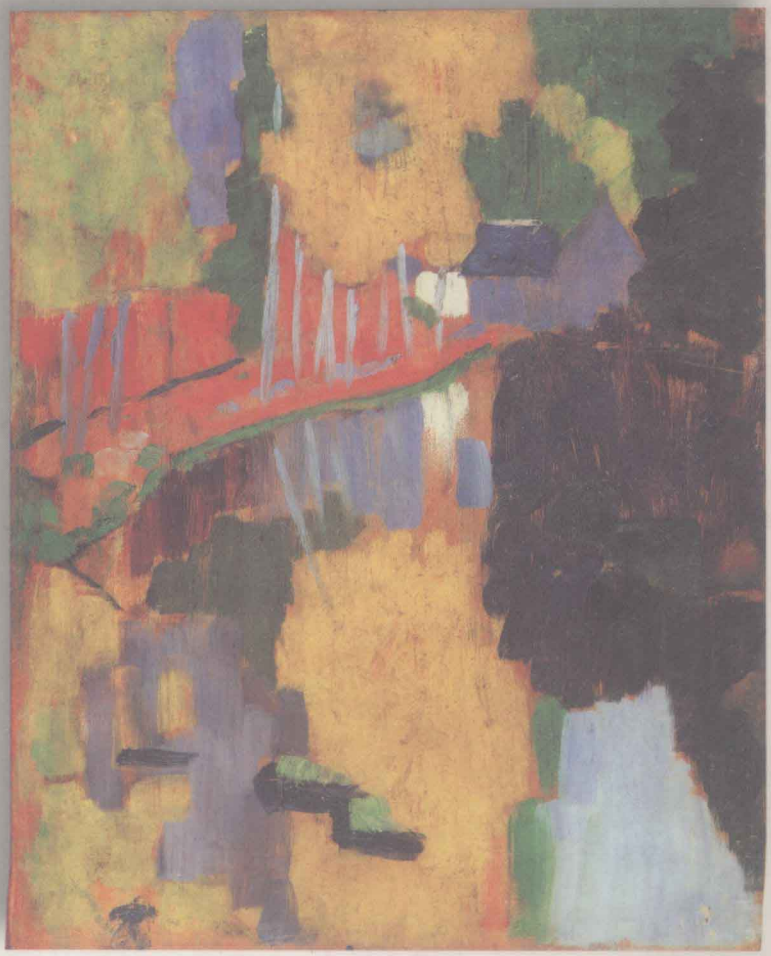
Natural Language Processing and Knowledge Representation

Language for Knowledge and Knowledge for Language



edited by
Łucja M. Iwańska
and Stuart C. Shapiro

Natural Language Processing and Knowledge Representation

Language for Knowledge and Knowledge for Language

Edited by

Łucja M. Iwańska and Stuart C. Shapiro

AAAI PRESS / THE MIT PRESS

MENLO PARK, CALIFORNIA, CAMBRIDGE, MASSACHUSETTS, LONDON, ENGLAND

Copublished and distributed by The MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts and London, England.

Library of Congress Cataloging-in-Publication Data

Natural language processing and knowledge representation : language for knowledge and knowledge for language / edited by Łucja M. Iwańska and Stuart C. Shapiro.

p. cm.

Includes bibliographical references and index.

ISBN 0-262-59021-2 (pbk. : alk. paper)

1. Natural language processing (Computer science) 2. Knowledge representation (Information theory). I. Iwańska, Łucja, 1958-. II. Shapiro, Stuart Charles. III. Title.

QA76.9.N38 N3843 2000

006.3'5—dc21 99-087360

CIP

Printed on acid-free paper in Canada.

10 9 8 7 6 5 4 3 2 1

For Tomasz — Ł. M. I.
and Caren — S. C. S.

Contents

Preface

Why this Book?

