Syllabus

Descrizione corso

Titolo insegnamento	Large Language Models and Information Retrieval
Codice insegnamento	73072
Titolo aggiuntivo	
Settore Scientifico-Disciplinare	INFO-01/A
Lingua	Inglese
Corso di Studio	Corso di laurea magistrale in Informatica per la Data Science
Altri Corsi di Studio (mutuati)	
Docenti	dr. Andrea Rosani, Andrea.Rosani@unibz.it https://www.unibz.it/en/faculties/engineering/academic-staff/person/43727
Assistente	
Semestre	Primo semestre
Anno/i di corso	2
CFU	6
Ore didattica frontale	40
Ore di laboratorio	20
Ore di studio individuale	90
Ore di ricevimento previste	18
Sintesi contenuti	- Web and mobile search - Boolean and vector-space retrieval models - Efficient document indexing, document mining and topic modelling - Traditional and machine learning-based ranking approaches - Large Language Models for Information Retrieval - Evaluation of Information Retrieval Systems
Argomenti dell'insegnamento	This course provides a comprehensive introduction to the principles and techniques of Information Retrieval (IR), focusing on both

	traditional methods and modern advancements. - Web and Mobile Search: Techniques for indexing, ranking, and retrieving information in large-scale web environments and mobile contexts, including challenges like personalization, context-awareness, and interface constraints. - Boolean and Vector-Space Models: Fundamental retrieval models including Boolean logic and vector space approaches, forming the basis for understanding document representation and relevance scoring. - Efficient Indexing, Document Mining, and Topic Modelling: Methods for building scalable indexing structures, mining valuable information from text corpora, and uncovering latent topics using models like LDA. - Ranking Algorithms: Traditional and Machine Learning-Based: Examination of classic ranking methods (e.g., BM25, PageRank) alongside machine learning-based techniques, including learning-to-rank and neural models for improved relevance and user satisfaction. - Large Language Models for Information Retrieval: Application of large pre-trained language models (e.g., BERT, GPT) in IR tasks such as semantic search, question answering, and conversational retrieval. - Evaluation of IR Systems: Approaches to measuring the effectiveness of retrieval systems using metrics like precision, recall, MAP, and nDCG, as well as methods for conducting user-centered evaluations.
Parole chiave	Indexing and Ranking Machine Learning for IR LLMs Evaluation Metrics
Prerequisiti	Programming and algorithm data structures skills, Linear algebra, probability theory, basic machine learning concepts.
Insegnamenti propedeutici	

Modalità di insegnamento	Frontal lectures, exercices, lab.
Obbligo di frequenza	The attendance is not compulsory, but students are highly encouraged to attend.
Obiettivi formativi specifici e risultati di apprendimento attesi	The course belongs to the type "caratterizzanti – discipline informatiche". The objective of this course is to present the scientific underpinnings of the field of Information Retrieval (IR). The student will study fundamental, mathematically sophisticated IR concepts first and then more advanced techniques for information filtering and decision support, including transformer-based solutions and LLMs. This course provides students with a rich and comprehensive catalogue of information search and text processing techniques that can be exploited for the design and implementation of modern IR applications. Knowledge and understanding: - D1.4 - Basic knowledge of storing, querying and managing large amounts of data and the associated languages, tools and systems Applying knowledge and understanding: - D2.2 - Ability to address and solve a problem using scientific methods Making judgments - D3.2 - Ability to autonomously select the documentation (in the form of books, web, magazines, etc.) needed to keep up to date in a given sector Communication skills - D4.1 - Ability to use English at an advanced level with particular reference to disciplinary terminology.
Obiettivi formativi specifici e risultati di apprendimento attesi (ulteriori info.)	
Modalità di esame	Final Project with report + oral exam The project will cover the learning outcome D2.2 and D3.2. It will consist of the design of an IR system in a specific application domain selected by the students.

	The project domain, the attacked problem, the techniques, and the obtained results must be described in a report (max. 10 pages). The project report will cover the learning outcome D1.4 and can be done in groups of 2-3 people. The oral exam will cover the learning outcome D4.1. It is composed by the discussion of the project and some individual questions on the content of the project itself.
Criteri di valutazione	Evaluation criteria - Project: 50% of the mark - Report: 30% of the mark - Final oral exam: 20% of the mark. Important note: both project and exam are required to be passed. Criteria for awarding marks Project: ability to implement data workflow to apply IR to real-world problems, correctness and clarity of the solution, experimental results, ability to solve IR problems with the appropriate technique. Report: ability to describe the proposed solution, with a critical approach describing the methodology and the results. Oral exam: ability to present and explain information retrieval concepts, methods and algorithms. ability to select appropriate solutions for IR problems.
Bibliografia obbligatoria	The suggested book for the introduction to information retrieval topics is: • C. D. Manning, P. Raghavan and H. Schutze. Introduction to Information Retrieval, Cambridge University Press, 2008. (Online: http://informationretrieval.org) Papers about the most recent advancements with regards to algorithms, information access modalities and interfaces will be

	provided during the course in electronic format. Copy of the slides will be available as well. Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it
Bibliografia facoltativa	
Altre informazioni	Software used: Python as programming language
Obiettivi di Sviluppo Sostenibile (SDGs)	Istruzione di qualità