




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



University of Strathclyde
Science

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

CIS

University of Strathclyde
Science

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.

CIS

University of Strathclyde
Science

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

CIS

University of Strathclyde
Science

The Seven pillars of information literacy

- The ability to recognise 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.

CIS

University of Strathclyde
Science

```

graph TD
    A[Task Definition  
(Determine what is needed)] --> B[Evaluation  
(Evaluation of the Entire Process)]
    B --> C[Information Seeking Strategies  
(Database selection, search strategies, etc.)]
    C --> D[Location and Access  
(Find the information and access it)]
    D --> E[Use of Information  
(Decide what is valuable)]
    E --> F[Synthesis of Information  
(Information restructuring and repackaging)]
  
```

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

CIS

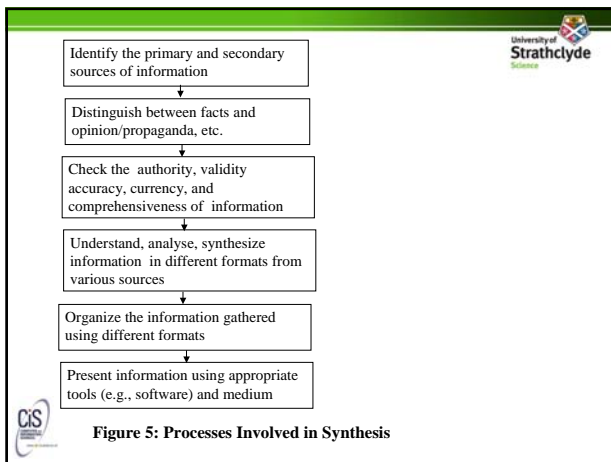
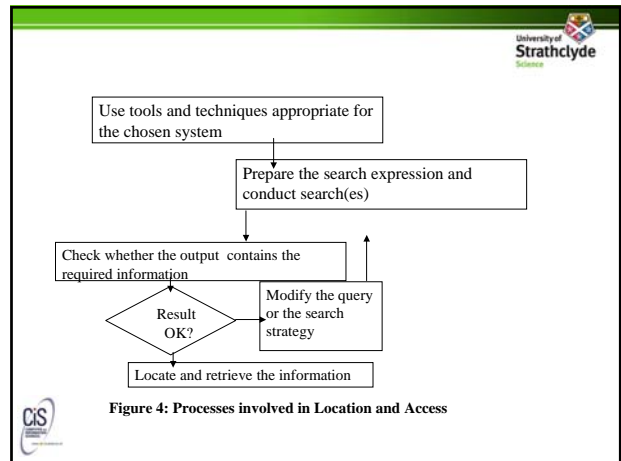
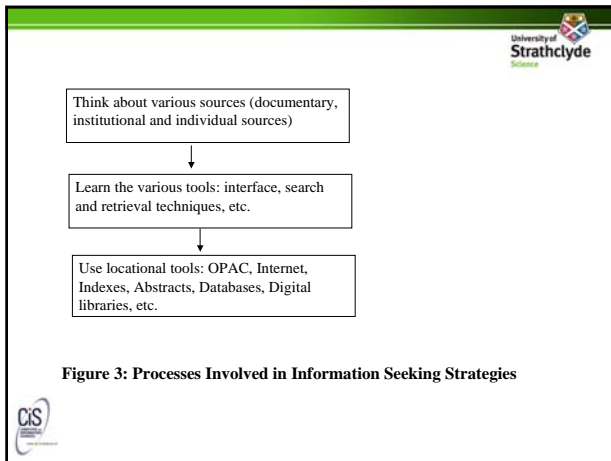
University of Strathclyde
Science

```

graph TD
    A[Determine the exact information required to solve a problem, to accomplish an assignment, etc.] --> B[Identify keywords using dictionaries, thesauri, etc.]
    B --> C[Use WH questions (Who, Which, What, When, Where, etc.) to focus on the topic]
    C --> D[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.]
    D --> E[Prepare the search strategy]
  
```

Figure 2: Processes Involved in Task Definition

CIS



Evaluation of the process


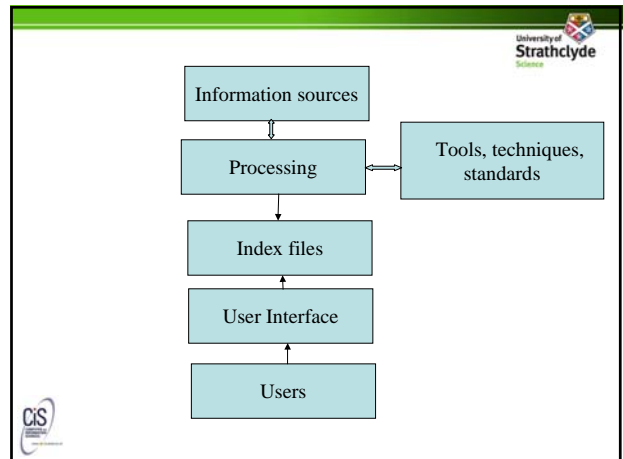
The following questions may be asked at this stage:

- 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?

