

Domain Driven Technologies for Natural Language Processing

Alfio Massimiliano Gliozzo
ITC-irst

Abstract: Semantic Domains are a matter of recent interest in Computational Linguistics. Domain Models allow us to represent lexical ambiguity and variability by inducing soft clusters of terms from corpora. Semantic Domains shows many interesting properties: lexical ambiguity inside a domain is sensibly reduced, paradigmatic relations are mainly established among terms in the same domain, semantic domains are stable among languages. These properties have been exploited to develop innovative technologies for a wide range of different Natural Language processing tasks, such as Text Categorization, Word Sense Disambiguation, Cross Language Text Categorization and Ontology Learning from Texts. The Domain Driven systems we developed are minimally supervised and can be easily ported across domains and languages as they rely on semi supervised learning techniques based on Kernel Methods, reporting the state of the art performances in many cases.