Information Organization and Access in the Digital Age

Dr. Gobinda Chowdhury
Graduate School of Informatics
University of Strathclyde
Glasgow, UK

Me and My Research
- My education, career and research activities
- Areas of research
- Information organization
- Information access
- Information use and users
- Digital libraries

Lecture Outline
- Information Literacy
  - Basic concept of information literacy and why it is important; the building blocks of information access and use; major problems and issues
  - Information retrieval
    - the basic process, What is IR and what goes under the bonnet; importance of language in textual information retrieval
- Information retrieval from different systems: OPACs to the web
  - Different approaches to information access from different systems: OPACs, online databases, e-journals and aggregator services, web and digital libraries
  - Tools, techniques, problems and issues

Lecture Outline (2)
- Problems of information access in the digital age
  - the personalization issue
  - recommender systems
  - DWE design and demonstration
- IR languages: controlled Vs. natural languages
  - The debate of natural vs. controlled indexing languages
  - Examples of typical controlled IR languages: Classification and vocabulary control tools – Live examples from WebDewey, LCSH and online databases/digital libraries like BUBL, PubMed, SOSIG and ACM digital library

Lecture Outline (3)
- New research approach to the Semantic web: meaning-based information access
  - The problems of information organization and access in the digital age; new approach to the semantic web; using subject classification and categorization tools as the backbone of the semantic web; the user approach to information access
- Information users
  - human information behaviour studies – what, why, and types, etc. information literacy and digital divide; barriers to information access
  - ICT differentials and its impact on information access
  - Barriers to information access

Lecture Outline (4)
- Trends and new research
- Meaning-based information organization and access on the web
- Systems for answering questions rather than retrieving a large number of documents (possibly) containing the required answer
- Personalized information access systems
Information and digital literacy skills

Information literacy: What?
• Information literacy is the ability to access, evaluate and use information from a variety of sources.
• An information literate person is expected to have acquired the necessary skills to retrieve information from a variety of sources, print as well as electronic, to meet the information needs at any given point of time.
• ALA Presidential Committee on Information Literacy recommend that to be information literate, a person must be able to recognize when information is needed and have the ability to locate, evaluate and use effectively the needed information.
• Terms like “Information Skills” and “Information mediacy” are also used.

Information literacy? Why?
Information literacy helps us in a number of ways, for example,
• To be able to recognize the need for information and define the information needed for problem-solving and decision-making
• To be able to identify potential sources of information
• To be able to apply appropriate search strategies to retrieve information using the available technologies and tools
• To be able to review, select, interpret and evaluate relevant information critically and to make meaning of this information
• To be able to organize and present information effectively and creatively
• To appraise the process and product of an information research
• To develop a habit of reading for information and for pleasure
• To continually improve and update knowledge
• To demonstrate initiative in information problem-solving and openness to learning
• To collaborate with others for information problem-solving.

Information skills for students
• The US higher education standards emerged from work done for the ALA on information literacy:
• The information literate student determines the nature and extent of the information needed.
• The information literate student accesses needed information effectively and efficiently.
• The information literate student evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.
• The information literate student, individually or as a member of a group, uses information effectively to accomplish a specific purpose.
• The information literate student understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.

The Seven pillars of information literacy
• The ability to recognize a need for information
• The ability to distinguish ways in which the information ‘gap’ may be addressed
• The ability to construct strategies for locating information
• The ability to locate and access information
• The ability to compare and evaluate information obtained from different sources
• The ability to organise, apply and communicate information to others in ways appropriate to the situation
• The ability to synthesise and build upon existing information, contributing to the creation of new knowledge.

Task Definition
( Determine what is needed)

Information Seeking Strategies
(Database selection, search strategies, etc.)

Location and Access
(Find the information and access it)

Use of Information
(Decide what is valuable)

Synthesis of Information
(Information restructuring and repackaging)

Figure 1: Six Major Steps in an Information Problem-Solving Process

Determine the exact information required to solve a problem, to accomplish an assignment, etc.

Identify keywords using dictionaries, thesaurus, etc.

