From Research to Commercial Applications: Making NLP Work in Practice

Proceedings of a Workshop
Sponsored by the
Association for Computational Linguistics

Edited by
Jill Burstein
and
Claudia Leacock

12 July 1997
Universidad Nacional de Educación a Distancia
Madrid, Spain

From Research to Commercial Applications: Making NLP Work in Practice

Proceedings of a Workshop
Sponsored by the
Association for Computational Linguistics

Edited by
Jill Burstein
and
Claudia Leacock

12 July 1997
Universidad Nacional de Educación a Distancia
Madrid, Spain

© 1997, Association for Computational Linguistics

Order additional copies from:

ACL
P.O. Box 6090
Somerset, NJ, 08875 USA
+1-908-873-3898
acl@bellcore.com

PREFACE

This volume contains the papers written for and presented at the "From Research to Commercial Applications: Making NLP Technology Work in Practice" workshop. The workshop, held on July 12, 1997 at the Universidad Nacional de Educacion a Distancia in Madrid, Spain, was sponsored by Association for Computational Linguistics and the European Chapter of the Association for Computational Linguistics.

Success in the marketplace is one form of validation for NLP techniques and underlying theories. The purpose of this workshop is to bring together researchers to discuss commercial and commercial-bound systems that use NLP. Our interest is to learn about systems that show promise in re-using NLP techniques, share experiences of the process of technology transfer for NLP applications, and explore industry-based practical considerations involving NLP technology.

The workshop is intended to open a dialogue among the research community about experiences and problems -- technical, logistic, or cultural -- encountered by people working on operational and commercial NLP applications. These include issues in the re-use of domain-specific techniques so that new applications can get leverage from using existing ones. The ability to re-use technology for diverse applications should both give the application a solid grounding, and save time and money. Other issues include technology transfer and industry-based practicalities that often guide the design of NLP technology. These include problems in customization and the trade-off between accuracy and other requirements, such as speed, and ease of use.

The goal of the workshop is to create connections between academic and industry-based research efforts in order to build a solid infrastructure for NLP technology re-use and lead to a deeper understanding of commercial NLP potential.

We wish to thank the members of the programming committee who helped review the large number of excellent papers that were submitted:

Andrew Golding, Mitsubishi Electric
Mary Dee Harris, Language Technology, Inc.
Kevin Knight, USC/ISI
Karen Kukich, Educational Testing Service
Lisa Rau, SRA International
Yael Ravin, IBM, T.J. Watson Research Center
Susanne Wolff, Educational Testing Service
Wlodek Zadrozny, IBM, T.J. Watson Research Center

We are grateful to Harald Trost for making the workshop preparation go smoothly. We thank to Lisa Hemat who handled many of the correspondences with reviewers, and who contributed greatly to the prompt preparation of the workshop proceedings.

Finally, we are extremely grateful to Educational Testing Service for support during all stages of planning and preparation of the workshop

Jill Burstein
Educational Testing Service MS 11R
Princeton, New Jersey 08541
(609) 734- 5823
jburstein@ets.org

Claudia Leacock
Princeton University
221 Nassau Street
Princeton, New Jersey
(609) 258-2824
leacock@clarity.princeton.edu

Final Program

9:00 - 9:30 (Coffee and Registration)

9:30 - 10:00: "Re-Use of a Proper Noun Recognition System in Commercial and Operational NLP Applications" (Chinatsu Aone and J. Maloney, SRA)

10:00 - 10:30: "Developing a new grammar checker for English as a second language" (Cornelia Tschichold, Franck Bodmer, Etienne Cornu, Francois Grosjean, Lysiane Grojean, Natalie Kubler, Nicolas Lewy and Corinne Tschumi, Universite of Neuchatel)

10:30 - 11:00: "Software Re-Use and Evolution in Text Generation Applications" (Karen Kukich and Rebecca Passaneau, Bellcore; Kathleen McKeown, Dragomir Radev, Vasileios Hatzivassiloglou and Hongyan Jing, Columbia University)

11:00 - 11:20: Coffee Break

11:20 - 11:50: "Estimating the True Performance of Classification-Based NLP Technology" (James Nolan, Siena College)

11:50 - 12:20: "Tailored patient information: some issues and questions" (Ehud Reiter and Liesl Osman, University of Aberdeen)

12:20 - 1:30: Lunch Break

1:30 - 2:00: "Practical Considerations in Building a Multi-Lingual Authoring System for Business Letters" (John Tait and Huw Sanderson,University of Sunderland; Jeremy Ellman, MARI Computer Systems, Ana Maria Martinez San Jose, Sistemas y Tratamiento de Informacion, SA; Peter Hellwig, University of Heidelberg; Periklis Tsagheas, SENA, SA)

2:00 - 2:30: "A Language Identification Application Built on the Java Client/Server Platform" (Gary Adams, Sun MicroSystems Labs. and Philip Resnik, University of Maryland)

2:30 - 3:00: "Practical Experience with Grammar Sharing in Multilingual NLP" (Michael Gamon, Carmen Lozano, Jessie Pinkham and Tom Reutter, Microsoft Research)

3:00 - 3:20: Coffee Break

3:20 - 3:50: "NLP and Industry: Transfer and Reuse of Technolgies" (Leo Obrst and Krishna Jha, Boeing Defense and Space Group)

3:50 - 4:20: "Recycling Lingware in a Multilingual MT System" (Manny Rayner and David Carter, SRI; Ivan Bretan, Roberts Eklund and Mats Wiren, Telia Research; Steffen Leo Hansen, Sabine Kirchmeier-Andersen, Finn Sorensen and Hanne Erdman Thomsen,Handelshojskolen i Kobenhavn, Institut for Datalingvistik)

4:20 - 5:00: Follow-up Discussion

Table of Contents

"Re-Use of a Proper Noun Recognition System in Commercial and Operational NLP Applications"....1
Chinatsu Aone and J. Maloney

"Developing a new grammar checker for English as a second language"....7
Cornelia Tschichold, Franck Bodmer,Etienne Cornu, Francois Grosjean, Lysiane Grojean, Natalie Kubler, Nicolas Lewy and Corinne Tschumi

"Software Re-Use and Evolution in Text Generation Applications"....13
Karen Kukich and Rebecca Passaneau, Bellcore; Kathleen McKeown, Dragomir Radev, Vasileios Hatzivassiloglou and Hongyan Jing

"Estimating the True Performance of Classification-Based NLP Technology"....23
James Nolan

"Tailored patient information: some issues and questions"....29
Ehud Reiter and Liesl Osman

"Practical Considerations in Building a Multi-Lingual Authoring System for Business Letters"....35
John Tait and Huw Sanderson,University of Sunderland; Jeremy Ellman, Ana Maria Martinez San Jose, Peter Hellwig, and Periklis Tsagheas

"A Language Identification Application Built on the Java Client/Server Platform"....43
Gary Adams and Philip Resnik

"Practical Experience with Grammar Sharing in Multilingual NLP"....49
Michael Gamon, Carmen Lozano, Jessie Pinkham and Tom Reutter

"NLP and Industry: Transfer and Reuse of Technolgies"....57
Leo Obrst and Krishna Jha

"Recycling Lingware in a Multilingual MT System"....65
Manny Rayner, David Carter, Ivan Bretan, Roberts Eklund, Mats Wiren, Steffen Leo Hansen, Sabine Kirchmeier-Andersen, Finn Sorensen and Hanne Erdman Thomsen