

FSMNLP 2013

Proceedings of the

11th

International Conference

on

Finite State Methods and

Natural Language

Processing

July 15–17, 2013
University of St Andrews
St Andrews, Scotland

Sponsors:



University of
St Andrews

600
YEARS

sicsa* The Scottish Informatics &
Computer Science Alliance



Scotland's National Tourism Organisation
Buidheann Turasachd Nàiseanta na h-Alba

©2013 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

These proceedings contain the papers presented at the 11th International Conference on Finite-State Methods and Natural Language Processing (FSMNLP 2013), held in St Andrews, Scotland (UK), July 15–17, 2013.

This series of conferences is the premier forum of the ACL Special Interest Group on Finite-State Methods (SIGFSM). It serves researchers and practitioners working on:

- natural language processing (NLP) applications or language resources,
- theoretical and implementational aspects, or
- their combinations

that have obvious relevance or an explicit relation to finite-state methods.

This volume contains the 8 long and 9 short papers presented at the conference. In total, 26 papers (11 long and 15 short papers) were submitted and double-blind refereed. The overall acceptance rate was 69%, noting one paper was withdrawn after acceptance. Each long paper was reviewed by at least 3 programme committee members and each short paper by at least 2. The programme committee was composed of internationally leading researchers and practitioners selected from academia, research labs, and companies.

I would like to thank the programme committee for their hard work, the referees for their valuable feedback, the invited speakers and the presenters of tutorials for their contributions and the many local staff and students for their tireless efforts. We are particularly indebted to the Scottish Informatics & Computer Science Alliance for their financial support. Most generous support was received from VisitScotland.

MARK-JAN NEDERHOF

Chair:

Mark-Jan Nederhof (University of St Andrews)

Local Organizing Committee:

Per Ola Kristensson (University of St Andrews)

Martin McCaffery (University of St Andrews)

Shyam Reyal (University of St Andrews)

Vinodh Rajan (University of St Andrews)

Invited Speakers:

Alexander Clark (King's College London)

Bill Byrne (University of Cambridge)

Tutorials by:

Ruth Hoffmann (University of St Andrews)

Bevan Jones (University of Edinburgh)

Kousha Etesami (University of Edinburgh)

Program Committee:

Iñaki Alegria (University of the Basque Country)
Francisco Casacuberta (Instituto Tecnológico de Informática, Spain)
Jan Daciuk (Gdańsk University of Technology, Poland)
Frank Drewes (Umeå University, Sweden)
Dale Gerdemann (University of Tübingen, Germany)
Mike Hammond (University of Arizona, USA)
Colin de la Higuera (University of Nantes, France)
Mans Hulden (Ikerbasque, Basque Country)
André Kempe (Nuance Communications, Germany)
Marco Kuhlmann (Uppsala University, Sweden)
Andreas Maletti (Universität Stuttgart, Germany)
Kemal Oflazer (Sabanci University, Turkey)
Maite Oronoz Anchordoqui (University of the Basque Country)
Laurette Pretorius (University of South Africa)
Strahil Ristov (Rudjer Boskovic institute, Croatia)
Frederique Segond (Viseo Innovation, France)
Max Silberstein (Université de Franche-Comté, France)
Richard Sproat (Google, USA)
Heiko Vogler (Technical University Dresden, Germany)
Anssi Yli-Jyrä (University of Helsinki, Finland)
Menno Van Zaanen (Tilburg University, Netherlands)
Lynette Van Zijl (Stellenbosch University, South Africa)

Additional reviewers:

Fabienne Braune (Universität Stuttgart, Germany)
Matthias Büchse (Technical University Dresden, Germany)
Jean-Christophe Janodet (University of Evry, France)
Johannes Osterholzer (Technical University Dresden, Germany)
Nina Seemann (Universität Stuttgart, Germany)

Table of Contents

<i>Computing the Most Probable String with a Probabilistic Finite State Machine</i>	
Colin de la Higuera and Jose Oncina	1
<i>Stochastic Bi-Languages to model Dialogs</i>	
M. Inés Torres	9
<i>ZeuScansion: a tool for scansion of English poetry</i>	
Manex Agirrezabal, Bertol Arrieta, Aitzol Astigarraga and Mans Hulden	18
<i>A Convexity-based Generalization of Viterbi for Non-Deterministic Weighted Automata</i>	
Marc Dymetman	25
<i>Processing Structured Input with Skipping Nested Automata</i>	
Dominika Pawlik, Aleksander Zabłocki and Bartosz Zaborowski	30
<i>Synchronous Regular Relations and Morphological Analysis</i>	
Christian Wurm and Younes Samih	35
<i>Parsing Morphologically Complex Words</i>	
Kay-Michael Würzner and Thomas Hanneforth	39
<i>Optimizing Rule-Based Morphosyntactic Analysis of Richly Inflected Languages - a Polish Example</i>	
Dominika Pawlik, Aleksander Zabłocki and Bartosz Zaborowski	44
<i>Finite State Morphology Tool for Latvian</i>	
Daiga Deksnė	49
<i>Modeling Graph Languages with Grammars Extracted via Tree Decompositions</i>	
Bevan Keeley Jones, Sharon Goldwater and Mark Johnson	54
<i>Finite State Methods and Description Logics</i>	
Tim Fernando	63
<i>Using NooJ for semantic annotation of Italian language corpora in the domain of motion: a cognitive-grounded approach</i>	
Edoardo Salza	72
<i>Multi-threaded composition of finite-state-automata</i>	
Bryan Jurish and Kay-Michael Würzner	81
<i>On Finite-State Tonology with Autosegmental Representations</i>	
Anssi Yli-Jyrä	90
<i>A Finite-State Approach to Translate SNOMED CT Terms into Basque Using Medical Prefixes and Suffixes</i>	
Olatz Perez-de-Vinaspre, Maite Oronoz, Manex Agirrezabal and Mikel Lersundi	99