Use WH questions (Who, Which, What, When, Where, etc.) to focus on the topic

Use strategies such as KWL (What I Know, What I would like to know, What I have already Learnt, etc.) to determine the topics, subtopics, etc.

Prepare the search strategy

Figure 2: Processes Involved in Task Definition
Think about various sources (documentary, institutional and individual sources)

Learn the various tools: interface, search and retrieval techniques, etc.

Use locational tools: OPAC, Internet, Indexes, Abstracts, Databases, Digital libraries, etc.

Figure 3: Processes Involved in Information Seeking Strategies

Use tools and techniques appropriate for the chosen system

Prepare the search expression and conduct search(es)

Check whether the output contains the required information

Result: OK?

Modify the query or the search strategy

Locate and retrieve the information

Figure 4: Processes Involved in Location and Access

Identify the primary and secondary sources of information

Distinguish between facts and opinion/propaganda, etc.

Check the authority, validity, accuracy, currency, and comprehensiveness of information

Understand, analyse, synthesize information from different formats from various sources

Organize the information gathered using different formats

Present information using appropriate tools (e.g., software) and medium

Figure 5: Processes Involved in Synthesis

Evaluation of the process

The following questions may be asked at this stage:
- Was the information problem solved?
- Was the need met?
- Was the decision made?
- Was the situation resolved?
- Does the product satisfy the requirements as originally defined?
Lecture 2

Information Retrieval Systems: From OPACs to the Web

Text Retrieval Systems Vs. DBMSs

<table>
<thead>
<tr>
<th>Text Retrieval</th>
<th>DBMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic</td>
<td>Deterministic</td>
</tr>
<tr>
<td>Language/text-based</td>
<td>Value-based</td>
</tr>
<tr>
<td>Relevance</td>
<td>Exact data</td>
</tr>
<tr>
<td>Vocabulary control</td>
<td>Exact Keys</td>
</tr>
<tr>
<td>Partial match</td>
<td>Exact match</td>
</tr>
<tr>
<td>Content analysis</td>
<td>Exact values</td>
</tr>
</tbody>
</table>

Types of IRSs

Historically Information Retrieval Systems have grown from simple keyword indexes to the modern digital libraries in the following order:
- Keyword indexes
- Online databases
- OPACs
- CD-ROM databases
- WWW and Digital Libraries

Types of IRS (2)

Information Retrieval Systems may be also be categorised as:
- Bibliographic information retrieval systems
- Full text retrieval systems
- Multimedia retrieval systems
- Knowledge-based systems

IR: Issues and Problems

- Information Sources: type, content, format, volume, etc.
- Indexing: policy, file organization, vocabulary control
- Users and User interface: nature, options, usability
- Queries: formulation, terminologies, query expansion/modification, vocabulary control
- Retrieval: matching – Boolean, best match, etc., filtering, presentation, etc.
Information Searching: Various Options

- OPACs
- Online databases
- Electronic Journals
- Electronic Books
- Digital Libraries
- Web search tools

Characteristics of OPACs

- Input: Metadata of Bibliographic collections of a library
- Processing: catalogue records of bibliographic materials
- Searching: limited search facilities on selected meta tags (author, title, keyword, subject, etc.); subject search is through LCSH
- Output: brief bibliographic details of records

Online Search Services: Characteristics

Information providers or database producers provide access to remote online databases
- Databases mostly contain textual information: abstracts or full text of documents
- A specially designed search interface is provided for searching many databases
- Users need to have communication links that connect them with the host and the database(s)
- Modern communications are web-based
- Successful searching needs some prior training
- Output of a search may contain abstract or full texts of items

E-journals and eBooks

- Electronic journals are accessible through a publisher or through an aggregator
- One or more journals can be searched or browsed through one interface
- Search output contains full text articles
- Many electronic books are just digital versions of printed books, such as encyclopedias, dictionaries, etc.
- Some eBooks are specially designed and need special readers

Web search tools

- Search engines and directories allow users to search for web information resources that may contain text or multimedia information
- Search engines have spider programs that select and index web pages
- Directories provide links to selected web pages
- Search output contains very brief information about the retrieved webpages with links

Search engines: basic Search features

- Natural language search
- Keyword and phrase search
- Limiting search
- Users can select a language to search
- Field search
- Boolean, truncation and proximity search
- Simple and advanced search interfaces
- Separate search screens for images, audio, video, etc.
Special features of search engines

- Link (in AltaVista, for example)
- Family filter
- Translate (in Alta Vista, for example)
- Statistics (in Google)
- Sorting of results (as in Alta vista, Hotbot, etc.)

