Course Outline

Introduction to the course

This course covers the following:

• Characteristics and structure of textual records: definition, content, structure and context; elements of record metadata.
• Databases of textual records: databases as collections of textual records, categorisation of database types, contrast and comparison with other types of databases, eg relational, electronic record keeping principles.
• Textual information retrieval principles: boolean operators, proximity operators, limit operators, truncation, inverted indexes, keyword versus phrase indexing, controlled vocabulary and thesaurus use versus uncontrolled keyword searching, retrieval command languages, set logic and construction for retrieval purposes.
• Construction and implementation of search strategies: search sequence diagrams, query expansion, broadening and narrowing search results, strategies to avoid information overload.
• Advanced retrieval features: relevance feedback, introduction to weighting and probabilistic retrieval.
• Information retrieval systems for specific information environments: libraries, records management systems, etc.
• Basic design and creation of text-based databases using information retrieval systems: data structures, documentary and management metadata elements and their properties, data entry or conversion requirements, data output techniques.

Objectives of the course

• To review the general knowledge of the role of description, classification and indexing in information management, and particular knowledge of how the principles may be applied to information retrieval (IR) systems;
• To provide an ability to apply the theories and principles of information organisation by designing and creating a text-based database using IR systems packages;
• To give an appreciation of the effects of file structure on IR capability;
• To compare and contrast the capabilities of commercial document or IR systems with those of Database Management Systems (DBMS);
• To provide experience in online searching on a range of Web and non-Web based bibliographic and full text databases using general and advanced IR principles.

Academic Staff

Lecturer-in-charge (LIC): A/Prof C.S. Wilson (Room: Quad 2087; Ph: 9385-7134; email: c.wilson@unsw.edu.au)
Tutor: Ms. Margaret Lo (Room: Quad 2100; Ph: 9385-7126; email: m.lo@unsw.edu.au)
Student Consultation Times

Mondays: 1 – 1:45 pm
Thursdays: 2 – 4:00 pm

Other consultation times may be arranged by emailing the lecturer. When sending an email, please state your question or concern, as often I can answer your questions, etc. by return email.

Course Web Page

The course lecture notes, exercises, other handouts and announcements will be available via WebCT to enrolled students at the following address: http://www.webct.unsw.edu.au
User name: <your student ID number preceded by the ‘z’>
Password: your Unipass number (remember this is case sensitive and consists of letters and numbers). After logging in, a list of the subjects you have enrolled in will show on the webpage. Click on IMGT5110 and all course material will be there.

Delivery Mode and Venue

Lectures are on Mondays from 2-4pm in RC (Red Centre) M032 (Mezzanine – Sch of Mathematics, upper end of the Red Centre Building); attendance is required and you will need to ‘sign-in’ each week.

Tutorials in computer laboratory are on Mondays from 4-5pm in Quad Lab 6; although attendance at the tutorial time is ‘not’ compulsory (that is, you do not have to ‘sign-in’), it is highly recommended. Assignments (e.g., weekly computer exercises and the Database report) can be done outside of the regular tutorial time slot; however, questions dealing with the exercises and the major assignment will be answered during the tutorials; therefore, students are expected to attend the tutorials to obtain help directly relating to the class assignments.

Computer Laboratories

Students will be attending computer laboratories from Weeks 1 to 14. Various exercises are set up for students to work through the tutorial-hour and later, in their own time. Two of these weekly exercises will be handed in (for assessment) via email attachments; due dates (generally) will be on the Fridays after the Monday tutorials. Students are required to bring their own formatted diskettes for saved files or for any downloading of information.

Workload Expectations

It is expected that you will spend at least ten hours per week studying this course. This time should be made up of reading and research; working on exercises, problems and assignments; performing computer tasks and attending classes. In periods where you need to complete assignments or prepare for the examination, the workload may be greater.

Over commitment has been a cause of failure for many students. You should take the required workload into account when planning how to balance study with employment and other activities.

Assessment/Assignment

<table>
<thead>
<tr>
<th>Task</th>
<th>Date Due</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database design, construction, evaluation and comparison of an information retrieval (IR) system for a specific information environment with that of a relational database management system (DBMS).</td>
<td>Week 10: Tuesday, 10\textsuperscript{th} of May, 5pm</td>
<td>45%</td>
</tr>
<tr>
<td>Information Retrieval Lab (selected) exercises: two, each worth 5%.</td>
<td>Submitted on Fridays after the tutorials</td>
<td>10%</td>
</tr>
<tr>
<td>Class participation in discussions</td>
<td>continuously</td>
<td>5%</td>
</tr>
<tr>
<td>Open book examination</td>
<td>Exam period</td>
<td>40%</td>
</tr>
</tbody>
</table>
Please Note

