

Conference Proceedings



26th Pacific Asia Conference on Language, Information, and Computation

PACLIC 26

Bali, November 7-10, 2012

Published by Faculty of Computer Science, Universitas Indonesia

**Proceedings of the
26th Pacific Asia Conference on
Language, Information and Computation
(PACLIC 26)**

**7 - 10 November 2012
Bali, Indonesia**

**© 2012 The PACLIC 26 Organizing Committee
and PACLIC Steering Committee**

All rights reserved. Except as otherwise expressly permitted under copyright law, no part of this publication may be reproduced, digitized, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, Internet or otherwise, without the prior permission of the publisher.

Copyright of contributed papers reserved by respective authors

ISBN: 978-979-1421-17-1

Published by Faculty of Computer Science, Universitas Indonesia

Welcome Message from Honorary Chairs

On behalf of the Organizing Committee of the 26th Pacific Asia Conference on Language, Information and Computation (PACLIC 26), we would like to extend our warm welcome to all of the participants and speakers, and in particular, we would like to express our sincere gratitude to our invited speakers.

This international conference is organized by the Faculty of Computer Science, Universitas Indonesia and is supported by the I-MHERE DIKTI project. We are very keen to host a conference about language processing fields which involves many researchers in this Asia Pacific region. We believe that this international conference will open up the opportunities for sharing and exchanging original research ideas and opinions, getting inspiration for future research, and broadening knowledge about various new topics and approaches in language study. We hope that in this conference, the attendees would have the opportunity to meet with new people and discuss the opportunity to collaborate together.

We chose to organize PACLIC 26 in Bali so that aside from attending this interesting conference, you can also enjoy the scenery and the culture of Bali. We realize that there might not be enough time to see all the nice places in Bali, but we hope that you can bring home some good memories.

We would like to express our sincere appreciation to the members of the Program Committee for a fruitful reviews of the submitted papers, as well as the Organizing Committee for the time and energy they have devoted to editing the proceedings and arranging the logistics of holding this conference. We would like to give an appreciation to the authors who have submitted their excellent works to this conference. Last but not least, we would like to extend our gratitude to the Ministry of Education and Culture of the Republic of Indonesia and the Dean of the Faculty of Computer Science at Universitas Indonesia for their continued support towards the PACLIC 26 conference.

Have a nice time in Bali and enjoy the conference.

Honorary Chairs:

Mirna Adriani (Universitas Indonesia)

I Wayan Arka (ANU / Universitas Udayana)

Welcome Message from Program Co-Chairs

Welcome to Bali! This is the first time that the PACLIC conference is being held in Indonesia, and we are very excited about this fact. By all accounts, Indonesia is a linguistic treasure trove, with over 700 living languages today according to the Ethnologue report. Moreover, with an increasing number of its 240 million population active on the Internet via the Web and social networks, clearly these are exciting times to be engaging in computational approaches towards the languages of Indonesia.

However, this PACLIC conference in 2012 is special for other reasons, most notably the commemoration of 25 years of the conference series. Over the years, the conference has developed into one of the leading conferences in the fields of theoretical and computational linguistics, extending beyond the Asia-Pacific region. This year, the specific research topics that the papers focus on can be classified into the following: discourse & pragmatics, grammar & syntax, information extraction, information retrieval, lexical semantics, machine translation, parsing, sentiment analysis, text summarization & paraphrasing, and word sense disambiguation & distributional semantics. Moreover, there is an interesting mix of both theoretical and computational approaches to almost all of the aforementioned topics.

We received paper submissions representing immense diversity, with authors from 29 countries or regions, namely Australia, Bahrain, Belgium, Canada, China, Czech Republic, Denmark, Egypt, France, Germany, Hong Kong, India, Indonesia, Japan, Korea, Macau, Malaysia, Pakistan, Philippines, Qatar, Singapore, Slovakia, Sri Lanka, Taiwan, Thailand, Tunisia, United Kingdom, United States, and Vietnam. To ensure that all accepted papers met the high quality standard of the PACLIC conference, all papers were sent to three reviewers. Of the 117 submissions that we received, 39 papers (33%) were accepted for oral presentation, and another 18 papers (15%) were accepted for poster presentation. We believe this has yielded an interesting, diverse, and high-quality collection of papers, and are confident that the conference will be successful as a result.

A successful conference is the result of many peoples efforts and contributions. Aside from the efforts of the authors who will be presenting their current work, thanks must be given to the tremendous efforts made by the program committee members in their paper reviews. Besides the oral and poster paper presentations, the conference is enriched by several invited speakers. Firstly there is a Special Session commemorating 25 years of PACLIC, which brings together Prof Kiyong Lee from Korea University, Prof Yuji Matsumoto from the Nara Institute of Science and Technology, and Prof Benjamin T'sou from the Hong Kong Institute of Education, three figures who have been instrumental in the formation of the PACLIC tradition. We have also scheduled invited talks from Prof I Wayan Arka from ANU & Universitas Udayana and Prof Tim Baldwin from the University of Melbourne. The expertise in the respective fields of all five speakers will undoubtedly provide us with new insights for research. On behalf of the program committee, we express our heartfelt thanks to them all. We would also like to thank the steering committee for their guidance, and the local organizing committee at Universitas Indonesia for their dedicated efforts and their excellent coordination with all parties, which has ensured that this conference will be a successful event.

Finally, we wish that you will all enjoy the conference presentations and resulting discussions between old and new friends, and also have some time to enjoy the wondrous setting that is the island of Bali.

