Syllabus

Descrizione corso

Titolo insegnamento	Information retrieval
Codice insegnamento	76057
Titolo aggiuntivo	
Settore Scientifico-Disciplinare	INFO-01/A
Lingua	Inglese
Corso di Studio	Corso di laurea magistrale in Ingegneria del Software
Altri Corsi di Studio (mutuati)	LM-18 Computing for Data Science
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	2025
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 • Foundation models • 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. • Foundation Models: 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	Information Retrieval, document indexing, foundation models.
Prerequisiti	Programming and algorithm data structures skills, Linear algebra, probability theory, basic machine learning concepts.
Insegnamenti propedeutici	
Modalità di insegnamento	Frontal lectures, exercises, lab, seminars.
Obbligo di frequenza	The attendance is not compulsory, but students are highly encouraged to attend.
Obiettivi formativi specifici e risultati di apprendimento attesi	Knowledge and understanding D1.4 have an in-depth knowledge of the principles, structures and use of processing systems for the automation of software systems; Applying knowledge and understanding D2.2 know how to design and carry out experimental analyses of

	software systems in order to acquire measurements of their behaviour and evaluate experimental hypotheses in different application fields, such as business, industry or research; Making judgements D3.1 ability to independently select documentation from various sources, including technical books, digital libraries, technical scientific journals, web portals or open source software and hardware tools; D3.5 be able to work with broad autonomy, including taking responsibility for projects and structures. Communication skills D4.4 ability to prepare and deliver presentations with technical content in English;
Obiettivi formativi specifici e risultati di apprendimento attesi (ulteriori info.)	The course belongs to the type "caratterizzanti – discipline informatiche" in the study path without curriculum". 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.
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	Gerhard. author Paaß, Foundation Models for Natural Language Processing Pre-trained Language Models Integrating Media , 1st ed. 4/4 2023. Cham: Springer International Publishing, 2023. doi: 10.1007/978-3-031-23190-2.
Altre informazioni	
Obiettivi di Sviluppo Sostenibile (SDGs)	