University of
Strathclyde
Science

Lecture 2


Information Retrieval Systems: From
OPACs to the Web

University of
Strathclyde
Science

Text Retrieval Systems Vs. DBMSs

Text Retrieval	DBMS
Probabilistic	Deterministic
Language/text-based	Value-based
Relevance	Exact data
Vocabulary control	Exact Keys
Partial match	Exact match
Content analysis	Exact values




University of
Strathclyde
Science

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




University of
Strathclyde
Science

Types of IRS (2)

Information Retrieval Systems may be also be categorised as:


- Bibliographic information retrieval systems
- Full text retrieval systems
- Multimedia information retrieval systems
- Knowledge-based systems



University of
Strathclyde
Science

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 searched or browsed through one interface
- Search output contains full text articles
- Many electronic books are just digital version of printed books, such as encyclopaedias, 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 contain 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.

University of Strathclyde
Science

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.)

CIS

University of Strathclyde
Science

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

CIS

The screenshot shows the AskJeeves search engine interface. The search query is "Where can I find comparisons and reviews of search engines?". The results list several websites, including About.com, Alltop, All in One, Alltop, Alltopweb, Advista, AOL, AskJeeves, and AskJeeves on Español. There are also sponsored links and a list of related search engines.

The screenshot shows the About.com search engine interface. The search query is "e-commerce". The results list several articles and guides, including "The ultimate internet marketing resource", "By these search terms to see more answers:", "Top 3 Search Engines", "Google Search Engine", "Top Search Engines", "Offline Search Engines", "Search Engine Optimization", "Local Search Engine", "Search Engine Submission", "Adult Search Engine", and "Search Engines". There are also sponsored links and a list of related search engines.

University of Strathclyde
Science

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

CIS

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

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




University of
Strathclyde
Science

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




University of
Strathclyde
Science

Strathclyde Univ. OPAC



Search Tips

- Title or Annual Title Browse**
Always use table search - ie, title, in, to, etc. eg. Title of two titles, not a list of two titles
For an electronic annual title, choose **Electronic Journals** from the **Quick Links** menu and then type your title in the search box as a **Journal Title Browse** search.
- Name Browse**
By personal names, type names first followed by that name in initial, eg. Birkley, T. For organisations, use central word only, eg. Law Society of Scotland.
All search conditions apply for name, eg. Colquhoun, M. Colquhoun.
NB: If you cannot find what you are looking for with a **Name Browse** search, try **Name keyword or phrase** search instead.
- Subject**
Subject headings are those used by Library of Congress, NB: older editions may not have subject headings
- Title/Name/Annual Title/Subject Keyword**
Type one keyword or a phrase.
use * for truncation, eg. *for business, business, business, business
use + for essential words, eg. +internet
- Keywords Anywhere Search**
Use quotes to search for a phrase - "world wide web"
use + for essential words - +internet
- Dewey/UDC class number**
use the numbering by the Library of Congress or UDC class number
The Dewey numbers with a D prefix, with D followed by a space, followed by the relevant class number eg. D 200 (D200) for books about **Travels in Scotland**.
For bookshelves, enter the number without a D, eg. 200.41. For higher education in Scotland, the numbers can be truncated by using *, eg. D 200 (D200) for all books on **Travels**.
- Quick Links**
NB: Items listed in Quick Links to make it clear
If the searching cannot be published in the last 7 or 10 days



University of
Strathclyde
Science


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



University of
Strathclyde
Science

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




University of Strathclyde Library
prev: Subject Browse - digital libraries
next: Displaying 1 through 40 of 40 entries.

#	Title	Heading	Heading Type
Display 1	17 Digital libraries		LC subject headings
Display 2	2 Digital libraries Administration		LC subject headings
Display 3	1 Digital libraries Canada		LC subject headings
Display 4	1 Digital libraries Collections and preservation		not applicable
Display 5	6 Digital libraries Congresses		LC subject headings
Display 6	1 Digital libraries Encyclopaedia		LC subject headings
Display 7	1 Digital libraries England Literature		LC subject headings
Display 8	1 Digital libraries Great Britain		LC subject headings
Display 9	4 Digital libraries Great Britain		LC subject headings
Display 10	2 Digital libraries United States Congresses		LC subject headings
Display 11	9 Digital libraries		LC subject headings

University of
Strathclyde
Science

Strathclyde Univ. OPAC: Subject Keyword Search Results



University of Strathclyde Library Catalogue


prev: Displaying 1 through 33 of 33 entries.