Program Co-Chairs:

Ruli Manurung (Universitas Indonesia)

Francis Bond (Nanyang Technological University)

PACLIC 26 Organizers

Steering Committee:

Jae-Woong Choe, Korea University
Yasunari Harada, Waseda University
Chu-Ren Huang, Hong Kong Polytechnic University
Rachel Edita Roxas, De La Salle University-Manila
Maosong Sun, Tsinghua University
Benjamin T'sou, City University of Hong Kong
Min Zhang, Institute for Infocomm Research

Honorary Chairs:

Mirna Adriani, Universitas Indonesia
I Wayan Arka, ANU/Universitas Udayana

Program Committee:

Co-Chairs:

Ruli Manurung, Universitas Indonesia
Francis Bond, Nanyang Technological University
Shu-Kai Hsieh, National Taiwan University
Donghong Ji, Wuhan University
Olivia Kwong, City University of Hong Kong
Seungho Nam, Seoul National University
Ryo Otoguro, Waseda University
Rachel Edita Roxas, De La Salle University-Manila

Members:

Wirote Aroonmanakun, Chulalongkorn University
Timothy Baldwin, University of Melbourne
Stephane Bressan, National University of Singapore
Hee-Rahk Chae, Hankuk University of Foreign Studies
Hsin-Hsi Chen, National Taiwan University
Eng Siong Chng, Nanyang Technological University
Siaw-Fong Chung, National Chengchi University
Beatrice Daille, University of Nantes
Mary Dalrymple, Oxford University
Danilo Dayag, De La Salle University

Minghui Dong, Institute for Infocomm Research
Rebecca Dridan, University of Oslo
Maria Flouraki, SOAS, University London
Guohong Fu, Heilongjiang University
Wei Gao, Chinese University of Hong Kong
Yasunari Harada, Waseda University
Munpyo Hong, Sungkyunkwan University
Shu-Kai Hsieh, National Taiwan Normal University
Xuanjing Huang, Fudan University
Kentoro Inui, Nara Institute of Science and Technology
Donghong Ji, Wuhan University
Nikiforos Karamanis, TouchType
Jong-Bok Kim, Kyung Hee University
Valia Kordoni, Saarland University
Sadao Kurohashi, Kyoto University
Oi Yee Kwong, City University of Hong Kong
Bong Yeung Tom Lai, City University of Hong Kong
Paul Law, City University of Hong Kong
Alessandro Lenci, University of Pisa
Gina-Anne Levow, University of Manchester
Haizhou Li, Institute for Infocomm Research
Qun Liu, Chinese Academy of Sciences
Qing Ma, Ryukoku University
Yanjun Ma, Baidu
Takafumi Maekawa, Hokusei Gakuen University Junior College
Yuji Matsumoto, Nara Institute of Science and Technology
Mathieu Morey, Universite d'Aix-Marseille & Nanyang Technological University
Yoshiki Mori, University of Tokyo
Seungho Nam, Seoul National University
Vincent Ng, University of Texas at Dallas
Jian-Yun Nie, Universite de Montreal
Toshiyuki Ogihara, University of Washington
David Yoshikazu Oshima, Nagoya University
Ryo Otoguro, Waseda University
Ceile Paris, Commonwealth Scientific and Industrial Research Organisation
Jong C. Park, Korea Advanced Institute of Science and Technology
Laurent Prevot, Universite de Provence
Long Qiu, Institute for Infocomm Research
Bali Ranaivo-Malancon, Universiti Malaysia Sarawak
Graeme Ritchie, University of Aberdeen
Rachel Edita Roxas, De La Salle University-Manila
Samira Shaikh, State University of New York - University at Albany
Sachiko Shudo, Waseda University
Melanie Siegel, Hochschule Darmstadt
Porn Siri Singhapreecha, Thammasat University
Virach Sornlertlamvanich, Thai Computational Linguistics Laboratory, NICT

Andrew Spencer, University of Essex
Jian Su, Institute for Infocomm Research
I-Wen Su, National Taiwan University
Keh-Yih Su, Behavior Design Corporation
Henry S. Thompson, University of Edinburgh
Takenobu Tokunaga, Tokyo Institute of Technology
Josef van Genabith, Dublin City University
Aline Villavicencio, Universidade Federal do Rio Grande do Sul
Haifeng Wang, Baidu
Houfeng Wang, Peking University
Hui Wang, National University of Singapore
Jiun-Shiung Wu, National Chiayi University
Jae Il Yeom, Hongik University
Satoru Yokoyama, Tohoku University
Min Zhang, Institute for Infocomm Research
Qiang Zhou, Tsinghua University
Michael Zock, Laboratoire d'Informatique Fondamentale de Marseille, C.N.R.S.
Chengqing Zong, Chinese Academy of Sciences

Local Organizing Committee:

Bayu Distiawan, Universitas Indonesia
Muhammad Hilman, Universitas Indonesia
Samuel Louvan, Universitas Indonesia
Lelya Rimadhiana, Universitas Indonesia
Clara Vania, Universitas Indonesia

The First Workshop on Generative Lexicon for Asian Languages (GLAL)

Organizers:

Shu-Kai Hsieh (Institute of Linguistics, National Taiwan University)
Zuoyan Song (School of Chinese Language and Literature, Beijing Normal University)
Kyoko Kanzaki (National Institute for Japanese Language and Linguistics)

Program Committee:

Toni Badia (Universitat Pompeu Fabra, Spain)
Christian BASSAC (Universit de Lyon2, France)
Pierrette Bouillon (ETI/TIM/ISSCO, Switzerland)
Nicoletta Calzolari (CNR-ILC, Italy)
Ann Copestake (University of Cambridge, UK)
Christiane Fellbaum (Princeton University, USA)

Catherine Havasi(MIT,USA)
Chu-Ren Huang (The Hongkong Polytechnic University, China)
Hitoshi Isahara (NICT, Kyoto, Japan)
Chungmin Lee(Seoul National University, Seoul, Korea)
Alessandro Lenci (Universita di Pisa, Pisa, Italy)
Kentaro Nakatani(Kounan University, Japan)
Seungho Nam (Seoul National University, Seoul, Korea)
Fiammetta Namer (ATILF-CNRS, University of Nancy, France)
Naoyuki Ono (Tohoku University, Sendai, Japan)
Laurent Prvot (Aix-Marseille Universit & CNRS, France)
James Pustejovsky (Brandeis University, USA)
Anna Rumshisky (Brandeis University, USA)
Patrick Saint-Dizier (CNRS, Toulouse, France)
Koichi Takeuchi(Okayama University, Japan)
Hongjun Wang (Peking University, Beijing, China)
Nianwen Xue (Brandeis University, Waltham, MA USA)
Yulin Yuan (Peking University, Beijing, China)
Seohyun Im (Brandeis University, USA)

Table of Contents

1. Invited Talks

<i>From All Possible Worlds to Small Worlds: A Story of How We Started and Where We Will Go Doing Semantics</i>	
Kiyong Lee	1
<i>Developing a Deep Grammar of Indonesian within the ParGram Framework: Theoretical and Implementation Challenges</i>	
I Wayan Arka	19
<i>Idiomatocity and Classical Traditions in Some East Asian Languages</i>	
Benjamin K Tsou	39
<i>Things between Lexicon and Grammar</i>	
Yuji Matsumoto	56
<i>Social Media: Friend or Foe of Natural Language Processing?</i>	
Timothy Baldwin	58

