



---

# INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE E LABORATORIO

12 CFU - 1° semestre

## Docenti titolari dell'insegnamento

**SIMONE FARO** - MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE - INF/01 - 6 CFU

**Email:** faro@dmi.unict.it

**Edificio / Indirizzo:** Dipartimento di Matematica e Informatica, Viale A.Doria n.6

**Telefono:** 095 7383053

**Orario ricevimento:** Disponibile all'indirizzo <http://www.dmi.unict.it/~faro/calendar.php>

**DOCENTE NON ANCORA ASSEGNATO** - MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE - INF/01 - 6 CFU

---

## MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

### ▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

L'insegnamento sarà svolto attraverso delle lezioni frontali in cui verranno presentati i contenuti del corso, anche attraverso delle dimostrazioni pratiche di programmazione in aula. L'insegnamento prevede l'applicazione dei concetti attraverso l'utilizzo del linguaggio Python. Lo studente avrà a disposizione ulteriori ore di lezione frontale (previste dal modulo di Laboratorio), durante le quali avrà la possibilità di perfezionare la propria preparazione sulla programmazione strutturata in Python. Inoltre lo studente avrà a disposizione una piattaforma di apprendimento attraverso la quale sarà possibile esercitarsi durante le ore di studio e autovalutarsi sui contenuti appresi a lezione. La medesima piattaforma fornisce un valido strumento per la preparazione all'esame.

## PREREQUISITI RICHIESTI

### ▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

Lo studente che accede al corso dovrà conoscere i fondamenti dell'Informatica. In particolare lo studente dovrà avere dimestichezza con il concetto di informazione e con i principi di funzionamento del sistema di elaborazione e del sistema operativo.

---

## FREQUENZA LEZIONI

### ▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

La frequenza delle lezioni è obbligatoria.

---

## CONTENUTI DEL CORSO

### ▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

Il corso prevede l'acquisizione da parte degli studenti della filosofia della programmazione strutturata e l'apprendimento delle principali tecniche per l'elaborazione del testo, attraverso l'utilizzo del linguaggio Python.

Il corso è diviso in due moduli principali. Nel primo modulo lo studente apprenderà le basi della programmazione e del linguaggio Python. Apprenderà l'utilizzo dei costrutti di base, delle funzioni, della ricorsione, e delle principali strutture di dati disponibili in Python. Il secondo modulo del corso ha lo scopo di fornire gli strumenti essenziali per l'elaborazione del testo e del linguaggio naturale, come la tokenizzazione, lo stemming e l'utilizzo del dizionario WordNet. Saranno acquisite inoltre competenze sul part-of-speech-tagging, il chunking e il named-entity-recognition. Lo studente imparerà come sono organizzati i corpora testuali e come crearne di personalizzati e sarà in grado di utilizzare svariate librerie Python per l'analisi e il processing del testo.

---

## TESTI DI RIFERIMENTO

### ▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

Il testo di riferimento al primo modulo è "Pensare in Python: come pensare da informatico" di Allen B. Downey, edito da O'Reilly Media (anno 2019, verranno svolte 213 di 259 pagine complessive). Il libro rappresenta uno strumento ideale per imparare le basi della programmazione, utilizzando il linguaggio Python. Nello specifico il testo introduce al linguaggio in modo graduale, iniziando con i concetti base della programmazione per poi passare a funzioni, ricorsione, strutture di dati e progettazione orientata agli oggetti. Gli esercizi, presenti in ogni capitolo, consentono di testare immediatamente i concetti appresi.

Il testo di riferimento al secondo modulo è "Python 3 Text Processing with NLTK 3 Cookbook" (in lingua inglese, verranno svolte 228 di 279 pagine complessive), di Jacob Perkins, edito da Packt Publishing (anno 2014). Il testo introduce lo studente alle tecniche essenziali di elaborazione del testo e del linguaggio naturale.

Alcuni temi di approfondimento saranno tratti dal testo "Applied Text Analysis with Python", di Benjamin Bengfort, Tony Ojeda e Rebecca Bilbro, edito da O'Reilly Media (anno 2018). Il libro introduce alcune tecniche per l'analisi del testo con Python, tra le quali la vettorizzazione, la classificazione, il topic modelling, la entity resolution e la rappresentazione grafica dei dati.

Di seguito la lista dei libri di testo utilizzati durante il corso:

[A] Allen B. Downey - Pensare in Python: come pensare da informatico - O'Reilly Media (2019)

[B] Jacob Perkins - Python 3 Text Processing with NLTK 3 Cookbook - Packt Publishing (2014)

[C] Benjamin Bengfort, Tony Ojeda, Rebecca Bilbro - Applied Text Analysis with Python: Enabling Language - Aware Data Products with Machine Learning - O'Reilly Media (2018)

## ALTRO MATERIALE DIDATTICO

### ▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

Durante il corso verrà fornito agli studenti il codice degli esempi e degli esercizi visti a lezione. Lo studente avrà a disposizione una piattaforma di apprendimento attraverso la quale sarà possibile esercitarsi durante le ore di studio e autovalutarsi sui contenuti appresi a lezione. La medesima piattaforma fornirà un valido strumento per la propria preparazione all'esame.

---

## PROGRAMMAZIONE DEL CORSO

### MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

	<b>Argomenti</b>	<b>Riferimenti testi</b>
1	Imparare a programmare	Cap.1 di [A]
2	Variabili, espressioni ed istruzioni	Cap.2 di [A]
3	Funzioni	Cap.3 di [A]
4	Istruzioni condizionali e ricorsione	Cap.4 di [A]
5	Funzioni produttive	Cap.5 di [A]
6	Iterazione	Cap.6 di [A]
7	Stringhe	Cap.7 di [A]
8	Liste	Cap.8 di [A]
9	Tuple	Cap.9 di [A]
10	Dizionari	Cap.10 di [A]
11	File ed eccezioni	Cap.11 di [A]
12	Tokenizzazione del Testo	Cap.1 di [B]
13	Utilizzo di WordNet	Cap.1 di [B]
14	Sostituzione e correzione di parole	Cap.2 di [B]
15	Creazione di Corpora	Cap.3 di [B]
16	Part of speech tagging	Cap.4 di [B]
17	Chunk Extraction	Cap.5 di [B]
18	Trasformazioni	Cap.6 di [B]
19	Classificazione	Cap.7 di [B]
20	Tecniche di visualizzazione	Cap.8 di [C]
21	Analisi	Cap.9 di [C]

---

---

## **VERIFICA DELL'APPRENDIMENTO**

### **MODALITÀ DI VERIFICA DELL'APPRENDIMENTO**

- **MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE**

L'esame si svolge in due prove. La prima prova di laboratorio consiste nello sviluppo di un programma in Python che prevede la risoluzione di alcuni problemi computazionali inerenti l'analisi del testo. La seconda prova consiste in un esame orale sui contenuti del corso.

### **ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI**

- **MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE**

I testi delle prove di laboratorio in Python saranno disponibili sul sito internet dei docente. Inoltre lo studente avrà a disposizione una piattaforma di apprendimento attraverso la quale sarà possibile esercitarsi durante le ore di studio e autovalutarsi sui contenuti appresi a lezione. La medesima piattaforma fornisce un valido strumento per la preparazione all'esame.

---