#	Title	Author	Publication Date
1	Building an electronic resource collection : a practical guide / Stuart D. Lee	Lee, Stuart D.	2002
2	There are various holdings for this title		
3	Creating a successful e-information service / Sheila Peatry and Peter Oulthwa	Peatry, Sheila	2002
4	Main Library Lending D 025 02/02 FANT	check shelves	
5	Digital futures : strategies for the information age / Marilyn Dwyer, Simon Tennant	Dwyer, Marilyn	2002
6	Main Library Lending D 025 02/02 DBER	various authors	
7	Digitising collections : storage issues for the information manager / Lorna Hughes and David Green	Hughes, Lorna	2002
8	Main Library J Week Loan		
9	Libraries, the Internet, and eScholarship : tools and trends converging / edited by Charles F. Thomas		2002

University of
Strathclyde
Science

IULA Lecture 3 (2)


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



University of
Strathclyde
Science

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



University of
Strathclyde
Science

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)



University of
Strathclyde
Science

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



University of
Strathclyde
Science

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




University of
Strathclyde
Science

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*.




University of Strathclyde
Science

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




University of Strathclyde
Science

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



University of Strathclyde
Science

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




University of Strathclyde
Science

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>




University of Strathclyde
Science

Proquest Basic Search Screen




The screenshot shows a search interface with a search bar at the top. Below it, there are several search criteria fields: 'Date range' (set to Current (2000 - Present)), 'Publication type' (set to All), and 'Search in' (set to All Fields). There are also checkboxes for 'Show results with full text availability only', 'Show articles that have removed publication info', and 'Show total number of articles'. A 'Basic Search' button is visible on the right.




University of Strathclyde
Science

Proquest Guided Search Screen



The screenshot shows a search interface similar to the basic search screen but with a 'Guided Search' button on the right. The search criteria fields are the same, but there are additional options for 'Search in' (e.g., All Fields, All Fields except Abstract, All Fields except Title, etc.) and 'Search type' (e.g., All, All except Abstract, All except Title, etc.). There are also checkboxes for 'Show results with full text availability only', 'Show articles that have removed publication info', and 'Show total number of articles'. A 'Guided Search' button is visible on the right.



University of Strathclyde
Science

Proquest Advanced Search Screen

CIS

University of Strathclyde
Science

Proquest Thesaurus

CIS

University of Strathclyde
Science

Proquest Thesaurus (2)

CIS

University of Strathclyde
Science

Ingenta Search Screen

CIS

University of Strathclyde
Science

Ingenta Advanced search Screen

CIS

University of Strathclyde
Science

Ingenta Search Results

CIS

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., Yahoo!igans, 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

University of
Strathclyde
Science

IULA Lecture 3 (3)


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



University of
Strathclyde
Science

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




University of
Strathclyde
Science

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



University of
Strathclyde
Science

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



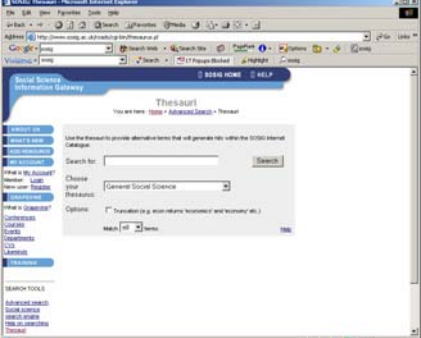

University of
Strathclyde
Science

SOSIG Home Page




University of
Strathclyde
Science

SOSIG Thesauri

University of Strathclyde
Science

Special Search Engines: Vivisimo

CIS

University of Strathclyde
Science

Special Search Engines: Kartoo

CIS

University of Strathclyde
Science

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

CIS

University of Strathclyde
Science

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


CIS

University of Strathclyde
Science

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

CIS





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/) (<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.



University of Strathclyde Science

PIE

(<http://www.dlib.org/dlib/march01/gambles/03gambles.html>)

CIS

University of Strathclyde Science

MyLibrary@NCState

a customizable HTML page accessible via any Web browser
 The process of creating a MyLibrary@NCState page begins when the user creates an account on the system.
 The account information includes rudimentary biographical data: name and email address. In addition, the creation of an account necessitates the selection of a primary academic interest from a pop-up menu.

CIS

University of Strathclyde Science

MyLibrary@NCState (2)

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:
<http://www.ted.cmis.csiro.au/sigir99/morgan/>

CIS

Message: MyLibrary

MY LIBRARY @NCState

Welcome back, Enter

"Who will own your scholarly work?" a guest column by Peggy Hoon in the Friday, January 21, 1999 issue of the *Journal* invites the campus community to the upcoming communication colloquium.