THIS BOOK CONTAINS the most recent theoretical and practical computational approaches to representing and utilizing the meaning of natural language. We believe that only such natural-language driven computational models allow the development of truly intelligent computer systems simulating, on a large scale, the critical role of natural language in human information and knowledge processing. Developing knowledge representation and reasoning systems based on natural language constitutes an exciting and, still, quite controversial new research direction in artificial intelligence. Natural language-based knowledge representation and reasoning systems and their large-scale implementations closely simulate the representational and inferential computational machinery of natural language, which, as argued by all the authors in this book, is dramatically different from any other theoretical and practical knowledge representation and reasoning system or automated reasoning system.

New Perspective: Natural Language as Knowledge Representation and Reasoning System, Not Just Interface

The new research direction of natural language based knowledge representation and reasoning systems emerged over the past few years. It grew out of concerns over the efficient handling of large-scale, general-purpose knowledge, reasoning, and the meaning of natural language. One motivation for this line of research was—and still is—the fact that a vast majority of knowledge representation and reasoning systems do not adequately reflect important characteristics of natural language and are representationally and inferentially impoverished relative to natural language. Notable exceptions include natural language motivated knowledge representation and reasoning systems of Schubert, Shapiro, and Sowa. Historically, the goal of knowledge representation and reasoning was to address representational and inferential needs of natural language pro-

cessing. However, the goals of knowledge representation and reasoning quickly and substantially diverged from that. In the past several years, there has been little interaction and interest between the knowledge representation and natural language processing communities. This book demonstrates that the natural language processing and knowledge representation communities have many common research goals.

The current shape of the natural language-based knowledge representation and reasoning systems direction evolved, in part, from discussions at a number of international symposia, workshops, and journal special issues involving the editors and contributors.

The research direction of natural language-based knowledge representation and reasoning systems constitutes a tremendous change in how we view the role of natural language in an intelligent computer system. The traditional view, widely held within the artificial intelligence and computational linguistics communities, considers natural language as an interface or front end to a system such as an expert system or knowledge base. In this view, inferencing and other interesting information and knowledge processing tasks are not part of natural language processing.

By contrast, the computational models of natural language presented in this book view natural language as a knowledge representation and reasoning system with its own unique, computationally attractive representational and inferential machinery. This new perspective sheds some light on the actual, still largely unknown, relationship between natural language and the human mind. Taken to an extreme, such approaches speculate that the structure of the human mind is close to natural language. In other words, natural language is essentially the language of human thought.

In the natural language-based knowledge representation and reasoning systems, general-purpose information and knowledge are (1) entered in computer systems via natural language in the form of texts or dialogs, (2) represented and combined via algorithms and data structures closely simulating the syntax and semantics of natural language, (3) reasoned about via inference mechanisms, which closely simulate inferences that humans make in natural language, and (4) exited from computer systems via natural language in the form of natural language answers to queries.

This means that all information and knowledge processing tasks supporting the computer system's intelligent behavior take place at the natural language level. The advantage of the natural language-like uniformity of representation and reasoning is a simple and powerful computer architecture. For not well formalized domains, it can be argued that natural language-based knowledge representation and reasoning are more advantageous than knowledge representation and reasoning not motivated by natural language.

Addressing Wide-Spread Misconceptions about Natural Language

This book sheds new light on the puzzling computational nature of natural language. Not only is natural language algorithmic, but its algorithmic buildup appears very different from both the traditional computer science number-crunching computation and the traditional artificial intelligence symbol-crunching computation. We believe that the research results presented in this book challenge a number of claims about natural language, including its alleged nonalgorithmicity and its remoteness from inference. Despite the fact that such claims have never been substantiated, they remain quite common in artificial intelligence symbol-crunching computation. For example, in his 1990 book *Representations of Commonsense Knowledge* (San Francisco: Morgan Kaufmann, p. 14), Ernest Davis states:

> There is a widespread agreement that AI programs should not use full natural language text as a knowledge representation language. Natural language cannot be easily manipulated algorithmically. It is full of ambiguities. Its meaning is context-dependent. Its syntax is extremely complex, and strongly dependent on semantics. Its connectives (prepositions, articles, and conjunctions) are remarkably vague and unsystematic. There are few powerful rules of inference on natural language strings.

