

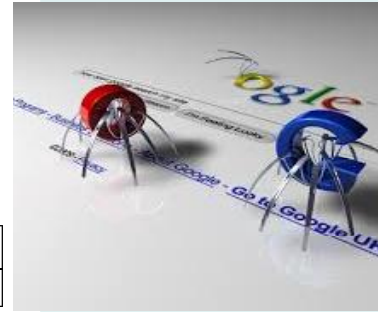
16CS401 SEARCH ENGINES

Hours Per Week :

L	T	P	C
3	-	-	3

Total Hours :

L	T	P	WA/RA	SSH/SHS	CS	SA	S	BS
45	-	-	5	40	5	8	5	5



Course Description and Objectives:

This course deals with the technologies/various methods used in finding relevant information on the web. The objective of this course is to understand the existing techniques of search engine and use the mathematical models to build a fair search engine.

Course Outcomes:

The student will be able to:

- demonstrate the effective use of a search engine to locate a website or web address.
- follow basic safe computing steps when using a search engine.
- understand how a search engine works.
- understand cyber threats caused by accessing unwanted websites and measures of protection.
- use information on safe computing presented in this lesson and previous lessons in making decisions about websites they come across.

SKILLS:

- ü Given a set of documents, find the term frequency & document frequency.
- ü Create an inverted index for a set of documents.
- ü Design and implement a secondary index for a set of documents.
- ü Build a Ranking model on a set of documents to retrieve relevant results.

ACTIVITIES:

- *Removal of stop words in a document.*
- *Create an index for the document.*
- *Understand the importance of building a secondary index.*
- *Analyse the attributes which contribute for Ranking the documents.*

UNIT - 1**L-09**

SEARCH ENGINE: Search Engine, Search Engineers, Architecture of a Search Engine - What is an Architecture, Basic Building Blocks, Breaking It, How Does It Really Work.

UNIT - 2**L-09**

CRAWLS AND FEEDS: Crawling - Document, Storing the Documents, Detecting duplicates, Removing noise; Processing Text - from Words to Terms; Text Statistics, Document parsing, Document Structure and Markup, Link Analysis, Information Extraction, Internationalization.

UNIT - 3**L-09**

INDEXES: Index and its Significance, Compression, Inverted index, Primary/Secondary/Ternary index construction.

UNIT - 4**L-09**

QUERIES AND INTERFACES: Information Needs, Query Transformation and Refinement, Showing the Retrieval Models - Overview of Retrieval, Probabilistic Models, Ranking based on Language, Application based Models.

UNIT - 5**L-09**

SEARCH ENGINE EVALUATION: Why Evaluate, The Evaluation, Logging, Effectiveness Metrics, Efficiency, Training, Testing and Statistics.

TEXT BOOK:

1. Bruce Croft Donald Metzler and Trevor Strohman, "Search Engines: Information Retrieval in Practice: International Edition", 1st edition, Pearson, 2009.

REFERENCE BOOKS:

1. Mark Levene, "An Introduction to Search Engines and Web Navigation", 2nd edition, WILEY, 2010.
2. Michael W. Berry and Murray Browne, "Understanding Search Engines", 2nd edition, SIAM, 2005.