Message from the librarian:

NORTH CAROLINA STATE UNIVERSITY LIBRARY SERVICES

The NC State University faculty, staff & student publications database is a listing of citations representing the published output of those individuals connected with North Carolina State University since 1997. It has been compiled and is maintained by the JCSU Librarians. Please check the public access listed under your name in the number listings and use the online forms to notify us of your publications.

5/1/99

Your Librarians:

There are many ways to get in touch with and get help from the library:

- The information needs contact: libref@lib.ncsu.edu (115-5908)
- To request a purchase contact: libref@lib.ncsu.edu (115-5900)

libref@lib.ncsu.edu

Personal Links:

- Home
- Get help/contact
- Search/Discover
- Log In/Sign Out
- My Library
- My Recent Activity
- My Recent Activity
- My Recent Activity

Quick Searches:

Dictionary

Search

Reference Shelf:

- Ask a Librarian
- Internet Library for Librarians
- Library Resources on the Internet

Bibliographic Databases:

- AGU/AGU/AGU/AGU
- AGU/AGU/AGU/AGU
- AGU/AGU/AGU/AGU
- AGU/AGU/AGU/AGU

University of Strathclyde Science

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

CIS

University of Strathclyde Science

DWE

CIS

Information for Specific User Categories

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

Glasgow Health Information Gateway (<http://www.ghi.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 of the various resource/service and (linking to the resources/services
- Personalization of the web page for information resource/service selection and/or display

University of
Strathclyde
Science

IULA Lecture 4, Part2: DWE

Gobinda Chowdhury



University of
Strathclyde
Science

User-Centred Research in the DL Environment

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




University of
Strathclyde
Science

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




University of
Strathclyde
Science

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



University of
Strathclyde
Science

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



University of
Strathclyde
Science

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

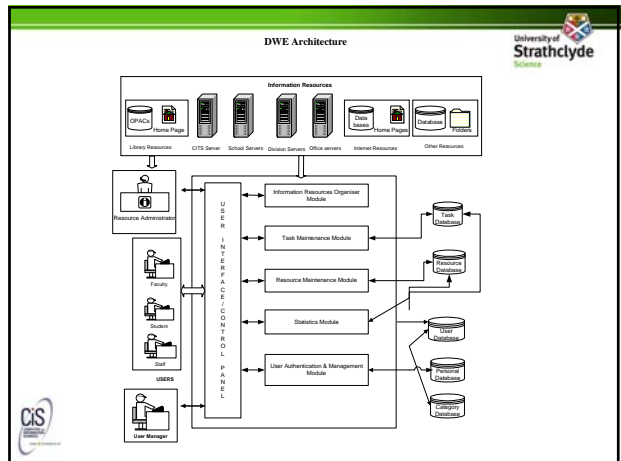


University of
Strathclyde
Science

DWE : an overview

Architecture
Components and workflow
Advantages
Problems
Prospects/future research

CIS



University of
Strathclyde
Science

DWE Components

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

CIS

University of
Strathclyde
Science

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

CIS

University of
Strathclyde
Science

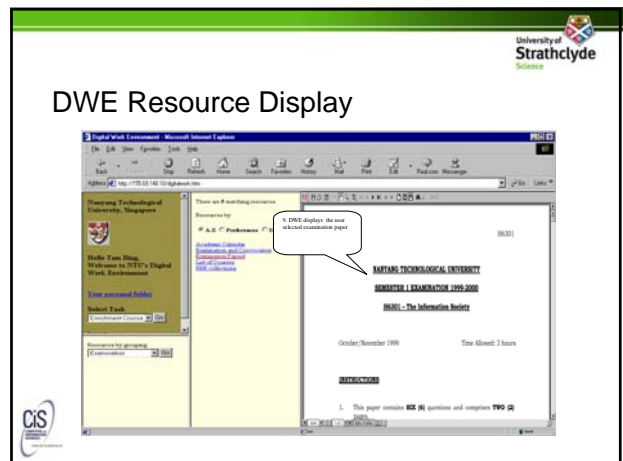
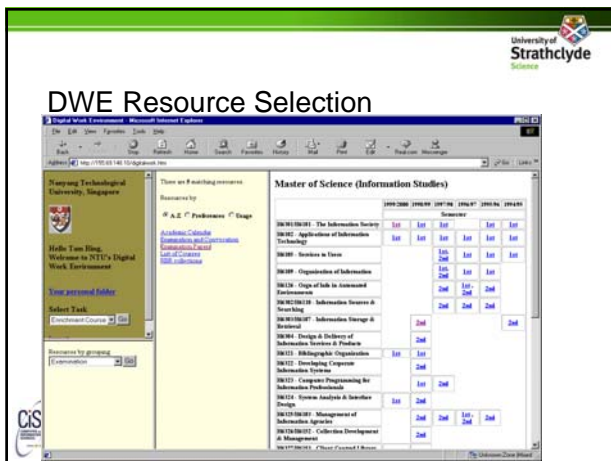
DWE Interface

