

---

## Preface

---

### 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 BSc and MSc Degrees in Department of Electronic Engineering from University of Tokushima, Japan, in 1974 and 1976, respectively, and a PhD Degree in Communication Engineering from University of Osaka, Japan. He is currently a Professor in Department of Information Science and Intelligent Systems, University of Tokushima, Japan. His research interests include design of an automatic selection method of key search algorithms based on expert knowledge bases, natural language processing.

---

## 1 Introduction

The special issue promotes exchange of opinions among experts working in different areas of the growing fields 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 systems 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 research.

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, 'An intelligent tool for syntactic annotation of Arabic corpora', by Chiraz Ben Othmane Zribi, Ferial Ben Fraj and Mohamed Ben Ahmed, proposes a new technique for semi-automatic syntactic annotation of Arabic texts. This paper describes a tool that takes a morpho-syntactic tagged corpus as an input and provides its syntactic annotation according to the Arab TAG formalism as output. In this study, the term 'intelligent' is used because this tool automatically learns and improves during elementary annotation (super-tagging) by applying a supervised classification method combining three classifiers (naïve Bayes, K-nearest neighbours, decision tree).

The second paper, 'Discovering Arabic structures from texts: what a formal analysis can tell us', by Hassina Aliane and Zaia Alimazighi, presents a new approach to discover

Arabic language structures from electronic texts. The method is based on a distributional analysis inspired from Arabic Grammar Tradition (AGT) and uses minimum knowledge about the Arabic language. The idea underlying this research is that people can learn about the structures of this language only by a formal analysis of raw corpora written in it, in the absence of a formal model of Arabic language and freely usable NLP tools for this language.

The third paper, 'Multi-document summarisation using genetic algorithm-based sentence extraction', by A. Kogilavani and P. Balasubramanie, proposes a system to produce an optimal summary by genetic algorithm-based sentence extraction strategies. Appropriate sentences are selected based on individual word weight as well as sentence-specific features. To produce the optimal summary, fitness function is employed which is based on summary quality criteria, such as maximising length, coverage and informativeness while minimising the redundancy.

The fourth paper, 'Optimising of support plans for new graduate employment market using reinforcement learning', by Keiko Mori and Setsuya Kurahashi, focuses on the problem of the new graduate market in Japan, where there is the gap of required level between the companies and the students. The paper proposes a multi-agented simulation system with reinforcement learning. Two types of reinforcement learning are adopted: the first is profit sharing and the second is actor-critic. The two adopted types are effective in searching for the best support plan for the employment market.

The fifth paper, 'A fast search method of similar strings from dictionaries', by Masao Fuketa, El-Sayed Atlam, Nobuo Fujisawa, Hiroshi Hanafusa, Kazuhiro Morita and Jun-ichi Aoe, proposes a structure of a dictionary for retrieving similar strings and a method to produce the edit distance of strings in descending order at high speed. This method can retrieve all similar strings and the speed of this method is faster than n-gram methods.

The last paper, 'An e-mail filtering method based on multi-attribute values of user's profile', by Masami Shishibori, Samuel Sangkon Lee and Kenji Kita, constructs the personal profile from existing email documents that have been already received by each user. This profile

consists of multi-attribute values such as 'Sender', 'Theme' and 'Type' of each document. By using the profile, this method can distinguish whether a new email document is important or not for the user.