



# UNIVERSITÀ DI PISA

---

## LINGUISTICA COMPUTAZIONALE

**ALESSANDRO LENCI**

Anno accademico 2016/17  
CdS INFORMATICA UMANISTICA  
Codice 305LL  
CFU 12

Moduli	Settore/i	Tipo	Ore	Docente/i
LINGUISTICA COMPUTAZIONALE	L-LIN/01	LEZIONI	72	FELICE DELL'ORLETTA ALESSANDRO LENCI

### Obiettivi di apprendimento

#### *Conoscenze*

lo studente avrà acquisito conoscenze in merito agli strumenti e alle metodologie per l'analisi computazionale dei testi e le tecnologie per il trattamento automatico della lingua

#### *Modalità di verifica delle conoscenze*

Per l'accertamento delle conoscenze saranno svolte prove in itinere, un elaborato scritto previsto all'inizio di ogni sessione d'esame e un progetto di analisi computazionale del testo

#### *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 legate all'analisi quantitativa e computazionale del testo e al trattamento automatico della lingua

Saranno acquisite accuratezza e precisione nella raccolta e 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

#### *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
- corpora annotati
- 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)



## UNIVERSITÀ DI PISA

---

- 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
- metodi di valutazione
- lessici semantici computazionali: Wordnet e FrameNet

### 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, 4.10-4.11, 5.1-5.4, 12, 13.1-13.3, 13.5, 19)

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

Gli studenti non frequentanti DEVONO contattare preventivamente il docente

### Modalità d'esame

Esame scritto + progetto

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.

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.

*Ultimo aggiornamento 14/11/2016 17:27*