



INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE E LABORATORIO

12 CFU - Insegnamento annuale

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>

LORENZO DI SILVESTRO - MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE - INF/01 - 6 CFU

Email: disilvestro@dmi.unict.it

Edificio / Indirizzo: DMI - I blocco - 1° piano - Stanza 200

Telefono: 000

Orario ricevimento: Quasi sempre, previo appuntamento

OBIETTIVI FORMATIVI

▪ 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.

▪ MODULO B LABORATORIO DI 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.

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.

Qualora l'insegnamento venisse impartito in modalità mista o a distanza potranno essere introdotte le necessarie variazioni rispetto a quanto dichiarato in precedenza, al fine di rispettare il programma previsto e riportato nel syllabus.

▪ **MODULO B LABORATORIO DI 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.

Qualora l'insegnamento venisse impartito in modalità mista o a distanza potranno essere introdotte le necessarie variazioni rispetto a quanto dichiarato in precedenza, al fine di rispettare il programma previsto e riportato nel syllabus.

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.

▪ **MODULO B LABORATORIO DI 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**

Obbligatoria.

▪ **MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE**

Obbligatoria.

CONTENUTI DEL CORSO

▪ **MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE**

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.

▪ **MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE**

Questo modulo del corso ha lo scopo di fornire gli strumenti essenziali per l'elaborazione del testo e del linguaggio naturale, come la divisione in token, lo stemming e l'utilizzo del dizionario WordNet. Saranno acquisite inoltre competenze sul part-of-speech tagging, il chunking e named-entity-recognition. Lo studente imparerà come realizzare ogni task richiesto per l'analisi e l'interpretazione del testo in maniera automatica mediante la libreria NLTK e il linguaggio di programmazione Python. Successivamente la stessa pipeline di lavorazione verrà replicata mediante la libreria SpaCy, confrontando il funzionamento e le finalità delle due librerie. Con quest'ultima verranno anche provati due semplice sistemi di visualizzazione dei risultati.

TESTI DI RIFERIMENTO

▪ **MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE**

Il testo di riferimento è "Pensare in Python: come pensare da informatico" di Allen B. Downey, edito da O'Reilly Media (anno 2019, verranno svolte le prime 213 di 259 pagine). 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.

Si ricorda che, ai sensi dell'art. 171 della legge 22 aprile 1941, n. 633 e successive disposizioni, fotocopiare libri in commercio, in misura superiore al 15% del volume o del fascicolo di rivista, è reato penale.

Per ulteriori informazioni sui vincoli e sulle sanzioni all'uso illecito di fotocopie, è possibile consultare le Linee guida sulla gestione dei diritti d'autore nelle università (a cura della Associazione Italiana per i Diritti di Riproduzione delle opere dell'ingegno - AIDRO).

I testi di riferimento possono essere consultati in Biblioteca.

▪ **MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE**

Il testo di riferimento 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. La seconda parte del secondo modulo, relativa al Natural Language Processing tramite SpaCy verrà trattata seguendo dispense fornite dal docente e la documentazione ufficiale di SpaCy reperibile online (<https://spacy.io/usage>).

Si ricorda che, ai sensi dell'art. 171 della legge 22 aprile 1941, n. 633 e successive disposizioni, fotocopiare libri in commercio, in misura superiore al 15% del volume o del fascicolo di rivista, è

reato penale.

Per ulteriori informazioni sui vincoli e sulle sanzioni all'uso illecito di fotocopie, è possibile consultare le Linee guida sulla gestione dei diritti d'autore nelle università (a cura della Associazione Italiana per i Diritti di Riproduzione delle opere dell'ingegno - AIDRO).

I testi di riferimento possono essere consultati in Biblioteca.

ALTRO MATERIALE DIDATTICO

▪ **MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE**

Altri materiali per lo studio e l'approfondimento dei contenuti trattati durante il corso potranno essere forniti attraverso la piattaforma Studium.

▪ **MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE**

Altri materiali per lo studio e l'approfondimento dei contenuti trattati durante il corso potranno essere forniti attraverso la piattaforma Studium.

PROGRAMMAZIONE DEL CORSO

MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

	Argomenti	Riferimenti testi
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]

MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE

	Argomenti	Riferimenti testi
1	Tokenizzazione del testo	Cap.1

2	Rimozione stopwords	Cap. 1
3	Utilizzo di WordNet	Cap. 1
4	Collocations	Cap. 1
5	Sostituzione e corezione di parole	Cap. 2
6	Creazione di Corpora	Cap. 3
7	Part of speech tagging	Cap. 4
8	Chunk Extraction	Cap. 5
9	Classificazione	Cap. 7 e Dispense
10	Named Entity Recognition con SpaCy	Dispense
11	Dependency Tree visualizer	Dispense
12	NER visualizer	Dispense

VERIFICA DELL'APPRENDIMENTO

MODALITÀ DI VERIFICA DELL'APPRENDIMENTO

▪ MODULO A INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE

Prova pratica al PC e orale

Per la valutazione dell'esame si terrà conto della padronanza dei contenuti e delle competenze acquisite, dell'accuratezza linguistica e proprietà lessicale, nonché della capacità argomentativa dimostrata dal candidato.

La verifica dell'apprendimento potrà essere effettuata anche per via telematica, qualora le condizioni lo dovessero richiedere.

▪ MODULO B LABORATORIO DI LINGUISTICA COMPUTAZIONALE

Prova pratica al PC e orale.

Per la valutazione dell'esame si terrà conto della padronanza dei contenuti e delle competenze acquisite, dell'accuratezza linguistica e proprietà lessicale, nonché della capacità argomentativa dimostrata dal candidato.

La verifica dell'apprendimento potrà essere effettuata anche per via telematica, qualora le condizioni lo dovessero richiedere.
