Editorial

Jun-ichi Aoe

Department of Information Science and Intelligent Systems,
University of Tokushima,
Minami-Josanjima-Cho Tokushima-shi Tokushima, 770-8506, Japan
E-mail: aoe@is.tokushima-u.ac.jp

Biographical notes: Jun-ichi Aoe received his BSc and MSc in Electronic Engineering from the University of Tokushima, Japan, in 1974 and 1976, respectively, and PhD in Communication Engineering from the University of Osaka, Japan. Since 1976, he has been with the University of Tokushima, where he is now a Professor in the Department of Information Science and Intelligent Systems. His research interests include design of an automatic selection method of key search algorithms based on expert knowledge bases and natural language processing. He was the Editor of *Computer Algorithm Series* of the IEEE Computer Society Press.

1 Introduction

The special issue promotes exchange of opinions between experts working in different areas of the growing field of computational linguistics and intelligent text processing. The main purpose of the special issue is to bring together scientists representing linguistics, computer science and related fields, and sharing a common interest related to the advancement of computational linguistics and natural language processing. This issue covers a broad spectrum of disciplines working towards enabling intelligent system to interact with humans using natural language.

The research and development of these systems, that exploit knowledge in the target domain, is at the forefront of modern researches.

This special issue is intended to present applications of intelligent text processing. Submitted papers are expected to postulate diverse problems, models and solutions for these applications.

2 Papers in this issue

The first paper in this issue, 'Partial retrieval of compressed semi-structured documents', by Ashutosh Gupta and Suneeta Agarwal, describes a compression model called tri-structural contexts model (TSCM), for semi-structured documents. This model is based with the hope that the attribute name/attribute value and textual words may reduce the entropy by separating the start tag. They also combine the attributes with their values and use a separate container for them. They mainly focus on semi-static models, and test the idea using a word-based tagged code.

The second paper, 'Clarification of the price fluctuation mechanism in financial markets: disparity in forecast accuracy among investors and asset price fluctuations', by Hiroshi Takahashi, analyses the impact of the disparity in forecast accuracy among investors in financial markets. This analysis has found the following interesting phenomena:

1 Depending on market environments, investors with good forecast accuracy do not always survive in the market.

2 Where the performance measurement period is short, there may be a negative influence such as deviation from the fundamental value of trading prices.

The third paper, 'SSGL: a semi supervised grammar learner', by K. Sundarakantham, N. Sheena and S. Mercy Shalinie presents a comprehensive solution for efficient language acquisition by a novel semi-supervised algorithm that learns a streamlined representation of linguistic structures from a plain natural-language corpus. The input datasets are ATIS dataset and sentences from children's literature.

The fourth paper, 'Persian POS tagging using probabilistic morphological analysis', by Hakimeh Fadaei and Mehrnoush Shamsfard, presents a POS tagger for Persian. It exploits a hybrid approach which is a combination of statistical and rule-based methods to tag Persian sentences. The proposed tagger uses a novel probabilistic morphological analysis to tag unknown words. As a secondary result of this research, a knowledge base of Persian morphological rules with their probabilities is built according to a corpus.

The fifth one, 'Analysis of greenhouse gases trading system using conversations among stakeholders', by Setsuya Kurahashi and Masato Ohori, aims to reduce the free riders which is most important thing for success of greenhouse gas treading system in current by using the following conditions:

1 to derive an emergence mechanism of the free riders

2 to consider the reductive mechanism of the free riders.

In order to examine a relationship among some governments, some companies, and some consumers under the social dilemma of world warming, this paper applies the

agent-based approach using interaction among them as conversations.

The sixth paper, 'Local features based script recognition from printed bilingual document images', by S. Abirami and D. Manjula, presents architecture for script recognition of bilingual document images (Tamil, English), which specifically takes the challenges of recognition at character level by predicting the script of word image using its initial character, thereby adapting to various font faces and sizes. This recogniser models every character as *Tetra bit values* (TBV), which corresponds to the spatial spread, derived from the segmented grids of the character. They employed a *decision tree classifier* (DTC) for the classification of script on over the patterns generated from TBV. *Spatial features based script recogniser* (SFBSR) is trained and tested with bilingual document images, consisting of various Tamil and English words, to show its effectiveness towards script identification. Classification accuracy in training and testing sets are promising.

The seventh paper, 'Building of field association terms based on links', by Mahmoud Rokaya and El-Sayed Atlam, proposes studying co-word relations between field association terms (FA terms), and candidate terms provide valuable information which helps the machine to take the correct decision and to append the candidate terms in their suitable places inside FA terms dictionary. In this paper, a pure automatic tool to build a dynamic FA terms dictionary. The concept of link power as well as modifications of the rules used to determine the FA terms level depending on the concept of links has been presented.

The eighth paper, 'Multi-page document analysis based on format consistency and clustering', by Liangcai Gao, Zhi Tang, Xiaofan Lin and Jing Fang, presents a new document analysis method based on DCIFC, which is complementary to the traditional document analysis methods based on the visual characteristics of document elements. One key advantage of their method is that DCIFC is stable from a document to a document, and this is not impacted by layout variability, which is a major challenge in document analysis. Their method uses clustering techniques to build statistical models and then applies the models to labelling document components.

The ninth paper, 'Effective extraction method of the mutual understanding gap based on the egocentrism', by Nobuo Suzuki and Kazuhiko Tsuda, defines the intensity for properties of the egocentrism with the method that presumed the egocentrism from language expressions in their previous study and presumes mutual understanding gap in the text data of Q&A sites that have dialogue form by using the intensity and calculating the score of the egocentrism in each speech. This paper shows the characteristics of the gap are expressed by the shape of graphs with these scores. We also evaluate this method with text data in actual Q&A sites and confirm its effectiveness.

The tenth paper, 'An English and/or Japanese writing support tool based on a web search engine', by Kazuaki Ando, Yuichi Tsunashima, and Youhei Furukawa, proposes a writing support tool for learners of English and/or Japanese as a second language based on a web search engine. The proposed tool can support language learners writing a sentence in English and/or Japanese. The proposed tool consists of a composition and a learning support module. The composition support module searches for example sentences and word usage on the web, and verifies given sentences and expressions by the hit count and the statistical analysis on snippets provided by the search engine. It produces a table and a graph based on the hit count of those queries, and it provides example sentences based on learner's vocabulary level.

The last paper, 'A schema for ontology-based concept definition and identification', by Zhan Li and Marek Reformat, proposes a novel adaptive assignment of a term importance (AATI). This schema is an ontology-based approach for defining and identifying concepts. It includes definitions of relations between terms and concepts, and an iterative algorithm for determining importance of terms. AATI continuously updates importance of terms with 'unknown' web documents, which makes it appropriates for web applications.