UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Text Mining and Search

2324-2-FDS01Q013

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.
- 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 alcune applicazioni correlate al Text Mining: classificazione, clustering, *Topic Modeling* e riassunto automatico di testi.
- 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 (*BoW*, *Word Embedding*, Introduzione a tecniche di *Contextual Word Embedding*)
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 informazioni
8. Software "open-source" per Text Mining e ricerca di informazioni online

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”.
- Potranno essere previsti seminari tenuti da esperti a livello nazionale ed internazionale.

Materiale didattico

- Berry, M. W., & Kogan, J. (Eds.). (2010). Text mining: applications and theory. John Wiley & Sons.
- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
- Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. Fundamentals of artificial intelligence, 603-649.

Altri testi specifici su Text Mining accesibili 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, e****d è costituita da un insieme di domande a risposta aperta. L'obiettivo del progetto di gruppo, attraverso l'utilizzo di sofrware "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. La **valutazione** dell'esame scritto avverrà in trentesimi. A tale valutazione verranno aggiunti da 0 a 4 punti ottenibili dalla valutazione del progetto.

Orario di ricevimento

Previo appuntamento con i docenti

Sustainable Development Goals