenistic	Temple
Terrace of Dedications	Terrace with retaining wall; on the northwestern side of the Sanctuary of Amphiaraios	Late Classical	Terrace
Theater	Hellenistic theater; on the northwestern side of the Sanctuary of Amphiaraios, behind the west half of the Stoa	Hellenistic	Theater

There is a simple frame for each building; each row shows the frame for a single building, each column is a slot. If all buildings in a display have the same value, we do not need a column for that slot. In the example, the region slot for all buildings has the value Attica. In a display of all theatres, we could omit the column Type but should add a column Region.

Consumer reports contains many tables of test results that are built on the same principle. Features that have the same value for all tested products do not have a column.


FACET 1 SUBJECT
FACET 2 COUNTRY
364.15 - Offenses Against Persons
7 - North America
73 - U.S.
364.152 - Homicide
364.153 - Sex Offenses
7341 - Maine
364.152/3 - Murder
364.15:7 Offenses Against Persons/North America
364.15:73 - Offenses Against Persons/U.S.
364.15:7341- Offenses Against Persons/Maine
364.152:7 Homicide/North America
364.153:7 Sex Offenses /North America
364.152/3:7 Murder/North America
364.152:73 Homicide/U.S.
364.153:73 - Sex Offenses /U.S.
364.152/3:73 - Murder /U.S.
364.152:7341- Homicide/ Maine
364.153:7341- Sex Offenses /Maine
364.152/3:7341- Murder /Maine


FACET 1
SUBJECT
FACET 2
COUNTRY
364.15 – Offenses Against Persons
7 - North America
364.152 - Homicide
364.153 – Sex Offenses
364.15:7 Offenses Against Persons/North America
73 – U.S.
364.152/3 - Murder
364.152:7 Homicide / North America
364.153:7 Sex Offenses / North America
364.15:73 Offenses Against Persons/U.S.
7341 - Maine
364.152/3:7 Murder / North America
364.152:73 Homicide / U.S.
364.153:73 Sex Offenses / U.S.
364.15:7341 Offenses Against Persons/Maine
364.152/3:73 Murder/ U.S.
364.152:7341 Homicide / Maine
364.153:7341 Sex Offenses / Maine
364.152/3:7341 Murder/ Maine

**Example 6. Remove redundancy,
have each separate element stand out on its own for ease of scanning**

Poor example: Result display in the catalog of the Montgomery County Public Libraries using Sirsi Dynix www.montgomerycountymd.gov/libtmpl.asp?url=/content/libraries/find/findbooks.asp

J 599.789 BRE 2006
Giant pandas up close
Bredeson, Carmen.
24 copies available at Aspen Hill Library, Chevy Chase Library, Damascus Library, Davis Library, Fairland Library, Gaithersburg Library, Germantown Library, Kensington Park Library, Noyes Children's Library, Olney Library, Poolesville Library, Potomac Library, Quince Orchard Library, Rockville Library, Silver Spring Library, Twinbrook Library, White Oak Library, and Longbranch Library

Improved example

J 599.789 BRE 2006
Giant pandas up close
Bredeson, Carmen.

24 copies available at

Aspen Hill	Davis	Germantown	Olney	Potomac	Silver Spring
Chevy Chase	Fairland	Kensington	Park Noyes	Quince Orchard	Twinbrook
Damascus	Gaithersburg	Longbranch	Poolesville	Rockville	White Oak

Each library name should be a link to the page that shows location and opening hours

Example 6a
Good example: Arrange citations chronologically to trace the development of an idea in time

As detailed earlier in the section *A Brief Chronicle of Relevance Research*, Cuadra & Katter's ideas were taken up and further developed in the early 1990s when a large group of researchers took a naturalistic approach to studying relevance assessments by real users with a genuine information need in real situations and found a large number of factors beyond topicality which were less studied before, such as novelty, recency, and credibility (**Schamber, *et al.*, 1990; Schamber, 1991, 1994; Park, 1992, 1994; Barry, 1993, 1994; Cool, Belkin, & Kantor, 1993; Bruce, 1994; Froehlich, 1994; Hersh, 1994; Janes, 1994; Sutton, 1994; Wang, 1994; Barry & Schamber, 1998; Wang & Soergel, 1998; Wang & White, 1999; Choi & Rasmussen, 2003; Lawley, Soergel, & Huang, 2005; Xu & Chen, 2006)**. Together these studies identified over 80 "relevance factors" from users, as grouped in an often-cited table by Schamber (1994: p. 11). This table lists relevance criteria, types o f information used, and other factors affecting relevance judgments.

Poor example: EndNote and similar programs arrange citations in the text alphabetically by first author, which conveys no meaning.

pink **Lecture 6.1b** (40 min) ***March 2***

Document macrostructure, document templates
Inter-document relationships

➜ UBLIS 506 Information Technology

Learning objectives From objectives for Lectures 6.1a - 6.2b, p. ~156 - 157	**3** Be able to apply understanding of information structure and principles of document design and text structure to the **analysis, critique, assessment and evaluation of information and its design in all forms**. (P2.2.2,2.3) 3.1 Be able to assess quality and usability of documents for collection development. 3.2 Be able to match documents to user characteristics and needs.
	4 Be able to create good design for new documents. (all, foundation in L6.1a) Includes 6.3 (P2.3.1,4)
	5 Understand specific issues in document structure and be able to apply this understanding to document analysis and design, including text processing techniques. (P2.3.1,5)
	5.1 Understand the importance of document templates that codify document structure. (P2.3.1,5.1)
	5.2 Understand document type / document template systems with hierarchy and hierarchical inheritance. (P2.3.1,5.2)
	5.3 Be able to design a document template or template system at the conceptual level so it can be implemented by the IT department. (P2.3.1,5.3)

Over

Practical significance	1 Document templates make document creation so much easier and thus save a lot of work. 2 Good document structure makes reading and understanding documents easier. 3 Good document structure allows for pinpoint retrieval of relevant document sections. 4 Lesson plan templates facilitate creating and sharing lesson plans. The school library media specialist would be the person to introduce teachers to this idea and find good implementations for the school.55 5 Well-structured hypertext / hypermedia allows for reader-directed / learner-directed selection and sequencing of material. 6 If you see that the organization you are working in does not have efficient procedures for creating documents, suggest that they use templates and that they get a system that supports document creation using templates (even MS Word does in a limited way) and perhaps you will be promoted to Chief Information Officer (CIO).

Discussion questions on hypertext / hypermedia	1 In the context of a hypertext system, what is a text? (a) Each individual segment or (b) the total sequence of text segments (and perhaps images, etc.) the user/reader constructs in the interaction with the system? 2 How can we design hypermedia systems that support the user in constructing coherent documents? 3 When should sequence be in the writer's hands, and when should it be in the reader's hands?

Document/text macrostructure

Structure of a scientific text -
a frame for structuring information (in a full article or in an abstract)

One possible outline

1 **Background** (could also be called Problem)

1.1 General problem area (often including a review of the literature)

1.2 Specific problem. Purpose of the study, question to be answered

2 **Methods**

2.1 Discussion of the methods used in the study

2.2 Description of the actual conduct of the study

3 **Results**

4 **Conclusions**: Relationship to existing body of knowledge. Implications for decision making and/or further research

Knowing this structure makes it easier to read a journal article. Having a template with this structure makes it easier to write a journal article. The facing p. ~181 gives an example of an abstract using the general outline. All abstracts in the abstracting journal *Alcohol Research* follow this outline making it easy for the reader to quickly peruse each abstract.

Another list of journal article components
(from a study of the human indexing process, indexers look at these components to find index terms)

Journal title	Introduction
Title	Statement of purpose
Author	Materials and methods
Author's affiliation	Results and discussion
Keywords	Conclusions
Abstract	Figures, tables, and plates with captions
Table of contents (sometimes)	Acknowledgments
	Literature cited

Next page: Structured abstract from *Alcohol Research*, an extremely well designed abstracting journal.

CONIGRAVE KM abstract 1049

Conigrave KM, Saunders JB, Reznik RB. Predictive capacity of the AUDIT questionnaire for alcohol related harm. Addiction 90 (1995) 1479-1485.

'AUDIT can predict a range of harmful consequences of alcohol consumption'

Background

Drinking problems often are not recognized. Most of the people who become alcohol-dependent do not seek help until their problems are obvious. Late diagnosis is of particular concern because effective and low-cost methods of treating problem drinking at an early stage are now available. In 1989, the WHO published a brief 10-item screening questionnaire, the Alcohol Disorders Identification Test (AUDIT) specifically designed to identify problem drinkers before physical dependence or chronic problems have arisen. AUDIT has been reported to have a sensitivity of 92% and a specificity of 94% in detecting hazardous or harmful alcohol use. This study examined the ability of the AUDIT questionnaire to predict which subjects experience medical or social harm from their drinking.

Methods

Subjects were 350 patients who attended a hospital emergency ward in 1984-1985. They underwent a comprehensive assessment of medical history, alcohol use, dependence and related problems in an interview schedule; the AUDIT questions were interspersed among other items. Biochemical variables measured included y-glutamyltransferase (GGT) and mean corpuscular volume (MCV). Twenty subjects refused to be contacted after 2-3 years or were excluded because of malignant disease. Thus, a cohort of 330 subjects (212 men, 108 women) was left for the longitudinal study; 250 subjects were interviewed again after 2-3 years. Interviewers were blind to the results of the initial assessment. The AUDIT questions were scored from 0 to 4. Subjects who scored 8 or more were classified as potentially hazardous drinkers. AUDIT was examined for its ability to predict a number of end-points including alcohol-related medical disorders, health care utilization, social problems and hazardous drinking at the time of follow-up.

Results

Of those who scored 8 or more on AUDIT at the initial interview, 61% experienced alcohol-related social problems compared with 10% of those with lower scores. They also reported more frequently alcohol-related medical disorders and hospitalization. The AUDIT score was a better predictor of social problems and of hypertension than laboratory markers. Its ability to predict other alcohol-related illnesses was similar to the laboratory tests, but GGT was the only significant marker of mortality.

Conclusions

AUDIT is a brief and convenient questionnaire which can readily be incorporated into the standard medical history. It can predict a range of harmful consequences of alcohol consumption. AUDIT should prove a valuable tool in screening for hazardous and harmful alcohol use so that intervention can be provided to those at particular risk of adverse consequences.

K. M. Conigrave, Centre for Drug and Alcohol Studies, Royal Prince Alfred Hospital, Missenden Road, Sydney, NSW 2050, Australia.

From *Alcohol Research*

Preview of document templates: A simple mail merge example

The main document: A form letter, a specific case of a template

LITTLE PEOPLE SCHOOL

February 26, 2016

«NamePrefix» «FirstName» «LastName»
«Street»
«City», «State» «ZipCode»

Dear «NamePrefix» «LastName»,

According to our records, «StudentFirstName» does not have a current Emergency Card on file at our school. Because this form is essential to «GenderPossessive» safety while at the Lourie Center, «StudentFirstName» will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

Data source: A MS Access table

AddressTable

ID	Name Prefix	FirstName	LastName	Street	City	State	Zip Code	StudentFirst Name	Gender Possessive
1	Mr.	Eric	Smith	504 Flower Ct	Springfield	VA	22151	Rebecca	her
2	Mrs	Elizabeth	Kain	4801 Thames St.	Springfield	VA	22151	Alexander	his
3	Dr.	Sylvia	Campbell	3708 Duke St.	Alexandria	VA	22304	Mary	her

Resulting letters: See facing page and next page

The example shows how using a template, in this case a form letter, can save a lot of work. The form letter includes variables or slots whose values are filled in from the database. **Filling in data from a database wherever possible is key to efficient document creation**

LITTLE PEOPLE SCHOOL

February 26, 2016

Mr. Eric Smith
504 Flower Ct
Springfield, VA 22151

Dear Mr. Smith,

According to our records, Rebecca does not have a current Emergency Card on file at our school. Because this form is essential to her safety while at the Lourie Center, Rebecca will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

LITTLE PEOPLE SCHOOL

February 26, 2016

Mrs Elizabeth Kain
4801 Thames St.
Springfield, VA 22151

Dear Mrs Kain,

According to our records, Alexander does not have a current Emergency Card on file at our school. Because this form is essential to his safety while at the Lourie Center, Alexander will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

LITTLE PEOPLE SCHOOL

February 26, 2016

Dr. Sylvia Campbell
3708 Duke St.
Alexandria, VA 22304

Dear Dr. Campbell,

According to our records, Mary does not have a current Emergency Card on file at our school. Because this form is essential to her safety while at the Lourie Center, Mary will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

Note

Mail merge per se is not a topic in 571, just used as an example of document templates.
If for some other purpose you are interested in learning about mail merge, here is a useful introduction:
http://extension.oregonstate.edu/esoc/ectu/services/lessons/documents/MailMerge_000.pdf

I am also happy to send you the files I used for this example upon request.

Example. A simple document system

A frame/object hierarchy of document templates and documents

A document template is a frame with a slot (or element) for each part of the document (a part can be a single line or part of a line). Many slots have a procedure attached; the procedure obtains the information from a database, if it is available, or displays a menu of possible values, or asks the user a question. The document templates are arranged in a hierarchy, so that the slots in common to all documents of a class, such as meeting announcements, need to be specified only once; these slots then inherit down to all descendants of the class. **Hierarchical inheritance!**

Lecture 6.1b deals with implementing document templates in XML using XML schemas or the older Document Type Definition (DTD)

The simple document system consists of just five document types arranged in a hierarchy:

Generic memo
. Sales report memo
. . Content management sales report memo
. . Customer relations management sales report memo
. Self-assessment memo

.

For each document type, we give the template and a sample document. Conventions used:

Bold	A template slot (or element)
Arial	An instruction to be carried out when the template is applied to produce a document. Usually these instructions are attached to a slot.
<variable>	A variable to filled in with the appropriate value by the system
`Courier`	Text or data filled in by the system or selected by the user from a menu of options displayed by the system
Times Roman	Text entered by the user
Italics	Comments/explanations (not part of the document)
[ ], [[ ]]	Inherited, from one level up, two levels up Inheritance is indicated separately for the slot and the content of the slot (the slot may be inherited from the level above, yet the content can be specified at the current level)
/* . . . */	Comment

On each page: Top: example, Bottom: explanation

Document template 1: Generic memo

Subtype of / child of / inherits from:	Top level
Has subtypes / children / inherits to:	Sales report memo, Self-assessment memo
Metadata	
To:	
From:	<*name of person signed on to system*>, <*title of person*>
Subject:	
Date:	<*today's date*> /* from computer's clock */
Keywords:	
URI:	<*Universal Resource Identifier*> /* to be filled in by system */
MemoBody	
PlainText:	

Picture a company that makes software for business; *content management* and *customer relations management* are two of their product lines. (Just as a matter of interest. customer relations management software supports keeping customers happy, keeping track of their purchases and complaints, suggest new products they might want to buy, and managing email correspondence. One feature of such systems is this: Incoming email is analyzed for its general tone – friendly, laudatory, neutral, angry; an angry email message is moved up the queue for answering.)
Each of the product lines has a manager in charge. The director of sales needs to monitor sales in all areas, and so she must be updated every month on the sales in every product line and problems and opportunities. Having a sales memo template for each product lines greatly reduces the time it takes a product line manager to produce these memos every month.

This lecture introduces such a sales memo template step by step, starting from a plain memo template that should be familiar to you from sending email. Every time you send an email message you save time since the system fills in

1 your name and email address under From
2 the date.

Filling in data from a database wherever possible is key to efficient document creation.
Continued on facing page.

Document example 1: Generic memo

To: Sue Feldman, CIO (*Chief Information Officer*)

From: `Bob Boiko, content management specialist`

Subject: What XML (eXtensible Markup Language) can do for us

Date: `February 7, 2001`

Keywords: XML; content management; document structure; databases on the Web

URI: `www.jasca.com/bboiko/memo20010207-04`

XML allows us to define document structures that will make it easier to create documents. Once a document is created, it can be displayed in many different ways (Web page in multiple formats, print, etc.) through applying style sheets (the simple Cascading Style Sheets, CSS2, or the more powerful eXtensible Stylesheet Language for document Transformation, XSLT). A table of contents can be created automatically. Moreover, the document can be displayed selectively using just the parts most appropriate for a given audience. Parts of one document can be reused in another document. In retrieval, specific parts of the document can be targeted; for example, a user could search for just the *results* section of scientific reports.

With XML we can also define documents that hold database records to present databases on the Web. The boundary between text documents and formatted databases becomes blurred.

In the examples I use a hypothetical document management system with templating functionality. There are many real systems that can do most of the things described in the examples. The syntax is different for each system.
< > indicates a value the system will fill in, from the computer's clock or a database
/* */ indicates a comment that explains a feature of the template
The memo above is an example of a memo created using the template. Note the `different typeface` for values filled in by the computer system.
Read the text of the memo; it tells you something useful.

Document template 2: Sales report memo

Subtype of / child of / inherits from:	Generic memo
Has subtypes / children / inherits to:	Content management sales report memo Customer relations management sales report memo
[**Metadata**]	
[**To:**]	<*name of director of sales*>, <*value* = “Director of Sales”>
[**From:**]	[<*name of person signed on to system*>, <*title of person*>]
[**Subject:**]	/* to be filled in by memo designer of child template */ <last_month>
[**Date:**]	[<*today’s date*> /* from computer’s clock */]
[**Keywords:**]	
[**URI:**]	[<*Universal Resource Identifier*> /* to be filled in by system */]
[**MemoBody**]	
[**PlainText:**]	
Sales data table:	header <*value* = “Sales”> <*last_month*> <*value* = “in $1,000”> Run query /* query to be filled in by designer of child templates */
Data analysis:	
Recommendations:	

Creating document templates for the many kinds of documents in an organization is in itself laborious. Using the generic memo template as a starting point for more specific memos saves work. Slots inherit down from the generic memo to more specific memos, and so on in a hierarchy of document types and their associated templates. A prime example of **hierarchical inheritance**. This system has the further advantage that all memos have common features making it easier for the reader.

[ ] around a slot name indicates that the slot is inherited from the next level up

[ ] around a slot filler indicates that the slot filler is inherited from the next level up

Sometimes only a slot inherits, sometimes a slot with its value. For example, in all sales repot memos the To: slot is filled with the director of sales as found in the personnel database. The clause <*value* = “Director of Sales”> means that the string “Director of Sales” is to be displayed here in every memo created using this template

See bottom of page

The company in our example has many different products; for each of these a monthly sales report memo is needed. Document template 2 is merely a stepping stone to the more specific sales memo templates that follow. Defining a template that includes everything that is in common to all sales memos makes it easier to define the specific sales memo templates and makes sure that sales memos for all products share a common structure.

The features of this template are explained in the following examples where there is a sample report with text

Again, there is no document example. People just use this template to make more specific templates with values for their specific sales report already filled in, as in the template for Content management sales report memo. Making these specific templates is much easier if one can start from the more general sales memo template .

Document template 3: Content management sales report memo

Subtype of / child of / inherits from:	Sales report memo
Has subtypes / children / inherits to:	No children
[[**Metadata**]]	
[[**To:**]]	[<*name of director of sales*>, <*value* = "Director of Sales">]
[[**From:**]]	[[<*name of person signed on to system*>, <*title of person*>]]
[[**Subject:**]]	<*value* = "Content management sales report"> [<*last_month*>]
[[**Date:**]]	[[<*today's date*> /* from computer's clock */]]
[[**Keywords:**]]	<*value* = "content management software">
[[**URI:**]]	[[<*Universal Resource Identifier*> /* to be filled in by system */]]
[[**MemoBody**]]	
[[**PlainText:**]]	
[**Sales data table:**]	[header <*value* = "Sales"> <*last_month*> <*value* = "in $1,000">] [Run query] "monthly-CM-sales"
[**Data analysis:**]	
[**Recommendations:**]	

Underline: Added to the sales report memo template [[ ]] inherited from two levels up.

This system can run a database query specified in the template and insert the results into the document. **MS Office can include live database query results from MS Access in a MS Word document**. Again, filling in data from a database wherever possible is key to efficient document creation**.**

The system takes care of assembling and arranging all the data.
The sales manager can focus on **Data analysis** and **Recommendations** and writing the text

Again: **Templates and inheritance**. A slot defined in a broad template, such as the *generic memo* template, occurs in all subordinate templates, such a the *sales report* and *self-assessment memo* templates. The slot may inherit just as a bare shell for content (only the slot name is enclosed in []) or it may inherit with some or all of its content specifications, such as default value, limitations on values, or a procedure to be used to get the content (slot content specification enclosed in []). For example, the From slot always inherits down with the attached procedure: put in the name of the person signed on to the computer. The To slot inherits as an empty shell; the *sales report* template and the *self-assessment memo* template each has its own procedure for filling in a value. However, from *sales report* to *content management sales report* the To slot inherits with the attached procedure.

Document example 3: Content management sales report

To: Joe Bush, Director of Sales
From: Cindy Weaver, Sales Associate
Subject: Content management sales report January 2001
Date: February 5, 2001
Keywords: Content management software
URI: www.jasca.com/rweaver/memo20010210-13

Sales January 2001 in $1,000

		Fed. Gov.	State& local	Fortune 500	Small comp.	Total
TeamSite	Dec. 2000	500	150	700	200	1,550
	Jan. 2001	**700**	**200**	**900**	**300**	**2,100**
Templating	Dec. 2000	250	30	350	50	680
	Jan. 2001	**350**	**40**	**450**	**75**	**915**
Metatagger	Dec. 2000	100	20	200	30	350
	Jan. 2001	**150**	**30**	**250**	**50**	**480**
Metafinder	Dec. 2000	100	10	130	30	270
	Jan. 2001	**80**	**0**	**90**	**20**	**190**
Total	Dec. 2000	950	210	1,380	310	2,850
	Jan. 2001	**1,280**	**270**	**1,690**	**445**	**3,685**

Data analysis:
Smaller organizations make proportionately less use of Templating. Conversations with some customers showed that they do not have the expertise to construct sophisticated templates that would bring great efficiency to their work.
Sales of metafinder are languishing.

Recommendations:

Offer training in the use of Templating and also a consulting service where the consultant would set up the templates for use by the organization's staff.

Promote Metafinder more aggressively through demonstrations of search improvements achieved through its spelling correction and thesaurus lookup features. Also offer a large generic thesaurus with the software so that an organization does not have the expense of constructing its own thesaurus form scratch.

Document template 4: Customer relations management sales report memo

Subtype of / child of / inherits from:	Sales report memo
Has subtypes / children / inherits to:	No children
[[**Metadata**]]	
[[**To:**]]	[<*name of director of sales*>, <*value* = "Director of Sales">]
[[**From:**]]	[[<*name of person signed on to system*>, <*title of person*>]]
[[**Subject:**]]	<*value* = " Customer relations management sales report"> [<*last_month*>]
[[**Date:**]]	[[<*today's date*> /* from computer's clock */]]
[[**Keywords:**]]	<*value* = "Customer relations management software">
[[**URI:**]]	[[<*Universal Resource Identifier*> /* to be filled in by system */]]
[[**MemoBody**]]	
[[**PlainText:**]]	
[**Sales data table:**]	[header <*value* = "Sales"> <*last_month*> <*value* = "in $1,000">] [Run query] "monthly_CRM_sales"
[**Data analysis:**]	
[**Recommendations:**]	

Underline: Added to the sales report memo template [[ ]] inherited from two levels up.

Another example just to reinforce how this template system works.

Document example 4: Customer relations management sales report

To: Joe Bush, Director of Sales
From: James Barry, Sales Associate
Subject: Customer relations management sales report January 2001
Date: February 5, 2001
Keywords: Customer relations management software
URI: www.jasca.com/jbarry/memo20010210-13
Sales January 2001 in $1,000

		Fed. Gov.	State& local	Fortun e 500	Small comp.	Total
Product 1	Dec. 2000	500	150	700	200	1,550
	Jan. 2001	**700**	**200**	**900**	**300**	**2,100**
Product 2	Dec. 2000	250	30	350	50	680
	Jan. 2001	**350**	**40**	**450**	**75**	**915**
Product 3	Dec. 2000	100	20	200	30	350
	Jan. 2001	**150**	**30**	**250**	**50**	**480**
Product 4	Dec. 2000	100	10	130	30	270
	Jan. 2001	**80**	**0**	**90**	**20**	**190**
Total	Dec. 2000	950	210	1,380	310	2,850
	Jan. 2001	**1,280**	**270**	**1,690**	**445**	**3,685**

Data analysis:

Smaller organizations make proportionately less use of Product 2. Conversations with some customers showed that they do not have the expertise or training that would allow them to utilize the software.

Sales of Product 4 are very low.

Recommendations:

Offer customer training and a consulting service.

Promote Product 4 through demonstrations, perform user studies and solicit feedback.

Document template 5: Self-assessment memo

Subtype of / child of / inherits from:	Generic memo
Has subtypes / children / inherits to:	No children
[**Metadata**]	
[**To:**]	<*supervisor of person signed on to the system*>, <*title of supervisor*>
[**From:**]	[<*name of person signed on to system*>, <*title of person*>]
[**Subject:**]	<*value* = “Self-assessment for year”> <*LastYear*>
[**Date:**]	[<*today’s date*> /* from computer’s clock */]
[**Keywords:**]	<*some subject keywords filled in from job description in database*>
[**URI:**]	[<*Universal Resource Identifier*> /* to be filled in by system */]
[**MemoBody**]	
[**PlainText:**]	
Accomplishments:	header <*value* = “Accomplishments in year”> <*LastYear*>
Goals:	header <*value* = “Goals for year”> <*ThisYear r*>
Training needs:	header <*value* = “Training needs for year”> <*ThisYear*>

Another document type with its template.

Using the self assessment template has many advantages, among them:

- The supervisor can assimilate each of the memos more quickly
- The system can produce a report that shows just the goals for each employee s the supervisor can compare the goals
- The system can compile the **Training need** sections from all self-assessment memos (and even use text analysis to sort these training needs into categories. This report will be very useful for the company's training coordinator

.

Document example 5: Self-assessment memo

To:	`Sue Feldman, CIO`
From:	`Bob Boiko, content management specialist`
Subject:	`Self-assessment for year 2012`
Date:	`February 7, 2001`
Keywords:	Content management; planning; XML; intranet; Web site
URI:	`www.jasca.com/bboiko/memo20140115-07`

Accomplishments in year 2012:

Developed a content management master plan.

Started the development of logical templates for the most important document types and implementation through XML document type definitions.

Developed specifications for the acquisition of content management software and selected a vendor.

Goals for year 2014:

Begin implementation of the content management master plan.

Install software and train staff in intranet-based document creation, deployment, and search.

Redesign the company Web site and use new software to streamline deployment of content on the Web

Training needs:

A course in information architecture

A course in advanced methods in XML, including XLink, XPointer, XPath, and XSLT (eXtensible Stylesheet Language for document Transformation)

For another example of a document template hierarchy, see supplement SLecture 6.1b

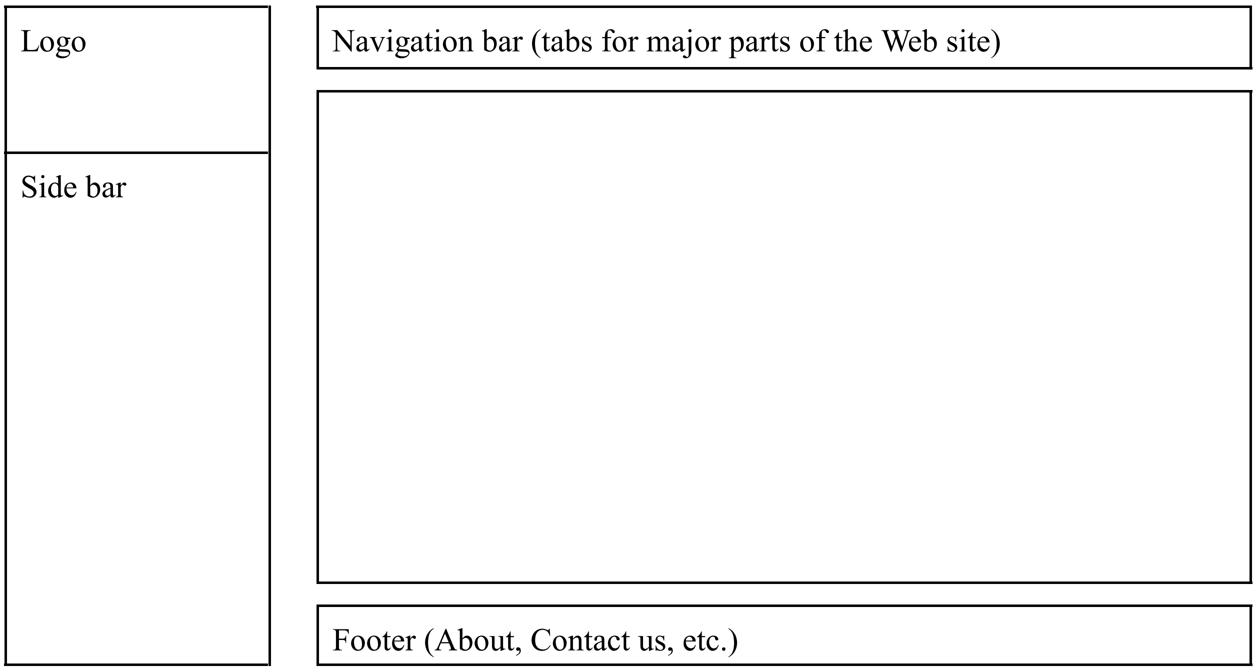
Web templates *Just read*

Web templates are very useful for creating and maintaining Web pages. The templates shown have slots and fillers just like the memo templates, but the focus here is on the display area allocated to the data in each slot. The underlying definition is not shown.

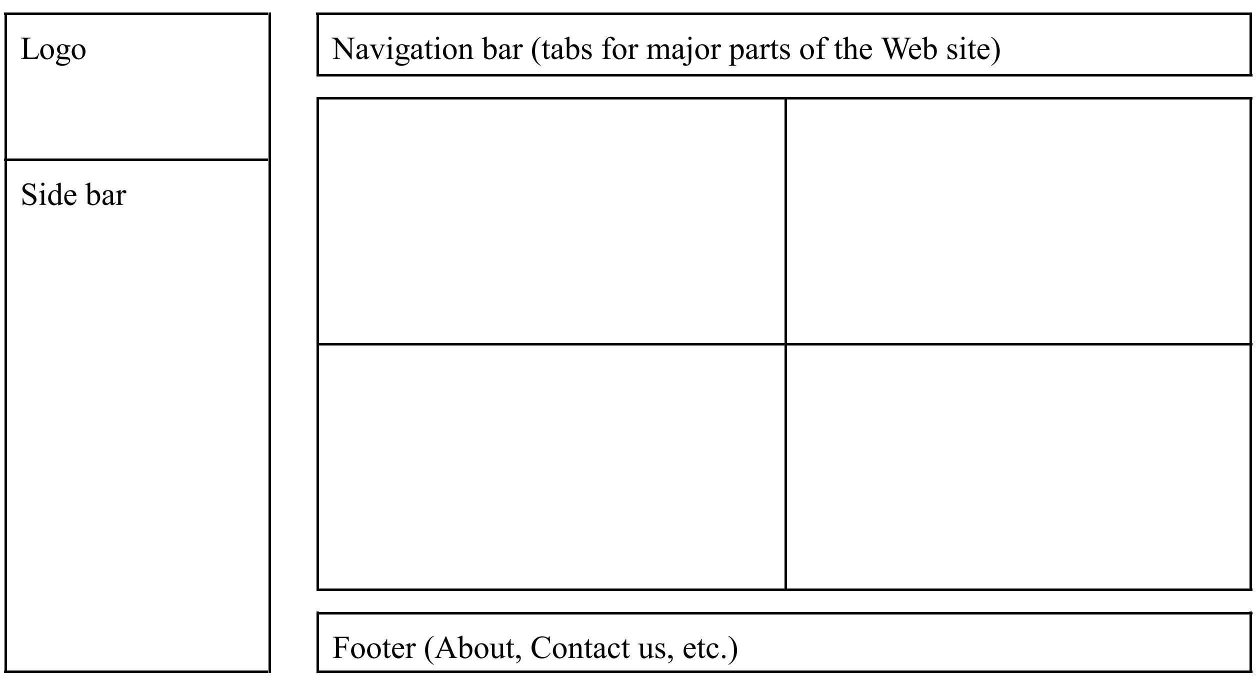
When one of these templates is used to create a Web page, each area of the template is linked to a document to be displayed in that area. This may be any multimedia document (text, image, or combination) or it may be a report that is created dynamically from a database. When a document is updated, the update is reflected immediately in all Web pages that link to the document.

In the design of an entire web**site**, you would include both Web **page** templates so that someone adding a new Web page could choose the template most suitable for the Web page to be created. You might even have more than two Web page templates.

Web template 1



Web template 2



Hypermedia/hypertext ➜ UBLIS 506 Information Technology *Just read*

Linear text vs. hypertext	**Typical text is linear in a sequence set by author**: "Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop." Lewis Carroll, Alice in Wonderland, Chapter XII **Hypertext / hypermedia is a collection of text pieces (and images and sound files) with links; the reader can / must establish her own order through the text** (if indeed the reader goes through the text); this is accomplished by treating the text in blocks (or by establishing nodes/locations within the document) and by supplying/permitting links between nodes by which the reader can navigate the text in his or her own order; the reader constructs his or her own text. A hypertext can give suggested linear sequences: <next> and <previous> links.
Major features of hypertext	• fragmented non-linear text form whose components can be rapidly accessed via machine-supported links/relationships under direction of user • interactive • malleable, modular: it is easy to add or revise small pieces • no strong document boundaries (at least in large hypertexts)

Hypertext examples	
World Wide Web Wikipedia	Primary example of hypertext: the **World Wide Web**; the presence of links gives the web metaphor. Hypertext functionality has existed long before, e.g., a research paper with footnotes/ bibliography/ tables/figures; the Web makes links convenient to use.
A holy book in hypertext format	such as the Koran, the Bible, the Sruti, the Tipitaka, the Five Classics, Bk of Mormon (see http://en.wikipedia.org/wiki/Religious_text) (or several combined), translations Links from chapter, group of verses, verse, or word to • "Original" version(s); manuscript image(s) • Alternative translations • Other holy book passages that are related in some way (within one book or across) • Commentary passages • Sermons about the passage (published or own) • Entry in Hebrew/Greek/Arabic/Sanskrit/Pali/Chinese dictionary/grammar • Map • Archaeological evidence
Fiction	"interactive fiction," "Choose your own adventure", adventure games

Inter-document structures, relationships between works (for more, see Lect. 4.1, Reading 2)

Relationships. Mentioned in cataloging rules	Continuation / sequel Answer key Parody	Critical review Abstract Concordance	Index Bibliography Translation	Citation relationships see Suppl. SLecture 6.1b

pink **Lecture 6.2a** (20 min)(Very brief, see ➜506 for more detail) ***March 2***

Formatting documents for interpretation by computer programs. Document markup languages

HTML (Hypertext Markup Language) and XML (eXtensible Markup Language)

Learning objectives (from Lectures 6.1a-6.2b p. ~156 - 157)	4 **Be able to create good design for new documents**. (P2.3.1,4)
	5 **Understand specific issues in document structure and be able to apply this understanding to document analysis and design, including text processing techniques.** (P2.3.1,5)
	5.4 Understand the principles of markup languages (HTML, specialized markup languages defined using XML) and markup / template definition languages (XML) and their importance for the implementation of good document design. (P2.3.1,5.4)
	5.5 Understand information reuse and repurposing: a single properly structured internal document can, through transformation, give rise to many external and internal representations of all or selected content. (P2.3.1,5.5)
	5.6 Be able to recognize a situation where information reuse and repurposing would be useful and suggest a solution in general terms so it can be implemented by the IT department. (P2.3.1,5.6)
Learning objectives Elaboration	**You will be expected to know the principles of XML, in particular that XML** • supports the definition of tailor-made templates that divide a document into meaningful sections with all the advantages of templates (Lecture 6.1b); • allows for transforming documents in many different ways for display or export to other systems (information reuse and repurposing; • supports a nested structure of document sections You will be expected to know how to use these characteristics of XML strategically in document management and information reuse in an organization. You **will not** be expected to know the details of defining document templates in XML or creating an XSLT style sheet for document display or transformation. XSLT = eXtensible Stylesheet Language Transformations)
Practical significance	See next page

Practical significance	**Databases of machine-readable tex**t are undergoing an explosion, not only on the Web, but also in intranets and in efforts of creating large text corpora for linguistic and literary studies (the *Text Encoding Initiative*). Conventions for marking the structure of documents are a prerequisite for creating such databases and for common access and data exchange. Academic librarians need to assist users in accessing and working with such texts and might work on creating text databases. Note: A **text corpus** is a (usually large) body of text in digital form, often with annotations, such as indicating the meaning of each homonym. Examples: The Brown corpus, *A Standard Corpus of Present-Day Edited American English,*, originally created in 1964 at Brown University and updated several times, see http://en.wikipedia.org/wiki/Brown_Corpus Also see www.lancs.ac.uk/fss/courses/ling/corpus/blue/l02_1.htm XML markup enables **reuse or repurposing** of all kinds of text, images, and data, bringing a whole new dimension to content management. **Beyond marking up text, markup languages are now expanded to specify any kind of data structure**, blurring the boundary between text and formatted data. There are database management systems that format data with XML-defined tags, for example http://mlu.marklogic.com/ . See Lecture 5.1 for the rise of linked data on the Web.

Outline

Brief introduction and basic principles

Definition and general introduction
Principles
HTML and XML

Examples

HTML example (simple)
 Document with HTML tags ("under the hood"), done directly by the author
 Document displayed

XML example
 a XML schema, defining document templates
 b Document with XML tags
 c Style sheet defining appearance
 d Document with HTML tags ("under the hood") produced from XML document
 e Document displayed

Note: a-c give XML its power. Not used in the much simpler but less powerful HTML. Style sheets can be used in HTML to display the same HTML document in different ways (Cascading Style Sheet, CSS); not shown in this lecture.

Brief introduction and basic principles

Just read through p. ~204

Definition and general introduction

Definition	Markup is the insertion of markup tags (codes) into a document text or other data stream to specify a structure such as: identify a phrase as the name of an organization, identify a part of the text as the introduction, or identify a text segment as <BOLD>. This structure can then be used for further processing: manipulating data, rearranging text and images, and controlling the appearance (or rendering) of a document when it is printed or displayed on a screen. Note 1: ***Tag*** **has many meanings**, in the context of UBLIS 571 primarily two: (1) **Mark-up tag.** A string inserted into a document text to specify a structure. Meaning of "tag" in this lecture. (2) **Indexing tag**. An index term assigned to a text document, an image, or whatever, especially if the assignment is made by a user (not a cataloger/indexer), as in social tagging. Note 2: The term *markup* derives from typesetting. An editor put marks in a manuscript that specified for the typesetter the fonts to be used for a portion of text and other matters of appearance. The meaning of the term has much expanded since then, particularly in the last few years.
General introduction	HTML markup tags are designed primarily to direct the display of documents. HTML tags also specify links to other documents to be included automatically at display time (such as images) or available to the user by clicking on the link symbol. Tags defined through XML are much more powerful for expressing document and data structure.

Principles

Physical markup	Tags specify actual appearance properties, such as ***<indent .3">***, ***<center>***, ***<bold>***, ***<font*** Times Roman 12> Problem: What if the display device cannot show Times Roman?
Logical markup	Tags specify the logical structure of the document, including importance of certain pieces of text. The display is done by a program, possibly in conjunction with style sheets, that renders logical elements in a format determined at output time. **Formal (or syntactic) logical elements**. Tags specify formal units such as ***<heading level 1>***, ***<paragraph>***, ***<numbered list>***, ***<emphasize>*** **Content logical elements**. Tags specify content units such as ***<from>***, ***<to>***, ***<subject>***, ***<recommendations>***, ***<warning>***, ***<methods>***, ***<conclusion>*** defining the content structure of a document. These tags can be used to define record formats even for highly structured data. (XML is used increasingly as a language to define the structure of data in Web-based database applications.) Logical markup allows for rearranging and repurposing complex content and for more precise retrieval (for example, search only in the conclusion of documents). The **display program then determines the physical appearance** in accordance with the capabilities of the display device and the preferences set by the user. **Examples:** A ***<heading level 1>*** may appear in Times Roman 16 pt bold or in all caps. A new ***<paragraph>*** may start with a blank line and no indention (block style) or without a blank line with the first line indented. The document element ***<warning>*** may be displayed in a box with light gray background and a heading **Warning**. Since logical content markup makes the logical structure of a document explicit, it can be used for information organization and retrieval as well. It can be used to define record formats for straightforward data to be processed by a database management system or to define templates for complex documents (see the examples in Lecture 6.1b). Organizations use markup languages defined in XML to organize large databases of document content, including text and images.

HTML and XML

HTML	HTML is a markup language; all tags are predefined. HTML emphasizes logical markup, but the logical elements are primarily formal, and HTML includes an increasing number of physical markup tags (but still not enough to provide tight control over the appearance of a page). An author uses HTML tags to describe the way she wishes the page to display, but parsing and interpretation of the HTML tags is dependent on the Web browser used to display the page. The browser may or may not implement all the features in the same way. For example, look at a complex web page side-by-side with Internet Explorer and FireFox.
XML	XML (see the main XML Web site at www.w3.org/XML/) is not a markup language but a language that can be used to define one's own tags, one's own markup language; XML is a **markup metalanguage**: **there are no predefined tags; authors and system administrators define their own tags**. Many specific markup languages can be defined using XML This makes it possible to represent more of a document's semantic structure than HTML does. HTML is one of many markup languages that can be defined in terms of XML. **Standards expressed in terms of XML**. There are many domains where multiple users have similar kind of documents or data. They need a format for structuring these documents, for metadata describing documents, and substantive data. There are many communities that use XML to define markup languages (domain-specific tagging schemes) for their own domain (with discussion in the whole user community) as a standard to be used by the community; examples are MathML, NewsML, HR-XML for human resource data, etc., financial documents or biological processes (tags for structuring data). This saves thousands of people from having to "reinvent the wheel" for their domain.

The HTML example on the next page is part of the introduction.

HTML example (With XML, more steps are required, see XML process diagram)
Should give even a completer novice a sense of the structure of an HTML document.

Document with HTML tags ("under the hood") (File **d** in the XML process diagram)
Document pieces are indicated by a begin tag < > and a corresponding end tag </>
Here: ***bold*** = this is a tag, non-bold part is explanation, e.g <***H**eading**1***>, <***H1***> is the tag
Read the text; it has useful information

<HTML>
  <HEAD> /* Note: The <HEAD> section is not displayed. It has metadata*/
    <TITLE>What XML can do for us***</TITLE>***
    <META NAME="creator" ***CONTENT***="Bob Boiko">
    <META NAME="keywords" ***CONTENT***="XML; content management; document structure; databases on the Web">
  </HEAD>
  <BODY> /* Note: The <BODY> section is displayed.*/
    <*H*eading*1*><*C*enter> Memorandum ***</Center></H*eading*1*>**
    To: Sue Feldman, CIO <Line***BR***eak>From: Bob Boiko***
*** /* Note: there is no </**BR**> */
    Date: February 7, 2003***

***
    Subject: **<*EM*phasize>** What XML can do for us**</*EM*phasize>**
    <*P*aragraph>XML allows us to define document structures that will make it easier to create documents. Once a document is created, it can be displayed in many different ways (Web page in multiple formats, print, etc.) through applying style sheets (the simple Cascading Style Sheets, CSS2, or the more powerful eXtensible Stylesheet Language for document Transformation, XSLT). A table of contents can be created automatically. Moreover, the document can be displayed selectively using just the parts most appropriate for a given audience. Parts of one document can be reused in another document. . . .**</*P*aragraph>**
  </BODY>
</HTML>

Document displayed by the Web browser under the control of HTML tags (File **e** in the diagram)

Memorandum

To: Sue Feldman, CIO
From: Bob Boiko
Date: February 7, 2003
Subject: **What XML can do for us**

XML allows us to define document structures that will make it easier to create documents. Once a document is created, it can be displayed in many different ways (Web page in multiple formats, print, etc.) through applying style sheets (the simple Cascading Style Sheets, CSS2, or the more powerful eXtensible Stylesheet Language for document Transformation, XSLT). A table of contents can be created automatically. Moreover, the document can be displayed selectively using just the parts most appropriate for a given audience. Parts of one document can be reused in another document. . . .

End introduction

Start of explanation of XML. There are three things you need to attend to for best understanding, so you want to have them spread out to look at:

(1) P. ~205 - 208 give explanations; take them out of the ring binder.
(2) P. ~209 shows the all-important figure **XML Files and Process.** This figure gives the overall framework; each kind of file is illustrated by an example. You need to refer to the figure throughout. Take it out of the ring binder.
(3) Example files start at p. ~211; leave them in the binder.
The explanations take you through the examples.

