

# Lecture Notes in Artificial Intelligence 7614

## Subseries of Lecture Notes in Computer Science

### LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

### LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

Hitoshi Isahara Kyoko Kanzaki (Eds.)

# Advances in Natural Language Processing

8th International Conference on NLP, JapTAL 2012  
Kanazawa, Japan, October 22-24, 2012  
Proceedings



Springer

## Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany  
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

## Volume Editors

Hitoshi Isahara  
Toyohashi University of Technology  
Information and Media Center  
1-1 Hibarigaoka, Tenpakucho  
Toyohashi 441-8580, Japan  
E-mail: isahara@tut.jp

Kyoko Kanzaki  
Toyohashi University of Technology  
Information and Media Center  
1-1 Hibarigaoka, Tenpakucho  
Toyohashi 441-8580, Japan  
E-mail: kyoko.kanzaki@gmail.com

ISSN 0302-9743 e-ISSN 1611-3349  
ISBN 978-3-642-33982-0 e-ISBN 978-3-642-33983-7  
DOI 10.1007/978-3-642-33983-7  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012948107

CR Subject Classification (1998): I.2, H.3, H.4, H.5, H.2, J.1, I.5

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

The 8th International Conference on Natural Language Processing (JapTAL 2012) took place in the city of Kanazawa, a leading tourist destination where 7 million tourists visit every year. It was a great castle town ruled by an influential leader from the seventeenth century to the second half of the nineteenth century. Kanazawa has not suffered from any war devastation or big natural disaster up to now. Therefore, it maintains rows of historical houses and inherits traditional handicrafts and traditional performing arts.

JapTAL was the eighth in the series of the TAL conferences, following IceTAL 2010 (Reykjavik, Iceland), GoTAL 2008 (Gothenburg, Sweden), FinTAL 2006 (Turku, Finland), EsTAL 2004 (Alicante, Spain), PorTAL 2002 (Faro, Portugal), VexTAL 1999 (Venice, Italy), and FracTAL 1997 (Besancon, France). This was the first time that the TAL conference was held outside Europe.

The main purpose of the TAL conference series is to bring together scientists representing linguistics, computer science, and related fields, sharing a common interest in the advancement of computational linguistics and natural language processing.

The conference consists of invited talks, oral and poster presentations, and special sessions on the applications and theory of natural language processing and related areas. It provides excellent opportunities for the presentation of interesting new research results and discussions about them, leading to knowledge transfer and the generation of new ideas.

In the reviewing process of the main conference track, we received 42 submissions. Among them, we selected 27 submissions as long papers and five submissions as short papers. Therefore, the acceptance ratio for the long papers is 64% and the total acceptance ratio including short papers is 76%.

We had two special sessions, i.e., “Game and NLP” and “Student/Young Researcher Session.” Papers submitted for these sessions were treated separately and accepted papers were presented in each session but they are not included in this volume.

As conference organizers of JapTAL 2012, we would like to thank all PC members who reviewed submissions very tight schedule and all local staff who helped us during the preparation of JapTAL. We would like to thank the International Exchange Program of the National Institute of Information and Communications Technology (NICT) for its support of JapTAL 2012.

Hitoshi Isahara  
Kyoko Kanzaki

# Organization

## Program Chairs

Hitoshi Isahara  
Kyoko Kanzaki

## Program Committee

Takako Aikawa	Microsoft Research, USA
Johan Bos	University of Groningen, The Netherlands
Pierrette Bouillon	Geneva University, Switzerland
Caroline Brun	Xerox Corporation, France
Sylviane Cardey	University of Franche-Comté, France
Key-Sun Choi	KAIST, Korea
Christiane Fellbaum	Princeton University, USA
Filip Ginter	University of Turku, Finland
Peter Greenfield	University of Franche-Comté, France
Philippe de Groote	INRIA, France
Yurie Iribe	Toyohashi University of Technology, Japan
Katsunori Kotani	Kansai Gaidai University, Japan
Krister Lindén	University of Helsinki, Finland
Hrafn Loftsson	Reykjavik University, Iceland
Qing Ma	Ryukoku University, Japan
Bente Maegaard	University of Copenhagen, Denmark
Mathieu Morey	Nanyang Technological University, Singapore
Masayuki Okabe	Toyohashi University of Technology, Japan
Guy Perrier	INRIA, France
Kiyooki Shirai	JAIST, Japan
Virach Sornlertlamvanich	NECTEC, Thailand
Koichi Takeuchi	Okayama University, Japan
Midori Tatsumi	Toyohashi University of Technology, Japan
Izabella Thomas	University of Franche-Comté, France
Noriko Tomuro	DePaul University, USA
Masatoshi Tsuchiya	Toyohashi University of Technology, Japan
Jose Luis Vicedo	University of Alicante, Spain
Simo Vihjanen	Lingsoft Ltd., Finland
Xiaohong Wu	Qinghai University for Nationalities, China
Kazuhide Yamamoto	Nagaoka University of Technology, Japan
Yujie Zhang	Beijing Jiaotong University, China
Tiejun Zhao	Harbin Institute of Technology, China

# Table of Contents

## Machine Translation

The Impact of Crowdsourcing Post-editing with the Collaborative Translation Framework . . . . .	1
<i>Takako Aikawa, Kentaro Yamamoto, and Hitoshi Isahara</i>	
Translation of Quantifiers in Japanese-Chinese Machine Translation . . . . .	11
<i>Shaoyu Chen and Tadahiro Matsumoto</i>	
Toward Practical Use of Machine Translation . . . . .	23
<i>Hitoshi Isahara</i>	
Phrase-Level Pattern-Based Machine Translation Based on Analogical Mapping Method . . . . .	28
<i>Jun Sakata, Masato Tokuhisa, and Jin'ichi Murakami</i>	

