

Preface

Computational syntax and computational semantics have been at the core of natural language processing technology since its inception in 1960s, and have always been considered a classical research field in computational linguistics. However, these fields continue to be actively developed. While computational syntax can be considered a well-understood area, computational semantics is still to find its way towards reconciling expressive power with technically affordable ways of text analysis.

Recent decades have seen a strong trend of moving away from rich text representations, in favor of simple, or even simplistic, modeling of texts as feature vectors. Such simplistic representations have two advantages: on the one hand, they are relatively simple to obtain from real-life texts with all their inconsistencies and grammatical peculiarities, and, on the other hand, they permit very complex and efficient processing algorithms, specifically, machine learning-based classifiers ranging from support vector machines to modern-day deep neural networks. Generalizing such machine-learning algorithms to text representations richer than feature vectors is non-trivial.

However, a disadvantage of purely statistical machine-learning methods based on feature-vector text representation is their inability to support complex reasoning, which has been long considered the basis of human language understanding and a necessary pre-requisite of any program aimed to flexible and reasonably complete natural language understanding.

It is thus on the agenda of natural language processing research now includes the search of ways of combining the advantages of the two worlds: the flexibility and efficiency of machine-learning methods and the power of logical reasoning over structured text representations. I see two ways to do it: first, extracting more complex features, such as recently introduced syntactic n-grams, from structured semantic representations, and feeding these features into machine-learning algorithms; secondly, and more interestingly, using machine-learning algorithms for decision-making in graph- and network-based logical reasoning algorithms.

This justifies a closer look at syntax and semantics of natural language, as well as on their applications in various tasks of natural language processing, such as question answering and machine translation.

An important and still unresolved issue in natural language processing technologies is the addressing of languages outside the major European languages. This includes attention to languages typologically quite different from Germanic and Romance languages, which have received most of attention in computational linguistics. This also includes looking at European languages from the perspective of a grammar tradition outside of the Greek-Latin mainstream. Accordingly, I selected for this issue some papers devoted to Arabic and two Turkic languages, as well as one paper devoted to the description of English grammar within a Sanskrit-based linguistic tradition.

The papers included in this special issue were selected basing on a double-blind review procedure, with participation of 126 leading experts in the topic from 38

countries. A list of international reviewers that participated in the selection process is given at the end of this volume.

This special issue will be useful to researchers, students, and engineers working in natural language processing and computational linguistics, especially in the subfields of computational syntax and computational semantics as well as their practical applications.

November 2016

Alexander Gelbukh,
Guest Editor