2010 Workshop on Applications of Tree Automata in Natural Language Processing

Proceedings of the Workshop

16 July 2010
Uppsala University
Uppsala, Sweden
Preface

We are pleased to present the proceedings of the ACL 2010 Workshop on Applications of Tree Automata in Natural Language Processing, which will be held on July 16 in Uppsala, Sweden, following the 48th Annual Meeting of the Association for Computational Linguistics (ACL).

The theory of tree automata has always had a close connection with natural language processing. In the 1960s, computational linguistics was the major driving force for the development of a theory of tree automata. However, the number of successful applications of this theory to natural language processing remained small during the 20th century. This situation has now changed. Applications of tree automata in natural language processing can be found in work on topics as diverse as grammar formalisms, computational semantics, language generation, and machine translation. Researchers in natural language processing have recognized the usefulness of tree automata theory for solving the problems they are interested in, and theorists are inspired by the resulting theoretical questions.

The goals of this workshop are to provide a dedicated venue for the presentation of work that relates the theory of tree automata to natural language processing, and to create a forum where researchers from the two areas can meet and exchange ideas. Specifically, the workshop aims at raising the awareness for theoretical results useful for applications in natural language processing, and at identifying open theoretical problems raised by such applications.

We are very happy that Kevin Knight (ISI/University of Southern California, USA), certainly the most recognized ambassador for tree automata techniques in machine translation, agreed to give one of his stimulating invited lectures.

For the workshop, authors were invited to submit full papers and proposals for quickfire presentations, the latter being a means for triggering discussions and an exchange of ideas. After a thorough reviewing process with three reviews per full paper, six full papers and four quickfire presentations were accepted for the workshop. The quality and diversity of the papers accepted ensures an interesting and inspiring workshop that we look forward to.

We thank the members of the program committee for their support, and in particular for being careful reviewers of the papers submitted. Furthermore, we would like to thank the program chairs, Sandra Carberry and Stephen Clark, as well as the workshop chairs, Pushpak Bhattacharyia and David Weir, for their friendly and professional assistance.

We hope that all participants of the workshop will experience an inspiring event characterized by curiosity and an open-minded atmosphere, and that all readers of these proceedings will gain new insights that make a difference.

Frank Drewes
Marco Kuhlmann
May 2010
Organizers:
Frank Drewes, Umeå University, Sweden
Marco Kuhlmann, Uppsala University, Sweden

Program Committee:
Parosh Aziz Abdulla, Uppsala University, Sweden
Leonor Becerra-Bonache, Yale University, USA
Chris Callison-Burch, Johns Hopkins University, USA
David Chiang, ISI/University of Southern California, USA
Loek Cleophas, University of Pretoria, South Africa, and Eindhoven University of Technology, the Netherlands
Trevor Cohn, University of Sheffield, UK
François Denis, Université de Provence, France
Johanna Högberg, Umeå University, Sweden
Liang Huang, ISI/University of Southern California, USA
Stephan Kepser, codecentric AG, Germany
Alexander Koller, Saarland University, Germany
Andreas Maletti, Universitat Rovira i Virgili, Spain
Sebastian Maneth, NICTA and University of New South Wales, Australia
Jonathan May, ISI/University of Southern California, USA
Brink van der Merwe, University of Stellenbosch, South Africa
Mark-Jan Nederhof, University of St Andrews, UK
Joachim Niehren, INRIA, France
Kai Salomaa, Queen’s University, Canada
Anoop Sarkar, Simon Fraser University, Canada
Giorgio Satta, University of Padua, Italy
Stuart Shieber, Harvard University, USA
Magnus Steinby, University of Turku, Finland
Marc Tommasi, INRIA, France
Heiko Vogler, Technische Universität Dresden, Germany

Invited Speaker:
Kevin Knight, ISI/University of Southern California, USA
# Table of Contents

*Preservation of Recognizability for Synchronous Tree Substitution Grammars*
  Zoltán Fülop, Andreas Maletti and Heiko Vogler ............................................. 1

*A Decoder for Probabilistic Synchronous Tree Insertion Grammars*
  Steve DeNeefe, Kevin Knight and Heiko Vogler ............................................. 10

*Parsing and Translation Algorithms Based on Weighted Extended Tree Transducers*
  Andreas Maletti and Giorgio Satta ................................................................. 19

*Millstream Systems – a Formal Model for Linking Language Modules by Interfaces*
  Suna Bensch and Frank Drewes ................................................................. 28

*Transforming Lexica as Trees*
  Mark-Jan Nederhof ................................................................. 37

*n-Best Parsing Revisited*
  Matthias Büchse, Daniel Geisler, Torsten Stüber and Heiko Vogler ............. 46
Workshop Program

Friday, July 16, 2010

09:00–09:15   Opening Remarks
09:15–10:30   Invited Talk by Kevin Knight
10:30–11:00   Coffee Break

**Full Paper Session 1**

11:00–11:30   *Preservation of Recognizability for Synchronous Tree Substitution Grammars*
               Zoltán Fülöp, Andreas Maletti and Heiko Vogler

11:30–12:00   *A Decoder for Probabilistic Synchronous Tree Insertion Grammars*
               Steve DeNeefe, Kevin Knight and Heiko Vogler

12:00–12:30   *Parsing and Translation Algorithms Based on Weighted Extended Tree Transducers*
               Andreas Maletti and Giorgio Satta

12:30–14:00   Lunch Break

**Full Paper Session 2**

14:00–14:30   *Millstream Systems – a Formal Model for Linking Language Modules by Interfaces*
               Suna Bensch and Frank Drewes

14:30–15:00   *Transforming Lexica as Trees*
               Mark-Jan Nederhof

15:00–15:30   *n-Best Parsing Revisited*
               Matthias Büchse, Daniel Geisler, Torsten Stüber and Heiko Vogler

15:30–16:00   Coffee Break
Friday, July 16, 2010 (continued)

**Quickfire Presentations**

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>16:00–16:15</td>
<td><em>Tree Automata Techniques and the Learning of Semantic Grammars</em></td>
<td>Michael Minock</td>
</tr>
<tr>
<td>16:15–16:30</td>
<td><em>Do We Really Want a Single Tree to Cover the Whole Sentence?</em></td>
<td>Aravind Joshi</td>
</tr>
<tr>
<td>16:30–16:45</td>
<td><em>The Tree Automata Workbench ‘Marbles’</em></td>
<td>Frank Drewes</td>
</tr>
<tr>
<td>16:45–17:00</td>
<td><em>Requirements on a Tree Transformation Model for Machine Translation</em></td>
<td>Andreas Maletti</td>
</tr>
<tr>
<td>17:00–17:30</td>
<td><strong>Discussion</strong></td>
<td></td>
</tr>
</tbody>
</table>