Foreword

The term "Named Entity", now widely used in Natural Language Processing, was coined for the Sixth Message Understanding Conference (MUC–6). It initially included names of persons, locations and organizations, and numeric expressions including time, date, money and percentage expressions. Identifying references to these entities in text was recognized as one of the important sub-tasks of IE and was called "Named entity recognition and classification (NERC)".

Named Entities provide critical information for many NLP applications. For example, consider the following two sentences:

(1) *A woman met with a man to discuss an issue.*
(2) *Condoleezza Rice met Sunday with President Mahmoud Abbas to discuss reviving the Israeli-Palestinian peace process.*

Obviously, (1) is only a general description of an event, which can occur anywhere, but (2) conveys significant information people are interested in. NERC (Named Entity Recognition and Classification) is a key technology for constructing NLP systems that can recognize the value of such information.

The seven papers in this special issue cover various interesting and informative aspects of NERC research. The first paper, by David Nadeau and Satoshi Sekine, is an extensive survey of past NERC technologies, which should be a very useful resource for new researchers in this field. The second paper, by Andrew Smith and Miles Osborne, describes a machine learning model which tries to solve the over-fitting problem. The third paper, by Pawel Mazur and Robert Dale, tackles a common problem of NE and conjunction. As conjunctions are often a part of NEs or appear close to NEs, this is an important practical problem. Then we have three papers about analysis and implementation of NERC for different languages, namely for Spanish by Sofia N. Galicia-Haro and Alexander Gelbukh, for Bengali by Asif Ekbal, Sudip Kumar Naskar and Sivaji Bandyopadhyay, and for Serbian by Dusko Vitas, Cvetana Krstev and Denis Maurel. These papers provide insight into NERC problems regarding interesting aspects of different languages. The last paper, by Ralf Steinberger and Bruno Pouliquen, reports on a real WEB application where NERC technology is one of the central components. It needs multilingual NERC in order to identify the occurrences of people, locations and organizations in newspapers in different languages.

Finally, we would like to thank all the contributors who made this special issue possible. We received many more submissions than we expected. Many good papers could not be included, because of the page limitation of the journal. We are very sorry about that. The guest editors appreciate all the scientific committee members who provided reviews, and also thank the journal for allowing us to publish this interesting special issue.

Scientific Committee Members:

Masayuki Asahara, Roberto Basili, Eckhard Bick, Samuel Eleuterio, Robert Gaizauskas, Jun Goto, Ralph Grishman, Kentaro Inui, Dimitrios Kokkinakis, Nicolas Nicolov, Thierry Poibeau, Elisabete Ranchhod, Dan Roth, Satoshi Sekine, Mário J. Silva, Ralf Steinberger, Kentaro Torisawa, Christa Womser-Hacker

The guest editors:

Satoshi Sekine   Elisabete Ranchhod
New York University   University of Lisbon
sekine@cs.nyu.edu   e_ranchhod@fl.ul.pt

April 2007