



UNIVERSITÀ DEGLI STUDI DI PALERMO

| | | | |
|---|---|-------------------------------|------------------|
| DIPARTIMENTO | Matematica e Informatica | | |
| ANNO ACCADEMICO OFFERTA | 2022/2023 | | |
| ANNO ACCADEMICO EROGAZIONE | 2023/2024 | | |
| CORSO DILAUREA MAGISTRALE | DATA, ALGORITHMS, AND MACHINE INTELLIGENCE | | |
| INSEGNAMENTO | INFORMATION RETRIEVAL AND NATURAL LANGUAGE PROCESSING | | |
| TIPO DI ATTIVITA' | B | | |
| AMBITO | 50341-Discipline Informatiche | | |
| CODICE INSEGNAMENTO | 22449 | | |
| SETTORI SCIENTIFICO-DISCIPLINARI | INF/01 | | |
| DOCENTE RESPONSABILE | PILATO GIOVANNI | Professore incaricato esterno | Univ. di PALERMO |
| ALTRI DOCENTI | | | |
| CFU | 6 | | |
| NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE | 102 | | |
| NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA | 48 | | |
| PROPEDEUTICITA' | | | |
| MUTUAZIONI | | | |
| ANNO DI CORSO | 2 | | |
| PERIODO DELLE LEZIONI | 2° semestre | | |
| MODALITA' DI FREQUENZA | Facoltativa | | |
| TIPO DI VALUTAZIONE | Voto in trentesimi | | |
| ORARIO DI RICEVIMENTO DEGLI STUDENTI | PILATO GIOVANNI Mercoledì 10:00 13:00 ICAR-CNR Via Ugo La Malfa 153, Palermo, 2 piano corpo B, PilatoPossibilita di incontro su Teams | | |

| | |
|--|---|
| PREREQUISITI | Conoscenza di base di matematica. Familiarità con i concetti di base di calcolo delle probabilità e algebra lineare. Conoscenza di Fondamenti di Informatica e familiarità con almeno un linguaggio di programmazione. |
| RISULTATI DI APPRENDIMENTO ATTESI | <p>Conoscenza e capacità di comprensione Il corso si propone di fornire agli studenti le conoscenze necessarie per comprendere e affrontare varie questioni relative all'analisi intelligente di dati presenti sul Web.</p> <p>Capacità di applicare conoscenze e comprensione Capacità di analizzare le tematiche relative all'analisi intelligente di dati web, con particolare riguardo ai dati di tipo testuale e proporre soluzioni progettuali per la gestione e l'analisi dei dati in tali contesti.</p> <p>Autonomia di giudizio Capacità di analizzare e valutare soluzioni per la gestione di grandi quantità di testi. Capacità di progettare sistemi di analisi dei dati avanzati in grado di comprendere la semantica dei documenti. Ogni studente avrà anche la possibilità di arricchire la propria autonomia di giudizio attraverso la realizzazione di un progetto di studio che contribuirà alla valutazione finale.</p> <p>Abilità comunicative Il corso svilupperà l'interazione tra gli studenti e la capacità di lavorare in gruppo, al fine di affrontare i problemi e trovare soluzioni sulla base delle conoscenze acquisite durante il corso. L'acquisizione delle capacità di comunicazione sarà realizzata attraverso la partecipazione attiva degli studenti alle lezioni, nonché l'esposizione dei risultati ottenuti su temi o problemi proposti dal docente.</p> <p>Capacità di apprendimento Il materiale didattico proposto in aula farà sviluppare le capacità di apprendimento degli studenti che saranno in grado di "interrogare" in modo integrato proprie conoscenze-competenze per risolvere i problemi affrontati. Gli studenti saranno guidati ad una conoscenza più approfondita e critica dei sistemi di Web Data Analytics.</p> <p>Competenze acquisite: - Implementare i metodi di base degli algoritmi di information retrieval, di natural language processing, e di sentiment analysis per risolvere un problema industriale e selezionare l'algoritmo più adatto ad un contesto specifico. - Mostrare le differenze tra l'elaborazione di dati di tipo strutturato e non strutturato e descrivere a un gruppo di professionisti i vantaggi e gli svantaggi dei sistemi automatici di elaborazione di dati di tipo non strutturato. - Mostrare le differenze tra i vari modelli di rappresentazione di dati di tipo testuale e presentare un'applicazione a un gruppo di professionisti finalizzata l'uso di sistemi basati sull'elaborazione del linguaggio naturale.</p> |
| VALUTAZIONE DELL'APPRENDIMENTO | <p>Progetto di studio, prova orale. Viene assegnato un progetto di studio agli studenti. Tale compito richiede la progettazione e la realizzazione di un sistema di elaborazione del linguaggio naturale orientato all'analisi di dati provenienti dal web o al recupero di informazioni semantiche. Il lavoro verrà svolto in gruppi composti da un massimo di 3 studenti.</p> <p>Il progetto coadiuva e stimola anche la capacità di interazione e di lavoro in gruppo. Gli studenti sono tenuti a descrivere la loro soluzione, mostrando le loro capacità di comunicazione.</p> <p>Verrà dato un punteggio complessivo al progetto, tale punteggio sarà ponderato per ogni studente a seconda del contributo individuale.</p> <p>Una valutazione positiva dipenderà l'originalità della soluzione proposta, dal rigore metodologico, dalla dimostrazione di avere acquisito e fatte proprie le competenze tecniche fornite dal corso.</p> <p>La valutazione di tale attività è espressa nell'intervallo [0-30] secondo la tabella di valutazione di seguito descritta e rappresenta per il 60% del punteggio finale.</p> <p>Prova orale: servirà per verificare le conoscenze acquisite durante il corso, l'autonomia dello studente nell'approfondire contenuti anche complessi e la capacità di trovare soluzioni individuali ai problemi proposti.</p> <p>La prova orale inizierà con la presentazione del progetto di studio, poi saranno fatte allo studente almeno tre domande per valutare il grado di conoscenza dello stesso sugli argomenti trattati nel corso.</p> <p>La valutazione di tale attività è espressa nell'intervallo [0-30] secondo la tabella di valutazione di seguito descritta e rappresenta il 40% del punteggio finale.</p> <p>Tabella di valutazione: Valutazione: Eccellente Voto: 30 - 30 e lode Esito: Ottima conoscenza degli argomenti, ottima proprietà di linguaggio, buona</p> |

