

Scheda proposta attività seminariale Disum per l'anno accademico 2021-22

Titolo	Company profiling system: approcci e tecniche data-driven e language-driven per la creazione di classificatori di aziende
Numero di ore	18
Docente proponente	Antonio Di Silvestro
Nome docente/i incaricato/i	Valentina Russo, Azzurra Mancini, Marco Grazioso
Link per iscrizione	https://tinyurl.com/cucdemjw
Contatto docente incaricato (telefono)	vrusso@logogramma.com
Breve profilo docente	<p>Valentina Russo e Azzurra Mancini sono Dottori di Ricerca in Teoria delle lingue e del linguaggio e fondatrici della start-up Logogramma che si occupa di soluzioni digitali basati sull'elaborazione automatica del linguaggio.</p> <p>Marco Grazioso è Dottorando in Information Technologies and Electrical Engineering presso l'Università degli Studi Federico II di Napoli e specializzato in tecniche di linguistica computazionale.</p>
Descrizione dell'attività seminariale	<p>Il seminario passa in rassegna diverse tecniche per la costruzione di classificatori di testi a partire da approcci prettamente data-driven ed approcci basati su analisi linguistiche, fino ad arrivare ad approcci integrati, ponendo il caso pratico di un Company Profiling System. Sarà articolato in 6 lezioni da 3 ore ciascuno, a carattere prevalentemente laboratoriale:</p> <ol style="list-style-type: none"> 1) Costruzione del corpus e tecniche di preprocessing dei dati (scraping, sentence splitting, etc.) 2/3) Costruzione di un modello di classificazione data-driven (dati già annotati, BERT, sentence embedding, clustering) 4/5) Costruzione di un modello di classificazione a partire da analisi linguistiche (parsing, entity extraction, cosine similarity, etc.) 6) Confronto dei risultati: verso un approccio integrato

Obiettivi formativi	Il seminario intende fornire agli studenti competenze per la creazione di classificatori di testi a partire da un approccio integrato, sfruttando varie tecniche di linguistica computazionale.
Requisiti minimi per accedere al seminario	Non si richiedono competenze informatiche precedenti, ma è consigliata una conoscenza di nozioni di base di linguistica computazionale.
Date inizio e fine prenotazioni	Dal 16 dicembre 2021 al 17 gennaio 2022
Periodo previsto e date indicative di inizio e fine attività seminariale	24 gennaio/26 febbraio 2022.
Luogo in cui si prevede lo svolgimento del seminario	Aula A10 / 252 e/o su piattaforma Teams.
Prova finale (Esempio: elaborato scritto, prova orale, testo creativo, performance, test, etc...) Giudizio finale (Idoneo/Non idoneo)	Progetto. Gli studenti dovranno creare un modello di classificazione su un dataset che verrà loro fornito, utilizzando una o più delle tecniche introdotte nel seminario.
Rilascio attestato finale firmato da consegnare ai responsabili amministrativi	L'accreditamento presso il DISUM prevede il rilascio di un'attestato formale su carta intestata recante il titolo del laboratorio, data, firma del proponente, nonché nome cognome e matricola di ciascun partecipante.