



# UNIVERSITÀ DI PISA

## INFORMATION RETRIEVAL

### PAOLO FERRAGINA

Academic year	2022/23
Course	INFORMATICA
Code	289AA
Credits	6

Modules	Area	Type	Hours	Teacher(s)
INFORMATION RETRIEVAL	INF/01	LEZIONI	48	PAOLO FERRAGINA

#### Obiettivi di apprendimento

##### **Conoscenze**

Lo studente che completerà con successo il percorso didattico avrà l'abilità e le conoscenze per progettare un semplice motore di ricerca e/o uno strumento di IR, utile nelle applicazioni moderne per il Web o i Big Data.

##### **Modalità di verifica delle conoscenze**

Gli studenti verranno valutati in base alle loro abilità e conoscenze acquisite sui contenuti del corso, attraverso una prova scritto/orale, accompagnata da eventuali prove intermedie.

Ulteriori informazioni possono essere trovate nella [home page](#) del corso.

##### **Capacità**

Gli studenti alla fine del percorso didattico saranno in grado di progettare e valutare motori di ricerca e IR tool, adottando soluzioni algoritmiche allo stato dell'arte.

##### **Modalità di verifica delle capacità**

Esame scritto e orale, con la possibilità di svolgere prove intermedie

##### **Comportamenti**

Gli studenti saranno esposti durante il corso al contesto dell'IR su Big Data testuali e non, le loro sfide, le scelte algoritmiche, e apprezzeranno l'impatto che le soluzioni efficienti in tempo e spazio viste in classe hanno sull'analisi, la ricerca e il mining di Big Data nelle applicazioni pratiche.

##### **Modalità di verifica dei comportamenti**

Esame scritto e orale, con la possibilità di svolgere prove intermedie

##### **Prerequisiti (conoscenze iniziali)**

Esami sul progetto di algoritmi, e conoscenze di programmazione e Math.

##### **Indicazioni metodologiche**

Lezioni di didattica frontale, e attraverso lo studio individuale.

##### **Programma (contenuti dell'insegnamento)**

Studio, progetto e analisi di sistemi di Information Retrieval che risultano efficienti ed efficaci nel processare, analizzare, memorizzare, ricercare, classificare e raggruppare documenti testuali e non. Le lezioni descriveranno

- le principali componenti di un motore di ricerca moderno: Crawler, Parser, Compressor, Indexer, Query resolver, Query and Document annotator, Results Ranker;
- le tecniche algoritmiche che sono alla base del progetto di applicazioni di IR per la compressione, l'indicizzazione e lo sketching di documenti;
- alcuni IR tool che sono utilizzati come componenti di motori di ricerca sofisticati oppure come tool indipendenti in applicazioni per processano Big Data, quali: Classification, Clustering, Recommendation, Random Sampling, Locality Sensitive Hashing.



# UNIVERSITÀ DI PISA

## Bibliografia e materiale didattico

C.D. Manning, P. Raghavan, H. Schutze. Introduction to Information Retrieval. Cambridge University Press, 2008  
Chapter 2 "Text compression" of Managing Gigabytes, I.H. Witten and A. Moffat and T.C. Bell, Morgan Kauffman, Second edition, 1999.  
Alcune note in Inglese scritte dal docente

## Indicazioni per non frequentanti

Come per i frequentanti

## Modalità d'esame

Esame scritto e orale, con la possibilità di svolgere prove intermedie

## Pagina web del corso

<http://didawiki.di.unipi.it/doku.php/magistraleinformatica/ir/start>

*Ultimo aggiornamento 29/07/2022 18:56*