UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Social Media Analytics

2122-2-F9101Q021

Obiettivi

Alla fine del corso, lo studente avrà appreso i principali concetti che stanno alla base della gestione dei dati originati nei media sociali (accesso, pre-processing, modellazione) e della loro successiva analisi. Lo studente sarà in grado, in particolare, di collezionare, processare e analizzare dati provenienti dai principali media sociali, utilizzando le tecnologie più adatte allo scopo. Sarà inoltre in grado di fornire una rappresentazione di strutture sociali complesse al fine di estrarne informazione utile.

Contenuti sintetici

1. Introduzione al Web sociale e alla terminologia utilizzata;
2. Il reperimento dei dati nei media sociali e problematiche connesse;
3. La rappresentazione dei dati sociali mediante strutture sociali complesse (teoria dei grafi);
4. Descrizione di problematiche di interesse (diffusione dell'informazione, credibilità);
5. Analisi delle reti sociali;
6. Analisi della soggettività, del "sentiment" e dell'ironia;
7. Named-entity Recognition e Linking;
8. Visualizzazione di dati provenienti da social media: problemi aperti e tecniche.

Programma esteso

PRIMA PARTE

1. Introduzione

- Introduzione al Web sociale e alla terminologia utilizzata: Web sociale, reti sociali, media sociali, contenuto generato dagli utenti.

- Gli “oggetti informativi” sul Web sociale: 1) testi: post, blog, microblog, 2) immagini, 3) audio, 4) video.
- *Social Media Analytics*: definizione e obiettivi, i concetti di auto-presentazione e auto-rivelazione.

2. Il reperimento dei dati nei media sociali

- Le principali piattaforme, le interfacce di programmazione, il processo di crawling.
- *Pre-processing* e memorizzazione di dati sociali.
- Cenni di problematiche legali legate alla raccolta dei dati (il GDPR).

3. Il problema della rappresentazione dei dati sociali

- Strutture dati elementari e complesse.
- Rappresentazione di strutture sociali mediante grafi (teoria dei grafi, tipologie di reti, elementi di analisi delle reti: *link analysis*, *web link analysis*, misure di centralità).

4. Descrizione di alcune problematiche di interesse

- Analisi dell’ evoluzione dell’informazione nel tempo.
- Valutazione della credibilità dell'informazione online.

SECONDA PARTE

5. Analisi delle reti sociali

- *Network clustering*: algoritmi di community detection.
- Modelli di influenza e contagio nelle reti sociali.

6. Analisi di soggettività, “sentiment”, emozioni ed ironia

- Lessici.
- Modelli supervisionati e semi-supervisionati.
- Tecniche di *natural language processing.*

7. Named-entity Recognition e Linking

- Modelli grafico probabilistici per *named entity recognition.*
- Metodi di *linking*: apprendimento vs *retrieval.*
- *Word Sense Disambiguation.*

8. Visualizzazione di dati provenienti da social media: problemi aperti e tecniche

Prerequisiti

Conoscenza di base dei principi di algebra lineare, statistica, programmazione, database relazionali e NoSQL.

Modalità didattica

- Lezioni
- Esercitazioni
- Laboratori

- Il corso verrà tenuto in lingua inglese

Materiale didattico

- Rahman, Md Saidur. Basic graph theory. Springer, 2017.
- Reinhard Diestel. Graph Theory. Springer, 2017.
- Suliman Hawamdeh, Hsia-Ching Chang. Analytics and Knowledge Management. CRC Press. 2018.
- John Scott. Social Network Analysis. Sage, 2012.
- Bing Liu. Sentiment Analysis and Opinion Mining. Morgan & Claypool. 2016.
- Eneko Agirre, Philip Edmonds. Word Sense Disambiguation: Algorithms and Applications (Text, Speech and Language Technology). Springer. 2007.

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

- **Prova scritta** con esercizi e domande aperte.
    - La prova scritta ha come obiettivo la valutazione estensiva ed intensiva delle competenze teoriche e teorico-pratiche acquisite durante il corso.
- **Progetto di gruppo** (con presentazione orale).
    - Il progetto ha come obiettivo la valutazione della capacità degli studenti di tradurre in ambiti applicativi reali le competenze acquisite durante il corso, attraverso lo sviluppo e l'impiego di soluzioni tecnologiche per l'analisi dei dati nei social media.

- La prova scritta viene valutata su una scala da 0 a 24.
- Gli studenti devono ottenere una valutazione maggiore o uguale a 12 nella prova scritta.
- Il progetto, con relativa discussione orale, viene valutato su una scala da 0 a 9.

Il voto finale sarà dato dalla somma del voto ottenuto nell'esame scritto e dal voto relativo al progetto.

Orario di ricevimento

Su appuntamento:

- Elisabetta Fersini (elisabetta.fersini@unimib.it)
- Marco Viviani (marco.viviani@unimib.it)