



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Text Mining and Search

2223-2-F9101Q015

---

#### Obiettivi

L'obiettivo del corso è fornire un'introduzione ai concetti fondamentali relativi a tecniche di Text Mining e rappresentazione di testi; si presenteranno inoltre alcune applicazioni di Text Mining: Classificazione e clustering di Testi, Topic Modeling, Riassunto Automatico di Testi. Si faranno accenni a sistemi quali i Motori di Ricerca e i Sistemi per la Raccomandazione di Informazioni.

#### Contenuti sintetici

Il corso fornirà inizialmente la definizione di Text Mining e indicherà le principali differenze tra Data Mining e Text Mining.

Il corso introdurrà quindi alcune applicazioni correlate al Text Mining: riassunto automatico di testi, e classificazione di testi. Tecniche di pre-processing di testi verranno presentate e il problema dell'indicizzazione di testi e della loro rappresentazione formale verrà affrontato. Il corso introdurrà quindi le applicazioni precedentemente citate. Si introdurranno alcuni software open source per la definizione di applicazioni di Text Mining.

#### Programma esteso

1. Definizione di Text Mining e delle principali differenze tra Text Mining e Data Mining
2. Breve introduzione di alcune applicazioni correlate al Text Mining
3. Pre-Processing, indicizzazione e rappresentazione formale di testi
4. Classificazione e clustering di testi
5. Topic Modelling
6. Riassunto automatico di testi

7. Introduzione ai Motori di ricerca testuali e ai Sistemi per la Raccomandazione di Informazione
8. Software Open Source

## **Prerequisiti**

Conoscenza di base di statistica e di linguaggi di programmazione.

## **Modalità didattica**

Il corso è tenuto in lingua inglese e prevede sia lezioni sia esercitazioni; in laboratorio sarà spiegato e sperimentato l'utilizzo di software "open source".

Sono previsti seminari tenuti da esperti a livello nazionale ed internazionale.

## **Materiale didattico**

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.

Testi specifici su Text Mining accessibili online verranno indicati durante il corso.

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

Primo semestre

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

Prova scritta e realizzazione di un progetto di laboratorio (project work) che è possibile svolgere in gruppo (sino a tre studenti)

La prova scritta ha come obiettivo la valutazione del livello di comprensione degli aspetti di base dell'insegnamento erogato, ed è costituita da un insieme di domande a risposta aperta.

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 i docenti

## Sustainable Development Goals

---