COURSE OUTLINE

1. GENERAL

SCHOOL	SCHOOL OF ENGINEERING, UNIVERSITY OF PATRAS		
DEPARTMENT	DEPARTMENT OF COMPUTER ENGINEERING AND INFORMATICS		
LEVEL OF COURSE	UNDERGRADUATE, CORE ELECTIVE		
COURSE CODE	CEID_NE5597	**SEMESTER OF STUDIES**	WINTER
COURSE TITLE	INFORMATION RETRIEVAL		

INDEPENDENT TEACHING ACTIVITIES *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for thewhole of the course, give the weekly teaching hours and the total credits*	**TEACHING HOURS PER WEEK**	**ECTS CREDITS**
Lectures	2	2
Recitation sections	1	1
Laboratory Exercises	2	2
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).		

COURSE TYPE *general background, special background, specialised general knowledge, skills development*	Skill develoment
PREREQUISITE COURSES:	There are no prerequisites. Recommended prerequisite knowledge are the courses of data structures (CEID_23Y210), algorithms (CEID_23Y205), and data base design (CEID_24Y344)
TEACHING AND ASSESSMENT LANGUAGE:	Greek. Instruction may be given in English if foreign students attend the course.
THE COURSE IS OFFERED TO ERASMUS STUDENTS	YES
COURSE WEBPAGE (URL)	https://eclass.upatras.gr/courses/CEID1037/

2. LEARNING OUTCOMES

Leraning outcomes

The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.

Consult Appendix A

- *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area*
- *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B*
- *Guidelines for writing Learning Outcomes*

At the end of this course, the student should be able to:

1. explain basic information storage and retrieval concepts and understand the issues that are specific to the design of efficient information retrieval systems.
2. explain the differences between alternative information retrieval models, and analyze why a particular model is appropriate for a specific application.
3. understand the basic design issues that are involved in the development of an efficient web search engine.
4. design and implement a small to medium size information storage and retrieval system.
5. understand the concept of word embeddings and its use in various applications

At the end of the course the student will have further developed the following skills/competences:

1. ability to understand the various concepts of information retrieval systems, and how these are interrelatedwith the performance to a working system

2. ability to apply methodologically these concepts in order to design and implement efficient working systems
3. ability to work cooperatively in order to solve problems that arise during the construction of a full-fledged working software system.

General Abilities

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology	*Project planning and management*
Adapting to new situations	*Respect for difference and multiculturalism*
Decision-making	*Respect for the natural environment*
Working independently	*Showing social, professional and ethical responsibility and sensitivity to gender issues*
Team work	*Criticism and self-criticism*
Working in an international environment	*Production of free, creative and inductive thinking*
Working in an interdisciplinary environment	.....
Production of new research ideas	*Others…*

Search for, analysis and synthesis of data and information, with the use of the necessary technology
Decision-making
Working independently
Team work
Production of new research ideas
Project planning and management
Production of free, creative and inductive thinking

3. COURSE CONTENT

1. Introductory notions (user modeling, document logical representation, retrieval process).
2. Performance evaluation metrics (recall, precision, average precision, R-precision, precision histograms, NDCG metric, harmonic median, user oriented metrics).
3. Information retrieval modeling.
4. Set-oriented models (boolean models, fuzzy set model, extended boolean model), algebraic models (vector space models, latent semantic indexing model, topic models), probabilistic models (classical and language models).
5. Web information retrieval and its peculiarities.
6. Web search engines (crawler, indexer). HITS algorithm (Hyperlink-induced topic search). Google search engine (the PageRank metric). The SALSA algorithm, variants in web searching
7. Machine Learning Techniques and Neural Models in Information Retrieval (Learning to Rank, vector representation of words and word embeddings such as word2vec, CBOW, skipgram, Transformers, BERT, GPT, Large Language Models and sparse vs dense search, Vector Search as in e.g. FAISS (HNSW etc.), using dense and/or sparse search in Retrieval Augmented Generation (RAG), Search Engines against Reasoning Engines).
8. Storage Techniques in Distributed Information Retrieval (MapReduce, Apache Spark)
9. Indexing structures (inverted files, signature files, bitmaps).
10. Full indexing structures in main memory (suffix trees, suffix arrays, acyclic directed graphs (DAWG) for strings), and in secondary memory (supra-suffix array, prefix B-tree, string B-tree).
11. Compression algorithms for text and for indexing structures.
12. Text Mining and Graph based Models (graph embeddings).

4. TEACHING AND LEARNING METHODS - ASSESSMENT

TEACHING METHOD *Face-to-face, Distance learning, etc.*	Face-to-face
USE OF INFORMATION AND COMMUNICATION TECHNOLOGIES *Use of ICT in teaching, laboratory education, communication with students*	We use Information and Communications Technology in communication with students. We use e_class, e_mail.
TEACHING ORGANIZATION *The manner and methods of teaching are described in detail. Lectures, seminars, laboratory practice,*	

Δραστηριότητα	*Φόρτος Εργασίας Εξαμήνου*
Lectures	60
Recitation sections	30

fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc. *The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS*	Laboratory exercises: 60 ***Total number of hours for the Course***: 150
STUDENT ASSESSEMNT *Description of the evaluation procedure* *Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open- ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other* *Specifically-defined evaluation criteria are given, and if and where they are accessible to students.*	• composing and presenting an essay by groups of 1-2 students, concerning the critical presentation, implementation and analysis of algorithms in a set of scientific papers that deals with a specific topic of the course; as an alternative the students can be involved in the in depth implementation of working parts of a retrieval ranking system (45% of the final mark, taken into account only when the student secures a minimum mark in the final written examination), • written examination (55% of the final mark).

5. RECOMMENDED LITERATURE

1. Christofer D. Manning, Prabhakar Raghavan, Hinrich Schutze Introduction to Information Retrieval, Cambridge University Press, 2008.
2. R. Baeza-Yates, B. Ribeiro-Neto, Modern Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition) · January 2011 with 973 Reads ISBN 978-0-321-41691-9 Publisher: Pearson Education Ltd., Harlow, England
3. Omar Alonso and Ricardo Baeza-Yates, Information Retrieval: Advanced Topics and Techniques, December 2024, Association for Computing Machinery, New York, United States, ISBN:979-8-4007-1050-6
4. Jimmy Lin, Rodrigo Nogueira, and Andrew Yates, Pretrained Transformers for Text Ranking: BERT and Beyond by University of Waterloo, University of Campinas, University of Amsterdam) Morgan & Claypool (Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, volume 53), 2021.
5. Dan Jurafsky and James H. Martin Speech and Language Processing (3rd ed. draft) Jan 2, 2025 draft (https://web.stanford.edu/~jurafsky/slp3 /)
6. Mohammad Taher Pilehvar , Jose Camacho-Collados, Embeddings in Natural Language Processing Theory and Advances in Vector Representations of Meaning, Morgan Claypool, 2020 (https://sites.google.com/view/embeddings-in-nlp , https://sites.google.com/view/embeddings-in-nlp/tutorial)
7. Stefan Büttcher Charles L. A. Clarke, Gordon V. Cormack, Information Retrieval: Implementing and Evaluating Search Engines (MIT Press) Paperback – February 12, 2016
8. Amy N. Langville, Carl D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings Princeton University Press Princeton, NJ, USA 2012
9. I. Witten, A. Moffat, T. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images, Morgan Kaufmann Publishers, 1999.