CIS

University of
Strathclyde
Science

DWE Task Frame

CIS



DWE Advantages

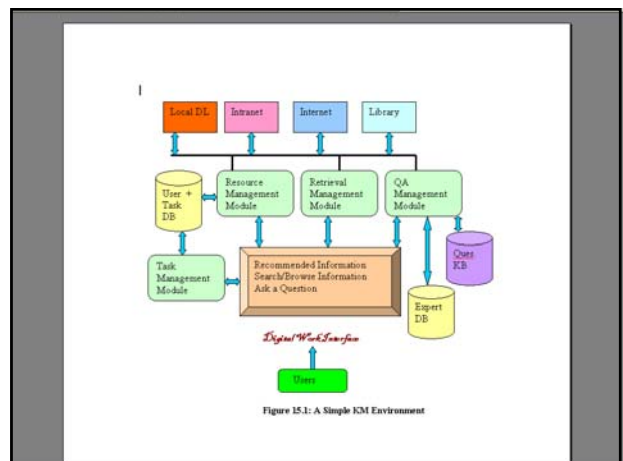
- One-stop window for access to information
- Recommended information resources based on user tasks
- Search facility based on metadata or full text
- Personal note space for every user

DWE Problems

- Identification and mapping of information resources : manual and resource-intensive
- Creation of metadata for every resource is 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



University of
Strathclyde
Science

IULA Lecture 5: Language issues in IR


Gobinda Chowdhury



University of
Strathclyde
Science

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




University of
Strathclyde
Science

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



University of
Strathclyde
Science

Dewey Main Classes


000	Computer science, information & general works
100	Philosophy & psychology
200	Religion
300	Social sciences
400	Language
500	Science
600	Technology
700	Arts & recreation
800	Literature
900	History & geography



University of
Strathclyde
Science

Main Classes: Second level


000	Computer science, information & general works
000	Computer science, knowledge & systems
010	Bibliographies
020	Library & information sciences
030	Encyclopedias & books of facts
040	[Unassigned]
050	Magazines, journals & serials
060	Associations, organizations & museums
070	News media, journalism & publishing
080	Quotations
090	Manuscripts & rare books



University of
Strathclyde
Science

Main Classes

000	Computer science, information & general works
000	Computer science, knowledge & systems
000	Computer science, information, general works
001	Knowledge
002	The book
003	Systems
004	Data processing Computer science
005	*Computer programming, programs, data
006	*Special computer methods
[007]	[Unassigned]
[008]	[Never assigned]
[009]	[Never assigned]



University of Strathclyde
Science

Main Classes

000 Computer science, information & general works

000 Computer science, knowledge & systems

005 ***Computer programming, programs, data**

005.019 Psychological principles

005.0684 Software consultants

005.1-005.5 Computer programming and programs

005[.6] Microprogramming and microprograms

005.7 *Data in computer systems

005.8 *Data security

CIS

University of Strathclyde
Science

Computer science, information & general works

Computer science, knowledge & systems

***Computer programming, programs, data**

Computer programming and programs

***Programs**

Standards

Testing and measurement

Maintenance and repair

Commercial miscellany

Special topics

Computer software--marketing management

Programs for digital computers

Programs for specific processing modes

Programs for specific operating systems and for specific

*Programs for hybrid and analog computers

CIS

University of Strathclyde
Science

LC: main classes

[1. GENERALITIES](#)

[2. PHILOSOPHY, PSYCHOLOGY](#)

[3. RELIGION, THEOLOGY](#)

[4. SOCIAL SCIENCES](#)

[5. NATURAL SCIENCES](#)

[6. TECHNOLOGY](#)

[7. THE ARTS](#)

[8. LANGUAGE, LINGUISTICS, LITERATURE](#)

[9. GEOGRAPHY, BIOGEOGRAPHY, HISTORY](#)

CIS

University of Strathclyde
Science

UDC Main classes

[0GENERALITIES](#)

[1PHILOSOPHY, PSYCHOLOGY](#)

[2RELIGION, THEOLOGY](#)

[3SOCIAL SCIENCES](#)

[4VACANT](#)

[5NATURAL SCIENCES](#)

[6TECHNOLOGY](#)

[7THE ARTS](#)

[8LANGUAGE, LINGUISTICS, LITERATURE](#)

[9GEOGRAPHY, BIOGEOGRAPHY, HISTORY](#)

CIS

University of Strathclyde
Science

Use of Classification in Internet IR

- Several classification schemes have been used to organize and to 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