This is a short intro to XML	This is a very short introduction to XML and how XML documents are displayed. You should have learned this in UBLIS 506. In XML one can define one's own tags corresponding to meaningful sections of a document; thus one can define document templates. Often it is possible to copy or modify someone else's template, but here we start from scratch.
Verbal outline of the XML schema definition, documents and process (on p. ~209 as a diagram)	Each box in this figure is a kind of file giving information needed in the process, as explained with examples on the pages that follow it. **Outline** a XML on its own has no tags. To create documents (such as the memos in Lecture 6.1b), one must first define appropriate tags in an *XML schema definition*; together the tags specify a *document schema or template*. b Now many *document instances* (individual documents in which the template slots are filled with text, images, or sound) can be produced using the tags defined. c But how to display these document instances, or documents for short? The browser does not understand the tags we defined. So the document must be transformed into an HTML document that the browser will display in the desired appearance. The instructions for this transformation are given in an XSLT *style sheet* (c). An XSLT processing program can read the instructions in the XSLT style sheet and apply them to a document instances. A document instance includes the location of the style sheet **XSLT** (= eXtensible Stylesheet Language Transformations d The result is a document with HTML tags, e which the browser then displays on the screen
a, XML schema definition, p. ~212	The box here and p. ~213 (with picture) tell the same story. Read both in either sequence. The audio tells the same story in a still different way. An XML document consists of *elements* that are nested into each other; each element is enclosed in a pair begin tag **. . .** end tag begin tag end tag (starts with </) ***<someName>* . . . *</someName>***

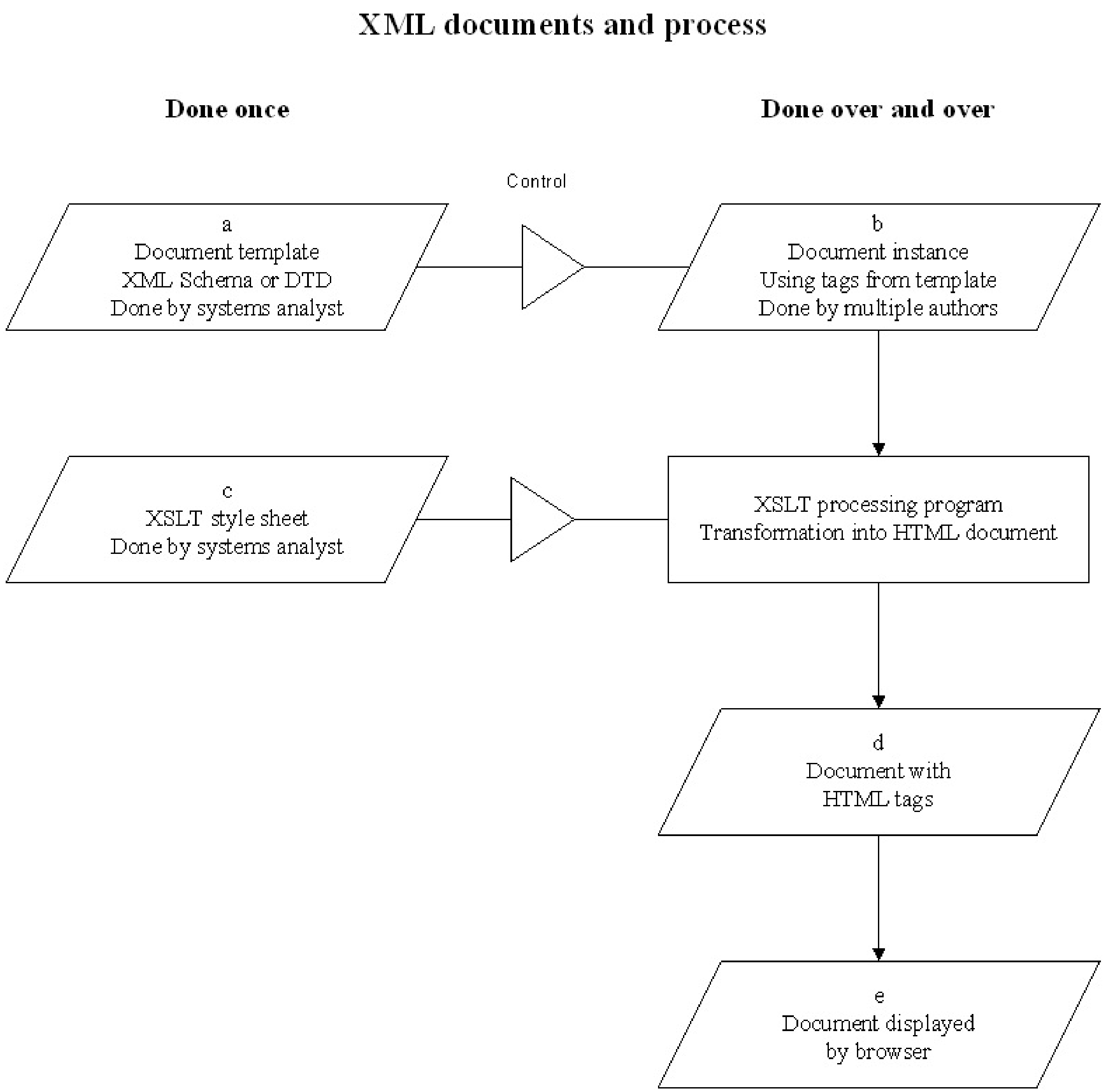
<table>
<tr>
<td>a,
XML
schema
definition,
continued</td>
<td>
The outermost element is the document itself, named memo. Elements have a name, that will be the name of the tag. Now that we have told the system that there is a big box, the element called memo, we need to tell it what is inside the box. This is done by associating the element with a type; the type specifies what is in the box. Each type needs a name, such as memoContent or metadataContent. Looking at the definition of memoContent, you (and the system) can tell that the outer box memo has nested in it two boxes called metadata and memoBody. But what is in the box labeled with tag <metadata>? To find out, look at the definition of the type metadataContent; you can see the sequence of elements to, from, subject, date, and keyword. The definitions of all document's elements make up a document template (document schema, document type). The schema definition is a special kind of XML document using tags predefined by the XML designers and prefixed by xsd: (XML Schema Definition). An XML processor reads an XML schema document and stores an internal definition of the schema / document template.
File a is such a schema definition. We now revisit it from the beginning:
Line 1 gives the XML version (versions 1.0 and 1.1 are both used)
Line 2 says this file is an XML schema definition.
Line 3 is a comment <! … >.
The following lines define elements, starting with the most encompassing element memo, the document itself. We first give the element a name (used as a tag that identifies the element in document instances) and assign it a type
<p><i><xsd:element name</i>="memo" <i>type</i>="memoContent"/></p>
The type specifies the internal structure or content of the element. The type needs a name as illustrated above. (Note on syntax: the predefined tag <xsd:element ... /> is begin and end at the same time, so it encloses no text but has information inside it, called "parameters", in this case "name" and "type".)
<p>To define the <i>type memoContent</i>, we use this predefined format:
<i><xsd:complexType name</i>="memoContent">
. . .
<i></xsd:complexType</i>></p>
The structure or content of memo consists of a sequence of two elements, named metadata and memoBody. Of course, for each of these elements we need to indicate its content by specifying a type.
The definition of metadataContent follows the pattern you just saw: Its structure is a sequence of five elements, each having a type. There is one difference: Defining types has to stop somewhere, otherwise we would end up in an infinite regress. Some basic types are pre-defined in XML. We use the built-in types string and date; you can guess from the names what they mean.
You should be able to follow the rest of the schema definition.
File a, which defines the tags for memo, is stored in www.jasca.com/cm/memo.xs
</td>
</tr>
</table>

Document b p. ~212	This is simply a document instance using the tags we just defined (which together make up a document template). In a real system there would be many such document instances. The first two tags are marked as special by the ?. They tell the version of XML used and what style sheet (see below) to use for display. The first regular tag **<*memo xmlns*=**"www.jasca.com/cm/memo.xs"> tells that this is document instance of memo. ***xmlns*** stands for XML Name Space; this is a file that contains the schema definition a; it is the space where all the tags used in the memo instance are defined. Digression: Elaboration of the name space idea. Everybody and their brother can define XML tags. Imagine the mess on the **World Wide** Web where billions of documents are posted containing tags defined by different people meaning different things but given the same name. So tag definitions are divided into *name spaces*; within a name space there can be only one tag associated with a name. The full name of a tag is *namespace:localname*; the full names are unique across the Web. For convenience, for a document instance, a name space can be declared at the beginning and all tag names are assumed to be from that namespace, unless explicitly given as full names. Note: The users producing document instances do not need to worry about tags. There are systems that display an input form, the user enters information in the appropriate places, and the system takes care of the XML, much like similar programs for defining Web pages.
More on the structure of XML components	Page ~214 shows document **b** as an arrangement of of nested boxes. This presentation makes it easier to follow the XSLT style sheet **c**. Note: XML schema definitions and XML style sheets are technically XML documents using pre-defined tags. However, to avoid confusion, in this lecture "document" refers to a file that has content to be seen by people or processed by a computer system.

c **XSLT style sheets made easy** p. ~215	Now comes the hard part. To display this document in a browser we need to **transform it into a HTML document**. This is done through an **XSLT style sheet**. Such style sheets can be exceedingly complex, but the style sheet used here is actually quite simple once you get the hang of it. An XSLT style sheet is again a special kind of XML document using tags pre-defined by the XML designers in the name space **xsl:**. A computer program designed to process XSLT style sheets interprets these tags and applies the appropriate transformations to the input XML document (in our case **b**). The operative part of the style sheet is between ***<xsl:template match="/">*** and ***</xsl:template>*** Any text that is not preceded by **<xsl:** > is transferred to the HTML document as is. This text, highlighted in **bold,** is a **formatting template** (or skeleton) f**or all document instances of type memo as transformed into HTML documents** (just like the form letter in the introductory example, p.~182). So now we have to worry about getting data into the template (putting meat on the skeleton). *Flashback:* In the form letter example, data come from a database; in the form letter each piece of data is identified by the field name used in the data base. *Here*: The data come from XML documents. The XSLT style sheet specifies the specific piece of data to be extracted from the input document **b** into the output document **d**. For example, what text should go after **From:** The following is the format of an extraction specification: ***<xsl:value-of select***=”...”/> In the specific case: ***<xsl:value-of select***=”memo/metadata/from”/> **memo/metadata/from** is called a **path,** starting from the root of the document going down the hierarchy of nested elements. To see where that path leads just look at ▸ in the nested boxes representation of Document **b** and verify that the correct text was inserted into the target Document **d** (the HTML document). For another example, what text goes between <TITLE> </TITLE>? This time the path is ***<xsl:value-of select***=”memo/metadata/subject”/> Again, it is not hard to figure out where that path leads and verify that the correct text was inserted into the target document **d** (the HTML document) Now that we have an HTML document, a browser can display it You can see that through an XSLT style sheet we can rearrange the text (and images, and sound) in an XML document any way we want to. Style sheets can be used to transform an XML documents into any file format imaginable. Data can be sorted and processed in many ways. As an example, consider a database of records on foods displayed with a table of content and an index produced based on definitions in an XSLT style sheet; s. Suppl., SLecture 6.1a.

The header of the figure should be

XML files and process



The example XML files used in the explanation of XML are on p. ~212 -215.

P. ~216 has another example, the XML schema for the self-assessment memo (see p. ~194 - 195)

****

Concluding notes on XML and related issues

Read after you have gone through the explanation and looked at all examples.

Important

Reuse is a big theme in the application of XML; reuse can be implemented in several ways.. There are whole collections of defined data types, such as a data type for US states, with the values restricted to a list of two-letter abbreviations of US states, or data types for US address, UK address, France address, Germany address, etc. (all derived from a generic address as a common parent). These type definitions are collected into *vocabularies*, each in its own name space, from which they can be included in any XML document schema, saving the schema creator a lot of work.

Optional

An XML schema is technically itself a document that follows XML syntax and tags defined in the W3C Recommendation XML Schema (2014-10-15), www.w3.org/2001/XMLSchema

Note: The syntax of the XML examples may not be entirely correct, but it gives the general idea.

There is supplemental material on XML and RDF in the supplement, Lecture 6.1a and 7.2b. In particular a fully worked out example of using XML and the sam example using RDF to represent the food data from Lecture 2.2 and using style sheets to create several different outputs (a table of contents, a detailed listing, and an alphabetical index) from this database in XML.

We noted in Lecture 5.1 that RDF could be expressed using XML, but that many people prefer the simpler Turtle notation. Likewise, for some applications, especially for structuring substantive data, many people prefer a simpler alternative to XML, JSON (Java Script Object Notation), see, for example

http://defiantjs.com/
http://en.wikipedia.org/wiki/JSON

Also, instead of using XSLT (which does not work with data in the JSON format), programmers often use JavaScript to format output documents.

a. Schema definition for document type *memo* (created once by systems analyst, defines tags)

An XML schema defines a document structure and identifies each element of the structure by a tag. This XML code creates a **memo schema or template.** The documents in the memo class must contain one top-level element, *memo*, which in turn consists of two subordinate elements, *metadata* and *memoBody* (exactly one of each in this order), which in turn contain subordinate elements.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="www.w3.org/2001/XMLSchema">
<!-- w3 schema file defines an XML name space; we use prefix xsd. ->
  <xsd:element name="memo" type="memoContent"/>
  <xsd:complexType name="memoContent">
      <xsd:sequence>
          <xsd:element name="metadata" type="metadataContent"/>
          <xsd:element name="memoBody" type="memoBodyContent"/>
      </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="metadataContent">
      <xsd:sequence>
          <xsd:element name="to" type="xsd:string"/>
          <xsd:element name="from" type="xsd:string"/>
          <xsd:element name="subject" type="xsd:string"/>
          <xsd:element name="date" type="xsd:date"/>
          <xsd:element name="keywords" type="xsd:string"/>
      </xsd:sequence>g
  </xsd:complexType>
  <xsd:complexType name="memoBodyContent">
      <xsd:sequence>
          <xsd:element name="plainText" type="xsd:string"/>
      </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```

b A document instance of type *memo* (done over and over by authors using tags defined in the memo template)

```
<?xml version="1.0"?>          [Note: See p. ~187 and p. ~204 where the same document is used as example]
<?xml:stylesheet type=”text/XSLT”
      xlink:href=”www.jasca.com/cm/memo.xslt”?>
<memo xmlns=”www.jasca.com/cm/memo.xs”>
      <metadata>
            <to>Sue Feldman, CIO</to>
            <from>Bob Boiko</from>
            <subject>What XML can do for us</subject>
            <date>February 7, 2003</date>
            <keywords>XML; content management; document structure; databases on the Web"</keywords>
      </metadata>
      <memoBody>
            <plainText>XML allows us to define document structures that will make it easier to create
            documents. Once a document is created, it can be displayed in many different ...</plainText>
      </memoBody>
</memo>
```

Explanation of the structure of a document using XML-defined tags

A document structured using XML-defined tags consists of a hierarchy of nested boxes called **elements**. Each box has

a **tag** that labels the box

a **type** that defines the content of the box

<***xsd:element name***="memo" ***type***="memoContent"/>
defines a box (element) with the name (tag) "memo"
The content of that box follows the pattern or structure of the type "memoContent"

memo (name or tag of the box)

memoContent (a type)
the type that gives the structure of the box content. Defined in

<***xsd:complexType name***="memoContent">

As the definition of the type shows, the **memo** box contains two boxes nested in it named with the tags

metadata (a tag)

metadataContent (a type)
the type that gives the structure of the content of this box (to, from, subject, date, keywords) as defined in
<***xsd:complexType name***="metadataContent">

memoBody (a tag)

memoBodyContent (a type)
defined in
<***xsd:complexType name***="memoBodyContent">

```
<xsd:complexType name="memoContent">
    <xsd:sequence>
        <xsd:element name="metadata" type="metadataContent"/>
        <xsd:element name="memoBody" type="memoBodyContent"/>
    </xsd:sequence>
</xsd:complexType>
```

The sample XML document shown as nested boxes ‣ A "path" down to a specific element

‣***memo***

‣***metadata***

to

Sue Feldman, CIO

‣***from***

Bob Boiko

subject

What XML can do for us

date

37658

keywords

XML; content management; document structure; databases on the Web

memoBody

plainText

XML allows us to define document structures that will make it easier to create documents. Once a document is created, it can be displayed in many different ...

‣**Path**: memo/metadata/from = Bob Boiko

The XSLT style sheet defines a different kind of template – the bolded text which goes in all HTML documents produced by this style sheet.

To fill each slot in the template, the XSLT stylesheet uses paths to pick out from the whole document the element that goes into the slot.

c XSLT style sheet (Controls the display of the document.) An XSLT processor program uses XML tags to identify pieces of data.. Determines selection of data to be displayed, their arrangement, and the appearance of each element. In the example, the output is an HTML document. But many other types of formatting are possible, e.g., to Wireless Markup Language (WML) for display on a handheld device
*** Here: **Bold** = text that goes as is from the XSLT style sheet into the output document, here HTML.

```
<xsl:stylesheet
  xmlns:xsl=”www.w3.org/TR/WD-XSLT”
  xmlns=”www.w3.org/TR/REC-html40”>
  <xsl:template match=”/”>
      <HTML>
          <HEAD>
          <TITLE><xsl:value-of select=”memo/metadata/subject”/></TITLE>
              <META NAME="creator" CONTENT="{<xsl:value-of select=”memo/metadata/from”/>}"/>
              <META NAME="keywords"
                  CONTENT="{<xsl:value-of select=”memo/metadata/keywords”/>}"/>
          </HEAD>
          <BODY>
              <H1><Center> Memorandum </Center></H1>
              To: <xsl:value-of select=”memo/metadata/to”/><BR/>
              From: <xsl:value-of select=”memo/metadata/from”/><BR>
              Date: <xsl:value-of select=”memo/metadata/date”/><BR><BR>
              Subject: <EM><xsl:value-of select=”memo/metadata/subject”/></EM>
              <P><xsl:value-of select=”memo/memoBody/plainText”/></P>
          </BODY>
      </HTML>
  </xsl:template>
</xsl:stylesheet>
```

d HTML document (same as p. ~212) (done over and over, produced by an XSLT processor program)

```
<HTML>
    <HEAD>
        <TITLE>What XML can do for us</TITLE>
        <META NAME="creator" CONTENT="Bob Boiko">
        <META NAME="keywords" CONTENT="XML; content management; document
        structure; databases on the Web">
    </HEAD>
    <BODY>
        <H1><Center> Memorandum </Center></H1>
        To: Sue Feldman, CIO <BR>
        From: Bob Boiko<BR>
        Date: February 7, 2003<BR><BR>
        Subject: <EM>What XML can do for us</EM>
        <P>XML allows us to define document structures that will make it easier to create
        documents. Once a document is created, it can be displayed in many different ways. . . .
        </P>
    </BODY>
</HTML>
```

e Document displayed by the Web browser under the control of HTML tags, see on p. ~212

XML schema for a self-assessment memo (see p. ~194 - 195) (DTD in Suppl., SLecture 6.1a)

Since a s*elf-assessment memo* is a specific type of memo, we can define its schema by adding to the *memo* schema; the *memo* schema is **reused**.

```
<?xml version="1.0"?>
<xsd:schema  xmlns:xsd="www.w3.org/2001/XMLSchema"
  xmlns="www.jasca.com/cm/memo.xsd">

  <include xsd:schemaLocation="www.jasca.com/cm/memo.xsd"/>
    <!--This schema includes a definition of the type metadataContent, which is used as is
    below and the type memoBodyContent which is modified by adding 3 elements. -->

  <xsd:element name="selfAssessmentMemo" type="selfAssessmentMemoContent"/>

  <xsd:complexType name="selfAssessmentMemoType">
    <xsd:sequence>
      <xsd:element name="metadata" type="metadataContent"/>
      <xsd:element name="memoBody" type="memoBodyContent"/>
    <xsd:sequence>
  </xsd:complexType>

  <!– redefinition of memoBodyContent –>
  <xsd:complexType name="memoBodyContent">
    <xsd:complexType>
      <xsd:extension base="memoBodyContent">
        <xsd:sequence>
          <xsd:element name="accomplishments" type="xsd:string"/>
          <xsd:element name="goals" type="xsd:string"/>
          <xsd:element name="trainingNeeds" type="xsd:string"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexType>
  </xsd:complexType>
</xsd:schema>
```

pink **Lecture 6.2b** ***March 2***

Document analysis for retrieval and information extraction (40 minutes)

Learning objectives (from Lectures 6.1a - 6.2b, p. ~156 - 157)	4 **Be able to create good design for new documents**. (P2.3.1,4)
	5 **Understand specific issues in document structure and be able to apply this understanding to document analysis and design, including text processing techniques.** (P2.3.1,5) 5.7 Understand the importance of noun phrases as carriers of meaning in English (P2.3.1,5.7) 5.8 Understand the characteristics of context that determine the meaning of an ambiguous word and how these characteristics can be used for automatic word sense disambiguation (WSD). (P2.3.1,5.8) 5.9 Understand text coherence and cohesion and their role in text understanding by people and by computer programs. (P2.3.1,5.9) 5.9.1 Understand how the idea of structuring information in frames or in a semantic network applies to discerning the structure of text. (P2.3.1,5.9.1) 5.9.2 Understand the problems of anaphoric reference and co-reference and their impact on information retrieval, esp. proximity searching, and information extraction. (P2.3.1,5.9.2) 5.10 Understand patterns used in text to represent relationships and how such patterns can be used for information extraction. (P2.3.1,5.10) 5.11 Understand information extraction into frames. (P2.3.1,5.11)
	6 **Have a general understanding of prevalent document analysis methods and their uses.** 6.1 **Be aware of the many useful applications of document analysis in dealing with vast quantities of tex**t, including automatic indexing for information retrieval, automatic abstracting and summarization, information extraction, and automated translation. 6.2 Have a general understanding of the potential of natural language processing for providing better, more pinpointed answers to users. 6.3 **Be able to recognize situations where information extraction and other methods of document analysis would be useful**, to suggest the use of appropriate software, and to participate in the selection of such software. (P2.3.3,1) 6.4 Have a general idea of how a computer program could do word sense disambiguation (WSD).
	[5.1 [For advanced study: Have some idea how automatic syntactic parsing (sentence diagraming) and semantic parsing work] (covered only in Supplement SL6.1b) (P2.3.3,1.1).
Practical significance	See **Overview of document analysis and it uses** • Understanding text coherence and text cohesion is important for evaluating texts and for good writing. • Knowing about anaphoric references points out limitations in using adjacency commands in free-text searching and the need for linguistic processing to overcome these limitations. ➜ UBLIS 518 Reference Sources and Services: Query formulation

Note **The importance of document analysis**	Users, especially in business, do not want to read large amounts of text. They demand specific data, just the information they need, preferably processed to show relationships and the large picture. Also, it is no longer feasible to manually index all documents, so automated techniques need to come to the rescue. Huge amounts of information are stored in written and audio text, where it is hard to access and process and correlate with other information. So transforming the information from text into a more structured form where specific data can be easily retrieved and where data can be processed and correlated is an extremely important function and one companies will pay for. (Again, text is not unstructured, just not structured enough for easy processing.) So document analysis, specifically Natural Language Processing (NLP), is booming both in commercial applications and in academic research, using sophisticated linguistic tools for intelligent indexing and data extraction, see list of some software on the next page. This course gives just a broad overview. If you want to learn more, perhaps take a course in linguistics, talk to me. Supplement Lecture 6.1b has some detailed examples of how automated syntactic and semantic parsing might work to demonstrate that such processing is feasible. An audio explanation is planned, ask me if you are interested.

****

Examples of Natural Language Processing (NLP) Software

www.copernic.com/en/products/**summarizer**/

Links to NLP tools, nicely classified: www-a2k.is.tokushima-u.ac.jp/member/kita/NLP/nlp_tools.html

There is much commercial information retrieval software and content management software that uses NLP techniques. A few examples

Temis - Luxid Semantic content enrichment goes mainstream
www.temis.com/index.php?id=208&selt=6

HP IDOL
HP's (former Autonomy's) next-generation information platform is a single processing layer that enables organizations to extract meaning and act on all forms of information, including audio, video, social media, email and web content, as well as structured data such as customer transaction logs and machine-based sensor data. The platform combines Autonomy's infrastructure software for automatically processing and understanding unstructured data with the high-performance, real-time analytics engine for structured data from Vertica, an HP Company.
www.autonomy.com/products/idol

ConceptSearching www.conceptsearching.com/wp/

Alchemy - Transforming text into knowledge www.alchemyapi.com/

1 Overview of document analysis

Text analysis. Natural Language Processing (NLP) (partially covered through examples)	**Text analysis, especially using linguistic/NLP and statistical / machine learning techniques, is incredibly useful for many functions:** 1 Preparing a description of the document, automatic cataloging / **automatic metadata generation** 1.1 Descriptive cataloging (e.g. from scanned title page) 1.2 Determining authorship or other characteristics of the origin of the document 1.3 Automatic or computer-assisted subject indexing /classification, multiple index terms or assigning a class (from Dewey or DMOZ or Chemical Abstract category), also called document categorization 1.4 Automatic abstracting, automatic summarization, including creating unified summaries from multiple documents (e.g., multiple news stories on the same topic or event) 2 **Determining the reading level** of a document (more generally: the audiences for which the document is appropriate) 3 **Sentiment analysis** Determining the attitudes, beliefs, or emotions underlying the document (content analysis in sociology and political science or in psychoanalytical methods). A hot topic today. 4 **Automatic** or computer-assisted **essay-grading** (controversial) 5 **Named entity recognition and information extraction** (related to 1) **5**.1 **Named entity recognition** (entity identification or extraction). Identify people, organizations, places, dates, chemical substances, diseases, ... Needed for 6.2 Information extraction. 5.2 **Information extraction** (fact extraction, relationship extraction). Automatic extraction of formatted data from text. Representing the relationships expressed in a document in a more explicit and more easily manipulated way. 6 **Question-answering**: Within a large document, find the specific sentence or paragraph that answers a question 7 **Assistance with document creation** **7**.1 **Automatic translation**, machine translation, for example on-the-fly translation of Web documents (e.g., www.google.com/language_tools) 7.2 Convert a document to hypertext, incorporate into a larger hypertext 7.3 Improve optical character recognition and speech recognition through post-processing the electronic text with NLP methods, e.g. spell check or determine which of two variant "readings" is more likely. 7.4 **Intelligent spell check** (for example, recognize *form here on* or *fro the benefit of* as typos. Grammar check. 7.5 Creating textual answers from the data returned by a database query 8 Assistance with query formulation, e.g. find noun phrase in string of words 9 Improve search, e.g. anaphora resolution for better recall in proximity search 10 Natural language interaction with software and systems

OCR (not covered)	**Optical character recognition** Converts text from graphical image (scanned) to character -encoded text that takes much less storage space and can be edited and processed through text analysis. For standard fonts in the Latin alphabet, much software is available for personal computers, accuracy > 99%. You probably have software on your computer. I use ABBYY FineReader http://www.abbyy.com Also standard software for major alphabets (Cyrillic, Arabic, Chinese, ...). Specialized software for other alphabets and old texts, for example http://gamera.informatik.hsnr.de/
Speech recognition (not covered)	Converts text from acoustic signal to character -encoded text that takes much less storage space and can be edited and processed through text analysis. Huge amounts of spoken text Commercial products: Dragon Naturally Speaking Speaker identification.

Image analysis
(not covered)
	Intellectual analysis:
	Reading Images: The Grammar of Visual Design. 2. ed.
	by Gunther Kress and Theo van Leeuwen
	
	Based on processing of pixels: **Content-Based Image Retrieval (CBIR)**
	http://en.wikipedia.org/wiki/Content-based_image_retrieval
	http://en.wikipedia.org/wiki/Image_analysis
	Many systems. Google image search uses text (caption) - based retrieval and CBIR
	Analyzes
	• colors
	• general shapes (triangle, square, circle, ...)
	• configuration of general shapes
	• specific shapes
	  • finger prints (searching a finger print database, finger print locks)
	  • faces: face recognition (also used for security)
	  • mood recognition
	  • features in medical images, indicators of disease
	  E.g. http://www.imageanalysis.org.uk/what-we-do/image-analysis
	Segmentation of movies or videos into scenes

Literary and artistic analysis	

2 Approaches to text analysis

Most of these techniques are used by human readers and machine systems alike for the purposes outlined under *Practical significance*.
Human readers may analyze a text for indexing, abstracting, extracting a specific fact or proposition, or for assimilating all the facts or propositions expressed in the text.

1 **Statistical** (used mostly by computer systems, but also implicitly by human readers

- 1.1 Word / phrase / concept frequency
- 1.2 Frequency of words that connote an attitudinal/emotional dimension (**content analysis** in psychology/sociology/political science).
- 1.3 Differential frequency.
- 1.4 Association of words with classes / document categories
- 1.5 Looking for the unexpected (such as weighting rare words highly in ranking retrieval results), as in AltaVista's ranking method, Lecture 5.2

2 **Based on text macrostructure - positional approach**

For example:

- 2.1 Introduction and conclusions useful source for abstract.
- 2.2 Section headings and figure captions useful source for index terms.
- 2.3 First and last paragraphs of sections, first and last sentences of paragraphs

3 **Cue words, cue phrases, and cue sentences**

- 3.1 For example, "method", "important result", "new"

4 **Morphologic, syntactic and semantic analysis**

- 4.1 Stemming, lemmatization (very important for search and before using statistics)
- 4.2 Noun phrase and verb phrase recognition
- 4.3 Parsing of sentences (sentence diagraming) or partial parsing to detect noun phrases
- 4.4 Parsing with semantic interpretation
- 4.5 Inter-sentence parsing, resolution of anaphoric references
- 4.6 Word sense disambiguation (WSD) ((Subject area of document or disambiguation rules based on semantic rules (such as laugh takes only animate subjects)

5 **Converting natural language statements into entity-relationship expressions**. Applying verb case frames. Using cue words to discover type of relationship between two entities, such as *because* or *therefore* indicating causation (Section 3.5.3 this lecture).

5 **Slot filling in frames using parsing or cues** used by human readers and by machines

Take this page out of your binder so you have the overview (on the back) as we work through the examples in Section 3.

Linguistic technique	**Notes on uses and importance** (read with the examples)
3.1 **Searching for a word or phrase and its synonyms** (query expansion)	No example here, but had examples in the Medline exercise and many examples later, such as in Textbook Chapter 12.
3.2 **Recognizing noun phrases**. **Importance for** **word sense disambiguation, retrieval,** **data extraction**, and other p. ~223 Supplement Lecture 6.1b deals with parsing to detect noun phrases	Noun phrases carry a lot of meaning in English. Take the component words (such as pool, blue, pressure, or program) alone out of context, and their meaning thins out into many possibilities. In formulating queries, think about using noun phrases (in " "). When a user puts in more than one word, Google looks for the occurrence of one or more noun phrases in what the user entered and boosts the rank of documents that contain the noun phrase rather than just the component words separately.
3.3 **Word sense disambiguation (WSD)** (disambiguation of homonyms), **semantic interpretation**. **Importance for retrieval, automated translation**, and other functions p. ~224 Not shown in this example: Identifying the part of speech (part-of-speech (POS) tagging, e.g., *monitor* as a noun or a verb) also helps w. WSD.	WSD is very hard; it requires much knowledge. People do it automatically, but computer programs have trouble. How do you know that in "white students" *white* is a race/ethnicity and in "white cars" it is a color. What knowledge do you bring to bear? Same question for the other examples. The last example, *white fountain*, is the toughest. It requires knowledge that in the segregationist South drinking fountains were separated by race, so that *white* refers to the people who were allowed to drink from that fountain, while in Italy *white* refers to the color of the stone the fountain is made of. Without word sense disambiguation, a free-text search for a word that has many meanings only one of which is of interest to the user has low precision.
3.4 Named-entity recognition (NER) (entity identification or extraction) p. ~226	Identify people, organizations, places, dates, chemical substances, diseases, **...** , surround them with XML tags and insert the preferred name or URI in the background. Semantic enrichment or semantic tagging.
3.5 Text cohesion and coherence. Resolving anaphoric references. **Expression of relationships in text** p. ~228	
3.5.1 Importance in free-text searching with proximity operators p. ~230	Searching for the occurrence of two words <u>in the same sentence</u> or paragraph is an important tool for high-precision searching. But without resolving anaphoric references recall suffers. In the *Blepharitis* example, the second sentence contains *eyelids* in spirit but not the string "eyelids". The search engine must recognize that *they* refers to eyelids to find the *Blepharitis* entry. You should be able to identify the issues in the second example. How does the human reader figure out that in the last paragraph *The disease* refers to osteoporosis? She applies this heuristic: In a handbook article, if the heading is a disease, then the reference *The disease* most likely refers to that disease, particularly if it appears at the beginning of a paragraph, so that there is no antecedent disease mentioned in the paragraph. The reader must know that osteoporosis is a disease.
3.5.2 E**xtracting substantive data from text (*information extraction* or *relationship extraction*)** p. ~231	The problems are very similar to 3.5.1; you should be able to figure it out using the **bolded** and *italicized* words as your guide. Makes use of devices used in text to express cohesion and coherence, see the Example 3.5.3
3.5.3 How relationships are expressed in text p. ~232	Find the many ways in which a text may express relationships, e.g., causation, to design a computer program that can extract causal relationships from text.
3.6 Extracting substantive data through **slot-filling in frames** p.~236	Read Box 1. The system designer constructed a disaster frame with rules a program can use to fill the slots. Applying these rules to the story on top of p. ~237 results in the frame instance for Event 345. Used to create a data-base of disasters gleaned from news. Some errors: The aftershocks did not kill 10 people, as the system asserts, the main earthquake did.

3 The need for text analysis through linguistic techniques.

Elements of text structure and how they relate to text analysis. Explored through examples

3.1 Query expansion (synonym expansion, hierarchic expansion) No example

3.2 Recognizing noun phrases. Importance for word sense disambiguation, retrieval, data extraction, and other functions

Approximation: search for contiguous strings of words, for example "information retrieval".

Example 3.2.1
information retrieval, retrieval of information, retrieval of legal information but: information on the retrieval of sunken treasures

Example 3.2.2. Noun phrases expressing a unit of meaning

hepatitis A
vitamin A
twelve-step program
route of administration (also known as administration route, medication route, route of drug entry, method of drug application)
gene pool
breath test
motivational interviewing
blue law
social control
boundary layer flow (aerodynamics)
data link layer (data communication)

peer pressure, pressure by peers social pressure vapor pressure benefits program safety program conference program computer program	Meaning of polysemous words determined by context in a phrase

Example 3.2.3. Importance of parsing complete sentences for noun phrase identification

1	The green vegetables supply calcium.	**NP** The green vegetables **V** supply
2	The green vegetables supply calcium to the body.	
3	The green vegetables supply digestible calcium.	
4	The green vegetables supply determines sufficiency of calcium.	**NP** The green vegetables supply

3.3 Word sense disambiguation (disambiguation of homonyms), semantic interpretation. Importance for retrieval, automated translation, and other functions

Example 3.3.1: Importance of semantic interpretation for disambiguating homonyms in searching (sense disambiguation, meaning disambiguation)

Query statement / information need: **Passages referring to white (race/ethnic group)**

Query formulation to search free-text: white

Passages retrieved:

White students were found to hold prejudices against their black and Hispanic peers.

White cars are preferred by middle-aged buyers.

The **white** dishwasher laughs

The **white** dishwasher is broken.

The black congresswoman won election in a majority **white** district.

Douglas **White** won the race.

A **white** knight came to the rescue of CSX Corporation in its take-over fight.

The family unit is the basis for American society. **White** units make up 53% of all family units in the state.

GE makes microwave ovens. Half the units sold are **white**.

The **white** drinking fountain

a. In a story set in the historic segregationist South
b. In a travel guide to Italy

A sophisticated free-text retrieval system would analyze the text to determine the meaning of **white** in each passage and tag the passage accordingly. It would ask the user what meaning of **white** she was after and find only properly tagged passages. Mistakes in the analysis may cause retrieval of erroneous passages and rejection of relevant passages.

This is also known as Word Sense Disambiguation (WSD)

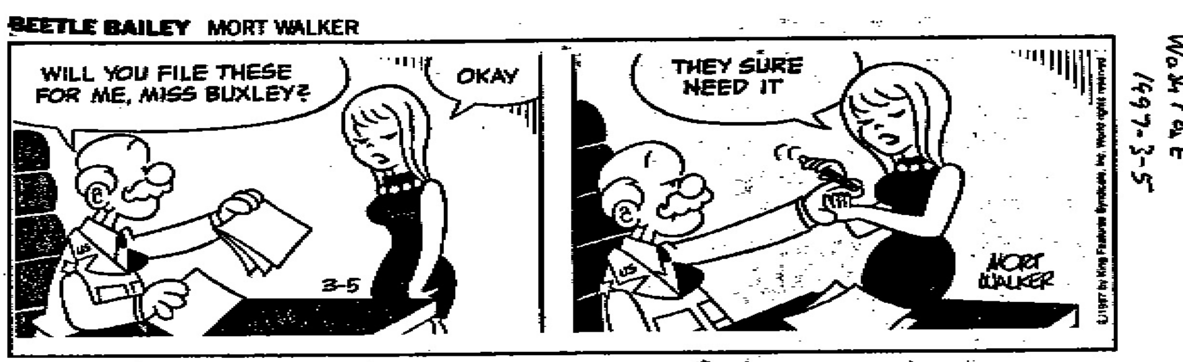
Example 3.3.2: Importance of semantic interpretation for automated translation

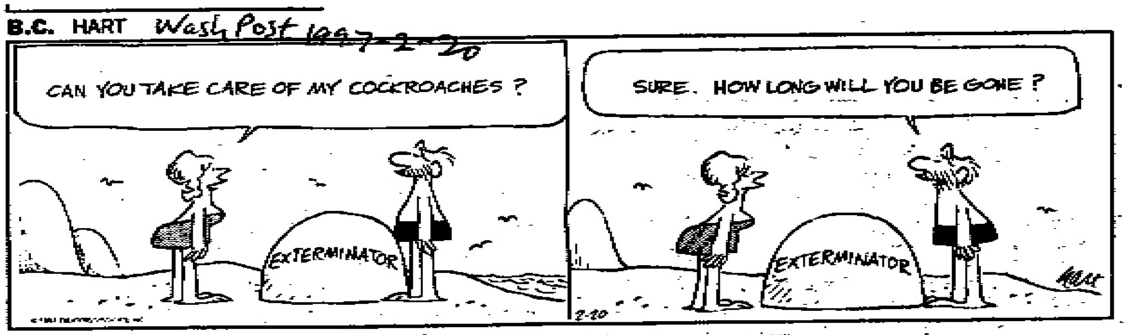
The white *dishwasher* laughs.	**The white *dishwasher* is broken.**
German: Der weisse *Tellerwäscher* lacht.	German: Die weisse *Spülmaschine* ist kaputt.
French: Le *plongeur* blanc rit.	French: Le *lave-vaisselle* blanc est détraqué.

Semantic interpretation often requires parsing (diagraming) of complete sentences (SLect 6.1b).

More examples and applications to retrieval in supplement, SLecture 6.1b

Example 3.3.3: Semantic interpretation rules illustrated in three comic strips.
The effect of jokes and comics is often based on word ambiguity (thus they are hard or impossible to translate). In jokes, the meaning of a word in a given context is often deliberately misconstrued. Thus, jokes can focus the attention of the language analyst on words with multiple meanings and on the semantic interpretation rules that ---distinguish between these meanings. With this in mind, analyze the comic strip episodes below. Enjoy the examples.





3.4 Named entity recognition

Example 3.4.1. Inxight ThingFinder (thing = named entity)

FAWIZ AL (RABBATI) PURCHASED TEN 1-TON TRUCKS (NFI) AND GETS SMUGGLERS TO CROSS THE BORDER APPROXIMATELY 10-15 KILOMETERS OUTSIDE OF KHASON. RABBATI RECRUITED JAN ANTON KRACZEWKI (AL-KIELBASA) TO WORK FOR HIM. KRACZEWKI IS APPROXIMATELY 53 YEARS OLD, AND 180 CENTIMETERS (CM) TALL. HE DRIVES A FOUR-DOOR 1984 GREEN SUBARU. KRACZEWKI USES HIS BACKGROUND AS AN ELECTRICIAN TO CREATE SOPHISTICATED BOMBS.

Person	FAWIZ AL (RABBATI), RABBATI, JAN ANTON KRACZEWKI (AL-KIELBASA), KRACZEWKI
Vehicle	TEN 1-TON TRUCKS, FOUR-DOOR 1984 GREEN SUBARU
Person_Common	SMUGGLERS, ELECTRICIAN
Measure	10-15 KILOMETERS, 150 CENTIMETERS (CM)
City	KHASON
Weapon	SOPHISTICATED BOMBS
Buy Artifact	FAWIZ AL (RABBATI) PURCHASED TEN 1-TON TRUCKS (NFI)
Travel across Border	SMUGGLERS TO CROSS THE BORDER APPROXIMATELY 10-15 KILOMETERS OUTSIDE OF KHASON
Recruit	RABBATI RECRUITED JAN ANTON KRACZEWKI ((AL-KIELBASA))
Person Appearance: Age	KRACZEWKI IS APPROXIMATELY 53 YEARS OLD
Person Appearance: Height	KRACZEWKI IS 180 CENTIMETERS (CM) TALL
Person Attributes: Vehicle	HE (KRACZEWKI) DRIVES A FOUR-DOOR 1984 GREEN SUBARU
Make Artifact	KRACZEWKI USES HIS BACKGROUND AS AN ELECTRICIAN TO CREATE SOPHISTICATED BOMBS

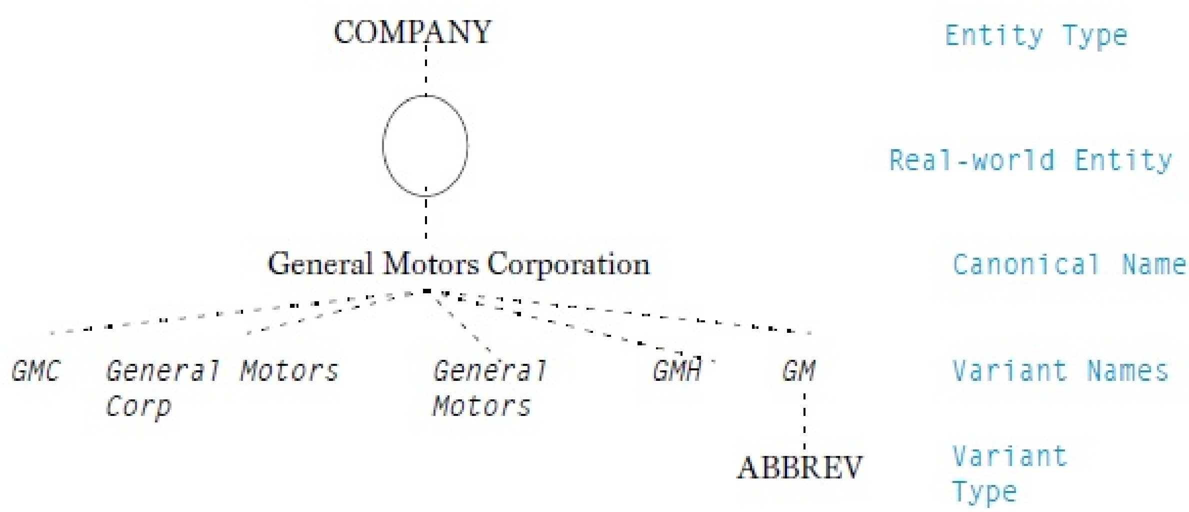
The proposed merger between Mega, Inc. and CNA Systems, Incorporated, has been postponed, Mega CEO Joe Smith said in an analyst call. "CNA's 1st quarter revenue dropped by 32%, and they lost 23 million dollars," Smith explained. CNA Systems sources blame weak sales in China. CNA shares (CNAI) fell 47 percent to $9.84 on May 12, the first trading day after the announcement.

Company	Mega, Inc., CNA Systems, Incorporated
Date	May 12
Person	Joe Smith
Person Position	CEO
Currency	23000000 USD and 9.84 USD
Measurement	32%, 47%
Country	China
Noun Group	proposed merger, analyst call, 1st quarter revenue weak sales, first trading day

Entity types in ThingFinder (selected)

- Person
- SSN
- Organization
- Address_Internet
- Phone
- Language
- Place
  - Country
  - State
  - City
- Date
  - Year
  - Month
- Holiday
- Product
  - Vehicle

Figure 2-3 :Name Catalog hierarchy



For more information, refer to the *ThingFinder Programmer's Guide and Reference*.

3.5 Text structure: cohesion and coherence. Resolving anaphoric references *Just read*

General introduction interwoven with application examples

Cohesion and coherence are the key devices in determining the internal structure of texts.

Cohesion (Grammatical)	Elements of a text are properly linked grammatically: Properly structured sentences Inter-sentence relationships
Anaphoric reference	Use of a pronoun or general noun to refer to an object, action, or thought previously identified in the text. Example: President Bill Clinton gave a speech at Concord High School. **He** emphasized the need for crime prevention and for the restoration of family values. **This** made the Republicans angry. **They** accused **him** of stealing **their** issues. Meanwhile, Buchanan addressed a rally in Manchester. **He** hammered away at the theme that the jobs of American workers must be protected from low-wage foreign competition. **This theme** has propelled him to the front in the polls. Importance in the context of information systems: - Detecting the relationships in a text. - Proximity searching.
Coherence (Lexical-semantic)	Does the document make sense? Does an argument proceed in a logical fashion? • If a section requires background knowledge the reader cannot be expected to possess, does the document provide this background knowledge before the reader gets to that section? • Are there proper transitions to prepare the reader's mind set for new information? • Do illustrations fit with the text? • In a conversation: Is a question properly answered? Does the contribution of one participant build on previous contributions? Importance in the context of information systems: Design hypermedia systems that support the user in constructing coherent documents Coherence related to document/text macrostructure

Incohesive text
President Bill Clinton gave a speech at Concord High School. **They** emphasized the need for crime prevention and for the restoration of family values. **This** made the Republicans angry. **She** accused **him** of stealing **her** issues. Meanwhile, Patrick Buchanan addressed a rally in LA. **She** hammered away at the theme that the jobs of American workers must be protected from low-wage foreign competition. **This scandal** has propelled him to the front in the polls.

Cohesive but incoherent text
President Bill Clinton gave a speech at Concord High School. **He** talked about playing the saxophone and mused about Plato. **This** made the Republicans angry. **They** climbed the Mount Everest. Meanwhile, Patrick Buchanan addressed a rally in LA. **He** ran down the street smashing cars. **This courageous action** has propelled him to the front in the polls.

Two principles from composition related to forms of knowledge representation:

Frame-style paragraph	Sentences in such a paragraph all have the same grammatical subject or main focus as established in the first sentence of the paragraph, the topic sentence. The paragraph presents a frame focused on one entity; each sentence is a frame slot giving information on that entity, allowing the user to maintain focus rather than jumping back and forth.
	Example: **Cattle** are domesticated ungulates, a member of the subfamily Bovinae of the family Bovidae. **They** are raised as livestock for meat (called beef and veal), dairy products (milk), leather and as draft animals (pulling carts, plows and the like). In some countries, such as India, **they** are subject to religious ceremonies and respect. **Cattle** are estimated to number 1.3 billion in the world.

Spreading activation paragraph	Alternatively, the sentences in a paragraph can be strung together so that the entity mentioned in the previous sentence is taken up at the beginning of the next sentence, like a path through a semantic network. The endpoint of one sentence gives the starting point for the next, keeping the user on the path.
	Example: **Cattle** are raised for beef and **milk**. Their **milk** is an important source of **calcium**. **Calcium** is important for growing strong bones and healthy teeth in children, preventing osteoporosis, and many other functions in the body, for example, **muscle contraction**, **which** is especially important for athletes.

3.5.1 Resolving *anaphoric references*. Needed for free-text searching with proximity operators

Proximity operators used here (syntax varies from system to system)

WS	two words occurring in the same sentence
WP	two words occurring in the same paragraph

Texts are from the Columbia University College of Physicians and Surgeons *Complete medical home guide*

Example 3.5.1,1. *They* stands for eyelid

Query statement / information need:	What to do about sticky eyelids
Query formulation to search free-text:	eyelid! WS stick!

BLEPHARITIS

Blepharitis is an infection of the edges of the **eyelids**. *They* become red, **sticky**, and crusty, and sometimes the victim has to unstick them to see anything in the morning.

The WS query misses this entry: *eyelid* and *sticky* do not occur in the same sentence.

Example 3.5.1,2.

Query formulation: calcium WS excret!	WS within same sentence (in para. 2)
Query formulation: osteoporosis WP vertebr!	WP within same paragraph (para. 4)

OSTEOPOROSIS

BONES NEED CALCIUM to maintain their strength, hardness, and to stay healthy. Milk, the main source of calcium in the diet, is important for the growing skeletons of children and adolescents as well as the bone-forming cells of adults. Regular daily consumption of at least 1 cup of skim or low-fat milk is essential for adults who want to keep their bones strong and to help prevent osteoporosis, a disease in which the body's bone mass decreases and bones become thin and brittle. Bones weakened by osteoporosis, a disease common to postmenopausal women, are prone to fracture if a person falls.

When **calcium** enters the body, *it* is absorbed into the bloodstream. If there is any excess, it is deposited in the end of the bone shafts where it is stored until the body needs to tap this reserve. (*Some* is also **excreted** via the kidneys.) When the calcium supply is deficient, the blood must take it back from the bones. If calcium intake remains inadequate over a long period of time, the bones eventually become porous and weak.

It is not known why calcium loss occurs. That postmenopausal women tend to get osteoporosis points in the direction of a hormonal disorder. Estrogen therapy is one treatment but its ability to decrease calcium loss may last only several years. More calcium and exercise are other therapies.

The disease most frequently affects the spinal column, causing backaches and rounded shoulders. in severe cases, the bone becomes as porous as a sponge and can collapse as a result. Collapsing **vertebrae**, which can cause sudden and sharp backaches, is one reason why elderly people tend to get shorter.

3.5.2 **Resolving anaphoric references**. Importance for **extracting data from text**

Example 3.5.2,1.
Consider the following text; find statements on *cranial arteritis* (only one example is treated in class; explore the other **bold** / *italic* pairs on your own):

VASCULITIS

VASCULITIS, as the name implies, is an inflammation of the blood vessels — both the arteries and the veins. Diseases in this category are relatively rare and comprise some of the most baffling and poorly understood disorders in medicine. Very often, the diagnosis remains unsuspected for long periods because of the variable way in which these disorders behave.

Inflammation of a blood vessel, particularly a small artery, can cause a narrowing of its lumen (internal diameter). If the vessel becomes completely closed, the tissue normally nourished by the diseased artery will die or be severely damaged.

Some forms of vasculitis are believed to result from an allergy or hypersensitivity, such as an adverse reaction to certain drugs. Sulfa drugs were very common causes of vasculitis, particularly in the early days of their use when the preparations were more crude and the dosages given were higher than today.

Patients with vasculitis, particularly when it involves widespread areas in the body, many be extremely ill with a generally poor prognosis. One particular type of vasculitis, which affects older people, involves inflammation of the cranial or temporal arteries, the vessels that serve a portion of the facial, jaw, and tongue muscles, the scalp, and most important, the retina. **Cranial arteritis** is the most common cause of sudden blindness in the elderly. Usually only one eye is involved but sometimes it occurs in both. *This condition* is successfully treated with corticosteroids, provided that treatment is started before there is significant loss of vision. *It* is often associated with a syndrome of severe muscle pain and stiffness called **polymyalgia rheumatica**. *This illness* is also largely confined to the elderly. It is almost always associated with a very high sedimentation rate, which measures the amount of inflammation, and it usually responds dramatically to cortisone-type drugs in low doses. Polymyalgia may occur without cranial arteritis, but because of the association, arteritis should be suspected in patients with polymyalgia.

Another form of vasculitis is called **Wegener's granulomatosis**. *This* is an extremely rare disorder which attacks the respiratory tract, the nasal sinuses, and the kidney in a progressively destructive process. Wegener's granulomatosis was once invariably fatal but now most patients can be treated successfully with cytotoxic or immunosuppressive drugs.

Patients with generalized or systemic vasculitis will often have paralysis of a foot or a wrist as a result of loss of blood supply to the peripheral nerve serving that limb. The blood vessels of the lung may also be affected, resulting in asthmalike symptoms. The development of asthma relatively late in life is very unusual, and may signify vasculitis.

There is another type of vasculitis known as **Takayasu's disease**, which occurs almost exclusively in young women. The inflammation is largely restricted to the branches of the great artery which leaves the heart (the aorta). *It* has also been called "pulseless" disease, for the diseased arteries may be so narrowed that a pulse cannot even be detected at the wrist. Patients with *this disease* will very frequently have symptoms of dizziness, light-headedness, weakness, and difficulty in using the arms, due to muscle pain from even slight physical effort. This is a direct result of lack of oxygen to the muscles, as the narrowed arteries are unable to deliver the increased amount of blood required during muscular effort. Corticosteroid therapy may be effective against Takayasu's disease, but the disease may go into remission without treatment.

These diseases are a few examples of the very broad spectrum of disorders included in the category of vasculitis. They are often difficult to diagnose, for their onset and evolution may be vague and ill-defined. The more classic types are easier to identify, but because of their relative rarity they are often not suspected until late in the course of the illness. Biopsy of an involved organ such as the kidney, muscle, or liver may be required in order to establish that a vasculitic process is indeed present

3.5.3 Introduction to the analysis of how relationships are expressed in text Application to information extraction

Information extraction is extremely important for systems to provide answers rather than pointers to where answers can be found. It helps people and organization cope with ever-increasing volumes of text. Google's Knowledge Graph is a huge database of statements, many extracted from text (using technology being further developed by Google). This database is used now to add substantive data to search results and will be used later to support question-answering.

We talk about how relationships are expressed in text to better understand systems that can extract assertions (facts) from text. Such systems exist now commercially (see later readings). They can be very useful for collecting a large number of assertions for the knowledge base of an expert system (such as the drug prescription expert system discussed in Week 1) and to answer questions from users by giving the actual answer (substantive data) rather than just pointing to one or more documents from which the user needs to laboriously extract the answer. Relationship extraction (also called information extraction) from text is one way to cope with information overload. You should consider acquiring such systems for your patrons. We will soon see such systems on the Web.

To extract relationships from text, a computer program needs to be able to figure out what an anaphoric reference refers to. This problem is illustrated in the cohesion part of both Crombie examples through the arrows that point from an anaphoric reference to the person, thing, or action referred to (the referent).

Then we need to analyze the many ways in which relationships can be expressed in text and generate patterns that we could give to a computer program. Look at Crombie example 2 and identify patterns (indicator words, grammatical patterns) that indicate a causative relationship so that a program could extract assertions of the form

A <*causes*> B

Example of an indicator word: Because

You may post your observations to the forum thread

First try to find examples of your own.

Then look at the examples on the page following the Crombie pages

XXX Include one paragraph without the lines so that students can draw their own. Include para with lines also.