2. Regular Papers

<i>Towards a Semantic Annotation of English Television News - Building and Evaluating a Constraint Grammar FrameNet</i>	
Eckhard Bick	60
<i>Compositionality of NN Compounds: A Case Study on [NI+Artifactual-Type Event Nouns</i>	
Shan Wang, Chu-Ren Huang and Hongzhi Xu	70
<i>Automatic Domain Adaptation for Word Sense Disambiguation Based on Comparison of Multiple Classifiers</i>	
Kanako Komiya and Manabu Okumura	80
<i>Calculating Selectional Preferences of Transitive Verbs in Korean</i>	
Sanghoun Song and Jae-Woong Choe	89
<i>Extracting and Visualizing Semantic Relationships from Chinese Biomedical Text</i>	
Qingliang Miao, Shu Zhang, Bo Zhang and Hao Yu	99
<i>Entity Set Expansion using Interactive Topic Information</i>	
Kugatsu Sadamitsu Sadamitsu, Kuniko Saito, Kenji Imamura and Yoshihiro Matsuo	108
<i>Improving Chinese-to-Japanese Patent Translation Using English as Pivot Language</i>	
Xianhua Li, Yao Meng and Yao Meng	117
<i>Combining Social Cognitive Theories with Linguistic Features for Multi-genre Sentiment Analysis</i>	
Hao Li, Yu Chen, Heng Ji, Smaranda Muresan and Dequan Zheng	127

<i>Indonesian Dependency Treebank: Annotation and Parsing</i>	
Nathan Green, Septina Dian Larasati and Zdenek Zabokrtsky	137
<i>Handling Indonesian Clitics: A Dataset Comparison for an Indonesian-English Statistical Machine Translation System</i>	
Septina Dian Larasati	146
<i>Two Types of Nominalization in Japanese as an Outcome of Semantic Tree Growth</i>	
Tohru Seraku	153
<i>Semantic Distributions of the Color Terms, Black and White in Taiwanese Languages</i>	
Huei-ling Lai and Shu-chen Lu	163
<i>Language Independent Sentence-Level Subjectivity Analysis with Feature Selection</i>	
Aditya Mogadala and Vasudeva Varma	171
<i>Annotation Scheme for Constructing Sentiment Corpus in Korean</i>	
Hyopil Shin, Munhyong Kim, Hayeon Jang and Andrew Cattle	181
<i>Lexical Gaps and Lexicalization: Implications for Word Segmentation Systems for Chinese NLP</i>	
Chan-Chia Hsu.....	191
<i>Extracting Keywords from Multi-party Live Chats</i>	
Su Nam Kim and Timothy Baldwin	199
<i>Extracting Networks of People and Places from Literary Texts</i>	
John Lee and Chak Yan Yeung.....	209
<i>Pre- vs. Post-verbal Asymmetries and the Syntax of Korean RDC</i>	
Daeho Chung	219
<i>Pattern Matching Refinements to Dictionary-Based Code-Switching Point Detection</i>	
Nathaniel Oco and Rachel Edita Roxas	229
<i>An Adaptive Method for Organization Name Disambiguation with Feature Reinforcing</i>	
Shu Zhang, Jianwei Wu, Dequan Zheng, Yao Meng and Hao Yu.....	237
<i>Predicting Answer Location Using Shallow Semantic Analogical Reasoning in a Factoid Question Answering System</i>	
Hapnes Toba, Mirna Adriani and Hisar Maruli Manurung	246
<i>On the Alleged Condition on the Base Verb of the Indirect Passive in Japanese</i>	
Tomokazu Takehisa.....	254
<i>Comparing Classifier use in Chinese and Japanese</i>	
Yue Hui Ting and Francis Bond.....	264
<i>Nominative-marked Phrases in Japanese Tough Constructions</i>	
Akira Ohtani and Maria del Pilar Valverde Ibanez	272

<i>Emotional Tendency Identification for Micro-blog Topics Based on Multiple Characteristics</i> Quanchao Liu, Chong Feng and Heyan Huang	280
<i>Product Name Classification for Product Instance Distinction</i> Hye-Jin Min and Jong C. Park	289
<i>Automatic Detection of Gender and Number Agreement Errors in Spanish Texts Written by Japanese Learners</i> Maria del Pilar Valverde Ibanez and Akira Ohtani	299
<i>A Reranking Approach for Dependency Parsing with Variable-sized Subtree Features</i> Mo Shen, Daisuke Kawahara and Sadao Kurohashi	308
<i>Applying Statistical Post-Editing to English-to-Korean Rule-based Machine Translation System</i> Ki-Young Lee and Young-Gil Kim	318
<i>A Model of Vietnamese Person Named Entity Question Answering System</i> Mai-Vu Tran, Duc-Trong Le, Xuan- Tu Tran and Tien-Tung Nguyen	325
<i>Towards a Semantic Annotation of English Television News - Building and Evaluating a Constraint Grammar FrameNet</i> Shaohua Yang, Hai Zhao and Bao-liang Lu	333
<i>Emotion Estimation from Sentence Using Relation between Japanese Slangs and Emotion Expressions</i> Kazuyuki Matsumoto, Kenji Kita and Fuji Ren	343
<i>Can Word Segmentation be Considered Harmful for Statistical Machine Translation Tasks between Japanese and Chinese?</i> Jing Sun and Yves Lepage	351
<i>Introduction of a Probabilistic Language Model to Non-Factoid Question Answering Using Example Q&A Pairs</i> Kosuke Yoshida, Taro Ueda, Madoka Ishioroshi, Hideyuki Shibuki and Tatsunori Mori	361
<i>Answering Questions Requiring Cross-passage Evidence</i> Kisuh Ahn and Hee-Rahk Chae	371
<i>Thai Sentence Paraphrasing from the Lexical Resource</i> Krittaporn Phucharasupa and Ponrudee Netisopakul	381
<i>Anaphora Annotation in Hindi Dependency TreeBank</i> Praveen Dakwale, Himanshu Sharma and Dipti M Sharma	391
<i>Improving Statistical Machine Translation with Processing Shallow Parsing</i> Hoai-Thu Vuong, Vinh Van Nguyen, Viet Hong Tran and Akira Shimazu	401
<i>Psycholinguistics, Lexicography, and Word Sense Disambiguation</i> Oi Yee Kwong	408