## Multilingual Issues

Parallel Texts Extraction from Multimodal Comparable Corpora . . . . .	40
<i>Haithem Afli, Loïc Barrault, and Holger Schwenk</i>	
A Reliable Communication System to Maximize the Communication Quality . . . . .	52
<i>Gan Jin and Natallia Khatseyeva</i>	
DANIEL: Language Independent Character-Based News Surveillance . . . . .	64
<i>Gaël Lejeune, Romain Brixtel, Antoine Doucet, and Nadine Lucas</i>	
OOV Term Translation, Context Information and Definition Extraction Based on OOV Term Type Prediction . . . . .	76
<i>Jian Qu, Akira Shimazu, and Le Minh Nguyen</i>	
Exploiting a Web-Based Encyclopedia as a Knowledge Base for the Extraction of Multilingual Terminology . . . . .	88
<i>Fatima Sadat</i>	
Segmenting Long Sentence Pairs to Improve Word Alignment in English-Hindi Parallel Corpora . . . . .	97
<i>Jyoti Srivastava and Sudip Sanyal</i>	
Shallow Syntactic Preprocessing for Statistical Machine Translation . . . . .	108
<i>Hoai-Thu Vuong, Dao Ngoc Tu, Minh Le Nguyen, and Vinh Van Nguyen</i>	

**Resources**

Linguistic Rules Based Approach for Automatic Restoration of Accents on French Texts .....	118
<i>Paul Brillant Feuto Njonko, Sylviane Cardey-Greenfield, and Peter Greenfield</i>	
Word Clustering for Persian Statistical Parsing .....	126
<i>Masood Ghayoomi</i>	
Building a Lexically and Semantically-Rich Resource for Paraphrase Processing .....	138
<i>Wannachai Kampeera and Sylviane Cardey-Greenfield</i>	
Tagset Conversion with Decision Trees .....	144
<i>Bartosz Zaborowski and Adam Przepiórkowski</i>	
Fitting a Round Peg in a Square Hole: Japanese Resource Grammar in GF .....	156
<i>Elizaveta Zimina</i>	

**Semantic Analysis**

Arabic Language Analyzer with Lemma Extraction and Rich Tagset ....	168
<i>Ahmed H. Aliwy</i>	
Tracking Researcher Mobility on the Web Using Snippet Semantic Analysis .....	180
<i>Jorge J. García Flores, Pierre Zweigenbaum, Zhao Yue, and William Turner</i>	
Semantic Role Labelling without Deep Syntactic Parsing .....	192
<i>Konrad Gołuchowski and Adam Przepiórkowski</i>	
Temporal Information Extraction with Cross-Language Projected Data .....	198
<i>Przemysław Jarzębowski and Adam Przepiórkowski</i>	
Word Sense Disambiguation Based on Example Sentences in Dictionary and Automatically Acquired from Parallel Corpus .....	210
<i>Pulkit Kathuria and Kiyooki Shirai</i>	
A Study on Hierarchical Table of Indexes for Multi-documents .....	222
<i>Tho Thi Ngoc Le, Minh Le Nguyen, and Akira Shimazu</i>	
Finding Good Initial Cluster Center by Using Maximum Average Distance .....	228
<i>Samuel Sangkon Lee and Chia Y. Han</i>	

Applying a Burst Model to Detect Bursty Topics in a Topic Model . . . . .	239
<i>Yusuke Takahashi, Takehito Utsuro, Masaharu Yoshioka, Noriko Kando, Tomohiro Fukuhara, Hiroshi Nakagawa, and Yoji Kiyota</i>	

UDRST: A Novel System for Unlabeled Discourse Parsing in the RST Framework . . . . .	250
<i>Ngo Xuan Bach, Nguyen Le Minh, and Akira Shimazu</i>	

## Sentiment Analysis

Long-Term Goal Discovery in the Twitter Posts through the Word-Pair LDA Model . . . . .	262
<i>Dandan Zhu, Yusuke Fukazawa, Eleftherios Karapetsas, and Jun Ota</i>	

Finding Social Relationships by Extracting Polite Language in Micro-blog Exchanges . . . . .	268
<i>Norinobu Hatamoto, Yoshiaki Kurosawa, Shogo Hamada, Kazuya Mera, and Toshiyuki Takezawa</i>	

Twitter Sentiment Analysis Based on Writing Style . . . . .	278
<i>Hiroshi Maeda, Kazutaka Shimada, and Tsutomu Endo</i>	

Extraction of User Opinions by Adjective-Context Co-clustering for Game Review Texts . . . . .	289
<i>Kevin Raison, Noriko Tomuro, Steve Lytinen, and Jose P. Zagal</i>	

## Speech and Generation

Automatic Phone Alignment: A Comparison between Speaker- Independent Models and Models Trained on the Corpus to Align . . . . .	300
<i>Sandrine Brognaux, Sophie Roekhaut, Thomas Drugman, and Richard Beaufort</i>	

A Story Generation System Based on Propp Theory: As a Mechanism in an Integrated Narrative Generation System . . . . .	312
<i>Shohei Imabuchi and Takashi Ogata</i>	

Automatic Utterance Generation by Keeping Track of the Conversation's Focus within the Utterance Window . . . . .	322
<i>Yusuke Nishio and Dongli Han</i>	

<b>Author Index . . . . .</b>	<b>333</b>
-------------------------------	------------