| | |
|---------------------------------------|--|
| | <p>capacità analitica, efficace capacità di interazione e di lavoro in gruppo, lo studente è in grado di applicare le conoscenze per risolvere i problemi proposti. Valutazione: Molto buono Voto: 27-29</p> <p>Esito: Buona padronanza degli argomenti, piena proprietà di linguaggio, buona capacità di interazione e di lavoro in gruppo, lo studente è in grado di applicare le conoscenze per risolvere i problemi proposti. Valutazione: Buono Voto: 24-26</p> <p>Esito: Conoscenza di base dei principali argomenti, discreta proprietà di linguaggio, discreta capacità di lavoro in gruppo, con limitata capacità di applicare in modo autonomo le conoscenze alla soluzione dei problemi proposti. Valutazione: Soddisfacente Voto: 21-23</p> <p>Esito: Non ha piena padronanza degli argomenti principali dell'insegnamento ma ne possiede le conoscenze, soddisfacente proprietà di linguaggio, scarsa capacità di saper applicare in modo autonomo le conoscenze acquisite. Valutazione: Sufficiente Voto: 18-20</p> <p>Esito: Minima conoscenza di base degli argomenti principali dell'insegnamento e del linguaggio tecnico, scarsissima o nulla capacità di saper applicare in modo autonomo le conoscenze acquisite. Valutazione: Insufficiente Esito: Non possiede una conoscenza accettabile degli argomenti trattati nell'insegnamento.</p> |
| OBIETTIVI FORMATIVI | <p>Il corso si propone di fornire una panoramica di base dell'analisi intelligente di dati sul web e le tecnologie di base dei moderni sistemi di opinion mining, sentiment analysis ed emotion detection from text.</p> <p>Tra gli argomenti verranno affrontati: progettazione e realizzazione di sistemi di Information Retrieval, tecniche automatiche di analisi del testo, tecniche di classificazione di dati di tipo testuale, modelli di gestione della conoscenza strutturata, fondamenti di opinion mining e sentiment analysis, standard rilevanti per la rappresentazione dei dati e della conoscenza. Per dimostrare l'implementazione degli approcci verrà adoperato il linguaggio Python.</p> |
| ORGANIZZAZIONE DELLA DIDATTICA | <p>Lezioni frontali. Al fine di facilitare la partecipazione, verranno implementate metodologie di apprendimento collaborativo.</p> |
| TESTI CONSIGLIATI | <p>Testi di riferimento/Reference books:</p> <p>Ricardo Baeza-Yates, Berthier Ribeiro-Neto - Modern Information Retrieval: The Concepts and Technology Behind Search, Addison-Wesley Professional, ISBN 0321416910</p> <p>Steven Bird, Ewan Klein, and Edward Loper - Natural Language Processing with Python - https://www.nltk.org/book/</p> <p>Testi suggeriti per approfondimento/Suggested texts for further study:</p> <p>Jens Albrecht, Sidharth Ramachandran, Christian Winkler - Blueprints for Text Analytics Using Python: Machine Learning-Based Solutions for Common Real World (NLP) Applications - O'Reilly - ISBN 149207408X</p> <p>Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, Antonio Feraco - A Practical Guide to Sentiment Analysis - Springer - ISBN: 3319856480</p> |