Meta and Speciality Search engines

Meta search engines conduct search on more than one search engines at a time

- Examples: AskJeeves, Mamma, Dogpile
Some search engines provide some special features

- For example, AskJeeves comes up with a set of pre-defined questions on the topic given by a the users

Information Searching in Digital Libraries

- One or more type of information resources
- Information may be located at one place, or may be distributed
- Cross database and cross site searching
- One or more different interfaces to search the same or specific parts of a DL
- Hybrid libraries provide facilities for searching traditional library resources as well as digital resources
IULA Lecture 3

Information Organization and access: From OPACs to online databases and e-journals

Queries
Most information retrieval systems are query-based
Users need to submit a query representative of their information needs
One or more search terms are to be entered with appropriate search operators
Selection of search terms for the formulation and modification of queries is often difficult
Ways of formulating and submitting a query depends on the IRS and its interface

IR: The Basic Process
Candidate Document terms are identified
Index terms are organized in an Inverted file
Document reference, position, frequency of occurrence, etc. for each term is stored in the Inverted index file
Search terms are matched against the Inverted file and matching items are retrieved

Basic Search Techniques
- Bibliographic Vs. Non-bibliographic databases
- Known Vs. Unknown item search
- Keyword vs. Subject search
- Search operators
- Search interfaces and options

OPACs: Features
Relatively simple interfaces; nowadays available through the web
Records can be retrieved through browsing or searching
Keyword and phrase search facilities are common
Bibliographic records can be searched through selected keys – author, title, ISBN, call number, etc.; these are searched as phrases, and are usually automatically right truncated

OPACs: Features (2)
Records are subject indexed through subject headings, and therefore users have to use the right subject heading to do a subject search
Boolean search: with the Keyword search option
Proximity search: with the keyword search option
OPACs: Features (3)

Some searches can be limited by date, collection, language, etc. Sometimes users can search more than one collection (within the same library or in different libraries) ANSI Z39.50 standard is used for the IR. Results are usually not ranked.

Strathclyde Univ. OPAC: Keyword and Subject Search Options

**Subject:** Subject headings are those used by Library of Congress

**Title/Name/Journal Title/Subject Keyword**
- type one keyword or a phrase
- use ? for truncation
**Keywords Anywhere Search**
- use quotes to search for a phrase - "world wide web"
- use + for essential words - +internet

Strathclyde Univ. OPAC: Subject Search Showing the Browseable List of Subject Headings

Strathclyde Univ. OPAC: Subject Keyword Search Results
IULA Lecture 3 (2)

Information Organization and access: From OPACs to online databases and e-journals

Online Databases: Characteristics

Novice and expert search mode
Word and phrase search
Truncation
Proximity search
Searching variant forms of a word (e.g., in LEXIS-NEXIS)
Boolean and Nested Boolean Search

Online Databases: Characteristics (2)

Field Specific Search
Limiting Search
Range Search
Frequency of Occurrence of words (e.g., Ovid Online)
Searching through one index for the selection of databases (e.g., DialIndex)

Online databases: Characteristics (3)

Thesaurus support
Cross database search
Discarding duplicate results
Display results in one or more chosen format
Sorting of search results
Ranking results

Online Searching: Strategy

1. Study the search topic and develop a clear understanding of the information requirement
2. Get access to an online search service
3. Log on to the service provider
4. Select appropriate database(s) to search
5. Formulate search expressions
6. Select the appropriate format for display
7. Reformulate your query, if necessary
8. Select the mode of delivery

E-journals

two major categories of e-journals:
– one that have their printed counterparts, for example, the *Journal of Documentation*, and
– the other that are available only in electronic format, for example, the *D-Lib Magazine*. 
E-journals: Access

- by publishers themselves or
- by aggregators.

