



Automatic Annotation by Categories for Text Information Extraction

Textual information extraction, and particularly the extraction of information from web-based text, requires the annotation of a great number of documents very quickly, using standard categories: syntactical, grammatical (identifying tenses and aspects), lexical (the identification of transfer verbs, donation verbs, localization verbs) and communicative categories (identifying relations of definition, of causality, identifying quotations, comments, images). Present annotation procedures are either automatic and not based on standard categories, or, conversely, based on standard categories but not automatic. We need automatic procedures accompanied by appropriate linguistic resources making use of these annotation categories. Using these annotations of text elements, verbs, sentences, phrases, propositions, paragraphs, titles, definitions, causality, and so on, should simplify the process of extracting target information from text. In other words, we need models that marry, within the same computational model, a logical approach and an heuristic one with an eventual search of its cognitive relevance. The aim of this track is to provide an international forum for discussing the latest approaches in annotation, its relations to natural language processing, and to AI.