COLING 2012

# 24th International Conference on Computational Linguistics

# Proceedings of the Workshop on Question Answering for Complex Domains

**Workshop chairs:**
**Nanda Kambhatla, Sachindra Joshi, Ganesh Ramakrishnan, Kiran Kate and Priyanka Agrawal**

**9 December 2012**
**Mumbai, India**

## Diamond sponsors

Tata Consultancy Services
Linguistic Data Consortium for Indian Languages (LDC-IL)

## Gold Sponsors

Microsoft Research
Beijing Baidu Netcon Science Technology Co. Ltd.

## Silver sponsors

IBM, India Private Limited
Crimson Interactive Pvt. Ltd.
Yahoo
Easy Transcription & Software Pvt. Ltd.

*Proceedings of the Workshop on Question Answering for Complex Domains*
Nanda Kambhatla, Sachindra Joshi, Ganesh Ramakrishnan, Kiran Kate and Priyanka Agrawal (eds.)
Revised preprint edition, 2012

Published by The COLING 2012 Organizing Committee
Indian Institute of Technology Bombay,
Powai,
Mumbai-400076
India
Phone: 91-22-25764729
Fax: 91-22-2572 0022
Email: pb@cse.iitb.ac.in



Also available online in the ACL Anthology at http://aclweb.org

**Preface**

Significant progress has been made in building Question Answering systems that are focussed on providing precise answers to specific questions in one shot. In many real world situations, however, the information need may be specified vaguely and some sort of interaction may be required for a complete specification of the problem. Further, the correct answer of the question may be a procedure or passage rather than a factoid. For example, someone having trouble with battery life for her smartphone may express a query as 'battery dying too soon for my XYZ phone'. The intent of the user here, obviously, is to seek a resolution to her problem. The resolution in question may be a procedure consisting of a sequence of steps. In order to recommend the correct resolution, a system may first have to engage in a dialog with the user to ascertain all the symptoms and match the correct resolution ('answer').

In this workshop, we are looking to explore such advanced QA systems that seek to resolve more general user problems in an interactive manner. We are specifically interested in the problem of rapidly bootstrapping such QA systems with limited data annotation. We invited researchers to submit papers discussing techniques for rapid annotation of Q/A pairs, information extraction, machine learning and interactive dialog for building such systems. The workshop topics include but are not limited to:

- Semi-automated data collection for building QA systems
- Crowdsourcing techniques applied to building QA systems
- Learning apriori or on the fly domain models for QA systems
- Information Extraction of problem resolutions from text
- Dialog systems for interactive question answering - clarification
- sub-dialogues, error correcting dialogues
- Building dialogue models using conversation transcripts
- System descriptions of large QA systems
- Evaluation: user centered evaluation, percent of cognitive load
- compared to search, effectiveness of interaction, quality of results

# Keynote presentation
# Answering Questions from Conversations

Douglas W. Oard
University of Maryland

**Abstact**

Perhaps unsurprisingly, the early research on question answering at the Text Retrieval Conferences (TREC) focused on answering questions from formal written text, often in the form of news stories. As is well known, that early work initially posed one-shot questions and asked for factual answers. Although our interests as a community have since grown to encompass a richer and more complex array of questions and desired answer types, formal written text is still where we most often look to find those answers. But that need not be so, and going forward it probably should not be so. In this talk I will suggest that conversations, both spoken and written, offer substantial scope for future question answering research. To make that case, I will argue from two key perspectives: (1) that some of what we seek to find can be found only in conversations, and (2) that simply retrieving parts of conversations will in many cases not be sufficient. Along the way I will illustrate my points with examples from some the work that has been done to date on meeting browsers, focused retrieval from interviews, and question answering from threaded discussion lists. I'll then look for inspiration to three emerging trends in content indexing: (1) automated characterization of social dynamics, (2) so-called "learning by reading" in which conformal semantic representations are automatically constructed from natural language, and (3) a diverse array of techniques for indexing spoken content. To wrap up the talk, I will invite us to imagine together what kinds of question answering systems we will be able to build for conversational content as these technologies mature.

## Organizing Committee:

Nanda Kambhatla (IBM Research - India)
Sachindra Joshi (BM Research - India)
Ganesh Ramakrishnan (IIT Bombay)
Kiran Kate (IBM Research - Singapore)
Priyanka Agrawal (IBM Research - India)

## Program Committee:

Eric Brown (IBM T. J. Watson Research Center)
Jennifer Chu-Carroll (IBM IBM T. J. Watson Research Center)
Raghavendra Udupa (Microsoft Research, India)
Doug Oard (University of Maryland, College Park)
Carolyn Rose (CMU)
Indrajit Bhattacharya (IISc)
Li Haizhou (Institute for Infocomm Research, Singapore)
Cong Gao (NTU, Singapore)
Sutanu Chakraborti (IIT Madras)
Rebecca Passonneau (Columbia University)
Balaraman Ravindran (IIT Madras)
Vasudeva Varma (IIIT Hyderabad)
Ullas Nambiar (EMC India)
Shourya Roy (Xerox Research, India)
Radhika Mamidi (IIIT Hyderabad)

## Invited Speakers:

Doug Oard (University of Maryland, College Park)
Nanda Kambhatla (IBM Research - India)

# Table of Contents

# Workshop on Question Answering for Complex Domains

## Program

**Sunday, 9 December 2012**

| | |
|---|---|
| 09:30–10:30 | **Invited keynote:**<br>*Answering Questions from Conversations*<br>Douglas W. Oard, University of Maryland |
| 10:30–11:00 | *Simple or Complex? Classifying Questions by Answering Complexity*<br>Yllias Chali and Sadid A. Hasan |
| 11:00–11:30 | *Question Classification and Answering from Procedural Text in English*<br>Somnath Banerjee and Sivaji Bandyopadhyay |
| 11:30–12:00 | Tea break |
| 12:00–12:30 | *Structured and Logical Representations of Assamese Text for Question-Answering System*<br>Shikhar Kr. Sarma and Rita Chakraborty |
| 12:30–13:00 | *Towards a thematic role based target identification model for question answering*<br>Rivindu Perera and Udayangi Perera |
| 13:00–13:30 | *Assessment of Answers: Online Subjective Examination*<br>Asmita Dhokrat, Hanumant Gite and C. Namrata Mahender |
| 13:30–14:30 | Lunch |
| 14:30–15:30 | **Invited talk**<br>*Overview of Jeopardy! winning Watson system*<br>Nanda Kambhatla, IBM Research - India |
| 15:30–16:00 | *WikiTalk: A Spoken Wikipedia-based Open-Domain Knowledge Access System*<br>Graham Wilcock |
| 16:00–16:30 | Tea break |
| 16:30–17:30 | Open discsussion |