The benefits of getting access to an individual publisher's journals are:
- value-added features and
- lack of intermediaries

E-journals: Access (2)

Aggregators, such as, Blackwell's Electronic Journal Navigator, Swetsnet, Ebsco Online, etc., conglomerate journals of several publishers under one interface and search system. Each publisher and aggregator of e-journals has a proprietary retrieval engine and search interface that can be used to search one or more e-journals.

E-journals: IR features

- Users can browse each issue or can search the entire collection
- There are usually novice and expert search modes
- Word and phrase search facilities are commonly available
- Common search facilities include: Boolean search, truncation, field search, limiting search and range search

E-journals: IR features (2)

Searches can be conducted on metadata (author, title, etc.) or on the full texts or articles.
- Output is available in one or more formats – HTML, PDF, etc.

For a list of Electronic Journals: Services and Providers see http://www.lib.strath.ac.uk/ejservices.htm

Proquest Basic Search Screen

Proquest Guided Search Screen
Web Searching

There are basically two ways to search the web:

– by conducting a search using what is known as a search engine, or
– by following the links in a specially designed list called a directory.

Search engines

Search engines allow the user to enter search terms – keywords and/or phrases – that are run against a database containing information on the web pages collected automatically by programs called spiders. A search engine retrieves web pages from its database that match the search terms entered by the searcher.

Search engines (2)

It is important to note that when a user conducts a search using a search engine, the later does not search for the information across the entire web at the given instance. Instead, it searches a fixed database, which is updated at a regular interval according to the specific criteria employed by the search engine, located at the search engine’s website containing information on selected web pages.

Search engine: components

The spider, i.e., the program that automatically collects information about millions of pages on the web, the index that stores information collected by the spider on the various web pages, and the search engine software and interface with which the users interact to conduct a web search.

Search engine: types

• The Major Search Engines, e.g., Altavista, AOL Search, Google, HotBot, etc.
• Kids Search Engines, e.g., Yahooligans, KidsClick, et.
• News Search Engines, Altavista News, Ananova, Yahoo News, etc.
• Specialty Search Engines, AskJeeves, AllExperts.com, CNETDownload.com, etc.

Search engine: types (2)

• Multimedia Search Engines, e.g., AltaVista Photofinder, FAST multimedia search, Ditto, Napster, Gnutella, etc.
• Search Utilities, Copernic, LexiBot, SearchWolf, Subject Search Spider, etc.
• Paid Listings Search Engines, e.g., Google AdWords, FindWhat.com, Espotting.com, etc.
• Metacrawlers, Kartoo, Query Server, Profusion, InfoGrid, etc.
• Regional Search Engines, e.g., IndiaInfo.com
Search engine: IR features

- Word, phrase and natural language search facilities are available
- Simple and Advanced Search options are available
- Special options are available for image, audio and video search
- A number of categories are available for browsing
- Most search engines support multilingual search

Search engines: IR features (2)

- Boolean Search: AND, OR, AND NOT; + and – signs are used to indicate that the following search term must and must not appear; Parentheses can be used for nested Boolean search. In the Advanced search mode there are three options, and one can be chosen after entering a word/phrase: Must have, Good to Have and Must Not have.
- Other search facilities include: proximity search, truncation, meta tag search (field search)

Search engines: IR features (3)

- A search can be constrained by different criteria, for example, by language, date of publication, collection type, etc.
- The search output is ranked based on some criteria set by the search engine software output
IULA Lecture 3 (3)

Information Organization and access: From OPACs to online databases and e-journals

Subject Gateway

Virtual Libraries or subject gateways are essentially online libraries containing links and pages on a variety of categorized topics. Similar, in a very general way, to web directories, they are a good resource for users trying to do serious online research on a specific area of interest.

