



Special Track:

Automatic Annotation and Information Retrieval

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. Moreover, some of the present annotation methods are based on statistical procedures namely word occurrences, collocation, distances between words. 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. The logical structures and logical organizations which are on the basis of annotation have to be studied as well.

These points of view are defended by two invited speakers: Jean-Pierre Desclés from the Sorbonne, who speaks about the relation between ontologies, semantic maps, and cognitive schemes; and Jean-Guy Meunier from the University of Québec in Montreal, who speaks about annotation and text analysis.

Papers presented in this track are focused on annotation systems based on linguistic procedures. EXCOM-MOCXE is a system building semantic metatext structure. This system, developed at the Sorbonne, is an indexation system based not on terms (as most of the present annotation systems) but on semantic relations. Two applications of this system are presented: the first one is an application to automatic summarization and the second is an application to a terminology extraction and construction of an ontology. Among linguistic tools for annotation, there are verb semantics used to annotate texts and categorial grammars. Segmentation as the preceding step to annotation is represented by a system based on categorial grammars to solve the problems of sentence coordination and a tagger for complex sentences applied to Czech. A system of building concept maps based on annotation constructed at University of Québec in Montreal and an annotation system based on machine learning to annotate events, developed in the University of Tunis are also presented in this special track.