Predicting Answer Location Using Shallow Semantic Analogical Reasoning in a Factoid Question Answering System

Hapnes Toba, Mirna Adriani, and Ruli Manurung

Faculty of Computer Science

Universitas Indonesia

Depok 16424, Indonesia

hapnes.toba@ui.ac.id, {mirna, maruli}@cs.ui.ac.id

Abstract

In this paper we report our work on a factoid question answering task that avoids named-entity recognition tool in the answer selection process. We use semantic analogical reasoning to find the location of the final answer from a textual passage. We demonstrate that without employing any linguistic tools during the answer selection process, our approach achieves a better accuracy than a typical factoid question answering architecture.

1 Introduction

The task of a question answering system (QAS) is to provide a single answer for a given natural language question. In a factoid QAS, the system tries to give the best answer of an open-domain fact-based question. For example, the question “*Where was an Oviraptor fossil sitting on a nest discovered?*”. A QAS should return ‘*Mongolia’s Gobi Desert*’ as the final answer.

A typical pipeline architecture in a fact-based QAS consists of four main processes, i.e.: question analysis, query formulation, information retrieval and answer selection. The main source of complexity in a QAS lies in the question analysis and answer selection process rather than in the information retrieval (IR) phase, which is usually achieved by utilizing third-party modules such as Lucene, Indri, or a web search engine.

The question analysis process seeks to determine the type of a given question, which in turn provides the expected answer type (EAT) of that question as a specific fact type, such as person, organization or

location. The EAT will be used to select the best answer during the answer selection process, usually by utilizing a named-entity recognizer (NER) tool in a factoid QAS (Schlaefel et al., 2006). Different approaches have been used in order to improve the performance of the answer selection component. Ko et al. (2010) employed probabilistic models for answer ranking of NER-based answer selection by utilizing external semantic resources such as WordNet. More advanced techniques utilizing linguistic tools have been proposed in Sun et al. (2005), which uses syntactic relation analysis to extract the final answer, and Moreda et al. (2010), which employs semantic roles to improve NER-based answer selection. Recent work by Moschitti and Quarteroni (2011) proposed classification of paired texts that learn to select answers by applying syntactic tree kernels to pairs of questions and answers.

In our current work, we try to reduce the dependency of the answer selection process on linguistic tools such as NER systems. Our main concern is that in reality we do not always have a complete NER tool for every fact type. In our example mentioned above, the answer has a fact type which is neither an exact location, person nor an organization, i.e.: ‘*Mongolia’s Gobi Desert*’. In such case, a NER-based system might fail to extract the answer. Further, if we have a complete NER-tool, it is still a complex problem to predict the location of the exact answer in a retrieval result.

We propose an approach which we call semantic analogical reasoning (SAR). Our approach tries to predict the location of the final answer in a textual passage by employing the analogical reasoning

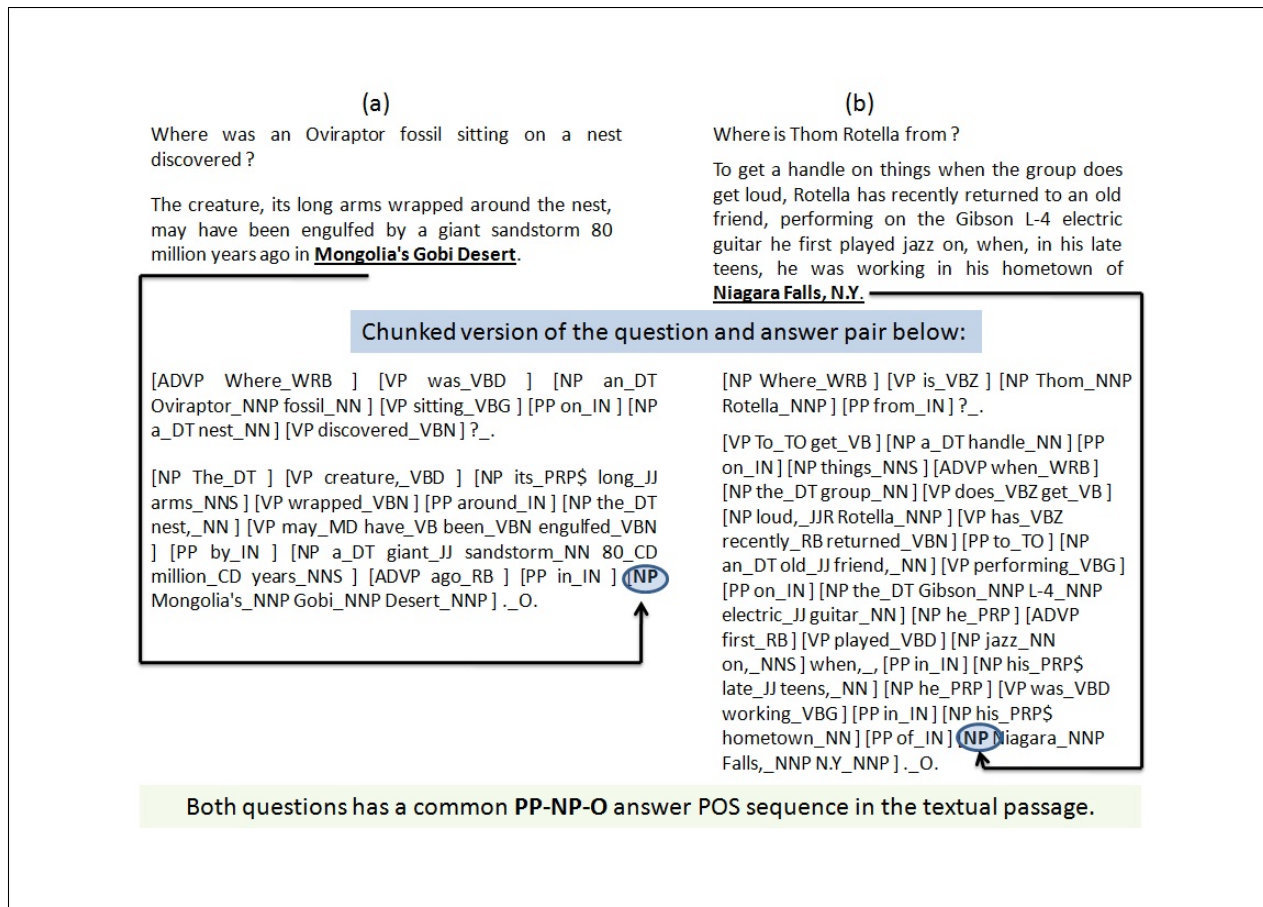


Figure 1: Idea of Semantic Analogical Reasoning

framework from Silva et al. (2010). We hypothesize that similar questions give similar answers. Based on the retrieved similar questions, our approach tries to provide the best example of question-answer pairs and use the influence level (weights) of the semantic features to predict the location of the final answer.