- 80% attendance at lectures is required and an attendance sheet must be signed at every lecture. This is a general University policy.
- Two Lab exercises will be selected for assessment.
- Late submission of assignments will incur a penalty of 10 percent of the maximum assessment mark per day. An extension (without penalty) on the time of submission will only be granted under exceptional circumstances by the Lecturer-in-charge. In all cases documented evidence must be provided.
- The database design, construction, evaluation (and comparison) assignment must be submitted as a printed report including a title page and stapled in the top left-hand corner or placed in a binder. The title page should include the subject name, student name, student number, assignment details and the following statement confirming originality of the work:
  "I certify that this assignment is my own work in which my sources are acknowledged and which I submit for the first time."
- More detailed information for the database assignment will be given to students by Week 4.
- Students must obtain a Pass (that is 50% or better) in the two major components (worth 45% and 40%) of the course in order to pass the course.

Academic Misconduct

Students are reminded that the University regards academic misconduct as a very serious matter. Students found guilty of academic misconduct are usually excluded from the University for 2 years. Because of the circumstances in individual cases the period of exclusion can range from one session to permanent exclusion from the University.

The following are some of the actions which have resulted in students being found guilty of academic misconduct in recent years:

- taking unauthorised materials into an examination;
- submitting work for assessment knowing it to be the work of another person;
- improperly obtaining prior knowledge of an examination paper and using that knowledge in the examination.
- failing to acknowledge the source of material in an assignment, e.g., the database report for this course.

Students are reminded of the penalties for plagiarism and the fact that action will be taken in all cases where plagiarism is suspected. Plagiarism entails taking and using as one's own, the thoughts or writings of another without acknowledgement, including:

- where paragraphs, sentences, a single sentence or significant part of a sentence which are copied directly, are not enclosed in quotation marks and appropriately footnoted;
- where direct quotations are not used, but ideas or arguments are paraphrased or summarised, and the source of the material is not acknowledged either by footnoting or other reference within the text of the paper; and
- where an idea, which appears elsewhere in print, film or electronic medium, is used or developed without reference being made to the author or the source of the idea.

For further information on academic misconduct, please see the following URLs:


https://my.unsw.edu.au/student/academiclife/assessment/AcademicMisconductStudentMisconduct.html
Citations and Referencing

If you are uncertain about how to cite (that is, reference) the work of others please refer to the collection of resources about citations and referencing on the UNSW Library Web Site. This can be found at the following URL:

http://www.library.unsw.edu.au/links/Research_and_Study_Skills/Citing_References/ The style you choose (for example, Referencing: The 'In-Text' or Harvard Method) should be applied consistently throughout your report.

Education Development Unit

Learning support, tailored to the needs of FCE students, is available from the Education Development Unit (EDU) in the Faculty. The EDU offers a range of services for FCE students including:

- Academic skills workshops run throughout the session;
- Printed and on-line study skills resources e.g. referencing guide, report writing and exam preparation;
- A drop-in resource centre containing books and audio visual material that can be borrowed;
- A limited consultation service for students with individual or small group learning needs.

More information about the EDU services including on-line resources, workshop details and consultation request forms are available from the EDU website.

Contacts and location:
EDU Web:   http://education.fce.unsw.edu.au
EDU Location:   Room 2039, Level 2 Quadrangle Building
EDU Email:   edu@unsw.edu.au
EDU Phone:  9385 5584
EDU Fax:  9385 6061
EDU Opening Hours:  Mon 1-5pm; Tues, Wed, Thurs 10-5pm; Fri 10-1pm

EDU services are free and confidential and are available to students of the Faculty of Commerce and Economics.

Other UNSW Support

In addition to the EDU services, the UNSW Learning Centre provides academic skills support services for students. The Learning Centre is located on Level 2 of the Library and can be contacted by phone: 9385 3890; by email; l.puni@unsw.edu.au; or through their website: http://www.lc.unsw.edu.au/. Students experiencing problems of an academic or personal nature are encouraged to contact the Counselling Service at UNSW. This service is free and confidential and run by professional counsellors. The Counselling Service is located on Level 2, Quadrangle East Wing, and can be contact on 9385 5418.

References and reading material

Books (in alphabetical order by first author):

Anderson, J.D.; Pérez-Carballo, J. Information retrieval design: Principles and options for information description, organization, display and access in information retrieval databases, digital libraries, catalogs and indexes, St. Petersburg, FL: Ometeca Institute. 2005. [on order]


Inmagic DB/TextWorks User's Manual. (There are online versions available – latest is Version 8).