Text analysis example 1 (from Crombie)

Cohesion

Apart from the disputes over the windmill, there was the question of the defence of (the farm). It was fully realized **that though** (the human beings) (had been defeated) in the Battle of the Cowshed **they** (might make another and more determined effort to recapture the farm and reinstate Mr Jones.) **They** had all the more **reason** for doing **so because** the news of **their defeat** had spread across the countryside and **made** the animals on the neighbouring farms **more restive than ever**. As usual, (Snowball) and (Napoleon) were **in disagreement**. According to **Napoleon**, what (the animals) must (do) was to **procure** (**fire-arms**) **and train themselves to use them**. **According to Snowball, they** must send out more and more pigeons and stir up rebellion among the animals on the other farms. **The one** argued that if (they) could not (**defend**) **themselves they** were bound to be conquered, **the other** argued that if rebellions happened everywhere (they) would have no need to **defend themselves**. (**The animals**) listened **first** to **Napoleon, then** to **Snowball**, and—could not make up their minds **which** was right; **indeed**, (they) always found **themselves** in agreement with **the one who** was speaking at the moment.

(George Orwell, *Animal Farm*.)

Coherence

Bonding (Rhetorical Coupling)	Apart from the disputes over the windmill, / there was the question of the defence of the farm.	
Concession–Contraexpectation	It was fully realized that / though the human beings had been defeated / in the Battle of the Cowshed / they might make another and more determined attempt / to recapture the farm / and reinstate Mr Jones. (Bonding (Coupling))	Amplification (Predicate Specification)
General Causative (Result–Reason)	They had all the more reason for doing so because / the news of their defeat had spread across / the countryside / and made the animals on the neighbouring farms / more restive than ever.	General Causative (Reason-Result)
Simple Contrast	As usual, Snowball and Napoleon were in disagreement. / According to Napoleon what the animals must do / was to procure fire-arms / and train themselves to use them. (Coupling) / According to Snowball, they must / send out more and more pigeons / and stir up rebellion among the animals / on the other farms. (Coupling)	Amplification (Predicate Specification)
Simple Contrast: Condition–Consequence	The one argued that / if they could not defend themselves / they were bound to be conquered,	Amplification (Predicate Specification)
Simple Contrast: Condition–Consequence	the other argued that / if rebellions happened everywhere / they would have no need to defend themselves.	Amplification (Predicate Specification)
Chronological Sequence	The animals listened first to Napoleon, / then to Snowball / and could not make up their minds	Bonding (Coupling)
Bonding (Rhetorical Coupling)	which was right; / indeed, they always found themselves / in agreement with the one who was speaking / at the moment.	

Cohesion

Text analysis example 2 (from Crombie)

Because the highest [electrical (conductivity)] can be obtained only with (pure) (metal), (copper) is extensively used (pure), about half the output of (copper) being used for electrical (requirements). For such purposes as long-span overhead cables high [electrical conductivity] is required, but the (metal) must also be (strong) to support its own considerable weight and to withstand additional stresses due to the effect of wind and accumulation of ice. (The strength) can be improved (by alloying), but any appreciable increase of (strength) thus obtained is usually offset by a serious lowering of (conductivity); for example, the addition of ten per cent aluminium more than doubles the (strength) of (copper) but reduces the (conductivity) to about one-sixth that of [(pure) (copper)]. There is, however, one element, (cadmium), which allows a useful compromise to be effected. When about 0.9 per cent of cadmium is alloyed with copper the conductivity is reduced to only about nine-tenths that of [pure copper] and at the same time the strength is much increased.

(W. Alexander and A. Street, *Metals in the Service of Man.*)

Coherence

Bonding (Statement–Exception)

Bonding Statement–Exemplification)

General Causative (Reason–Result)
- Because the highest electrical conductivity can be obtained only with pure metal,
- copper is extensively used pure, about half the output of copper being used for electrical requirements. — Bonding (Coupling)

Purpose–Means
- For such purposes as long-span overhead cables
- high electrical conductivity is required, but the metal must also be strong — Bonding (Rhetorical Coupling)

Means–Purpose
- to support its own considerable weight and to withstand additional stresses — Bonding (Coupling)
- due to the effect of wind and accumulation of ice. — Bonding (Coupling)
- General Causative (Result–Means)

Bonding (Rhetorical Coupling)
- The strength can be improved by alloying — General Causative: (Result–Means)
- but any appreciable increase of strength thus obtained is usually offset by a serious lowering of conductivity; — General Causative (Result–Means)
- Bonding (Contrastive Coupling)

Bonding (Contrastive Coupling)
- for example, the addition of ten per cent aluminium more than doubles the strength of copper but reduces the conductivity to about one sixth that of pure copper. — General Causative (Means–Result)

There is, however, one element, cadmium, which allows a useful compromise to be effected.

General Causative: Condition–Consequence
- When about 0.9 per cent of cadmium is alloyed with copper
- the conductivity is reduced to only about nine-tenths that of pure copper and at the same time the strength is much increased. — Temporal Overlap *And* Bonding (Contrastive Coupling)

Examples of patterns for extracting causative assertions

Text	Pattern	Extracted assertion
Because the highest electrical conductivity can be achieved only with pure metal, copper is extensively used pure	*Because* Cause, Effect could also be Effect *because* cause	{highest electrical conductivity can be achieved only with pure metal} <*causes*> {copper is used pure}
Stress *due to* {the effect of} wind	Effect *due to* Cause	Wind <*causes*> Stress
The strength can be improved *by* alloying	Effect *by* Cause	Alloying <*causes*> Material strength
The addition of ten percent aluminum more than *doubles* the strength of copper	Cause *doubles* Effect (Also *reduce*, *increase*, others)	Addition of ten percent aluminum <*causes*> Doubled strength of copper

Example of information extraction (entity-relationship statement extraction, relationship extraction) from the previous text. Done by hand to illustrate what we want a machine to do.
Legend: EntityValue anaphoric reference *relationship*

Patients with vasculitis, particularly when it involves widespread areas in the body, *may be* extremely ill *with* a generally poor *prognosis*. *One particular type of* vasculitis, which *affects* older people, involves inflammation of the cranial or temporal arteries, the vessels that serve a portion of the facial, jaw, and tongue muscles, the scalp, and most important, the retina. Cranial arteritis is the most common *cause of* sudden blindness in the elderly. Usually only one eye is involved but sometimes it occurs in both. This condition is successfully *treated with* corticosteroids, provided that treatment is *started before* there is significant loss of vision. It is often *associated with* a syndrome of severe muscle pain and stiffness *called* polymyalgia rheumatica. This illness is also largely *confined to* the elderly. It usually *responds* dramatically *to* cortisone-type drugs in low doses.	Vasculitis <*mayCause*> Extreme illness Vasculitis <*hasPrognosis*> Poor Cranial arteritis <*isa*> Vasculitis Cranial arteritis <*definedAs*> Vasculitis that involves inflammation of the cranial or temporal arteries, the vessels that serve a portion of the facial, jaw, and tongue muscles, the scalp, and most important, the retina Cranial arteritis <*occursIn*> Elderly Cranial arteritis <*mayCause*> (Sudden blindness, Elderly) (Sudden blindness, Elderly) <*causedBy*> (Cranial arteritis, High percentage of cases) [Blindness <*causedBy*> Cranial arteritis] <*preventedWith*> (Corticosteroids, Given early) Cranial arteritis <*associatedWith*> Polymyalgia rheumatica Polymyalgia rheumatica <*definedAs*> Syndrome of severe muscle pain and stiffness Polymyalgia rheumatica <*occursIn*> Elderly Polymyalgia rheumatica <*causedBy*> (Cranial arteritis, Medium percentage of cases) Polymyalgia rheumatica <*treatedWith*> (Cortisone-type drugs, *dosage*: Low, *response*: Very good)

Sample extraction patterns

Patients with Disease *may have* Condition ➜ Disease <*causes*> Condition

A type of A *is* B ➜ B <*isa*> A

Disease *responds to* ChemSubstance ➜Disease <*treatedWith*> ChemSubstance

3.6 Extracting data through slot-filling in frames: examples

Understanding and summarizing stories by machine

Based on distinguishing types of stories, such as *corporate merger*, *disaster*, *state visit*.
Each type of story has a list of items to be included in a summary; these are arranged in **a frame** specific for that type of story.

The summarizing process then proceeds in two steps:

1 Detect basic type of story, for example *story about disaster*, and pull up the proper frame

2 For filling each slot, fill the slots following the instructions given

Disaster frame – general pattern

Slot	**Instructions: What to look for to find slot fillers**
Type of disaster	indicator word such as *earthquake*, *aftershock*, *hurricane*
Where	place name (from a large dictionary of place names)
When	date line plus words such as *today, yesterday, Sunday*, *recent*
Number of dead	*killed* or *dead* or *fatality*, and a number close by
A*mount of damage*	(*$* or *dollar* or and number before or after) or *much* or *heavy*, esp. when close to *damage* or *worth* or *destroyed*

Disaster frame – Event 345

Slot	**Slot filler** (for story on facing page)
Type of disaster	earthquake aftershocks
Where	central Italy
When	October 6, 1997
Number of dead	10
Amount of damage	$1 billion

Disaster frame – Event 406

Slot	**Slot filler** (for story on facing page)
Type of disaster	hurricane
Where	Mexico's Pacific Coast, Acapulco
When	October 9, 1997
Number of dead	120
Amount of damage	untold millions of dollars

Frame for extracting data from pesticide reports see Supplement SLecture 6.1b

Aftershocks Jar Central Italy; Repair Cost Put at $1 Billion

Associated Press

ROME, **Oct 6**—The ground in **central Italy** rumbled again **today**, and officials said repairing buildings **damaged** by a series of earthquakes could cost more than **$1 billion.**

The aftershocks in the Umbria and Marches regions have prompted more people to seek temporary shelter, 11 days after a pair of quakes **killed 10** people. The National Geophysics Institute said today's tremors hit about every 30 minutes before dawn, the strongest with a magnitude of 3. No new destruction was report-ed.

The Sept. 26 quakes damaged the beloved Basilica of St. Francis in Assisi, along with thousands of other buildings.

The less severely damaged buildings will be repaired so that as many people as possible can return to their homes before winter, civil defense chief Franco Barberi said at a news conference. He said it will cost $875 million to $1.15 billion to repair damaged buildings.

The government will move about 3,000 units of prefab housing into the region in the next few weeks. Tents and camping vehicles already in place can shelter as many as 50,000 people.

Hurricane Devastates Mexico's Pacific Coast

Floods Kill at Least 120, Most in Acapulco

By Chris Kraul
and Mary Beth Sheridan

Los Angeles Times

ACAPULCO, Mexico, **Oct 9**-
Bearing 115 mph winds and torrential rain, Hurricane Pauline roared out of the Pacific through this coastal resort region before dawn **today**, leaving at least **120** people **dead**, thousands homeless and **untold millions of dol-lars** in **damage**.

Most of the dead were counted in and around Acapulco, a sunny port city usually filled with carefree Mexican and foreign tourists. The powerful storm left Acapulco, a city of about 1 million people, "unrecognizable," according to one report—a tangle of uprooted trees, downed power lines, overturned cars and bodies.

Morning light revealed corpses and garbage and the wreckage of countless wood-frame homes floating in oily, four-foot-deep floodwa-ters that coursed through the streets and washed over La Costera Miguel Aleman, a fabled promenade skirting Acapulco's ocean-front. City officials said there had been some isolated instances of looting, and army units were called out to patrol the streets. A deluge of rain—20 inches in less than 24 hours—sent floodwater, mud, gravel and boulders rushing down drought-parched hills surrounding Acapulco through several slum neighborhoods, smashing poorly constructed shanties and more substantial houses to flinders and washing away anything not firmly anchored. At least seven mudslides reportedly caused heavy property damage around the city, and local officials fear thick layers of mud coating many neighborhoods may conceal dozens of bodies as well.

While the official death stood at 120 late last night local authorities said it would certainly climb—and perhaps double—as search parties -comb through the debris left by the storm. The U.S. Embassy in Mexico City said that no Americans were reported among the dead or missing.

The Red Cross issued a plea for

See HURRICANE, A29, Col. 4

From the Washington Post, October 7, 1997 and October 10, 1997, respectively

red

Things to do in Week 7, W March 2 - March 9

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	None	
Readings after Lectures 6.1 - 6.2b	1 Kamps, T.; Hüser, C.; Möhr, W.; Schmidt, I. **Knowledge-based information access for hypermedia reference works: Exploring the spread of the Bauhaus movement**. In Agosti, M.; Smeaton, A. *Information retrieval and hypertext.* Boston: Kluwer; 1996. Ch. 10, p. 225-256. Nice and easy to understand example that uses all the technologies discussed here. Presages Google Knowledge Graph.	☐
	2 Elhadad, N.; Kan, M.; Klavans, J.; McKeown, K. **Customization in a unified framework for summarizing medical literature**. Journal of Artificial Intelligence in Medicine, 33(2):179-198, 2005 www.cs.columbia.edu/nlp/papers/2005/elhadad_al_05a.pdf. Find documents based on a patient record, then extract relevant pieces and present in a multi-document summary..	○
	Assignments due W March 9 ☐ required ○ optional	√
Assignment(s)	Ass.7: Apply linguistic techniques to retrieval problems (2 h)(assigned L6.2b, `Mar. 4)	☐

	New topics this week

	7.1a General introduction to metadata (10 min)	
Readings	1 Lectures 7.1a-7.2 Objectives etc. (pink) (pink sheets for Lectures 7.1a-7.2b are not for advanced reading). Read Knowledge Updates on p.~243. Also have a look at Assignments 8-10.	☐
Lecture	**Lecture 7.1a slides (10 min)**	☐

	7.1b Bibliographic and record control. General issues (40m))	
Readings	1 Tillett**,** Barbara **What is FRBR?: A Conceptual Model for the Bibliographic Universe.** www.loc.gov/cds/FRBR.html. In packet, **Required**	☐
	2 **Functional Requirements for Bibliographic Records: Final Report**. 1998 www.ifla.org/VII/s13/frbr/frbr.pdf. **Optional**	○
	3 **Statement of international cataloguing principles**. IFLA 2009. Retrieved on Sept. 18, 2010 from www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf	☐
Lecture	**Lecture 7.1b slides (40 min)**	☐

over

	7.1c Bibliographic and record control. Description (20 min)	
Readings	1 Oliver. Introducing RDA	☐
	2 Descriptive cataloging codes	☐
Lecture	**Lecture 7.1c slides (15 min)**	☐

	7.2a Bibliographic and record control: Entries and access (40 min)	
Readings	1 **Lubetzky's conditions for author entry** (from Needham, Organizing knowledge in libraries, 1971), rearranged by DS	☐
	2 **Excerpts from AACR2,** ***Part 2*** **and RDA side-by-side** (Familiarize yourself with the general layout and some major rules)	☐
Lecture	**Lecture 7.2a slides (80 min)**	☐

	7.2b Dublin Core (DC) (20 min**)**	
Readings	*1 Lecture 7.2b Objectives etc. (pink)*	☐
Lecture	**Lecture 7.2b slides (10 min)**	☐

Learning blog	Learning blog Week 7 due W March 9	○

	Assignments assigned W March 9	
Assignments assigned	▸**Assignment 8, Descriptive cataloging practice** (3.5 hrs) (complete when you can, no later than `May 18) ▸**Assignment 9, Problems of entry** (4 hours) (complete when you can, no later than `May 18) ▸**Assignment 10, Index three documents,** **prepares you for Lecture 8.1** (2 hours) (due `Mar. 23)	
MIDTERM	Closed-book midterm exam distributed (1.5 hours) (due `March 23)	

pink **Lectures 7.1a- 7.2b.** ***March 9***

Cataloging and metadata. Bibliographic and record control

(no Textbook chapter)

Learning objectives	**1 Understand the nature and importance of metadata and metadata creation (cataloging)**. (P2.3.5,1)	L7.1a
	1.1 Understand the use of metadata for finding, interpreting, and using any kind of data source (in some interpretations: any kind of object).(P2.3.5,1.1)	L7.1a
	1.2 Be aware of the range of metadata schemas for various types of entities, with the MARC format and RDA being one example, and be able to locate and use an appropriate metadata schema for cataloging outside the scope of MARC and RDA. (P2.3.5,1.2)	L7.1a L7.2b
	1.2.1 Be familiar with the Dublin Core and be aware that there are other bibliographic metadata standards used on the Web.	L7.2b
	1.3 Know to reuse existing cataloging data / metadata whenever possible, know when to amend such metadata to meet local user needs, and know how to locate and use resources for sharing metadata. (P2.3.5,1.3)	L7.1a
	2 **Understand that solving the fundamental problems of descriptive cataloging of documents and other entities (bibliographic control) is an application of general principles of Information Organization**. (P2.3.6)	L7.1b
	3 Understand fundamental issues in bibliographic control (P2.3.6,1)	L7.1b
	3.1 Understand the problems of defining "document" and of defining the relationships between several versions of a document. Understand the complexity of defining bibliographic units.(P2.3.6,1.1)	L7.1b
	3.2 Be able to apply this understanding to the analysis and design of cataloging codes. (P2.3.6,1.2)	L7.1b
	3.3 Understand the complexities in determining the useful entries (access points) for a document. Be able to recognize cases where making the proper entry is tricky.(P2.3.6,1.3)	L7.2a
	3.4 Understand that there are multiple solutions to problems of bibliographic description and be aware of the variety of codes for bibliographic description (such as the Dublin Core). (P2.3.6,1.4)	L7.1b - L7.2b
	4 Be able to catalog (create MARC records) consulting the record format (MARC, but also others) and the cataloging rules (AACR2 or RDA, others). Be able to quickly improve cataloging skills on the job. (P2.3.6,2)	L7.1c L7.2a L7.2b
	4.1 Know sources about record formats and cataloging rules and know one's way around these sources to find field formats and rules.	
	5 Be able to exploit knowledge of catalog structure for searching, especially the record format (such as MARC) and the cataloging rules (such as AACR2 or RDA) used. (P2.3.6,3)	L7.1c L7.2a

Practical significance see back. **Pink sheets for Lect. 7.1a-7.2b are not for advance reading**
In this section, Learning objectives are not repeated for each individual lecture.

pink

Practical significance	• Good catalogs are more important than ever, just look at the World Wide Web. It is now fashionable to call data about documents (i.e., data about data-carrying objects) *metadata*. • The question "what is a document" is important for library catalogs but even more important for electronic records, where different versions of the same document proliferate rapidly, especially on the Web. A Uniform Resource Locator or URL does not identify a document but a file storing a document. Many files in different locations can store the same document, creating a burden on the user. There are efforts to define a **Uniform Resource Identifier** or **URI** that would identify a **document** no matter where it is stored (like an ISBN). A URI identifies a *document, intellectual work,* a URL identifies a *document, physical volume* (as defined in Chapter 3). However, a system of URIs, while beneficial to users, introduces many difficulties: Who would assign URIs, using what rules? Who would maintain the database(s) with links from each URI to all its physical volumes (URLs)? What happens if one of these files is slightly modified? • Controlling different versions of a document is important for managing document production and access afterwards and for preserving legal and historical evidence. • Good catalogs that are widely usable and that can share data require standardized cataloging rules. In the World Wide Web domain, there are efforts to agree on a metadata standard. (A minimal standard, the *Dublin Core* has already found wide acceptance.) In the domain of geographic data there is the *Content Standard for Digital Geospatial Metadata* issued by the Federal Geographic Data Committee. 1 In some (not all) jobs you need to be able to catalog or learn it quickly. 2 Knowledge of descriptive cataloging is useful in searching library catalogs.

Overview of Lectures 7.1a - 7.2b

7.1a General introduction to metadata and data standards

7.1b **Basic concepts in bibliographic control**.

7.1c&7.2a The **conceptual data schema of a library catalog**:

7.1c **Description** (what data describing a document to capture and how for format them) (AACR2, Part 1) and

7.2a **Access** (what entries to make under which a document can be retrieved and what form of names and other entries to use) (AACR2, Part 2)

7.2b **Bibliogr. data on the Web: BibFrame, Dublin Core, Open Archives Initiative, ...**

Knowledge updates

Background knowledge needed to understand cataloging examples

Laurence Kerr Olivier, Baron Olivier, OM [Order of Merit] (22 May 1907 – 11 July 1989) was an English actor, director, and producer. One of the most famous and revered actors of the 20th century,[1] he was the youngest actor to be knighted and the first to be elevated to the peerage.[2] Actor Spencer Tracy said that Olivier was 'the greatest actor in the English-speaking world'.[3][http://en.wikipedia.org/wiki/Laurence_Olivier

Goethe's Faust

Note: This is an example of "Know the reader". To help you appreciate the complexity of the relationship between bibliographic units, I need a rich example. Goethe's Faust, one in a family of many works based on a medieval German legend, is such an example, but I was told that even among university students few Americans know about this work (or Goethe, for that matter) and its context. So now that I know my readers I know that I need to provide some cultural background.

Johann Wolfgang von Goethe 1749 – 1832) was a German writer, pictorial artist, biologist, theoretical physicist, and polymath.[2] He is considered the supreme genius of modern German literature.[3] His works span the fields of poetry, drama, prose, philosophy, and science. **His Faust has been called one of the greatest dramatic works of modern European literature.[3**] His other well-known literary works include his numerous poems, the Bildungsroman Wilhelm Meister's Apprenticeship, and the epistolary novel The Sorrows of Young Werther [recently made into the movie *Young Goethe in Love*, www.filmjournal.com/filmjournal/content_display/reviews/specialty-releases/ e3i6bd64ad3379a97770a1ee84a33191727]. http://en.wikipedia.org/wiki/Goethe

Goethe's Faust is one of the great works of world literature, his most famous work and considered by many to be one of the greatest works of German literature" http://en.wikipedia.org/wiki/Goethe's_Faust

The work is one of many based on a classic German legend:
Faust or **Faustus** (Latin for "auspicious" or "lucky") is the protagonist of a classic German legend. Though a highly successful scholar, he is dissatisfied, and makes a deal with the devil, exchanging his soul for unlimited knowledge and worldly pleasures. Faust's tale is the basis for many literary, artistic, cinematic, and musical works. The meaning of the word and name has been reinterpreted through the ages. *Faust*, and the adjective ***Faustian***, are often used to describe an arrangement in which an ambitious person surrenders moral integrity in order to achieve power and success: the proverbial "deal with the devil". The terms can also refer to an unquenchable thirst for knowledge.[1]
Plays and comic puppet theatre loosely based on this legend were popular throughout Germany in the 16th century. http://en.wikipedia.org/wiki/Faust

Also, many, many works are based on this legend; to give just two examples, Marlowe's Doctor Faustus and Gounaud's opera Faust (http://en.wikipedia.org/wiki/Faust_(opera)). Please look at the long list in http://en.wikipedia.org/wiki/Works_based_on_Faust. Goethe's Faust was translated into many languages (it is read in high school in China) and is in turn the basis for derivative works, including Randy Newman's musical *Faust* (1993) and Kamelot's albums *Epica* (2003) and *The Black Halo* (2005). A bit further removed is Thomas Mann's Doctor Faustus, "a re-shaping of the Faust legend set in the context of the first half of the 20th century and the turmoil of Germany in that period." (http://en.wikipedia.org/wiki/Doctor_Faustus_(Thomas_Mann_novel)).

If you think **cataloging all these works could be more efficient using hierarchical inheritance**, you are exactly right.

pink **Lecture 7.1a** (10 min) ***March 9***

General introduction to metadata and data standards

For online: The audio elaborates on this section

Metadata
Synonymous with cataloging data or pointer data as defined in Organizing Information, Chapter 2.
= Data used to **describe other data** and **give context for other data** for:

- retrieval
- assessment
- interpretation and use.

Metadata and data standards a hot topic in the context of Linked Data and the Semantic Web.

Important: There is no intrinsic difference between substantive data and metadata. If data are used for the purpose of retrieving, assessing, interpreting, or using other data, they are used as metadata. Often substantive data that are clearly formatted are erroneously called metadata (e.g data about products on ecommerce websites that look just like cataloging data for a book). The data modeling mechanisms are the same no matter what the data modeled are used for.

Efficiency and cost-effectiveness mandate data sharing, and data sharing requires standards, as long recognized in the library world (hence the *Anglo-American Cataloging Rules (AACR*) (since 1908), now *Resource Description and Access (RDA)*, the Library of Congress catalog Card Distribution Service (since 1902), and the MARC format (since the 1960s). Computers and the Web have in-creased the opportunities for data sharing and the need for data standards. In the standard world, data and metadata are treated together. For example, both the *IMS Global Learning Consortium* and the *IEEE Learning Technology Standards Committee (LTSC)* develop standards for cataloging learning objects (such as lesson plans) and for structuring learning objects (mostly by tags defined using XML), so the whole learning object or a part of it (such as a specific activity in a lesson plan) can be reused. Schemas focusing on metadata are maintained for many types of "data containers" by many user communities. The W3C is a hub of standardization activity

A standard defines a conceptual data schema for a given domain, usually in the form of either XML tags or of RDF classes (entity types) and properties (relationship types). Local applications often mix and match elements from several standards into a local schema called an *application profile*. Remember name spaces from Lecture 5.1.

Important: Principle of reuse.

Always **reuse existing cataloging data / metadata when available, possibly amended to meet local user needs**, rather than cataloging / creating metadata from scratch. Find out about resources for sharing metadata and use them. For example: perform copy cataloging on existing records from OCLC vs. creating your own. You may need to enlighten management about this. For example, doing original cataloging in a small school or public library is so 19th century.

Review Lecture 4.2. **Conceptual data schemas and input, storage, and presentation formats**
Then look over the **examples of data schemas** / data standards (many expressed in XML) on the next page to get an idea of a range of data standards **with focus on metadata**.
Pick one in accordance with your interests to examine in a bit more depth – follow links (use Lecture 7.1a Reading 1 to just click on the URLs).

pink

General overview. www.schema.org , http://www.w3.org/wiki/WebSchemas/SchemaDotOrgProposals http://www.w3.org/wiki/WebSchemas

[Bibliographic data: MARC / RDA. Dublin Core. Bibliographic Ontology (BIBO)] Done later, not now

Text: **The Text Encoding Initiative (TEI)**
A comprehensive standard for describing (literary) texts, both metadata and the structure of the actual text, http://www.tei-c.org/ *For intro:*
http://teibyexample.org/examples/TBED00v00.htm, read first screen, then click 6. TEI P5 (XML)
http://teibyexample.org/modules/TBED01v00.htm (perhaps not the whole thing)
Bibliography: http://www.tei-c.org/Support/Learn/tei_bibliography.xml
Projects using TEI: http://www.tei-c.org/Activities/Projects/ (both give a flavor of TEI use)

Archival data: **Encoded Archival Description (EAD)**
http://www.dlib.org/dlib/november99/11pitti.html (still a good introduction)
http://en.wikipedia.org/wiki/Encoded_Archival_Description
http://www.loc.gov/ead/ Example repository using EAD: http://www.cdlib.org/inside/projects/oac/
EAD Updates overview:
http://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/ead-revision
Several projects: http://dcevents.dublincore.org/IntConf/index/pages/view/2014-archives

Museum data: The **CIDOC Conceptual Reference Model (CRM)**, related to FRBR
Introduction: http://cidoc.ics.forth.gr/docs/cidoc_crm_meeting_Prato-1.ppt
http://www.dsoergel.com/UBLIS571DS-07.1a-1Reading2Solanki-semantic-web-in-cultural-heritage-and-archaeology.pdf
Full specification: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf

Learning objects (instructional materials): IEEE Standard for Learning Object Metadata, IMS Global
Introduction: **A Comparative Study of Learning Object Metadata**
http://www.tmrfindia.org/eseries/ebookV2-C6.pdf Follow some links from there.

Geospatial data: will be covered later. Ask me if you are interested

Not for the faint of heart

TV programs: TV Anytime Forum. TV Anytime is a set of specifications for the controlled delivery of multimedia content to a user's personal device (Personal Video Recorder (PVR))
www.etsi.org/deliver/etsi_ts/102800_102899/1028220301/01.07.01_60/ts_1028220301v010701p.pdf
BBC:www.bbc.co.uk/rd/publications/whitepaper192
European Broadcasting Union (EBU): Overview of metadata in broadcasting
https://tech.ebu.ch/docs/events/metadata14/MDN2014_programme_web.pdf
Over/pmeta/WIP/ESCORT/ESCORT2006.htm (A faceted classification)

Multimedia: **MPEG-7**,
"Multimedia Content Description Interface", is a standard for describing features of multimedia content: catalog data (e.g., title, creator, rights), semantic data (e.g., the who, what, when, where information about objects and events) and structural data (e.g., the color histogram - measurement of the amount of color associated with an image or the timbre of a recorded instrument). Builds on AV data representation defined by MPEG-1, 2 and 4. www.mpeg.org/MPEG/starting-points.html
Update: http://mpeg.chiariglione.org/ (Standards tab)
Part of this family: MPEG-1 or MPEG-2 Audio Layer III,[4] more commonly referred to as MP3

POWDER (Platform for Web Description Resources) http://www.w3.org/TR/powder-primer/
Paper introducing: users.iit.demokritos.gr/~konstant/dload/Pubs/ijmso.pdf

pink **Lecture 7.1b** (40 min) ***March 9***

Bibliographic control. General issues

Introduction	This lecture deals with control of all kinds of documents (all kinds of materials): Regular books and reports, serials, journal and newspaper articles, organizational records, images, sound documents. New dimension of problem: Electronic documents. Ease of copying and modification, cryptic filenames, and online accessibility of electronic documents create special difficulties. A number of general principles of Organization of Information are applicable to the control of any kind of concrete object or "thought object." Each type of material presents its own challenging problems in applying these general principles. Parts of the thinking on descriptive cataloging and the resulting practices are still valid. Other parts have been made obsolete through the greater power of automated systems. Control is mainly access, but also inventory control, including preservation. A distinction is generally made between description and access, but the two are more closely intertwined than many people realize.
What is a catalog?	A catalog is a database that contains identifying/descriptive data about objects, such as books (or, more broadly, documents) or data sets (such as geospatial data sets) or merchandise. The coverage of a catalog may be limited to a given physical collection (the books for which physical copies are held in the library, the merchandise items available from a catalog store); that is, the catalog contains only data referring to objects in the given collection. Often the term *catalog* is defined in this sense of being tied to a physical collection as distinguished from a *bibliography*, which may include data about documents no matter where physical copies are held. A *union catalog* refers to objects in multiple collections. If the objects referred to in the catalog are information sources, the catalog data are used primarily as *pointer data* (see Section 2.5 of Textbook,), now commonly called *metadata*. However, remember, that the distinction between substantive data (data contributing directly to the problem solution) and pointer data lies not in the nature of the data themselves but rather in their use. If data are used to find other data/information sources, they are used as pointer data. If data are directly applied to problem solution, they are used as substantive data.

pink

Objectives of the library catalog according to Cutter

1. To enable a person to find a book for which either
   A. the author
   B. the title
   C. the subject
   is known
2. To see what a library has
   D. by a given author
   E. on a given subject
   F. in a given kind of literature
3. To assist in the choice of a book
   G. as to its edition (bibliographically)
   H. as to its character (literary or topical)

Problems with Cutter's objectives

(a) Is the user interested in a particular *book* or in the *work* that is embodied in the book? And what is a *book* anyhow?

(b) In today's world of electronic access, what is a library?

Need to address these before restating objectives.

Fundamental problem in bibliographic control: What are the units we are dealing with?

Look at the examples on the following pages to get a feel for the problem.

****

Sample documents illustrating problems in defining bibliographic units and in cataloging. Facing page, some questions to think about here.
Share observations on the examples on the discussion board.

Examine entries (1) – (10) and think how they are related to each other.

How many works?

Who is the author of (4). When cataloging (4), should Halliday be mentioned? If so, how?

How about (8)? What is the status of the marginal notes?

Sample documents illustrating problems in defining bibliographic units

(1) *The man I killed*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1935.

(2) *The man I killed*, by Michael Halliday (i.e. John Creasey). Large print edition. London: Society for Assistance to the Blind; 1938.

(3) *The man I killed*, by Michael Halliday (i.e. John Creasey). Audiotape, read by Sir Lawrence Olivier. New York: Books on Tape; 1966.

(4) *The man I killed*, play by Christopher Wern, based on the novel by Michael Halliday.

(5) *The man I killed*, a movie version of the play by Christopher Wern, based on the novel by Michael Halliday. On videotape.

(6) An individual copy of (1) as originally printed.

(7) An individual copy of (1), produced by making a copy of (6).

(8) An individual copy of (1), owned by Sir Lawrence Olivier, with many marginal notes in ink.

(9) A facsimile edition of (8), published by Marx Brothers.

(10) *The man I killed, completely revised and with a new ending* , by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1941.

(11) A legal document with original signatures

(12) A copy of the same

(13) A notarized copy of the same

See facing page. Are these all different works? Or is there an identity preserved through all the changes? If so, what does this identity pertain to? A work? ⇨

****

Continuation of lecture notes text

In all these examples, we are confronted with two fundamental questions:

1 What are the units, the types of entities we must deal with in bibliographic control

2 How are bibliographic entities related.

Pages 228 - 238 deal with these questions

See Knowledge update: Goethe's Faust, p.~243

Edward T. O'Neill and Diane Vizine-Goetz

Table II. Publishing History for *Guide to Reference* (adapted and updated by DS)

Edition	Date	Authors	Publisher	Title
	1902	Kroeger	Houghton, Mifflin & Company	Guide to the study and use of reference books; a manual for librarians, teachers, and students
Title edition	1904	Kroeger	American Library Association Publishing Board	Guide to the study and use of reference books; a manual for librarians, teachers, and students
2d ed., rev. and enl.	1908	Kroeger, Mudge	American Library Association	Guide to the study and use of' reference books
3d ed., rev. throughout and much enlarged	1917	Kroeger, Mudge	"	Guide to the study and use of reference books
[4th ed.]	1923	Mudge	"	New guide to reference books
5th ed.	1929	Mudge, Kroeger	"	Guide to reference books
6th ed.	1936	Mudge, Winchell	"	"
7th ed.	1951	Winchell, Mudge	"	"
8th ed.	1967	Winchell, Mudge, Sheehy	"	"
9th ed.	1976	Sheehy, Keckeissen, Mcllvaine, Winchell	"	"
10th ed.	1986	Sheehy (ed.	"	"
11th ed.	1996	Balay (ed.), Carrington, Martin	"	"
12th ed.	2008 -	Kieft	"	Guide to Reference [Online; includes Web sources] www.guidetoreference.org/ || http://en. wikipedia.org/wiki/ Guide_to_Reference

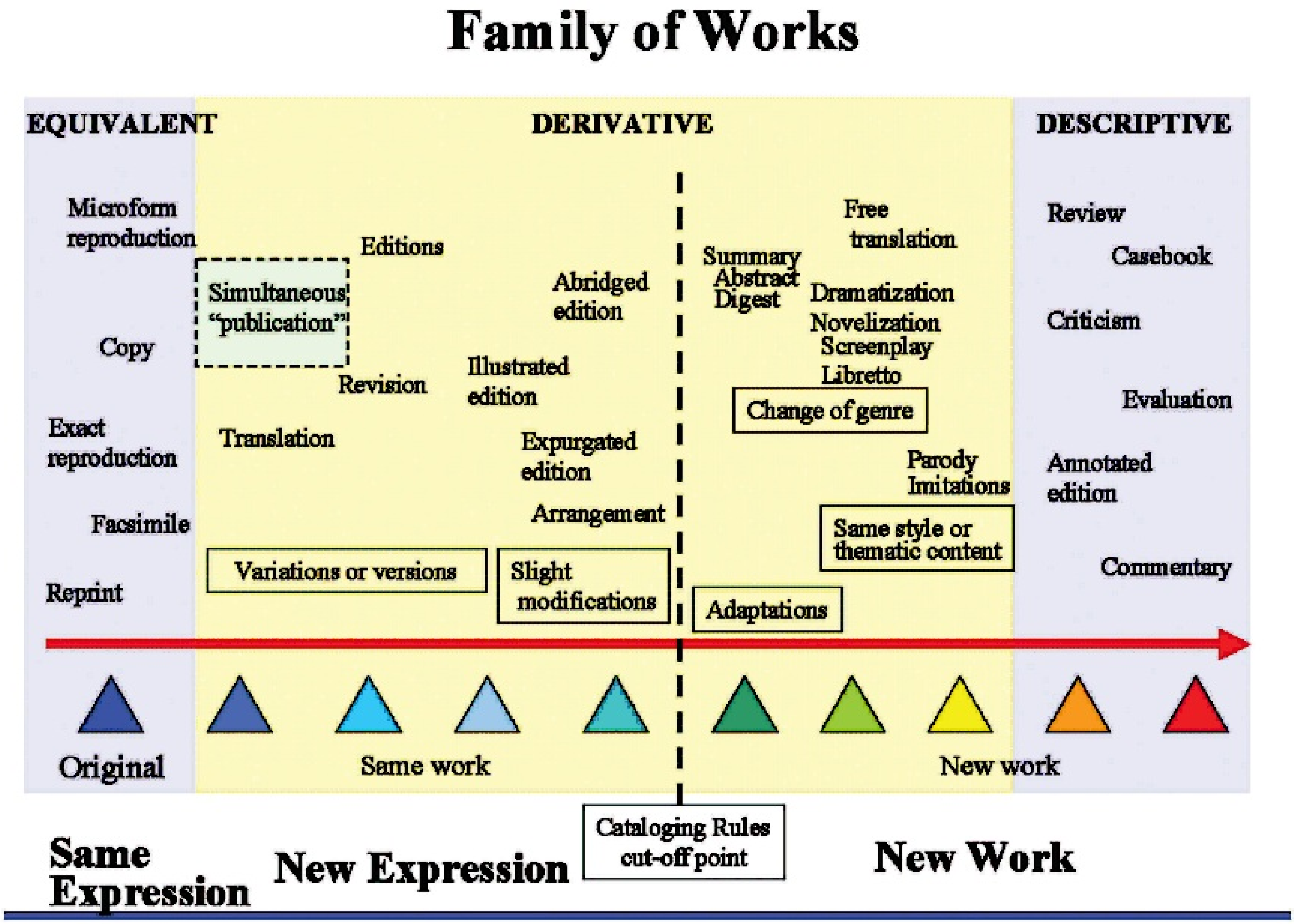
Notes: This listing does not include supplements issued between editions. Up to edition 9, the person(s) associated with the work are listed thus: *compiled by*, from edition 10 onward *edited by*.

Definition of units in bibliographic and record control

Soergel draft As the most inclusive term that is superordinate to all of the types defined here we will use *document* (in the broadest sense) or, even broader, *resource*.	**FRBR** **F**unctional **R**equirements for **B**ibliographic **R**ecords
Work Intellectual or artistic entity, as the abstract essence or as a text, image, or piece of music. Range: *A basic story or theme* the legend of Faust the myth of the Great Flood *A text telling the story*, such as Goethe's Faust the account of the Great Flood in the Bible (original Hebrew), a specific version of that text, a Latin version, the account of the same myth in another culture. *A novel* build on the general theme of the legend of Faust but set in a different time with different characters.	**Work** A distinct intellectual or artistic creation. **Expression** The specific intellectual or artistic form that a work takes each time it is ‘realized'
Manifestation (also called edition in one meaning of edition) A specific rendering of a work by means of a graphical image or sound, taken in the abstract; the idea of such an expression. Examples: The text of Goethe's Faust presented in a particular typeface and layout. (A performance at which the text is recited also renders the text but is a separate, but related, work.) A specific score of a given version of Schubert's Fifth. (A performance of that version of Schubert’s Fifth also renders the piece of music but is a separate, but related, work) Also the recording of a work in the form of digital storage that can be transformed to a graphical image or sound, again taken as the abstract pattern of digital signals. The actual file is an item`	**Manifestation** The physical embodiment of an expression of a work. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form.
Printing A set of books printed at the same time or printed at different times containing no more than slight variations	
Item (also individual copy or simply copy or physical copy) The embodiment of a manifestation in a physical object. We can perceive the content of a manifestation only through an individual item (copy) of it (unless we have memorized the contents of a manifestation and can conjure it up from memory). There are works that have only one manifestation of which there is only one item.	**Item** A single exemplar of a manifestation. The entity defined as item is a concrete entity

FRBR Functional Requirements for Bibliographic Records	253
Work	A work is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work. When we speak of Homer's Iliad as a work, our point of reference is not a particular recitation or text of the work, but the intellectual creation that lies behind all the various expressions of the work.
Expression	An expression is the specific intellectual or artistic form that a work takes each time it is "realized." Expression encompasses, for example, the specific words, sentences, paragraphs, etc. that result from the realization of a work in the form of a text, or the particular sounds, phrasing, etc. resulting from the realization of a musical work. . . . excludes aspects of physical form, such as typeface and page layout, that are not integral to the intellectual or artistic realization of the work as such. When an expression is accompanied by augmentations, such as illustrations, notes, glosses, etc. that are not integral to the intellectual or artistic realization of the work, such augmentations are considered to be separate expressions of their own separate work(s). Such augmentations may, or may not, be considered significant enough to warrant distinct bibliographic identification. Inasmuch as the form of expression is an inherent characteristic of the expression, any change in form (e.g., from alpha-numeric notation to spoken word) results in a new expression. Similarly, changes in the intellectual conventions or instruments that are employed to express a work (e.g., translation from one language to another) result in the production of a new expression. If a text is revised or modified, the resulting expression is considered to be a new expression.
Manifestation	The physical embodiment of an expression of a work. . . . encompasses a wide range of materials, including manuscripts, books, periodicals, maps, posters, sound recordings, films, video recordings, CD-ROMs, multimedia kits, etc. . . . . represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form. When a work is realized, the resulting expression of the work may be physically embodied on or in a medium such as paper, audio tape, video tape, canvas, plaster, etc. That physical embodiment constitutes a manifestation of the work. . . . Whether the scope of production is broad (e.g., in the case of publication, etc.) or limited (e.g., in the case of copies made for private study, etc.), the set of copies produced in each case constitutes a manifestation. All copies produced that form part of the same set are considered to be copies of the same manifestation. [Changes in physical form result in a new manifestation; examples:] changes affecting display characteristics (typeface, size of font, page layout, etc.), changes in physical medium (e.g., change from paper to microfilm), changes in the container (e.g., change from cassette to cartridge as the container for a tape).
Item	A single exemplar of a manifestation. . . . a concrete entity. It is in many instances a single physical object (e.g., a copy of a one-volume monograph, a single audio cassette, etc.). There are instances, however, where an item comprises more than one physical object (e.g., a monograph issued as two separately bound volumes, a recording issued on three separate compact discs, etc.). . . . variations may occur from one item to another, even when the items exemplify the same manifestation, where those variations are the result of actions external to the intent of the producer of the manifestation (e.g., damage occurring after the item was produced, binding performed by a library, etc.).

For still another definition of these types of entities see Supplement, SLecture 7.1b



This diagram from Tillet 2004 illustrates the boundary between work and expression in FRBR

The relationships between the bibliographic entities (Group 1) in FRBR are

Work	*<isRealizedThrough>*	Expression	(1:N)
Expression	*<isEmbodiedIn>*	Manifestation	(1:N)
Manifestation	*<isExemplifiedBy>*	Item	(1:N)

(A work can have many expressions, but an expression is always of one work)

FRBR includes other entity types, namely

Group 2. person (an individual) and **corporate body** (an organization or group of individuals and/or organizations).

Person or Corporate Body <*creates*>	Work	(N:M)
Person or Corporate Body <*realizes*>	Expression	(N:M)
Person or Corporate Body <*produces*>	Manifestation	(N:M)
Person or Corporate Body <*owns*>	Item	(N:M)

Group 3. Entities that serve as the **subjects of works**. The group includes **concept** (an abstract notion or idea), **object** (a material thing), **event** (an action or occurrence), and **place** (a location).

Work	<*hasAsSubject*>	Concept, Object, Event, Place	(N:M)
Work	<*hasAsSubject*>	Work, Expression, Manifestation, Item, Person, Corporate Body	(N:M)

You have read a description of
FRBR (Functional Requirements for Bibliographic Records)
and some more information is provided on the previous pages.

For better or for worse, FRBR is the conceptual basis for the revision of cataloging systems. So you need to be familiar with it.

FRBR takes a small step towards basing cataloging on an entity-relationship model, but is still rooted too much in the past, and its conceptual analysis could be improved.

The FRBR distinction between the entity types *work* and *expression* is poorly conceived. The distinction is so hard to make that it requires a lot of effort from catalogers for no benefit. All the information that is captured in this distinction can be captured better through proper relationships within one entity type *work*. The generic *Story of Faust* is a work, *Goethe's Faust* is a work, and an *English translation of Goethe's Faust* is a work; the relationships between these works are obvious.

Different pieces of information should be pegged at the right level.

Question: In the hierarchy of works, where should we record
<*dealswith*> Pact with the devil

Answer: The *Story of Faust.* This piece of information applies to all works that are based on the *Story of Faust*, so it can be entered into a (properly structured) catalog once and will then inherit down to all the works based on the *Story of Faust*.

Question: Where should <*authoredBy*> Goethe, Johann Wolfgang von appear?

Answer: Once with the record of this work. It will inherit down to the many manifestations of this work.

Question: Where **should** <*publishedby*> Cotta, <*publishedInDate*> [1903], <*printedIn*> [Fraktur, 11 point] (important when large print) appear?

Answer: In the record for a specific manifestation, one of many that renders the text of the work *Goethe's Faust*. Any of these manifestations would do to get the text, provided the prospective reader can read the type face.

An inventory number would be assigned to an individual item

In the cataloging of rare books information on individual items (individual physical copies) is important.

Look at the court case on electronic vs. printed copies of email on the next page as an example of the practical importance of considering different versions of a document.

This ruling is a good example of the importance of discussing the problem of different versions of a document. Emphasis added.

Public Citizen v. John Carlin, Archivist of the United States Oct. 1997
Overturned by Court of Appeals for the District of Columbia Circuit August 1999, but the reasoning is nevertheless important in our context.

Washington Post, Thursday, October 23, 1997, p. A21

Judge Nullifies Rule on Computer Data

Archivist Criticized for Letting Agencies Eliminate Electronic Records

By George Lardner Jr.
Washington Post Staff Writer

A federal judge held yesterday that the head of the National Archives ignored his duties and acted illegally in issuing a regulation that authorizes all government agencies to wipe out their electronic mail and other computerized records regardless of content.

In a 36-page ruling sharply critical of Archivist John W. Carlin, U.S. District Judge Paul L. Friedman declared the controversial rule "null and void" and called the government's defense of it "irrational on its face."

The two-year-old regulation, known in bureaucratic jargon as "GRS [General Records Schedule]-20," permitted all agencies, from the Executive Office of the President on down, to destroy e-mail and wordprocessing records once they have been copied on paper or some other format and deemed "no longer needed for updating and revision."

Historians, researchers and journalists represented by the nonprofit advocacy group Public Citizen denounced the provision as an "electronic shredder" and filed suit, accusing Carlin of abdicating his responsibilities to appraise the value of the records on an agency-by-agency basis.

Friedman agreed. "Simply put," he said, **"electronic communications are rarely identical to their paper counterparts; they are records unique and distinct from printed versions of the same record."**

Citing an example from the Iran-contra scandal, the judge pointed out that so-called PROF notes-computerized messages between national security adviser John M. Poindexter and White House aide Oliver L North-played an important role in the trials of both men.

"Admiral Poindexter, a computer expert set up a special channel known as "Private Blank Check," which allowed North and Poindexter to relay messages to each other without those messages being accessible to other NSC staff," noted Friedman, who was once an Iran-contra prosecutor. "The communication itself was clearly important to investigators, but the mode of communication and the special channel through which it was sent, which would not have been reflected in paper printouts of the messages, was also important."

In promulgating GRS-20 in 1995, the judge said, Carlin categorically determined that electronic records possess no administrative, legal, research or historical value beyond paper printouts of the same document. In doing this, "the Archivist has absolved both himself and the federal agencies he is supposed to oversee of their statutory duties to evaluate specific electronic records as to their value."

Carlin, the judge said, also exceeded his authority in giving agencies "carte blanche" to destroy electronic versions" whenever agency officials believe they are no longer needed.

The government had argued that GRS-20 was soundly based because such government-wide rules were meant for records of common form, such as "electronic" media.

Lawyers for Carlin had also protested that most federal agencies are not yet equipped to preserve records in electronic format. Friedman said this was " an important concern" but observed that "computers have now become a significant part of the way the federal government conducts its business" and the government must adapt to that reality.

The Archives had no immediate comment.

Structure of a better catalog - next page

Example illustrating the idea of an interlinked catalog

Consider this document from the list of sample documents given earlier:

(4) *The man I killed*, play by Christopher Wern, based on the novel by Michael Halliday

The novel referred to is Document (1)

(1) *The man I killed*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1935

How should (4) be cataloged?

Who is the author? Most would say Wern. But should this work also be found in a search for works authored by Halliday? Some users may want to find it, others may not.

In the present system where each book (each manifestation) is cataloged on its own, the cataloging rules must prescribe whether to make the Wern play findable under Halliday or not; the user is then stuck with this decision. The choice is made by the system (the cataloging rules), not the user.

The better solution would be this:

- Catalog (1), with author Halliday (whether Halliday or Creasy is a question addressed in **7.2a**).
- Catalog (4) with author Wern and a <*basedOn*> link to (1) but no further information. Of course, (4) <*basedOn*> (1) could also be written as (1) <*basisFor*> (4).
- Have a search system where the user looking for books authored by a person, in the example Halliday, can choose whether to limit the search to books where Halliday is listed directly as the author or whether to follow links <*basisFor*> to more documents such as (4).

Two principles can be derived from this example:

1 **For all information systems**: **Leave choices to the user**. This means

1.1 the systems data need to describe the real situation faithfully and precisely without any distortions due to system rules and limitations;

1.2 the system must give the user many options of searching, combining the data in various ways.

2 **For bibliographic catalogs**: **No bibliographic unit (work, manifestation) is an island.** Bibliographic units are linked in a complex network that should be faithfully represented in the catalog and used in searching. (The hierarchical inheritance examples discussed earlier also fall under this principle.)

How to design a better catalog system

The key to designing an efficient database structure for a catalog lies in analyzing and applying the relationships between bibliographic entities. The root cause for the complexity of many cataloging rules is the attempt to force data with very complex relationships into a simple-minded data structure.

Elements of a conceptual data schema for a database with data about documents.

