

Master of science in Theory and Technology of Communication a.a. 2017-18

year	code	course name	ECTS	type	semester	educational activity type	ECTS	hours	faculty
2	F9201P031	INFORMATION RETRIEVAL	6	optional	first semester	lecture	5	40	Pasi Gabriella
						laboratory	1	12	Pasi Gabriella

CV: <http://www.unimib.it/go/176181440>

Contents This course aims at introducing the basic concepts, the formal models and the main techniques to define and design Information Retrieval Systems (also called Search Engines, and in particular Web Search Engines when working on the Web to the aim of retrieving Web pages) and Information Filtering (IF) systems. In particular, various techniques for the analysis and the indexing of texts will be presented, also including a basic introduction to multimedia documents indexing. Moreover, the issue of estimating the relevance of documents to a query will be addressed: several models finalised at the assessment of a numeric estimate of relevance (degree or probability) of a document to a query will be explained. Among the advanced IR topics, the language models will be presented, as well as the main approaches to personalised search. The course will also introduce additional applications related to text analysis and mining, such as the crawling and analysis of user generated content on Social Media (e.g. Twitter, Facebook, etc.). The important issue of the evaluation of the credibility of the content generated by users in social media will be also presented.

Textbooks Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
John Scott, Social Network Analysis (Third Ed.), SAGE, 2013.

Course objectives The objective of this course is to provide an introduction to the main concepts, to the formal models, and to the techniques at the basis of the definition of systems that allow to locate information relevant to specific users' needs. Such systems include Information Retrieval Systems (better known as Search Engines; in particular Web Search Engines when the considered documents are Web pages), and Information Filtering Systems, i.e. systems for the recommendation of information/services. In this context, the main problem is the assessment of the relevance of documents with respect to the information needs formulated in a user's query. The students will acquire the capability of understanding and defining algorithms for documents indexing and retrieval, and to use open source software to implement ad hoc search engines. They will also develop a search engine application by using open source software.

Prerequisites Basic knowledge of statistics and of linear algebra.

Teaching methods The course will be composed of lectures introducing the main topics as well as of sessions in a laboratory where exercises will be undertaken and where the usage of an open source software for the implementation of search engines will be explained and experienced.

Learning assessment Written and oral examination, definition of a laboratory project

Extended Syllabus

1 "Introduction to Information Retrieval (IR) and Information Filtering (IF).

- Documents and information needs, and their representation.
- The notion of relevance.
- Efficiency and effectiveness. Evaluation of the effectiveness of an IRS and of an IFS.

2 Introduction to an open source software platform for the development of search engines.

3 Text Indexing techniques. Basic concepts of multimedia indexing techniques.

4 Information Retrieval models: basic models (Boolean model, Vector Space model, probabilistic models). Introduction to multimedia information retrieval.

5 Web Search Engines: crawling, link analysis and other factors for estimating relevance of Web pages.

6 The evaluation of Search Engines.

7 "Advanced Information Retrieval topics:

- Personalized Search
- Analysis of the content generated by users on the Web
- The assessment of the "credibility" of user's generated content.