



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## COURSE SYLLABUS

### Information Retrieval

2122-2-F1801Q110

---

#### Obiettivi

L'obiettivo del corso è fornire un'introduzione ai concetti fondamentali, ai modelli formali, e alle tecniche per la realizzazione di sistemi per il reperimento automatico di documenti in forma digitale (sistemi di "Information Retrieval", detti Motori di Ricerca o Motori di Ricerca su Web quando i documenti da reperire sono costituiti da pagine Web) e per la raccomandazione di informazioni (Recommender Systems). In questo contesto il principale problema da affrontare è quello della valutazione della rilevanza dei documenti rispetto alle necessità informative dell'utente. Al termine del corso lo studente sarà in grado di progettare tecniche per il reperimento, il trattamento e l'indicizzazione di testi semi-strutturati, e di utilizzare software "open source" per la definizione di applicazioni di Information Retrieval. Il laboratorio sarà finalizzato alla realizzazione di un'applicazione.

#### Contenuti sintetici

Il corso introdurrà un insieme di tecniche per la progettazione e la realizzazione di motori di ricerca, e per la definizione di sistemi per la raccomandazione di informazioni (Information Filtering).

In particolare saranno presentate tecniche per il trattamento, l'analisi e l'indicizzazione di testi, con accenni all'indicizzazione di documenti multimediali; saranno inoltre presentati alcuni modelli per la determinazione della stima (grado, o probabilità) di rilevanza di un documento rispetto alle necessità informative dell'utente. Verranno inoltre presentate alcune tecniche per la personalizzazione della ricerca.

Il corso introdurrà il problema della collezione, l'analisi e reperimento di contenuto generato dagli utenti sui Social Media (ad esempio Twitter, Facebook, ecc.).

## **Programma esteso**

1. Definizione di Text Mining e delle principali differenze tra Text Mining e Data Mining.
2. Introduzione di alcune applicazioni correlate al Text Mining
3. Pre-Processing, indicizzazione e rappresentazione formale di testi
- 4 Modelli di sistemi di Information Retrieval: i modelli base (Booleano, Vettoriale, modelli Probabilistici). Modelli avanzati (ad esempio modelli neurali). Accenni a motori di ricerca per documenti multimediali.
- 5 I motori di ricerca su Web: crawling, link analysis e altri fattori per la stima della rilevanza di pagine Web.
- 6 La valutazione dei motori di ricerca.
- 7 Argomenti avanzati:
  - Personalizzazione della ricerca.
  - Social Search
  - Domain Specific Search
- 8 Introduzione a software open source per la definizione di motori di ricerca

## **Prerequisiti**

Nozioni di base di Statistica e di Algebra Lineare.

## **Modalità didattica**

Il corso verrà tenuto in lingua inglese e prevede lezioni ed esercitazioni in laboratorio. Si organizzeranno seminari tenuti da esperti a livello internazionale.

## **Materiale didattico**

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.

## **Periodo di erogazione dell'insegnamento**

Primo semestre

## **Modalità di verifica del profitto e valutazione**

Prova scritta individuale costituita da esercizi e domande aperte sui contenuti del corso; esame orale che verte su una discussione dello scritto e su eventuali domande inerenti il contenuto del corso. Realizzazione di un progetto di laboratorio che è possibile svolgere in gruppo (sino a tre studenti).

La prova scritta ha come obiettivo la valutazione del livello di comprensione degli aspetti teorici e tecnici di base dell'insegnamento erogato.

L'obiettivo del progetto di gruppo, attraverso l'utilizzo di software open source, è lo sviluppo di soluzioni tecnologiche a problemi affrontati a lezione. In particolare, si considerano ambiti applicativi reali che necessitano della definizione di sistemi i cui fondamenti sono stati presentati a lezione.

## **Orario di ricevimento**

Previo appuntamento con la docente.

---