

FSMNLP 2012

Proceedings of the

10th

International Workshop on

Finite State Methods and

Natural Language

Processing

July 23–25, 2012
University of the Basque Country
Donostia-San Sebastián

Sponsors:



©2012 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

Preface

These proceedings contain the papers presented at the 10th International Workshop on Finite State Methods and Natural Language Processing (FSMNLP 2012), held in Donostia-San Sebastián (Basque Country), July 23–25, 2012.

The workshop covers a wide range of topics from morphology to formal language theory. A special theme was chosen for FSMNLP 2012: “practical issues in finite-state technology,” including:

- Practical implementations of linguistic descriptions with finite-state technology
- Software tools and utilities for finite-state NLP
- Finite-state models of linguistic theories
- Applications of finite-state-based NLP in closely related fields

This volume contains the 7 long and 12 short papers presented at the workshop. In total, 31 papers (13 long and 18 short papers) were submitted and double-blind refereed. Each paper was reviewed by 3 program committee members. The overall acceptance rate was 61 per cent.

The program committee was composed of internationally leading researchers and practitioners selected from academia, research labs, and companies.

The organizing committee would like to thank the program committee for their hard work, the referees for their valuable feedback, the invited speaker and the presenters of tutorials for their contributions and the local organizers for their tireless efforts. We are particularly indebted to the University of the Basque Country (UPV/EHU) and the Basque Government (Eusko Jauriaritza) for significant financial support and to the Cursos de Verano/Uda Ikastaroak and the IXA research group for their support in organizing the event.

IÑAKI ALEGRIA
MANS HULDEN

Local Organizing Committee:

Iñaki Alegria (University of the Basque Country)
Koldo Gojenola (University of the Basque Country)
Izaskun Etxeberria (University of the Basque Country)
Nerea Ezeiza (University of the Basque Country)
Mans Hulden (University of the Basque Country / Ikerbasque, Basque Foundation for Science)
Amaia Lorenzo (University of the Basque Country)
Esther Miranda (University of the Basque Country)
Maite Oronoz (University of the Basque Country)

Invited Speaker:

Kimmo Koskenniemi (University of Helsinki)

Tutorials by:

Tommi Pirinen (University of Helsinki)
Iñaki Alegria (University of the Basque Country)
Mans Hulden (University of the Basque Country / Ikerbasque, Basque Foundation for Science)
Miikka Silfverberg (University of Helsinki)
Josef Novak (University of Tokyo)

Program Committee:

Kenneth R. Beesley (SAP Business Objects, USA)
Francisco Casacuberta (Instituto Tecnológico De Informática, Spain)
Jan Daciuk (Gdańsk University of Technology, Poland)
Frank Drewes (Umea University, Sweden)
Dale Gerdemann (University of Tuebingen, Germany)
Mike Hammond (University of Arizona, USA)
Thomas Hanneforth (University of Potsdam, Germany)
Colin de la Higuera (University of Nantes, France)
Jan Holub (Czech Technical University in Prague, Czech Republic)
Mans Hulden (Ikerbasque, Basque Country)
André Kempe (CADEGE Technologies & Consulting, France)
Andras Kornai (Eotvos Lorand University, Hungary)
Andreas Maletti (University of Stuttgart, Germany)
Mark-Jan Nederhof (University of St Andrews, Scotland)
Kemal Oflazer (Carnegie Mellon University, Qatar)
Maite Oronoz (University of the Basque Country)
Laurette Pretorius (University of South Africa, South Africa)
Strahil Ristov (Ruder Boskovic Institute, Croatia)
Frederique Segond Frederique (ObjectDirect, France)
Max Silberztein (Université de Franche-Comte, France)
Richard Sproat (University of Illinois at Urbana-Champaign, USA)
Trond Trosterud (University of Tromsø, Norway)
Shuly Wintner (University of Haifa, Israel)
Anssi Yli-Jyra (University of Helsinki, Finland)
Menno van Zaanen (Tilburg University, Netherlands)
Lynette van Zijl (Stellenbosch University, South Africa)

Additional Reviewers:

Alicia Pérez (University of the Basque Country)
Suna Bensch (Umeå University, Sweden)
Emad Mohamed (Carnegie Mellon University, Qatar)
Johanna Björklund (Umeå University, Sweden)

Table of Contents

<i>Effect of Language and Error Models on Efficiency of Finite-State Spell-Checking and Correction</i> Tommi A. Pirinen and Sam Hardwick	1
<i>Practical Finite State Optimality Theory</i> Dale Gerdemann and Mans Hulden	10
<i>Handling Unknown Words in Arabic FST Morphology</i> Khaled Shaalan and Mohammed Attia	20
<i>Urdu - Roman Transliteration via Finite State Transducers</i> Tina Bögel	25
<i>Integrating Aspectually Relevant Properties of Verbs into a Morphological Analyzer for English</i> Katina Bontcheva	30
<i>Finite-State Technology in a Verse-Making Tool</i> Manex Agirrezabal, Iñaki Alegria, Bertol Arrieta and Mans Hulden	35
<i>DAGGER: A Toolkit for Automata on Directed Acyclic Graphs</i> Daniel Quernheim and Kevin Knight	40
<i>WFST-Based Grapheme-to-Phoneme Conversion: Open Source tools for Alignment, Model-Building and Decoding</i> Josef R. Novak, Nobuaki Minematsu and Keikichi Hirose	45
<i>Kleene, a Free and Open-Source Language for Finite-State Programming</i> Kenneth R. Beesley	50
<i>Implementation of Replace Rules Using Preference Operator</i> Senka Drobac, Miikka Silfverberg and Anssi Yli-Jyrä	55
<i>First Approaches on Spanish Medical Record Classification Using Diagnostic Term to Class Transduction</i> A. Casillas, A. Díaz de Ilarraza, K. Gojenola, M. Oronoz and Alicia Pérez	60
<i>Developing an Open-Source FST Grammar for Verb Chain Transfer in a Spanish-Basque MT System</i> Aingeru Mayor, Mans Hulden and Gorka Labaka	65
<i>Conversion of Procedural Morphologies to Finite-State Morphologies: A Case Study of Arabic</i> Mans Hulden and Younes Samih	70
<i>A Methodology for Obtaining Concept Graphs from Word Graphs</i> Marcos Calvo, Jon Ander Gómez, Lluís-F. Hurtado and Emilio Sanchis	75
<i>A Finite-State Temporal Ontology and Event-Intervals</i> Tim Fernando	80