In the remainder of this paper, our basic idea and related works of semantic analogical reasoning will be presented in Section 2. The system architecture, procedures, experiments, and performance evaluation will be presented in Sections 3 and 4. Finally, our conclusions and future work will be drawn in Section 5.

2 Semantic Analogical Reasoning

The basic idea of semantic analogical reasoning is to find a portion of text in a passage which is considered useful during the answer selection process. Consider the two pairs of question and answer in Figure 1. Both questions need a fact type as the final

answer, i.e. *Mongolia's Gobi Desert* (a) and *Niagara Falls, N.Y.* (b). We postulate that both questions have a high probability to share common answer features that will be useful to find the location of the final answer.

If we investigate the structure of the answer passage of question (a), we can see that the final answer is a noun phrase (NP), which is surrounded by a preposition (PP) and a stop sign (O). In question (b), we also found that the final answer is located in a sequence of PP-NP-O. Thus, if we can learn these kinds of related structures between question answer pairs for any EAT, we will have useful information to predict the location of the final answer in a textual passage. In this sense, we focus our work in learning the relational feature similarity between question answer pairs.

Silva et al. (2007; 2010) has investigated a statistical-based analogical reasoning (AR) framework. It is a method for ranking relations based

on the Bayesian similarity criterion. The underlying idea of AR is to learn model parameters and priors from related objects (question and answer pairs in our case), and update the priors during the retrieval process of a query. The objective of the AR framework is to obtain a marginal probability that relates a new object pair (query) with a set of objects that have been learnt.

Most methods of classification or similarity measures focus on the similarity between the features of objects in a candidate pair and the features of objects in a query pair. AR focuses instead on the similarity between functions that map pairs to links. To some extent, this is the main reason that AR is appropriate for our idea.

Wang et al. (2009) has shown that AR is effective in retrieving similar question-answer pairs in a community-based QAS. They use statistical features such as term frequency, common n -gram length, and question answer length ratio. In contrast to our approach which tries to validate the location of a final answer; their work is limited to the retrieval of similar question.

The SAR approach, that we develop in this research is an extension of our previous work (Toba et al., 2011), which showed that AR can be used to construct EAT patterns. In our previous research, we used named-entity occurrences as features to relate the question and answer pairs. This time, instead of using named-entities as features, we use semantic information - which is based on syntactic features - to predict the corresponding named-entities.

Moschitti and Quarteroni (2011) use predicate argument structures, syntactic and shallow semantic tree kernel features to train question and answer pairs on SVM rank. Two consequences of using complex linguistic features is high computational cost and the requirement to have access to adequate linguistic resources and tools. For these reasons, we propose to use a simpler feature set, i.e. the trigram sequences of syntactic chunk. Unlike the research in Moschitti and Quarteroni (2011) that uses the whole syntactic tree, in this research we only keep the order of the root of any partial tree segment in trigrams of part-of-speech (POS) sequences.

In short, we develop a set of procedures to determine the best similar question-answer pair and predict the final answer location of a given factoid ques-

Question Features	Answer Features
Question word (W5H) of a question and its syntactic chunk. Example: <i>[NP which_WDT]</i>	POS-tagger and syntactic chunk of the final answer. Example: <i>NP-NNP</i>
Trigram of syntactic chunk sequence which appears in the question. Example: <i>PP-NP-PP, VP-NP-VP</i>	Trigram of the final answer, [left - answer - right chunk] (during training). Example: <i>PP-NP-VP</i> . Trigram chunk sequences of the whole answer passage (during testing)

Table 1: Question Answer Semantic Features used in the Analogical Reasoning

tion. Our SAR approach extends the above mentioned related works in the following aspects:

1. We extend the AR framework from Silva, et al. (2007; 2010) to re-rank the AR retrieval process according to the most influential semantic features.
2. We extend the question-answer retrieval process of Wang, et al. (2009) and Moschitti and Quarteroni (2011) to find the most possible final answer location in a textual passage by utilizing POS sequences as semantic features.

3 System Architecture

Our architecture is depicted in Figure 2. There are two main process flows in the architecture. The first one is the training process (noted by the dashed lines), and the second one is the question answering process (noted by the solid lines).

3.1 Question Answering Framework

During the training process, the semantic features as described in Table 1 will be extracted and used in the AR training module. The training process will produce an AR model. Another important step in the training process is the evaluation of the features importance level. We need to know which semantic feature has the most influence in the model. This information will be important to select the best question answer pair later in the re-ranking process.

In the question answering step, the shallow semantic features of the question and the related answer passages - which have been retrieved during

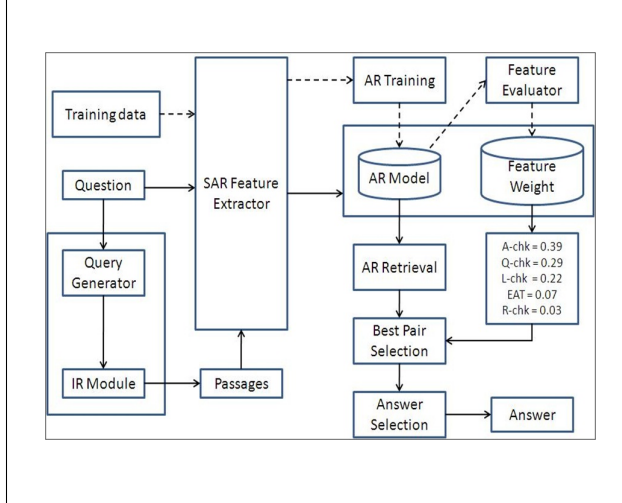


Figure 2: System Architecture

the IR process - are extracted. In this step, we will have a collection of ranked similar question answer pairs from the learnt AR model. Each similar pair needs to be evaluated (re-ranked), to make sure that we will have the best similar pair. In the final step, based on the best similar question answer pair, we search for the location of the answer from the textual passage by matching the sequence of the answer chunk to produce the final answer.