<*isVersionOf*>, more specific <*isTranslationOf*>

<*isPartOf*>

These relationships may hold between works, between manifestations, or between items.

Work <*isRenderedIn*> Manifestation or (inverse) Manifestation <*isRenderingOf*> Work

FRBR chain: Work <*isRealizedThrough*> Expression,

Expression <*isEmbodiedIn*> Manifestation

Manifestation <*generatedFrom*> (Manifestation, RenderingProcess)

(Examples of rendering processes: different screen renderings from same HTML source text through different browsers; facsimile; Optical Character Recognition. The distinction between a rendering process and a reproduction process is fluid.)

Manifestation <*isInstantiatedIn*> Item or (inverse) Item <*isInstantiatedIn*> Manifestation

FRBR: Manifestation <*isExemplifiedBy*> Item

Item <*reproducedFrom*> (Item, ReproductionProcess)

Some problems in a conceptual data schema for bibliographic and record control

Records, originals vs. copies

Permanent copy vs. fleeting copy

Specific printing may use different paper - preservation!

Performance of a work may be more than a mere manifestation since it brings separate creative elements. Perhaps a performance should be considered a work of its own, with the tape (or audio file) on which it is captured being an item of a manifestation of that work (remastering such a tape would create another manifestation).

Definition of "catalog" - elaboration

A work is *covered by a catalog* if the catalog contains data about the work, or any manifestation of the work, or any item (individual copy) of any manifestation of the work. The collection linked with a catalog may be either a collection of items or merely a list of works, manifestations, or copies; a manifestation is said to be *represented in a collection* of items if any item of the manifestation is in the collection. In the electronic age the concept of *collection* becomes more and more fluid. Is the whole Web one collection, or is a collection confined to the documents (files) stored on one Web server? Likewise, the concept of *library* becomes more and more fluid; there are now digital libraries whose "collections" may be distributed over many sites (whence the term *virtual library*). In fact, there is no sharp distinction between a digital library and any computerized information system. The functional distinctions made in Section 2.6 of the Textbook are useful to clarify some of the issues here.

Objectives of the library catalog - restatement by D. Soergel

The catalog (of a library, a book seller, ...) should be an efficient instrument for ascertaining

(1) **Criterion search (intellectual access)** [4.1.2]
Which works, manifestations, or items are helpful to a given user for a given purpose, to wit

(1a) which works, manifestations, or items covered by the catalog meet a combination of criteria relating to provenance (including authorship), subject, artistic characteristics, and/or other criteria (**retrieval** or **identification**) (in some cases only certain manifestations or individual items may meet the user's search criteria) (**find** a set of resources) [4.1.2];

(1b) whether a work, manifestation, or item meets the needs of the user and how several suitable works, manifestations, or items should be ranked (**selection**) [4.3];

(1c) how a work, manifestation, or item relates to (another) work, manifestation, or item (**relation**) (for example, <*isRevisionOf*>, <*isReprintOf*>, <*isBasedOn*>);

(2) **Search for a known** work, manifestation, or item (**find** a single resource) [4.1.1]
(Confusingly, this is called *known-item search*, a term coined before FRBR)
To ascertain that a resource given in the catalog is the same as the resource in hand (**identity**) [4.2]

(2a) whether a known work is covered by the catalog and, if so, which are the manifestation(s) of the work that are covered in the catalog (**coverage**);

(2b) whether a known manifestation is covered in the catalog;

(2c) whether a known item (specific copy) is covered in the catalog (important for rare books)

(3) how the user can get **physical access** to some item (copy) of the work (method of access, time, cost) (**acquire** or **obtain)** [4.4].

The objectives are arranged by decreasing complexity and increasing concreteness, not by importance. After achieving objective 1 a user must achieve objective 2 and then objective 3.

The Statement of International Cataloging Principles (a reading), Section 4, presents a somewhat different organization of objectives of the catalog; their numbers given in []

The Statement lists 4.5 to **navigate**, but navigation is a means for achieving any of the objectives, just as query-based search, so it does not belong in this list. Of course, a catalog should support both

Note: Many of these objectives apply to searching for people, organizations, software, or whatever.

pink **Lecture 7.1c** (25 min) ***March 9***

Bibliographic and record control: Description

Describing texts and documents in a more general context

General principles of description; their application to bibliographic and record control; their implementation in ISBD/AACR2; relationship to the MARC format. User-oriented analysis of elements of description needed.

Description: What needs to be known about an entity?

Relates to catalog objective 1b, ascertaining relevance. Also relates to objective 3, ascertaining whether a given manifestation is indeed the same manifestation that is covered in a catalog record. (The "given manifestation" may be the manifestation requested by the user or the manifestation of which the item in hand is an instance.)

Data about bibliographic entities - conceptual data schema.

Peg each piece of data to the correct bibliographic level (work, manifestation, item).

How to you decide at which bibliographic level (work, manifestation, item) a piece of information should be given? (See also the discussion right after the FRBR picture above)

In used book seller catalog (or union catalog, such as www.abebooks.com) there are descriptions of used books for sale. They often describe the condition of the book, such as

Book Condition: Used - Good.. Shows some signs of wear, and may have some markings on the inside

Hardcover. No dust jacket. Used, good. EX-LIBRARY - has usual library wear/markings/attachments. Small tears/creases on spine/cover. Cover has some edge wear

What bibliographic level does this part of the description pertain to?

Note: For rare books description is much more extensive.

Looking ahead: It is efficient to mark up the description in such a way that entries (access points) can be extracted easily. Cataloging in the MARC format achieves this

Sources of information for cataloging data

Primary: Title page and verso (back of title page),
Secondary: Preface, last page, cover page of a journal issue, etc. (data shown in [])

Which is source is most authoritative?

Sometimes it requires considerable inquiry to find information required by the conceptual data schema (the cataloging rules) but not known from the representation in the item being examined. In the extreme, this involves long research into the authorship of a play or of a painting.

pink

Arrangement of cataloging data

- **in a record** - record format such as MARC21 (eventually to be replaced by BibFrame)
- **in a display** (printed or on a screen)

Many different styles: AACR2, ANSI Standard, Turabian, American Psychological Association (over 4,000 and counting).

Let a computer program, such as Endnote, do the work!

MARC is just one record format. There are many others. We will talk later about Dublin core and other Web-based initiatives

Open Archives Initiative www.openarchives.org/

The record format should store data fine-grained so that many different displays can be supported (for example, journal volume, issue number, year, month, day, beginning page, end page should each be a separate piece of data because different display formats arrange these in different ways.

Display format

An OPAC (Online Public Access Catalog) should give the user considerable control over the display format (but they rarely do); systems like Dialog (now part of Proquest) that provide access to bibliographic and other databases are usually much better at this.

Publishers and journals prescribe one of the many available display formats for use with their publications; you should be familiar with some of them and be able to assist users with preparing bibliographies. In many courses you are required to use APA style, a very common style, in formatting bibliographic records in your bibliography.

Unfortunately, these are mostly based on personal preferences of someone in charge or who was in charge 50 years ago rather than on empirically based analysis of what is most functional for users. Thoreau' s quote applies to many of these: "Any fool can make a rule, and every fool will mind it". For example, for scanning a bibliography the most important piece of information is the title of the document itself, yet APA style requires highlighting (in italics) the title of the journal in which an article appeared. I always bold the title of the document.

There are many **bibliographic citation managers**, among them

Endnote (on your computer and the Web, www.endnote.com/),

Library Master (more powerful than most but fewer users, www.balboa-software.com),

Zotero (entirely Web-based but may have a way to store data on your computer as well, www.zotero.org)

Mendeley (on your computer and the Web, www.mendeley.com)

These programs have style sheets (many premade, but you can also make your own) for any conceivable display format.
You should use one of these bibliography managers for your work. UB has a site license for Endnote, use version X4 (http://library.buffalo.edu/libraries/endnote/)

pink **Lecture 7.2a** (40 min) ***March 9***

Bibliographic and record control 2: Entries

General principles of access; their application to bibliographic and record control; their implementation in AACR2 choice and form of entry. Authority files. User-oriented analysis of access points needed.

Definition	An *entry* is an element, such as an author name, a title, a series title, or a subject descriptor under which a document (or another object) can be found in a catalog or index. (The term comes from book or card catalogs, where an entry for a document is made by writing or filing a card.) Determining entries is a problem of data structure and access.
Two issues:	A Which of the data in the description should be made access points for lookup searching? (The answer to that question might have repercussions for description if a data element is important for access but not for ascertaining the relevance of an item.) B What form should each entry take? (Rules for entity values)
Main entry **Added entry**	A document may have many authors / contributors. Most of the time, this does not present a problem: just list them all and provide access from all ("make an entry," as in a card catalog, for all). But sometimes we want to list a document record only once: In a printed bibliography, in a listing of search results arranged by author, in a card catalog before reproduction equipment when every card had to be written or typed by hand. In that case, we need to select the most significant author / contributor, the one under which the one entry should be made. This is the <u>main entry</u>. (This concept was quite important in the age of card catalogs; it is less important now but still has applications.)
In-lecture exercise	**Problems of determining author entry analyzed according to Lubetzky** Lubetzky was the foremost thinker on bibliographic cataloging rules.
Advanced	Thinking about rules for corporate entry

The Author approach: Conditions and cases (Lubetzky after Needham)
See next page. Lubetzky's discussion of possible solutions is found in the readings.

pink **Condition**	**Issue A: What entries to make**	**Issue B: What form the entry should take**
1. Documents having more than one Author	1.1 Document prepared by an author with the aid of collaborators or contributors.	
	1.2 Document composed by an editor or compiler from the writings of several other people	
	1.3 Document by several authors with no one author more responsible for it than any of the others.	
	1.4 Document in which the writer reports the communication of another person (real or fictitious).	
2. Authors having more than one name		2.1 The author has changed his or her name in consequence of marriage, adoption of new citizenship, joining a religious order, or for any other reason.
		2.2 The author always writes under an assumed name different from his real name, or under his title of nobility, or under part of his name.
		2.3/4 Author uses more than one name in successive documents
		2.5 Authors whose names appear in translation in varying forms.
3. Dependent Documents		
4. Corporate authors	4.1 The reports and statements of a corporate body are usually prepared by one of its officers or by another person engaged to prepare the statement for it.	4.2 Many corporate bodies have no proper identifying names of their own but only generic names describing their type and common to most bodies of that type e.g. public library, historical society, dramatic club, etc.
		4.3 Change of name in corporate bodies.
		4.4 An organization may act or speak as a whole or through one of its branches, divisions, offices, etc.

Sample documents for analyzing author entry according to Lubetzky

For the sample documents, think about

(1) Who should be listed as author for purposes of finding the document?
(2) For the people selected as authors, in what form should their named be recorded

How can sample documents (e) and (t) be used to illustrate the idea of a interlinked catalog?

There is an "interactive" PowerPoint presentation for this exercise,

(a) *The record guide* by Edward Sackville-West and Desmond Shaw-Taylor, with Andrew Porter and William Mann.

(b)* *Studies in the social psychology of adolescence*, by J. E. Richardson, J. F. Forrester, J. K. Shukla, and P. J. Higginbotham; edited with a foreword by C. M. Fleming.

(c) *The tropics*, by Edgar Aubert de la Rue, Francois Bourliere, Jean-Paul Harroy.

(d) *Ambit* (a periodical), edited by M. C. O. Bax and Edwin Brock.

(e)* *Chisholm's handbook of commercial geography*, entirely rewritten by L. Dudley Stamp and S. Carter Gilmour.

(f)* *Making magical apparatus*, by Jane Reid (i.e. Mrs. David Johnstone).

(g) *Lord Jim*, by Joseph Conrad (i.e. Josef Theodor Konrad Korzeniowski).

(h) *The far country*, by Neville Shute (i.e. Neville Shute Norway).

(i) *The trimmed lamp, and other stories*, by O. Henry (i.e. William Sydney Porter).

* designates items to be analyzed in Assignment 9

(j) *The scene of the crime*, by John Creasey.

(k) *The man I killed*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1935

(l) *A branch for the baron*, by Anthony Morton (i.e. John Creasey).

(m)* *Schubert: thematic catalogue of all his works in chronological order*, by Otto Erich Deutsch in collaboration with Donald R. Wakeling.

(n)* *A concordance to the poems of William Wordsworth*, by Lane Cooper.

(o)* *The poetical works of Wordsworth*, edited by E. de Selincourt.

(p) *Oxford book of English verse, 1250-1918*, chosen and edited by Sir Arthur Quiller-Couch.

(q) *Shakespeare's 'Much ado about nothing'*, by N. T. Carrington (Notes on chosen English texts). Text and commentary.

(r) *The man I killed*, by Michael Halliday (i.e. John Creasey). Audiotape, read by Sir Lawrence Olivier. New York: Books on Tape; 1966.

(s) *The man I killed*, play by Christopher Wern, based on the novel by Michael Halliday.

(t) *The Aeneid of Virgil*, retold by N. B. Taylor.

(u) *Billy Budd* (libretto), adapted from the story by Herman Melville by E. M. Forster and Eric Crozier.

(v) *Iban agriculture: a report on the shifting cultivation of hill rice*, prepared for the Colonial Office by John Derek Freeman.

(w) *Essays and studies, 1962: being volume fifteen of the new series of Essays and studies*; collected for The English Association by Beatrice White.

(x) *The Library Association Record*, edited by Edward Dudley.

(y) Ministry of Education. 15-18. (The Crowther report).

(z) *Yearbook* of the Institution of Agricultural Engineers (originally the Institution of British Agricultural Engineers).

(aa) National Physical Laboratory. *Mathematical tables*. (The National Physical Laboratory is a branch of the Department of Trade and Industry.)

(bb) *Farm business statistics for south-east England*. Wye College (London University).

(cc) *Annual report* of the Association of Assistant Librarians (a division of the Library Association).

Observations on form of name

The principle **Leave choices to the user** and **Record data faithfully** applies here as well. Consider John Creasey, who wrote under three names: John Creasey, Michael Halliday, and Anthony Morton. Some catalog rules prescribe that in all cases John Creasey should be used as the form of name, the identifier for this one person. However, a user who liked one book by Anthony Morton and wants to find similar books may not be pleased at all by finding books John Creasy wrote under his real name because those books might be quite different (there might be three authorial spirits living in John Creasey's body, and for this user the entity of interest is the authorial spirit, not the actual person). For another example consider someone who has read Alice in Wonderland and wants more books by the same author, listed as Lewis Carroll, and is handed a mathematical treatise the same person wrote under his real name Charles Lutwidge Dodgson. The solution is as follows:

(1) Catalog books under the name of the author as it appears on the title page.

(2) Create an authority file of personal names (a database with information about people, including their names) that links together all names that refer to the same person. Such a database exists at the Library of Congress (http://authorities.loc.gov) and is used by many libraries. It includes birth and death dates and all names, with time periods during which a name was in effect, if necessary, and other information.

This leaves the choice to the user: She can look for all books authored by the "authorial spirit" Anthony Morton or expand the search to all names listed with the person John Creasey.

Problems of entry for works emanating from corporate bodies

What is a corporate body?

The **definition of a corporate body** according to AACR2:

"21.1B Entry under corporate body

21.1Bl. Definition. A corporate body is an organization or a group of persons that is identified by a particular name and that acts, or may act, as an entity. Consider a corporate body to have a name if the words referring to it are a specific appellation rather than a general description. If, in a script and language using capital letters for proper names, the initial letters of the words referring to a corporate body are consistently capitalized, and/or if, in a language using articles, the words are always associated with a definite article, consider the body to have a name. Typical examples of corporate bodies are associations, institutions, business firms, nonprofit enterprises, governments, government agencies, projects and programs, religious bodies, local churches, and conferences.

Note that some corporate bodies are subordinate to other bodies (e.g. the Peabody Museum of Natural History is subordinate to Yale University.)

Consider ad hoc events (such as athletic contests, exhibitions, expeditions, fairs, and festivals) and vessels (e.g., ships and spacecraft) to be corporate bodies."

Issue A. **When to make an entry under a corporate body** (corporate entry)?
(When to establish a relationship between a work and a corporate body?)
One question to be asked: When does a corporate body have a role similar to the responsibility of an author (corporate authorship)?

AACR2:

"21.1B2. Footnote 2. Consider a work to have emanated from a corporate body if it is issued by that body or has been caused to be issued by that body or if it originated with that body."

A corporate body can

- be fully responsible, as in a law enacted by a government or the official statement of an organization;
- have commissioned a work;
- have issued / published a work;
- provided the environment in which a work was created, such as a university or Rand providing an environment for a researcher (usually handled, if at all, as author affiliation).

Easiest solution: Make a good list of these relationship types and use the specific relationship when cataloging a work. This does not fit into the prevailing system of cataloging; there are just the MARC fields 110 and 710 for corporate names (corporate body in an author-like role). AACR2 gives rule under what circumstances such entries should be made.

Issue B. What form of name should be used for the corporate body?

See AACR2 Chapter 24

Problems:

- Corporate bodies change their names, for example
  Bureau of Foods became *Center for Food Safety and Applied Nutrition*.
- More complex problem: Corporate bodies change, cease to exist, or are merged with other corporate bodies.
- Corporate bodies are part of other corporate bodies, for example
  US. Department of Health and Human Services. Public Health Service. Food and Drug Administration. Center for Food Safety and Applied Nutrition. Technical Operations Branch.
  Some levels in such a hierarchy are often better known than others, for example,
  Food and Drug Administration is better known than
  Public Health Service
  (possibly a phenomenon similar to basic level concepts)
  There are other relationships, such as corporation B being a wholly-owned subsidiary of corporation A (interesting if you want to sue corporation B for damages).
- Corporate bodies often have generic names that are meaningful only in conjunction with a place, such as
  Metropolitan Museum of Art
  Some rules suggest to put the name of the place first:
  New York Metropolitan Museum of Art
  Cleveland Metropolitan Museum of Art
- Corporate bodies are often better known under an acronym or short popular name, for example
  FDA is better known than *Food and Drug Administration*

These problems occur in any information system that deals in any way with corporate bodies. The best solution is to have a database of corporate bodies including all their names, their life span, their relationships, and information on place. Such a database could be used in conjunction with a bibliographic catalog or other information systems.

Observations on names of corporate bodies
At one point it was thought that users should be able to find all documents emanating from a corporate body under one name, the most recent one, even over time. In a card catalog, that meant when the name of a corporate body changed, all the catalog cards for documents cataloged under the previous name had to be changed to the new name, a process known as superimposition. But then the prevailing opinion changed to prefer the previous system which used the name of the corporate body as found in the document cataloged; so all these catalog cards had to be changed back, a process known as desuperimposition. Now that computers are available, we can have our cake and eat it too: Make an authority file for corporate names, a database with information about corporate bodies, including their names over time. Such a database is included in the Library of Congress name authority (http://authorities.loc.gov). With such a database, a user could search for all the names under which a corporate body is or was known, or she could restrict her search to one particular name and find only documents associated with that name. It would be even better if it included hierarchical and other relationships among corporate bodies: subordinate corporate bodies as in the hierarchy of the federal government or the various subordinate organizations within ALA (such as the American Association of School Librarians (AASL)(or the relationship between a holding company and the companies it holds. If the system had that information, the user could do an inclusive search for all documents emanating from a corporate body or any of its subordinate bodies.

Hierarchy of corporate bodies
The federal government is hierarchically organized, as shown in this example: **Department of Health and Human Services** Public Health Service **Food and Drug Administration** Center for Drugs Center for Food Safety and Applied Nutrition Technical Operations Branch The major departments are generally known in the population. But below that there are some levels that stand out, such as FDA, while others above or below these agencies are not often known to the general public. This is a phenomenon similar to basic level concepts.

Advanced exercise: Thinking about rules for corporate entry.

See Supplement, SLecture 7.2a

Concluding note

Cataloging rules were originally developed for the card catalog. AACR2 shows this clearly, and even though RDA was developed explicitly for the electronic age, it still carries some of the legacy.

My suggestions come from a radical rethinking of cataloging rules "from scratch", serving user needs with what is possible using the power of modern computer systems, specifically the ability to produce results by combining many different kinds of data and the ability to tailor the selection and display of data to specific needs.

pink **Lecture 7.2b** (25 min) ***March 9***

Bibliogr. data on the Web: BibFrame, Dublin Core, Open Archives Initiative, ...

1 Bibliographic Framework Initiative (BibFrame)

"Initiated by the Library of Congress, BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world. This site presents general information about the project, including presentations, FAQs, and links to working documents. In addition to being a replacement for MARC, BIBFRAME serves as a general model for expressing and connecting bibliographic data. A major focus of the initiative will be to determine a transition path for the MARC 21 formats while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades."
http://www.loc.gov/bibframe/ http://en.wikipedia.org/wiki/BIBFRAME

2 Dublin Core (DC) (http://dublincore.org/)

The Dublin Core is a minimal standard for the description of "document-like objects"

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

Definitions in Supplement, SLecture 7.2b

"Plain" Dublin Core has just 15 properties (relationship types), simplicity both good and bad. The Dublin Core list of properties can be implemented in many ways, among them:

In the *meta* section of an HTML document (see next page and the model catalog)

In plain XML

In XML, using the RDF enhanced syntax

The Dublin Core is a conceptual data schema for bibliographic description that is simpler than MARC and could thus be used widely by people not trained in cataloging. Bibliographic description is understood to cover a broad range of documents and "document-like objects", especially Web pages. The DC set of fields is shown in the box. Expansions and refinements can and have been created for different types of materials, such as learning objects.

You must be familiar with the Dublin Core, but do not memorize the DC data fields.

See http://dublincore.org/, esp. DCMI specifications

The following examples of how the Dublin Core is used are optional (starting next page).

pink **Here are a few lines of an HTML document with Dublin Core metadata**

```
<HTML>
    <HEAD>
      <META name="dc.creator" content="Renato Iannella">
      <META name="dc.creator.affiliation" content="DSTC">
      <META name="dc.subject" content="Cats, Fur, Purr">

      Specification of the scheme from which subject descriptors were taken, here LCSH
      <META name="dc.subject" content="(scheme=LCSH) Animals~Felines">

      Language of title or subject descriptors can be specified in like manner
    </HEAD>
    <BODY>
      . . .

    </BODY>
</HTML>
```

pink **Example of an RDF resource description: simple document description**

Here is a simple example of RDF syntax used to describe a resource. This example uses properties defined in the Dublin Core (DC). RDF tags (prefixed with **rdf:** and Dublin Core (DC) tags (prefixed with **dc:**) are defined in files found at the URLs given and then used in the record. Name spaces ensure that there is no collision between tag names in the two syntaxes (imagine what happens when two language syntaxes, each defined in XML, use the same tag name, like "description", to define different data fields. Without the use of distinct name spaces, things would get horribly confusing, both to document creators and the automated systems parsing XML documents).

```
<?xml version="1.0"?>
<rdf:RDF
   xmlns:rdf="www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.0/">
   <rdf:description  rdf:about="www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/">
         <dc:title> Guidance on expressing the Dublin Core within RDF </dc:title>
         <dc:creator> Eric Miller </dc:creator>
         <dc:creator> Paul Miller </dc:creator>
         <dc:creator> Dan Brickley </dc:creator>
         <dc:subject> Dublin Core; RDF; XML </dc:subject>
         <dc:publisher> Dublin Core Metadata Initiative </dc:publisher>
         <dc:contributor> Dublin Core Data Model Working Group </dc:contributor>
         <dc:date> 1999-07-01 </dc:date>
         <dc:format> text/html </dc:format>
         <dc:language> en </dc:language>
   </rdf:description>
</rdf:RDF>
```

Note: This is an extended frame: Each ***<dc:*** is a statement about the document given through the URL after ***about*=**, such as www.ukoln. **. . .** <hasTitle> " Guidance on **. . .**

More complex example illustrating more features of RDF and refinements of the Dublin Core.

```
<?xml version='1.0'?>
<rdf:RDF
    xmlns:rdf="www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.0/"
    xmlns:dcq="http://purl.org/dc/qualifiers/1.0/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/>  /* foaf = friend of a friend, http://xmlns.com/foaf/spec/ */
    <rdf:description  rdf:about="www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/">
        . . .
        <dc:creator>
            <rdf:Description>
                <rdf:type rdf:resource="http://purl.org/dc/terms/1.0/creator/class/Person"/>
                <dcq:creatorType rdf:resource="http://purl.org/dc/terms/1.0/creator/type/Editor"/>
                <rdf:value rdf:resource="http://411.com/EricMiller"/>
            </rdf:Description>
        </dc:creator>
        . . .
    </rdf:description>

    <rdf:description rdf:about="http://411.com/Eric Miller">
        <foaf:name>Eric Miller </foaf:name>
        <foaf:homepage>rdf:resource="www.oclc.org/~emiller" </foaf:homepage>
        <foaf:img rdf:resource="/images/emiller.jpg" />
    </rdf:Description>
 </rdf:RDF>
```

3 The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

A simple format for metadata describing resources on the Web. The idea is that providers of resources (for example, a large repository or digital library of reports produced by a national research laboratory) include with each resource metadata in the prescribed format or store these metadata separately on their website. A system maintaining a large bibliographic database can then "harvest" these metadata from many repositories and build its database without laborious cataloging, allowing users to search for documents from many repositories in one place. http://www.openarchives.org/pmh/

4 Other schemes

There are a number of other schemes, for example:

BIBO. The Bibliographic Ontology http://bibliontology.com/
"The Bibliographic Ontology Specification provides main concepts and properties for describing citations and bibliographic references (i.e. quotes, books, articles, etc) on the Semantic Web."

From **schema.org: Thing > CreativeWork > Article > ScholarlyArticle** Excerpts on next page
see http://schema.org/ScholarlyArticle for full list
Covers many of the types of information that MARC provides. The excerpt includes mostly items that go beyond bibliographic information commonly recorded in the library world.

Thing > CreativeWork > Article > ScholarlyArticle

Property	Expected Type	Description
Properties from Article		
articleBody	Text	The actual body of the article.
articleSection	Text	Articles may belong to one or more 'sections' in a magazine or newspaper, such as Sports, Lifestyle, etc.
pageEnd	Integer or Text	The page on which the work ends; for example "138" or "xvi".
pageStart	Integer or Text	The page on which the work starts; for example "135" or "xiii".
pagination	Text	Any description of pages that is not separated into pageStart and pageEnd; for example, "1-6, 9, 55" or "10-12, 46-49".
wordCount	Integer	The number of words in the text of the Article.
Properties from CreativeWork		
about	Thing	The subject matter of the content.
accountablePerson	Person	Specifies the Person that is legally accountable for the CreativeWork.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
audience	Audience	The intended audience of the item, i.e. the group for whom the item was created.
author	Person or Organization	The author of this content.
award	Text	An award won by this person or for this creative work. Supercedes awards.
citation	CreativeWork or Text	A citation or reference to another creative work, such as another publication, web page, scholarly article, etc.
comment	Comment	Comments, typically from users, on this CreativeWork.
commentCount	Integer	The number of comments this CreativeWork has received.
discussionUrl	URL	A link to the page containing the comments of the CreativeWork.
editor	Person	Specifies the Person who edited the CreativeWork.
educationalAlignment	AlignmentObject	An alignment to an established educational framework.
educationalUse	Text	The purpose of a work in the context of education; for example, 'assignment', 'group work'.
P**roperties from Thing**		
alternateName	Text	An alias for the item.
description	Text	A short description of the item.
image	URL	URL of an image of the item.

More specific Types MedicalScholarlyArticle.

Midterm exam sample questions

There will be 5 questions. 15-20 minutes = number of points each, for a total of 90 minutes.

(20) 1. **Document template**

P2.3.1,5.3# 571-L6.1b Be able to design a document template or template system at the conceptual level so it can be implemented by the IT department.

A congregation has a calendar database, which lists for each event (religious service, committee meeting, etc.) the title, time and place, and main participants (celebrant, speaker, etc.). They also have a separate calendar with the dates and names of holy days. They need to produce the following documents.

1a A congregational bulletin that lists the week's events.

1b A Web page that is meant to attract prospective members and that gives the week's events.

2 For each service a leaflet with the program (date and name of holy day, scripture passages, hymns, main participants).

Sketch a document template for document 1a or 1b and for document 2.

(15) 2. **Conceptual data schema**

P2.3.4,2# 571-L2.2 Graduates are able to select or design the knowledge structure for a new info. system
571-L4.2 Specifically, graduates are able to select or design a conceptual data schema for an information system based on user requirements.

Describe the process of developing a conceptual data schema for an employment service (an information system on jobs/positions and on people seeking jobs with the purpose of finding matches between jobs and people). Illustrate your discussion with examples, including some of the key entity types and relationship types.

(15) 3. **Search with two databases**

2.3.2,1#	571-L1.2 571-L2.2 571-L4.1	Graduates are able to combine different kinds of data to find an answer (inference, chained searches) (BT 2.5.2.1

Assume you have

- a database of recipes that gives ingredients (basic foods) and their amounts for many prepared dishes, and
- a nutrient database that gives for each basic food the amount of each nutrients (proteins, carbohydrates, fats, vitamins and minerals) it contains.

You want to find all prepared dishes that contain both vitamin A and vitamin D. Describe a process that could be followed by an automated retrieval system to accomplish such a search.

(20) 4. **Restructuring a semantic network**

2.3.1,2.1#	571-L2.2 571-A4 571-L5.1	Graduates understand the principle of hierarchical inheritance and are able to apply it to achieve more compact internal storage – in the computer and in the mind – and more compact and more easily grasped external representation.

There will be a question on restructuring a semantic network with application to bibliographic records.

(20) 5. **Linguistic techniques. Free-text searching refinements**

P2.3.3,1#	571-L6.1b	Have a general understanding of prevalent document analysis methods, including word sense disambiguation (WSD). Be aware of the many useful applications of document analysis in dealing with vast quantities of text, including automatic indexing for information retrieval, automatic abstracting and summarization, information extraction, and automated translation. Be able to recognize situations where information extraction and other methods of document analysis would be useful, to suggest the use of appropriate software, and to participate in the selection of such software. RT 2.3.1.5.7 – 11 (a list of methods can be found there)

This question has to do with possible improvements in retrieval through linguistic techniques. Below is an example consisting of a query and some short passages of text. Assume a straightforward free-text search system that searches for **words**; all passages that contain all query words joined by AND are retrieved. As a refinement, the system provides the proximity operator **WS**, which means **within the same sentence**. Thus, the query formulation *forest* **WS** *fire* retrieves all passages in which the two words occur in the same sentence; such passages are considered **retrieved**. See the instructions on the next page.

In the table of passages below, check all passages that are relevant to the user's need as expressed in the query. Then check all passages that are retrieved by the query formulation. Based on these checks, fill in the 3x3 grid and compute recall and discrimination. What refinements could be used to improve retrieval performance? For each refinement, list the passages whose retrieval is affected (give passage numbers). Which of these refinements improve recall, which improve discrimination, and which improve both?

Query topic: Forest fires

Query formulation: forest **WS** fire* (fire* finds fire or fires)

Passage	Relevant	RetrIrved
1 Forest fires in Indonesia cause serious air pollution in South East Asia.		
2 The fire in Yellowstone Park destroyed 25% of the forest.		
3 The fire station is located behind the city forest.		
4 With fire in her eyes she chased him through the forest.		
5 The soldiers opened fire into the forest.		
6 The fire went out of control. It reached the forest and destroyed many acres.		
7 The animal got scared by the fire burning in the field. It ran into the forest.		
8 He asked whether he should fire the forest workers.		
9 Many square miles of forest in the West are burning.		
10 The dry wooded area went up in flames.		

	Relevant	Not relevant	All
Retrieved			
Not retrieved			
All			

red

Weeks 8 - 13. March 9 - April 27

Part 5.
Knowledge Organization Systems (KOS)
Classification and subject access

Strand 1: General principles of KOS functions and structure

Strand 2: Analysis of highly used Knowledge Organization Systems (KOS)

red

Things to do in Week 8, W March 9 - March 23

	Review answer keys. None ☐ required ○ optional	√
	Assignments due W March 23 ☐ required ○ optional	√
Assignment(s)	10 Indexing of three documents (2 hours) (assigned. `Mar. 9)	☐
	Midterm exam (distributed. `Mar. 9)	☐

	New topics this week	

	8.1 Explorations in subject access (Strand 1. General KOS principles and Strand 2. Analysis of highly used KOS)	
Readings	1 Lecture 8.1 Objectives etc. (pink)	☐
	2 Read / look over all pages for Lecture 8.1a in the Lecture notes starting p. ~291 (tasks, alpha index, LCSH, LCC)	☐
Lecture	**Lecture 8.1 slides** (120 min)	☐

	8.2a KOS functions (Strand 1. General KOS principles)	
Readings	1 Lecture 8.2a Objectives etc. (pink). Also have a look at Assignment 11.	☐
	2 Textbook Ch. 13 **Index language functions**	☐
	3 Mooers, Calvin; Brenner, Claude W. **A case history of a Zatocoding IR system**. In: Punched Cards: Their Applications to Science and History, 2nd ed., R. S. Casey; J. W. Perry; M. M. Berry; and A .Kent, eds., New York: Reinhold, 1958. p. 346-352. This is the **seminal article on request-oriented indexing** (called *filtering technique* in the article. Mooers coined the terms *information retrieval* and *descriptor*	☐
	4 Soergel, **Functions of a thesaurus / classification / ontological KB**. Part of Reading 5.	☐
	5 Soergel, Dagobert. **Knowledge Organization Systems. Overview.** 2009.	○
Lecture	**Lecture 8.2a slides** (40 min)	☐

	8.2b Vocabulary control and lexical relationships (Strand 1. General KOS principles)	
Readings	1 Lecture 8.2b Objectives etc. (pink)	☐
	2 Textbook Ch. 12 **Terminological control**	☐
Lecture	**Lecture 8.2b slides** (20 min)	☐

over

Learning blog	Learning blog Week 8 due W March 23	○

	Assignments assigned W March 23	
Assignments assigned	▸**Assignment 12.1 Semantic factoring** (1.5 hrs) (assigned `Apr. 1, due `Apr. 6) ▸**Assignment 12.2 Building a hierarchy of elemental concepts** (1.5 hours) (assigned `Mar. 30, due `Apr. 6)	

pink

Part 5

March 23 - April 27

Knowledge Organization Systems (KOS). Classification and subject access

Learning objectives`	**0 Understand the application of general information structure principles to the subject cataloging/ indexing/ coding of docu-ments and other entities.** Understand the issues in creating and using a good classification / index language / coding scheme This is a general heading encompassing 1 and 2 (P2.3.7)	
	1 Understand the problems and principles of vocabulary control / authority control and be able to apply these principles to indexing and searching (P2.3.8)	L8.1 L8.2a
	1.1 **Understand the retrieval and communication problems caused by variety and ambiguity in language** – synonymy, homonymy, polysemy – including any kind of names, such as names of organizations. (P2.3.8,1)	L8.1 L8.2a
	1.2 **Understand and be able to apply vocabulary control** to remedy these problems, either through using a controlled vocabulary **in indexing** or – through query term expansion – **in searching**. . (P2.3.8,2)	L8.1 L8.2a
	1.3 Understand the structure of a thesaurus with its **synonym-homonym structure** (all terms), **classificatory structure** (concepts expressed by preferred terms), **index language** (concepts and corresponding preferred terms selected as subject descriptors), and **lead-in vocabulary** (all terms that are not subject descriptors). . (P2.3.8,3)	L8.1 L8.2a
	2 Understand the principles of classification and indexing and be able to apply these principles for the benefit of users (P2.3.9)	L8.1 see specific
	2.1 Understand the **pervasive role of classification** throughout the human endeavor. (P2.3.9,1)	L2.1 L2.2 L8.1+
	2.2 Understand the **functions of classification (more broadly, Knowledge Organization Systems, KOS) for a wide variety of tasks in a wide variety of systems.** (P2.3.9,2)	L8.1 L8.2b L10.2a +
	2.2.1 Understand the **use of classification** to support learning. (P2.3.9,2.1)	L8.2b
	2.2.2 Understand the **use of classification** to help users get oriented in a subject and clarify their information needs. (P2.3.9,2.2)	L8.2b

	2.2.3 Understand the **use of classification for** request-oriented indexing and inclusive searching. (P2.3.9,2.3)	L8.2b
	2.3 Understand the principles of the **structure of subject classification**, in particular facet organization and hierarchy and of methods for presenting this structure, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. (P2.3.9,3)	Part 5
	2.3.1 Understand the nature of hierarchical relationships among concepts. (P2.3.9,3.1)	L8.1 L9.1
	2.3.2 Understand the principle of concept componentiality: Elemental concepts can be combined into compound concepts and compound concepts can be analyzed (often non-obviously) into their elemental components resulting in a concept formula (semantic factoring, facet analysis). (P2.3.9,3.2)	L8.1 L9.1
	2.3.3 Be able to use the entity-relationship approach, specifically facet analysis, to discern the conceptual structure of a subject domain. **Be able to discern the facet structure of a subject domain**. (There are facets everywhere.) (P2.3.9,3.3)	L8.1 L9.1
	2.3.3.1 Understand the discernment of abstract concepts that apply across subject disciplines through semantic factoring. (P2.3.9,3.3.1)	L9.1
	2.3.4 Understand inclusive (hierarchically expanded) searching and be able apply inclusive searching in any system. (P2.3.9,3.4)	L9.2
	2.3.5 Understand how the relationship between two compound concepts can be inferred by comparing the concept formulas. Understand the complex hierarchies that result from combining hierarchically structured facets. Understand compound concepts as nodes in a semantic network. Be able to apply this understanding to broadening or narrowing query formulations and to the analysis of classification systems such as DDC, LCC, and subject heading systems, such as LCSH. (P2.3.9,3.5)	L10.1 L12.1
	2.3.6 Solidify understanding of postcombination and precombination - more generally, the degree of precombination — and how they relate to the retrieval mechanism used. Understand the nature of precombined descriptors as new nodes in a semantic network. (P2.3.9,3.6)	L12.1
	2.3.7 Understand the effect of precombination on index language structure and searching and be able to apply this understanding to the analysis of classification schemes such as DDC and LCC and improved searching with such schemes. (P2.3.9,3.7)	L12.1

	2.3.8 Be able to match the index language structure to the database organization and search mechanism available. (P2.3.9,3.8)	L12.1
	2.3.9 Understand the access mechanisms that help a user find the proper descriptors in a large classification scheme with many precombined descriptors, in particular **cross-references** and a **descriptor-find index**. (P2.3.9,3.9)	L12.1
	2.3.10 Understand principles of meaningful arrangement of search results. (P2.3.9,3.10)	L12.1
	2.4 **Know and understand actual schemes**. Get an overview of the wide range of different types of classification schemes and other Knowledge Organization Systems (KOS) and their uses. Grasp the structure of these schemes by applying the general conceptual framework developed earlier in the course to their analysis. (P2.3.9,4)	L8.1 L10.2a L10.2b L11.1 L11.2 L13.1 L13.2a
	2.4.1 Be acquainted with major KOS used on the Web or in American libraries, such as Yahoo (or DMOZ) Classification, Dewey Decimal Classification, Library of Congress Classification, Library of Congress Subject Headings ERIC Thesaurus (P2.3.9,4.1)	L10.2a L10.2b L11.1 L11.2 L13.2a
	2.4.2 Have started to learn to use some KOS used in American libraries for cataloging (indexing) and query formulation for searching. (P2.3.9,4.2)	L10.2a L10.2b L11.1-2 L13.2a
	2.4.3 Be cognizant of the wide range of different types of classification schemes and other Knowledge Organization Systems (KOS) and their many uses. Be acquainted with some important example KOS. (P2.3.9,4.3)	L13.1
	2.5 **Understand indexing characteristics and their effect on system performance** and be able to apply this understanding to the analysis and design of databases and to database selection and query formulation. (P2.3.9,5)	L13.2b
	2.5.1 Understand the concepts of **exhaustivity and specificity of indexing** and their effect on searching. (P2.3.9,5.1)	L13.2b
	2.5.2 Be able to ascertain the exhaustivity and specificity of indexing in a given system and apply this knowledge to appropriate query formulation. (P2.3.9,5.2)	L13.2b

	2.5.3	Understand the concepts of weights in indexing and query formulation. (P2.3.9,5.3)	L13.2b
	2.5.4	Be able to apply indexing weights in query formulation (including analogous techniques in free-text searching). (P2.3.9,5,4)	L13.2b
	2.5.5	Be able to determine the proper levels of exhaustivity and specificity of indexing for a new IR system based on user requirements. (P2.3.9,5.5)	L13.2b
	2.6	Be able to extend classification and indexing principles to **entity types other than Subject**, for example to a hierarchy of organizations and organizational units. (P2.3.9,6)	L8.1

Practical significance	• The practical significance of vocabulary control in indexing and, more importantly, in free-text searching is detailed in Lecture 8.1. • The multiple and pervasive uses of classification have been detailed in the reading for Lecture 8.1. Also remember Lectures 2.1-2.2, The nature of knowledge and knowledge representation. • For IR systems specifically: The index language – the set of subject descriptors used in an IR system and their interrelationships – underlies all activities in subject retrieval. Understanding index language functions and structure - facet structure and hierarchy - is, therefore, at the heart of understanding IR systems.	

pink **Lecture 8.1a Interactive** ***March 23***

Explorations in subject access (based on Assignment 10) (120 min)

Read p. ~291-293 beforehand and have a look at the rough alpha index

<table>
<tr><td>Learning objectives</td><td>All Learning objectives of Part 5 (p. ~285 - 288) are addressed in this exploration

1 Through your own analysis and discussion, arrive at an appreciation of the complexities of subject access and identify the major problems. You do this by working through realistic examples.

2 This practical experience and problem awareness form the basis for the treatment of solutions at a more theoretical level in lectures and readings.

Note: We have not yet discussed nor have you read about what solutions might exist for these problems. The whole point of this interactive exercise is for you to think on your own and figure out solutions to subject access problems yourself. The rest of the course will address each problem in turn.</td></tr>
<tr><td>Tasks</td><td>There are 5 tasks, all exploring problems of the index language in an IR system.

Tasks a - c deal with a sample collection of transportation documents constructed for this assignment.

Task a Formulate queries to search the rough alphabetical index

Task b Design a good alphabetical index

Task c Establish an index language for a computerized IR system (no more than 100 descriptors)

Tasks d and e deal with the Library of Congress Subject Headings (LCSH) and the Library of Congress Classification (LCC), respectively.

Task d From the Library of Congress Subject Headings (LCSH) extract a list of headings dealing with Transportation and Traffic.

Task e From the Library of Congress Classification (LCC) extract a list of classes dealing with Transportation and Traffic.</td></tr>
<tr><td>Materials (attached)</td><td>• Rough alphabetical subject index to the sample collection
• Excerpts from the Library of Congress Subject Headings (LCSH)
• Excerpts from the Library of Congress Classification (LCC)

Note: LCSH and LCC are two different systems for subject access.

The Library of Congress Subject Headings: an alphabetical list of descriptors (subject headings) used in online searching or to arrange a card catalog.

The Library of Congress Classification: a systematic arrangement of descriptors for all areas of knowledge; it is used to arrange books on shelves or links in a Web subject directory.

At the Library of Congress, one cataloger assigns one or more LC subject headings to a book, another the one LC class that best fits.</td></tr>
</table>

In this interactive exercise emphasis is on identifying the problem and possible solutions. There is not enough time to actually implement a solution. We will do just enough practical work to understand the problem.

Tasks a-c: **Dealing with the sample collection**

Tasks a-c deal with subject access problems in a sample collection created by other students assigning subject descriptors to about 200 documents (six for each student) without any guidance whatsoever, as you did in Assignment 10. The rough alphabetical index is simply an alphabetized list of the terms assigned in this "instructionless" indexing, with document numbers following each term. (Many terms have been culled to make the index shorter.)

Task a: **Identify the terms that should be used in the rough alphabetical index to search for the following queries** (no more than 30 terms for each query): (15 min.)

1. Harbors for large tankers.
2. Air cushion craft.
3. The consequences of the development of new types of vehicles for terminal design.
4. Simulation of passenger flow over transportation networks

Task b: **Design a good alphabetical index** to the sample collection (20 min). Your experience with Task a revealed problems with the rough alphabetical index. How would you design an alphabetical index that would address these problems and make searching easier? Start making some revisions to the index, just enough to get some experience that enables you to address the following points:

- What should a good alphabetical index look like?
- How would you go about transforming the rough index into a good index?

Task c: Establish an **index language** for a computerized IR system in the field of transportation (35 min). Assume that the list of terms in the rough alphabetical index is representative of the topics to be searched. So that users can remember all descriptors, the index language is limited to 100 descriptors (not counting place names, they are "free"), yet must allow searching for most of the concepts expressed in the rough alphabetical index without loss of specificity. Remember from Chapter 11 that in a computerized IR system it easy to search with Boolean AND (good manipulative power). See whether you can come up with an idea on the nature of such an index language and apply it to a few examples.

Result of the discussion: A sketch of what the index language might look like

Tasks d-e: **Dealing with existing systems used in American libraries.**

The Library of Congress, and many libraries in the US and around the world (esp. academic and research libraries), use **two separate schemes for subject access**.

1 The first is an alphabetical list of subject terms, the **Library of Congress Subject Headings (LCSH)**. These are used to index books, originally for an alphabetical subject catalog on cards and now for search in an **O**nline **P**ublic **A**ccess **C**atalog (OPAC). Usually several subject headings are assigned to a book to provide multiple access points. There are about half a million subject headings, listed in five large red volumes (to be found in Baldy 14A, older editions are fine).

2 The second is a systematically arranged scheme of subject classes, the **Library of Congress Classification (LCC)**; each class is identified by a class number marking its place in the classified arrangement, for example, BJ 2139 *Etiquette for airplane travel*. All books on this topic are shelved together at BJ 2139. LCC is used for the systematic subject arrangements of books on the shelves. Since books are customarily shelved in only one place (even if there are multiple copies), only one LCC class is assigned to a book, providing only one access point. For example, consider a book titled *The history of State Street in Boston, 1870-1930*. This book could be classed under F73.5 *History of Boston* > *1865-1950* or under F73.67.P3 *History of Boston* > *Streets. Bridges. Railroads* > *Park Street*, but the cataloger has to choose one of these. The 400,000 LCC classes are listed in 30 volumes of *classification schedules* (McKeldin Library Government Documents SU Docs LC 26.9 and McKeldin Reference Z696.U5). XXX Check UB location

The task is assembling a list of headings from LCSH and a list of class numbers from LCC that deal with transportation and traffic. These lists can serve the following purposes:

- A list of subject headings or a classification, respectively, for a transportation library.
- A query formulation to search the OCLC WorldCat database every week for new records in the area of transportation as the basis for book selection in a transportation library. All elements of your list would be connected by OR and the resulting query formulation would retrieve all items on transportation either based on the subject headings or based on the class numbers assigned.
- An aid to a user who wants to search a general catalog for transportation topics. Such a user can find the appropriate subject heading(s) or class numbers much more quickly in a selected list than in the full LCSH or full LCC.

The task is the same as that of any user approaching the subject catalog or the shelves with a question, only magnified by the breadth of your topic.

You can get a feel for both schemes from the sample pages in the assignment materials, but you should also look at the actual schemes. Look at the LCC Outline (a thin separate volume).

Task d: Think about what you would do to put together a list of LCSH headings dealing with **Transportation and traffic**. (25 min) (Start with *Transportation*; you might also try *Ships*, *Railroads*, and other broad terms. Explore from there.)

A list of 15 relevant subject headings, at least 5 of which are not simply taken from the cross-references listed under a very broad heading such as *Transportation*, *Ships*, *Railroads*, or *Air transport*. (Cross-references are the Broader Term (BT), Narrower Term (NT), and Related Term (RT) cross-references given in LCSH, as well as the USE and UF cross-references, which are not of interest here.) You may not include subject headings that start with either *transportation* or *traffic*.

Task e: Think about what you would do to put together a list of classes dealing with **Transportation and traffic** (25 min). Can you restrict your efforts to one or two volumes (After all, a classification is supposed to bring all related subjects together)? Why not?

In a given branch of the hierarchy always list the broadest class that falls under transportation, for example, do not list *TF840-851 Technology > Railroads > Elevated railways and subways*, but go up to the broader level still included in transportation, *TF XXX Technology > Railroads*.

