

Lecture Notes in Artificial Intelligence 4938

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Takenobu Tokunaga Antonio Ortega (Eds.)

# Large-Scale Knowledge Resources

Construction and Application

Third International Conference  
on Large-Scale Knowledge Resources, LKR 2008  
Tokyo, Japan, March 3-5, 2008  
Proceedings



Springer

## Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

## Volume Editors

Takenobu Tokunaga  
Tokyo Institute of Technology  
Department of Computer Science  
Tokyo Meguro Oookayama 2-12-1  
152-8552 Japan  
E-mail: take@cl.cs.titech.ac.jp

Antonio Ortega  
University of Southern California  
Signal and Image Processing Institute  
Department of Electrical Engineering  
Los Angeles, CA 90089-2564, USA  
E-mail: antonio.ortega@sipi.usc.edu

Library of Congress Control Number: 2008920448

CR Subject Classification (1998): I.2.7, I.2, I.4, I.5, H.3, H.5

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743  
ISBN-10 3-540-78158-7 Springer Berlin Heidelberg New York  
ISBN-13 978-3-540-78158-5 Springer Berlin Heidelberg New York

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.

Springer is a part of Springer Science+Business Media  
springer.com

© Springer-Verlag Berlin Heidelberg 2008  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper SPIN: 12227532 06/3180 5 4 3 2 1 0

# Preface

At the start of the 21st century, we are now well on the way towards a knowledge-intensive society, in which knowledge plays ever more important roles. Thus, research interest should inevitably shift from information to knowledge, with the problems of building, organizing, maintaining and utilizing knowledge becoming central issues in a wide variety of fields. The 21st Century COE program “Framework for Systematization and Application of Large-scale Knowledge Resources (COE-LKR)” conducted by the Tokyo Institute of Technology is one of several early attempts worldwide to address these important issues. Inspired by this project, LKR2008 aimed at bringing together diverse contributions in cognitive science, computer science, education and linguistics to explore design, construction, extension, maintenance, validation and application of knowledge.

Responding to our call for papers, we received 38 submission from a variety of research areas. Each paper was reviewed by three Program Committee members. Since we were aiming at an interdisciplinary conference covering a wide range of topics concerning large-scale knowledge resources (LKR), each paper was assigned a reviewer from a topic area outside the main thrust of the paper. This reviewer was asked to assess whether the authors described the motivation and importance of their work in a comprehensible manner even for readers in other research areas. Following a rigorous reviewing process, we accepted 14 regular papers and 12 poster papers.

While the technical program covered a broad range of application areas, papers can be roughly classified into four major areas. First, some authors reported on their experience building large-scale resources, with a particular emphasis on language resources and ontologies. Second, a major focus was in the extraction of knowledge from large sets of data, ranging from mining of text data to semantic analysis of content such as images and video. Third, some papers reported on experiences with operating actual LKR systems, including question and answer systems, inductive reasoning systems, etc. Finally, several papers focused on infrastructure problems to enable efficient LKR systems to be deployed.

We would like to thank the Program Committee members for their hard work. In addition to these technical papers, this volume includes six papers by the invited speakers. We hope the conference and this volume will contribute to opening up a new interdisciplinary research area on LKR.

Antonio Ortega  
Tokunaga Takenobu

# Organization

Conference Chair	Furui, Sadaoki (Tokyo Institute of Technology)
Local Organizing Chair	Shinoda, Koichi (Tokyo Institute of Technology)
Program Co-chairs	Ortega, Antonio (University of Southern California) Tokunaga, Takenobu (Tokyo Institute of Technology)
Publicity Co-chairs	Yokota, Haruo (Tokyo Institute of Technology) Yonezaki, Naoki (Tokyo Institute of Technology)

## Program Committee

Calzolari, Nicoletta (ILC-CNR, Italy)  
Harman, Donna (NIST, USA)  
Huang, Chu-Ren (Academia Sinica, Taiwan)  
Ishizaki, Syun (Keio University, Japan)  
Juang, Bing-Hwang (Georgia Institute of Technology, USA)  
Kusumi, Takashi (Kyoto University, Japan)  
Matsumoto, Yuji (Nara Institute of Science and Technology, Japan)  
Nakazawa, Jin (Keio University, Japan)  
Nilsson, Jorgen Fischer (Technical University of Denmark, Denmark)  
Rigoll, Gerhard (Munich University of Technology, Germany)  
Russell, Stuart (University of California, Berkeley, USA)  
Sakamoto, Takashi (JAPET, Japan)  
Sekine, Satoshi (New York University, USA)  
Shibano, Koji (Tokyo University of Foreign Studies, Japan)  
Sugiura, Masatoshi (Nagoya University, Japan)  
Sugiyama, Shinya (Keio University, Japan)  
Theeramaunkong, Thanuruk  
(Sirindhorn International Institute of Technology, Thailand)  
Thompson, Henry (University of Edinburgh, UK)  
Tobe, Yoshito (Tokyo Denki University, Japan)  
Tomimori, Nobuo (Tokyo University of Foreign Studies, Japan)  
Washio, Takashi (Osaka University, Japan)  
Yamada, Tuneo (National Institute of Multimedia Center, Japan)  
Yano, Tamaki (Doshisha University, Japan)

# Table of Contents

## Keynote Talk

From Information to Intelligence: The Role of Relative Significance in Decision Making and Inference.....	1
<i>Biing-Hwang Juang</i>	