PROGRAMMA

| ORE | Lezioni |
|-----|---|
| 1 | Introduzione al corso Il Web - Il problema dell'Information Retrieval - Data e Information Retrieval - Un Web comprensibile alle machine - Il Web Semantico. |
| 4 | Information Retrieval - IR Systems -IR Models - Information Retrieval Classica - Modelli basati sugli insiemi. |
| 4 | Modelli algebrici alternativi - Modelli probabilistici alternativi - Retrieval di testo strutturato - Modello Cranfield - Metriche di retrieval. |
| 5 | Classificazione dei testi - Caratterizzazione della classificazione di testi - Algoritmi non supervisionati . Algoritmi supervisionati Selezione delle caratteristiche o riduzione di dimensionalità - Metriche di valutazione - Organizzazione delle classi - Costruzione di tassonomie. |
| 3 | Indicizzazione e ricerca sul web - Indici Invertiti - Web Crawling - Tassonomia di Crawlers - Crawlers focalizzati e algoritmi correlati. |
| 4 | Elementi di linguaggio Python - Usare l'interprete Python - Numeri - Stringhe - Liste - Strumenti per il controllo del flusso - Definizione di funzioni - Strutture dati - Tuple e sequenze - Sets - Dizionari - Input e Output. |
| 5 | Python per l'elaborazione dei testi - Uso della libreria NLTK - Accedere a corpora e risorse lessicali - Elaborare testo grezzo - Categorizzare ed etichettare parole - Apprendere come classificare testi. |
| 4 | Estrarre informazioni dal testo - Analizzare la struttura di una frase - Costruire grammatiche basate su caratteristiche - analizzare il significato delle frasi. |

PROGRAMMA

| ORE | Lezioni |
|-----|---|
| 2 | Introduzione al Sentiment Analysis e alla individuazione delle emozioni nei testi - Applicazioni dei Sentiment Analysis. |
| 1 | Generazione del lessico dei sentimenti - Approccio basato sul dizionario - Approccio basato sul corpus - Fatti desiderabili e indesiderati. |
| 3 | Tecniche di individuazione delle emozioni dal testo. |
| 2 | Classificazione del sentimento nei testi - Classificazione del sentimento tramite apprendimento supervisionato - Classificazione del sentimento tramite apprendimento non supervisionato - Classificazione del sentimento su domini diversi - Classificazione del sentimento su lingue diverse. |
| 2 | Soggettività della frase e classificazione del sentimento - Classificazione della soggettività - Classificazione del sentimento della frase - Trattare le frasi condizionali - Trattare le frasi sarcastiche - Soggettività e classificazione del sentimento in più lingue - Uso delle informazioni sul discorso per la classificazione del sentimento. |
| 2 | Riassunto delle opinioni - Diversi tipi di opinioni - Soggettività ed emozione - Punto di vista dell'autore e del lettore. |
| 3 | Analisi del sentimento basata sugli aspetti - Classificazione del sentimento basato sugli aspetti - Regole di base delle opinioni e semantica compositiva - Estrazione degli aspetti - Identificazione dell'aspetto dell'uso delle risorse - Espansione simultanea del lessico delle opinioni ed estrazione degli aspetti - Raggruppamento degli aspetti in categorie - Estrazione di entità, titolari delle opinioni e tempo - Risoluzione della coreferenza e disambiguazione del senso delle parole. |
| 3 | Estrarre conoscenza dal Web e dai Social Networks. |