UNIVERSITÀ DI PISA

HUMAN LANGUAGE TECHNOLOGIES

LUCIA PASSARO

Academic year	2023/24
Course	INFORMATICA
Code	649AA
Credits	9

Modules	Area	Type	Hours	Teacher(s)
HUMAN LANGUAGE TECHNOLOGIES	INF/01	LEZIONI	72	LUCIA PASSARO

Obiettivi di apprendimento

Conoscenze
Al termine del corso:

- Gli studenti avranno acquisito conoscenze sulle tecniche, gli algoritmi e i modelli fondamentali utilizzati nell'elaborazione del linguaggio naturale.
- Gli studenti comprenderanno l'architettura delle tipiche applicazioni sul testo e le librerie per costruirle.
- Gli studenti avranno competenze nella progettazione, implementazione e valutazione di applicazioni basate sull'analisi, la comprensione e la trasformazione dei testi.

Modalità di verifica delle conoscenze
La valutazione continua monitorerà i progressi attraverso test periodici, presentazioni e discussioni che coinvolgeranno il docente e i gruppi di studenti che sviluppano il progetto. Gli studenti saranno valutati in base alla loro capacità di discutere i contenuti del corso utilizzando la terminologia appropriata e di applicare le tecniche di elaborazione del linguaggio naturale.

Capacità
Al termine del corso, lo studente sarà in grado di progettare, implementare e valutare applicazioni che coinvolgano l'analisi, l'interpretazione e la trasformazione dei testi. Lo studente sarà in grado di presentare in una relazione scritta i risultati dell'attività svolta per la realizzazione di un progetto pratico su un compito NLP.

Modalità di verifica delle capacità
Agli studenti verranno assegnati dei compiti (come parte dell'esame intermedio o come progetti finali) per valutare le loro capacità di sviluppo e di comprensione della ricerca.

Comportamenti
Gli studenti saranno in grado di analizzare un problema di elaborazione del testo, selezionare i metodi corretti per risolverlo e implementare una soluzione funzionante. Saranno consapevoli di diversi problemi legati all'elaborazione del testo, tra cui l'affidabilità dei risultati.

Modalità di verifica dei comportamenti
Durante lo sviluppo del progetto, vengono verificate le modalità di definizione del problema, di gestione delle responsabilità, e l'organizzazione delle fasi progettuali.

Prerequisiti (conoscenze iniziali)
Prerequisiti utili:

- Capacità di programmazione, conoscenza del linguaggio di programmazione Python.
- Calcolo elementare e algebra lineare (es. corso “Computational Mathematics for learning and data analysis” (646AA))
- Probabilità e statistica (es. corso "Calcolo delle Probabilità e Statistica" (269AA))
- Apprendimento automatico (es. corso "Machine Learning" (654AA))

Indicazioni metodologiche
La lingua ufficiale del corso è l'inglese: tutti i materiali, i riferimenti e i libri sono in inglese. Le diapositive delle lezioni saranno disponibili, insieme

alle letture suggerite, sul sito web del corso.
Attività di apprendimento:

- partecipazione a lezioni
- partecipazione a seminari di aziende e/o esperti
- partecipazione a discussioni
- studio individuale
- esercitazioni
- progetto di gruppo

Programma (contenuti dell'insegnamento)

Il corso presenta principi, modelli e tecniche all'avanguardia per l'analisi e la generazione del linguaggio naturale, concentrandosi principalmente su approcci di apprendimento automatico statistico e in particolare sull'apprendimento profondo (Deep Learning). Gli studenti impareranno come applicare queste tecniche in una vasta gamma di applicazioni utilizzando moderne librerie di programmazione.
Fondamenti dell'Elaborazione del Linguaggio Naturale

- Introduzione all'NLP, Sviluppi Storici, Applicazioni Chiave
- Metodi Statistici in NLP: Language Models, Hidden Markov Model, Viterbi Algorithm, Generative vs. Discriminative Models
- Elementi Essenziali di Linguistica: Parole, Lemmi, Morfologia, Parti del Discorso (PoS), Frasi e Struttura delle Frasi, Riconoscimento di Entità Nominative (NER)
- Tecniche di Parsing: Constituency Parsing, Dependency Parsing

Metodi e Tecniche di NLP

- Pipeline di Annotazione, Annotazione Linguistica, Annotazione Semantica, Strumenti e Framework per l'annotazione del testo
- Semantica Lessicale: Collocazioni e multi-word expressions, Corpora and Corpus Linguistics, Thesauri e Gazetteers
- Semantica Distribuzionale: Word Embeddings, Character Embeddings, Similarità Semantica
- Deep Learning per l'elaborazione del Linguaggio naturale

Applicazioni e Studi di Caso

- Estrazione di Informazioni: Riconoscimento di Entità, Collegamento di Entità
- Classificazione del Testo e Riassunto: Tecniche di Classificazione, Modelli di Riassunto
- Applicazioni Avanzate di NLP: Analisi delle Opinioni e Sentiment Analysis, Risposte alle Domande, Interfacce Dialogiche (Chatbot), Traduzione Automatica

Il programma potrebbe subire variazioni in base al ritmo del corso e alle tendenze emergenti.

Bibliografia e materiale didattico

- D. Jurafsky, J.H. Martin, Speech and Language Processing. 3rd edition, Prentice-Hall, 2021.
- S. Bird, E. Klein, E. Loper. Natural Language Processing with Python.
- Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016.

Additional Material
C. Manning, H. Schutze. Foundations of Statistical Natural Language Processing. MIT Press, 2000.
S. Kubler, R. McDonald, J. Nivre. Dependency Parsing. 2010.
P. Koehn. Statistical Machine Translation. Cambridge University Press, 2010.

Indicazioni per non frequentanti

Gli studenti non frequentanti non possono svolgere i compiti intermedi. L'esame consisterà nello sviluppo di un progetto (con presentazione finale) e in una discussione orale sia sul progetto che sulla parte teorica.

Modalità d'esame

L'esame del corso per gli studenti che frequentano le lezioni si svolge in due fasi, comprendenti compiti intermedi, lo sviluppo di un progetto (con presentazione finale) e una discussione orale. Gli esami intermedi forniscono una dispensa per la discussione orale.

Pagina web del corso

https://classroom.google.com/c/NjUxNDAyMzczNTA2?cjc=dmwt6sk

Ultimo aggiornamento 05/02/2024 10:38