Rough alpha index

based on "instructionless" indexing of a sample collection on transportation by students in previous classes

For some terms, the document number got lost but the terms are still important

Access criteria **Airport**

Access criteria 182
Access study
Accessibility 92
Accommodations 100
Advance acquisition 92
Aerial camera system 160
Aerial car transit(act) 195
Aerial photography 160, 179
Aero engines 188
Aerodynamic 79
Aerodynamics 103, 379
Aerodynamics improvements 145
Aesthetics 92
Air cargo 218, 376, 76
Air cargo flow 127
Air cargo traffic 127
Air cushion craft 109
Air cushion vehicles 150, 188
Air flight paths 337
Air force 155
Air freight 127, 128
Air freight directory 385
Air freight statistics 385
Air passengers 200
Air pressure 89
Air resistance 83
Air rights 185
Air shipping statistics 84
Air terminal 16
Air traffic 16, 162, 316, 346
Air traffic control 13, 16, 67, 133, 325, 346, 367, 372
Air traffic control -SST 194
Air traffic routes 342
Air traffic automation 72
Air traffic -systems and methods 72
Air transport 65, 204
Air transportation 29, 45, 91, 93, 110, 372, 373, 374
Air transportation policy 13
Airbus 91
Aircraft 58, 84, 85, 103, 118, 138, 163,211,354,384
Aircraft accommodation 204
Aircraft design 199, 358
Aircraft development 169
Aircraft developments 85
Aircraft engineering 193
Aircraft in Japan 46
Aircraft maintenance 98
Aircraft marketing research 169
Aircraft navigation 358
Aircraft production 98
Aircraft tests 54
Aircraft-wingless 163
Aircraft--design and construction 47
Aircraft--future 46, 47
Airfield master plans 215, 219
Airline fares 73
Airline flight paths 37
Airline networks 37
Airline operations 113
Airline parking, new plans 325
Airline passenger models 337
Airline passenger transportation 37
Airline rates 143
Airline route maps 93
Airline schedules 73
Airline schedules and fares 373, 374
Airline scheduling 37
Airlines 143
Airlines-timetables, maps, schedules 74
Airplane design 128
Airplanes-passenger information 73
Airport 162

Airport access **Commercial aircraft survey**

Freight stations **International passenger transport**

International railroads **Pipelines**

Library of Congress Subject Headings

Air passenger fares
Airlines—Fares
— — **Law and legislation**
(May Subd Geog)
— **Reservation systems** *(May Subd Geog)*
UF Aeronautics, Commercial—Passenger traffic—Reservations
Airline reservation systems
Airlines—Reservations *[Former heading]*
BT Reservation systems
NT Apollo (Computer system)
SABRE (Computer system)
— Reservations
USE Airlines—Reservation systems
— Stewardesses
USE Flight attendants
— Stewards
USE Flight attendants
— Strikes and lockouts
USE Strikes and lockouts—Airlines
— Time-tables
USE Airlines—Timetables
— **Timetables**
UF Aeronautics, Commercial—Timetables
Airlines—Time-tables *[Former heading]*
Airlines, Local service
USE Local service airlines
Airmen *(May Subd Geog)*
UF Air force personnel
BT Soldiers
Airmobile operations (Military science)
BT Aeronautics, Military
Tactics
Transportation, Military
Airparks
USE Airports
Airparks, Residential
USE Residential airparks
Airphotos
USE Aerial photographs
Airplane accessories
USE Airplanes—Equipment and supplies
Airplane accidents
USE Aircraft accidents
Airplane accidents and drinking
USE Drinking and airplane accidents
Airplane ambulances *(May Subd Geog)*
[RA996.5 (Ambulance service)]
[TL722.8 (Aeronautical engineering)]
UF Air ambulances
BT Aeronautics in medicine
Ambulances
RT Aviation medicine
NT Aeronautics—Relief service
— **Law and legislation** *(May Subd Geog)*
Airplane carriers
USE Aircraft carriers
Airplane child restriant systems
USE Child restraint systems in aircraft
Airplane collisions
USE Aircraft accidents
Airplanes—Collision avoidance
Airplane crashes
USE Aircraft accidents
Airplane engines
USE Airplanes—Motors
Airplane equipment industry
USE Aircraft supplies industry
Airplane exhaust emissions
USE Aircraft exhaust emissions

UF Aircraft industry—Automation
Airplane fuels
USE Airplanes—Fuel
Airplane industry
USE Aircraft industry
Airplane landing mats
USE Landing mats
Airplane parts
USE Airplanes—Parts
Airplane pressurization
USE Airplanes—Pressurization
Airplane racing *(May Subd Geog)*
[GV759]
UF Airplanes—Racing
BT Aeronautical sports
Aeronautics—Competitions
Racing
NT Gordon Bennett Cup Race
Model airplane racing
— **Nevada**
NT National Championship Air Races, Reno, Nev.
Airplane recognition
USE Airplanes—Recognition
Airplane seats
USE Airplanes—Seats
Airplane size
USE Airplanes—Sizes
Airplane sounds
UF Airplanes—Sounds
BT Sounds
RT Airplanes—Noise
NT Fighter plane sounds
Jet plane sounds
Airplane speed records
USE Airplanes—Speed records
Airplane supplies industry
USE Aircraft supplies industry
Airplane theft
USE Aircraft theft
Airplane workers
USE Aircraft industry workers
Airplanes *(May Subd Geog)*
[TL670-TL723]
UF Aeroplanes *[Former heading]*
Aircraft, Fixed wing
Fixed wing aircraft
Planes (Airplanes)
BT Aeronautics
RT Aircraft industry
Flying-machines
SA *headings beginning with the word* Airplane
NT Aeronca airplanes
Air mail service
Airspeed airplanes
Airtankers (Forest fire control)
Amphibian planes
Armstrong Whitworth aircraft
Autogiros
Avro airplanes
BAC airplanes
Beagle airplanes
Bellanca airplanes
Blackburn aircraft
Boeing airplanes
Brewster airplanes
Bristol airplanes
British Aerospace airplanes
Bücker aircraft
Caproni airplanes
Convair airplanes
Convertiplanes
Curtiss aircraft

Fokker airplanes
Gliders (Aeronautics)
Gloster aircraft
Government aircraft
Grönland-wal (Airplane)
Grumman airplanes
GWF airplanes
Handley Page airplanes
Hawker airplanes
Heinkel airplanes
Helicopters
HFB airplanes
Hopfner aircraft
Hypersonic planes
Il'iushin airplanes
Jet planes
Jet transports
Junkers airplanes
Kawasaki airplanes
Lockheed airplanes
McDonnell airplanes
Messerschmitt 108 (Airplane)
Messerschmitt airplanes
Miles airplanes
Mitsubishi airplanes
Mooney airplanes
Multiplanes
Nieuport aircraft
North American airplanes (Military aircraft)
Northrop aircraft
Nuclear aircraft
Ōka (Airplane)
Parnall aircraft
Photography of airplanes
Piggyback aircraft
Piper airplanes
Pitts aircraft
Private planes
Propeller-driven aircraft
Research aircraft
Rocket planes
Rockwell airplanes (Civil aircraft)
Ryan airplanes
Saab aircraft
Savoia-Marchetti airplanes
Seaplanes
Short airplanes
Short take-off and landing aircraft
Sikorsky aircraft
Sopwith airplanes
Stearman airplanes
Supersonic planes
Taildragger airplanes
Training planes
Transonic planes
Tupolev aircraft
Twin-engine airplanes
Ultralight aircraft
Used aircraft
Vertically rising aircraft
Vickers aircraft
Voyager (Airplane)
Vultee airplanes
Waco airplanes
Welch airplanes
Westland airplanes
Wight aircraft
Wright Flyer (Airplane)
— Accidents
USE Aircraft accidents
— Aerodynamics
USE Aerodynamics
— Ailerons

 LCSH-1

Coagulation factors
USE Blood coagulation factors
Coagulation in water purification
USE Water—Purification—Coagulation
Coagulation of blood
USE Blood—Coagulation
Coagulopathies
USE Blood coagulation disorders
Coahuila Indians
USE Cahuilla Indians
Coahuila language
USE Cahuilla language
Coahuilan box turtle
USE Terrapene coahuila
Coahuiltec languages
USE Pakawan languages
Coahuiltecan Indians *(May Subd Geog)*
[E99.C834]
UF Coahuilteco Indians
BT Indians of Mexico
Indians of North America—Texas
NT Payaya Indians
— Names
USE Names, Coahuilteco
Coahuiltecan language
USE Coahuilteco language
Coahuiltecan languages
USE Pakawan languages
Coahuilteco Indians
USE Coahuiltecan Indians
Coahuilteco language *(May Subd Geog)*
[PM3681]
UF Coahuiltecan language
Pajalate language
Tejano language
Texas language
BT Mexico—Languages
Pakawan languages
Texas—Languages
— Names
USE Names, Coahuilteco
Coahuilteco names
USE Names, Coahuilteco
Coaiker Indians
USE Cuaiquer Indians
Coaiker language
USE Cuaiquer language
Coaiquer Indians
USE Cuaiquer Indians
Coaiquer language
USE Cuaiquer language
Coaker family
USE Coker family
Coakley family *(Not Subd Geog)*
UF Cokeley family
Cokely family
Coal *(May Subd Geog)*
[HD9540-HD9559 (Economics)]
[TN799.9-TN834 (Mineral industries)]
UF Coal lands
BT Caustobioliths
Fossil fuels
NT Anthracite coal
Bituminous coal
Cannel coal
Carbonization
Char
Coal reserves
Coaling
Coke
Electric power-plants—Conversion to coal
Furnaces—Conversion to coal
Lignite
Maceral
Sapropelites
Torbanite
— Analysis
[TP325-TP329]
— Carbonization
UF Coking
Low temperature carbonization of coal
BT Carbonization
Coal—Testing
Coke
— — Pilot plants *(May Subd Geog)*
BT Pilot plants
— Chlorine content *(May Subd Geog)*
BT Chlorine
— Combustion
NT Combined combustion of coal and gas
— — Law and legislation
(May Subd Geog)
— Composition
NT Coal—Mineral inclusions
— Desulphurization *(May Subd Geog)*
— — Law and legislation
(May Subd Geog)
— Gasification
USE Coal gasification
— Geology *(May Subd Geog)*
UF Coal geology
BT Geology
— Liquefaction
USE Coal liquefaction
— Magnetic separation *(May Subd Geog)*
BT Magnetic separation
— Methane content
USE Coalbed methane
— Microbiology *(May Subd Geog)*
[QR53.5.C73]
— Mineral inclusions *(May Subd Geog)*
UF Inclusions of minerals in coal
Mineral impurities in coal
Mineral inclusions in coal
Mineral matter in coal
BT Coal—Composition
— Optical properties
— Organic constituents
USE Maceral
— Pipe lines
USE Coal slurry pipelines
— Research *(May Subd Geog)*
UF Coal research
[Former heading]
— — Law and legislation
(May Subd Geog)
— Research grants *(May Subd Geog)*
— — Law and legislation
(May Subd Geog)
— Reserves
USE Coal reserves
— Spectra
— Storage *(May Subd Geog)*
— Sulphur content *(May Subd Geog)*
BT Sulphur
— Taxation *(May Subd Geog)*
UF Coal taxation
Coal trade—Taxation
— — Law and legislation
(May Subd Geog)
— Testing
[TP325-TP329]
NT Coal—Carbonization
— Thermal properties *(May Subd Geog)*
— Transportation *(May Subd Geog)*
[HE199.5.C6 (General)]
[HE595.C6 (Ships)]
[HE2321.C6 (Railroads)]
NT Coal slurry pipelines
— — Law and legislation
(May Subd Geog)
— Weathering
USE Coal-weathering
— Montana
NT Fort Union Coal Region (Mont.)
Coal, Pulverized *(May Subd Geog)*
[TP328]
UF Pulverized coal
NT Fuel, Colloidal
Coal apple
USE Coal balls
Coal ash *(May Subd Geog)*
UF Ash, Coal
Ashes, Coal
NT Fly ash
Coal ash dumps
USE Coal ash sites
Coal ash landfills
USE Coal ash sites
Coal ash sites *(May Subd Geog)*
UF Ash sites, Coal
Coal ash dumps
Coal ash landfills
Coal ash storage facilities
Dumps, Coal ash
Landfills, Coal ash
Sites, Coal ash
Storage facilities, Coal ash
BT Ash disposal
Sanitary landfills
— Maryland
NT Faulkner Coal Ash Storage Facility (Md.)
Coal ash storage facilities
USE Coal ash sites
Coal augers *(May Subd Geog)*
UF Auger-type continuous miners
BT Augers
Coal-mining machinery
Coal balls *(May Subd Geog)*
[QE990]
UF Balls, Coal
Coal apple
Permineralized peat
BT Concretions
Coal-burning furnaces
USE Coal-fired furnaces
Coal-burning power plants
USE Coal-fired power plants
Coal Canyon Wilderness (Utah)
UF Coal Canyon Wilderness Study Area (Utah)
BT National parks and reserves—Utah
Wilderness areas—Utah
Coal Canyon Wilderness Study Area (Utah)
USE Coal Canyon Wilderness (Utah)
Coal-carrying vessels *(May Subd Geog)*
[VM401]
BT Bulk carrier cargo ships
Cargo ships
Merchant ships
Ships
Coal-cutting bits *(May Subd Geog)*
[TN813-TN814]
BT Bits (Drilling and boring)
Coal-mining machinery
Coal export terminals
USE Coal shipping terminals
Coal family
USE Cole family
Coal-fired furnaces *(May Subd Geog)*
UF Coal-burning furnaces
BT Furnaces
Coal-fired power plants *(May Subd Geog)*
UF Coal-burning power plants
Power plants, Coal-fired
BT Fossil fuel power plants
NT Electric power-plants—Conversion to coal
Geothermal-coal hybrid power plants
— Fuel
BT Fuel
— Italy
NT Centrale di Pietrafitta (Italy)
— Maryland
NT Morgantown Generating Plant (Md.)
Coal-gas
USE Gas

LCSH-2

Electric power-plants
— Germany *(Continued)*
NT Kraftwerk Neuhof (Hamburg, Germany)
— Maryland
NT Chalk Point Generating Station (Md.)
Electric power-plants, Offshore
UF Offshore electric power plants
BT Offshore structures
NT Offshore nuclear power plants
Electric power-plants, Portable
UF Portable electric power-plants
Electric power-plants, Underground
USE Underground electric power plants
Electric power pooling
USE Interconnected electric utility systems
Electric power production *(May Subd Geog)*
[TK1001-TK1831]
UF Power production, Electric
BT Electric engineering
Electric power systems
Electrification
NT Cogeneration of electric power and heat
Direct energy conversion
Electric generators
Hydroelectric power plants
Hydrothermal electric power systems
Photovoltaic power generation
Remote area power supply systems
Total energy systems (On-site electric power production)
— Magnetohydrodynamic generation
[TK2970]
UF Magnetohydrodynamic power generation
BT Direct energy conversion
Magnetohydrodynamics
NT Magnetohydrodynamic generators
— — Law and legislation
(May Subd Geog)
Electric power production from chemical action
[TK2901]
BT Direct energy conversion
NT Fuel cells
Electric power supplies to apparatus
UF Power supplies to apparatus, Electric
BT Electric power
SA *subdivision* Power supply *under apparatus, components, etc., e.g.* Computers—Power supply
NT Electric batteries
Electric current converters
Electric current rectifiers
Electric generators
Emergency power supply
Storage batteries
Switching power supplies
Electric power supply
USE Electric power
Electric power supply, Constant-current
UF Constant-current power supply
Power supply, Constant-current
BT Electric power
Electric power system control
USE Electric power systems—Control
Electric power system stability
[TK1010]
UF Stability of electric power systems
BT Transients (Electricity)
RT Electric power systems—Control
Electric power systems *(May Subd Geog)*
Here are entered works on the complex assemblage of equipment and circuits for generating, transmitting, transforming, and distributing electric energy.
BT Electric engineering
Energy industries
NT Electric power distribution
Electric power-plants
Electric power production
Electric power transmission
Emergency power supply
Interconnected electric utility systems
Pulsed power systems
— Communication systems
NT Electric lines—Carrier transmission
— Control
UF Control of electric power systems
Electric power system control
RT Electric power system stability
— Earthquake effects *(May Subd Geog)*
BT Earthquakes
— Electric losses *(May Subd Geog)*
UF Electric losses in power systems
— Law and legislation *(May Subd Geog)*
— Load dispatching
UF Load dispatching (Electric engineering)
— Management
— Natural disaster effects
(May Subd Geog)
BT Electric power failures
Natural disasters
— Protection *(May Subd Geog)*
NT Protective relays
— State estimation
UF State estimation in electric power systems
BT Estimation theory
— Valuation *(May Subd Geog)*
Electric power transmission
(May Subd Geog)
[TK3001-TK3521]
UF Electricity—Transmission
Power transmission, Electric
BT Electric power systems
RT Electric lines
Electric power distribution
NT Satellite power transmission
— Alternating current
[TK3141-TK3171]
BT Electric currents, Alternating
— Direct current
[TK3111]
UF Direct current power transmission
BT Electric currents, Direct
— Equipment and supplies
NT Electric power transmission equipment industry
— New Mexico
NT Four Corners-Ambrosia-Pajarito 500 kV Transmission Project
Electric power transmission equipment industry *(May Subd Geog)*
[HD9695]
BT Electric power transmission—Equipment and supplies
Electric power wheeling
USE Electricity wheeling
Electric precipitation
USE Electrostatic precipitation
Electric properties
USE *subdivision* Electric properties *under individual materials, e.g.* Metals—Electric properties
Electric properties, Lunar
USE Moon—Electric properties
Electric propulsion
BT Electric motors
Electric propulsion of space vehicles
USE Space vehicles—Electric propulsion systems
Electric prospecting *(May Subd Geog)*
UF Geo-electric prospecting
Prospecting—Electric methods
BT Magnetotelluric prospecting
Prospecting—Geophysical methods
Electric protective apparatus
USE Electric apparatus and appliances—Protection
Electric pumps
USE Pumping machinery, Electric
Electric radiation
[QC483]
UF Radiation, Electric
NT Blackbody radiation
Electric railroad accidents *(May Subd Geog)*
UF Electric railroads—Accidents
[Former heading]
BT Railroad accidents
Electric railroads *(May Subd Geog)*
[HE4201-HE5300 (Street-railroads)]
[HE5351-HE5600 (Interurban railroads)]
[TF855-TF1126 (Engineering)]
For general works only. Works on electric street-railroads in a particular locality are entered under Street-railroads—[local subdivision]
UF Electric street-railroads
Railroads, Electric
BT Electricity in transportation
Public utilities
Railroads
RT Railroads—Electrification
Street-railroads
NT Trolley buses
— Accidents
USE Electric railroad accidents
— Brakes
[TF949.B7]
— Cars
[TF920-TF949]
UF Street-cars
Street-railroads—Cars
Streetcars
BT Railroads—Cars
RT Electric railroads—Rolling-stock
NT Near-side Car (Streetcar)
PCC Car (Streetcar)
Peter Witt Car (Streetcar)
Trolley cars
— — Maintenance and repair
BT Electric railroads—Maintenance and repair
— — Museums *(May Subd Geog)*
BT Railroad museums
— — Sounds
BT Railroad sounds
Sounds
— Clearances
UF Clearances (Electric railroads)
— Communication systems
— Construction
USE Electric railroads—Design and construction
— Cost of construction
USE Electric railroads—Design and construction—Costs
— Design and construction
[TF863-TF912]
UF Electric railroads—Construction
[Former heading]
— — Costs
UF Electric railroads—Cost of construction
[Former heading]
— Electronic equipment
BT Electric railroads—Equipment and supplies
Electronic apparatus and appliances
— Equipment and supplies
[TF920-TF952]
NT Electric railroads—Electronic equipment
— — Catalogs
[TF952]

LCSH-3

Proto-Eskimo-Aleut language
Yupik languages
Eskimo leatherwork *(May Subd Geog)*
UF Eskimos—Leather work *[Former heading]*
Leatherwork, Eskimo
BT Leatherwork
Eskimo literature *(May Subd Geog)*
Here are entered collections of works in two or more Eskimo languages. Collections of works in an individual Eskimo language are entered under the specific literature, e.g. Inuit literature. Collections of works by Eskimo authors in non-Eskimo languages are entered under the specific literature with the subdivision Eskimo authors, e.g. American literature—Eskimo authors.
BT Arctic regions—Literatures
NT Eskimo poetry
Eskimo masks *(May Subd Geog)*
UF Eskimos—Masks *[Former heading]*
Masks, Eskimo
BT Masks—Arctic regions
Eskimo medical personnel
USE Eskimos in medicine
Eskimo music
USE Eskimos—Music
Eskimo mythology *(May Subd Geog)*
UF Mythology, Eskimo
Eskimo newspapers *(May Subd Geog)*
Eskimo painting *(May Subd Geog)*
UF Painting, Eskimo
BT Painting, American
Painting, Canadian
Eskimo periodicals *(May Subd Geog)*
Eskimo philosophy *(May Subd Geog)*
UF Eskimos—Philosophy *[Former heading]*
Philosophy, Eskimo
BT Philosophy, American
Philosophy, Canadian
Eskimo poetry *(May Subd Geog)*
BT Eskimo literature
Eskimo poetry (American)
USE American poetry—Eskimo authors
Eskimo prints
USE Prints, Eskimo
Eskimo proverbs
USE Proverbs, Eskimo
Eskimo sculpture *(May Subd Geog)*
UF Eskimos—Sculpture *[Former heading]*
Sculpture, Eskimo
BT Sculpture, American
Sculpture, Canadian
Sculpture, Greenlandic
Sculpture, Russian
Eskimo women *(May Subd Geog)*
UF Eskimos—Women *[Former heading]*
Women, Eskimo
BT Women—North America
Eskimo youth *(May Subd Geog)*
UF Eskimos—Youth *[Former heading]*
Youth, Eskimo
BT Youth—Arctic regions
Eskimos *(May Subd Geog)*
[E99.E7]
Here are entered works discussing collectively the Inuit peoples and the related Eskimo peoples of southern and western Alaska and adjacent regions of Siberia, or works for which the individual group cannot be identified. Works limited to the indigenous Arctic peoples of Greenland, Canada, and northern Alaska are entered under Inuit.
UF Eskimauan Indians
Esquimaux
BT Arctic peoples
Indians of North America
NT Inuit
Libraries and Eskimos
—Alcohol use *(May Subd Geog)*
——Law and legislation *(May Subd Geog)*
BT Liquor laws
—Amusements
USE Eskimos—Games
—Anthropometry *(May Subd Geog)*
—Antiquities
NT Thule culture
—Art
USE Eskimo art
—Arts
USE Eskimo arts
—Basket making
USE Eskimo baskets
—Boats *(May Subd Geog)*
NT Umiaks
—Business enterprises
USE Eskimo business enterprises
—Cartography
USE Eskimo cartography
—Children
USE Eskimo children
—Costume *(May Subd Geog)*
UF Eskimos—Costume and adornment *[Former heading]*
NT Kamiks
—Costume and adornment
USE Eskimos—Costume
—Dances
USE Eskimo dance
—Dolls
USE Eskimo dolls
—First contact with Europeans *(May Subd Geog)*
BT First contact of aboriginal peoples with Westerners
—Fishing *(May Subd Geog)*
—Food *(May Subd Geog)*
—Games
UF Eskimos—Amusements
Eskimos—Recreations
—Hunting *(May Subd Geog)*
—Implements *(May Subd Geog)*
—Languages
——Writing
UF Eskimos—Writing *[Former heading]*
—Leather work
USE Eskimo leatherwork
—Masks
USE Eskimo masks
—Music
UF Eskimo music
—Philosophy
USE Eskimo philosophy
—Recreations
USE Eskimos—Games
—Sculpture
USE Eskimo sculpture
—Treaties
BT Treaties
—Urban residence *(May Subd Geog)*
BT Urbanization
—Women
USE Eskimo women
—Writing
USE Eskimos—Languages—Writing
—Youth
USE Eskimo youth
—Alaska
NT Aleuts
Kuuvanmiit Eskimos
Nunamiut Eskimos
Pacific Gulf Yupik Eskimos
Ugalakmiut Eskimos
Yupik Eskimos
—Canada, Northern
NT Ahiarmiut Eskimos
—Greenland
NT Ammassalimiut Eskimos
Polar Eskimos
—Northwest Territories
NT Caribou Eskimos
Copper Eskimos
Iglulik Eskimos
Inuvialuit Eskimos
Mackenzie Eskimos
Netsilik Eskimos
Utkuhigjalingmiut Eskimos
—Russia (Federation)
NT Yuit Eskimos
—Yukon Territory
NT Inuvialuit Eskimos
Mackenzie Eskimos
——Antiquities
Eskimos and libraries
USE Libraries and Eskimos
Eskimos in literature *(Not Subd Geog)*
Eskimos in medicine *(May Subd Geog)*
UF Eskimo medical personnel
BT Medical personnel
Eskimos in motion pictures
BT Motion pictures
Eskola family *(Not Subd Geog)*
ESL
USE English language—Study and teaching—Foreign speakers
English language—Textbooks for foreign speakers
Esla River (Spain)
UF Río Esla (Spain)
BT Rivers—Spain
Esler family
USE Essler family
Esley family
USE Easley family
Eslick family
USE Eastlack family
Esmeraldas River (Ecuador)
UF Río Esmeraldas (Ecuador)
BT Rivers—Ecuador
Esmeraldas River Valley (Ecuador)
UF Esmeraldas Valley (Ecuador)
BT Valleys—Ecuador
Esmeraldas Valley (Ecuador)
USE Esmeraldas River Valley (Ecuador)
Eso (African people)
USE Topoke (African people)
Esocesidae
USE Needlefishes
Esocidae *(May Subd Geog)*
[QL638.E7]
UF Pike family (Fishes)
BT Salmoniformes
NT Esox
Esocidae, Fossil *(May Subd Geog)*
[QE852.S2]
BT Salmoniformes, Fossil
ESOL
USE English language—Study and teaching—Foreign speakers
English language—Textbooks for foreign speakers
ESOP (Employee stock option plans)
USE Employee stock options
ESOP (Employee stock ownership plans)
USE Employee ownership
Esophageal atresia
USE Esophagus—Atresia
Esophageal hernia
USE Hiatal hernia
Esophageal motility
USE Esophagus—Motility
Esophageal motility disorders
USE Esophagus—Motility—Disorders
Esophageal motor disorders
USE Esophagus—Motility—Disorders
Esophageal reflux
USE Gastroesophageal reflux
Esophageal speech
BT Speech, Alaryngeal

LCSH-4

Hubbard Creek Reservoir (Tex.)
USE Hubbard Creek Lake (Tex.)
Hubbard Creek Watershed (Tex.)
BT Watersheds—Texas
Hubbard family *(Not Subd Geog)*
UF Hubbart family
Hubbert family
Hubberts family
Hubert family
Huebert family
RT Hiebert family
Houpert family
Hubert-Valleroux family
Hubbard Lane (New York, N.Y.)
BT Streets—New York (State)
Hubbard model
UF Model, Hubbard
BT Energy-band theory of solids
Hubbardiidae *(May Subd Geog)*
[QL458.692.H82]
UF Schizomidae *[Former heading]*
BT Schizomida
NT Trithyreus
Hubbardston, Battle of, 1777
USE Hubbardton, Battle of, 1777
Hubbardton, Battle of, 1777
[E241.H8]
UF Hubbardston, Battle of, 1777
BT Burgoyne's Invasion, 1777
United States—History—Revolution, 1775-1783—Campaigns
Vermont—History—Revolution, 1775-1783
Hubbart family
USE Hubbard family
Hubbel family
USE Hubbell family
Hubbell family *(Not Subd Geog)*
UF Hubald family
Huball family
Hubbel family
Hubble family
Hubbold family
Hubel family
Huble family
Hubbell Trading Post National Historic Site (Ganado, Ariz.)
BT Historic sites—Arizona
National parks and reserves—Arizona
Trading posts—Arizona
Hubbert, T. S. (Fictitious character) *(Not Subd Geog)*
UF Hubbert (Fictitious character)
T. S. Hubbert (Fictitious character)
Hubbert (Fictitious character)
USE Hubbert, T. S. (Fictitious character)
Hubbert family
USE Hubbard family
Hubberts family
USE Hubbard family
Hubble family
USE Hubbell family
Hubble-Sandage variables
USE Luminous blue variables
Hubbold family
USE Hubbell family
Hubb's beaked whale *(May Subd Geog)*
[QL737.C438]
UF Mesoplodon carlhubbsi
Hubbs family *(Not Subd Geog)*
Hubel family
USE Hubbell family
Huber family *(Not Subd Geog)*
Hubert family
USE Hubbard family
Hubert-Valleroux family *(Not Subd Geog)*
RT Hubbard family
Huberta (Hippopotamus) *(Not Subd Geog)*
BT Hippopotamus
Hubertus House (Amsterdam, Netherlands)
USE Hubertushuis (Amsterdam, Netherlands)
Hubertushuis (Amsterdam, Netherlands)
UF Hubertus House (Amsterdam, Netherlands)
BT Apartment houses—Netherlands
Single parents—Dwellings—Netherlands
Hubinger family *(Not Subd Geog)*
Huble family
USE Hubbell family
Hubner family
USE Huebner family
Hubnerite
USE Huebnerite
Hubrecht family *(Not Subd Geog)*
Hubrechtella *(May Subd Geog)*
[QL391.N6]
BT Hubrechtidae
Hubrechtidae *(May Subd Geog)*
[QL391.N6]
BT Palaeonemertea
NT Hubrechtella
Hubris (The Greek word)
USE Hybris (The Greek word)
Hubsugul Lake (Mongolia)
USE Hövsgöl Lake (Mongolia)
Hucha de Oro prize
USE Premio Hucha de Oro
Huchen *(May Subd Geog)*
[QL638.S2]
UF Hucho hucho
Huchingson family
USE Hutchinson family
Hucho *(May Subd Geog)*
[QL638.S2]
BT Salmonidae
Hucho hucho
USE Huchen
Huck embroidery
USE Huckaback darning
Huck family *(Not Subd Geog)*
Huckaback
BT Textile fabrics
Huckaback darning
[TT778.H83]
UF Darning, Huckaback
Huck embroidery
BT Embroidery
Huckabay family
USE Huckaby family
Huckabee family
USE Huckaby family
Huckaby family *(Not Subd Geog)*
UF Huckabay family
Huckabee family
Huckerby family
Huckel family *(Not Subd Geog)*
UF Huckle family
Hückel molecular orbitals
UF HMO (Chemistry)
BT Molecular orbitals
Huckerby family
USE Huckaby family
Huckle family
USE Huckel family
Huckleberries *(May Subd Geog)*
[QK495.E58 (Botany)]
[SB386.H83 (Berries)]
BT Vaccinium
RT Cookery (Huckleberries)
Huckleberry Finn (Fictitious character)
USE Finn, Huckleberry (Fictitious character)
Huckstep family *(Not Subd Geog)*
UF Hookstep family
Hooksteppe family
Hucsteppe family
Hucksters
USE Peddlers and peddling
Hucsteppe family
USE Huckstep family
Hucul horse *(May Subd Geog)*
[SF293.H78]
UF Hutsul horse
Hutzul horse
Huzul horse
BT Horse breeds
Huculs
USE Hutsuls
Huculszczyzna (Ukraine)
USE Hutsulshchyna (Ukraine)
Hudak family
USE Hudec family
Huddleson family
USE Huddleston family
Huddleston family *(Not Subd Geog)*
UF Huddleson family
Huddlestone family
Hudelston family
Hudleson family
Hudleston family
Hudliston family
Hudlustun family
Huddlestone family
USE Huddleston family
Hüde 1 Site (Germany)
BT Germany—Antiquities
Hudec family *(Not Subd Geog)*
UF Hudak family
Hudek family
Hudik family
Hudek family
USE Hudec family
Hudelston family
USE Huddleston family
Hudgens family
USE Hutchins family
Hudgin family
USE Hutchins family
Hudgins family
USE Hutchins family
Hudhayl tribe *(May Subd Geog)*
UF Banū Hudhayl
BT Arabs
Hudiburg family
USE Hudiburgh family
Hudiburgh family *(Not Subd Geog)*
UF Heudebourg family
Hudiburg family
Hudik family
USE Hudec family
Hudleson family
USE Huddleston family
Hudleston family
USE Huddleston family
Hudliston family
USE Huddleston family
Hudlow family *(Not Subd Geog)*
Hudlustun family
USE Huddleston family
Hudnut family *(Not Subd Geog)*
Hudon family *(Not Subd Geog)*
Hudson, Lake (Mayes County, Okla.)
UF Lake Hudson (Mayes County, Okla.)
Markham Ferry Reservoir (Okla.)
BT Lakes—Oklahoma
Reservoirs—Oklahoma
Hudson (Bomber) *(Not Subd Geog)*
UF A-29 bomber
BT Bombers
Lockheed airplanes
Hudson (Locomotive) *(Not Subd Geog)*
UF Baltic (Locomotive)
BT Locomotives
Hudson automobile *(Not Subd Geog)*
[TL215.H]
BT American Motors automobiles
Hudson Bay
UF Hudson's Bay
BT Bays—Canada

LCSH-5

Railroad transportation
  USE Railroads
Railroad travel *(May Subd Geog)*
  UF Rail travel
    Railroads—Travel
    Routes of travel
  BT Transportation
    Travel
    Voyages and travels
Railroad travel in literature
  (Not Subd Geog)
Railroad tunnels *(May Subd Geog)*
  [TF230-TF238]
  BT Tunnels
  — Ventilation *(May Subd Geog)*
    [TF235]
    BT Ventilation
  — Colorado
    NT Atlantic and Pacific Tunnel (Colo.)
      Moffat Tunnel (Colo.)
  — England
    NT Channel Tunnel (England and France)
  — France
    NT Channel Tunnel (England and France)
  — Germany
    NT Landrückentunnel (Germany)
  — Germany (West)
  — Japan
    NT Seikan Tonneru (Japan)
  — Pennsylvania
    NT Staple Bend Tunnel (Pa.)
  — Switzerland
    NT Lötschberg-Tunnel (Switzerland)
  — Washington (State)
    NT Cascade Tunnel (Wash.)
Railroad Tycoon (Game)
  [GV1469.25.R34]
  BT Computer games
Railroad vandalism
  USE Railroads—Vandalism
Railroad workers
  USE Railroad construction workers
    Railroads—Employees
Railroad Workers' Strike, Bucharest, Romania, 1933
  BT Strikes and lockouts—Railroads—Romania
Railroad wrecks
  USE Railroad accidents
Railroad yards
  USE Railroads—Yards
Railroads *(May Subd Geog)*
  [HE1001-HE5600 (Economics)]
  [TF (Technology)]
  UF Lines, Railroad
    Rail lines
    Rail transportation
    Railroad lines
    Railroad transportation
    Railways
  BT Communication and traffic
    Concessions
    Public utilities
    Transportation
    Trusts, Industrial
  SA *names of individual railroads; and headings beginning with the word* Railroad
  NT Broad gauge railroads
    Cipher and telegraph codes—Railroads
    Classification—Books—Railroads
    Electric railroads
    Express service
    Horse railroads
    Military railroads
    Mine railroads
    Monorail railroads
    Mountain railroads
    Narrow gauge railroads
    Pacific railroads
    Photography of railroads
    Rack-railroads
    Railroad holding companies
    Railway mail service
    Scenic railways
    Ship-railroads
    Street-railroads
    Subways
    Train robberies
  — Abandonment *(May Subd Geog)*
    UF Abandonment of railroad lines
    BT Railroads and state
  — Accidents
    USE Railroad accidents
  — Air-brakes
    USE Air-brakes
  — Airspace utilization
    UF Airspace over railroads
    BT Railroads—Right of way
  — Automatic speed control
    USE Railroads—Automatic train control
  — Automatic stop
    USE Railroads—Automatic train control
  — Automatic train control
    [TF638]
    UF Automatic train control
      Railroads—Automatic speed control
      Railroads—Automatic stop
    BT Railroads—Electric equipment
      Railroads—Safety appliances
  — Baggage
    USE Railroads—Baggage handling
  — Baggage handling *(May Subd Geog)*
    [HE2556 (Economics)]
    [TF656 (Railroad operations)]
    UF Baggage handling on railroads
      Railroads—Baggage
        [Former heading]
  — Bills of lading
    USE Bills of lading
  — Brakes
    [TF415-TF430]
    UF Railroad brakes
      Railroad car brakes
      Railroads—Cars—Brakes
    BT Brakes
      Railroads—Safety appliances
    NT Locomotives—Brakes
      Skates (Railroads)
  — Branch lines
  — Bridges
    USE Railroad bridges
  — Buildings and structures
    [TF270-TF320]
    NT Hump yards
      Railroad stations
      Railroad terminals
      Railroads—Repair-shops
      Roundhouses (Railroads)
  — — Foundations
      BT Foundations
  — Cabooses
    USE Cabooses (Railroads)
  — Car reporting systems
    [TF606]
    UF Car reporting systems (Railroads)
    BT Railroads—Cars
  — Cars
    [HE1830 (Car service)]
    [TF371-TF499 (Car building)]
    [TF600-TF606 (Operation)]
    UF Carriages, Railway
      Cars, Railroad
      Cars and car building
      Railway carriages
    RT Railroads—Rolling-stock
    NT Cabooses (Railroads)
      Car axles
      Car-couplings
      Car fenders
      Car-springs
      Car trucks (Railroads)
      Car-wheels
      Dump cars
      Electric railroads—Cars
      Handcars
      Railroad motor-cars
      Railroads—Car reporting systems
      Railroads—Dining-car service
      Railroads—Freight-cars
      Railroads—Passenger-cars
      Railroads—Private cars
      Railroads—Track-inspection cars
      Railway mail service—Cars
      Refrigerator cars
      Tank cars
      Work cars (Railroads)
  — — Bearings
  — — Brakes
      USE Railroads—Brakes
  — — Construction
      USE Railroads—Cars—Design and construction
  — — Design and construction
      UF Railroads—Cars—Construction
        [Former heading]
  — — Dynamics
      BT Dynamics
  — — Electric equipment
  — — Fittings
  — — Heating and ventilation
      (May Subd Geog)
      [TF445-TF449]
      UF Car heating
        Car ventilation
  — — Insulation *(May Subd Geog)*
      BT Insulation (Heat)
  — — Lighting *(May Subd Geog)*
      [TF445-TF449]
      UF Car lighting
      BT Electric lighting
        Railroads—Electric equipment
        Railroads—Lighting
  — — Lubrication
  — — Maintenance and repair
      UF Cars and car building
        Railroads—Cars—Repairing
      BT Railroads—Repair-shops
  — — Models *(May Subd Geog)*
  — — Painting
  — — Repairing
      USE Railroads—Cars—Maintenance and repair
  — — Sanitation
      USE Railroads—Sanitation
  — — Vibration *(May Subd Geog)*
  — — Welding *(May Subd Geog)*
  — Centralized traffic control
    USE Railroads—Signaling—Centralized traffic control
  — Cipher and telegraph codes
    USE Cipher and telegraph codes—Railroads
  — Circus trains
    USE Circus trains
  — Claim departments
    [HE1795]
    UF Railroad accidents—Claims
  — Clearances
    UF Clearances (Railroads)
  — Collective bargaining
    USE Collective bargaining—Railroads
  — Collective labor agreements
    USE Collective labor agreements—Railroads
  — Colonies
    USE Railroads, Colonial
  — Communication systems

LCSH-6

Transport planes *(May Subd Geog)*
UF Airliners
Cargo aircraft
Cargo planes
Freight planes
Mail planes
Transport aircraft
BT Aeronautics, Commercial
NT An-26 (Transport plane)
Beagle B-206 (Transport planes)
Beaver (Transport planes)
Boeing 247 (Transport plane)
Britten-Norman Islander (Transport planes)
Canadair DC-4M (Transport plane)
Commuter aircraft
Constellation (Transport planes)
Convair transport planes
Curtiss CW-20 (Transport plane)
Dart Herald (Transport planes)
Douglas DC-3 (Transport plane)
Douglas DC-6 (Transport plane)
Douglas transport planes
Dove (Transport plane)
Ford Three-engined Monoplane (Transport plane)
Heron (Transport plane)
Jet transports
Junkers 290 (Transport plane)
Junkers Ju-52 (Transport plane)
Messerschmitt 323 (Transport planes)
Provider (Transport planes)
Skyvan (Transport plane)
StarLifter (Transport planes)
Supersonic transport planes
Turboprop transports
V-22 Osprey (Transport plane)
Viking (Transport plane)
Vimy (Bomber)
Yukon (Transport planes)
— Decoration *(May Subd Geog)*
BT Decoration and ornament
— Fuel consumption
— Identification marks
USE Transport planes—Markings
— Landing
— Loading and unloading *(May Subd Geog)*
BT Loading and unloading
— Markings *(May Subd Geog)*
UF Identification marks on transport planes
Markings on transport planes
Transport planes—Identification marks *[Former heading]*
— Take-off
Transport proteins
USE Carrier proteins
Transport theory
[QC175.2]
UF Boltzmann transport equation
Transport phenomena
BT Mathematical physics
Particles (Nuclear physics)
Radiation
Statistical mechanics
NT Collision integrals
Electrodiffusion
Electron transport
Energy transfer
Enskog equation
Exciton theory
Galvanomagnetic effects
Ion swarms
Mass transfer
Momentum transfer
Neutron transport theory
Photon transport theory
Radiative transfer
— Carlson method
USE Transport theory—Discrete ordinates method
— Discrete ordinate method
USE Transport theory—Discrete ordinates method
— Discrete ordinates method
UF Carlson method in transport theory
Discrete ordinate method in transport theory
Discrete ordinates method in transport theory
Transport theory—Carlson method
Transport theory—Discrete ordinate method
Transport theory—Wick-Chandrasekhar method
Wick-Chandrasekhar method in transport theory
BT Finite differences
— Wick-Chandrasekhar method
USE Transport theory—Discrete ordinates method
Transport workers *(May Subd Geog)*
[HD8039.T7]
UF Transportation—Employees
Transportation workers
NT Carters
Cheese carriers (Persons)
Highway transport workers
Railroads—Employees
Rickshaw men
Sailors
Stevedores
Street-railroads—Employees
Teamsters
Transit police
Water carriers (Persons)
— Collective labor agreements
USE Collective labor agreements—Transport workers
— Strikes and lockouts
USE Strikes and lockouts—Transport workers
— Trade-unions
USE Trade-unions—Transport workers
— Wages
USE Wages—Transport workers
Transport Workers' Strike, Bulgaria, 1919-1920
BT Strikes and lockouts—Transport workers—Bulgaria
Transport Workers' Strike, Liverpool, England, 1911
USE General Transport Strike, Liverpool, England, 1911
Transport Workers' Strike, New York, N.Y., 1966
UF Transit Strike, New York, N.Y., 1966
BT Strikes and lockouts—Transport workers—New York (State)
Transportable breakwaters
USE Breakwaters, Mobile
Transportation *(May Subd Geog)*
[GT5220 (Manners and customs)]
[HE (General)]
UF Public transportation
Transport
Transportation—Economic aspects
Transportation, Primitive *[Former heading]*
BT Locomotion
RT Commerce
Communication and traffic
Storage and moving trade
SA *subdivision* Transportation *under subjects*
NT Acids—Handling and transportation
Air travel
Bridges
Cab and omnibus service
Carriages and carts
Carriers
Choice of transportation
Commuting
Delivery of goods
Electricity in transportation
Electronics in transportation
Emergency transportation
Employer-sponsored transportation
Express service
Freight and freightage
Harbors
High speed ground transportation
Intra-airport transportation
Local transit
National parks and reserves—Transportation
Ocean travel
Pack transportation
Passes (Transportation)
Pipelines
Pneumatic-tube transportation
Postal service
Railroad travel
Railroads
Roads
Route surveying
Rural transportation
Shipping
Steam-navigation
Terminals (Transportation)
Transit, International
Urban transportation
Vehicles
Waterways
— Advertising
USE Advertising—Transportation
— Climatic factors *(May Subd Geog)*
— Collective bargaining
USE Collective bargaining—Transportation
— Consolidation
USE Transportation—Mergers
— Earthquake effects *(May Subd Geog)*
BT Earthquakes
— Economic aspects
USE Transportation
— Effect of inflation on *(May Subd Geog)*
BT Inflation (Finance)
— Employees
USE Transport workers
— Energy conservation *(May Subd Geog)*
[TJ163.5.T7]
— Equipment and supplies
— — Tariff
USE Tariff on transportation equipment
— Fares *(May Subd Geog)*
UF Fares, Transportation
BT Transportation—Rates
— Federal aid
USE Federal aid to transportation
— Finance
NT Federal aid to transportation
— — Law and legislation *(May Subd Geog)*
— Freight
USE Freight and freightage
— Government policy
USE Transportation and state
— Handbooks, manuals, etc.
USE Shippers' guides
— History
— Industrial capacity *(May Subd Geog)*
— Law and legislation *(May Subd Geog)*
UF Transportation—Laws and regulations *[Former heading]*
Transportation—Laws and regulations, International *[Former heading]*
BT Trade regulation

Transportation
— **Law and legislation** *(Continued)*
— — **Criminal provisions**
— Laws and regulations
USE Transportation—Law and legislation
— Laws and regulations, International
USE Transportation—Law and legislation
— Medical aspects
USE Transportation medicine
— **Mergers** *(May Subd Geog)*
UF Transportation—Consolidation *[Former heading]*
Transportation mergers
BT Consolidation and merger of corporations
— Museums
USE Transportation museums
— Noise
USE Transportation noise
— **Passenger traffic**
UF Passenger traffic
NT Passenger conveyors
Segregation in transportation
— Passes
USE Passes (Transportation)
— **Physiological aspects**
BT Transportation medicine
— **Planning**
UF Transportation planning *[Former heading]*
— Public relations
USE Public relations—Transportation
— **Rates** *(May Subd Geog)*
[HE195.4-HE195.5]
UF Carriers—Rates
NT Canals—Rates and tolls
Shipping conferences
Transportation—Fares
— **Records and correspondence**
— — **Abstracting and indexing** *(May Subd Geog)*
UF Transportation—Records and correspondence—Indexing *[Former heading]*
— — Indexing
USE Transportation—Records and correspondence—Abstracting and indexing
— **Research** *(May Subd Geog)*
NT Local transit—Research
— Route choice
USE Route choice
— **Safety measures**
[HE194-HE194.5]
UF Transportation accidents—Prevention
Transportation safety
— **Security measures** *(May Subd Geog)*
— — **Equipment and supplies**
NT Transportation security equipment industry
— **Social aspects** *(May Subd Geog)*
UF Society and transportation
— **Taxation** *(May Subd Geog)*
UF Transportation tax
NT Freight and freightage—Taxation
— — **Law and legislation** *(May Subd Geog)*
— Time-tables
USE Transportation—Timetables
— **Timetables**
UF Transportation—Time-tables *[Former heading]*
— **Vocational guidance** *(May Subd Geog)*
[TA1160 (Engineering)]
Transportation, Atomic-powered
USE Nuclear propulsion

Transportation, Automotive *(May Subd Geog)*
[HE5601-HE5725]
UF Automotive transportation
Highway transportation
Motor carriers
Motor transportation
Road transportation
RT Automobiles—Social aspects
NT Automobile parking
Automobiles
Automobiles, Company
Bus lines
Buses
Freight-cars on truck trailers
Limousine services
Motor vehicles
Motorization, Military
Piggyback transportation
School children—Transportation
Taxicabs
Traffic safety
Truck terminals
Trucking
Trucks
— Accidents
USE Traffic accidents
— Collective labor agreements
USE Collective labor agreements—Trucking industry
— **Communication systems**
— **Cost of operation**
— **Dispatching**
— Employees
USE Highway transport workers
— Freight
USE Trucking
— Freight classification
USE Trucking—Classification
— **Law and legislation** *(May Subd Geog)*
UF Transportation, Automotive—Laws and regulations *[Former heading]*
Transportation, Automotive—Laws and regulations, International *[Former heading]*
NT Traffic regulations
— Laws and regulations
USE Transportation, Automotive—Law and legislation
— Laws and regulations, International
USE Transportation, Automotive—Law and legislation
— **Rates** *(May Subd Geog)*
UF Motor carrier rates
— **Taxation** *(May Subd Geog)*
UF Highway taxes
Highway user taxes
Road user taxes
— — **Law and legislation** *(May Subd Geog)*
— Traffic control
USE Traffic signs and signals
Transportation, Choice of
USE Choice of transportation
Transportation, Military *(May Subd Geog)*
[UC270-UC360 (General)]
[UH500-UH505 (Medical service)]
[VC550-VC580 (Naval)]
UF Military transportation
Motor vehicles in war
BT Communications, Military
Logistics
Military art and science
RT Stream crossing, Military
SA *subdivisions* Transport of sick and wounded, Transport service *and* Transportation *under names of individual military services, e.g.* United States. Army—Transport of sick and wounded; United States. Army—Transport service; United States. Army—Transportation
NT Airdrop
Airlift, Military
Airmobile operations (Military science)
Hospital trains
Landing craft
Landing operations
Military bridges
Military railroads
Military sealift
Motorization, Military
Pack transportation
Transports
Vehicles, Military
— **Cold weather conditions**
UF Arctic transportation
Cold weather operations (Military)
BT Winter warfare
— **Research** *(May Subd Geog)*
BT Military research
Transportation, Piggyback
USE Piggyback transportation
Transportation, Prehistoric *(May Subd Geog)*
[GN799.T73]
UF Prehistoric transportation
Transportation, Primitive
USE Transportation
Transportation, Rural
USE Rural transportation
Transportation, Urban
USE Urban transportation
Transportation (Punishment)
USE Prisoners, Transportation of
Transportation accidents *(May Subd Geog)*
[HE194-HE194.5]
BT Accidents
NT Aircraft accidents
Railroad accidents
Traffic accidents
— Prevention
USE Transportation—Safety measures
Transportation and state *(May Subd Geog)*
UF State and transportation
Transportation—Government policy
Transportation policy
NT Railroads and state
Urban transportation policy
— **Africa**
NT United Nations Transport and Communications Decade in Africa, 1978-1988
Transportation buildings *(May Subd Geog)*
[NA6290]
BT Buildings
Terminals (Transportation)
NT Airport buildings
Boathouses
Local transit stations
Parking garages
Service stations
Subway stations
Transportation centers
USE Terminals (Transportation)
Transportation consultants *(May Subd Geog)*
BT Consultants
Transportation engineering *(May Subd Geog)*
BT Civil engineering
Engineering
NT Highway engineering
Railroad engineering
Traffic engineering