3.2 Analogical Reasoning, Re-rank Process and Final Answer Selection

In this part we summarize first the AR framework as introduced by Silva et al. (2007; 2010). The framework consists of two phases, i.e. the training and retrieval process. Consider a collection of related objects with some unseen labels L_{ij} 's, where $L_{ij} \in \{0, 1\}$ is an expected indicator of the existence of a relation between two related objects i and j . Consider then that we also have K -dimensional vectors, each consisting of features which relates the objects i and j : $\Theta = [\Theta_1 \dots \Theta_k]^T$. In general, this vector will represent the presence or absence of relation between two particular objects.

Given the vectors of features Θ , the strength of the relation between two objects i and j is computed by performing logistic regression estimation as follows:

$$P(L^{ij}|x^{ij}, \Theta) = \text{logistic}(\Theta^T X^{ij}) \quad (1)$$

where $\text{logistic}(x)$ is defined as:

$$\frac{1}{1 + e^{-x}} \quad (2)$$

During AR training phase, the framework learns the weight (prior) for each feature by performing the following equation:

$$P(\Theta) = N(\tilde{\Theta}, (c\tilde{T})^{-1}) \quad (3)$$

where $\tilde{\Theta}$ is the logistic estimator of Θ , and $N(m, v)$ is a normal of mean m and variance v . Matrix \tilde{T} is the empirical second moment's matrix of the link object features, and c is a smoothing parameter which is set by the user.

During the AR retrieval phase, a final score that indicates the rank of predicted relations between two new objects i and j (query) and the related objects that have been learnt in a given set S is compute as follows:

$$\text{score}(Q^i, A^j) = \log \frac{P(L^{ij}|X^{ij}, S, L^S = 1)}{P(L^{ij} = 1|X^{ij})} \quad (4)$$

Silva et al. (2010) use the variational logistic regression approach to compute the scoring function in equation 4. This score gives the rank of similarity of how "analogous" a new query is to other related objects in a given learnt set S .

A drawback of AR as mentioned in Silva et al. (2010) is that by conditioning on the link indicators, the similarity score (eq. 4) between two objects $i:j$, and other objects $x:y$, is always a function of pairs (i,j) and (x,y) that is not in general decomposable as similarities between i and x , and j and y . Due to this limitation, we propose to evaluate the importance level (weight) of each feature which is used to relate the objects, and use the weights to re-rank the similarity score.

We empirically calculate the weighting factors for each feature set in Table 1, with respect to the expected answer-type, by performing chi-square (χ^2) evaluation ((Manning et al., 2008), pp. 255-256) of overlapped features. The chi-square evaluation of the weighting factors are computed from the AR-retrieval results of the training data. To calculate the importance of each feature, we performed a top-10 retrieval for each question during the training phase on several parameter settings.

No.	Feature Set	Weight
1.	Answer Chunk	0.39
2.	Question Word + Chunk	0.29
3.	Left of Answer Chunk	0.22
4.	Expected Answer Type	0.07
5.	Right of Answer Chunk	0.03

Table 2: Weighting Factors of the Feature Sets. ‘*Expected Answer Type*’ is not part of the extracted and learnt feature set in the AR model. The information about EAT during the experiments is provided by the gold standard.

As suggested in Silva et al. (2010), the value of the smoothing parameter is set as the ‘*number of positive links*’ in the training data. We took variations of this smoothing parameter by multiplying the ‘*number of positive links*’ by a factor of: 0.1, 0.5, 2, 4, 8, 10 and 16 during the chi-square evaluation. Finally, we compute an average value to form the final weighting factor of each feature, as can be seen in Table 2.

The final answer selection strategy is started by selecting the best question-answer analogous pair. To select the best pair we performed first a top-10 AR retrieval, re-ranked them by using the feature weighting factors, and finally took the best score pair. This pair is considered as the best pair which has the most overlapped features to the new question. To select the final answer in a passage, we performed a feature matching process of the answer features, i.e.: the overlap of trigram POS chunks sequences ‘*[left chunk - answer chunk - right chunk]*’.

4 Experiments and Evaluation

The main objectives of our experiments are two-fold: on one hand, we try to find the importance level of the feature set that we use. On the other hand, we evaluate the potential of our approach to locate factoid answers in snippets and document retrieval scenarios without using any NER-tool. For the second objective we run two kinds of experiments. The first one is by using the gold standard snippets and the second one is by performing a document retrieval process.

In our experiments we use the question answer pairs from CLEF¹ English monolingual of the year

¹Question Answering at Cross Language Evaluation Forum (<http://celct.fbk.eu/ResPubliQA/index.php>)

2006, 2007 and 2008. For the training data we use the 2007 and 2008 collections. In total it consists of 321 factoid question answer pairs. For the testing data we use the 2006 collection, consisting of 75 factoid questions (Magnini et al., 2006).

In our empirical experiments, by performing chi-square statistics, we find that the answer chunk is the most important feature. The right-chunk of an answer is the least significant feature. The complete weighting factors of the feature set can be seen in Table 2. In our experiments, we also add the EAT parameter as one of the factors which will be important in the re-ranking process.

We use the accuracy metric during evaluation (Schlaefler et al., 2006) (Peñas et al., 2010), which covers the proportion the number of questions correctly answered in the test set. We choose this kind of evaluation because we are interested in the potential of our approach to predict the location of an answer in a given snippet / document.

4.1 Gold Standard Snippets

In this first experiment we assume that the IR process performed perfectly and returns the best snippet which covers the final answer. We choose Open Ephyra (Schlaefler et al., 2006) as our competing pipeline. This decision is based on the fact that Open Ephyra employs two types of NER integrated in it. The first type is the model-based NER which consists of OpenNLP² and Stanford NER³. The second type is a dictionary-based NER that was specially design for TREC-QA competition. To maintain the fairness of the evaluation, we decided to only use the first type (model-based NER) and build a special trained answer-type classifier for CLEF datasets as described in Toba et al. (2010). In short, we hold the QA components of our approach and those of Open Ephyra all the same, except for the final answer selection.

