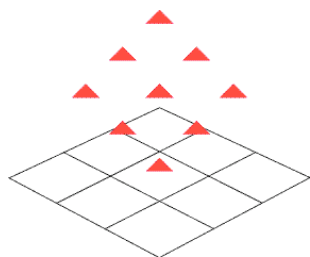


Acquiring (and Using) Linguistic (and World) Knowledge for Information Access

Papers from the 2002 AAAI Spring Symposium
Technical Report SS-02-09



AAAI Press

American Association for Artificial Intelligence

Acquiring (and Using) Linguistic (and World) Knowledge for Information Access

Papers from the 2002 AAAI Symposium

*Jussi Karlgren, Chair
Björn Gambäck and Pentti Kanerva, Cochairs*

March 25–27, Stanford, California
Technical Report SS-02-09

AAAI Press
Menlo Park, California

Copyright © 2002, AAAI Press

The American Association for Artificial Intelligence
445 Burgess Drive
Menlo Park, California 94025

ISBN 1-57735-192-4 SS-02-09

AAAI retains the right of first refusal to any publication arising from this AAAI event, and retains compilation copyright. Please do not make any inquiries or arrangements for hardcopy or electronic publication of all or part of the papers contained in this technical report without first exploring the options available through AAAI Press and *AI Magazine*. A signed release of this right by AAAI is required before publication by a third party.

Manufactured in the United States of America

Organizing Committee

Jussi Karlgren (chair), Swedish Institute of Computer Science

Pentti Kanerva (cochair), CSLI, Stanford University

Björn Gambäck (cochair), Swedish Institute of Computer Science

Marti Hearst, SIMS, University of California, Berkeley

Robert Hecht-Nielsen, HNC Software

Tony Plate, Bios Group

Contents

Foreword and Challenge: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access or Theory for Systems; Application for Theories / 1	<i>Jussi Karlgren, Björn Gambäck, and Pentti Kanerva</i>
On What Latent Semantic Analysis (LSA/LSI) Does and Doesn't Do / 2	<i>Thomas Landauer</i>
Using Proper Names to Cluster Documents / 3	<i>Dan Winchester and Mark Lee</i>
Document Content Analysis through Inverted Generation / 10	<i>Aurélien Max and Marc Dymetman</i>
Towards a Flexible Model of Word Meaning / 15	<i>Magnus Sahlgren</i>
Current Issues in Markup-Based Knowledge Extraction / 20	<i>Udo Kruschwitz</i>
OntoQuery: Ontology-based Querying of Texts / 26	<i>Troels Andreasen, Per Anker Jensen, Jørgen Fischer Nilsson, Patrizia Paggio, Bolette Sandford Pedersen, and Hanne Erdman Thomsen</i>
Construction of Thematic Representations of Texts Based on Domain-Specific Thesaurus / 30	<i>Natalia V. Loukachevitch and Boris V. Dobrov</i>
Foreground and Background Text in Retrieval / 36	<i>Jussi Karlgren and Timo Järvinen</i>
Underspecified Semantic Interpretation in an E-Mail Speech Interface / 43	<i>Björn Gambäck and Maria Cheadle</i>
The Public Acquisition of Commonsense Knowledge / 47	<i>Push Singh</i>
Representing Textual Content in a Generic Extraction Model / 53	<i>Nancy McCracken</i>
Symposium Summary: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access / 56	<i>Jussi Karlgren</i>