Subject Gateways: Information organization

The fundamentals of this approach are as follows:
- Various Internet resources are selected by human experts for their quality and relevance to a particular target audience.
- They are then reviewed and resource descriptions created, which are stored, generally with the associated metadata, and generally in a structured format.
- These efforts are made to improve the recall and especially the precision of information searching on the web by the users in a particular subject or discipline.

SOSIG (www.sosig.ac.uk)

SOSIG, Social Science Information Gateway, aims to provide access to high-quality Internet information for researchers and practitioners in the field of social sciences, business and law. It uses social science academics and information specialists who select and describe sites after an evaluation of their quality. It is a unique Internet search tool that enables users to find high quality Internet sites and resources in social sciences.

SOSIG Home Page

SOSIG Thesauri
Biz/ed (www.bized.ac.uk)

Biz/ed, the Business and Economics information gateway, is a unique service for students, teachers and lecturers of business, economics, accounting, leisure and recreation and travel and tourism.
It allows users either to browse resource descriptions within subject headings or to use keyword searching of the descriptions.
OMNI

OMNI (Organising Medical Networked Information) is a gateway to evaluated, quality Internet resources in health and medicine, aimed at students, researchers, academics and practitioners in the health and medical sciences.
Information Searching in Digital Libraries

One or more type of information resources
Information may be located at one place, or may be distributed
Cross database and cross site searching
One or more different interfaces to search the same or specific parts of a DL
Hybrid libraries provide facilities for searching traditional library resources as well as digital resources

Selected Digital Library types

California Digital Library
New Zealand Digital Library
Networked Digital Library of Theses and Dissertations
ACM Digital Library
British Library
Alexandra Digital Library

Personalised Information Services in DLs

Services that allow users to personalise services through a personal webpage, e.g., MyLibrary, PIE, etc.
Task-based Information systems, e.g., MyLibrary@NCState, DWE
Customised interface for specific user categories, e.g., Glasgow health Information Gateway
IULA Lecture 4: Personalization

Issues

Gobinda Chowdhury

Why Personalization?

- Digital information resources and services are provided in the same way for everyone
- Every user has a special need and information behaviour
- Every user has a choice of specific information resources and services
- Users need to conduct similar searches over and over again and on a daily basis most of which are routine in nature
- Users often find it difficult to guess and decide where to search for information related to a specific need
- Users often need to go back to a previous search and result sets
- Personalization can help us resolve many of these problems

Personalised Information Services in DLs

- Services that allow users to personalise services through a personal webpage, e.g., MyLibrary, PIE, etc.
- Task-based information systems, e.g., MyLibrary@NCState, DWE
- Customised interface for specific user categories, e.g., Glasgow health Information Gateway

MyLibrary@Cornell
(http://mylibrary.cornell.edu/)

MyLibrary is a collection of personal electronic services, developed by the Cornell University Library, that can be customized to reflect a given user’s personal interests and research needs. MyLibrary consists of three services: MyLinks, MyUpdates and MyContents.

MyLinks

MyLinks can be used to collect and organize user’s frequently used electronic resources. It consists of an individualized “homepage” containing links to the user’s frequently used resources, ones that the user had selected from either Cornell’s Library Gateway or anywhere else on the Internet.

Reading
http://www.dlib.org/dlib/april00/mistlebauer/04mistlebauer.html

MyUpdates and MyContents

MyUpdates can be used for bi-weekly notification of new books, journals, and other media added to the Library Catalog that meets the user’s specific research and personal interest needs. MyContents is a tables of contents (TOC) service that delivers TOC of the user’s choice. The journal tables of content may be received in a variety of formats.
MyLibrary@NCState

Once an account is created, the user’s MyLibrary@NCState page is dynamically built and displayed. Afterwards other links may be added to or subtracted from the pre-configured list of recommendations to create a personal digital library

– Reading:

Digital Work Interface (DWE)

A research project aimed at the organisation of, and access to, digital information based on user tasks

Users can select a specific task, and the system will list (and open, if appropriate) one or more information resources useful for accomplishing the chosen task
Information for Specific User Categories

Glasgow Health Information Gateway
(http://www.accessglasgowhealth.org.uk/)

