

FACOLTÀ	Facoltà di Scienze della Formazione
ANNO ACCADEMICO	2014/2015
CORSO DI LAUREA MAGISTRALE	LM 92: Teorie della comunicazione
INSEGNAMENTO	Laboratorio di elaborazione del linguaggio naturale
TIPO DI ATTIVITÀ	Altre conoscenze utili per l'inserimento nel mondo del lavoro
AMBITO DISCIPLINARE	Altre conoscenze utili per l'inserimento nel mondo del lavoro
CODICE INSEGNAMENTO	16555
ARTICOLAZIONE IN MODULI	NO
NUMERO MODULI	
SETTORI SCIENTIFICO DISCIPLINARI	ING-INF/05
DOCENTE RESPONSABILE	Roberto Pirrone (PA) Università degli Studi di Palermo
CFU	3
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	55
NUMERO DI ORE RISERVATE ALLE ATTIVITÀ DIDATTICHE ASSISTITE	20
PROPEDEUTICITÀ	Nessuna
ANNO DI CORSO	2
SEDE DI SVOLGIMENTO DELLE LEZIONI	http://portale.unipa.it/facolta/scienzeform.
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni frontali, Esercitazioni in aula
MODALITÀ DI FREQUENZA	Facoltativa
METODI DI VALUTAZIONE	Elaborato progettuale
TIPO DI VALUTAZIONE	Voto in trentesimi
PERIODO DELLE LEZIONI	Primo semestre
CALENDARIO DELLE ATTIVITÀ DIDATTICHE	http://portale.unipa.it/facolta/scienzeform.
ORARIO DI RICEVIMENTO DEGLI STUDENTI	Il martedì dalle 11.00 alle 13.00, salvo impegni istituzionali, presso il Dip. Ing. Chimica, Gestionale, Informatica, Meccanica, Ed. 6, III piano, stanza 8

Conoscenza e capacità di comprensione

Lo studente, al termine del laboratorio, conoscerà i fondamenti teorici ed i più semplici pacchetti software per l'analisi automatica del testo nonché avrà una comprensione generale delle motivazioni che giustificano il ricorso a tali strumenti e dei contesti applicativi in cui questi vengono utilizzati. Questa capacità verrà valutata attraverso la documentazione a corredo dell'elaborato progettuale di esame.

Capacità di applicare conoscenza e comprensione

Lo studente, al termine del laboratorio, sarà in grado di utilizzare autonomamente le principali funzionalità degli applicativi studiati, con particolare riferimento alle tecniche di pre-analisi, alla ricerca all'interno delle più comuni risorse linguistiche annotate e alla individuazione di entità annotate nel testo, al fine di condurre semplici indagini su testi di proprio interesse culturale. Questa capacità sarà valutata lungo tutto il corso laboratoriale stimolando il dibattito e la

riflessione degli studenti sulle tematiche oggetto di studio nonché l'analisi dell'elaborato progettuale di esame.

Autonomia di giudizio

Lo studente saprà giudicare autonomamente su quali categorie di strumenti dovranno essere utilizzate in dipendenza dallo scopo dell'indagine su un dato testo. Questa capacità sarà valutata lungo tutto il corso laboratoriale stimolando il dibattito e la riflessione degli studenti sulle tematiche oggetto di studio nonché l'analisi dell'elaborato progettuale di esame.

Abilità comunicative

Lo studente acquisirà il lessico proprio dell'analisi automatica dei testi che saprà integrare con le sue competenze comunicative in ambito più strettamente linguistico. Quest'abilità sarà valutata attraverso la documentazione a corredo dell'elaborato progettuale di esame.

Capacità d'apprendimento

Lo studente sarà in grado di approfondire autonomamente qualunque testo anche avanzato riguardante l'analisi automatica dei testi e ne coglierà il significato generale. Egli dovrà però avere il supporto di un esperto di dominio per poter approfondire correttamente le basi teoriche, matematiche e statistiche, delle tecniche investigate. Questa capacità sarà valutata lungo tutto il corso laboratoriale stimolando il dibattito e la riflessione degli studenti sulle tematiche oggetto di studio nonché l'analisi dell'elaborato progettuale di esame.

OBIETTIVI FORMATIVI DEL MODULO

Il "Laboratorio di elaborazione del linguaggio naturale" si propone di fornire allo studente delle competenze di base sull'analisi automatica dei testi per le applicazioni più comuni quali quelle di linguistica computazionale, ma anche di information retrieval e di text e web mining.

Il laboratorio prevede di affrontare i vari passi dell'analisi linguistica automatica: la fase di pre-analisi per la rimozione delle parole troppo comuni e la determinazione delle radici morfologiche dei termini (stemming), l'analisi morfo-sintattica automatica attraverso software addestrati con corpora documentali annotati (POS tagging) il supporto all'analisi lessicale per via dei tesauri informatizzati nelle varie lingue (Wordnet, Framenet, Verbnet) e gli strumenti di analisi semantica automatica quali la Named Entity Recognition (NER) e la Latent Semantic Analysis (LSA). Infine si accennerà alle infrastrutture software per lo sviluppo di applicazioni di analisi del testo quali GATE e UIMA.

Gli argomenti teorici saranno trattati in maniera semplice per via delle loro complesse implicazioni matematiche ed informatiche, ma verranno illustrati i concetti base degli strumenti software più diffusi per condurre le analisi citate di cui si spiegherà il funzionamento e le possibili interazioni.

Il singolo studente finalizzerà la propria attività laboratoriale alla realizzazione di un semplice progetto di analisi automatica di testi concordato con gli altri docenti del Corso di Laurea Magistrale appartenenti all'area linguistica. La valutazione del progetto con la relativa documentazione sarà l'oggetto della prova di esame.

MODULO	Laboratorio di elaborazione del linguaggio naturale
ORE FRONTALI	LEZIONI FRONTALI
1	Introduzione all'analisi automatica del testo; principali applicazioni: linguistica computazionale, information retrieval, text e web mining.
1	Strumenti e tecniche di preprocessing: rimozione delle stop word, stemming.
1	POS tagger: analisi morfo-sintattica di corpora annotati.
1	Risorse linguistiche: WordNet, Framenet, Verbnet.
2	Analisi semantica automatica del testo: concetti base della Latent Semantic Analysis e della Named Entity Recognition.
1	Cenni ai framework GATE e UIMA.
	ESERCITAZIONI
2	Uso di strumenti per la pre-analisi automatica di un testo.
3	Uso di strumenti per l'analisi automatica morfo-sintattica e lessicale di un testo: Stanford Parser, Wordnet, Framenet, Verbnet.
2	Uso di strumenti per la Named Entity Recognition su un testo.
2	Uso di strumenti per l'analisi della semantica latente su un testo.
4	Realizzazione di un piccolo progetto di analisi automatica di un testo mirata ad obiettivi specifici concordati con i docenti di area linguistica.
TESTI CONSIGLIATI	Materiale didattico in forma elettronica disponibile sul sito web di Facoltà Principali siti web di riferimento per gli strumenti di attività laboratoriale: http://verbs.colorado.edu/~mpalmer/projects/verbnet.html http://verbs.colorado.edu/verb-index/ http://www.wordnet-online.com/ http://wordnet.princeton.edu/ http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=834/vers=ita http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=820/vers=ita http://nlp.stanford.edu/software/lex-parser.shtml http://www.taltac.it/it/index.shtml http://nlp.stanford.edu/software/CRF-NER.shtml http://gate.ac.uk/ http://uima.apache.org/