More recently, John McCarthy gave an invited talk at the 1997 AAAI Fall Symposium on Context in Knowledge Representation and Natural Language. He claimed that natural language has little to do with inferencing. His transparency read:

> Language is froth on the surface of thought:
> - A human's basic knowledge of the world is not represented linguistically in the brain. Linguistic expression is often possible.
> - Language is for communication with people who cannot see through your eyes or duplicate your reasoning. Therefore, it cannot express thought directly.

We stress that the exact division of the representational and inferential labor between natural language and other systems such as the human mind, database or knowledge base remains an open research question to be extensively investigated.

Knowledge Representation, Reasoning, and Acquisition in Large-Scale, General Purpose Natural Language Processing Systems

In this book, we present contributions concerning the representing of, reasoning with and acquisition of different types of knowledge from large-scale natural language inputs. Some contributions represent the

traditional approach in which natural language plays a lesser representational and inferential role and the natural language processing system is supported by a nonnatural language-motivated knowledge representation and reasoning system.

A number of contributors present very interesting methods of automatic knowledge acquisition from medium-to-large-scale natural language inputs. These acquisition methods are based on natural language and incorporate its various algorithmic aspects. They have the potential to replace hand-crafting knowledge from textual documents, a costly and error-prone process. These methods raise hope that we will no longer need humans to postverify and clean-up system-acquired garbage.

Natural language processing researchers and practitioners are constantly faced with the inadequacy of data structures for representing natural language because their syntax and semantics do not fit the syntax and semantics of natural language well. Inference in natural language is not well simulated by traditional reasoning algorithms. This further complicates the task of the natural language processing-required automatic knowledge acquisition.

Computer Science-Style Natural Language Processing: Theory and Serious, Rich-Data Implementation

The authors of this book are computer scientists who focus clearly on the computational, computer science-style aspects of natural language. Most of them have strong interests in and extensive knowledge of such related disciplines as linguistics, psychology, and philosophy. They often incorporate and extend theories which originate in those fields. Some authors present theories of natural language phenomena for which no adequate theory exists in those related disciplines.

Another distinct characteristic is that all the authors view implementation as a necessary and, to some extent, independent piece of evidence of the correctness and generality of their theoretical findings. Their ideas have been implemented in existing, and, in most cases, mature natural language processing systems, which were developed over years at academic and industrial natural language research groups. The scope and scale of implementations vary significantly. Some authors present complete systems capable of a particular type of processing on a large scale. Others discuss a proof-of-concept implementation, which demonstrates a computational feasibility of the proposed theoretical ideas.

The natural language processing systems the authors discuss also differ significantly in terms of their test and development data. Some sys-

tems have been designed and tested on artificial, toy examples of English utterances (texts, dialogs and their fragments) illustrating the researched problems. Other systems use plausible simulation data. Finally, some systems are capable of handling large-scale corpora of real data.

Potential for Practical Applications: An In-Depth Processing of Knowledge in Huge Volumes of Internet Texts

The potential for practical applications of the work presented in our book is tremendous. We believe that natural language-based knowledge representation and reasoning systems allow one to uniquely combine the human-like quality of knowledge processing with computer quantitative advantages. Such systems allow computers to process knowledge in the form of natural language similar to people and then to combine it with nonhuman computer capabilities to be precise, fast, systematic and virtually memory-unlimited. We believe that only such human and nonhuman combinations of information and knowledge processing give hope for an in-depth processing of information and knowledge in the huge volumes of natural language inputs. Such huge volumes include documents such as texts and transcribed dialogs freely available on the world wide web. This constantly growing volume is currently estimated to be billions of textual documents.

About this Book

The first part of this book presents natural language based knowledge representation and reasoning systems, or formal, computational models of natural language, whose original motivation was the inherent limitations of the representation of natural language based on first-order classical logic. The second part of the book discusses large-scale approaches to representing reasoning with and acquiring different types of knowledge for general-purpose natural language processing systems.

