Theory and Applications
of Natural Language Processing

Aims and Scope

The field of Natural Language Processing (NLP) has expanded explosively over the past decade: growing bodies of available data, novel fields of applications, emerging areas and new connections to neighboring fields have all led to increasing output and to diversification of research.

"Theory and Applications of Natural Language Processing" is a series of volumes dedicated to selected topics in NLP and Language Technology. It focuses on the most recent advances in all areas of the computational modeling and processing of speech and text across languages and domains. Due to the rapid pace of development, the diversity of approaches and application scenarios are scattered in an ever-growing mass of conference proceedings, making entry into the field difficult for both students and potential users. Volumes in the series facilitate this first step and can be used as a teaching aid, advanced-level information resource or a point of reference.

The series encourages the submission of research monographs, contributed volumes and surveys, lecture notes and textbooks covering research frontiers on all relevant topics, offering a platform for the rapid publication of cutting-edge research as well as for comprehensive monographs that cover the full range of research on specific problem areas.

The topics include applications of NLP techniques to gain insights into the use and functioning of language, as well as the use of language technology in applications that enable communication, knowledge management and discovery such as natural language generation, information retrieval, question-answering, machine translation, localization and related fields.

The books are available in printed and electronic (e-book) form:

* Downloadable on your PC, e-reader or iPad
* Enhanced by Electronic Supplementary Material, such as algorithms, demonstrations, software, images and videos
* Available online within an extensive network of academic and corporate R&D libraries worldwide
* Never out of print thanks to innovative print-on-demand services
* Competitively priced print editions for eBook customers thanks to MyCopy service http://www.springer.com/librarians/e-content/mycopy

For other titles published in this series, go to
www.springer.com/series/8899

Peter Spyns • Jan Odijk
Editors

Essential Speech and Language Technology for Dutch

Results by the STEVIN programme

Springer

Editors
Peter Spyns
Nederlandse Taalunie
The Hague
The Netherlands

Jan Odijk
UiL-OTS
University of Utrecht
Utrecht
The Netherlands

Foreword by
Linde van den Bosch
Nederlandse Taalunie
The Hague
The Netherlands

ISSN 2192-032X
ISBN 978-3-642-30909-0
DOI 10.1007/978-3-642-30910-6
Springer Heidelberg New York Dordrecht London

ISSN 2192-0338 (electronic)
ISBN 978-3-642-30910-6 (eBook)

Library of Congress Control Number: 2012955243

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The STEVIN programme was not only an important scientific endeavour in the Low Countries, but also a quite rare case of a tight inter-institutional cross-border collaboration within the Dutch-speaking linguistic area. Four funding agencies, three ministerial departments and one intergovernmental organisation in Flanders and the Netherlands were involved in this programme. STEVIN is an excellent illustration of how a medium European language can set an example in the domain of language (technology) policy.

It remains extremely important that citizens can use their native language in all circumstances, including when they deal with modern ICT and leisure devices. For example, a very recent trend is that devices such as smart-phones and television sets become voice-controlled. But usually English speaking people are the first to benefit from such an evolution; other linguistic communities have to wait – some for ever? Not only does this pose a danger of reducing the overall functionality of a language (and an impoverishment of an entire culture), but also it threatens those groups in society that do not master the universal language. For example, elderly or disabled people, who deserve most to enjoy the blessings of modern technology, are in many cases the last ones to benefit from it. Therefore, R&D programmes that support the local language are needed. Also in the future, the Dutch Language Union will continue to emphasise this issue.

Many individuals have contributed to make STEVIN a success story, of all which I sincerely want to thank for their commitment. A particular mention goes to the funding government organisations from the Netherlands and Flanders.

I am confident that the STEVIN results will boost research in academia and technology development in industry so that the Dutch language can continue to "serve" its speakers well under all circumstances. Hence, it is with great pleasure that I invite you to discover the scientific results of the STEVIN programme.

The Hague, The Netherlands

Linde van den Bosch
General Secretary of the
Dutch Language Union (2004–2012)

Preface

Summarising a research programme that lasted for more than 6 years is a demanding task due to the wealth of deliverables, publications and final results of each of the projects concerned. In addition to the content-related topics, which interest scientists, research programmes also lead to new insights for policy makers and programme managers. The former want to discover advances in the state of the art, while the latter are eager to learn good practices in programme governance and management.

