ACL 2012

**50th Annual Meeting of the
Association for Computational Linguistics**

**Proceedings of the Workshop on Detecting Structure in
Scholarly Discourse**

Preface

As this year is a celebration of the 50th ACL conference, we are delighted to be able to include the work presented in the first Workshop on Detecting Structure in Scholarly Discourse (DSSD) as part of these 50th anniversary proceedings.

Discourse structure, as a field of research within computational linguistics, is attracting renewed research interest, due to its increasing relevance to diverse fields such as bio-medical text analysis, ethnography, and scientific publishing. Much effort is directed at detecting and modeling a range of discourse elements at different levels of granularity and for different purposes. Such elements include: the statement of facts, claims, and hypotheses; the identification of methods and protocols; and the detection of novelty in contrast to the re-stating of previous existing work. More ambitious long-term goals include the modeling of argumentation, rhetorical structure, and narrative structure.

A broad variety of approaches and of features are used to identify discourse elements, including verb tense/mood/voice, semantic verb class, speculative language or negation, various classes of stance markers, text-structural components, or the location of references. The choice of features is often motivated by linguistic inquiry into the detection of subjectivity, opinion, entailment, inference, as well as author stance, author disagreement, motif and focus.

Six submissions were selected for presentation at the workshop. The submissions represent three fundamental perspectives of research concerning discourse structure: taxonomy and annotation, exploiting cross-document structure in text mining, and detecting discourse elements in scholarly texts. Further development of discourse models and of systems is likely to bring together and integrate aspects from all three. At the same time, these three perspectives give rise to interesting contrasts and different research questions, for instance: Are explicit taxonomies and annotation levels necessary for text mining and for the identification of particular types of discourse elements? or, more generally: How do these different perspectives all relate to a central theory of discourse? The workshop aims to be a forum for discussion of these exciting questions.

Along with our fellow workshop organizers Anita de Waard, Agnes Sandor, and Sophia Ananiadou, we would like to thank all the authors for the hard work that they put into their submissions. We are grateful to the members of the program committee for their thorough reviews. Special thanks go out to Eduard Hovy for his support, and to the ACL-2012 workshop co-chairs Massimo Poesio and Satoshi Sekine for their help with administrative matters. We also gratefully acknowledge the AMICUS (Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts) network for endorsing and supporting the workshop.

Antal van den Bosch, Nijmegen, and Hagit Shatkay, Delaware
May 25, 2012

Organizers:

Sophia Ananiadou, School of Computer Science, University of Manchester, UK
Antal van den Bosch, Centre for Language Studies, Radboud University Nijmegen, The Netherlands
Ágnes Sándor, Xerox Research Europe, Grenoble, France
Hagit Shatkay, Dept. of Computer and Information Sciences, University of Delaware, USA
Anita de Waard, Elsevier Labs, USA

Program Committee:

Catherine Blake, University of Illinois at Urbana-Champaign, USA
Kevin Cohen, University of Colorado, USA
Nigel Collier, National Institute of Informatics, Japan
Walter Daelemans, University of Antwerp, Belgium
Robert Dale, Macquarie University, Australia
Kjersti Fløttum, University of Bergen, Norway
Rocana Girju, University of Illinois at Urbana-Champaign, USA
Lynette Hirschman, MITRE, USA
Halil Kilicoglu, Concordia University, Canada
Jin-Dong Kim, The University Of Tokyo, Japan
Anna Korhonen, Cambridge University, UK
Maria Liakata, Aberystwyth University, UK
Roser Morante, University of Antwerp, Belgium
Raheel Nawaz, University of Manchester, UK
Dragomir Radev, University of Michigan, USA
Dietrich Rebholz-Schuhmann, EBI, UK
Andrey Rzhetsky, University of Chicago, USA
Caroline Sporleder, Saarland University, Germany
Padmini Srinivasan, University of Iowa, USA
Simone Teufel, University of Cambridge, UK
Paul Thompson, University of Manchester, UK
Jun'ichi Tsujii, Microsoft Research Asia, China
Lucy Vanderwende, Microsoft Research, USA

Table of Contents

Conference Program

Thursday, July 12, 2012

Session 1: Exploiting Discourse Structure

9:00–9:45   *Identifying Comparative Claim Sentences in Full-Text Scientific Articles*
Dae Hoon Park and Catherine Blake

9:45–10:30   *Identifying Claimed Knowledge Updates in Biomedical Research Articles*
Ágnes Sándor and Anita de Waard

10:30–11:00   Coffee Break

Session 2: Detecting Discourse Elements

11:00–11:45   *Detection of Implicit Citations for Sentiment Detection*
Awais Athar and Simone Teufel

11:45–12:30   *Open-domain Anatomical Entity Mention Detection*
Tomoko Ohta, Sampo Pyysalo, Jun'ichi Tsujii and Sophia Ananiadou

12:30–2:00   Lunch

Session 3: Taxonomies and Annotation

2:00–2:45   *A Three-Way Perspective on Scientific Discourse Annotation for Knowledge Extraction*
Maria Liakata, Paul Thompson, Anita de Waard, Raheel Nawaz, Henk Pander Maat and Sophia Ananiadou

2:45–3:30   *Epistemic Modality and Knowledge Attribution in Scientific Discourse: A Taxonomy of Types and Overview of Features*
Anita de Waard and Henk Pander Maat

3:30–4:00   Coffee Break

4:00–5:00   Discussion between the authors on detecting and using discourse structure for scholarly text

5:00–5:30   Wrapup and close