We have attempted to make this book self-contained text. We provide appendices with basic computational and mathematical concepts needed to successfully follow the material. We discuss the up-to-date utility of first-order classical logic for representing the meaning of natural language. Until very recently, first-order classical logic was a primary non-adhoc computational representation of natural language. Throughout the book, we comment on the inherent limitations of first-order classical logic for capturing the meaning of natural language. We also provide, in

Subject	Professional Activity More Information	Year
Knowledge representation for natural language processing in implemented systems	AAAI Symposium wy.smsu.edu/pub/www/kr_nlp_impl_sys.html	1994
Context in natural language processing	IJCAI Workshop www.cs.wayne.edu/~lucja/context-w1.html	1995
Knowledge representation systems based on natural language	AAAI Symposium www.cs.wayne.edu/~lucja/nlkr-w1.html	1996
Context in knowledge representation and natural language	AAAI Symposium www.cs.wayne.edu/~lucja/context-w2.html	1996
Knowledge representation for natural language	*Minds and Machines* 3(1) Kluwer Academic Publishers	1993
Knowledge representation and inference for natural language processing	*Expert Systems* 9(1) JAI Press	1996
Knowledge representation for natural language processing in implemented systems	*Natural Language Engineering* 3(2) Cambridge University Press	1997
Context in natural language processing	*Computational Intelligence* 13(2) Blackwell Publishers	1997

the appendices, many examples and data illustrating representational and inferential challenges of natural language. These examples and data put to a hard test any general-purpose knowledge representation and reasoning system, natural language-motivated or not, as well as any general-purpose natural language processing system.

Acknowledgments

Many people directly and indirectly contributed to this book. We would like to thank the contributors of the individual chapters as well as the organizers, program committee members and the participants of the professional activities that provided a friendly, but tough discussion forum for many ideas presented here. Our book is a coherent collection largely due to the fact that the editors and the contributors were actively involved in the closely related professional activities, conferences and journal special issues listed in the table above.

- Łucja M. Iwańska and Stuart C. Shapiro

Contributors

Syed S. Ali
Department of Mathematical Sciences
University of Wisconsin-Milwaukee
tigger.cs.uwm.edu/~syali/

Karen Ehrlich
Computer Science Department
SUNY College at Fredonia
www.cs.fredonia.edu/~ehrlich/

Bonnie J. Dorr
Department of Computer Science
University of Maryland, College Park
www.umiacs.umd.edu/~bonnie/

Robert Givan
Department of Electrical
& Computer Engineering
Purdue University
dynamo.ecn.purdue.edu/~givan/

Susan M. Haller
Computer Science Department
University of Wisconsin—Parkside
cs.uwp.edu/staff/haller/

Sanda Harabagiu
Department of Computer Science
and Engineering
Southern Methodist University
www.seas.smu.edu/~sanda/

Chung Hee Hwang
Microelectronics & Computer Technology
Corporation (MCC)
hwang@mcc.com

Łucja M. Iwańska
Department of Computer Science
Wayne State University
www.cs.wayne.edu/profs/iwanska.html

Kellyn Kruger
Department of Computer Science
Wayne State University
kkk@cs.wayne.edu

Naveen Mata
Department of Computer Science
Wayne State University
osm@cs.wayne.edu

David McAllester
AT&T Labs Research
www.research.att.com/~dmac/

David D. McDonald
Gensym Corporation
davidmcdonald@alum.mit.edu

Susan W. McRoy
Department of Electrical Engineering and
Computer Science
University of Wisconsin—Milwaukee
www.cs.uwm.edu/~mcroy/

Dan Moldovan
Department of Computer Science
and Engineering
Southern Methodist University
www.seas.smu.edu/~moldovan/

William J. Rapaport
Department of Computer Science
and Engineering
State University of New York at Buffalo
www.cse.buffalo.edu/~rapaport/

Lenhart Schubert
Department of Computer Science
University of Rochester
www.cs.rochester.edu/u/schubert/

Stuart C. Shapiro
Department of Computer Science
and Engineering
State University of New York at Buffalo
www.cse.buffalo.edu/~shapiro/

Clare R. Voss
Information Science and Technology Directorate, Army Research Laboratory,
Adelphi, MD
voss@arl.mil

Section One