The STEVIN programme is no exception. In this work, the collaborators of each STEVIN R&D project have selected and summarised their scientific achievements. Even though the scientific accomplishments are the main focus of this volume, we have also added descriptions of some other particular aspects of the programme as a whole, such as its rationale, IPR management and the main conclusions of its final evaluation.

This volume is the result of a great deal of dedicated and hard work by many individuals, who, unfortunately, we cannot all mention by name as the list would be too long. We would first like to thank our colleagues of the Nederlandse Taalunie (NTU – Dutch Language Union), the members of the HLT steering board and the STEVIN programme office, the participants of the various STEVIN committees and related working groups, the project collaborators for their dedicated work and, of course, the funding organisations.

Additionally, we gratefully acknowledge everyone who has been involved in creating this volume. There are the authors of the various chapters. Also, the following members of the STEVIN international assessment panel (IAP) were so kind to, in addition to project proposals earlier on, review contributions to this volume as their last official duty for STEVIN:

- Gilles Adda – LIMSI (Paris)
- Nicoletta Calzolari – ILC (Pisa)
- Paul Heisterkamp – DaimlerChrysler (Ulm)
- Stelios Piperidis (& Sotiris Karabetsos) – ILSP (Athens)
- Gábor Prószéky – Morphologic (Budapest)

For this latter task, much-appreciated help came from the following internationally renowned researchers:

- Etienne Barnard (& Marelie Davel) – CSIR (Pretoria)
- Núria Bel – IULA (Barcelona)
- Nick Campbell – TCD (Dublin)
- Thierry Declerck – DFKI (Saarbrücken)
- Koenraad De Smedt – UIB (Bergen)
- Cédric Fairon – CENTAL (Louvain-la-Neuve)
- Steven Krauwer – UiL-OTS (Utrecht)
- Bente Maegaard – CST (Copenhague)
- Wim Peters (& Diana Maynard) – DCS-NLPG (Sheffield)
- Louis Pols – UvA (Amsterdam)
- Laurette Pretorius – UNISA (Pretoria)
- Steve Renals – ILCC (Edinburg)
- Justus Roux – CTEXT (Potchefstroom)
- Khalil Sima'an – UvA-ILLC (Amsterdam)
- Dan Tufiş – RACAI (Bucarest)
- Josef van Genabith – DCU (Dublin)
- Gerhard van Huyssteen – NWU (Potchefstroom)
- Werner Verhelst – ETRO-DSSP (Brussels)

Finally, we are also indebted to Springer-Verlag's editorial staff for their help, namely Dr. Olga Chiarcos and, in particular, Mrs. Federica Corradi Dell'Acqua.

It is our sincere hope and conviction that this volume will be of great interest to an international audience of researchers in human language technologies (HLT), in particular those who work on Dutch, to government officials active in HLT or language policy and to funders of science, technology and innovation programmes in general.

The STEVIN[1] programme was funded by the Flemish and Dutch governments (www.stevin-tst.org). Its results are presented at (www.stevin-tst.org/etalage) and are available via the HLT Agency (www.tst-centrale.org).

The Hague, The Netherlands — Peter Spyns
STEVIN programme coordinator

Utrecht, The Netherlands — Jan Odijk
Chair of the STEVIN programme committee

[1] STEVIN stands for 'Essential Speech and Language Technology Resources'. In addition, Simon Stevin was a seventeenth century applied scientist who, amongst other things, introduced Dutch terms for mathematics and physics concepts. He worked both in Flanders and the Netherlands. Hence, his name is a perfect acronym for this joint programme. And he became famous for building a land yacht for Prince Maurice of Orange.

Acknowledgements

All projects reported on in this volume were funded by the STEVIN programme. STEVIN was organised under the auspices of the Dutch Language Union and jointly financed by the Flemish and Dutch governments. The Flemish government was represented by the Department of Economy, Science and Innovation (EWI), the Agency for Innovation (IWT), and the Research Foundation – Flanders (FWO). The Dutch government was represented by the Ministry for Economy, Agriculture and Innovation (E,L&I), the Ministry of Education, Culture and Science (OCW), the Netherlands Organisation for Research (NWO) and the Agency NL.

Contents