

PROPOSTA DI ATTIVITA' DI TIROCINIO CURRICULARE

Area disciplinare: Linguistica Computazionale

Tutor: Manuela Angioni (CRS4)

Durata prevista: minimo 75 ore

Argomento:

Il CRS4 propone la possibilità a uno o più studenti delle Classi di Laurea di: Lettere; Storia e Filosofia; Lingue e culture per la mediazione linguistica; Lingue e comunicazione; di attivare il proprio tirocinio nell'ambito della Linguistica Computazionale la disciplina che affronta lo studio del linguaggio naturale (il linguaggio umano) per mezzo di strumenti automatici, attraverso la definizione di algoritmi che permettono l'analisi automatica dei testi e l'estrazione del significato in essi contenuto.

Il progetto in cui il tirocinante sarà inserito prevede lo sviluppo, a livello prototipale, di alcune applicazioni rivolte all'analisi sintattica e semantica delle frasi che compongono il testo analizzato. A tale scopo si rende necessaria la disponibilità di particolari risorse linguistiche annotate sintatticamente specifiche per la lingua adoperata nel testo e si prevede la partecipazione alla realizzazione di una risorsa linguistica per la lingua sarda definita sulla base di WordNet, il database lessicale inizialmente realizzato per la lingua inglese e successivamente esteso a numerose altre lingue, sviluppato sotto la direzione di George A. Miller presso l'Università di Princeton (<http://wordnet.princeton.edu/>). In WordNet, i sostantivi, i verbi, gli aggettivi e gli avverbi sono raggruppati secondo insiemi di sinonimi chiamati synset, ciascuno dei quali esprime un concetto distinto. I synset sono collegati tra loro attraverso le relazioni lessicali e semantiche che costituiscono la rete semantica di WordNet.

Obiettivi del Tirocinio:

Il tirocinante parteciperà in modo diretto alla realizzazione della risorsa linguistica per la lingua sarda attraverso la realizzazione di un thesaurus legato ad alcune aree tematiche e prevede le seguenti attività: la creazione e gestione della rete semantica della lingua sarda, che consentirà la navigazione tra le relazioni semantiche esistenti tra i concetti presenti nei thesaurus; il *mapping* di termini in lingua sarda con le lingue preesistenti su WordNet; la partecipazione, quando possibile, alle pubblicazioni sul tema di ricerca svolto.

Prerequisiti:

Corso di studio delle Classi di Laurea di: Lettere; Storia e Filosofia; Lingue e culture per la mediazione linguistica; Lingue e comunicazione.

Conoscenze di base d'informatica. Conoscenza della lingua inglese; Capacità di lavorare in gruppo.

Modalità di svolgimento del tirocinio:

Sede: CRS4, Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna - Pula

Unità Organizzativa: ISOC/NAT

Monte ore settimanale: 15

Facilitazioni previste:

Servizio mensa e biglietto A/R autobus di linea da Cagliari per Piscina Manna (Pula)

Per il **colloquio di selezione** i candidati potranno contattare il tutor mediante telefono: 070.9250322 o email: angioni@crs4.it