CIS

University of Strathclyde
Science

Advanced Search

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

000 Generalities
Includes: computing, Internet, libraries, informatics science

100 Philosophy and psychology
Includes: ethics, parapsychology, phenomenology

200 Religion
Includes: Bible, religions of the world

300 Social sciences
Includes: sociology, politics, economics, law, education

400 Language
Includes: linguistics, language learning, specific languages

500 Science and mathematics
Includes: physics, chemistry, earth sciences, biology, ecology

600 Technology
Includes: medicine, engineering, agriculture, management

700 The arts
Includes: art, planning, architecture, music, sport


800 Literature and rhetoric
Includes: literature of specific languages

900 Geography and history
Includes: travel, genealogy, urbanology

BIBEL uses the Dewey Decimal Classification system as the primary organization structure for its catalogue of Internet resources.
The Dewey Decimal Classification is (c) 1996-2002 OCLC Online Computer Library Center. Used with permission.


[E-LIB | OCLC | Contacts and Credits](#)

BUBL Information Service, Centre for Digital Library Research, Strathclyde University, Glasgow G1 1XH, Scotland
Tel 0141 548 4752 Email bubl@strath.ac.uk



IULA Lecture 5: Language issues in IR, Part 2

Gobinda Chowdhury





Search for | [Advanced Search](#) | [Feedback](#)

by Subject [Map of the SOGSI sections](#)

epidology	Law
ew and Management	Philosophy
omics	Politics
ation	Psychology
amental Science	Research Tools and Methods
ean Studies	Social Welfare
aphy	Sociology
ment Policy	Statistics
	Women's Studies

Home | [Database](#) | [URL](#)

[SOGSI Home](#) | [About IULA](#) | [About Feedback](#) | [Site Map](#) | [Contact Us](#)

© 2005 SOGSI  All rights reserved. 

Broader terms	Narrower terms	Related terms
<ul style="list-style-type: none"> ECONOMICS 	<ul style="list-style-type: none"> CAPITAL CAPITAL FLOW COMMUNICATION FINANCE CREDIT FINANCIAL ACCOUNTING FINANCIAL COMPANIES FINANCIAL EXPECTATIONS FINANCIAL INCENTIVES FINANCIAL INSTITUTIONS FINANCIAL MANAGEMENT FINANCIAL MARKETS FINANCIAL RESOURCES FINANCING HOUSEHOLD BUDGETS INFORMATION LIBRARY FINANCE INSURANCE INVESTMENTS LABOR AND FREEDOM FINANCE MONETARY ECONOMICS MONEY OVERSEAS TRANSACTIONS PENSION CONTRIBUTIONS REGIONAL FINANCIAL MANAGEMENT PUBLIC FINANCE REAL ESTATE FINANCE RESEARCH FINANCE SAVINGS SHARES UNEMPLOYMENT 	<ul style="list-style-type: none"> FINANCIAL POLICY INCOME INSURANCE CLAIMS INSURANCE PREMIUMS

To search the catalogue for one or more terms, check boxes to include Narrower terms and/or Related terms in your search then click on the "Search on" button to explore the database further, select a highlighted term.

Summary | [Show](#) | [Sort by](#) | [Send to](#)

1: [Ortiz-Roldan D, Ojeda L, Tzeng L, Shera C, Ortiz AM, Ocasio C.](#)
A Formal Total Synthesis of Drotrecovirin Using the Ring-Closing Metathesis (RCM) Reaction of a Densely Functionalized Diene as the Key Step: Investigation of the Human RCM Stereochemistry
Chemistry 2005 Sep 20; 19(36):6671-6676
PMID: 16171000 [PubMed - as supplied by publisher]

2: [Shanov A, Kucharski AV, Tarkenton GP, Choudhury AA, Choudhury N, Akhmedov LA, Bakov D, Akhmedov M, Bakov YH, Olenovskiy VV, Bepko VV, Rubtsov BH, Hovsepian JV.](#)
Molecular epidemiology of the hepatitis C virus in Western Siberia
J Med Virol 2005 Sep 19; 77(9):362-369 [Epub ahead of print]
PMID: 16172011 [PubMed - as supplied by publisher]

3: [Mitsunaka M, Chagnac AN, Lippman R.](#)
Tuberculosis of the gastrointestinal tract: imaging features with pathological correlation.
Gastroenterology 2002 Oct; 123(4):1067-75.
PMID: 16172011 [PubMed - as supplied by publisher]

4: [Cohen E, Eyalon JL, Shitka E, Pines E, Moshiri AC, Shitka AA.](#)
Elevated serum cholesterol in a patient with acquired immunodeficiency syndrome treated by zalcitabine: a case report.
Rev Soc Bras Med Trop 2005 Sep-Oct; 38(4):438-40. Epub 2005 Sep 12.
PMID: 16172013 [PubMed - as supplied by publisher]

