SUBJECT DESCRIPTION FORM

Subject Title: Internet Information Retrieval

Subject Code: COMP5324

Credit Value: 3

Pre-requisite (Subject title and code no, if any) :

Nil

Recommended background knowledge:

Familiar with C/C++ or Java

Mutual Exclusions: Internet Information Retrieval (COMP533)

Learning Approach:

42 hours of class activities including - lecture, tutorial, lab, workshop seminar where applicable

Assessment:

Continuous Assessment	55%
Examination	45%

Objectives:

1. To introduce the fundamental knowledge and techniques in information retrieval (IR) and information extraction (IE);
2. To apply the IR and IE fundamentals to various internet applications; and
3. To explore the use of information retrieval technology in advanced IR internet applications, like information filtering;

Learning Outcomes:

After completing this subject, students should be able to:

1. be aware of various classical information retrieval models;
2. comprehend the main difference between classical information retrieval and Internal information retrieval;

The Department reserves the right to update the syllabus contents. Please note that the learning approach for the same subject could vary slightly due to different delivery modes.

3. handle the problems particularly associated to Internal information retrieval;
4. design and implement effective retrieval systems;
5. apply retrieval evaluation techniques to improve retrieval system; and
6. extract important pieces of information from the retrieved text and convert from unstructured text to structured database

---

Keyword Syllabus:

Presentation of Information in the Internet
Tagging and Processing: HTML, XML and SGML

Classical Information Retrieval
Architecture, IR models, Term selection and weighting, Ranking, Query processing, Evaluation techniques, indexing and search engine fundamentals

Information Extraction
Extraction: Keyword identification, NP extraction, String pattern extraction

Distributed Information Retrieval
Web-graph analysis, Server ranking, Meta search engines

Applications
Digital library, Wireless information access

Advanced Information Retrieval
Relevance feedback, Advanced indexing techniques, issues in Multilingual/Multimedia information retrieval, Information filtering and text categorization

---

Indicative reading list and references:

G.G. Chowdhury, 2004, *An Introduction to Modern Information Retrieval*, London, Facet.
R. Baeza-Yates, B. Ribeiro-Neto., 1999, *Modern Information Retrieval*, ACM Press.
D.A. Grossman and O. Frieder., 1998, *Information Retrieval: Algorithms and Heuristics*, Kluwer.
H.I. Witten, A. Moffat and T.C. Bell., 1999, *(2nd edition) Managing gigabytes : compressing and indexing documents and images*, New York : Van Nostrand Reinhold.
W.B. Frakes and R. Baeza-Yates (Eds.), 1992, *Information Retrieval: Data Structures and Algorithms. Prentice-Hall*, Englewood Cliffs, NJ, 241-263.
G. Grefenstette (ed.), 1998, *Cross-language information retrieval*, Dordrecht, The Netherlands: Kluwer Academic Press.

ACM Transactions on Internet Technology
ACM Transactions on Information Systems
IEEE Transactions on Knowledge and Data Engineering
Information Retrieval
ACM Transactions on Database Systems
IEEE Transactions on Software Engineering
Information Processing and Management
Journal of the American Society for Information Sciences & Technology
ACM SIGIR Conference Proceedings
WWW Conference Proceedings