

ITTIG
Istituto di Teoria e Tecniche dell'informazione
giuridica

Via dei Barucci 20 Firenze
Aula telematica piano 1

Seminario 30 marzo 2007
Ore 10.30

Giulia Venturi
(Istituto Linguistica Computazionale)

"Metodi e strumenti per l'estrazione automatica di informazione semantico-lessicale da un corpus documentale giuridico in materia ambientale. Dalla costruzione di un glossario terminologico di dominio all'acquisizione di classi di eventi giuridici e ambientali".

Nel seminario verranno illustrati metodi e strumenti per il riconoscimento e l'estrazione automatica dai testi di informazione giuridica da rappresentare in un'ontologia di dominio. L'approccio seguito sarà pertanto quello di isolare e acquisire in modo incrementale i concetti e le relazioni fondamentali costitutive un sistema di organizzazione strutturata della conoscenza di questo tipo.

Lo studio di fattibilità realizzato su di un corpus di atti normativi e amministrativi, europei, statali e regionali, in materia ambientale, ha l'obiettivo di dimostrare in che modo tecniche di trattamento automatico del linguaggio possano potenziare strumenti di estrazione dell'informazione semantica codificata in una base documentale. L'intuizione fondamentale è che sia possibile indagare il testo partendo da categorie di analisi estratte dal testo stesso (bootstrapping). Gli studi più recenti in linguistica computazionale assumono che in un documento i concetti rilevanti sono veicolati dalle unità terminologiche semplici e complesse statisticamente più significative (acquisite in modo automatico, dopo una fase di parsing linguistico del testo stesso) presenti nella base documentale. Partendo di qui, il seminario intende presentare una metodologia per l'acquisizione semi-automatica di classi di eventi di dominio dopo una preliminare indagine sulla reale distribuzione sintagmatica delle co-occorrenze di verbi e nomi nella frase (legati da relazioni sintattico-funzionali minime di soggetto e oggetto). Analizzando cioè classi di verbi simili, legati da relazioni di similarità semantica rispetto alla propria struttura argomentale, si vuole proporre un'ipotesi per la costruzione di una mappa semantica multidimensionale del corpus di testi di partenza, organizzata per nodi concettuali rappresentati da classi di eventi giuridici e ambientali.