The result of this experiment can be seen in Table 3. Our approach outperforms the overall result of Open Ephyra. The best accuracy of our approach is achieved in the OTHER and LOCATION answer-type, both 0.83, whereas the worst accuracy is for the TIME-typed questions, 0.45. In particular, our

²<http://opennlp.apache.org>

³<http://nlp.stanford.edu/ner/index.html>

Q.Type	#.Quest.	SAR	OE-NER
measure	14	0.57	0.64
person	13	0.77	0.69
other	12	0.83	0.25
location	12	0.83	0.92
organization	13	0.62	0.54
time	11	0.45	0.73
all	75	0.68	0.63

Table 3: Gold Standard Experiment Accuracy

approach performs exceptionally well in the ‘OTHER’ type. We believe this is due to the fact that our strategy finds the location of an expected answer without depending on the performance of an NER-tool. An example of ‘OTHER-typed’ questions is (CLEF 2006 #8): “*What is the Bavarian National Anthem?*”. The expected answer for this question is: “*God be with you, land of Bavarians*”. The answer chunk constituent in the gold standard snippet is a sequence of “VP-NP-O”, which comes from the following snippet: “*They ended their demonstration by singing the Bavarian Anthem “God be with you, land of Bavarians”. Then many of them moved on to support another Bavarian tradition – Oktoberfest.*”.

If we look deeper into the feature set which is used in the AR training in Table 1, our trigram chunk features actually consist of two bigrams: (*left+answer*)-chunk and (*answer+right*)-chunk. During the final answer selection we consider these left and right chunk-bigrams as part of the selection process, not only the trigram sequence. This strategy is to ensure that the answer could be covered in one of the possibilities: a chunk trigram, a chunk left-bigram, or a chunk right-bigram.

In this first experiment, the most difficult questions to be answered are the TIME and MEASURE question-types. The answer of the TIME answer-type can be in the form of: dd/mm/yy, dd-mmm-yy, a single year number, or in the form of hh:mm a.m./p.m. These variations give rise to problems during the feature extraction process, because sometimes the chunker recognizes variations as numbers or as nouns. This problem also occurred in the MEASURE-typed questions. A measurement can be written as numbers (for example: “40”) or as text (“forty”), and the chunker recognizes them differ-

ently, even though they express the same thing.

Figure 3(a) gives the number of expected “AR trigram sequences” in each answer-type which needs to be found in the snippets. We can see that for a factoid question answering task, the expected answers are mostly in the form of an NP (noun phrase).

4.2 Indri Document Retrieval

In our second experimental setting, we try to simulate our approach in a more realistic question answering system. In the real situation, we will not have any information about the semantic chunk of the final answer. We assume that the best pair (i.e. the top-1 pair after the re-ranking process) of the AR answer features will supply us with that information. We performed IR process by using Indri Search Engine to retrieve the top-5 documents and pass them on to Open Ephyra and our system.

In this experiment, we use the same AR feature set as in the first experiment during the training phase. However, unlike the first experiment, during the top-10 AR retrieval process, we only use the question feature set, i.e.: the question word and its chunk, and the question trigram of the semantic chunks. Due to the lack of the answer features, we need to adjust the way of the re-ranking process. We use a scoring function (*sf*) which takes the AR and Indri retrieval results into consideration. We use the formula in equation 5. We adjust the weight of the parameters to fit the question features during the re-ranking process of the AR retrieval results.

$$sf = \{\alpha OV(a_i, a_j) + \beta OV(b_i, b_j)\} * \log(AR) * IR * \frac{1}{100} \quad (5)$$

where:

- a = question chunk
- b = expected answer type
- α = weight of overlap question word and chunk (0.92)
- β = weight of overlap expected answer type (0.08)
- AR = score of the AR retrieval (see eq. 4)
- IR = the weight of the Indri top-5 retrieval rank (5 to 1)
- $OV(x,y)$ = 1, if there is overlap between x and y in a question i and its analogy pair j , otherwise 0

The result of this second experiment can be seen in Table 4. Both the SAR approach and Open E-

(a)							(b)						
Patterns 3-g	Meas	Pers	Oth	Loc	Org	Time	Patterns 3-g	Meas	Pers	Oth	Loc	Org	Time
ADVP-NP-NP	1	0	0	0	0	0	ADVP-NP-O	1	0	0	0	0	0
NP-NP-O	0	0	0	0	0	1	NP-NP-NP	0	1	0	0	0	0
NP-NP-VP	0	1	1	0	0	0	NP-NP-PP	2	0	0	0	0	0
O-NP-O	2	0	0	0	0	1	NP-NP-VP	1	0	0	0	0	0
O-NP-PP	0	3	0	1	1	0	O-NP-NP	0	2	0	1	0	0
O-NP-VP	2	3	1	0	2	0	O-NP-O	0	1	2	2	2	0
PP-NP-PP	1	0	0	0	0	1	O-NP-PP	0	1	1	0	1	0
PP-NP-VP	0	0	2	4	1	3	O-NP-VP	1	0	1	0	2	0
PP-NP-ADJP	1	0	0	0	0	0	PP-NP-NP	0	0	0	1	0	0
PP-NP-ADVP	0	0	0	2	0	1	PP-NP-O	2	5	3	5	5	8
PP-NP-NP	0	0	1	2	1	0	PP-NP-PP	2	1	0	1	0	2
PP-NP-O	1	3	1	3	5	1	PP-NP-VP	1	1	0	0	1	0
PP-NP-PP	1	1	1	0	0	0	VP-NP-O	2	1	3	2	0	0
VP-NP-ADVP	0	1	0	0	1	0	VP-NP-PP	0	0	2	0	2	1
VP-NP-NP	0	0	1	0	0	0	VP-NP-VP	2	0	0	0	0	0
VP-NP-O	2	0	2	0	2	0	# Questions	14	13	12	12	13	11
VP-NP-PP	2	1	2	0	0	3	# of Patterns	9	8	6	6	6	3
VP-NP-VP	1	0	0	0	0	0							
# Questions	14	13	12	12	13	11							
# of Patterns	10	7	9	5	7	7							

ADVP = Adverb phrase
NP = Noun phrase
PP = Prepositional phrase
O = Begin/End of a sentence;
or a coordinating conjunction

Figure 3: Number of Chunk Trigram Sequences in each EAT: (a) Snippet Experiment (b) Indri Retrieval Experiment

phra have a lower accuracy compared to the first experiment. Once again, our approach achieves a higher accuracy. In the NER-based system, the errors are mainly caused by the model in the NER tools which cannot find the appropriate answer. For example for a person name “Carl Lewis”, the NER tools can only recognize it either as Carl or Lewis, but not the whole name.

