



UNIVERSITÀ DI PISA

LINGUISTICA COMPUTAZIONALE

ALESSANDRO LENCI

Anno accademico	2022/23
CdS	INFORMATICA UMANISTICA
Codice	305LL
CFU	12

Moduli	Settore/i	Tipo	Ore	Docente/i
LINGUISTICA COMPUTAZIONALE	L-LIN/01	LEZIONI	72	ALESSANDRO BONDIELLI ALESSANDRO LENCI

Obiettivi di apprendimento

Conoscenze

Lo studente acquisirà conoscenze in merito agli strumenti e alle metodologie per l'analisi computazionale dei testi, alle tecnologie per il trattamento automatico della lingua, e al linguaggio di programmazione Python per le analisi testuali

Modalità di verifica delle conoscenze

L'accertamento delle conoscenze verrà svolto attraverso:

- prova scritta (divisa in due parti), nella quale lo studente dovrà dimostrare la conoscenza dei temi svolti a lezioni e dei testi di esame rispondendo a domande aperte e risolvendo esercizi. La prova scritta si terrà in occasione di ogni appello di esame. In alternativa, **solo gli studenti frequentanti** potranno sostenere due prove in itinere
- realizzazione di un progetto sui temi delle esercitazioni di laboratorio
- prova orale, in cui verrà discussa la prova scritta, il progetto e verrà verificata la conoscenza del programma di esame

Capacità

Lo studente saprà utilizzare le tecniche di base per l'analisi computazionale dei testi e imparerà a usare l'ambiente di programmazione Python NLTK per lo sviluppo di moduli per il trattamento automatico della lingua

Modalità di verifica delle capacità

Lo studente realizzerà un progetto di costruzione e analisi computazionale di un corpus con NLTK

Comportamenti

Lo studente potrà acquisire e/o sviluppare sensibilità alle problematiche relative all'analisi quantitativa e computazionale del testo e al trattamento automatico della lingua

Saranno acquisite accuratezza e precisione nella raccolta e nell'analisi computazionale di dati linguistici

Modalità di verifica dei comportamenti

Durante le sessioni di laboratorio saranno valutati il grado di accuratezza e precisione delle attività svolte

Prerequisiti (conoscenze iniziali)

Nozioni di base di linguistica generale (morfologia, sintassi e semantica)

Nozioni di base di programmazione

Indicazioni metodologiche

Lezioni: frontali

Modalità di apprendimento:

- frequenza delle lezioni
- studio individuale
- attività di laboratorio

Frequenza: raccomandata



UNIVERSITÀ DI PISA

Metodi di insegnamento:

- lezioni
- laboratorio

Programma (contenuti dell'insegnamento)

Il corso ha lo scopo di presentare i temi principali della linguistica computazionale e del "Natural Language Processing" (NLP) e di familiarizzare lo studente con gli strumenti di base per l'analisi quantitativa e computazionale del testo.

Lezioni teoriche

1. Analisi computazionale dei dati linguistici

- corpora: tipologia e uso
- codifica e rappresentazione dei dati linguistici
- analisi statistica dei dati linguistici
- linguaggio e probabilità: modelli di markov, entropia, legge di Bayes
- metodi computazionali per l'esplorazione dei dati linguistici: concordanze, collocazioni e misure di associazione

2. Natural Language Processing (NLP)

- metodi a regole e di machine learning
- corpora annotati per lo sviluppo di strumenti di NLP
- analisi morfologica: Automi e Trasduttori a Stati Finiti
- Part of Speech Tagging
- analisi sintattica: chunking e shallow parsing, grammatiche context-free, cenni di parsing a costituenti e a dipendenze, probabilistic context free grammars
- metodi di valutazione
- lessici semantici computazionali: Wordnet e FrameNet
- word sense disambiguation

Esercitazioni

- espressioni regolari
- introduzione a NLTK
- analisi computazionali del testo con NLTK
- annotazioni semantiche del testo

Bibliografia e materiale didattico

Testi la cui conoscenza è oggetto di verifica all'esame:

- A. Lenci, S. Montemagni, V. Pirrelli, Testo e computer. Elementi di linguistica computazionale, Carocci, 2005
- D. Jurafsky & J.H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (second edition), Prentice Hall, 2008 (capp. 1, 2, 3, 4.1-4.3, 4.5.1, 4.10-4.11, 5.1-5.4, 12, 13.1-13.3, 13.5, 14.1, 19, 20.1-20.5)

Testo di consultazione per le esercitazioni:

- Bird, S., Klein, E., & E. Loper, Natural Language Processing with Python (Analyzing Text with the Natural Language Toolkit), O'Reilly Media, 2009

Indicazioni per non frequentanti

Testi la cui conoscenza è oggetto di verifica all'esame:

- A. Lenci, S. Montemagni, V. Pirrelli, Testo e computer. Elementi di linguistica computazionale, Carocci, 2005
- D. Jurafsky & J.H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (second edition), Prentice Hall, 2008 (capp. 1, 2, 3, 4.10-4.11, 5.1-5.4, 12, 13.1-13.3, 13.5, 19)

Testo di consultazione per la realizzazione del progetto:

- Bird, S., Klein, E., & E. Loper, Natural Language Processing with Python (Analyzing Text with the Natural Language Toolkit), O'Reilly Media, 2009

Modalità d'esame

Esame scritto + progetto + esame orale

L'esame scritto, diviso in due prove, può essere sostenuto o come compiti in itinere durante il corso oppure in occasione degli appelli scritti (le due prove scritte possono essere sostenute anche in appelli differenti). Le prove in itinere possono essere effettuate SOLO dagli studenti frequentanti.



UNIVERSITÀ DI PISA

Il progetto viene presentato e discusso in occasione degli appelli orali. Il superamento dell'esame scritto (18/30 in ciascuna prova scritta) è condizione strettamente necessaria per presentarsi all'orale e discutere il progetto.

Il progetto dà diritto a un bonus fino a due punti, che si aggiunge alla media dei punteggi ottenuti nelle prove scritte. In occasione della prova orale, verrà verificata la conoscenza del programma di esame con ulteriori domande.

Note

Commissione di esame:

- Alessandro Lenci
- Alessandro Bondielli

NB: Tutte le lezioni si svolgeranno in presenza. Solo le lezioni di Laboratorio del mercoledì saranno trasmesse anche in streaming (senza registrazione) [su Teams](#)

INIZIO LEZIONI: lunedì 19 settembre

Ultimo aggiornamento 20/09/2022 10:06