**Title (Units):** **COMP4105 Web Search Principles and Technology (3,2,1)**

**Course Aims:** To provide a study of the techniques and mechanisms for searching diverse types of information such as text documents, images, video and audio information. The powerful features and functions of popular current search systems will be examined. Emphasis is placed on large information sources and databases on the Internet.

**Prerequisite:** COMP2045 Programming and Problem Solving AND
COMP2046 Problem Solving Using Object Oriented Approach

**Course Intended Learning Outcomes (CILOs):**
Upon successful completion of this course, students should be able to:

| No. | Course Intended Learning Outcomes (CILOs) |
|---|---|
| | **Knowledge** |
| 1 | Describe the characteristics of different forms of textual and multimedia information |
| 2 | Distinguish between the requirements of different types of search tasks |
| 3 | Explain the techniques used for information retrieval |
| 4 | Describe the principles for producing and ranking of Web search results |
| | **Professional Skill** |
| 5 | Make use of the advanced features of popular search systems to locate different types of information |
| 6 | Carry out analysis and ranking of documents and interlinked Web pages |

**Calendar Description:** This course provides a comprehensive examination of different popular search systems for diverse types of data such as text, image, video and audio information. Students will be introduced to the powerful features in these systems, as well as the technology underpinning them. Students will learn how large information repositories are efficiently organized, managed and searched, and the principles of Web search engines and information retrieval.

**Teaching and Learning Activities (TLAs):**

| CILOs | Type of TLA |
|---|---|
| 1 - 4 | Lectures, laboratory classes, exercises and assignments |
| 1, 3 | Lectures, laboratory classes |
| 4 - 6 | Problem and laboratory classes, exercises and assignments, Web information search tasks |

**Assessment:**

| No. | Assessment Methods | Weighting | CILOs to be addressed | Description of Assessment Tasks |
|---|---|---|---|---|
| 1 | Continuous Assessment | 50% | 2, 4, 5, 6 | This includes assessed tasks and assignments, which are designed to assess the students' ability to apply the techniques learned to carry out search tasks and manage information. |
| 2 | Examination | 50% | 1 - 6 | The final examination is designed to measure the extent to which the students have reached all of the course intended learning outcomes. Students are required to have a good mastery of the concepts, techniques, methodologies, and applications of information search techniques to different situations. |

**Assessment Rubrics:**

| | |
|---|---|
| **Excellent (A)** | • Achieves all six CILOs, demonstrating a good mastery of both the theoretical and practical aspects of the knowledge and skills associated with Web search principles and technology<br>• Able to develop and present sound arguments and correct solutions to problems, accompanied by in-depth analysis and insight<br>• Demonstrates a thorough understanding and solid knowledge of the concepts, algorithms, and methodologies associated with Web search principles and technology<br>• Able to draw on a variety of techniques and relevant knowledge and appropriately apply them to new situations and problems |
| **Good (B)** | • Achieves all six CILOs, demonstrating a good understanding of the associated concepts and underlying methodologies<br>• Able to develop solutions to problems, accompanied by adequate explanations<br>• Demonstrates a competent level of knowledge of Web search principles and technology<br>• Ability to make use of appropriate techniques and knowledge and apply them to familiar situations and problems |
| **Satisfactory (C)** | • Achieves most of the six CILOs, demonstrating a basic level of understanding of the associated concepts and underlying methodologies<br>• Able to provide acceptable solutions to problems<br>• Demonstrates an adequate level of knowledge of Web search principles and technology<br>• Ability to make use of some techniques and knowledge and apply them to familiar situations |
| **Marginal Pass (D)** | • Achieves most of the six CILOs, with minimal understanding of the associated concepts and underlying methodologies<br>• Able to provide solutions to simple problems<br>• Demonstrates a basic level of knowledge of Web search principles and technology<br>• Ability to apply some techniques and knowledge to a limited number of typical situations |
| **Fail (F)** | • Achieves less than three of the six CILOs, with little understanding of the associated concepts and underlying methodologies<br>• Unable to provide solutions to simple problems<br>• Knowledge of Web search principles falling below the basic minimum level<br>• Unable to apply techniques and knowledge to situations or problems |

**Course Content and CILOs Mapping:**

| **Content** | | **CILO No.** |
|---|---|---|
| I | Text Operation and Retrieval | 1, 3 |
| II | Web Search Engines | 2, 4, 5, 6 |
| III | Audio, Video and Image Search | 1, 2, 4, 5 |
| IV | Case Studies and Applications | 1 - 6 |

**References:**


- M. Levene, An Introduction to Search Engines and Web Navigation, 2nd Edition, Wiley, 2010.
- I. Witten, M. Gori, and T. Numerico, Web Dragons: Inside the Myths of Search Engine Technology, Morgan Kaufmann, 2007
- B. Croft, D. Metzler, and T. Strohman, Search Engines: Information Retrieval in Practice, 1st Edition, Pearson, 2009
- P. Muneesawang, N. Zhang, and L. Guan, Multimedia Database Retrieval: Technology and Applications, Springer, 2014
- M. Maybury, Multimedia Information Extraction: Advances in Video, Audio, and Imagery Analysis for Search, Data Mining, Surveillance and Authoring, 1st Edition, Wiley, 2012


**Course Content:**

**Topic**

I. Text Operation and Retrieval
   A. Taxonomy of information retrieval models
   B. Classical and probabilistic models
   C. Evaluation measures
   D. Document clustering
   E. Document indexing

II. Web Search Engines
   A. Search engine architecture, features and advanced usage
   B. PageRank algorithm and Google search
   C. Browsing and relevance feedback
   D. Semantic web

III. Audio, Video and Image Search
   A. Multimedia information and metadata properties
   B. Query by example
   C. Content-based and tag-based search
   D. Web image and video search
   E. Query by humming and MIDI search

IV. Case Studies and Applications
   Popular applications such as Google Earth, YouTube, Book Search, Blog Search, Patent Search, and Google Directory will be examined.