5: [Mitsunaka M.](#)
Outbreak of hepatitis C virus infection in patients with acquired immunodeficiency syndrome.
Eur J Clin Microbiol Infect Dis 2004 Sep; 29(9):165-70.
PMID: 16172014 [PubMed - as supplied by publisher]

6: [Tobias AM, Aguilera CA, Rosen DG, Cohen JE, DE Bekos C, Tolosa A, Arango M, Cian JF, Sampedro J, Partridge A.](#)
DISSEMINATED HISTOPLASMA: A COMPARATIVE STUDY BETWEEN PATIENTS WITH ACQUIRED IMMUNODEFICIENCY SYNDROME AND NON-HUMAN IMMUNODEFICIENCY VIRUS-INFECTED INDIVIDUALS
Am J Trop Med Hyg 2005 Sep; 73(3):336-342.
PMID: 16172014 [PubMed - as supplied by publisher]

7: [Huang SP, Swales CT, Sun JH, Kohli A, Jayaraman A, He CT, Cohen DE, Owsen EA, Han Y, Che H, Hwang S, Chen JF.](#)
Asymptomatic hepatitis B virus infection among human immunodeficiency virus-infected patients in Taiwan.
Am J Trop Med Hyg 2005 Sep; 73(3):330-2.
PMID: 16172015 [PubMed - as supplied by publisher]

Full | [Show](#) | [Sort by](#) | [Send to](#)

1: [Viral Vaccines](#)
Vaccines or candidate vaccines containing inactivated HIV or some of its component antigens and designed to prevent or treat AIDS. Some vaccines containing antigens are *Year introduced: 1993*

Subheadings: This list includes those paired at least once with this heading in MEDLINE and may not reflect current rules for allowable combinations.

administration and dosage adverse effects antigens and adjuvants biosynthesis blood chemical synthesis chemistry classification control genetics history immunology isolation and purification metabolism pharmacokinetics pharmacology standards supply and distribution toxicity

Restrict Search to Major Topic headings only
 Do Not Exclude this term (x), do not include MeSH terms found below this term in the MeSH tree).

Entry Terms:

- HIV Vaccines

Previous heading:

- [Vaccines \(1987-1993\)](#)

All MeSH Categories

- Chemical and Drug Category
- Immunologic and Biological Factors
- Biological Products
- Vaccines

Viral Vaccines

[AIDS Vaccines](#)
[Cytomegalovirus Vaccines](#)
[Ebola Vaccines](#)
[Hepatitis Vaccines](#)
[Herpesvirus Vaccines](#)
[Influenza Vaccines](#)
[Japanese Encephalitis Vaccines](#)
[Measles Vaccines](#)
[Mumps Vaccines](#)
[Rubella Vaccines](#)
[Tuberculosis Vaccines](#)
[Yellow Fever Vaccines](#)

Full | [Show](#) | [Sort by](#) | [Send to](#)

1: [Viral Vaccines](#)
Suspensions of attenuated or killed viruses administered for the prevention or treatment of infectious viral diseases. *Year introduced: 1965*

Subheadings: This list includes those paired at least once with this heading in MEDLINE and may not reflect current rules for allowable combinations.

administration and dosage adverse effects analysis antigens and adjuvants biosynthesis blood chemical synthesis chemistry contraindications diagnostic use economics genetics history immunology isolation and purification metabolism pharmacokinetics pharmacology reduction effects serovirus standards supply and distribution therapeutic use toxicity

Restrict Search to Major Topic headings only
 Do Not Exclude this term (x), do not include MeSH terms found below this term in the MeSH tree).

Entry Terms:

- Vaccines, Viral

All MeSH Categories

- Chemical and Drug Category
- Immunologic and Biological Factors
- Biological Products
- Vaccines


Viral Vaccines

[AIDS Vaccines](#)
[Cytomegalovirus Vaccines](#)
[Ebola Vaccines](#)
[Hepatitis Vaccines](#)
[Herpesvirus Vaccines](#)
[Influenza Vaccines](#)
[Japanese Encephalitis Vaccines](#)
[Measles Vaccines](#)
[Mumps Vaccines](#)
[Rubella Vaccines](#)
[Tuberculosis Vaccines](#)
[Yellow Fever Vaccines](#)