Setting up Virtual Information Services

- User and user needs identification
- Information Service (Digital Libraries, Reference Services, Databases, e-journals, etc.) Identification
- Categorisation of information resources/services
- Design of webpages for (1) the display of the various resource/service and (linking to the resources/services
- Personalization of the web page for information resource/service selection and/or display
IULA Lecture 4, Part2: DWE

Gobinda Chowdhury

User-Centred Research in the DL Environment

Gobinda G. Chowdhury
Graduate School of Informatics
Dept. of Computer and Information Sciences, Univ. of Strathclyde

Background

Typical Characteristics (problems?) of DLs:
- Volume, Variety, and location of information resources
- Creation/conversion of information resources
- Highly dependent on IT and networks
- Users need special skills
- So far the focus has been on building enabling technologies rather than services and users

A Possible Solution

Digital Work Environment (DWE)
- A task-based system for the organization of, and access to, digital information
- Users only need to identify a task, and the system will automatically identify, and provide access to, the appropriate information resources
- A one-window shop for getting access to heterogeneous information resources

DWE

A one-stop window for access to all sorts of digital information:
library, intranet and internet
Task-based access to information resources
Information filtering as well as search system

Access to DL resources: Various Options

Through the specially designed search interfaces of a sp. DL, e.g., NCSTRL, NDLTD
Through the specific interface of each DL resource/service, e.g., the Strathclyde library webpage
Through a single search interface, e.g., SearchLight in California DL
Through personalised interfaces, e.g., HeadLine PIE, MyLibrary@NCState
DWE : an overview

Architecture
Components and workflow
Advantages
Problems
Prospects/future research

DWE Components
Information Resources Organiser Module
Task and Resource Maintenance Modules
User Authentication and Management Module
Statistics Module
User Interface

DWE Functions
Identification of Information Resources
Identification of various User Tasks
Mapping of resources on user tasks
Interface linking a number of back-end databases

DWE Interface

DWE Task Frame
DWE Resource Selection

DWE Resource Display

DWE Advantages
One-stop window for access to information
Recommended information resources based on user tasks
Search facility based on metadata or full text
Personal note space for every user

DWE Problems
Identification and mapping of information resources: manual and resource-intensive
Creation of metadata for every resource: resource-intensive
Maintenance of resource and user/task databases: resource-intensive (though the amount of work will reduce over time)

DWE: further research
Use of an ontology for query expansion and item/query mapping
More domain- and task-specific research
Automatic resource identification and/or filtering mechanisms based on a sample set of resources organised a/c to user tasks
Building KM tools/environment
IULA Lecture 5: Language issues in IR

Gobinda Chowdhury

IR: Language issues

- Information access is based on words/phrases or headings representing the content of the information resource
- The natural vs. artificial language debate
- Pre- vs. post coordinate indexing languages
  - Pre co-ordinate indexing
  - Post coordinate indexing
- Natural language processing
  - Lexical and syntactic analysis
  - Semantic analysis
  - Pragmatic analysis

Controlled IR languages

Two types of controlled languages have been used in IR
- Artificial notations
  - Used in classification schemes like DDC, UDC, LC, etc.
- Natural language – pre-coordinated
  - Used in subject heading lists like LCSH
  - Used in thesauri

