Proceedings of the

2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora

Held in conjunction with
The 38th Annual Meeting of the
Association for Computational Linguistics

Sponsored by
Behavior Design Corporation
GroupFire Inc
Intel China Research Center
LEXIS-NEXIS

Edited by

Hinrich Schütze
and
Keh-Yih Su

7-8 October 2000
Hong Kong University of Science and Technology (HKUST)
Hong Kong

Proceedings of the

2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora

Held in conjunction with
The 38th Annual Meeting of the
Association for Computational Linguistics

Sponsored by
Behavior Design Corporation
GroupFire Inc
Intel China Research Center
LEXIS-NEXIS

Edited by
Hinrich Schütze
and
Keh-Yih Su

7-8 October 2000
Hong Kong University of Science and Technology (HKUST)
Hong Kong

Order copies of this and other ACL workshop proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901
USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
acl@aclweb.org

SPONSORS:

Behavior Design Corporation
GroupFire Inc
Intel China Research Center
LEXIS-NEXIS

ORGANIZERS:

Chair: Hinrich Schütze, GroupFire Inc
Co-Chair: Keh-Yih Su, Behavior Design Corporation

Proceedings: David Yarowsky, Johns Hopkins University

PROGRAM COMMITTEE:

Einat Amitay, Macquarie University & CSIRO
Sophia Ananiadou, University of Salford
Susan Armstrong, University of Geneva
Thorsten Brants, Saarland University
Eric Brill, Microsoft Research
Jason Chang, National Tsing Hua University
Francine Chen, Xerox Palo Alto Research Center
Key-Sun Choi, Korea Adv. Inst. of Science & Technology
David Elworthy, Microsoft Research Cambridge
Tomaz Erjavec, Institute Jozef Stefan
Pascale Fung, Hong Kong University of Science & Technology
Eric Gaussier, Xerox Research Centre Europe
Niyu Ge, Brown University
Nancy Ide, Vassar College
Martin Jansche, Ohio State University
Andy Kehler, UC San Diego
Geunbae Lee, Pohang University of Science & Technology
Lillian Lee, Cornell University
Dekang Lin, University of Alberta
Kim-Teng Lua, National University of Singapore
Chris Manning, Stanford University
Yuji Matsumoto, Nara Inst. of Science & Technology
Helen Meng, Chinese University of Hong Kong
Masaaki Nagata, NTT Cyber Space Labs
Dragomir Radev, University of Michigan
Hae-Chang Rim, Korea University
Maosong Sun, Tsinghua University
Bing Swen, Peking University
Mark Wasson, Lexis-Nexis
Yorick Wilks, University of Sheffield
Jakub Zavrel, University of Antwerp

FURTHER INFORMATION:

Hinrich Schütze
GroupFire Inc
3600 Bridge Parkway
Redwood City, CA 94065, USA

Table of Contents

PRELIMINARY PROGRAM

Saturday, October 7

9:00-9:10 Welcome

9:10-9:35 *Pattern-Based Disambiguation for Natural Language Processing*
Eric Brill

9:35-10:00 *What's Yours and What's Mine: Determining Intellectual Attribution in Scientific Text*
Simone Teufel and Marc Moens

10:00-10:25 *Japanese Dependency Structure Analysis Based on Support Vector Machines*
Taku Kudo and Yuji Matsumoto

10:25-10:50 *Coaxing Confidences from an Old Friend: Probabilistic Classifications from Transformation Rule Lists*
Radu Florian, John Henderson and Grace Ngai

11:10-12:25 Invited Speaker

12:25-12:50 *Topic Analysis Using a Finite Mixture Model*
Hang Li and Kenji Yamanishi

12:50-13:15 *Sample Selection for Statistical Grammar Induction*
Rebecca Hwa

13:15-13:40 *A Uniform Method of Grammar Extraction and Its Applications*
Fei Xia, Martha Palmer and Aravind Joshi

13:40-15:00 LUNCH

15:00-15:25 *Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger*
Kristina Toutanova and Christopher D. Manning

15:25-15:50 *Error-driven HMM-based Chunk Tagger with Context-dependent Lexicon*
GuoDong Zhou and Jian Su

15:50-16:15 *Nonlocal Language Modeling based on Context Co-occurrence Vectors*
Sadao Kurohashi and Manabu Ori

16:15-16:40 *Detection of Language (Model) Errors*
K.Y. Hung, R.W.P. Luk, D. Yeung, K.F.L. Chung and W. Shu

17:00-17:25 *Cross-lingual Information Retrieval Using Hidden Markov Models*
Jinxi Xu and Ralph Weischedel

17:25-17:50 *Query Translation in Chinese-English Cross-Language Information Retrieval*
Zhang Yibo, Sun Le, Du Lin and Sun Yufang

17:50-18:15 *Word Alignment of English-Chinese Bilingual Corpus Based on Chunks*
Sun Le, Jin Youbing, Du Lin and Sun Yufang

18:15-18:40 *Empirical Term Weighting and Expansion Frequency*
Kyoji Umemura and Kenneth W. Church

Sunday, October 8

9:10-10:25 Invited Speaker

10:25-10:50 *A Machine Learning Approach to Answering Questions for Reading Comprehension Tests*
Hwee Tou Ng, Leong Hwee Teo and Jennifer Lai Pheng Kwan

10:50-11:15 *Automated Construction of Database Interfaces: Integrating Statistical and Relational Learning for Semantic Parsing*
Lappoon R. Tang and Raymond J. Mooney

11:15-11:40 *Automatic WordNet Mapping Using Word Sense Disambiguation*
Changki Lee, Geunbae Lee and Seo Jung Yun

11:40-12:05 *A Real-time Integration Of Concept-based Search and Summarization of Chinese Websites*
Joe F. Zhou and Weiquan Liu

12:25-12:50 *A Statistical Model for Parsing and Word-Sense Disambiguation*
Daniel M. Bikel

12:50-13:15 *Reducing Parsing Complexity by Intra-Sentence Segmentation based on Maximum Entropy Model*
Sung Dong Kim and Byoung Tak Zhang

13:15-13:40 *An Empirical Study of the Domain Dependence of Supervised Word Sense Disambiguation Systems*
Gerard Escudero, Lluís Màrquez and German Rigau

13:40-15:00 LUNCH

15:10-16:30 Panel

16:30-16:55 *Combining Lexical and Formatting Cues for Named Entity Acquisition from the Web*
Christian Jacquemin and Caroline Bush

16:55-17:20 *A Query Tool for Syntactically Annotated Corpora*
Laura Kallmeyer

17:20-17:45 *Statistical Filtering and Subcategorization Frame Acquisition*
Anna Korhonen, Genevieve Gorrell and Diana McCarthy

17:45-18:10 *One Sense per Collocation and Genre/Topic Variations*
David Martinez and Eneko Agirre