Editorial

SEMANTIC WEB and intelligent text processing technologies involved in it are crucial for today's information infrastructure. Brilliant success of information retrieval and machine translation, as well as rapid progress in opinion mining and sentiment analysis for decision making and recommender systems, make these topics, as well as their underlying technology and research, especially important for the readers of this journal. It is my pleasure to present to the readers an issue of Polibits featuring a thematic section on Semantic Web and Intelligent Text Processing.

The thematic section contains nine papers written by authors from seven countries: France, India, Mexico, Romania, Spain, Switzerland, and USA. In addition, the issue contains a regular paper.

The first two papers are directly connected with treatment of information in webpages.

López, Silva, and Insa (Spain) address a very important problem of automatic separation of useful text on a webpage from noise, mainly advertisement, which nowadays floods nearly any webpage. They describe a simple yet powerful idea of using the HTML tree structure to identify large blocks of text, which very probably represent useful text and not advertisements and other noise.

Schilder, Kondadadi, and Kadiyska (USA) explain how to extract tabular data from pieces of text that look like tables for humans but are not structured well enough to allow trivial technique for recognition of the structure. They use geometric approach based on spatial proximity of the pieces of text representing table cells in two dimensions, starting from a corner of the table and iteratively pulling the cells adjacent to those already identified.

The next two papers deal with string alignment problems: when you have two different strings and you want to identify what they have in common.

Dănăilă, Dinu, Niculae, and Șulea (Romania) present a detailed survey of a number of different string comparison measures and study their performance in an important task: identifying different expressions that refer to the same thing, such as different spellings of the name of the same person or product, or different ways to express the same address. This is an important task because the presence of huge amount of such duplicates in large databases prevents us from correct analysis and handling of those databases.

Nicolas Béchet and Marc Csernel (France) use string alignment technique for comparison of different versions of the same, or nearly the same, text in Sanskrit. A particular difficulty of comparing Sanskrit documents is that text in Sanskrit is written without spaces between words, and some parts of text can be freely moved without changing the meaning of the text. I believe their technique will be interested not only for historians and philologists but also for those who deal with genetic sequences: DNA structures exhibit similar properties.

The next two papers deal with very important extralinguistic phenomena that are present in huge quantities in the (semantic) Web: emotions and references to locations.

Loza-Pacheco, Torres-Ruiz and Guzmán-Lugo (Mexico) identify the location to which a map, a photo, or a toponym (name of a place) belongs. They use knowledge-based techniques and ontologies to reason about spatial relationships between objects and thus their names and thus to bring order and meaning in geographically-related databases. While the paper is written in Spanish, it provides a English abstract.

Das and Bandyopadhyay (India) extend analysis of emotions expressed by the authors of blog messages to a new language, in this case Bengali, which is, according to different accounts, the fourth or fifth world's most spoken language, accounting for approximately 200 million speakers. Analysis of emotions and sentiments expressed in blogs and social networks is currently probably the hottest topic in natural language processing and web-related studies.

The next three papers deal with semantics of natural language.

Martins (Switzerland) discusses issues related to an extension of the Universal Networking Language (UNL, see www.cicling.org/2005/UNL-book for an introduction) to a knowledge representation language called XUNL, and draws some important conclusions and guidelines about the desired structure and properties of such semantic language.

Castro-Sánchez and Sidorov (Mexico) present a novel method for building formal syntactico-semantic structures that describe the roles of nouns that they play in a situation expressed by a verb, and how these constructions are expressed in natural language texts. They build such a formal lexical resource out of existing dictionaries oriented to human readers, which do not allow direct use of the information they contain by computer programs.

Dinu (Romania) closes the thematic section on Semantic Web and Intelligent Text Processing with a paper devoted to particular issues in formal semantics, such as scope taking and quantification, expressed in precise mathematical form. These issues, that have been in the core of formal semantics research during decades if not centuries, are discussed in the paper in the context of a specific semantic theory called continuation semantics.

Apart from the thematic section, this issue of Polibits includes a regular paper unrelated to the thematic section but directly related with the topic of the journal.

Jiménez, Sossa, Cuevas, and Gómez apply well-known artificial intelligence technique called particle swarm optimization to an important practical technical problem: interferometry, which is a task of non-destructive optical measuring of dimensions of a physical object with very high precision comparable with the wavelength of light. They introduce the reader to the practical task with a clear explanation of the problem, and then explain how the artificial intelligence technique is used to solve this practical problem.

I would like to thank the Editorial Board of Polibits for inviting me to serve as a guest editor of the journal and express my hope that the papers selected for this issue will prove to be interesting and useful for all readers working in, or interested in, Computer Science in general, Artificial Intelligence, and specifically Text Processing and Semantic Web.

Dr. Niladri Chatterjee

Associate Professor,
Indian Institute of Technology Delhi,
New Delhi, India

Guest Editor