LCSH-8

— **Awards** *(May Subd Geog)*
— — **United States**
NT Elmer A. Sperry Award
Transportation engineers *(May Subd Geog)*
BT Engineers
Transportation equipment industry *(May Subd Geog)*
[HD9709]
BT Manufactures
NT Bicycle industry
Railroad equipment industry
Traffic control equipment industry
Truck equipment industry
Transportation geography *(May Subd Geog)*
[HE323-HE328]
UF Transport geography
BT Geography
Transportation in art
Transportation in literature *(Not Subd Geog)*
Transportation in numismatics *(May Subd Geog)*
[CJ161.T73]
BT Numismatics
Transportation insurance
USE Insurance, Inland marine
Insurance, Marine
Transportation libraries *(May Subd Geog)*
[Z675.T7]
UF Libraries, Transportation
BT Technical libraries
Transportation markings *(May Subd Geog)*
UF Markings, Transportation
BT Signs and symbols
NT Buoys
Daymarks
Landing aids (Aeronautics)
Road markings
Traffic signs and signals
Transportation medicine *(May Subd Geog)*
[RC1030-RC1035]
UF Medicine, Transportation
Transportation—Medical aspects
NT Automotive medicine
Aviation medicine
Space medicine
Transportation—Physiological aspects
Transportation mergers
USE Transportation—Mergers
Transportation museums *(May Subd Geog)*
[TA1006]
UF Transport museums
Transportation—Museums
Vehicles—Museums
BT Museums
SA *subdivision* Museums *under individual land vehicles, e.g.* Automobiles—Museums
NT Aeronautical museums
Railroad museums
Streetcar museums
Transportation noise *(May Subd Geog)*
[TD893.6.T7]
UF Transportation—Noise
BT Noise
NT Airport noise
Traffic noise
Transportation of animals by air
USE Animals, Air transportation of
Transportation of criminals
USE Penal colonies
Transportation of livestock by rail
USE Railroads—Livestock transportation
Transportation of logs
USE Log transportation
Transportation of merchandise in bond
USE Bonded warehouses and goods
Transportation of prisoners
USE Prisoners, Transportation of
Transportation on postage stamps
BT Postage stamps
Transportation passes
USE Passes (Transportation)
Transportation planning
USE Transportation—Planning
Transportation policy
USE Transportation and state
Transportation radio stations *(May Subd Geog)*
BT Radio stations
Transportation safety
USE Transportation—Safety measures
Transportation security equipment industry *(May Subd Geog)*
[HD9999.S45-HD9999.S454]
BT Transportation—Security measures—Equipment and supplies
Transportation tax
USE Transportation—Taxation
Transportation terminals
USE Terminals (Transportation)
Transportation to airports
USE Access to airports
Transportation to harbors
USE Harbors—Access
Transportation to recreation areas
USE Recreation areas—Access
Transportation workers
USE Transport workers
Transporter-bridges
[TG435]
BT Bridges
Transports *(May Subd Geog)*
[UC325]
UF Troopships
BT Transportation, Military
NT Embarkation (Military science)
Transposable elements
USE Transposons
Transposition (Genetics)
USE Translocation (Genetics)
Transposition (Music)
[MT68]
BT Musical pitch
Transposition of great arteries
USE Transposition of great vessels
Transposition of great vessels *(May Subd Geog)*
[RC687]
UF Great arteries, Transposition of
Great vessels, Transposition of
Heart—Blood-vessels—Transposition
Transposition of great arteries
Transposition of the great arteries
BT Aorta—Abnormalities
Congenital heart disease
Pulmonary artery—Abnormalities
Transposition of the great arteries
USE Transposition of great vessels
Transposons
UF Tn elements [Former heading]
Transposable elements
BT Mobile genetic elements
Transputers
[TK7895.T73]
BT Integrated circuits—Very large scale integration
Microcomputers
Transsexual surgery
USE Sex change
Transsexualism *(May Subd Geog)*
[HQ77.7-HQ77.9 (Sociology)]
[RC560.G45 (Psychiatry)]
UF Transsexuality
BT Gender identity disorders
— Law and legislation
USE Transsexuals—Legal status, laws, etc.
— Patients
USE Transsexuals
Transsexuality
USE Transsexualism
Transsexuals *(May Subd Geog)*
UF Transexuals
Transsexualism—Patients
BT Persons
— **Legal status, laws, etc.** *(May Subd Geog)*
UF Transsexualism—Law and legislation
Transtasman Singlehanded Race
BT Yacht racing
Transtentorial herniation, Traumatic
USE Traumatic tentorial herniation
Transubstantiation
[BX2220 (Catholic Church)]
Here are entered works on the Catholic doctrine that the substance of the elements of bread and wine is transformed into the substance of the body and blood of Christ after receiving priestly consecration in the Mass. Works on the doctrine that Christ is present in the sacrament of the Lord's Supper are entered under Lord's Supper—Real presence.
BT Lord's Supper—Real presence
NT Mass
Transubstantiation in literature *(Not Subd Geog)*
Transudates
USE Exudates and transudates
Transuranic elements
USE Transuranium elements
Transuranic wastes
USE Alpha-bearing wastes
Transuranium elements
UF Transuranic elements
BT Actinide elements
RT Uranium
NT Neptunium
Plutonium
Transplutonium elements
Transuranium elements in the body
Transurethral prostatectomy
USE Prostatectomy, Transurethral
Transurethral resection
USE Prostatectomy, Transurethral
Transvaal (South Africa)
— **Capital and capitol**
NT Raadsaal (Pretoria, South Africa)
— **Description and travel**
UF Transvaal (South Africa)—Description and travel—To 1910 [Former heading]
— — To 1910
USE Transvaal (South Africa)—Description and travel
— General Strike, 1913
USE General Strike, Transvaal, South Africa, 1913
— **History**
— — **To 1880**
— — **War of 1880-1881**
UF Boer War, 1880-1881
Transvaal War, 1880-1881
NT Majuba Hill (South Africa), Battle of, 1881
— — **1880-1910**
NT Jameson's Raid, 1895-1896
— — **1910-**
— **Politics and government**
— — **To 1880**
— — **1880-1910**
— — **1910-**
Transvaal daisy
USE Gerbera
Transvaal Ndebele language
USE Ndebele language (South Africa)
Transvaal Sotho (African people)
USE Pedi (African people)
Transvaal Sotho language
USE Northern Sotho language
Transvaal War, 1880-1881
USE Transvaal (South Africa)—History—War of 1880-1881

 LCSH-9

Trochesset family *(Not Subd Geog)*
UF Trauchessec family
Trochesec family
Trossecet family
Trochilidae
USE Hummingbirds
Trochiliidae (Insects)
USE Clearwing moths
Trochilus colubris
USE Ruby-throated hummingbird
Trochilus rufus
USE Rufous hummingbird
Trochochaetidae *(May Subd Geog)*
[QL391.A6]
UF Disomidae
BT Spionida
Trochodendraceae *(May Subd Geog)*
[QK495.T75]
BT Magnoliales
Trochoids
USE Cycloids
Epicycloids and hypocycloids
Trocholinidae
USE Involutinidae
Trochomorphidae
USE Zonitidae
Trochophore
BT Larvae
Trochosa *(May Subd Geog)*
[QL458.42.L9]
BT Wolf spiders
NT Trochosa ruricola
Trochosa ruricola *(May Subd Geog)*
[QL458.42.L9]
BT Trochosa
Trochotrons *(May Subd Geog)*
[TK7871.79.T7]
BT Vacuum-tubes
Trochu family *(Not Subd Geog)*
Trochus shell fisheries *(May Subd Geog)*
[SH377.5]
BT Mother-of-pearl
Shellfish fisheries
Troclar de Sainte-Sigolène Site (France)
UF Sainte-Sigolène Site, Troclar de (France)
Site du Troclar de Sainte-Sigolène (France)
BT France—Antiquities
Trodos (Cyprus)
USE Troodos Mountains (Cyprus)
Troed family
USE Trued family
Troedsson family
USE Trued family
Troelsen family *(Not Subd Geog)*
Troelsgaard family *(Not Subd Geog)*
Trofast family *(Not Subd Geog)*
Troféo automobile *(Not Subd Geog)*
[TL215.T]
UF Oldsmobile Troféo automobile
BT Oldsmobile automobile
Trogden family
USE Trogdon family
Trogdon family *(Not Subd Geog)*
UF Throgden family
Trogden family
Troghton family
Troglin family
RT Troughton family
Troghton family
USE Trogdon family
Trogidae
USE Scarabaeidae
Troglin family
USE Trogdon family
Troglodytes *(May Subd Geog)*
[GN783-GN784]
BT Cave dwellers
Troglodytidae
USE Wrens
Troglohyphantes *(May Subd Geog)*
[QL458.42.L55]
BT Linyphiidae
Trogoderma *(May Subd Geog)*
[QL596.D4]
BT Dermestidae
Trogoderma granarium
USE Khapra beetle
Trogonidae
USE Trogons
Trogoniformes *(May Subd Geog)*
[QL696.T67]
BT Birds
NT Trogons
Trogonoptera *(May Subd Geog)*
[QL561.P2]
UF Birdwing butterflies
BT Papilionidae
Trogons *(May Subd Geog)*
[QL696.T7]
UF Trogonidae
BT Trogoniformes
Trogositidae *(May Subd Geog)*
[QL596.T8]
UF Bark-gnawing beetles
Lophocateridae
Ostomatidae
Ostomidae
Peltidae
Temnochilidae
Trogossitidae
BT Beetles
Trogossitidae
USE Trogositidae
Troides *(May Subd Geog)*
[QL561.P2]
UF Birdwing butterflies
BT Papilionidae
Troilism
USE Group sex
Troilite *(May Subd Geog)*
BT Sulphide minerals
Troilite ores *(May Subd Geog)*
[TN538.T74]
BT Ores
— Geology *(May Subd Geog)*
[TN538.T74]
BT Geology
Troilus (Greek mythology)
USE Troilus (Legendary character)
Troilus (Legendary character)
(Not Subd Geog)
UF Troilus (Greek mythology)
[Former heading]
BT Folklore—Greece
Folklore—Rome
Troilus (Legendary character) in literature
(Not Subd Geog)
Trois couronnements (Tapestry)
USE Three coronations (Tapestry)
Trois Glorieuses, France, 1830
USE France—History—July Revolution, 1830
Trois Pignons Forest (France)
UF Forêt des Trois Pignons (France)
Forêt domaniale des Trois Pignons (France)
BT Forests and forestry—France
Trois-Rivières River (Québec)
USE Saint Maurice River (Québec)
Troitsa (Icon)
USE Rublev, Andreĭ, Saint, d. ca. 1430. Trinity
Trojan (Training plane)
USE T-28 (Training plane)
Trojan Nuclear Plant (Or.)
USE Trojan Nuclear Power Plant (Or.)
Trojan Nuclear Power Plant (Or.)
UF Trojan Nuclear Plant (Or.)
BT Nuclear power plants—Oregon
Trojan War
BT Mythology, Greek
Troy (Extinct city)—History
Trojans
[DF221.T8]
Troldheimen (Norway)
USE Trollheimen (Norway)
Troll dolls *(May Subd Geog)*
BT Dolls
Trolls
Troll-Maxen (Legendary character)
(Not Subd Geog)
[GR226.J]
Trolle family *(Not Subd Geog)*
Trolley buses *(May Subd Geog)*
[TL232]
UF Trackless trolleys
BT Buses
Buses, Electric
Electric railroads
Street-railroads
— Brakes
— Electric equipment *(May Subd Geog)*
— Law and legislation *(May Subd Geog)*
— Motors
BT Electric motors
— Valuation *(May Subd Geog)*
— Wires and wiring
BT Electric wiring
Trolley car lines
USE Street-railroads
Trolley car museums
USE Streetcar museums
Trolley cars *(May Subd Geog)*
BT Electric railroads—Cars
Trolley-wheels
USE Electric railroads—Trolley-wheels
Trollheimen (Norway)
UF Troldheimen (Norway)
BT Mountains—Norway
Trolling (Fishing) *(May Subd Geog)*
[SH457.7]
BT Fishing
Trolls *(May Subd Geog)*
[GR555]
UF Trulls
BT Fairies
NT Troll dolls
Trolls in art
Tromba marina
USE Trumpet marine
Trombetas River (Brazil)
UF Rio Trombetas (Brazil)
BT Rivers—Brazil
Trombetas River Watershed (Brazil)
BT Watersheds—Brazil
Trombiculidae *(May Subd Geog)*
[QL458.2.T75]
UF Trombiculids
BT Acariformes
Chiggers (Mites)
NT Euschongastia
Leptotrombidium
Pseudoschoengastia
Trombiculids
USE Trombiculidae
Trombidiidae *(May Subd Geog)*
[QL458.2.T76]
UF Thrombidiidae
BT Acariformes
Trombone *(May Subd Geog)*
[ML965-ML968 (History)]
[MT460-MT472 (Instruction)]
BT Brass instruments
NT Bass trombone
Sackbut
— Methods
— — Self-instruction
[MT468]
UF Trombone—Self-instruction
— Orchestra studies

1/SH-10

Library of Congress Classification

Broad Outline (Main classes)

A	General works	
B	Philosophy. Psychology. Religion	
C	Auxiliary sciences of history	
D	History: General and outside the Americas	
E-F	History of America	
	E	History: America General and United States General
	F	History: United States local, Canada, and Latin America
G	Geography	
H	Social sciences	
J	Political science	
K	Law	
L	Education	
M	Music and books on music	
N	Fine arts	
P	Language and literature	
Q	Science	
R	Medicine	
S	Agriculture	
T	Technology	
U	Military science	
V	Naval science	
Z	Bibliography and library science	

The following pages give first a detailed outline and then examples of classes dealing with or relevant to *transportation and traffic*.

Library of Congress Classification. Detailed Outline

A General works

AC Collections. Series. Collected works
AE Encyclopedias (General)
AG Dictionaries and other General reference works
AI Indexes (General)
AM Museums (General). Collectors and collecting (General)
AN Newspapers
AP Periodicals
AS Academies and learned societies (General)
AY Yearbooks. Almanacs. Directories
AZ History of scholarship and learning. The humanities

B Philosophy. Psychology. Religion

B-BJ Philosophy, incl. BF Psychology

B Philosophy (General)
BC Logic
BD Speculative philosophy
BF Psychology. Parapsychology. Occult sciences
BH Aesthetics
BJ Ethics. Social usages. Etiquette

BL-BX Religion

BL Religions. Mythology. Rationalism
BM Judaism
BP Islam. Bahaism. Theosophy
BQ Buddhism

BR-BX Christianity

BR Christianity
BS The Bible
BT Doctrinal theology
BV Practical theology
BX Christian denominations

C Auxiliary sciences of history

C Auxiliary sciences of history (General)
CB History of civilization
CC Archaeology (General)
CD Diplomatics. Archives. Seals
CE Technical chronology. Calendar
CJ Numismatics
CN Inscriptions. Epigraphy
CR Heraldry
CS Genealogy
CT Biography [General]

D History: General and Old World

D History (General). Europe (General)
DA Great Britain
DAW Central Europe
DB Austria, Hungary, Czech Republic, Slovakia
DC France
DD Germany
DE Mediterranean region. Greco-Roman World
DF Greece
DG Italy
DH Netherlands (low Countries). Belgium, Luxemburg
DJ Netherlands (Holland)
DJK Eastern Europe
DK Russia and former Soviet republics. Poland
DL Northern Europe. Scandinavia
DP Spain. Portugal
DQ Switzerland
DR Balkan peninsula
DS Asia
DT Africa
DU Oceania (South Seas) [Australia. New Zealand]
DX Roma (Gypsies)

E-F History of America

E1-143 America (General)
E151-857 United States (Gen.)
F1-957 United States: States and Local
F1001-1140 Canada
F1201- Other individual countries [mostly Latin America]

G Geography

G - GF Geography

G Geography (General). Atlases. Maps
GA Mathematical geography. Cartography
GB Physical geography
GC Oceanography
GE Environmental sciences
GF Human ecology. Anthropogeography
GN Anthropology
GR Folklore
GT Manners and customs (General)
GV Recreation. Leisure

H Social sciences

H Social sciences (General)
HA Statistics

HB-HJ Economics

HB Economic theory. Demography
HC-HD Economic history and conditions
HE Transportation and communication
HF Commerce
HG Finance
HJ Public finance

HM-HX Sociology

HM Sociology (General and theoretical)
HN Social history. Social problems. Social reform
HQ The family. Marriage. Woman
HS Societies: secret, benevolent, etc. Clubs
HT Communities. Classes. Races
HV Social pathology. Social and public welfare. Criminology
HX Socialism. Communism. Anarchism

J Political science

J General legislative and executive papers

JA-JC Political science

JA Collections and general works
JC Political theory

JF-JQ Political institutions and public administration

JF General works. Comparative works
JK United States
JL Brit. America. Latin America
JN Europe
JQ Asia. Africa. Australia. Oceania
JS Local government
JV Colonies and colonization. Emigration and Immigration
JX International law. International relations

No longer used at LC

K Law

K Law (General)
KD United Kingdom and Ireland
KDZ America. North America
KE Canada
KF United States
KG Central America, Caribbean
KH South America
KJ-KK Europe

L Education

L Education (General)
LA History of education
LB Theory and practice of educ.
LC Special aspects of education
LD-LG Individual institutions
LD United States
LE America except United States
LF Europe
LG Asia, Africa, Oceania
LH College and school magazines and papers
LJ Student fraternities and societies in the United States
LT [Multi-subject] Textbooks

M Music and books on music

M Music [instrumental and vocal]
ML Literature of music
MT Musical instruction and study

N Fine arts

N Visual arts (General)
NA Architecture
NB Sculpture
NC Drawing. Design. Illustration
ND Painting
NE Print media
NK Decorative arts. Applied arts. Decoration and ornament
NX Arts in General

P Language and literature

P Philology and linguistics (Gen.)
PA Classical languages and lit.
PB-PH Modern European lang.
PB Celtic languages and literature
PC Romance languages
PD-PF Germanic languages
PD Scandinavian. North Germanic
PE English
PF West Germanic
PG Slavic. Baltic. Albanian languages and literature
PH Finno-Ugrian. Basque l & l
PJ-PL Oriental languages & lit.
PJ Oriental. Semitic
PK Indo-Iranian
PL Languages and literatures of Eastern Asia, African, Oceania
PM Hyperborean, Indian, and artificial languages
PN-PZ Literature
PN Literary history and collections
PQ Romance literature
PR English literature
PS American literature
PT Germanic literature
PZ Children's literature

Q Science

Q Science (General)
QA Mathematics. [Computer science]
QB Astronomy
QC Physics
QD Chemistry
QE Geology
QH-QR Biology
QH Natural history (General). Biology (General)
QK Botany
QL Zoology
QM Human anatomy
QP Physiology
QR Microbiology

R Medicine

R-RL Medicine
R Medicine (General)
RA Public aspects of medicine
RB Pathology
RC Internal medicine. Practice of Medicine
RD Surgery
RE Ophthalmology
RF Otorhinolaryngology
RG Gynecology and obstetrics
RJ Pediatrics
RK Dentistry
RL Dermatology
RM-RZ Allied disciplines
RM Therapeutics. Pharmacology
RS Pharmacy and materia medica
RT Nursing
RV Botanic, Thomsonian, and eclectic medicine
RX Homeopathy
RZ Other systems of medicine [Chiropractic. Osteopathy. Mental healing]

S Agriculture

S Agriculture (General)
SB Plant culture
SD Forestry
SF Animal culture
SH Aquaculture. Fisheries. Angling
SK Hunting

T Technology

T Technology (General)
TA-TH General engineering and civil engineering
TA General
TC Hydraulic and ocean eng.
TD Environmental technology, sanitary engineering
TE Highway engineering
TF Railroads
TG Bridge engineering
TH Buildings
TJ-TL Mechanical group
TJ Mechanical engineering
TK Electrical engineering. Nuclear engineering
TL Motor vehicles. Aeronautics. Astronautics
TN-TR Chemical group
TN Mining, metallurgy
TP Chemical technology
TR Photography
TS-TX Composite group
TS Manufactures
TT Arts and crafts. Handicrafts
TX Home economics

U Military science

U Military science (General)
UA Armies: Organization, description, facilities, etc
UB Military administration
UC Maintenance and transportation
UD Infantry
UE Cavalry, armor
UF Artillery
UG Military engineering. Air forces. Air warfare
UH Other services

V Naval science

V Naval science (General)
VA Navies: Org., descr., fac., etc
VB Naval administration
VC Naval maintenance
VD Naval seamen
VE Marines
VF Naval ordnance
VG Minor services of navies
VK Navigation. Merchant marine
VM Naval engineering. Ship-building. Marine Engineering

Z Bibliography and library science.

Z4-115 Books (General). Writing. Paleography
Z116-659 Book industry & trade
Z662-1000 Libraries. [Library science. Information science]
Z1001-8999 Bibliography

Library of Congress Classification
Sample classes dealing with or relevant to *transportation*

The following list gives a sampling of LC classes dealing with or relevant for *transportation and traffic*. The example classes are in *italics*. For each, the hierarchical chain leading to it is given to provide a sense of context, but neighboring classes are shown only in a few cases for illustration. Some full pages from the classification are also included with examples underlined (unless the entire section is about transportation). The examples have been chosen to make it easy for you to detect patterns on your own.

B Philosophy. Psychology. Religion

BJ	**Ethics**
BJ1801-2195	. Social usages. Etiquette
BJ 2137	. . *Etiquette of travel*
BJ2139-2156	. . . *Special topics*
BJ2139	. . . . *Airplane travel*
BJ2140	. . . . *Bus travel*

BS	**The Bible**
BS1-680	. General (Whole Bible)
BS410-680	. . Works about the bible
BS620-672	. . . Auxiliary topics
BS647-649	. . . . Prophecy
BS649	. . . . . Prophecy of special future events, A-Z
BS649.S8	. . . . . . *Steam engines*

BV	**Practical theology**
BV5-530	. Worship (public and private)
BV205-287	. . Prayer
BV229-283	. . . Prayers
BV283	. . . . Other special prayers, A-Z
BV283.A4	. . . . . *Air pilots' prayers*
BV283.T7	. . . . . *.Traveller's prayers*
BV590-1652	. Ecclesiastical theology
BV900-1450	. . Religious societies, associations, etc.
BV950-1280	. . . Religious societies of men, brotherhoods, etc.
BV955-1280	. . . . By period
BV960-1280	. . . . . 19th-20th centuries
BV1000-1220	. . . . . . Young Men's Christian Associations
BV1160-1220	. . . . Work with special classes
BV1175	. . . . *Commercial travellers*
BV1200	. . . . *Railroad employees*

BV2002-3705	. Missions
BV2610-2695	. . Special types of missions
BV2660-2695	. . . Work among special classes, by occupation
BV2695	. . . . Other classes, A-Z
BV2695.R3	. . . . . *Railroad men*
BV4000-4470	. Pastoral theology
BV4200-4317	. . Preaching. Homiletics
BV4239-4317	. . . Sermons
BV4309-4316	. . . . Sermons and talks to special classes of persons
BV4316	. . . . . Other classes, A-Z.
BV4316.R3	. . . . . . *Railroad men*
BV4316.S3	. . . . . . *Sailors and seamen*
BV4400-4470	. . Practical church work. Social work. Work of the layman
BV4435-4470	. . . Church work with special classes
BV4457-4459	. . . . Soldiers and sailors
BV4458	. . . . . *Sailors and seamen*
BV4485-5099	. Practical religion. The Christian life
BV4527-4596	. . Religious works for special classes
BV4588-4591	. . . Soldiers. Soldiers and sailors
4590-4591	. . . . *Sailors and seamen*
4596	. . . Other, A-Z
4596.R3	. . . . *Railroad men*

C Auxiliary sciences of history

CB	**History of civilization**
CB156	. *Terrestrial evidence of interplanetary voyages*
CB440-481	. Relation to special topic
CB440	. . *Astronautics and civilization*

CJ	**Numismatics**
CJ1-4625	. Coins
CJ161	. . Symbols. devices, etc., A-Z
CJ161.B2	. . . *Bridges*
CJ 161.S5	. . . *Ships*
CJ 161.T73	. . . *Transportation*

	Massachusetts
	Regions, counties, etc., A-Z -- Continued
72.N2	Nantucket Co.
	Including Nantucket, Muskeget, and Tuckernuck Islands
72.N6	No Mans Land (Island)
72.N8	Norfolk Co.
	Pioneer Valley, see F72.C7
72.P6	Plum Island
72.P7	Plymouth Co.
	Including North River, South Shore
72.S9	Suffolk Co.
	Cf. F73, Boston
72.S94	Swift River Valley. Quabbin Reservoir. Ware River Valley
72.W9	Worcester Co.
	Including Wachusett Mountain
73	Boston
73.1	Periodicals. Societies. Collections
73.15	Museums. Exhibitions, exhibits
	For Foreign Exhibition, 1883, see T460
73.18	Guidebooks
73.25	Biography (Collective). Genealogy (Collective)
	Including vital records, epitaphs
73.27	Historiography
73.29	Study and teaching
73.3	General works. Histories
73.33	Juvenile works
73.35	Pamphlets, addresses, essays, etc.
73.36	Anecdotes, legends, pageants, etc.
73.37	Historic monuments (General). Illustrative material
73.39	Antiquities (Non-Indian)
	By period
73.4	Early to 1775
	Including the fires of 1711, 1737, 1760, etc.
	Including biography: William Blackstone, etc.
	Cf. E215, Events just prior to the Revolution
73.44	1775-1865
	Cf. E450, Fugitive slave riots
73.5	1865-1950
	Including fire of 1872
	1951-
73.52	General works
	Biography
73.53	Collective
73.54.A-Z	Individual, A-Z
	Sections. Localities. Districts, etc.
73.6	General works
73.61	Cemeteries
	Including Copp's Hill Burial Ground, Granary Burial Ground, King's Chapel Burial Ground, Mount Hope Cemetery

Massachusetts
Boston
Sections. Localities.
Districts, etc. -- Continued

73.62	Churches
73.625	Hotels, taverns, etc.
73.627	Places of amusement
73.63	Harbor
73.64	Monuments. Statues
	e.g.
73.64.A1	General works
73.64.B4	Beacon Hill Monument
73.64.W4	Wendell Phillips Statue
73.65	Parks. Squares. Circles
	Including Boston Common, Franklin Park
73.67	Streets. Bridges. Railroads, etc.
	e.g.
73.67.A1	General works
73.67.P3	Park Street
73.67.S7	State Street
73.67.T7	Tremont Street
73.67.W3	Washington Street
73.68.A-Z	Suburbs. Sections of the city. Rivers
73.68.B4	Beacon Hill
	Brighton, see F74.B73
73.68.C3	Castle Island
	Charlestown, see F74.C4
	Dorchester, see F74.D5
73.68.E2	East Boston
73.68.L43	Leather District
73.68.N65	North End
73.68.R2	Rainsford Island
	Roxbury, see F74.R9
73.68.S7	South Boston
73.68.T5	Thompson Island
	West Roxbury, see F74.W59
73.69	Wards
	Buildings
73.7	Collective
73.8.A-Z	Individual, A-Z
	e.g.
73.8.C9	Crown Coffee House
73.8.F2	Faneuil Hall Market
73.8.O4	Old State House
73.8.P3	Parker House
73.8.S8	State House
	Elements in the population
73.9.A1	Collective
73.9.A2-Z	Individual elements
	For list of racial, ethnic, and religious elements (with cutter numbers), see E184.A+

F1–97

	Massachusetts -- Continued
74.A-Z	Other cities and towns, etc., A-Z
	e.g.
74.B73	Brighton
74.C1	Cambridge
74.C4	Charlestown
74.C8	Concord
74.D5	Dorchester
74.G9	Groton
74.P8	Plymouth
74.P96	Provincetown
74.R9	Roxbury
74.S1	Salem
74.S8	Springfield (Table F2)
74.T2	Taunton
	Including Old Colony Historical Society
74.W59	West Roxbury
74.W9	Worcester
	Elements in the population
75.A1	Collective
75.A2-Z	Individual elements
	For a list of racial, ethnic, and religious elements (with cutter numbers), see E184.A+
	Rhode Island
76	Periodicals. Societies. Collections
76.5	Museums. Exhibitions, exhibits
77	Gazetteers. Dictionaries. Geographic names
77.3	Guidebooks
78	Biography (Collective). Genealogy (Collective)
78.2	Historiography
	Historians, see E175.5.A+
78.5	Study and teaching
79	General works. Histories
79.3	Juvenile works
79.5	Pamphlets, addresses, essays, etc.
79.6	Anecdotes, legends, pageants, etc.
80	Historic monuments (General). Illustrative material
81	Antiquities (Non-Indian)
	By period
82	Early to 1775
	Including First Rhode Island Charter, 1663; the Narragansett country; union of plantations at Newport, Portsmouth, Providence, and Warwick, 1636-1643
	Including biography: John Clarke, Samuel Gorton, Roger Williams, etc.
	Cf. E83.67, King Philip's War, 1675-1676
83	1775-1865
	Cf. E215.6, Gaspee affair
	Cf. E230+, Military operations and battles
	Cf. E263.R4, Rhode Island in the Revolution (General)
	Cf. E528.1+, Civil War, 1861-1865 (General)

Note
A document on
Railroads in Cambridge
is classified simply as
F74.C1

G Geography

GN	**Anthropology**
GN301-673	. Ethnology. Social and cultural anthropology
GN406-517	. . Cultural traits, customs, and institutions
GN406-442	. . . Technology. Material culture
GN438-442	. . . . *Transportation*
GN438	. . . . . *General works*
GN438.2	. . . . . *General special*
GN439	. . . . . *Routes of communication*
	Including trails, roads, bridges, etc.
GN440-440.2	. . . . . *Transportation by water. Navigation*
GN440	. . . . . . *General works*
GN440.2	. . . . . . *Boats*
GN440.2	. . . . . . *.Canoes*
GN441	. . . . . *Vehicles. Wheels*
GN442	. . . . . *Snowshoes. Skis*
GN448-450	. . . . Economic organization. Economic anthropology
	. . .
GT	**Manners and customs**
GT3400-5090	. Customs relative to public and social life
GT5010	. . Official ceremonies of royalty, nobility, etc.
	. . .
GT5220-5285	. *Customs relative to transportation and travel*
	Cf, BJ2137+Etiquette of travel
	G149+ Voyages and travels
	G540 Seafaring life
	GT490 Customs relating to wayfarers
	HE Transportation
GT5220	. . *General works*
	. . *By period*
GT5230	. . . *Ancient*
GT5240	. . . *Medieval*
GT5250	. . . *Modern, through 1800*
GT5260	. . . *1801-*
	. . *Vehicles. Chariots. Cars*
GT5280	. . . *General works*
GT5285	. . . *Sleighs and sledges*
	Horses, see GT5885

H Social sciences

HD	**Economic history and conditions** [See sample pages following]
HE	**Transportation and communications** [See sample pages following]
HF	**Commerce**
HF5001-6182	. Business
HF5601-5689	. . Accounting. Bookkeeping
HF5686	. . . By business or activity, A-Z
	A list of seven pages, including
HF5686.A38	. . . . *Air transportation. Airlines*
	Railways, see HE2241 [Accounting under Railways]
HF5686.T6	. . . . Tobacco
HF5686.T7	. . . . *Transportation*
	Transportation, air, see HF5686.A38
	Transportation, automotive,
	see HE5618 Automotive transportation > Finance, accounting, etc.
	Transportation, local,
	see HE4351 Street railways. Subways. Rapid transit systems > Finance. Accounting. Auditing
HF5686.T73	. . . . Travel agents
HF5686.T8	. . . . Trustees
HJ	**Public finance**
HJ2240-5957	. Revenue. Taxation. Internal revenue
HJ3231-3696	. . Taxation. Administration and procedure
	[Note: Transportation taxes used to be here, but have been moved to HE: HE196.9 [Taxation under Transportation in general] or HE384+ [Control, taxation, tolls, etc. under Water transportation]
HQ	**The family. Marriage. Woman**
HQ503-1064	. The family. Marriage. Home
HQ1060-1064	. . Aged, Gerontology (Social aspects). Retirement
HQ1063.5	. . . *Transportation*
HV	**Social pathology. Social and public welfare. Criminology**
HV697-4959	. Protection, assistance, and relief
HV1551-3024	. . Handicapped
HV1568.6	*Transportation and travel*
	For transportation of persons with specific handicaps, see the specific class of handicapped persons. [But not all have a subclass transportation.]
HV3011-3024	. . . Physically handicapped
HV3022	. . . . *Transportation and travel*
HV3025-3174	. . Special classes. By occupation
HV3025-3164	. . . *Seamen*

Labor. Work. Working class
By industry or trade, A-Z -- Continued
8039.T24 Telecommunication workers
Apply table at HD8039.A23
Telegraphers
Including radio operators, railroad telegraphers
8039.T25 Commercial
Apply table at HD8039.A23
8039.T27 Railroad
Apply table at HD8039.A23
8039.T3 Telephone employees
Apply table at HD8039.A23
8039.T38 Television station employees. Radio station employees
Apply table at HD8039.A23
8039.T4 Textile workers. Man-made fibers industry employees
Including cotton, silk and woolen workers, spinners and weavers
Apply table at HD8039.A23
Theatrical employees, see PN1+
8039.T52 Thimblemakers
Apply table at HD8039.A23
8039.T528 Tile industry employees
Apply table at HD8039.A23
8039.T53 Tile layers
Including mosaic tile layers
Apply table at HD8039.A23
Tin workers, see HD8039.I5
Tin plate workers, see HD8039.M5
8039.T57 Tinkers
Apply table at HD8039.A23
8039.T58 Tire industry workers
Apply table at HD8039.A23
8039.T6 Tobacco workers
Apply table at HD8039.A23
8039.T64 Tourist trade employees
Apply table at HD8039.A23
8039.T67 Toy makers. Toy industry workers
Apply table at HD8039.A23
Trackmen, see HD8039.R43
8039.T69 Tractor industry employees
Apply table at HD8039.A23
8039.T7 Transport workers
Cf. HD8039.M795, Motor-truck drivers
Apply table at HD8039.A23
Traveling sales personnel, see HD8039.C65
Truck drivers, see HD8039.M795
8039.T8 Turpentine industry workers
Apply table at HD8039.A23
8039.T95 Type founders
Apply table at HD8039.A23
Typographical unions, see Z120

Special industries and trades
Mechanical industries
Photographic equipment. Photographic services
Special equipment, A-Z -- Continued
9708.5.P56-P564 Phototypesetting equipment (Table H21)
Transportation vehicles and equipment
Class here works on the production of vehicles a equipment used for transportation
For works on transportation economics, see HE1
9709 General works (Table H20)
9709.5 Carriages and wagons (Table H20)
Automobiles. Motor vehicles
Including electric vehicles
Cf. HD9715.7, Trailer homes
Cf. HD9744.V43+, Military vehicles
9710 General works (Table H20)
9710.3 Automobile supplies, parts and accessories industry (Table H20)
9710.33 Limousines (Table H20)
9710.34 Buses (Table H20)
Including bus supplies, parts, etc.
9710.35 Trucks (Table H20)
Including truck supplies, parts, etc.
For industrial electric trucks, see HD9697.T73+
9710.37 Recreational vehicles (Table H20)
9710.38 Trailers (Table H20)
Including camping trailers, travel trailers, truck trailers, utility trailers
9710.4 Tractors (Table H20)
Cf. HD9486+, Agricultural implements and machinery industry
9710.5 Motorcycle industry (Table H20)
9710.6 Moped industry (Table H20)
Airplanes
Cf. TL500+, Aeronautical engineering
Cf. TL671.28, Airplanes (Manufacturing. Factory equipment and methods)
Cf. TL724+, Airplane industries
9711 General works (Table H20)
9711.2 Airplane supplies, parts and accessories industry (Table H20)
Cf. HD9744.A33+, Military aeronautics equipment
9711.25 Helicopters (Table H20)
9711.3 Lighter-than-air craft (Table H20)
Cf. TL667+, Technology
9711.5 Aerospace industry (Table H20)
Cf. TL780+, Rocket propulsion
Cf. TL787+, Astronautics
9711.75 Space industrialization (Table H20)
9712 Railroad rolling stock, equipment, etc. (Table H20)

Freight (General) -- Continued
199.A3-Z By region or country, A-Z
For freight of a special commodity in a specia place, see HE199.5.A+
199.5.A-Z Special commodities, A-Z
Under each commodity:
.x *General works*
.x2A-Z *By region or country, A-Z*
199.5.A33 Aggregates
199.5.A54 Animals
199.5.B35 Banana
199.5.B7 Bricks
199.5.C5 Chlorine
199.5.C6 Coal
199.5.C625 Coal-water fuel
199.5.C63 Coffee
199.5.C67 Corn
199.5.D3 Dangerous goods. Hazardous substances
199.5.F3 Farm produce
199.5.F6 Food
199.5.F8 Fuel
199.5.G3 Natural gas
199.5.G7 Grain
Hazardous substances, see HE199.5.D3
Iron, Sand, see HE199.5.S26
199.5.L43 Liquefied petroleum gas
199.5.L45 Liquids
199.5.L5 Livestock
199.5.L84 Lumber
199.5.M4 Metal products
199.5.M5 Milk
199.5.O27 Oats
199.5.O7 Ores
199.5.P4 Petroleum
199.5.R3 Radioactive substances
199.5.R43 Refrigerated goods
199.5.S26 Sand, Iron
199.5.S66 Soybeans
199.5.T43 Tea
199.5.W5 Wheat
199.9 Passenger traffic (General)
Cf. HQ1063.5, Aged
Cf. HV554, Evacuation of civilians in disast
Cf. HV3005.5, Mentally handicapped
Cf. HV3022, Physically handicapped
Cf. LB2864+, School children
For passenger traffic by special forms of transportation, see the form
By region or country
201 America
North America
202 General works
United States

Railways
Railway administration
Traffic
Rates. Tariff
Passenger tariff
Reduced rates, commutation rates, etc.
Rates for special classes
of passengers, A-Z -- Continued
1960.B5 Blind
1960.C5 Civil Service employees
1960.T6 Tourists
1960.W7 Workingmen
1961 Passes
1965 Transportation in exchange for advertisi
1971 Ticket brokerage
1976 Tickets
2001-2100 By region or country (Table H1)
Add country number in table to HE20
Freight tariff
Including schedules
2101 General works
Calculators, see TF664
Premium tables, see TF664
2116.A-Z Rates on particular articles, A-Z
For rates on articles in a particular
country, see HE2121+
2121-2220 By region or country (Table H1)
Add country number in table to HE212
Finance
2231 General works
2236 General special
2241 Accounting
2242 Bills of lading
2243 Collecting
2245 Disbursements
2246 Auditing
2251 Receiverships and reorganization
2261 Clearinghouse
Statistics
2271 Theory. Method
2273 General works
By region or country
see HE2713, HE2751, and HE2801-HE3560,
subdivisions (1) and (8) under each coun
Freight
2301 General works
2311 General special
2315 Weighing of freight
2316 Piggyback transportation
Refrigerator service, see TF667
2321.A-Z Special commodities, products, etc., A-Z
2321.A48 Aggregates
2321.A5 Agricultural machinery

coal

Railways
  Railway administration
    Traffic
      Freight
        Special commodities,
          products, etc., A-Z -- Continued

2321.A6	Alcohol
2321.A8	Automobiles
2321.B35	Berries
2321.B4	Beverages
2321.B7	Brick
2321.C25	Canned foods
2321.C35	Cement
2321.C45	Chemicals
2321.C5	Citrus fruits
2321.C6	Coal
	Corn, see HE2321.M3
2321.C7	Corpses
2321.C75	Cotton
2321.D25	Dairy products
2321.D3	Dangerous goods. Hazardous substances
2321.E5	Eggs
2321.E8	Explosives
2321.F3	Farm produce
2321.F4	Fertilizers
2321.F7	Fruit
2321.F8	Fuel
2321.G4	Glass
2321.G7	Grain
	Hazardous substances, see HE2321.D3
2321.I5	Insecticides
2321.I7	Iron
2321.L4	Leather
2321.L5	Lime
2321.L7	Livestock
2321.L8	Lumber
2321.M3	Maize
2321.M4	Meat
2321.M45	Metal products
2321.M6	Mineral products
2321.N3	Naval stores
2321.P3	Paper
2321.P4	Petroleum
2321.P5	Plaster
2321.P6	Potatoes
2321.R44	Refuse
2321.R5	Rice
2321.R8	Rubber
2321.S3	Salt
2321.S6	Soap
2321.S7	Steel
2321.S75	Stone
2321.S8	Sugar

Under this class, there is a six-page list of countries.
A similar lengthy list under Street railways. Subways. Rapid transit systems (starting at HE4201), a subclass of railways.

Railways
By region or country
starts at HE 2701
Other regions or countries
Latin America
South America -- Continued

2951-2960	Ecuador
2961	Guianas
2962	Guyana. British Guiana
2963	Surinam. Dutch Guiana
2964	French Guiana
2965.A-Z	Special companies, A-Z
2966-2970	Paraguay
2971-2980	Peru
2981-2990	Uruguay
2991-3000	Venezuela
3001-3010	Europe
3011-3020	Great Britain
	Including works on the British Empire, "Greater Britain", etc.
	England, see HE3011+
3031-3040	Scotland
3041-3050	Ireland
3051-3059.2	Austria
3059.2	Special railroads or companies, A-Z
3059.3	Czechoslovakia
3059.5	Hungary
3060.5	Poland
3061-3070	France
3071-3080	Germany
	Including West Germany
3080.5	East Germany
3081-3090	Greece
3091-3100	Italy
3100.3	Malta
3100.5	Benelux countries. Low countries
3111-3120	Belgium
3121-3130	Netherlands
3130.5	Luxembourg
3131-3140	Russia. Soviet Union. Former Soviet Reput
3140.2	Russia (Federation)
3140.3	Belarus
3140.4	Moldova
3140.5	Ukraine
	Baltic States
3140.6	General works
3140.65	Estonia
3140.7	Latvia
3140.75	Lithuania
3141-3150	Scandinavia
3151-3160	Denmark
3161-3170	Iceland
3171-3180	Norway
3181-3190	Sweden

	Railways
	By region or country
	Other regions or countries
	Europe -- Continued
3190.5	Finland
3191-3200	Spain
3201-3210	Portugal
3211-3220	Switzerland
3221-3230	Balkan States
3231-3240	Bulgaria
3241-3245	Yugoslavia
3251-3260	Romania
	Turkey, see HE3371+
3271-3280	Asia
3281-3290	China
3290.5	Taiwan
3291-3300	India
3300.3	Sri Lanka
3300.5	Pakistan
3300.6	Bangladesh
3301-3310	Southeast Asia. Indochina
3320.2	Cambodia
3320.3	Vietnam
3320.4	Laos
3321-3330	Malaysia
3331-3340	Indonesia
3341-3350	Philippines
3351-3360	Japan
3360.5	Korea
3361-3370	Iran
3371-3380	Turkey
3380.3	Saudi Arabia
3380.5.A-Z	Other divisions of Asia, A-Z
	For Siberia, see HE3140.2
3390.A-Z	Other special companies, A-Z
3391-3400	Africa
	North Africa
3400.5	General works
3401-3410	Egypt. United Arab Republic
3411	Morocco
3412	Algeria
3413	Tunisia
3414	Libya
3415	Sudan
	Northeast Africa
3415.5	General works
3416	Ethiopia
3417	Somalia
	Including British and Italian Somaliland
3418	French Territory of the Afars and Issas
	Southeast Africa
3418.5	General works
3419	Kenya

HE

	Transportation and communications
	Including roads, waterways, railways, postal service, telecommunication, telegraph, telephone, broadcasting, commercial aviation, etc.
	Cf. HF1+, Commerce
	Cf. P87+, Communication. Mass media
	Cf. TA1001+, Transportation engineering
	For transportation and communications by region or country, see HE201+
	Periodicals. Societies. Serials. By language of publication
1	English
3	French
5	German
7	Other languages (not A-Z)
8	Yearbooks
	Shippers' guides. Directories. Timetables, etc.
	Cf. G153+, Travelers' guides
8.9	General works
9.A-Z	By region or country, A-Z
11	Congresses
	Museums. Exhibitions
13	General
13.2.A-Z	By region or country, A-Z
	Under each country:
	.x *General works*
	.x2A-Z *Special. By city, A-Z*
	Collected works (nonserial)
131	Several authors
136	Individual authors
141	Dictionaries. Encyclopedias
147	Terminology. Abbreviations. Notation
	Theory. Method. Relation to other subjects
147.5	General works
147.6	Data processing
147.7	Mathematical models
148	Transportation and community development
148.5	Transportation and developing countries
149	Transportation and agriculture
	Cf. HE1042.8+, Railways and farming interests
151	General works
	Biography
151.4	Collective
151.5.A-Z	Individual, A-Z
152	Juvenile works
152.5	General special (Special aspects of the subject as a whole)
152.6	Addresses, essays, lectures
153	Use of animals for transportation
	Cf. SF180, Draft animals
	History
(155)	Primitive
	see GN438-GN442

	History -- Continued
	Ancient
159	General works
161	Egypt
163	Phoenicia
167	Assyro-Babylonian Empire. Persia
169	Other Oriental
	Greece and Rome
171	General works
173	Greece
175	Rome
177.A-Z	Other, A-Z
	Middle Ages
	Including works treating the ancient period and Middle ages combined
181	General works
	By region or country, see HE201+
	Modern, see HE151
	Statistics
191.4	Collections of statistics
191.5	Theory
	Information services
191.8	General works
191.82.A-Z	By region or country, A-Z
	Study and teaching
191.9	General works
192.A-Z	By region or country, A-Z
	Research
192.5	General works
192.55.A-Z	By region or country, A-Z
193	Public policy (General)
	Transportation safety. Transportation accidents
194	General works
194.5.A-Z	By region or country, A-Z
	Rate making
	Including costs
195.4	General works
195.5.A-Z	By region or country, A-Z
196	Concessions
196.5	Finance
	Taxation, tolls, etc.
	Cf. HE384+, Tolls on waterways
196.9	General works
197.A-Z	By region or country, A-Z
198	State ownership (General)
	For state ownership related to a subject, see the subject
	Freight (General)
	For freight related to a special form of transportation, see the form
199.A2	General works

Water transportation
Shipping
Marine insurance -- Continued
71 Salvage
Cf. VK1491, Naval science
Railways
For railways by region or country, see HE2701+
001 Periodicals. Serials
003 Societies
005 Congresses
007 Museums. Exhibitions
009 Directories
Collected works (nonserial)
011 Several authors
013 Individual authors
Dictionaries and encyclopedias, see TF9
021 History
031 General works
035 General special
Relation to other subjects
041 Railways and civilization
Railways and farming interests
042.8 General works
By region or country
043 United States
043.5.A-Z Other regions or countries, A-Z
Railways and industries
044 General works
045 Railways and coal interests
049 Railways and other carriers
Including waterways, airplanes, and automobiles
1050 Railways and foreign trade
Public policy (General). Government control
For other countries, see HE2801+
For United States, see HE2757
1051 General works
State aid
1059.8 General works
1061 United States
1062 Pacific railroads
Land grants
United States
1063.A6 By name of railroad, A-Z
1063.A7-Z General works
1064.A-Z By region or state, A-Z
Use in war, see UC310+
1067 Postal service
Cf. HE6175+, Railway mail service
Cf. HE6475+, Parcel post
Taxation, valuation, depreciation, etc.
For individual railroads, see HE2791.A+
1070.8 General works
By region or country

HE

J Political science

JF-JQ	**Political institutions and public administration**
JK	. United States
JK401-1685	. . Government. Public administration
JK468	. . . Other special, A-Z
JK468.T7	. . . . *Transportation*

K Law

KF	**United States — general**
KF1600-2940	. Regulation of industry, trade and commerce. Occupational law
KF2161-2849	. . *Transportation and communication* [10 pages, divided by type of transportation, for example]
KF2271-2379	. . . *Railroads* [with much detail]

KFC	**United States — California**
KFC390-547	. Regulation of industry, trade and commerce. Occupational law
KFC469-543	. . *Transportation and communication* *[3 pages, divided by type of transportation, but less detail.]*
KFC490-499	. . . *Railroads [only three classes under here]*

L Education

LC	**Special aspects of education**
LC65-245	. Social aspects of education
LC189-214.53	. . Educational sociology
LC213-214.53	. . . Educational equalization. Right to education
LC214-214.53	. . . . School integration
LC214.5-53	. . . . . Special means of integration
LC214.5-53	. . . . . . *Transportation. Busing*

M Music and books on music

M	**Music**
M1497-5000	. Vocal music
M1497-1998	. . Secular vocal music
M1900-1980	. . . Songs (part and solo) of special character
M1977-1978	. . . . By topic, A-Z [A three-page list, including]
M1977.R3	. . . . . *Railroads*
M1977.T87	. . . . . *Truck drivers*

N Fine arts

NA	**Architecture**
NA4100-8480	. Special classes of buildings
NA4170-8480	. . Classed by use
NA4170-7010	. . . Public buildings
NA6290-6370	. . . . *Transportation and storage buildings*
NA6290	. . . . . *General works*
NA6300-6307	. . . . . *Airport buildings*
	Divided like NA4410-4417
NA6310-6317	. . . . . *Railway stations*
NA6320-6327	. . . . . *Bus terminals*

NC	**Drawing. Design. Illustration**
NC760-825	. Special subjects (Technique, history, and collections)
NC825	. . Other subjects, A-Z
NC825.A4	. . . *Airplanes*
NC825.A8	. . . *Automobiles*
NC825.B6	. . . *Boats*
NC825.B7	. . . *Bridges*
	. . .

P Language and literature

PN	**Literary history and collections**
PN6147-6231	. Wit and humor, satire
PN6231	. . Collections on special topics
	[A ten-page list, including]
PN6231.T68	. . . *Traffic regulations*

Q Science

QC	**Physics**
QC251-338.5	. Heat
QC290-297	. . Calorimeters and calorimetry
QC293	. . . Special types of calorimeters, A-Z
QC293.F8	. . . . *Fuel* [related to transportation]

QH	**Natural history (General). Biology (General)**
QH540-549.5	. Ecology
Qh545.A1-Z	. . Influence of special factors in the environment
QH545.A2-Z	. . . Special, A-Z
QH545.A3	. . . . *Air pollution* [related to transportation]
	[Note: In Germany, speed limits were introduced to cut emissions thought harmful to forests]

R Medicine

RA	**Public aspects of medicine**
RA1-1270	. Public health. Hygiene. Preventive medicine
RA772	. . Other subjects of public health, A-Z
RA772.T7	. . . *Traffic accidents* Cf. HE5613.5+, Motor vehicles
RC	**Internal medicine**
RC952-1245	. Special situations and conditions
RC970-986	. . Military medicine. Naval medicine
RC981-986	. . . *Naval medicine* *Including merchant marine*
	. . .
RC1030-1160	. . *Transportation medicine*
RC1040-1045	. . . *Automotive medicine (and classes under it)*
RC1050-1097	. . . *Aviation medicine (and classes under it)*

S Agriculture

SF	**Animal culture**
SF277-359.7	. Horses
SF311-312	. . *Draft horses*
SH	**Aquaculture. Fisheries. Angling**
SH209-399	. Fisheries
SH337	. . *Packing, transportation, and storage*
SH337.5	. . *Fishing port facilities*

T Technology

TE	***Highway engineering***
TF	***Railroads***
TG	***Bridge engineering***
TL	***Motor vehicles, aeronautics, astronautics***

[See two sample pages following. Note difference in perspective from class HE.]