## Mining Knowledge

Comparing LDA with pLSI as a Dimensionality Reduction Method in Document Clustering .....	13
<i>Tomonari Masada, Senya Kiyasu, and Sueharu Miyahara</i>	
Identification of MCMC Samples for Clustering .....	27
<i>Kenichi Kurihara, Tsuyoshi Murata, and Taisuke Sato</i>	
TGC-Tree: An Online Algorithm Tracing Closed Itemset and Transaction Set Simultaneously .....	38
<i>Junbo Chen and Bo Zhou</i>	
Extracting Concepts from Religious Knowledge Resources and Constructing Classic Analysis Systems .....	51
<i>Hajime Murai and Akifumi Tokosumi</i>	
What Types of Translations Hide in Wikipedia? .....	59
<i>Jonas Sjöbergh, Olof Sjöbergh, and Kenji Araki</i>	
Initial Solution Set Improvement for a Genetic Algorithm in a Metadata Generation Support System for Landscape Photographs .....	67
<i>Tetsuya Suzuki and Takehiro Tokuda</i>	
Identifying Semantic Relations in Japanese Compound Nouns for Patent Documents Analysis .....	75
<i>Kiyoko Uchiyama, Shunsuke Aihara, and Shun Ishizaki</i>	
Extracting Prehistories of Software Refactorings from Version Archives .....	82
<i>Shinpei Hayashi and Motoshi Saeki</i>	

## Building Resources

Initiatives, Tendencies and Driving Forces for a “Lexical Web” as Part of a “Language Infrastructure” (Invited Paper) .....	90
<i>Nicoletta Calzolari</i>	

Corpus Annotation/Management Tools for the Project: Balanced Corpus of Contemporary Written Japanese (Invited Paper) . . . . .	106
<i>Yuji Matsumoto</i>	
Capturing the Structures in Association Knowledge: Application of Network Analyses to Large-Scale Databases of Japanese Word Associations . . . . .	116
<i>Terry Joyce and Maki Miyake</i>	
Construction of a Probabilistic Hierarchical Structure Based on a Japanese Corpus and a Japanese Thesaurus . . . . .	132
<i>Asuka Terai, Bin Liu, and Masanori Nakagawa</i>	
CHISE: Character Processing Based on Character Ontology . . . . .	148
<i>Tomohiko Morioka</i>	
Systematization of Knowledge about Performative Verbs: Capturing Speaker's Intention . . . . .	163
<i>Naoko Matsumoto, Hajime Murai, and Akifumi Tokosumi</i>	
Toward Construction of a Corpus of English Learners' Utterances Annotated with Speaker Proficiency Profiles: Data Collection and Sample Annotation . . . . .	171
<i>Yasunari Harada, Kanako Maebo, Mayumi Kawamura, Masanori Suzuki, Yoichiro Suzuki, Noriaki Kusumoto, and Joji Maeno</i>	
Filling the Gap between a Large-Scale Database and Multimodal Interactions . . . . .	179
<i>Masahiro Araki</i>	
Design and Prototype of a Large-Scale and Fully Sense-Tagged Corpus . . . . .	186
<i>Sue-jin Ker, Chu-Ren Huang, Jia-Fei Hong, Shi-Yin Liu, Hui-Ling Jian, I-Li Su, and Shu-Kai Hsieh</i>	
<b>Image and Video</b>	
Soccer Formation Classification Based on Fisher Weight Map and Gaussian Mixture Models . . . . .	194
<i>Toshie Misu, Masahide Naemura, Mahito Fujii, and Nobuyuki Yagi</i>	
Supervised Learning of Similarity Measures for Content-Based 3D Model Retrieval . . . . .	210
<i>Hamid Laga and Masayuki Nakajima</i>	
Automatic Score Scene Detection for Baseball Video . . . . .	226
<i>Koichi Shinoda, Kazuki Ishihara, Sadaoki Furui, and Takahiro Mochizuki</i>	

## Using Resources

Large Scale e-Language Laboratory Based on Web 2.0 (Invited Paper).....	241
<i>Kohji Shibano</i>	
Distant Collocations between Suppositional Adverbs and Clause-Final Modality Forms in Japanese Language Corpora .....	252
<i>Irena Srdanović Erjavec, Andrej Bekeš, and Kikuko Nishina</i>	
Using Singular Value Decomposition to Compute Answer Similarity in a Language Independent Approach to Question Answering .....	267
<i>Edward Whittaker, Josef Novak, Matthias Heie, Shuichiro Imai, and Sadaoki Furui</i>	
On the Representation of Perceptual Knowledge for Understanding Reference Expressions .....	280
<i>Philipp Spanger and Takenobu Tokunaga</i>	
A Computational Model of Risk-Context-Dependent Inductive Reasoning Based on a Support Vector Machine .....	295
<i>Kayo Sakamoto and Masanori Nakagawa</i>	
Order Retrieval .....	310
<i>Neil Rubens, Vera Sheinman, Takenobu Tokunaga, and Masashi Sugiyama</i>	
Stylistic Analysis of Japanese Prime Ministers' Diet Addresses .....	318
<i>Takafumi Suzuki and Kyo Kageura</i>	
Evaluation of Logical Thinking Ability through Contributions in a Learning Community .....	326
<i>Kayo Kawamoto and Noriko H. Arai</i>	

## Infrastructure

Web Architecture and Naming for Knowledge Resources (Invited Paper).....	334
<i>Henry S. Thompson</i>	
Towards Better Evaluation for Human Language Technology (Invited Paper).....	344
<i>Donna Harman</i>	
An Effective Scheduling Scheme for Information Searching with Computational Resources Scattered over a Large-Scale Network .....	351
<i>Shota Kondo, Shinji Sugawara, and Yutaka Ishibashi</i>	
<b>Author Index</b> .....	367