Van Rijsbergen, C.J. The geometry of information retrieval / C.J. Van Rijsbergen. Cambridge: Cambridge University Press. [P025.04/290]


Wolfram, D. Applied informetrics for information retrieval research. Westport, CN: Libraries Unlimited. 2003 [on order?]

Selected Journal Articles (in alphabetical order by author):


**Journals useful for browsing – use the Library’s electronic journals to search:**

*Information Processing & Management.*

*Online information review.*

*E-doc: Guide to technologies driving e-business.*

*Information Fusion.*

*Proceedings of the ... Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.*

*Proceedings of the ... International Conference on Information and Knowledge Management Information retrieval.*

*EContent.*

*Journal of the American Society for Information Science & Technology (JASIST)*

**Major databases useful for information on the subject of ‘information retrieval’:**

INSPEC (available through the Library and on the Dialog information system; database contains mostly journal and conference materials in Computer Science, Electrical and Electronic Engineering and Physics)

SCI – *Science Citation Index* (available through the Library as part of the *Web of Science* and on the Dialog information system)

SSCI – *Social Sciences Citation Index* (available through the Library as part of the *Web of Science* and on the Dialog information system)

LISA – *Library and Information Science Abstracts* (available through the Library and on the Dialog information system)

**Additional references:**

http://www.dialog.com and more specific links (to be distributed later in class)

http://www.cs.jhu.edu/~weiss/ir.html and other links

http://www.gslis.utexas.edu/~palmquis/inforets99/


http://www.lis.pitt.edu/~nhprc/evidence.html
## Lecture and Tutorial Outline

<table>
<thead>
<tr>
<th>Week</th>
<th>Lecture*</th>
<th>Computer Tutorial*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction and overview of IR Systems</td>
<td>Familiarity with the Faculty’s computer lab facilities; ‘Warm-up’ search exercise</td>
</tr>
<tr>
<td>2</td>
<td>Introduction to searching techniques; Keyword and concept searching; Electronic information sources; Boolean logic; Inverted files; Field searching – fields.</td>
<td>Allocation of Dialog user numbers/passwords Search exercise using various databases</td>
</tr>
<tr>
<td>3</td>
<td>Structure of IR System records. Dialog: Electronic Blue sheets; Web versus DialogLink (versus Telnet) access; Databases (types: bibliographic, statistical, full text, newspapers, etc.); Dialog command syntax</td>
<td>Dialog search exercise; Inverted Index exercise</td>
</tr>
<tr>
<td>4</td>
<td>Controlled vocabulary: Thesauri and Descriptor field; Free text searching; Phrase versus keyword searching.</td>
<td>Dialog exercises using controlled vocabulary fields.</td>
</tr>
<tr>
<td></td>
<td><strong>RECESS WEEK</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DB/TextWorks – database design and construction; Reports and screen designs. Advanced searching techniques including proximity operators, etc.</td>
<td>DB/TextWorks DB creation; reports and screen designs.</td>
</tr>
<tr>
<td>7</td>
<td>DB/TextWorks: Advanced (optional) features including images, semi-relational capability; Contrast between DB/TextWorks and Microsoft Access</td>
<td>DB/TextWorks continued and exploration of the Microsoft Access database system.</td>
</tr>
<tr>
<td>8</td>
<td><strong>Public Holiday (No class, however, see Week ‘15’ below)</strong></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>The Internet – Search engines and searching techniques on the Internet vis-à-vis other IR systems such as Dialog</td>
<td>Search exercise using various Internet search engines.</td>
</tr>
<tr>
<td>10</td>
<td>Citation indexing; Structure of a citation index.</td>
<td>Web of Science exercise</td>
</tr>
<tr>
<td>11</td>
<td>Citation indexing continued. Statistical and ranking features in Dialog.</td>
<td>Citation searching, ranking, etc.</td>
</tr>
<tr>
<td>12</td>
<td>Advanced features including Dialog Target and Rank features; Internet relevance feedback.</td>
<td>Target vs Advanced AltaVista searching</td>
</tr>
<tr>
<td>13</td>
<td>Advanced features of IR systems continued; Evaluation of IR systems: Recall and Precision</td>
<td>Further Advanced searching on the Dialog Information System</td>
</tr>
<tr>
<td>14</td>
<td>Current and future research in IR</td>
<td>Further Advanced searching in various IR systems</td>
</tr>
<tr>
<td>“15”</td>
<td>Study week – exam revision and answers to FAQ</td>
<td>Venue TBA</td>
</tr>
</tbody>
</table>

* Note that the sequence of lectures and tutorials may change.