University of
Strathclyde
Science

IULA Lecture 6: using controlled languages to build the Semantic Web


Gobinda Chowdhury



University of
Strathclyde
Science

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



University of
Strathclyde
Science

Semantic Web: the W3C vision


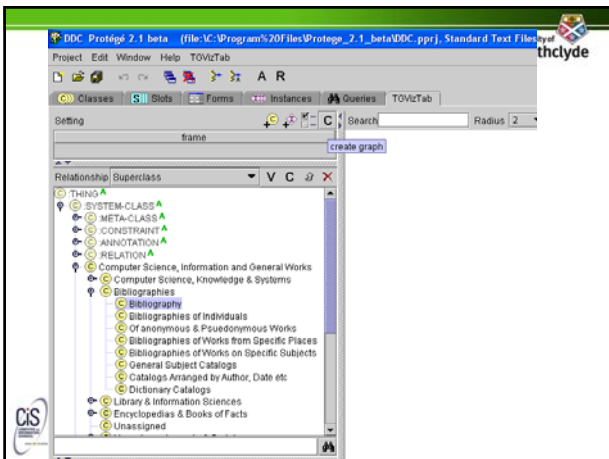
*The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description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



University of
Strathclyde
Science

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

DDC Protege 2.1 beta (file: C:\Program\201 files\Protege_2.1_beta\DDC.ppr) Standard Text Files

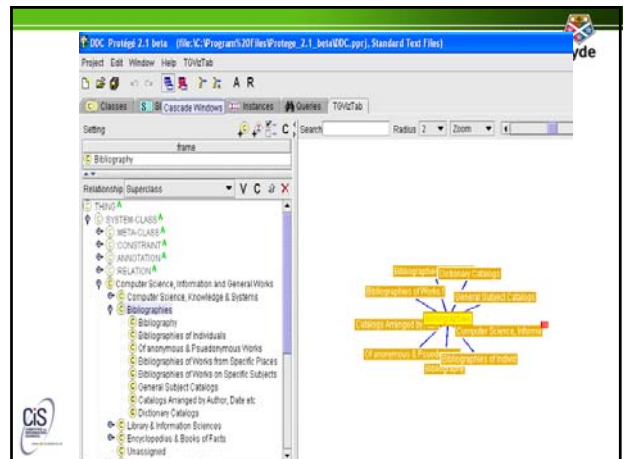
Project Edit Window Help TOVizTab

Classes Slots Forms Instances Queries TOVizTab

Setting frame create graph Radius: 2

Relationship Superclass V C X

- THING A
 - SYSTEM-CLASS A
 - META-CLASS A
 - CONSTRAINT A
 - ANNOTATION A
 - RELATION A
 - Computer Science, Information and General Works
 - Computer Science, Knowledge & Systems
 - Bibliographies
 - Bibliographies of Individuals
 - Of anonymous & Pseudonymous Works
 - Bibliographies of Works from Specific Places
 - Bibliographies of Works from Specific Subjects
 - General Subject Catalogs
 - Catalogs Arranged by Author, Date etc.
 - Dictionary Catalogs
 - Library & Information Sciences
 - Encyclopedias & Books of Facts
 - Unassigned



DDC Protege 2.1 beta (file: C:\Program\201 files\Protege_2.1_beta\DDC.ppr) Standard Text Files

Project Edit Window Help TOVizTab

Classes Cascade Windows Instances Queries TOVizTab

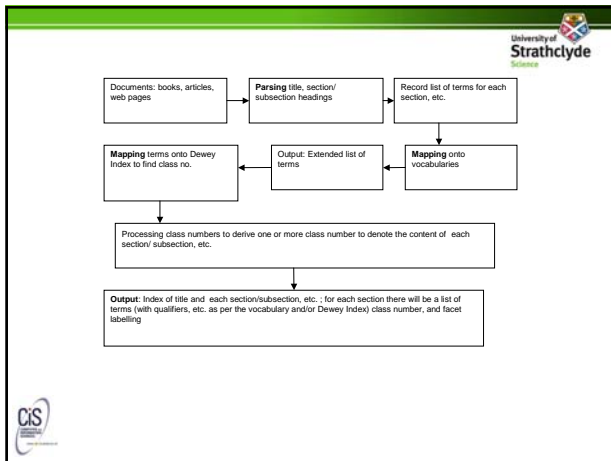
Setting frame Search Radius: 2 Zoom: 1

Relationship Superclass V C X

Bibliography

THING A

- SYSTEM-CLASS A
- META-CLASS A
- CONSTRAINT A
- ANNOTATION A
- RELATION A
- Computer Science, Information and General Works
 - Computer Science, Knowledge & Systems
 - Bibliographies
 - Bibliographies of Individuals
 - Of anonymous & Pseudonymous Works
 - Bibliographies of Works from Specific Places
 - Bibliographies of Works from Specific Subjects
 - General Subject Catalogs
 - Catalogs Arranged by Author, Date etc.
 - Dictionary Catalogs
 - Library & Information Sciences
 - Encyclopedias & Books of Facts
 - Unassigned



- ## 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