Dewey Main Classes

<table>
<thead>
<tr>
<th>Main Classes</th>
<th>Second level</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>Computer science, information &amp; general works</td>
</tr>
<tr>
<td>000</td>
<td>Computer science, knowledge &amp; systems</td>
</tr>
<tr>
<td>010</td>
<td>Bibliographies</td>
</tr>
<tr>
<td>020</td>
<td>Library &amp; information sciences</td>
</tr>
<tr>
<td>030</td>
<td>Encyclopedias &amp; books of facts</td>
</tr>
<tr>
<td>040</td>
<td>[Unassigned]</td>
</tr>
<tr>
<td>050</td>
<td>Magazines, journals &amp; serials</td>
</tr>
<tr>
<td>060</td>
<td>Associations, organizations &amp; museums</td>
</tr>
<tr>
<td>070</td>
<td>News media, journalism &amp; publishing</td>
</tr>
<tr>
<td>080</td>
<td>Quotations</td>
</tr>
<tr>
<td>090</td>
<td>Manuscripts &amp; rare books</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Main Classes</th>
<th>Second level</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>Computer science, information &amp; general works</td>
</tr>
<tr>
<td>001</td>
<td>Knowledge</td>
</tr>
<tr>
<td>002</td>
<td>The book</td>
</tr>
<tr>
<td>003</td>
<td>Systems</td>
</tr>
<tr>
<td>004</td>
<td>Data processing – Computer science</td>
</tr>
<tr>
<td>005</td>
<td>*Computer programming, programs, data</td>
</tr>
<tr>
<td>006</td>
<td>*Special computer methods</td>
</tr>
<tr>
<td>007</td>
<td>[Unassigned]</td>
</tr>
<tr>
<td>008</td>
<td>[Never assigned]</td>
</tr>
<tr>
<td>009</td>
<td>[Never assigned]</td>
</tr>
</tbody>
</table>
### Main Classes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>Computer science, information &amp; general works</td>
</tr>
<tr>
<td>005</td>
<td>Computer science, knowledge &amp; systems</td>
</tr>
<tr>
<td>005.019</td>
<td>&quot;Computer programming, programs, data&quot;</td>
</tr>
<tr>
<td>005.0684</td>
<td>Psychological principles</td>
</tr>
<tr>
<td>005.1-005.5</td>
<td>Software consultants</td>
</tr>
<tr>
<td>005[.6]</td>
<td>Computer programming and programs</td>
</tr>
<tr>
<td>005.7</td>
<td>Microprogramming and microprograms</td>
</tr>
<tr>
<td>005.8</td>
<td>&quot;Data in computer systems&quot;</td>
</tr>
<tr>
<td>005.8</td>
<td>&quot;Data security&quot;</td>
</tr>
</tbody>
</table>

### Use of Classification in Internet IR

- Several classification schemes have been used to organize and provide access to online information resources, e.g. BUBL, CyberDewey, ACM digital library, etc.
- We shall look at two specific examples, BUBL and the ACM DL.
Computer programming - visual basics

**Overview**

- Visual programming concepts
- Microsoft Visual Basic Express
- Basic syntax and structure
- Data types and variables
- Control structures: if-then-else, for, while
- Functions and procedures
- Arrays and collections
- File input/output

**Key Features**

- Interactive learning environment
- Step-by-step code snippets
- Debugging tools
- Project-based assignments

**Applications**

- Web application development
- Mobile app creation
- Game development
- Automation and control systems

**Resources**

- Official documentation
- Online tutorials
- Community forums
- Sample projects

**Conclusion**

Visual programming is a powerful tool for beginners and professionals alike, offering a intuitive approach to coding. With its rich set of features and comprehensive learning resources, Microsoft Visual Basic Express enables users to efficiently develop applications across various domains.
IULA Lecture 5: Language issues in IR, Part 2

Gobinda Chowdhury
Semantic Web

- It is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation
- It is a mesh of information linked up in such a way as to be easily processable by machines, on a global scale
- Two main objectives of the semantic web are:
  - to facilitate information access and retrieval by meaning, and
  - to build mechanisms that will facilitate machine communications for linking and processing information

Semantic Web: the W3C vision

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

The Subject approach

- It is proposed that Dewey coupled with some vocabulary mapping tools and techniques can play a major part in organizing and linking web resources for supporting semantics-based access and retrieval.
- This can be accomplished by:
  - Semantic querying
  - Semantic mapping of digital resources
  - Semantic retrieval
Summary and trends

- A number of issues are related to information access and use in today's digital world.
- Language plays the key role in IR.
- While NLP techniques have been tested and many NLP systems have been designed over the past decades, they appear to be more suitable for domain-specific applications.
- Simple NLP techniques can be useful in building QA systems.
- Word meanings have a great role to play in building the semantic web.
- Automatic semantic annotation of documents on the web is the key to realizing the goals of the semantic web.
- Controlled languages coupled with automatic semantic tagging and annotation systems can lead to a reasonable solution.