	Railroad engineering and operation
	Cf. HE1001+, Economic aspects of railroads
	Cf. TD893.6.R3, Noise control
	Periodicals and societies. By language of publication
1	English
2	French
3	German
4	Other languages (not A-Z)
5	Congresses
	Exhibitions. Museums
6.A1	General works
6.A2-Z	By region or country, A-Z
	Apply table at TA6.A2-Z
7	Collected works (nonserial)
9	Dictionaries and encyclopedias
12	Directories, etc.
	For directories of other railroads, see HE2801+
	For directories of purchasing agents, see TF359
	For directories of supply dealers, see TF355
	For directories of U.S. railroads, see HE2721
	History
15	General works
16	Antiquities
	Including early curiosities of railroad development and horse railroads
19	Nineteenth century
20	Twentieth century
21-127	Country divisions (Table T1)
	Class here only technical works, including works on the physical plants of individual railroads
	Add country table in table to TF0
	Biography
139	Collective
140.A-Z	Individual, A-Z
	General works
144	Early to 1850
145	1850-
146	Elementary textbooks
147	Popular works
148	Juvenile works
149	Pictorial works
151	Pocketbooks, tables, etc. (General)
153	General special
155	Addresses, essays, lectures
	Study and teaching. Research
	America
171	General works
173	United States
176	Canada
179	Latin America
180	Europe
181	Asia
182	Africa

	Study and teaching. Research -- Continued
183	Australia
190	Economics of location
	Class here only technical works
	Cf. HE1613+, Location, right of way, etc. (Economic aspects)
	Cf. TF215+, Railroad surveying (General)
193	Estimates, costs, etc.
195	Specifications
197	Model railways
	Safety measures, see TF610
	Railway construction
200	General works
	Including reconnaissance, surveying, and location
203	General special
205	Railroad engineering (Fieldbooks, tables, etc.)
208	Preliminary operations
	Railroad surveying
210	General works
212	Reconnaissance
213	Preliminary surveying
213.5	Aerial photography in railroad surveying
214	Plotting, profiles, etc.
	Location
215	General works
216	Curves and turnouts
217	Transition spiral, etc.
	Earthwork
	Including rock cuttings
220	General works
222	Tables, calculations, etc.
	Excavating machinery, steam shovels, etc.
225	General works
226	Catalogs
	Tunnels and tunneling
230	General works
232	General special
	Including special methods of tunnels, e.g. Pressure tunneling
	Maintenance and operating of tunnels
234	General works
235	Ventilation
236	Lighting
238.A-Z	Special tunnels, A-Z
	Permanent way. Superstructure. Roadway. Track
	Cf. TF530+, Maintenance and repair
240	General works
241	General special
243	Grades
244	Gage
245	Drainage
248	Tracklaying machinery
	Details of the permanent way

U Military science

UC	**Maintenance and transportation**
UC270-360	. *Transportation*
	[One-page broad classification of all modes of transportation, e.g.]
UC310-315	. . *Railroads*

V Naval science

[Almost all of this is relevant, see the detailed LCC outline. Especially]

VK	***Navigation. Merchant marine*** *and*
VM	***Naval engineering. Ship-building. Marine Engineering***
	[Both refer to civilian water transport]

Z Bibliography and library science.

Z662-1000	**Libraries**
Z665-718.8	. Library science. Information science
Z675	. . Classes of Libraries, A-Z [Three-page listing, including]
Z675.N3	. . . *Naval*
Z675.T7	. . . *Transportation libraries*
Z687-718.8	. . The collections. The books
Z693-Z695.83	. . . Cataloging
Z695.1	. . . . By subject, A-Z
	[four-page list, including]
Z695.1.N3	. . . . . *Naval art and science*
Z695.1.R34	. . . . *Railroads*
Z695.1.T73	. . . . . *Transportation*
Z696-697	. . . Classification and notation
Z697	. . . . By subject or form, A-Z [Two-page list, including]
Z697.T7	. . . . . *Transportation*
Z1001-8999	**Bibliography**
Z1001-1121	. General bibliography
Z1201-4980	. National bibliography
Z5051-7999	. Subject bibliography
	Subjects arranged in alphabetical sequence [sic!]
Z5811-14	. . Education
Z5814	. . . Special topics, A-Z [Four-page list, including]
Z5814.T7	. . . . *Transportation of pupils*
Z7231-7234	. . *Railroads*
Z8001-8999	. Personal bibliography
	Names of individuals arranged in alphabetical sequence

pink **LECTURE 8.1b (BRIEF)** ***March 23***

Short Media Streams Classification demo

XXX Video to be posted

An interesting classification for the purpose of indexing movie segments and video clips. In keeping with indexing visual materials it uses icons as descriptors

The video presentation will walk through the major sections of the classification and show examples of descriptors.

As you watch, reflect on the structure of the classification. Post comments on the discussion board.

pink

pink **Lecture 8.2a** (All students use online materials) ***March 23***

Index language functions (Textbook, Chapter 13) (60 min)

Subject analysis; abstracting and indexing; types and functions of abstracts

Learning objectives (expansion of Objective 2.2 from Part 5, p. ~285)	2.2 Understand the **functions of classification (more broadly, Knowledge Organization Systems, KOS) for a wide variety of tasks in a wide variety of systems.** (P2.3.9,2) 2.2.1 Understand the use of classification to support learning. (P2.3.9,2.1) 2.2.2 Understand the use of classification to help users get oriented in a subject and clarify their information needs. (P2.3.9,2.2) 2.2.3 Understand the principle of **request-oriented indexing** (user-centered indexing) and the fundamental role of the index language to communicate users' interests to the indexer. (P2.3.9,2.3) 2.2.4 Be able to **make intelligent decisions about the type of index language, indexing, and query formulation** to be used in a given IR system, considering costs and benefits. (P2.3.9,2.4) 2.2.5 Be able to recognize search requests that are difficult to handle in a system that does not use request-oriented indexing and be able to compensate, as far as possible, through creative pursuit of different avenues for the search. (P2.3.9,2.5) Also: XXX add to program objectives Understand the types and functions of abstracts and other document surrogates and be able to apply this understanding to writing abstracts and to the selection of databases based on the types of abstracts they contain.
Practical significance	Request-oriented indexing (also called problem-oriented indexing or user-centered indexing) is a special case of the maxim that the design and operation of information systems should be based on a thorough understanding of user requirements. Request-oriented indexing is the key to good system performance for the questions that matter to users. Yet in practice it is rarely used. Understanding this will enable students to make the best of existing systems and, more importantly, to go out and change practice.
Discussion question	How could request-oriented indexing be implemented in a reference tool addressed to a general audience, such as the *Reader's Guide to Periodical Literature*?

pink

Chapter highlights	• Derivation of the principle of **request-oriented indexing** from the problem-oriented approach to information systems introduced in Chapters 1 and 5 (Sections 13.1 and 13.2), • the role of index languages in searching and database organization (Sections 13.3 and 13.4), • design issues (Section 13.5), • review of index language functions (Section 13.6), • culminating in the recognition of an index language as a communication device from users to indexers, so that the indexers understand the users' interests (Section 13.7). Terminology: Filtering technique of indexing (Mooers 1958), Request-oriented indexing (DS 1974), problem-oriented indexing (DS), user-centered indexing (term in vogue now).
Questions	Your questions here
Discussion question (repeated)	How could request-oriented indexing be implemented in a reference tool addressed to a general audience, such as the *Reader's Guide to Periodical Literature*?

Document representation: purpose, structure, process of creation

Abstracts as a different form of document representation

Indicative abstract (also called descriptive abstract)
Merely indicates what the document is about or relevant for, pointer data.

Informative abstract
In addition, includes some of the substantive data given in the document or reports some generalization that can be derived from the document.

Both types of abstract assist the reader in deciding whether to pursue the document further (and incur any costs in doing so). An informative abstract often gives the substantive data needed and thus saves the user the trouble of having to consult the document itself.

Other categorization of abstracts: Reporting vs. analytical-critical. Book reviews

The structure of document representations (abstracts or lists of index terms)
Use a standard structure appropriate for the subject domain. The sample abstract illustrates one possibility. More discussion in the lecture on document structure.
Using terms from a controlled vocabulary (that should reflect the users' language) may increase readability and will increase findability.

Abstracting &indexing as a cognitive process. Empirical study of document-oriented indexing

Parts of the document considered

Method of information assimilation (reading, interpreting pictures)
Reading/scanning to identify subject matter of interest to users — request-oriented reading
Reading/scanning to fill slots of a frame
Building up mental image

Selecting topics to be included in the abstract or the index terms. Request-orientation comes into play here as well

Choosing a form of expression

Knowledge brought to bear on these operations - from own knowledge or tools (such as thesauri) consulted, for example
- General knowledge of the field
- Knowledge of user needs
- Frames for phenomena in the field
- Knowledge of terminology
- Knowledge of document structure, including knowledge of cue words

Automatic or computer-assisted abstracting and indexing

Conigrave KM, Saunders JB, Reznik RB. **Predictive capacity of the AUDIT questionnaire for alcohol related harm**. Addiction 90 (1995) 1479-1485.

Indicative abstract

This study deals with early identification of alcohol use disorders. It examined the ability of the Alcohol Disorders Identification Test (AUDIT) questionnaire published by the World Health Organization to predict which subjects experience medical or social harm from their drinking. Subjects were 350 emergency room patients who answered the AUDIT questions was part of a comprehensive medical assessment. 250 subjects were interviewed after 2-3 years to determine alcohol-related medical disorders, health care utilization, social problems and hazardous drinking at the time of follow-up. Audit is compared to biochemical indicators for its ability to predict these conditions.

Informative abstract

'AUDIT can predict a range of harmful consequences of alcohol consumption'
Background. Drinking problems often are not recognized. Most of the people who become alcohol-dependent do not seek help until their problems are obvious. Late diagnosis is of particular concern because effective and low-cost methods of treating problem drinking at an early stage are now available. In 1989, the WHO published a brief 10-item screening questionnaire, the Alcohol Disorders Identification Test (AUDIT) specifically designed to identify problem drinkers before physical dependence or chronic problems have arisen. AUDIT has been reported to have a sensitivity of 92% and a specificity of 94% in detecting hazardous or harmful alcohol use. This study examined the ability of the AUDIT questionnaire to predict which subjects experience medical or social harm from their drinking.
Methods. Subjects were 350 patients who attended a hospital emergency ward in 1984-1985. They underwent a comprehensive assessment of medical history, alcohol use, dependence and related problems in an interview schedule; the AUDIT questions were interspersed among other items. Biochemical variables measured included y-glutamyltransferase (GGT) and mean corpuscular volume (MCV). Twenty subjects refused to be contacted after 2-3 years or were excluded because of malignant disease. Thus, a cohort of 330 subjects (212 men, 108 women) was left for the longitudinal study; 250 subjects were interviewed again after 2-3 years. Interviewers were blind to the results of the initial assessment. The AUDIT questions were scored from 0 to 4. Subjects who scored 8 or more were classified as potentially hazardous drinkers. AUDIT was examined for its ability to predict a number of end-points including alcohol-related medical disorders, health care utilization, social problems and hazardous drinking at the time of follow-up.
Results. Of those who scored 8 or more on AUDIT at the initial interview, 61% experienced alcohol-related social problems compared with 10% of those with lower scores. They also reported more frequently alcohol-related medical disorders and hospitalization. The AUDIT score was a better predictor of social problems and of hypertension than laboratory markers. Its ability to predict other alcohol-related illnesses was similar to the laboratory tests, but GGT was the only significant marker of mortality.
Conclusions. AUDIT is a brief and convenient questionnaire which can readily be incorporated into the standard medical history. It can predict a range of harmful consequences of alcohol consumption. AUDIT should prove a valuable tool in screening for hazardous and harmful alcohol use so that intervention can be provided to those at particular risk of adverse consequences.

pink **Lecture 8.2b** (All students use online materials) ***March 30***

Vocabulary control (terminological control) and authority control

Learning objectives (from Learning objectives for Part 5, p. ~285 - 288,)	**1** Understand the problems and principles of vocabulary control / authority control (unique entity identifiers) and be able to apply these principles to indexing and searching (P2.3.8)
	1.1 **Understand the retrieval and communication problems caused by variety and ambiguity in language** – synonymy, homonymy, polysemy – including any kind of names, such as names of organizations. (P2.3.8,1)
	1.2 **Understand and be able to apply vocabulary control** to remedy these problems, either through vocabulary control **in indexing** or – through query term expansion – **in searching**. (P2.3.8,2)
	1.3 Understand the structure of a thesaurus with its **synonym-homonym structure** (all terms), **classificatory structure** (concepts expressed by preferred terms), **index language** (concepts and corresponding preferred terms selected as subject descriptors), and **lead-in vocabulary** (all terms that are not subject descriptors). (P2.3.8,3)
Practical significance	• Authority control is applied to terms designating subjects, to names of persons and organizations, to titles of often cited or reprinted works, and in many other cases. It is a major principle underlying many information retrieval systems, especially those used in libraries. • Lack of vocabulary control and authority control more generally is one of the most serious problems impeding the success of end-user searching in free-text searching. The solution lies in the design of systems, including search thesauri, that can assist end users.

Wider applications	**Vocabulary control as a special case of authority control** Vocabulary control is the control of subject identifiers. Similar problems arise in the control of the identifiers of other types of entities, such as persons or organizations; thesaurus of organizational names. In the broader sense one speaks of authority control (see Sections 9.1.1 and 9.1.2). The purpose of authority control can also be stated as **referential integrity**, that is, assuring a one-to-one correspondence between entity values and the character strings or other symbols that refer to them.

pink

Review of Textbook, Chapter 12 (20 min)

Lexical relationships (10 min.)

Paradigmatic relationships: Synonymy, antonymy, hyponymy

In linguistics: Relationships between terms based on their meanings, that is, on the concepts they designate. If a term has multiple meanings, only one of these meanings participates in the relationships discussed here.

In classification theory/knowledge representation: Relationships between concepts in a classificatory structure.

Paradigmatic relationships are contrasted with **syntagmatic relationships** that bind together words into phrases and sentences or elemental concepts into compound concepts, statements, or larger units of meaning.

<table>
<tr><td>Synonymy</td><td>Two terms designate the same concept. True synonyms can be used interchangeably in sentences without changing the meaning. Core meaning and connotations. Problem of shades of meaning and connotation.</td></tr>
<tr><td>Antonymy</td><td>Two terms designate opposite concepts. Opposites can be endpoints of a scale, such as light and dark, or exclusive categories, such as male and female</td></tr>
<tr><td>Hyponymy</td><td>Term A designates concept Concept A', term B designates concept Concept B', and Concept B' is more specific then Concept A'. Examples: flute (in one of its meanings) has as hyponym recorder (in one of its meanings); keyboard instrument has as hyponyms harpsichord and cembalo.

Note: In a thesaurus with a controlled vocabulary we would select a preferred term, for example harpsichord, and have the relationships

keyboard instrument Narrower Term harpsichord
harpsichord Synonymous Term cembalo

Antonymy and hyponymy are really concept relationships to be dealt with in Chapter 14; hyponymy is the relationship that defines a concept hierarchy. But all three relationships have in common that one term can be exchanged for the other in a sentence and still leave a sentence that has meaning.</td></tr>
</table>

Homonymy and polysemy

Note: The transition from homonymy to polysemy is gradual

<table>
<tr><td rowspan="2">Homonymy</td><td>Strict definition: two different words or phrases have the same spelling (homography) or the same pronunciation (homonymy in the narrowest sense).

Examples:
<table><tr><td>seal (marine mammal)
seal (document)</td><td>drill (bore a hole)
drill (furrow)
drill (fabric).</td></tr></table>

Note: While they are spelled the same, the words in each group have different etymological origin.</td></tr>
<tr><td>More expansive definition: The same word has two quite different meanings.

Examples:
<table><tr><td>drill (bore a hole)
drill (training)</td><td>seizure (disorder)
seizure (law enforcement).</td></tr></table>

Note: In each group, we have the same word (same etymological origin). The word acquired completely different meanings over time.</td></tr>
<tr><td>Polysemy</td><td>The same word has several meanings that can all be traced to a common core of meaning.

Example:
<table><tr><td>integration (mathematics)
integration (psychology)
integration (social groups)
integration (economic-political)
integration (curriculum)</td></tr></table>

All these meanings share a common core meaning: putting together pieces into whole where the pieces are held together in a larger structure.

Polysemy is often the result of metaphoric extension of the meaning of a term.

Example:
<table><tr><td>field (piece of land)
field (subject)
field (physics)
field (mathematics)</td></tr></table>

Even for drill (bore a hole) and drill (train) one can identify the core meaning of repetitive and persistent performance of an operation.</td></tr>
</table>

6 WordNet pages in supplement, SLecture 8.2a

red

Week 9. Things to do in Week 9, W March 23 - March 30

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	Assignment 10: Indexing of three documents (2 hours) (assigned L7.2, `Mar. 9)	☐
	Assignments due W March 30 ☐ required ○ optional	√
Assignment(s)	Assignment 11: Request-oriented indexing (2 hours) (assigned. L8.1a, `Mar. 23)	☐

	New topics this week	
	9.1 KOS structure 1: conceptual (70 min) (Strand 1. Gen. KOS principles)ⓘ ⓘ	
Readings	1 Lecture 9.1 Objectives etc. (pink). Also have a look at Assignments 12.1-12.2.	☐
	2 Textbook Ch. 14 **Index language structure1: conceptual**	☐
Lecture	**Lecture 9.1 sides** (50 min)	☐

	9.2 Application of KOS structure to searching (70 min.) (Strand 1. General KOS principles) ⓘ ⓘ	
Readings	1 Lecture 9.2 Objectives etc. (pink)	☐
	2 Textbook Section 14.4. **Application and illustration: Searching**	☐
Lecture	**Lecture 9.2 slides** (120 min)	☐

Learning blog	Learning blog Week 9 due W March 30	○

	Assignments assigned March 30	
Assignments assigned	▸**Assignment 12.1 Semantic factoring** (1.5 hrs) (assigned `Mar. 30, due `Apr. 6) ▸**Assignment 12.2 Building a hierarchy of elemental concepts** (1.5 hours) (assigned `Mar. 30, due `Apr. 6)	

pink **Lecture 9.1 Interactive** ***March 30***

Index language structure 1: conceptual (Textbook, Chapter 14)

Learning objectives	2.3 Understand the principles of the **structure of subject classification**, in particular facet organization and hierarchy, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. (P2.3.9,3)
	2.3.1 Understand the nature of hierarchical relationships among concepts. (P2.3.9,3.1)
	2.3.2 Understand the principle of concept componentiality: Elemental concepts can be combined into compound concepts and compound concepts can be analyzed (often non-obviously) into their elemental components resulting in a concept formula (semantic factoring, facet analysis). (P2.3.9,3.2)
	2.3.3 Be able to use the entity-relationship approach, specifically facet analysis, to discern the conceptual structure of a subject domain. **Be able to discern the facet structure of a subject domain**. (There are facets everywhere.) (P2.3.9,3.3)
	2.3.3.1 Understand the discernment of abstract concepts that apply across subject disciplines through semantic factoring. (P2.3.9,3.3.1)
Practical significance	This understanding provides a basis for • constructing an index language, a task required in setting up specialized information systems and, more importantly, in developing expert systems; • evaluating an index language to determine whether it is suitable for a given application; • indexing, particularly making sure that all applicable facets have been covered; • query formulation, facet analysis of queries.

pink

In-lecture exercises: Three steps in the **conceptual analysis and synthesis** in a subject:

Step 1.	Semantic factoring (results in a list of elemental concepts).
Step 2.	Arranging the elemental concepts in a well-structured faceted hierarchy.
Step 3.	Fit compound concepts into the framework of the hierarchy (if compound concepts need to be dealt with explicitly)

Step 1. In-lecture exercise: Semantic factoring

Semantic factor the concepts (Dewey classes) from the attached list.
Keep a running list of elemental concepts as they arise.

Step 2. In-lecture exercise: Building a hierarchy of elemental concepts

Sort elemental concepts into entity types or facets.

Arrange values within each entity type or facet in a meaningful structure.

Step3. In-lecture exercise: Fitting compound concepts into a hierarchy (in part)

In-lecture exercise, Textbook, Chapter 14: Semantic factoring
(Step 1)

Factor the following concepts (from Dewey Decimal Classification ed. XXX) **into their semantic components (semantic factors)**. If this is not possible, comment.
Keep a running list of the elemental concepts needed.

Note: A broader class is given in ( ) if necessary to specify the meaning of a term.

372.19	Curriculums of elementary schools
372.35043	Science in the elementary school curriculum
372.414	Methods of instruction for reading in elementary schools
372.72043	Arithmetic in the elementary school curriculum
373.19	Curriculums in secondary schools
373.243	Military schools (Secondary Education)
376.63	Secondary education of women
378.19	Curriculum of colleges and universities
378.33	Fellowships (Higher Education)
371.7	School health and safety
371.855	Men's social societies and fraternities (Generalities of Education)
371.856	Women's social societies and sororities
371.911	Blind and partially sighted students
371.912	Deaf and hard-of-hearing students
371.95	Curriculums for gifted students

In-lecture exercise: Building a hierarchy of elemental concepts

Sort elemental concepts into entity types or facets.

Arrange values within each entity type or facet in a meaningful structure.

Elemental concepts Running list (from Step 1)	**Elemental concepts in a meaningful structure**

In-lecture exercise: Fitting compound concepts into a hierarchy

pink **Lecture 9.2 Interactive** ***March 30***

Application of index language structure to searching (Organizing Info., 14.4)

Learning objectives (From Part 5, p. ~285)	2.3 Understand the principles of the **structure of subject classification**, in particular facet organization and hierarchy, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. (P2.3.9,3) 2.3.4 Understand **inclusive (hierarchically expanded) searching** and be able apply inclusive searching in any system. (P2.3.9,3.4)
Practical significance	Inherited from Part 5 plus Inclusive searching is an essential technique for achieving high recall.

In-lecture exercise: Retrieval of documents in a sample collection

The sample collection consists of about 200 documents on transportation and traffic and is indexed using the index language shown on the following pages (same as the index language used in Assignment 11, *Request-oriented indexing*).

Query statement: I need information on vehicles used in rail transport

Query formulation: E2.1 Vehicles AND B1.2 Rail transport

Search in a printed index: Look for document numbers listed for both E2.1 Vehicles and B1.2 Rail transport (they are marked with = in the entry for B1.2).

E2.1 Vehicles 10, 12, 13, 24, 25, 26, 30, 36, 40, 46, 47, 50, 53, 54, 58, 59, 62, 64, 70, 76, 77, 79, 80, 81, 85, 91, 92, 94, 95, 100, 101, 102, 103, 104, 105, 106, 108, 109, 1 10, 116, 118, 121, 122, 126, 127, 132, 133, 134, 138, 148, 150, 151, 153, 155, 168, 169, 170, 171, 173, 174, 176, 178, 180, 181, 186, 187, 188, 191, 192, 193, 194, 199, 202, 204, 205, 207, 210, 21 1, 212, 213, 214, 216, 218, 219, 322, 330, 332, 333, 336, 337, 340, 346, 347, 353, 354, 355, 356, 357, 358, 362

B1.2 Rail transport, general references =10, =24, 34, 41, 42, 44, =50, 89, 114, =126, 140, 149, 166, 184, =191, 195, =213, 310, 334

B1.2.1 Local rail transit 27, 56, 87, 99, =108, 111, 120, 123, 129, 130, =213, 327, 345, 350, =356

B1.2.2 Intercity railroads =10, 30, =46, =62, =64, =79, 82, 84, 97, =102, =108, 1 14, =122, =132, 156, 177, =213, 341, 362

Question: Did this search find all relevant documents? Would need to also find the documents in common to *E2.1 AND B1.2.1* and to *E2.1 AND B1.2.2*, in other words, search for

E2.1 AND (B1.2 gen. ref OR B1.2.1 OR B1.2.2)

Additional index entry to make this easier (B1.2 hierarchically expanded = B1.2 inclusive) includes all documents listed under B1.2.1 and B1.2.2 as well, index entry has Boolean OR "built in".

B1.2 Rail transport, inclusive =10, =24, 27, =30, 34, 41, 42, 44, =46, =50, 56, =62, =64, =79, 82, 84, 87, 89, 97, 99, =102, =108, 111, 114, 120, =122, 123, =126, 129, 130, =132, 140, 149, 156, 166, 177, 184, =191, 195, =213, 310, 327, 334, 341, 345, 350, =356, =362

Over

Pink (Only in seated class)

Search in a peek-a-boo file (some samples distributed). A small retrieval experiment

Like a printed index, but with more manipulative power. Today one uses computers. Here just used for visualization.

Each descriptor has its own card. Each document number has a position on the card.
In a printed index, the applicable document numbers are listed after the descriptor.
In a peek-a-boo file, the applicable document number positions are punched on the descriptor card.
In this particular implementation, document numbers are read off as follows: Find the column number (printed in tiny print), for example **12**). Find the row number (large single digits printed across each row, punched out if the position is punched), for example **6**. The document number is **126**. (One of the first uses of peek-a-boo cards: A bird guide. Retrieval of birds based on their features.)

This peek-a-boo file provides for inclusive searching: Each descriptor that has narrower descriptors under it has two cards: An **inclusive card** that includes all the documents from the narrower descriptors as well, and a **general references card** that includes only the documents indexed by the descriptor itself.

To find documents for

E2.1 Vehicles AND B1.2 Rail transport, inclusive

superimpose the two cards and read off the document numbers from the holes that still appear (document numbers in common to both cards).

We will broaden and narrow the search to observe the effects of hierarchy.

Lessons learned from
General reference search (find only documents that deal with all or most of a broad topis and inclusive search (include documents that deal with any minute subtopic of the broad topic) Difficulty of relevance judgments; titles are not enough. "The relevance judge does not live on titles alone." Important for reference librarians and end users. Reference librarians mus assist end users in determining relevance. Computer systems can help, for example by showing the abstract with query words (and their synonyms etc.) highlighted. In retrieval testing and evaluation, bad relevance judgments cause the results to be misleading for system design. Through better query formulation both recall and discrimination/precision can be improved (the universals inverse relationship is a myth) Search of the type Specific topic OR Broader topic, **general reference** A user searching for methods of *vegetable pickling* can find useful information in a textbook or encyclopedia of *food science* or in general encyclopedia. Checking indexing by making sure at least one descriptor from every facet (if only *not applicable*) is used.

File builders and searchers classification display

The descriptors shown in italics with numbers D1.xx are precombined descriptors. Each system using this index language would decide whether to use these precombined descriptors (such as *D1.20 Aircraft*) or whether to index with the corresponding elemental descriptors (in the example D1 Air transport and E6 Vehicles) instead. **In assignment 11 only elemental descriptors are used for indexing.**

Outline: Facets

B	**Mode of transportation**
E	**Transportation system elements**
F	Power supply for vehicles
G	Type of propulsion
H	Materials to build facilities or vehicles
J	**Passenger transport vs. freight transport**
K	Traffic operations
L	Transportation providers
M	Creation of traffic systems and components
N	Organization, administration
Q	General and other concepts
R	Geographic range
S	Geographic location

The three facets used for arrangement are shown in **bold**.

In the full display

+ signifies descriptors that have Narrower Terms under them

+A	**Transportation and traffic, inclusive**	
A	Transportation and traffic, general references	
+B	**Mode of transportation, inclusive**	
B	. Mode of transportation, general references	
+B1	**. Ground transport, inclusive**	
+B1	. . Ground transport, general references	
B1.1	. . Road transport	28*
B1.2	. . Rail transport, inclusive	
B1.2	. . . Rail transport, gen. ref.	25
B1.2.1	. . . Local rail transit BT R1[1]	26
B1.2.2	. . . Intercity railroads BT R2.1[1]	27
B1.3	. . Pipeline transport	
B1.4	. . Pedestrian mode	
B1.5	. . Multi-modal ground transport	
+B2	**. Water transport, inclusive**	
B2	. . Water transport, general references	
B2.1	. . Inland water transport	31
B2.2	. . Ocean Transport	32
+B3	**. Air transport, inclusive**	
B3	. . Air transport, general references	
B3.1	. . Supersonic air transport	35
B4	**. Air cushion transport**	
B5	**. Multi-modal transport**	
B8	. Other specific modes of transportation	
B9	. Mode of transportation not applicable	

D, E Free for expansion

+E	**Transportation system components, inclusive**	
E	Transportation system components, gen. ref.	**15**
+E1	**. Traffic facilities, inclusive**	
E1	. . Traffic facilities, gen. references	**17**
E1.1	. . Traffic routes	18
E1.2	. . Traffic stations	19
E1.3	. . Stationary equipment	
+E2	**. Methods to move persons or freight, incl.**	
E2	. Methods to move persons or freight, g. r.	
E2.1	. . Vehicles RT F, G	20
E3	. . Containers	
E4	. . Self-transport	
E8	. Other concepts	
E9	. Transp. system components not applicable	

\+ Inclusive (hierarchically expanded, finds documents on all narrower terms as well)

* Edge-notched card hole no.

[1] B1.2.1 (= B1.2 : R1) and B1.2.1 (= B1.2 : R2.1) are pre-combined descriptors, narrower then B1.2 by combination

+F	**Power supply f. vehicles, incl**. RT E2.1
F	. Power supply for vehicles, gen. ref.
+F1	. Hydrocarbons, inclusive
F1	. . Hydrocarbons, general references
F1.1	. . Gasoline
F1.2	. . Diesel fuel
F1.3	. . Hydrocarbons from renewable sources
F5	. Electric power
F6	. Nuclear power
F7	. Animate power
F8	. Other power supply
F9	. Power supply not applicable
+G	**Type of propulsion, inclusive** RT E2.1
G	. Type of propulsion, gen. references
+G1	. Engine, inclusive
G1	. . Engine, general references
+G1.1	. . Combustion engine, inclusive
G1.1	. . Combustion engine, general ref.
G1.2	. . Steam engine
G2	. Turbines
G3	. Walking
G8	. Other type of propulsion
G9	. Type of propulsion not applicable
+H	**Materials to build facilities or vehicles, inclusive**
H	. Materials to build facilities or vehicles, g.r.
H1	**. Materials by composition, inclusive**
H1	. . Materials by composition, gen. ref
H1.1	. . Soils, aggregates
H1.2	. . Bitumen
H1.3	. . Cement, Concrete
H1.4	. . Ceramics, glasses
H1.5	. . Wood, paper
H1.6	. . Fibers, textiles
H1.7	. . Plastics
H1.8	. . Rubbers
H2	**. Materials by origin, inclusive**
H2	. . Materials by origin, general references
H2.1	. . Petroleum products
H3	**. Materials by use, inclusive**
H3	. . Materials by use, general references
H3.1	. . Marking or coating materials
H3.2	. . Adhesives, seals
H8	. . Other specific materials
H9	. . Material not applicable

+J	**Passenger vs. freight transport, incl.**	
J	. Passenger vs freight transport, g.r.	8
J1	. Passenger transport	9
+J2	. Freight transport, inclusive	
J2	. . Freight transport, general references	11
J2.1	. . Transport of material of heavy weight	12
J2.2	. . Transport of bulk material	13
J9	. Passenger vs. freight transport not applicable	
+K	**Traffic operations, inclusive**	
K	. Traffic operations, general references	
+K1	**. Traffic communication, control, safety, I.**	
K1	. . Traffic comm., control, safety, gen. ref.	4
K1.1	. . Traffic communications	
K1.2	. . Traffic control	6
K1.3	. . Traffic safety	7
+K2	**. Routes and schedules, inclusive**	
K2	. . Routes and schedules, general references	
K2.1	. . Routes, route systems, traffic networks	2
K2.2	. . Schedules	
K3	. Handling, loading, unloading	
K8	. Other specific traffic operations	
K9	. Traffic operations not applicable	
+L	**Transportation providers, inclusive**	
L	. Transportation providers, gen. references	
L1	. Organizations, companies	
L2	. Personnel, operators	
L9	. Transportation providers not applicable	
+M	**Creation of traffic systems&comp.**,incl	
M	. Creation of traffic systems&components, g.r.	
+M1	**. Research, design, and evaluation, inclusive**	
M1	. . Research, design, and evaluation, g. ref.	
M1.1	. . Research and development	
M1.2	. . Planning	
M1.3	. . Design	
M1.4	. . Testing, demonstration, evaluation	
M2	. Manufacturing, construction	
M3	. Acquisition	
M4	. Training	
M5	. Maintenance	
M8	. Other specific activities in system creation	
M9	. System creation not applicable	
+N	**Organization, administration, incl.**	
N	. Organization, administration, gen. references	
N1	. Administration, management	
N2	. Costs, financing	
N3	. Marketing	
N4	. Legal aspects	
N8	. Other specific topics in organization	
N9	. Organization, administration not applicable	

+Q	**General and other concepts inclusive**	
Q	. General and other concepts, gen. references	
Q1	. Traffic flow	
Q2	. Simulation	3
+Q3	**. System characteristic, inclusive**	
Q3	. . System characteristics, general references	
Q3.1	. . Noise, vibration	
Q3.2	. . Pollution	
Q3.3	. . Quality, performance	
Q3.4	. . Durability, life, reliability	
Q3.5	. . Demand, use	
Q3.6	. . Human characteristics	
Q3.7	. . Community characteristics	
Q3.9	. . Other system characteristics	
+Q4	**. Small vs large capacity, inclusive**	
Q4	. . Small vs. large capacity, gen. references	
Q4.1	. . Small capacity	
Q4.2	. . Large capacity	
+Q5	**. Civilian vs military, inclusive**	
Q5	. . Civilian vs military, general references	
Q5.1	. . Civilian	
Q5.2	. . Military	
Q99	. Other concepts not applicable	
+R	**Geographic range, inclusive**	
R	. Geographic range, gen. references	
+R1	**. Local systems, inclusive**	
R1	. . Local Systems, general references	
+R1.1	. . Urban systems, inclusive	
R1.1	. . Urban systems, general references	
R1.2	. . Rural systems	
+R2	**. Beyond local systems, inclusive**	
R2	. . Beyond local systems, general references	
R2.1	. . Interurban systems	
R2.2	. . State-wide systems	
R2.3	. . National systems	
R2.4	. . International systems	
R8	. Other specific range	
R9	. Geographic range not applicable	
+S	**Geographic location, inclusive**	
S	. Geographic location, general references	
+S1	. North and Central America, inclusive	
S1	. . North and Central America, gen. ref.	
S1.1	. . Canada	
S1.2	. . U.S.	
S1.3	. . Central America	
S2	. South America	
S3	. Europe	
S4	. Asia	
S5	. Australia	
S6	. Africa	
S8	. Other geographic locations	
S9	. Geographic location not applicable	

In-lecture exercise: Retrieval access and hierarchy

Below are six documents which were indexed in the request-oriented approach you used in Assignment 11. Each descriptor is on a separate line. Using the hierarchy of the index language in the *File builder's and searcher's display* (see preceding pages), do the following:

1. For each descriptor (index term), list the descriptor(s) under which the document should be found on the basis of this index term.

2. Give some query formulations retrieving the document. The query formulations should illustrate how a search for a combination of two broad concepts finds documents indexed by more specific concepts.

Document 1 is a filled-in example

Document 1
Automatic control of freeway ramp traffic, P.J.ATHOL. SAE—Analysis & Control of Traffic Flow Symposium—Conf Proc. Jan 9-10 1968 paper 680172 p 61-5.

Major problem in operating transportation system is traffic overloading demands at peak periods; expressway Surveillance Project was formed to improve efficiency of highway system through application of electronic automation and traffic engineering to problem of traffic congestion.; by providing means for quick response in case of accidents and fast removal of hindrances. Volume capacity of freeways was effectively increased during peak periods; use of ramp metering controls achieved reduction in delay, safer merging characteristics, and reduced freeway accidents.

Descriptors assigned		**Descriptors under which the document should be found**
B1.1	Road transport	B1.1; B1 inclusive
E1.1	Traffic routes	E1.1; E1 inclusive
K1	Traffic communication, control, safety	K1 g**en. ref**.; K1 inclusive
M1.2	Planning	M1.2; M1 inclusive
M1.3	Design	M1.3; M1 inclusive,
M1.4	Testing, demonstration, evaluation	M1.4; M1 inclusive,
Q1	Traffic flow	Q1

Query formulations

B1 Ground tr., incl. AND E1 Tr. fac., incl.
B1.1 Road tr. AND K1 Tr. comm., control, safety incl.

Document 2

Antwerp's new container dock, K.W.Flitcroft for the Antwerp Harper Committee. Dock & Harbor Authority v 49 n 571 May 1968 p 28-30.

Dock described is protected by locks from rise and fall of tides; spreader is employed in lifting of containers and is adaptable in spread to handle both long and short types; containers can be stored on quay and special connections for powering of plants of refrigerated containers are set in concrete paving every 10 ft.; set of rail tracks runs along quay between high legs of container cranes to bring rail-hauled containers directly for lifting off.

Descriptors assigned		**Descriptors under which the document should be found**
B2.2	Ocean transport	
E1.2	Traffic stations	
J2	Transport of freight, material, cargo	
K3	Handling, loading, unloading	
R2.1	International system	
S3	Europe	
		Query formulations

Document 3. **Rolling Stock for London Transport's Victoria Line**

Descriptors assigned

B1.2.1 Local rail transit
E2.1 Vehicles
F5 Electric power
G1 Engine
M1.3 Design
M3 Acquisition
Q3.9 Other characteristics (automation)
R1.1 Urban systems
S3 Europe

Descriptors under which the document should be found

Query formulations

Document 4. **Air Transport 1975 and Beyond - Systems Approach**

Descriptors assigned

B3 Air transport
E Transportation system components
J Passenger transp. vs. freight transp.
K Traffic operations
M1.1 Research and development
N Organization, administration
Q5 Civilian vs. military
R2.3 National systems
S1.2 U.S.

Descriptors under which the document should be found

Query formulations

Document 5. **Technical and Economic Prospects of Air Cargo Traffic**

Descriptors assigned		**Descriptors under which the document should be found**
B3	Air transport	
E2.1	Vehicles	
F1.3	Hydrocarbons from renewable sources	
G2	Turbines	
J2	Transport of freight, material, cargo	
K3	Handling, loading, unloading	
M1	Research, design, and evaluation	
Q1	Traffic flow	
Q3.5	Demand, use	
Q3.9	Other characteristics (automation)	
R2.4	International system	
S	Geographic location	**Query formulations**

Document 6. **United States Subway Requirements 1968-1990**

Descriptors assigned		**Descriptors under which the document should be found**
B1.2.1	Local rail transit	
E1	Traffic facilities	
M2	Manufacturing, construction	
N2	Costs, financing	
N3	Marketing	
Q2	Simulation	
Q3	System characteristic	
R1.1	Urban systems	
S1.2	U.S.	**Query formulations**

red

Things to do in Week 10, W March 30 - April 6

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	Assignment 11: Request-oriented indexing (assigned L8.2b, Mar. 23)	☐
	Assignments due W April 6 ☐ required ○ optional	√
Assignment(s)	Assignment 12.1: Semantic factoring (1.5 hrs) (assigned. L 9.1, Mar. 30)	☐
	Assignment 12.2: Building a hierarchy of elemental concepts (1.5 hrs) (ass. L 9.1, Mar. 30)	

	New topics this week

	10.1 Constructing a hierarchy from facet combination (70 min) (Strand 1. General principles of KOS) ◑ ◑	
Readings	*1 Lecture 10.1 Objectives etc. (pink). Also have a look at Assignments 12.3a-12.3b.*	☐
Lecture	**Lecture 10.1 slides** (105 min)	☐

	10.2a Brief introduction to Assignments 13.1 - 13.4 (20 min) (Strand 2. Analysis of highly used KOS)		
Readings	1	Lecture 10.2a Objectives etc. (pink)	☐
	2	Assignments 13.1-4 Analysis of and practice with Knowledge Organization Systems (KOS) Introduction (gold), p. ~157-158, look over the materials that follow, assignment packet p. ~159 - 196	☐
	3	Assignment 13.1 DDC Introduction (gold), p. ~197, look over pages that follow, assignment packet p. ~198 - 234 (just to get a feel for what is there)	☐
	4	**Model Catalog** (very first Reading in packet) (gives examples how different KOS are used indexing)	☐
	5	Needham, Christopher D. **Organizing knowledge in libraries: An introduction to information retrieval**. 2nd ed. New York: Seminar Press, 1971. Ch. 7 **Review of classification principles**, p. 109-132 (facet analysis & synthesis) Ch. 8 Schemes of classification, esp. p. 133, **DDC**, p.140-152; **LCC**, p. 163-168 Still the clearest exposition of timeless classification principles. **Required**	☐
	6	Chan, Lois Mai. **Cataloging and Classification: An Introduction. 3. ed**. Scarecrow; 2007. 600p. ISBN-10: 0810860007 Part 3. Subject access in library catalogs, p. 153 - 254. General, LCSH, Sears SH, MeSH Part 4. Classification, p. 255 - 400. General, DDC, LCC, NLM Classification, others	☐
Lecture		**Lecture 10.2a slides** (20 min)	☐

over

	10.2b Intro. & in-class exercise: Assignment 13.1 Dewey Decimal Classification (70 min) (Strand 2. Analysis of highly used KOS)		
Readings	1	Lecture 10.2b Objectives etc. (pink) and Reading 3 from 10.2a	☐
	2	**On the new ed. 23**: www.oclc.org/dewey/versions/print/default.htm. Find on Web, **Optional**, after	☐
	3	Chan, Lois Mai; Comaromi; Mitchell, Joan Dewey Decimal Classification: Principles and Application. 3. ed. Dublin, Ohio : OCLC, 2003. xi, 216 p. Find in a library. **Optional**, after	☐
	4	London Education Classification (LEC) pdf	☐
Lecture	**Lecture 10.2b slides** (70 min)		☐

Learning blog	Learning blog Week 10 due W April 6	○

	Assignments assigned April 6	
Assignments assigned	Based on Lecture 10.1 ‣**12.3a** **Practice: Hierarchy from facet combination with education concepts** (2 hours) (due `Apr. 13)  ‣**12.3b** **Real: Hierarchy from facet combination with concepts from Assignments 12.1 / 12.2** (2 hrs) (due `Apr. 20) Start this after you received the answer key for Assignment 12.3a (Based on Lecture 10..2b) ‣**13.1** **Dewey Decimal Classification (DDC)** (4 hours) (due `Apr. 20) If possible, start working on the DDC worksheet `Apr. 6 with work in study group, then listen to presentation (online) or participate in class (seated) , complete by `Apr. 20	

pink

Lecture 10.1 Interactive

Hierarchy from Facets

April 6

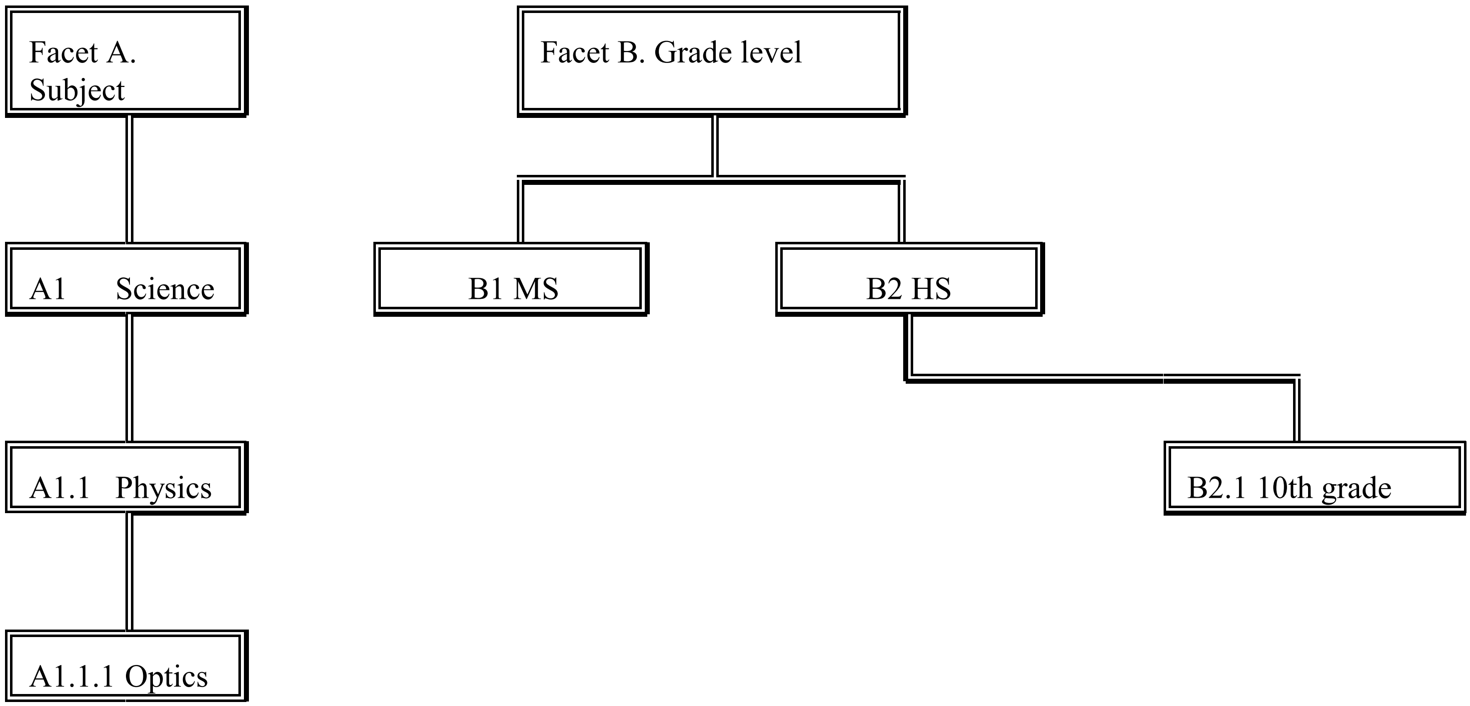
Information system of instructional materials. Two facets, only between-facet combinations
Understand complex hierarchies that result from combining hierarchically structured facets.

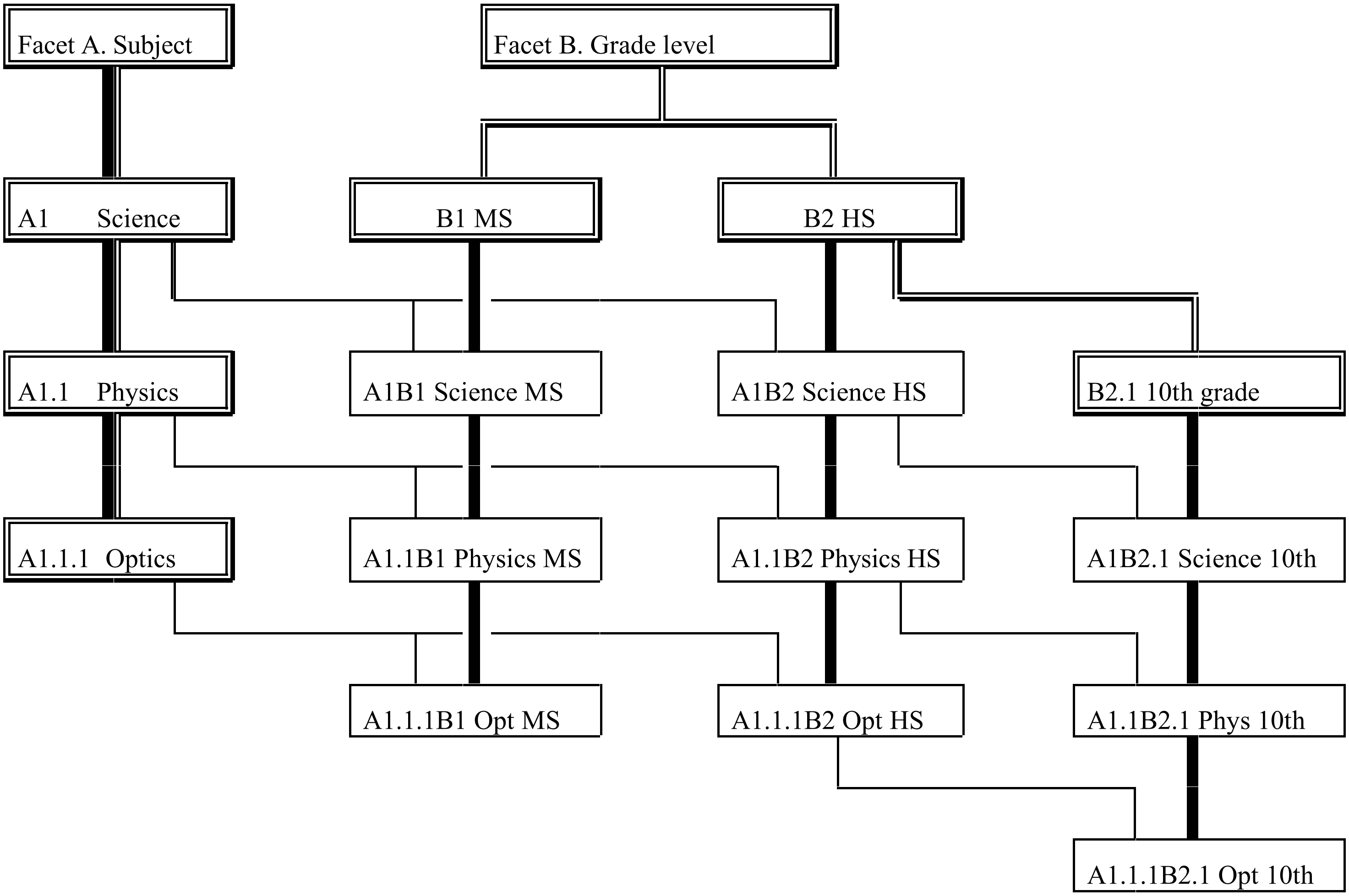
Learning objectives (from Part 5, p. ~285)	2.3 Understand the principles of the **structure of subject classification**, in particular facet organization and hierarchy, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. (P2.3.9,3)
	2.3.5 Understand how the relationship between two compound concepts can be inferred by comparing the concept formulas. Understand the complex hierarchies that result from combining hierarchically structured facets. Understand compound concepts as nodes in a semantic network. Be able to apply this understanding to broadening or narrowing query formulations and to the analysis of classification systems such as DDC, LCC, and subject heading systems, such as LCSH. (P2.3.9,3.5)
Practical significance	Basis for understanding search Basis for understanding structure of DDC, LCC, LCSH, and similar systems

Process	Step 1:	Form all possible between-facet combinations (do not combine with facet heads).
	Step 2:	Find all hierarchical relationships. (Specifying all BT one level up defines the hierarchy completely.)
	Step 3:	Represent the hierarchy
	Step 3a:	– as a two-dimensional graph
	Step 3b:	– as a linear arrangement with indention plus cross-references (in outline form).
Application to retrieval		In a system using only elemental descriptors In a system using precombined descriptors with multiple entry (such as LC Subject Headings) In a system using precombined descriptors with single entry (such as Library of Congress Classification)

Note: Look at p. ~373 Go through the exercise (seated) or the slides (online) before looking at the result on p. ~374

pink


Facet A. Subject
A1 Science
A1.1 Physics
A1.1.1 Optics
Facet B. Grade level
B1 MS
B2 HS
B2.1 10th grade



A Facet A. Subject

. A1 Science
. . A1B1 Science MS NT A1.1B1; BT B1
. . A1B2 Science HS NT A1.1B2; BT B2
. . . A1B2.1 Science 10th grade NT A1.1B2.1; BT B2.1
. . A1.1 Physics
. . . A1.1B1 Physics MS NT A1.1.1B1; BT A1B1
. . . A1.1B2 Physics HS NT A1.1.1B2; BT A1B2
. . . . A1.1B2.1 Phys 10th gr NT A1.1.1B2.1; BT A1B2.1
. . . A1.1.1 Optics
. . . . A1.1.1B1 Optics MS BT A1.1B1
. . . . A1.1.1B2 Optics HS BT A1.1B2
. . . . . A1.1.1B2.1 Optics 10th grade BT A1.1B2.1

B Facet B. Grade level

. B1 MS NT A1B1
. B2 HS NT A1B2
. . B2.1 10th grade NT A1B2.1

A Facet A. Subject

. A1 Science NT B1A1, B2A1
. . A1.1 Physics NT B1A1.1, B2A1.1
. . . A1.1.1 Optics NT B1A1.1.1, B2A1.1.1

B Facet B. Grade level

. B1 MS
. . B1 A1 MS Science BT A1
. . . B1 A1.1 MS Physics BT A1.1
. . . . B1 A1.1.1 MS Optics BT A1.1.1
. B2 HS
. . B2 A1 HS Science NT B2.1A1; BT A1
. . . B2 A1.1 HS Physics NT B2.1A1.1; BT A1.1
. . . . B2 A1.1.1 HS Optics NT B2.1A1.1.1; BT A1.1.1
. . B2.1 10th grade
. . . B2.1 A1 10th grade Science BT B2A1
. . . . B2.1 A1.1 10th grade Physics BT B2A1.1
. . . . . B2.1 A1.1.1 10th grade Optics BT B2A1.1.1

pink **Lecture 10.2a** *April 6*

Brief Introduction to Assignments 13.1-4. Examination of KOS

Also Lectures 10.2b, 11.1, 11.2, 12.2, 13.1

Analysis of Knowledge Organization Systems (ontologies, classification schemes, thesauri, etc.) based on their hierarchical structure, facet structure, and citation order. See the calendar or the assignment page for schedule of assignment activities and due dates.

