

NAACL HLT 2010

Workshop on Semantic Search

Proceedings of the Workshop

June 5, 2010
Los Angeles, California

USB memory sticks produced by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2010 The Association for Computational Linguistics

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Introduction

Welcome to the NAACL HLT Workshop on Semantic Search!

Information retrieval (IR) research has been actively driven by the challenging information overload problem and many successful general-purpose commercial search engines. While the popularity of the largest search engines is a confirmation of the success and utility of IR, the identification, representation, and use of the often-complex semantics behind user queries has not yet been fully explored.

In this workshop we target methods that exploit semantics in search-related tasks. One of the major obstacles in bridging the gap between IR and Natural Language Processing (NLP) is how to retain the flexibility and precision of working with text at the lexical level while gaining the greater descriptive precision that NLP provides. We have solicited contributions on automatic analysis of queries and documents in order to encode and exploit information beyond surface-level keywords: named entities, relations, semantic roles, etc.

This workshop is meant to accelerate the pace of progress in semantic search techniques by connecting IR and NLP, bridging semantic analysis and search methodologies, and exploring the potentials of search utilizing semantics. We also focus on forming an interest group from different areas of research, exploring collaboration opportunities, providing deeper insight into bringing semantics into search, and provoking or encouraging discussions on all of its potential.

We are interested in semantic search technologies including the following topics:

- Query Parsing and Semantic Tagging
- Query Suggestion and Recommendation
- Query Expansion and Intention Detection
- Web Query Analysis and Mining
- Semantic Annotation and Indexing
- Language Modeling for Information Retrieval
- Information Extraction and Summarization for Indexing and Search
- Question Answering
- Search Reranking Integrating Semantic Features
- Search Relevance Evaluation using Semantic Technology
- Topic Modeling and Semantic Tagging

We received 11 submissions and selected 6 papers after a rigorous review process. Each paper has been reviewed by at least three reviewers. We are pleased to present these papers in this volume.

Our workshop will start with a keynote speech by Ronald Kaplan (Powerset Division of Microsoft Bing). We will also hold a panel discussion on the potential to explore semantic search technologies.

We are very grateful to the Program Committee for their hard work, and the presenters for their excellent papers.

Best regards,
Donghui Feng, Jamie Callan, Eduard Hovy, and Marius Paşca
Workshop Organizers

Organizers:

Donghui Feng, AT&T Interactive
Jamie Callan, Carnegie Mellon University
Eduard Hovy, USC/Information Sciences Institute
Marius Paşca, Google Inc.

Program Committee:

Srinivas Bangalore, AT&T Labs-Research
Raman Chandrasekar, Microsoft Research
Jennifer Chu-Carroll, IBM Research
Nikesh Garera, Kosmix
Ralph Grishman, New York University
Patrick Haffner, AT&T Labs-Research
Liang Huang, USC/Information Sciences Institute
Heng Ji, City University of New York
Zornitsa Kozareva, USC/Information Sciences Institute
Hang Li, Microsoft Research Asia
Dekang Lin, Google Inc.
Yumao Lu, Yahoo! Labs
Hwee Tou Ng, National University of Singapore
Fuchun Peng, Yahoo! Labs
Ellen Riloff, University of Utah
Dan Roth, University of Illinois at Urbana-Champaign
James Shanahan, AT&T Interactive
Young-In Song, Microsoft Research Asia
Qin Wang, AT&T Interactive
Kenji Yamada, AT&T Interactive
Remi Zajac, AT&T Interactive

Invited Speaker:

Ronald Kaplan, Powerset Division of Microsoft Bing

Table of Contents

<i>LDA Based Similarity Modeling for Question Answering</i>	
Asli Celikyilmaz, Dilek Hakkani-Tur and Gokhan Tur	1
<i>Experts' Retrieval with Multiword-Enhanced Author Topic Model</i>	
Nikhil Johri, Dan Roth and Yuancheng Tu	10
<i>Query-based Text Normalization Selection Models for Enhanced Retrieval Accuracy</i>	
Si-Chi Chin, Rhonda DeCook, W. Nick Street and David Eichmann	19
<i>A Graph-Based Semi-Supervised Learning for Question Semantic Labeling</i>	
Asli Celikyilmaz and Dilek Hakkani-Tur	27
<i>Capturing the Stars: Predicting Ratings for Service and Product Reviews</i>	
Narendra Gupta, Giuseppe Di Fabrizio and Patrick Haffner	36
<i>Object Search: Supporting Structured Queries in Web Search Engines</i>	
Kim Pham, Nicholas Rizzolo, Kevin Small, Kevin Chen-Chuan Chang and Dan Roth	44

Workshop Program

Saturday, June 5, 2010

- 9:20-9:30 Opening Remarks
- 9:30-10:30 Invited Talk by Ronald Kaplan
- 10:30-11:00 Morning Break
- 11:00–11:30 *LDA Based Similarity Modeling for Question Answering*
Asli Celikyilmaz, Dilek Hakkani-Tur and Gokhan Tur
- 11:30–12:00 *Experts' Retrieval with Multiword-Enhanced Author Topic Model*
Nikhil Johri, Dan Roth and Yuancheng Tu
- 12:00–12:30 *Query-based Text Normalization Selection Models for Enhanced Retrieval Accuracy*
Si-Chi Chin, Rhonda DeCook, W. Nick Street and David Eichmann
- 12:30-1:45 Lunch Break
- 1:45–2:15 *A Graph-Based Semi-Supervised Learning for Question Semantic Labeling*
Asli Celikyilmaz and Dilek Hakkani-Tur
- 2:15-3:00 Panel Discussion
- 3:00-3:30 Afternoon Break
- 3:30–3:45 *Capturing the Stars: Predicting Ratings for Service and Product Reviews*
Narendra Gupta, Giuseppe Di Fabrizio and Patrick Haffner
- 3:45–4:00 *Object Search: Supporting Structured Queries in Web Search Engines*
Kim Pham, Nicholas Rizzolo, Kevin Small, Kevin Chen-Chuan Chang and Dan Roth
- 4:00-4:45 Workshop Wrap-up Discussion

