Question Answering in Restricted Domains

Proceedings of the ACL 2004 Workshop

25 July 2004
Barcelona, Spain

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
73 Landmark Center
East Stroudsburg, PA 18301
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

INVITED SPEAKER:

Daniel Marcu, University of Southern California, USA

PROGRAM COMMITTEE:

Diego Mollá, Macquarie University, Australia, Chair
José Luis Vicedo, Alicante University, Australia, Chair

Johan Bos, University of Edinburgh, UK
Antonio Ferrández, Alicante University, Spain
Juergen Franke, DaimlerChrysler AG, Germany
Julio Gonzalo, UNED, Spain
Sanda Harabagiu, University of Texas, USA
Donna Harman, NIST, USA
Michael Hess, University of Zurich, Switzerland
Lynette Hirschman, MITRE, USA
Graeme Hirst, University of Toronto, Canada
Jimmy Lin, MIT, USA
Bernardo Magnini, ITC-Irst, Italy
Mark Maybury, MITRE, USA
Manuel Palomar, Alicante University, Spain
Anselmo Peñas, UNED, Spain
Maarten de Rijke, University of Amsterdam, The Netherlands
Fabio Rinaldi, University of Zurich, Switzerland
Horacio Rodríguez, Universitat de Catalunya, Spain
Rolf Schwitter, Macquarie University, Australia
Richard Sutcliffe, University of Limerick, Ireland
Felisa Verdejo, UNED, Spain
Ellen Voorhees, NIST, USA
Bonnie Webber, University of Edinburgh, UK
Ingrid Zukerman, Monash University, Australia
Pierre Zweigenbaum, DIAM, France

WORKSHOP WEBSITE:

http://www.clt.mq.edu.au/Events/Conferences/acl04qa/

PREFACE

This volume contains the papers accepted for presentation at the workshop on Question Answering in Restricted Domains, which is part of the 42nd Annual Meeting of the Association for Computational Linguistics, held on July 21-26, 2004 in Barcelona, Spain.

Much of the current research in question answering systems is driven by programs such as AQUAINT and evaluation exercises such as TREC, NTCIR and CLEF, all of which focus on open-domain question answering. The availability of large volumes of data (e.g. documents extracted from the World Wide Web) has prompted the development of systems that focus on shallow text processing.

But there are many document sets in restricted domains that are potentially valuable as a source for question answering systems. For example, the documentation pages of Unix and Linux systems would make an ideal corpus for QA systems targeted at users that want to know how to use these operating systems. There is a wealth of information in other technical documentation such as software manuals, car maintenance manuals, and encyclopediae of specific areas such as medicine. Users interested in these specific areas would benefit from QA systems targeted to their areas of interest.

Restricted domains typically have limited data available and therefore conventional techniques based on data redundancy can simply not be applied in an effective way. The scarcity of data available seems to prompt for a more targeted, NLP-intensive approach to QA. The use of additional corpora such as the WWW raises a number of interesting questions. For instance, will these corpora help or obstruct the proper functioning of NLP-intensive approach to QA? And, how do we find good pockets of information that are appropriate to the chosen domains?

On the other hand, restricted domains (e.g. law, medicine) have specific stylistic conventions. Often these domains use terminology that is not stored in conventional lexica. Consequently NLP approaches devised for open-domain systems may choke on these specific domains, thus raising the question of how portable these systems can be.

In this workshop we aim at answering some of the following questions:

- Are open-domain question answering techniques appropriate for QA in restricted domains?
- Can we use generic large corpora and/or the WWW? How can we identify specific pockets of information in these generic corpora?
- How can we use specific sources such as CIA factbook, acronym lists, e-commerce sites (e.g., e-bay), and specialized glossaries and encyclopedia? How can we discover new specific sources?
- What types of question-answering techniques are best for what types of restricted domains?
- Is it easy/possible/worthwhile to develop domain-independent QA systems for restricted domains? What would be the cost of porting a QA system to other domains?
- Are restricted domains more suitable than open domains to drive research in NLP?
- Is evaluation of restricted-domain QA systems different than that of open-domain QA systems?

Of the 13 papers submitted, the programme committee selected 8 papers. We are very grateful to our programme committee for the effort they put in reviewing the full papers. We are also grateful to the ACL/EACL-2004 conference organisers on whom we could rely for the local organization.

Diego Mollá & José Luis Vicedo (editors)
June 2004

Table of Contents

Technical Program Schedule

Sunday, July 25

8:45-9:00 Welcome

9:00-10:00 *The Perils and Rewards of Developing Restricted Domain Applications*
Daniel Marcu

Coffee Break

10:30-11:00 *Evaluation of Restricted Domain Question-Answering Systems*
Anne R. Diekema, Ozgur Yilmazel and Elizabeth D. Liddy

11:00-11:30 *The Problem of Precision in Restricted-Domain Question Answering. Some Proposed Methods of Improvement*
Hai Doan-Nguyen and Leila Kosseim

11:30-12:00 *A Qualitative Comparison of Scientific and Journalistic Texts from the Perspective of Extracting Definitions*
Igal Gabbay and Richard F.E. Sutcliffe

Lunch Break

13:50-14:20 *BioGrapher: Biography Questions as a Restricted Domain Question Answering Task*
Oren Tsur, Maarten de Rijke and Khalil Sima'an

14:20-14:50 *Cooperative Question Answering in Restricted Domains: the WEBCOOP Experiment*
Farah Benamara

14:50-15:20 *A Practical QA System in Restricted Domains*
Hoojung Chung, Young-In Song, Kyoung-Soo Han, Do-Sang Yoon, Joo-Young Lee, Hae-Chang Rim and Soo-Hong Kim

Coffee Break

15:50-16:20 *Answering Questions in the Genomics Domain*
Fabio Rinaldi, James Dowdall, Gerold Schneider and Andreas Persidis

16:20-16:50 *Analysis of Semantic Classes in Medical Text for Question Answering*
Yun Niu and Graeme Hirst

16:50-17:15 Closing Words

THIS IS A BLANK PAGE PLEASE IGNORE

Author Index