Module 3:   Information Retrieval and Web Search

Stage	1
Semester	1
Module Title	Information Retrieval & Web Search
Module Number	3
Module Status	Mandatory
Module ECTS Credits	10
Module NFQ level	9
Pre-Requisite Module Titles	None
Co-Requisite Module Titles	None
Capstone Module?	No
List of Module Teaching Personnel	Dr Mark Scanlon

Contact Hours				Non-contact Hours			Total Effort (hours)
60				**140**			**200**
Lecture	Practical	Tutorial	Seminar	Assignment	Placement	Independent Work	
36		24		40		100	

Allocation of Marks (Within the Module)					
	Continuous Assessment	Project	Practical	Final Examination	Total
Percentage Contribution	40			60	100

Intended Module Learning Outcomes

On successful completion of this module the learner will be able to:

1. Contrast individual information retrieval algorithms
2. Implement various information retrieval strategies
3. Compare the performance of IR techniques through the use of evaluation metrics
4. Describe the notions of collection and data fusion and implement associated algorithms
5. Critically evaluate the various search strategies employed by commercial search engines
6. Analyse state-of-the-art information retrieval techniques

Module Objectives

This module introduces the learner to the concepts of information retrieval (IR) and web search. They encounter various techniques used in IR and means of evaluating their performance. Learners also gain an exposure to the practical design of large-scale IR systems that are commonly used in the web search domain. Current trends in IR, such as collection and data fusion are introduced through the use of academic papers.

Module Curriculum

- **Information Retrieval**
  What is information retrieval? / The IR process: stemming; stopwords; indexing; retrieval ; evaluation / Retrieval techniques: vector space; TF-IDF; boolean; ext-bool; LSI; fuzzy set.

- **Evaluation**
  Evaluating performance of IR systems: precision; recall; f-score; MAP; b-pref; and interpolated precision / Comparing system performances.

- **Information Fusion**
  Collection Fusion / Data Fusion / CombMNZ / probFuse / slideFuse

- **Web   Search**
  Competing models: keyword; Natural Language Processing; pagerank; Case studies of Google, Yahoo, etc.

Reading Lists and other learning materials

Recommended Reading

Manning C D., Raghavan P., 2013, *Introduction to Information Retrieval,* Cambridge University Press.

Liu B, 2011, *Web Data Mining@ Exploring Hyperlinks, Contents, and Usage Data (Data Centric Systems and Applications),* 2nd Edition, Springer

Buttcher S, Clarke C, Cormack G, 2010, *Information Retrieval: Implementing and Evaluating Search Engines,* MIT Press

Baeza-Yates, R., Ribeiro-Neto, R., *Modern Information Retrieval,* Addison-Wesley, 1999

Van Rijsbergen, K., *Information Retrieval*,* Butterworth-Heinemann, 1997

*available online

Secondary Reading

Russell-Rose T, Tyler T, 2012, *Designing the Search Experience: The Information Architecture of Recovery,* Morgan Kaufmann

Russell M A, 2011, *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn , and Other Social Media Sits,* 1st Edition, O'Reilly Media

Croft B, Metzler D, Strohman T, 2009, *Information Retrieval in Practice,* Pearson Education

Grossman D A, Frieder O, 2004, *Information Retrieval: Algorithms and Heuristics,* Springer

Chakrabarti S, 2002, *Mining the Web: Analysis of Hypertext and Semi Structured Data,* Morgan Kaufmann

Various Authors, *Proceedings of SIGIR Conferences*,* ACM, To Date

Various Authors, *Proceedings of ECIR Conferences*,* Springer, To Date

Various Authors, *The Journal of Information Retrieval*,* Springer, To Date

*available online

Additional reading as recommended by lecturer, appropriate to topic.

Module Learning Environment

Accommodation

Lectures are carried out in class rooms / lecture halls in the College. Computer Labs throughout the Campus are accessible for the purpose of completing assignments. There is no specific software required to deliver the programme.

Library

All learners have access to an extensive range of physical and electronic (remotely accessible) library resources. The library monitors and updates its resources on an on-going basis, in line with the College's Library Acquisition Policy. Lecturers update reading lists for this course on an annual basis as is the norm with all courses run by Griffith College.

Module Teaching and Learning Strategy

Each week, there are three classes:
Classes are used to explain the concepts, exemplify the techniques, and solve (in workshop style) a series of exercises and problems. Some classes involve the

discussion of seminal research papers in the IR domain. Learners are expected to read the material prior to class.

Module Assessment Strategy

Element No.	Weighting	Type	Description	Learning Outcomes Assessed
1.	20%	Assignment	For this assignment learners are required to implement a common model of Information Retrieval. Learners must create a program (using a programming language of their choice) that can process the documents contained in a sample document collection. The data will have to be pre-processed before the IR step.	2,4,5
2.	20%	Assignment	For this assignment, learners are required to implement three standard Information Retrieval evaluation metrics. Learner must create a program (using a programming language of your choice) that is capable of calculating the required metrics from a given document collection and the results for a number of queries from an IR system.	2,3,4,5
3.	60%	Examination	The examination will test the learners understanding of the theoretical aspects of the coursework.	1,3,4,5,6