Q.Type	#.Quest.	SAR	OE-NER
measure	14	0.57	0.36
person	13	0.62	0.08
other	12	0.17	0.08
location	12	0.17	0.42
organization	13	0.23	0.15
time	11	0.27	0.27
all	75	0.33	0.23

Table 4: Indri Retrieval Experiment Accuracy

We classify the error types of our approach in three groups, i.e.: (1) not covered by Indri retrieval, (2) decreasing rank of relevant document because of

the AR re-ranking score function, and (3) irrelevant example from the best AR pair. The frequency of these error groups can be seen in Table 5.

Error Class	Fre
AR Re-ranking (decreasing rank of relevant documents)	19
Irrelevant AR Example	17
Indri Retrieval	14
Total not found answers	50

Table 5: Frequency of Error Classification of SAR Approach

In our opinion, the main drawback of our approach is that it suffers from the variations of sentence structures - those of the snippets in the training set and those of the retrieved documents. These variations influence the AR re-ranking and matching process of chunk sequences. For instance, if the AR best pair suggests that the answer should be located at the end of a sentence, while that chunk could not be found in the retrieved document, then we will

have a negative result. An example of such case can be seen in Table 6. The complete occurrences of the expected trigram sequences in this second experiment can be found in Figure 3(b).

Question and Answer	SAR (found in relevant document)	Expected (AR retrieval)
Q: Who is head of Bank of Tokyo? (CLEF 2006 #52) A: Tasuku Takagaki	Sequence: PP-NP-VP Bank_B-NP of_B-PP Tokyo_B-NP president_I-NP Tasuku_I-NP Takagaki_I-NP said_B-VP	Sequence: VP-NP-O said_B-VP series_B-NP creator_I-NP Sherwood_I-NP Schwartz_I-NP ..O

Table 6: Influence of Sentence Structures

5 Conclusion and Future Work

In this paper we have shown that by learning analogical linkages of question-answer pairs we can predict the location of factoid answers of a given snippet or document. Our approach achieves a very good accuracy in the OTHER answer-type (cf. Section 4.1). It shows the potential of our approach for dealing with an answer-type with no available corresponding NER tool.

Another finding in our experiments is that there is no trigram answer chunk sequence that really dominates in each answer-type. This suggests that each question depends on the sentence structure of a given snippet, and has a different way to be answered. This fact also suggests that our approach could suffer from the variations of the sentence structures. In our opinion, this is one of the reasons why the accuracy drops when the AR retrieval does not guarantee the occurrence of an answer (cf. Section 4.2). However, our approach has achieved a higher accuracy than a pure NER-based question answering system.

For our future work, we plan to develop a hybrid method of our approach with NER-based methods on larger and different datasets with more answer-type variations. We also plan to conduct another research in which we consider the trained question answer pairs as a kind of rule set. In this sense we look forward to combining the statistical approach, i.e. the analogical framework, and the semantic approach, i.e. the knowledge (rule) acquisition from the trained question answer pairs.

References

- Alessandro Moschitti and Silvia Quarteroni. 2011. Linguistic Kernels for Answer Re-ranking in Question Answering Systems. *Information Processing and Management*, 47:825–842.
- Anselmo Peñas, Pamela Forner, Alvaro Rodrigo, Richard Sutcliffe, Corina Forascu, and Christina Mota. 2010. Overview of ResPubliQA 2010: Question Answering Evaluation over European Legislation. *CLEF Notebook Papers*.
- Bernardo Magnini, Danilo Giampiccolo, Pamela Forner, Christelle Ayache, Valentin Jijkoun, Petya Osenova, Anselmo Peñas, Paulo Rocha, Bogdan Sacaleanu, and Richard Sutcliffe. 2006. Overview of the CLEF 2006 Multilingual Question Answering Track. *CLEF Question Answering Working Notes*.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, USA.
- Hapnes Toba, Mirna Adriani, Ruli Manurung. 2010. Contextual Approach for Paragraph Selection in Question Answering Task. *CLEF Notebook Papers*.
- Hapnes Toba, Mirna Adriani, Ruli Manurung. 2011. Expected Answer Type Construction using Analogical Reasoning in a Question Answering Task. *Proc. of ICACSSIS*.
- Jeongwoo Ko, Luo Si, Eric Nyberg, and Teruko Mitamura. 2010. Probabilistic Models for Answer-Ranking in Multilingual Question-Answering. *ACM Trans. on Information Systems*, 28(3) article 16: 1–35.
- N. Schlaefter, P. Giesemann, T. Schaaf, and A. Waibel. 2006. A Pattern Learning Approach to Question Answering within the Ephyra Framework. *LNAI*, 4188: 687-694. Springer, Heidelberg.
- P. Moreda, H. Llorens, E. Saquete, and M. Palomar. 2010. Combining Semantic Information in Question Answering System. *Information Processing and Management*, doi: 10.1016/j.ipm.2010.03.008.
- Ricardo Silva, Katherine Heller, and Zoubin Ghahramani. 2007. Analogical Reasoning with Relational Bayesian-sets. *Proc. of AISTATS*.
- Ricardo Silva, Katherine Heller, Zoubin Ghahramani, and Eduardo M. Airolidi. 2010. Ranking Relations Using Analogies in Biological and Information Networks. *The Annals of Applied Statistics*, 4(2):615–644.
- Renxu Sun, Hang Cui, Keya Li, Min-Yen Kan, and Tat-Seng Chua. 2005. Dependency Relation Matching for Answer Selection. *Proc. of SIGIR*.
- X-J. Wang, X. Tu, D. Feng, and L. Zhang. 2009. Ranking Community Answers by Modeling Question-Answer Relationship via Analogical Reasoning. *Proc. of SIGIR*.