<i>Syncretism and How to Deal with it in a Morphological Analyzer: a German Example</i>	
Katina Bontcheva	104
<i>Finite State Approach to the Kazakh Nominal Paradigm</i>	
Bakyt M Kairakbay and David L Zaurbekov	108

Conference Program

Sunday, July 14, 2013

18:00–19:00 Wine Reception

Monday, July 15, 2013

8:30–9:20 Registration

9:20–9:30 Opening

9:30–10:30 Keynote Lecture: Alexander Clark: Towards a theory of context-free languages

10:30–11:00 Break

11:00–11:30 *Computing the Most Probable String with a Probabilistic Finite State Machine*
Colin de la Higuera and Jose Oncina

11:30–12:00 *Stochastic Bi-Languages to model Dialogs*
M. Inés Torres

12:00–12:30 *Zeuscansion: a tool for scansion of English poetry*
Manex Agirrezabal, Bertol Arrieta, Aitzol Astigarraga and Mans Hulden

12:30–13:15 Lunch

13:15–14:00 Business Meeting

14:00–14:20 *A Convexity-based Generalization of Viterbi for Non-Deterministic Weighted Automata*
Marc Dymetman

14:20–14:40 *Processing Structured Input with Skipping Nested Automata*
Dominika Pawlik, Aleksander Zabłocki and Bartosz Zaborowski

14:40–15:00 *Synchronous Regular Relations and Morphological Analysis*
Christian Wurm and Younes Samih

Monday, July 15, 2013 (continued)

15:00–15:30 Break/Posters/Demos

15:30–16:30 Tutorial: Ruth Hoffmann: Experimenting with finite state automata in GAP

Tuesday, 16th July 2013

09:00–10:00 Keynote Lecture: Bill Byrne: Pushdown Automata in Statistical Machine Translation

10:00–10:20 *Parsing Morphologically Complex Words*
Kay-Michael Würzner and Thomas Hanneforth

10:20–10:40 *Optimizing Rule-Based Morphosyntactic Analysis of Richly Inflected Languages - a Polish Example*
Dominika Pawlik, Aleksander Zablocki and Bartosz Zaborowski

10:40–11:00 *Finite State Morphology Tool for Latvian*
Daiga Dekšne

11:00–11:30 Break/Posters/Demos

11:30–12:00 *Modeling Graph Languages with Grammars Extracted via Tree Decompositions*
Bevan Keeley Jones, Sharon Goldwater and Mark Johnson

12:00–12:30 *Finite State Methods and Description Logics*
Tim Fernando

12:30–13:00 *Using NooJ for semantic annotation of Italian language corpora in the domain of motion: a cognitive-grounded approach*
Edoardo Salza

13:00–14:00 Lunch

14:15–16:30 Guided Walk and Visit of MUSA

18:15–21:00 Conference Dinner and Concert

Wednesday, 16th July 2013

- 9:00-11:00 Tutorial: Bevan Jones: Variational inference for tree automata and tree transducers
- 11:00-11:30 Break
- 11:30-12:00 *Multi-threaded composition of finite-state-automata*
Bryan Jurish and Kay-Michael Würzner
- 12:00-12:30 *On Finite-State Tonology with Autosegmental Representations*
Anssi Yli-Jyrä
- 12:30-13:30 Lunch
- 13:30-13:50 *A Finite-State Approach to Translate SNOMED CT Terms into Basque Using Medical Prefixes and Suffixes*
Olatz Perez-de-Vinaspre, Maite Oronoz, Manex Agirrezabal and Mikel Lersundi
- 13:50-14:10 *Syncretism and How to Deal with it in a Morphological Analyzer: a German Example*
Katina Bontcheva
- 14:10-14:30 *Finite State Approach to the Kazakh Nominal Paradigm*
Bakyt M Kairakbay and David L Zaurbekov
- 14:30-15:00 Break/Posters/Demos
- 15:00-17:00 Tutorial: Kousha Etessami: Algorithms and complexity of analyzing unrestricted stochastic context-free grammars
- 17:00-17:10 Closing