Learning objectives	0 **Understand the application of general information structure principles to the subject cataloging/indexing/coding of documents and other entities** Understand the issues in creating and using a good classification / index language / coding scheme This is a general heading encompassing 1 and 2. (P2.3.7) Specifically here: Solidify in your mind the general principles underlying all KOS and develop a more complete understanding of the general concepts of classification structure by applying them to several concrete schemes.
	2.2 Understand the **functions of classification (more broadly, Knowledge Organization Systems, KOS) for a wide variety of tasks in a wide variety of systems**. (P2.3.9,2)
	2.3 Understand the principles of the **structure of subject classification**, in particular facet organization and hierarchy and of methods for presenting this structure, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. (P2.3.9,3) Here: Improve this understanding by examining many examples.
	2.4 **Know and understand actual schemes**. Get an overview of the wide range of different types of classification schemes and other Knowledge Organization Systems (KOS) and their uses. Grasp the structure of these schemes by applying the general conceptual framework developed earlier in the course to their analysis. (P2.3.9,4)
	2.4.1 Be acquainted with major KOS used on the Web or in American libraries, such as Yahoo (or DMOZ) Classification, Dewey Decimal Classification, Library of Congress Classification, Library of Congress Subject Headings ERIC Thesaurus (P2.3.9,4.1)
	2.4.2 Have started to learn to use some KOS used in American libraries for cataloging (indexing) and query formulation for searching. (P2.3.9,4.2)
	2.4.3 Be acquainted with a number of other KOS illustrating the wide variety of KOS. (P2.3.9,4.3)

over

pink

Practical significance	• A good working knowledge of faceted classification principles is important for the conceptual analysis of queries as a basis for developing good query formulations in any system. • Knowledge of specific schemes is important for searching catalogs and indexes based on those schemes (including catalogs of Web documents). • Knowledge of the variety of schemes that exist for different purposes is important for being able to work in many different applications and for recognizing where classification could be useful. • Knowing the general principles that underlie all KOS will enable you to evaluate KOS, to improve existing KOS, and to build new KOS (after taking UBLIS 514 Indexing and Surrogation). • You will be able to "sell your skills" to a wider variety of organizations, increasing opportunities for work

Turn to Assignments 13.1-4, Assignments p. ~157
Read p. ~157 (information for all assignments 13.1-13.4).

Listen to presentation

Lecture 10.2b, going through the Dewey lecture/PowerPoint will make this clearer.

pink **Lecture 10.2b** ***April 6***

Introduction and in-lecture exercise: Assignment 13.1. Dewey Decimal Classification (DDC)

.

Learning objectives **Inherited from 10.2a**, re-read there	Specifically, gain a first understanding of the Dewey Decimal Classification with emphasis on the general structure. Knowledge where specific subjects are placed in DDC will come with practice.
Practical significance **Inherited from 10.2a**	Specifically, DDC is used in most public and school libraries in the US. It is also used in many other countries. OCLC, the owner of DDC, is pushing its use for organizing subject directories on the Web.

Read the golden page for Assignment 13.1, DDC

With a study group, start on the DDC worksheet and do as much as you can in no more than two hours (perhaps less). The idea is to get at least a start on figuring things out for yourself without banging your head against the wall.

Then go to the presentation, look at the pdf while listening to the mp3 audio

pink

red

Things to do in Week 11, W April 6 - April 13

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	Assignment 12.1: Semantic factoring (assigned L9.1, `Mar. 30) Assignment 12.2: Building a hierarchy of elemental concepts (assigned L9.1, `Mar. 30)	☐
	Assignments due W April 13 ☐ required ○ optional	√
Assignment(s)	Assignment 12.3a: Practice Hierarchy from facet combination with education concepts (2 hours) (assigned. L10.1, `Apr. 6)	☐

	New topics this week	

	11.1 Introduction and in-class exercise on Assignment 13.2 Yahoo (70 min) (Strand 2. Analysis of existing KOS) ⓘ ⓘ	
Readings	1 Look at https://web.archive.org/web/20131231053351/http://dir.yahoo.com (No Lecture Objectives etc.) Also have a look at Assignment 13.2. I may replace this with the Wikipedia classification	☐
Lecture	**Lecture 11.1 slides** (40 min)	☐

	11.2 Introduction and in-class exercise Assignment 13.2 LCC (65 min.) (Strand 2. Analysis of highly used KOS) ⓘ ⓘ	
Readings	(No Lecture Objectives etc.)	
	1 Needham, Ch. 8, **Schemes of classification**, p. 163-168 LCC (also **assigned for 10.2a**)	☐
	2 Look at Excerpts from LCC class H and Z for Lecture 11.2 (get pdf from the website)	☐
	3 Look at LCC volumes (if you can find them in an academic library)	○
	4 Chan, Lois Mai. **A guide to the Library of Congress Classification**. 5th ed. Englewood, Colo.: Libraries Unlimited, 1999. Z696.U4C47 1999 Find in a Library p. 1-14 The history of the classification p. 14-19 Focus and use p. 23-47 Principles, structure, and format (skim the examples). For further study, read more from this book	○
	5 Fundamentals of Library of Congress Classification. Manual for Trainees. ca 2007. www.loc.gov/catworkshop/courses/fundamentalslcc/pdf/classify-trnee-manual.pdf	○
Lecture	**Lecture 11.2 slides** (55 min)	☐

Learning blog	Learning blog Week 11 due W `April 13	○

	Assignments assigned W April 13	
Assignments assigned	**One of** ▸Assignment 13.2 **Yahoo** (due `Apr. 20) **OR** ▸Assignment 13.2 **LCC** (due `Apr. 20) **OR** ▸Assignment 13.2 **DDC2. More practice with DDC** (due `Apr. 20)	

pink **Lecture 11.1** ***April 13***

Introductory discussion and in-lecture exercise on Assignment 13.2 Yahoo

We will start going through the worksheet, index a document, and formulate a query.

Lecture 11.2 ***April 13***

Introductory discussion and in-lecture exercise on Assignment 13.2 LCC

We will start going through the worksheet, index a document, and formulate a query.

pink

red.

Things to do in Week 12, W April 13 - April 20

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	Assignment 12.3a Practice: Hierarchy from facet combination with education concepts (assigned L10.1, `Apr. 6)	☐
	Assignment 13.1: Dewey Decimal Classification (assigned L10.2b, `Apr. 6)	☐
	Assignments due W April 20 ☐ required ○ optional	√
Assignment(s)	Assignment 12.3b Real: Hierarchy from facet combination with concepts from Assignments 12.1/12.2 (assigned L10.1, `Apr. 6)	☐
	Assignment 13.2: Yahoo OR LCC OR DDC2 (assigned. L11.1/2, `Apr. 13)	
Preparation for the final	Submit questions / requests for topics to be included in the final review to dsoergel@buffalo.edu between now and Sunday May 1 midnight	○
	Submit answers to sample final questions for feedback (notes for Lecture 14, p. ~413+) to dsoergel@buffalo.edu between now and Sunday May 1 midnight	○

	New topics this week	

	12.1 KOS conceptual structure 2: Application to database organization (implementation) (70 min) (Strand 1. General KOS principles)	
Readings	1 Lecture 12.1 Objectives etc. (pink). Also have a look at Assignments 13.3-13.4.	☐
	2 Textbook Ch. 15 **Index language structure 2: database organization (for discussion)**	☐
Lecture	**Lecture 12.1 slides** (70 min)	☐

over

	12.2a Brief discussion of LCSH and Assignment 13.3 (5-10m) (Strand 2. Analysis of highly used KOS)	
Readings	1 Needham, Ch. 10, *The alphabetic subject catalog*, p. 199-223. In reading packet	○
	2 Chan, Lois Mai 2007. **Cataloging and classification**: **An introduction.** Lanham, Md. : Scarecrow, 2007. Chapter 8 on LCSH. Find in a library.	○
	3 Chan, Lois Mai 2005 **Library of Congress Subject Headings. principles of structure and application**. 4. ed. Englewood, CO: Libraries Unlimited; 2005. Find in a library. **F**or further study	○
	4 Perreault, Jean M. 1979 **Library of Congress Subject Headings: A New Manual**. International Classification 1979 Nov.; 6(3):158-169. Extensive review of an earlier version of Chan's book. Gives a good feel for some of the problems in LCSH. For further study.	○
Lecture	No online materials. **Ask questions**	☐

Learning blog	Learning blog Week 12 due W `April 20	○

	12.2b Brief discussion of ERIC and Assignment 13.4 (5-10m) (Strand 2. Analysis of highly used KOS)	
Readings	*No assigned readings*	
Lecture	No online materials. **Ask questions**	☐

	Assignments assigned April 20	
Assignments assigned	▸Assignment 13.3, **Library of Congress/Sears Subject Headings** (LCSH) (5 hours), due `Apr. 27) ▸Assignment 13.4, **ERIC Thesaurus** (3 hours) (due `Apr. 27)	

pink **Lecture 12.1** ***April 20***

KOS conceptual structure 2:
Application to database organization (implementation) (Organizing Info., Ch. 15)

Learning objectives (from Part 5, p. ~285)	2.3 Understand the principles of the **structure of subject classification**, in particular facet organization and hierarchy, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. (P2.3.9,3)
	2.3.5 Understand how the relationship between two compound concepts can be inferred by comparing the concept formulas. Understand the complex hierarchies that result from combining hierarchically structured facets. Understand compound concepts as nodes in a semantic network. Be able to apply this understanding to broadening or narrowing query formulations and to the analysis of classification systems such as DDC, LCC, and subject heading systems, such as LCSH. (P2.3.9,3.5)
	2.3.6 Solidify understanding of postcombination and precombination - more generally, the degree of precombination — and how they relate to the retrieval mechanism used. Understand the nature of precombined descriptors as new nodes in a semantic network. (P2.3.9,3.6)
	2.3.7 Understand the effect of precombination on index language structure and searching and be able to apply this understanding to the analysis of classification schemes such as DDC and LCC and improved searching with such schemes. (P2.3.9,3.7)
	2.3.8 Be able to match the index language structure to the database organization and search mechanism available. (P2.3.9,3.8)
	2.3.9 Understand the access mechanisms that help a user find the proper descriptors in a large classification scheme with many precombined descriptors, in particular **cross-references** and a **descriptor-find index**. (P2.3.9,3.9)
	2.3.10 Understand principles of meaningful arrangement of search results. (P2.3.9,3.10)
Practical significance Inherited from Lect. 9.1-11.1 plus these XXX no 9.1-11.1	In conjunction with Ch. 14, this lecture establishes the foundation for understanding • the structure of systems used in libraries — and increasingly for the arrangement and display of electronic information — such as the Dewey Decimal Classification (DDC), the Library of Congress Classification (LCC), the Yahoo classification, and the Library of Congress Subject Headings (LCSH); • the structure of Web directories designed for browsing; • ad-hoc arrangement of retrieval results based on the analysis of noun phrases as compound concepts, as in the next-generation Web search engines.
Discussion question	Consider the design of an interface to a public-access online catalog in an academic library that would assist users in finding the appropriate LC class number and the appropriate LC subject headings.

Over

pink

12.1a. Introductory example

12.1b. Discussion of Textbook, *Chapter 15*

Section 15.1 and 15.2	Further examination and explication of postcombination vs. precombination of the concepts chosen as descriptors and their relationship to database organization and search mechanism. Interpretation of postcombination and precombination in terms of the entity-relationship approach and semantic networks (see figures on the following two pages). Examples of applying these concepts to a better understanding of index languages such as the Library of Congress Classification and the Library of Congress Subject Headings. In-lecture exercise 12.1c: Hierarchical inheritance and precombined descriptors
15.3	Unified view of classification and indexing
15.4	Emphasis on looking at precombination as a matter of degree.
15.5	Methods for organizing an index language for access. Emphasis on understanding the idea of a descriptor-find index. Section 15.5.2 on how then arrangement of precombined concepts is important for understanding classification schemes and for arranging search output or any type of information, in print or online.
15.6	A look into the future: the idea of a conceptually unified index language for different search mechanisms.

12.1c. In-lecture exercise: Hierarchical inheritance and precombined descriptors

12.1d. In-lecture exercises concluding Lectures 8.1-10.1 and 12.1

12.1d1 Vocabulary control and hierarchical structure

12.1d2 Conceptual analysis and synthesis

Discussion question on OPAC (Online Public Access Catalog) interface

12.1c. In-lecture exercise: Hierarchical inheritance and precombined descriptors

Reorganize the database of six document records (from textbook p. 294) to take up less storage space. Step1, finding documents that are alike, already done. Answer on next page

Original database		DB reorganized using hierarchical inheritance
Document 1	Method of instruction Reading First Grade Good	**Document 1**
Document 5	Method of instruction Reading First Grade Good	**Document 5**
Document 3	Method of instruction Reading First Grade Bad	**Document 3**
Document 18	Method of instruction Reading First Grade Bad	**Document 18**
Document 6	Traffic station Ocean transport Freight Bad	**Document 6**
Document 7	Traffic station Ocean transport Freight Bad	**Document 7**

Answer. Identify what is in common to each group of documents and create a new "master record" with that common information. Each document record inherits from the appropriate master record. In this exercise, the information consists of elemental concepts assigned as index terms. Each master record can be considered a class, a precombined descriptor. Indexing a document with a class / precombined descriptor gives the same information as indexing it with several elemental descriptors

Original database	DB reorganized using hierarchical inheritance
	Group #1 Method of instruction Reading First Grade Good
Document 1 Method of instruction Reading First Grade Good	**Document 1** Inherits from Group #1 Also: Document 1 indexed with Group #1
Document 5 Method of instruction Reading First Grade Good	**Document 5** Inherits from Group #1
	Group #3 Method of instruction Reading First Grade Bad
Document 3 Method of instruction Reading First Grade Bad	**Document 3** Inherits from Group #3
Document 18 Method of instruction Reading First Grade Bad	**Document 18** Inherits from Group #3
	Group #5 Traffic station Ocean transport Freight Bad
Document 6 Traffic station Ocean transport Freight Bad	**Document 6** Inherits from Group #5
Document 7 Traffic station Ocean transport Freight Bad	**Document 7** Inherits from Group #5

12.1c. In-lecture exercise: Hierarchical structure

The following is a list of terms that have occurred in query statements and in document titles. Organize it for purposes of information retrieval.

Book Campaign Candidate Department of State Elections Foreign Office Issue Journal Movement Periodicals Roll-call vote Running for Governor Running for Office State Department Vote	

This task calls on you to apply your knowledge from Organizing Information, Chapters 12-15. Therefore, no further guidelines are provided. (You may have to do this on your own on a much larger scale, in real life.) Since the list of terms is so small, facet analysis and synthesis is not required in this task.

12.1d2. In-lecture exercise: Conceptual analysis and synthesis

Organize the following list of terms for purposes of information retrieval.

Terms to work on	**Additional terms** (just to think about)
U.S. Congress	Foreign Office
State Court	British Parliament
County administration	United Nations
State legislature	Prime minister
Federal court	House of Commons
U.S. Senate	House of Lords
U.S. House of Representatives	UN Secretary-General
State administration	UN Security Council
State senate	UN General Assembly
State assembly	World Court

Procedure: Facet analysis and synthesis

Step 1:	Factor concepts into semantic components, resulting in elemental concepts
Step 2:	Organize the resulting elemental concepts in facets
Step 3:	Combine the facets (form all combinations)

The resulting hierarchical structure is to be shown graphically as well as in a linear sequence with cross-reference.

Note: The combinations produced in step 3 show gaps in the original list of terms.

****

12.1d3 Discussion question

Consider the design of an interface to a public-access online catalog in an academic library that would assist users in finding the appropriate LC class number and the appropriate LC subject headings.

pink **Lecture 12.2a - b** ***April 20***

Lecture 12.2a Brief discussion of LCSH and Assignment 13.3

Lecture 12.2b Brief discussion of ERIC and Assignment 13.4

red

Things to do in Week 13, W April 20 - April 27

	Review answer keys ☐ required ○ optional	√
Review answer key(s)	Assignment 12.3b Real: Hierarchy from facet combination with concepts from Assignments 12.1/12.2 (assigned L10.1, `Apr. 6)	☐
	Assignment 13.2: Yahoo OR LCC OR DDC2 (assigned. L11.1/2, `Apr. 13)	☐

	Assignments due W April 27 ☐ required ○ optional	√
Assignment(s)	Assignment 13.3: **Library of Congress/Sears Subject Headings** (LCSH) (5 hours), (assigned. L12.2a, Apr. 20)	☐
	Assignment 13.4: **ERIC Thesaurus** (3 hrs) (assigned. L12.2b, `Apr. 20)	☐
Preparation for the final	Submit questions / requests for topics to be included in the final review to dsoergel@buffalo.edu between now and Sunday May 1 midnight	○
	Submit answers to sample final questions for feedback (notes for Lecture 14, p. ~413) to dsoergel@buffalo.edu between now and Sunday May 1 midnight	○

	New topics this week

	13.1 Exploration of Knowledge Organization Systems (70 min) (Strand 2. Analysis of highly used KOS) ➀ ➀	
Readings	1 Lecture 13.1 Objectives etc. (pink) and *Schemes to be examined* (p. ~395- 398, white)	☐
	2 The big stack of Reading s for Lect. 13.1 (End of Reading Packet, blue dividers) Have a look at them beforehand; not meant for word-by-word reading.	☐
	3 Krathwohl 2002, **A revision of Bloom's Taxonomy: An Overview**. *Theory into practice*, 2002 41(4); 212-264.	☐
Lecture	**Lecture 13.1 slides** (110 min)	☐

	13.2 Indexing and system performance (50 min.) (Strand 1. General KOS principles)	
Readings	1 Lecture 13.2 Objectives etc. (pink)	☐
	2 Textbook Chapter 16. **Indexing and system performance**	☐
	3 Soergel, **Indexing and retrieval performance: The logical evidence**. JASIS, 1994.9; 4(8): 589-599. Perhaps easier to read this article before Chapter 16.	☐
Lecture	**Lecture 13.2 slides** (60 min)	☐

Learning blog	Learning blog Week 13 due W April 27	○

	Assignments assigned None

pink **Lecture 13.1** ***April 27***

Exploration of Knowledge Organization Systems (KOS)
(ontologies, classification schemes, thesauri)

Read all pages for this lecture before listening to the audio

Learning objectives Inherit from Lecture 10.2a In addition	2.4.3 Be cognizant of the wide range of different types of classification schemes and other Knowledge Organization Systems (KOS) and their many uses. Be acquainted with some important example KOS. (P2.3.9,4.3) Also Have an improved understanding of general principles of KOS structure and of methods for presenting this structure and be able to apply these principles in analyzing KOS.
Practical significance Inherit from Lecture 10.2a In addition	• You will be able to "sell your skills" to a wider variety of organizations, increasing opportunities for work. • Knowing the general principles that underlie all KOS will enable you to evaluate KOS, to improve existing KOS, and to build new KOS (after taking UBLIS 514 Indexing and Surrogation).

over

pink

Materials and methods	The schemes to be explored are listed below Each scheme is represented by a judiciously chosen selection of pages. These sample pages are not meant to be read word for word but rather to be examined with the goal of forming an overall image of the scheme. In particular, look at any sample entries marked by underline or *; they are usually part of a coordinated example illustrating the structure and relationships between parts. The schemes are separated by a blue sheet. These pages are meant for exploration, not for word-to-word to reading. Get an overview and pick out some examples to understand the structure of a scheme. Two elements to consider when examining a system 1 Function: What is this system used for? What could it be used for 2 Structure: Facets, hierarchy The lecture will go through the schemes, discussing examples from each. You can examine the schemes before listening to the lecture (lecture will make more sense) and/or after listening to the lecture. Please explore all schemes and select two schemes for closer examination and post an observation on at least one scheme on the free write for this week. Select the two schemes according to your interests, For example, if you are in School Library Media Program, examine the *Taxonomy for learning, teaching, and assessing* (the revision of *Blooms Taxonomy of educational objectives*, if you are interested in art and/or museums, examine the *Art and Architecture Thesaurus*. The last page for this lecture has a listing of schemes illustrating more functions. Just read it. For more in-depth information, see www.dsoergel.com/SoergelKOSOverview.pdf

pink

Schemes to be examined

Bloom	Anderson, L. W., & Krathwohl, D. R. **A taxonomy for learning, teaching, and assessing:** **A revision of Bloom's taxonomy of educational objective**s. New York: Addison Wesley Longman; 2001.
SOC	**Standard Occupational Classification** 2000 Bureau of Labor Statistics (BLS) + other agencies http://stats.bls.gov/soc/home.htm The SOC is augmented by the **Occupational Information Network (O*NET)**, a database with additional occupational titles, definitions, and features of occupations. http://www.doleta.gov/programs/onet
CSDGM	**Content Standard for Digital Geospatial Metadata** 1998 Federal Geographic Data Committee (FGDC) http://www.fgdc.gov/metadata/contstan.html
AOD	**The Alcohol and Other Drug Thesaurus.** 3rd ed., 2000. National Institute on Alcohol Abuse and Alcoholism (NIAAA) http://etoh.niaaa.nih.gov/AODVol1/Aodthome.htm
MeSH	**Medical Subject Headings**. National Library of Medicine (NLM) No longer published in print www.nlm.nih.gov/mesh/meshhome.html
AAT	**Art and Architecture Thesaurus**. 2nd ed 1994 Getty Art History Information Program www.getty.edu/research/tools/vocabulary/aat/index.html
WordNet	**WordNet Lexical Database**. Version 3.0 Aug. 2003 Princeton University, Cognitive Science Laboratory (www.notredame.ac.jp/cgi-bin/wn.cgi)
CYC Ontology	(CYC Corporation) 1997 To get an idea of what is in CYC: http://cyc.com/cyc/technology/whatiscyc_dir/whatdoescycknow to search: http://sw.opencyc.org/
FOAF	Friend Of A Friend http://xmlns.com/foaf/spec/

pink

Additional KOS examples illustrating different functions

HS	**Harmonized Commodity Description and Coding System**. World Customs Organization, Brussels. Info: http://pacific.commerce.ubc.ca/trade/HS.html
NAICS	**North American Industrial Classification System** "common industry definitions for Canada, Mexico, and the US. Developed in cooperation with the US Economic Classification Policy Committee, Statistics Canada, and Mexico's Instituto Nacional de Estadistica, Geografia e Informatica to better compare economic and financial statistics and ensure that such statistics keep pace with the changing economy. NAICS will replace the countries' separate classification systems (in the US: Standard Industrial Classification, SIC) with one uniform system for classifying industries." Info: www.census.gov/epcd/www/naics.html, www.naics.com www.cdc.gov/nchs/about/major/dvs/icd10des.htm
ICD-10	**The International Statistical Classification of Diseases and Related Health Problems, tenth revision**. Produced by the World Health Organization. Published in many languages. Info: www.who.int/whosis/icd10/index.html, www.cdc.gov/nchs/about/major/dvs/icd10des.htm
CPT	**Physicians' Current Procedural Terminology.** www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page (Info: www.ama-assn.org/ama/pub/category/3113.html, listing of codes https://webstore.ama-assn.org/index.jhtml)
HCPCS	Health Care Finance Administration (HCFA) **Common Procedure Coding System (HCPCS) for Medicare reimbursement for hospital outpatient services**. It has three levels - CPT (level 1), HCPCS or National (level 2), and Local (level 3). In its data collection the Agency for Health Care Policy and Research (AHCPR) uses data standards that are based on those used by the Census Bureau, the American Hospital Association, the Health Resources and Services Administration (Area Resource File), the National Center for Health Statistics, and codes for clinical diagnosis and procedures such as ICD-10 and CPT 1998. These standards facilitate data analysis and use by ensuring comparability, quality and interoperability. Further, uniform health care data advance medical and health care services research, the efficiency of the private sector health care delivery system, and quality improvement measurement. Further type of classification: **biological taxonomies**. Used in biology, agriculture, food science, and medicine. Several rivaling schemes for major areas (kingdoms) and many publications on specific areas. www.itis.gov/ www.ucmp.berkeley.edu/help/taxaform.html See www.obofoundry.org/ for many biomedical ontologies

pink **Lecture 13.2** ***April 27***

Indexing and system performance (Textbook, Chapter 16)

Learning objectives From Part 5 Learning objectives, p. ~285	2.5 **Understand indexing characteristics and their effect on system performance** and be able to apply this understanding to the analysis and design of databases and to database selection and query formulation. (P2.3.9,5) 2.5.1 Understand the concepts of **exhaustivity and specificity of indexing** and their effect on searching. (P2.3.9,5.1) 2.5.2 Be able to ascertain the exhaustivity and specificity of indexing in a given system and apply this knowledge to appropriate query formulation. (P2.3.9,5.2) 2.5.3 Understand the concepts of weights in indexing and query formulation. (P2.3.9,5.3) 2.5.4 Be able to apply indexing weights in query formulation (including analogous techniques in free-text searching). (P2.3.9,5,4) 2.5.5 Be able to determine the proper levels of exhaustivity and specificity of indexing for a new IR system based on user requirements. (P2.3.9,5.5)
Practical significance	An examination of indexing parameters, especially **exhaustivity** and **specificity** and **term weighting**, their measurements, their effect on retrieval performance (which is often oversimplified), their dependence on various factors in the indexing process, and their costs (Chapter 16). A correct understanding of these relationships is important for optimal query formulation in online systems, including the more sophisticated Web search engines, as well as for system design.
Discussion questions	How could one gauge the exhaustivity of indexing in a database if indexers' instructions are not available? How could one tell if within one and the same database exhaustivity varies from subject to subject? Give examples of exhaustivity and specificity of indexing for other types of entities and relationships.

XXX Combine Ch. 16 and logical evidence paper into new Ch. 16

Over

pink

Discussion of Textbook *Chapter 16* and the reading

<table>
<tr><td>Textbook Section 16.2.1</td><td>Definitions of exhaustivity and specificity. Indexing weights.

Put in the context of the conceptual data schema of a system.

Indexing specificity has to do with the entity values for the entity type subject (or of other entity types, for example Date, to which the concept of specificity can be applied).

The rules for exhaustivity in indexing are a special case of rules for establishing relationships, such as relationships between a document and subjects. Analogous rules can be defined for many types of relationships.

Indexing with weights requires three-place relationships, such as

Document deals with or is relevant for (Subject, Weight)</td></tr>
<tr><td>16.3.1</td><td>Effects of indexing exhaustivity on retrieval performance

Important conclusion: The query formulation must be adapted to the exhaustivity of indexing for best retrieval results.</td></tr>
<tr><td>Other questions</td><td>Questions on the remainder of the chapter and the reading.</td></tr>
</table>

In-lecture exercise:

Compare indexing with Dewey Decimal Classification and the London Education Classification (a faceted classification) with respect to exhaustivity and specificity of indexing.

First read through p. ~403 - 405, then we will work jointly on Documents 1 - 3.

Exhaustivity and specificity of indexing are important parameters that affect retrieval performance. This exercise will help you gain a better understanding of these parameters. For three documents, you will compare DDC with the faceted *London Education Classification* (LEC, available at XXX).
Look at the example of Document 0 on the facing page. Our task is to compare the exhaustivity and specificity achieved with the LEC with that achieved with DDC (the one Dewey class used for the call number) as shown in the example. We will analyze Documents 1 - 3 in the same way. For all documents, the descriptors from LEC are already filled in; you do not need the LEC. The Dewey class is also given. As discussed in Chapter 16, in order to compare exhaustivity and specificity of two sets of descriptors assigned to the same document, one must first semantic factor the descriptors to arrive at two lists of elemental concepts, which can then be compared. The LEC concepts are already elemental**. Determine the semantic factors of the Dewey class**; **all we need for this is the class caption, such as** *Organization and activities in secondary education > curricula.* **Enter the resulting elemental concepts in column 2, matching them up with the LEC concepts where possible**. Use the blank row if there is an elemental concept resulting from semantic factoring the DDC class without a corresponding concept from LEC. The elemental concept from the Dewey class may be broader or narrower than the LEC concept. For some LEC concepts there may be no elemental concept from the Dewey class; conversely, an elemental concept from the DDC class may have no corresponding LEC concept. In column 3 circle or underline the scheme that has the more specific concept. Can you detect a pattern?

LEC outline		Some LEC examples	
B	Education: foundations, principles, policy, etc	Bap	Educational opportunities; Bept Statistics of educ. Betm Financial resources; Bid Government Center
D	School buildings, bldg. services, equipment		
F	The teaching profession, personnel management	Def	Music Room; Dvo Computer
G	Type of personnel in education	Fal	Responsibility; Fas Recruitment
H	Management of educational institutions	Gan	Dean; Get Librarian; Gon Nurse
Hs-z	Human biology. Health and hygiene	Hab	Management of education; Heb Admission
J	Psychology of education. Educ. measurement	Htw	Motor ability; Huv Neurosis
K	Educand's work (study method, interests, voluntary vs. compulsory work)	Jze	Emotion; Jed Student discipline; Jud Counseling
L	Teaching method	Kad	Study methods; Kib Children as writers
M-P	Subject taught (Curriculum)		
Rab-Sus	Grade level and type of institution	Lah	Team teaching; Lep Group work; Lus Library
Svb-Tvp	Educands (by age, by exceptional requirements, and other characteristics)	Mab	Curriculum, syllabus, in general; Mok Biology
		Rib	Comprehensive school; Rek Secondary school - upper
		Svg	School child, pupil; Teb Blind and partially sighted

Example. **Document 0**

373.19 **Organization and activities in secondary education > curricula**
Note: This class caption is all you need to semantic factor the DDC class to arrive at the elemental concepts to put in column 2.
Conant, James Bryant, 1893-1978
The comprehensive high school; a second report to interested citizens

Elemental LEC concepts	Elemental concepts from DDC class	Which scheme is more specific?		
Rib Comprehensive (type of school)		SAME LEC DDC		
Rek Secondary school-upper	Secondary education	SAME **LEC** DDC		
Mab Curriculum, syllabus, in general	Curricula	**SAME** LEC DDC		
Bept Statistics of education		SAME LEC DDC		
Bap Educational opportunities, access to education		SAME LEC DDC		
	No. of concepts	1	1	0

Exhaustivity		Specificity	
LEC elemental concepts #	5	LEC same as DDC #	1
DDC elemental concepts #	2	LEC more specific #	1
		DDC more specific #	0
More exhaustive (more el. concepts)**:** LEC		**Overall more specific:** LEC	

In this example: LEC indexing has 5 concepts, DDC indexing 2, thus LEC is more exhaustive
Of the two common concepts, one is more specific in LEC and one is the same specificity, thus, for this document, indexing with LEC is more specific overall.

Example, repeated

373.19 **Organization and activities in secondary education > curricula**
Conant, James Bryant, 1893-1978
The comprehensive high school; a second report to interested citizens

LEC concepts	Elemental concepts from DDC class	Which scheme is more specific?		
Rib Comprehensive (type of school)		SAME LEC DDC		
Rek Secondary school-upper	Secondary education	SAME LEC DDC		
Mab Curriculum, syllabus, in general	Curricula	SAME LEC DDC		
Bept Statistics of education		SAME LEC DDC		
Bap Educational opportunities, access to education		SAME LEC DDC		
	No. of concepts	1	1	0

Exhaustivity		Specificity	
LEC elemental concepts #	5	LEC same as DDC #	1
DDC elemental concepts #	2	LEC more specific #	1
		DDC more specific #	0
More exhaustive: LEC		**Overall more specific:** LEC	

In this example: LEC indexing has 5 concepts, DDC indexing 2, thus LEC is more exhaustive
Of the two common concepts, one is more specific in LEC and one is the same specificity, thus, for this document, indexing with LEC is more specific overall.

Document 1

378.111 **Higher Education > Organization and activities in higher education > Personnel management > Administrators**

Dibden, Arthur James, 1919-

The academic deanship in American colleges and universities

LEC concepts	Elemental concepts from DDC class	Which scheme is more specific?		
Sab Institutions of Higher Education		SAME LEC DDC		
Hab Management of Education		SAME LEC DDC		
Gan Dean		SAME LEC DDC		
Ban Sociology of Education		SAME LEC DDC		
		SAME LEC DDC		
	No. of concepts			

Exhaustivity		Specificity	
LEC elemental concepts #		LEC same as DDC #	
DDC elemental concepts #		LEC more specific #	
		DDC more specific #	
More exhaustive: (LEC or DDC)		**Overall more specific:** (LEC or DDC)	

Document 2

379.1214 **Public policy issues in education > Specific elements of support and control of public ed. > Support by specific level of government > Support by national governments > National support of higher ed.**

Wakefield, Rowan Albert, 1919-

Sources of Federal support for higher education. Experimental systems for a national information network.

LEC concepts	Elemental concepts from DDC class	Which scheme is more specific?		
Sab Higher education		SAME LEC DDC		
Dvo Computer		SAME LEC DDC		
Buxt Information services		SAME LEC DDC		
Bid Government: central		SAME LEC DDC		
Betm Financial resources		SAME LEC DDC		
Bepd Economics and education		SAME LEC DDC		
		SAME LEC DDC		
	No. of concepts			

Exhaustivity		Specificity	
LEC elemental concepts #		LEC same as DDC #	
DDC elemental concepts #		LEC more specific #	
		DDC more specific #	
More exhaustive: (LEC or DDC)		**Overall more specific:** (LEC or DDC)	

Document 3

027.8 **General libraries, archives, information centers > School libraries**

Delaney, Jack J. *The school librarian. Human relations problems*.

LEC concepts		Elemental concepts from DDC class	Which scheme is more specific?		
Svg	School child, pupil		SAME LEC DDC		
Rag	School education and systems		SAME LEC DDC		
Lus	Library		SAME LEC DDC		
Jed	Discipline (psychological aspects of relationship)		SAME LEC DDC		
Hab	Management of education		SAME LEC DDC		
Get	Librarian, teacher librarian		SAME LEC DDC		
Fal	Role, responsibility		SAME LEC DDC		
			SAME LEC DDC		
		No. of concepts			

Exhaustivity		Specificity	
LEC elemental concepts #		LEC same as DDC #	
DDC elemental concepts #		LEC more specific #	
		DDC more specific #	
More exhaustive: (LEC or DDC)		**Overall more specific:** (LEC or DDC)	

Briefly discuss what you saw about exhaustivity and specificity from these examples:

Discussion questions (repeated)

How could one gauge the exhaustivity of indexing in a database if indexers' instructions are not available? How could one tell if within one and the same database exhaustivity varies from subject to subject?

Give examples of exhaustivity and specificity of indexing for other types of entities and relationships.

red

Week 14. April 27 - May 4

Part 6
Conclusion

red

red

Things to do in Week 14, W April 27 - May 4

	Review answer keys ☐ required ○ optional	√
	Assignments due W May 4 ☐ required ○ optional	√
Review answer key(s)	Assignment 13.3: **Library of Congress/Sears Subject Headings** (LCSH) (assigned. L12.2a, Apr. 20, due Apr. 27)	☐
	Assignment 13.4: **ERIC Thesaurus** (assigned. L12.2b, Apr. 20, due Apr. 27)	☐
	Assignments due W May 4. None ☐ required ○ optional	√
Assignment(s)	No assignments due	
Preparation for the final	Submit questions / requests for topics to be included in the final review to dsoergel@buffalo.edu; was due on Sunday May 1 midnight	○
	Submit answers to sample final questions for feedback (notes for Lecture 14, p. ~413+) to dsoergel@buffalo.edu; was due on Sunday May 1 midnight	○
Learning blog	Learning blog week 13.	○

New topics this week

	14.1- 14.2 Final review (130 min) ⓘ ⓘ	
Readings None	*No assigned readings*	
Lecture	**Lectures 14.1 - 14.2 slides, Lectures 14.1 - 14.2 pdf**	☐

	Lecture 14.2 Final review, continued (65 min) ⓘ ⓘ	
Readings None	*No assigned readings*	
Lecture	**Lecture 14.1 - 14.2 slides continued**	☐

Learning blog	Learning blog Week 14 due W May 4	○

Assignment assigned None

red

pink **Part 6. Conclusion** ***May 4***

Lectures 14.1-14.2 Final review

Numbers at left margin indicate number of minutes = number of points in a 3 hr exam.

15 1 **Typology of Web documents**

There are a wide variety of "documents" on the World Wide Web ("Web pages" and "Web sites"). In a catalog of Web documents, it might be useful to include an indication of the type of document in the catalog record. **Develop a typology of Web documents for this purpose**. (A typology is a list or classification of types).

P2.3.4,2#	Graduates are able to select or design the knowledge structure for a new info. system Specifically, graduates are able to select or design a conceptual data schema for an information system based on user requirements.	571-L2.2 571-L4.2 571-L14Q1

P2.3.9,3#	Graduates understand the principles of the structure of subject classification, in particular facet organization and hierarchy and of methods for presenting this structure, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation.	571-A2 571-Part 5 571-A12 571-A13.1-4 L14Q1

30 2 **Home page design**

A naturalist organization keeps an inventory of (rare) plants and animals in Western NY New York.

They provide access to several general databases on plants and animals.

They maintain a database of locations where (rare) plants are found and of sightings of (rare) animals. They collect these data from individual members (or anybody, for that matter, with some check of credentials). Each occurrence/ sighting report is recorded by place, date, and time. Some of this information is kept confidential to prevent poaching.

They also have events (lectures, excursions) they announce with online registration. Design a home page for such an organization.

P2.3.1,4#	Graduates are able to apply understanding of information structure and principles of document design and text structure to information presentation, including the creation of good documents of all kinds (such as websites). Graduates understand the principles of good document design: developing a good conceptual structure for a body of knowledge and representing that structure for human understanding and for machine processing. (RT P2.2.2,2.2; P2.5.2,3)	571-L2.1 571-L5.2a-L6.1b 571-L14Q2

pink

15 3 **Reorganize thesaurus information to take less reading and less storage space.**

P2.3.1,2.1#	Graduates understand the principle of hierarchical inheritance and are able to apply it to achieve more compact internal storage YMBOL45\f"Symbol"\s10 in the computer and in the mind – and more compact and more easily grasped external representation.	571-L2.2 571-A4 571-L5.1 571-L14Q3

Example 1: Three entries from the ERIC Thesaurus:

Autoinstructional aids
RT Audiovisual aids
RT Computer assisted instruction
RT Courseware
RT Individualized instruction
RT Learner controlled instruction

Programmed instructional materials
RT Audiovisual aids
RT Computer assisted instruction
RT Courseware
RT Learner controlled instruction
RT Workbooks

Teaching machines
RT Computer assisted instruction
RT Courseware
RT Learner controlled instruction
RT Pacing

Another example

free participation
RT health care delivery and administration
RT health care economics

payment-based participation
RT health care delivery and administration
RT health care economics

subsidized payment
RT health care delivery and administration
RT health care economics

full cost-recovery payment
RT health care delivery and administration
RT health care economics

20 4 You have to **design a controlled-vocabulary IR system** (with human indexing) that gives the searcher the option of emphasizing either discrimination (one factor determining precision) or recall. List the features that are important for achieving this flexibility.

P2.3.9,5#	Graduates understand indexing characteristics and their effect on system performance and be able to apply this understanding to the analysis and design of databases and to database selection and query formulation.	571-L13.2b 571-L14Q4 571-L14Q5

5 **Query formulation for free text.** A user needs information on the following topic:

Validity of the evaluation of instructors through undergraduate students in social science courses.

A free-text search for this topic is to be made in a bibliographic database

(1) in database 1, searching is by terms occurring in the **title** of the document,

(2) in database 2, searching is based on terms that occur in the **title and/or the abstract of the document.**

P2.3.9,5#	Graduates understand indexing characteristics and their effect on system performance and be able to apply this understanding to the analysis and design of databases and to database selection and query formulation.	571-L13.2b 571-L14Q4 571-L14Q5

P2.3.8,2#	Graduates understand and be able to apply vocabulary control to remedy these problems, either through using a controlled vocabulary in indexing or through query term expansion in searching. RT P2.5.2,1.3.3	571-A7 571-A10 571-L8.1 571-L8.2a 571-L14Q5

20 5a For each database give the **conceptual query formulation** that you would use (do not worry about terminology at this point). Give your rationale.

10 5b **Give the free-text query formulation for database 2.** Assume that the search from the previous question is to be made in the system searching on **titles and abstracts** (system 2). Any word or phrase (multi-word term) occurring in the title or abstract can be used as descriptor for searching. Briefly describe how you would go about developing the query formulations in terms of descriptors (3 min.) Start doing it (7 min.)

20 6 This question deals with **retrieval in archives**; sufficient background is provided so that you can answer it even if you are not familiar with archives. Archives are a collection of documents (letters, memoranda, reports, etc.) produced by an organization, its various units, and the persons working in the units. (Assume an organization of the complexity of the Federal Government with many organizational units interrelated hierarchically and otherwise.) The organization of archives usually allows for easy retrieval of all documents produced by an organizational unit or a person; a document is linked to its producer at its creation so that the archivist need not do additional indexing to provide this type of access. Date when created, receiving organizational unit or person, and often related documents are also known for each document. It is usually too expensive to assign subject descriptors to individual documents, yet subject searches are frequent. The archivist doing a subject search uses her - more or less - complete knowledge of organizational units and persons and the subjects they have been dealing with at certain times to find relevant documents to look under appropriate units and persons.

Sketch a conceptual data schema for a computerized retrieval system for archives that implements in a formal way the approach described. Describe how this system does searches for subject.

P2.3.1,1.1.4#	Graduates are able to construct a conceptual data schema for a given information system BT P2.3.4,2	571-L4.2 571-A6 571-L14Q6

P2.3.2,1#	Graduates are able to combine different kinds of data to find an answer (inference, chained searches) (BT P2.5.2,1	571-L1.2 571-L2.2 571-L4.1 571-L14Q6

40 7 You are appointed as head of a medium-sized IR-system (about 200,000 documents) that uses three different systems for subject access:

(1) an alphabetical subject catalog of books;
(2) shelving books by subject;
(3) an independent classification scheme for filing newspaper clippings

Your analysis shows that the subject heading list and the shelving classification are both far from satisfactory. The subject headings have grown without control and no listing is available. But a cost-benefit analysis rules out major changes or revision, like introducing new schemes, especially in view of the large costs for re-indexing the old collection. On the other hand, the cost-benefit analysis also shows that some costs would be justified to improve the usability of the IR-system. What do you suggest should be done? How would you implement your suggestions?

P2.3.9,3.9#	Graduates understand the access mechanisms that help a user find the proper descriptors in a large classification scheme with many precombined descriptors, in particular cross-references and a descriptor-find index.	571-L12.1 571-L14Q7

XXX Associate learning objectives with the remaining questions.

40 8 You are given the task to design an IR system. One problem is to determine **how much money should be spent for indexing**. Discuss the data you need/the considerations on which you would base your decision.

40 9 You are given the task of **developing an index language and thesauru**s for
(1) a newly set up information center in a company, or
(2) a public information center in the inner city (choose **one)**.
What are the main points you have to take into consideration in performing this task?

20 10 **Assist users in coping with large Web search results**. A search in a Web directory, such as Yahoo or the Open Directory Project (http://dmoz.org/about.html), or a search engine, such as Google or Bing, often returns hundreds of documents. What could the system do to help the user to cope with these large numbers?

15 11 Discuss **exhaustivity** in the context **of hypertext links** made in a system.

15 12 A large subject index is to be put on microfiche. The system has two parts:

(1) The actual index on microfiche. This is an ordinary index: Under each descriptor the entries for the documents (or other retrieval objects) indexed by that descriptor are listed.

(2) To help the user find the appropriate microfiche, there is a hard copy "index to the index." This is simply a list of all descriptors, giving for each the microfiche number and the frame number on the microfiche.

Question: Should the subject index on microfiche be arranged in classified or in alphabetical order? How should the hard-copy "index to the index” be arranged?

Assume a microfiche reader where the user must manually insert the fiche and find the frame.

40 13 You are charged with the design and development of an online information retrieval system for courses at the University at Buffalo. The system should serve
(1) students in course selection and
(2) curriculum committees who want to know what courses exist in a given area (such as *statistics* or *communication in organizations* before approving a new course.
Discuss your approach (describe the workings of the system you propose to the extent feasible in 40 minutes; bulleted lists for some pieces are fine)

12 14 Compare a system using shelf arrangement based on an index language like LCC or DDC with a system based on postcombination (such as a computerized IR system) with respect to the exhaustivity and specificity of indexing that can be achieved. What can you say about retrieval performance in both cases?

30 15 Assume you have to **design a large lexical and classification database** that has the ambitious objective of serving as a tool for both natural language processing and indexing and retrieval. What information should be included for each term or concept?

Final review. Natural language processing (NLP). See Lecture 6.2b, p. ~217+

Also think about Knowledge required by NLP

Final review. Precombination vs postcombination See following pages

Both important

Final review. Precombination vs postcombination

1 Precombination vs postcombination in searching

1.1 Basic problem: Most searches are for topics or themes expressed as compound concepts, such as

the effect of alcohol on the liver or

how to improve test scores of minority children.

In a retrieval system that allows for combining descriptors (as an online system allowing for Boolean query formulations) an index language consisting only of elemental descriptors is sufficient: the user can combine the elemental descriptors that make up her search topic.

But in a retrieval system that allows only searches for single descriptors, as in a card catalog, printed index, shelve arrangement, or a Web subject directory (Yahoo, LookSmart, etc.) the system must provide precombined descriptors for the topics users want to search. The user who wants to find materials on a topic for which there is no precombined descriptor will have difficulty.

1.2-1.4 Additional reasons for introducing precombined descriptors even in a system that is mainly based on postcombination

1.2 With postcombination, the components of the query formulation may not have the right relationship in the document

Example: A search for

Air transport AND Vehicles

finds a document on

Vehicles used on the metro line to the airport

Need descriptor *Aircraft*

1.3 With postcombination, a combination of elemental descriptors might be ambiguous

Examples:

School	AND	Library	Need descriptors *School library* and *Library School*
Personnel	AND	Administration	Need descriptors *Personnel administration* and *Administrative personnel*

1.4 Requiring the user to combine elemental descriptors may be unnatural

2 Precombination vs postcombination in database organization

Documents are about topics/themes and can be usefully grouped by topics/themes

Examples:

- Shelf arrangement,
- Web subject directory,
- Organizing Web search results (often hundreds or thousands of items) or online catalog search results into meaningful groups.

Need precombined descriptors that define groups (classes).

Problem of arranging precombined descriptors / classes in a meaningful order

Other aspect of same problem: If document can be assigned only one descriptor, that descriptor should express as many of the document concepts as possible; it needs to be precombined.

Relationship top semantic networks; precombined descriptor as an abstraction of what is in common to a group of documents.

3 User problem with systems using a large number of precombined descriptors:

Finding all precombined descriptors under which to search (because of extensive polyhierarchy in a large set of precombined descriptors, one search often requires looking under many, as the DDC and LCC assignments demonstrated)

Solution: descriptor-find index