<i>A Finite-State Approach to Phrase-Based Statistical Machine Translation</i> Jorge González	90
<i>Finite-State Acoustic and Translation Model Composition in Statistical Speech Translation: Empirical Assessment</i> Alicia Pérez, M. Inés Torres and Francisco Casacuberta	99
<i>Refining the Design of a Contracting Finite-State Dependency Parser</i> Anssi Yli-Jyrä, Jussi Piitulainen and Atro Voutilainen	108
<i>Lattice-Based Minimum Error Rate Training Using Weighted Finite-State Transducers with Tropical Polynomial Weights</i> Aurelien Waite, Graeme Blackwood and William Byrne	116

Conference Program

Monday, July 23rd, 2012

15:00–18:00 TUTORIALS

15:00–16:00 Tommi Pirinen and Mans Hulden: Spelling and Grammar Correction with FSTs

16:00–17:00 Miikka Silfverberg: Probabilistic Parsing with Weighted FSTs

17:00–18:00 Josef Novak: Grapheme-to-Phoneme Training and Conversion with Weighted FSTs

Tuesday, July 24th, 2012

9:00 Opening

9:30–10:30 Invited Speaker: Kimmo Koskenniemi: The Simplicity of Two-Level Morphology

10:30–11:00 Coffee Break

11:00–12:00 LONG PAPERS I

Effect of Language and Error Models on Efficiency of Finite-State Spell-Checking and Correction

Tommi A. Pirinen and Sam Hardwick

Practical Finite State Optimality Theory

Dale Gerdemann and Mans Hulden

12:00–13:00 SHORT PAPERS I

Handling Unknown Words in Arabic FST Morphology

Khaled Shaalan and Mohammed Attia

Urdu – Roman Transliteration via Finite State Transducers

Tina Bögel

Integrating Aspectually Relevant Properties of Verbs into a Morphological Analyzer for English

Katina Bontcheva

Finite-State Technology in a Verse-Making Tool

Manex Agirrezabal, Iñaki Alegria, Bertol Arrieta and Mans Hulden

13:00–14:30 Lunch

Tuesday, July 24th, 2012 (continued)

14:30–15:30 SHORT PAPERS II

DAGGER: A Toolkit for Automata on Directed Acyclic Graphs

Daniel Quernheim and Kevin Knight

WFST-based Grapheme-to-Phoneme conversion: Open Source Tools for Alignment, Model-Building and Decoding

Josef R. Novak, Nobuaki Minematsu and Keikichi Hirose

Kleene, a Free and Open-Source Language for Finite-State Programming

Kenneth R. Beesley

Implementation of Replace Rules Using Preference Operator

Senka Drobac, Miikka Silfverberg and Anssi Yli-Jyrä

15:30–16:30 SHORT PAPERS III

First Approaches on Spanish Medical Record Classification Using Diagnostic Term to Class Transduction

A. Casillas, A. Díaz de Ilarraza, K. Gojenola, M. Oronoz and A. Pérez

Developing an Open-Source FST Grammar for Verb Chain Transfer in a Spanish-Basque MT System

Aingeru Mayor, Mans Hulden and Gorka Labaka

Conversion of Procedural Morphologies to Finite-State Morphologies: A Case Study of Arabic

Mans Hulden and Younes Samih

A Methodology for Obtaining Concept Graphs from Word Graphs

Marcos Calvo, Jon Ander Gómez, Lluís-F. Hurtado and Emilio Sanchis

16:30–18:00 POSTERS AND COFFEE

20:30 Dinner in the Old Town

Tuesday, July 25th, 2012

9:30–11:00 LONG PAPERS II

A Finite-State Temporal Ontology and Event-Intervals

Tim Fernando

A Finite-State Approach to Phrase-Based Statistical Machine Translation

Jorge González

Finite-State Acoustic and Translation Model Composition in Statistical Speech Translation: Empirical Assessment

Alicia Pérez, M. Inés Torres and Francisco Casacuberta

11:30–12:30 LONG PAPERS III

Refining the Design of a Contracting Finite-State Dependency Parser

Anssi Yli-Jyrä, Jussi Piitulainen and Atro Voutilainen

Lattice-Based Minimum Error Rate Training using Weighted Finite-State Transducers with Tropical Polynomial Weights

Aurelien Waite, Graeme Blackwood and William Byrne

12:30–13:30 Business Meeting

