UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Information Retrieval

2425-2-F9201P031

Obiettivi

Il corso è una mutuazione di "2425-2-FDS01Q013" - fate riferimento a quel corso onde evitare problemi di sincronizzazione dei contenuti

L'obiettivo del corso è quello di fornire un'introduzione ai concetti fondamentali relativi alle tecniche di rappresentazione del testo e di estrazione del testo; inoltre, nel corso verranno presentate alcune applicazioni di Text Mining: Text Classification and Clustering, Topic Modelling, e Text Summarization. Verrà fornita un'introduzione ai motori di ricerca e ai sistemi di raccomandazione.

Contenuti sintetici

Questo corso fornirà innanzitutto la definizione di Text Mining e indicherà le differenze fondamentali tra Data Mining e Text Mining.

Il corso introdurrà quindi alcune attività coinvolte nel Text Mining, che includono il riepilogo del testo e la classificazione del testo. Verranno affrontati i problemi della pre-elaborazione e analisi del testo e dell'indicizzazione e rappresentazione del testo. Quindi il corso introdurrà i compiti precedentemente menzionati. Verranno introdotti e praticati alcuni software open source per il Text Mining.

Programma esteso

1. Definizione di Text Mining e differenze fondamentali tra Data Mining e Text Mining
2. Introduzione ad alcune attività relative al Text Mining
3. Pre-elaborazione del testo, indicizzazione e rappresentazione formale

4. Classificazione e clustering del testo
5. Modellazione dell'argomento
6. Riepilogo del testo
7. Introduzione ai motori di ricerca testuali e ai sistemi di raccomandazione
8. Software Open Source per il Text Mining e la Ricerca

Prerequisiti

Conoscenza di base della statistica e dei linguaggi di programmazione.

Modalità didattica

Il corso sarà tenuto in lingua inglese, e sarà costituito sia da lezioni frontali introduttive agli argomenti principali, sia da sessioni in laboratorio in cui verranno spiegati e utilizzati gli strumenti open source.
Faranno parte del corso seminari tenuti da esperti di livello nazionale e internazionale.

Materiale didattico

Christopher D. Manning, Prabhakar Raghavan e Hinrich Schütze, Introduzione al recupero delle informazioni, Cambridge University Press. 2008.

Durante il corso saranno consigliati libri specifici sul text mining accessibili online.

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

Esame scritto individuale, definizione di un progetto di laboratorio (project work) che può essere sviluppato anche da gruppi di studenti (fino a tre studenti).
La prova scritta è finalizzata ad accertare il livello di comprensione degli aspetti di base insegnati durante il corso; è costituito da un insieme di domande aperte.
L'obiettivo del progetto di gruppo è l'utilizzo di software open source che verranno impiegati per sviluppare soluzioni tecnologiche ai problemi affrontati nel corso. In particolare verranno presi in considerazione reali ambiti applicativi, che richiedono la definizione di sistemi presentati durante il corso.

Orario di ricevimento

To be agreed with the teachers

Sustainable Development Goals