

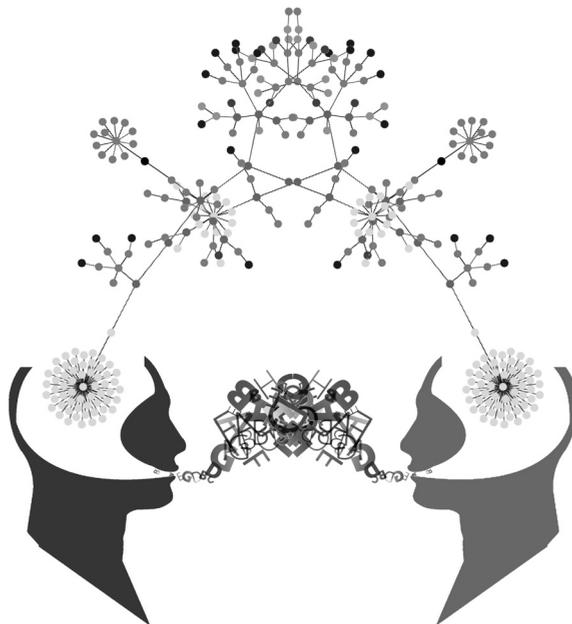


UNIVERSITÀ DEGLI STUDI DI SALERNO

Dottorato in Scienze della Comunicazione

Exploring Formal Models of Linguistic Data Structuring

Enhanced Solutions for Knowledge Management Systems
Based on NLP Applications



Federica Marano

Supervisor
Prof. Annibale Elia

Coordinator
Prof. Alessandro Laudanna

X Ciclo – Nuova Serie
2008-2011

ABSTRACT

Lo scopo principale di questa ricerca è descrivere in che modo i modelli formali per la strutturazione dei dati linguistici sono cruciali nelle applicazioni di Trattamento Automatico del Linguaggio (Natural Language Processing – NLP). Presteremo particolare attenzione ai Knowledge Management System (KMS) che sono progettati per Internet, e soprattutto alle soluzioni avanzate che essi possono richiedere. Per affrontare appropriatamente questi argomenti, descriveremo come ottenere applicazioni di linguistica computazionale utili agli esseri umani nello stabilire e gestire una relazione proficua con le tecnologie, in special modo con quelle che sono basate o che producono interazioni uomo-macchina in linguaggio naturale.

Esploreremo la relazione positiva tra Risorse Linguistiche ben formate e KMS allo scopo di stabilire che se l'architettura dell'informazione di un KMS è basata sulla formalizzazione dei dati linguistici, allora il sistema funziona meglio ed è più efficace.

Considerati gli argomenti da trattare, prima di tutto è indispensabile definire che per sviluppare strumenti di Information Retrieval (IR) efficienti ed efficaci, la prima operazione da compiere è capire e formalizzare i meccanismi combinatori del linguaggio naturale, soprattutto perché qualsiasi tipo di informazione prodotta dall'uomo su Internet è necessariamente un atto linguistico. Inoltre, in questo lavoro di ricerca tratteremo la strutturazione di un modello ibrido di formalizzazione linguistica basata sull'NLP che speriamo si riveli un utile strumento per supportare, migliorare e perfezionare i KMS.

Nello specifico, nella sezione 1 descriveremo come strutturare risorse linguistiche implementabili nei KMS, come esse possono migliorare le prestazioni di questi sistemi e come viene affrontato nei metodi di formalizzazione del linguaggio il problema della strutturazione dei dati linguistici. Nella sezione 2 con una breve introduzione alla linguistica computazionale si porrà particolare attenzione a software specifici come Intex, Unitex, NooJ e Cataloga, i quali sono sviluppati in base al metodo del Lessico-Grammatica (LG), una teoria linguistica fondata durante gli anni '60 da Maurice Gross.

Nella sezione 3 descriveremo alcuni lavori specifici per monitorare lo stato dell'arte nei modelli di strutturazione dei dati linguistici, nelle soluzioni avanzate per i KMS e nelle applicazioni di NLP per i KMS.

La sezione 4 è dedicata ai metodi di formalizzazione del linguaggio, descrivendo principalmente la Grammatica Generativo Trasformativa (GGT) e LG, più altri metodi basati su approcci statistici e ontologie.

Nella sezione 5 proporremo un modello ibrido per applicazioni di NLP allo scopo di creare efficaci soluzioni avanzate per i KMS. Attributi ed elementi specifici del nostro modello ibrido saranno mostrati attraverso alcuni risultati di lavori di ricerca sperimentali. Il caso di studio presentato è un problema di NLP molto complesso e ancora poco esplorato negli ultimi anni, quello del trattamento delle Multi Word Unit (MWU).

La sezione 6 chiude la ricerca valutandola con risultati e presentando possibili prospettive future.

Keywords

Knowledge Management System, Natural Language Processing, Linguistic Formal Model, Hybrid Formal Model.