

Preface

(Prefacio)

This volume of “Research in Computing Science” presents a selection of papers on computational linguistics by authors from 14 countries: Canada, China, Egypt, Finland, France (6), Germany (2), India (2), Japan, Mexico (3), New Zealand, Romania, USA, Spain, and Vietnam. The papers have been carefully chosen based on reviews by the members of the international reviewing committee of this volume.

Computational linguistics, also referred to as natural language processing or human language technologies, is a branch of science in the intersection of linguistics and artificial intelligence that has two main objects of study. On the one hand, it studies the structure of human languages using computational tools, which give to the linguist a lot more power than traditional paper-and-pencil techniques. Computers are capable of processing huge amounts of text to suggest to the linguist hypotheses about possible regularities in the language or to test the linguist’s own hypotheses about the structure and statistical properties of the human language. In addition, computers can easily adapt to different genres and thematic domains of texts.

On the other hand, natural language processing has the goal of enabling computers to accomplish meaningful tasks connected with text and speech by using linguistic knowledge, both encoded by human linguists in the form of grammars and dictionaries and learnt automatically. Modern machine learning methods are capable of automatically learning linguistic regularities from huge text corpora and using this information in order to accomplish the required natural language processing tasks.

In this volume, twenty papers present a wide range of problems and solutions in the area of computational linguistics and natural language processing. The first paper uses computational methods to explore an ancient writing system. The next papers evaluate word segmentation techniques in Chinese, propose treatment of named entities, suggest new approaches to tagging in languages of India, to measuring semantic relatedness, to word sense disambiguation, to detecting person’s names in Arabic, to solving ambiguities in discourse parsing and parsing of intentions, and to recognizing textual entailment.

The next group of papers is devoted to various aspects of lexicography, such as extracting parallel sentences from Wikipedia, improving biomedical term extraction, exploring various types of approaches to multilingual terminology extraction, and learning for relation extraction. Finally, several papers are devoted to various aspects of analysis of text, such as mining for suggestions and outdated information, analyzing changes in quotes over time, genre identification, building a collocation dictionary, and evaluation of existing approaches to text summarization.

Alexander Gelbukh

November 2013