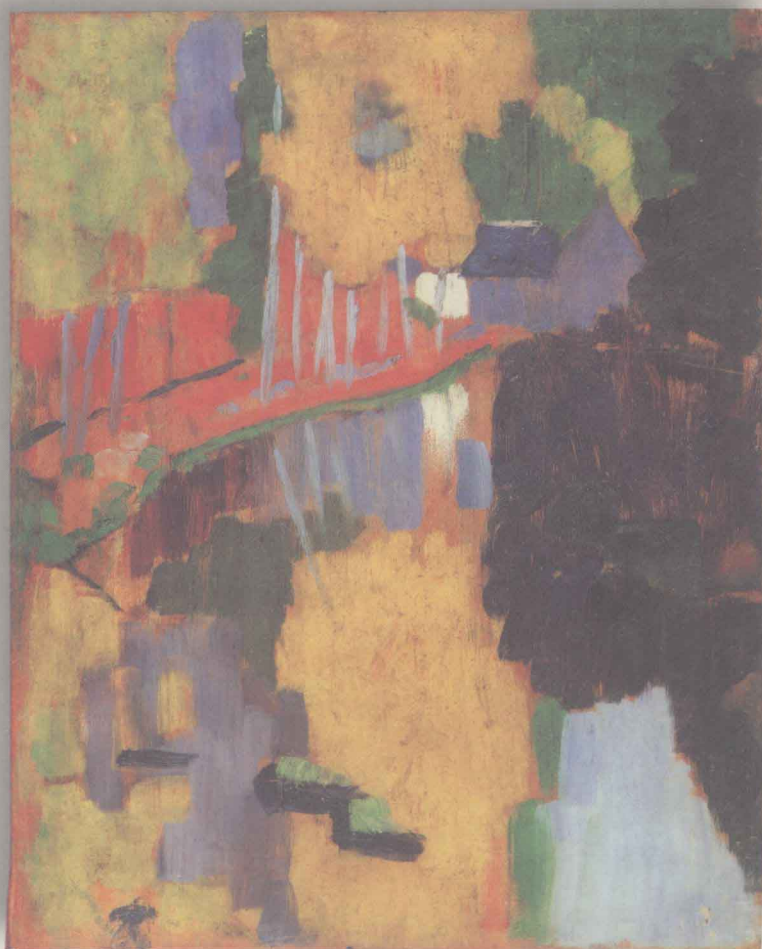


# Natural Language Processing and Knowledge Representation

---

Language for Knowledge  
and Knowledge for Language

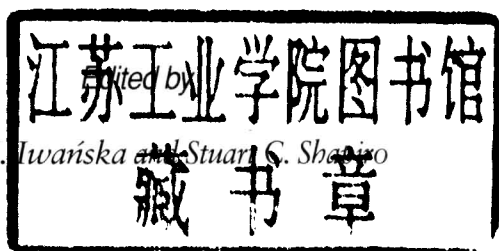


edited by  
**Łucja M. Iwańska**  
and **Stuart C. Shapiro**

# Natural Language Processing and Knowledge Representation

Language for Knowledge  
and Knowledge for Language

*Lucja M. Iwańska and Stuart C. Shapiro*



AAAI PRESS / THE MIT PRESS

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

Copyright © 2000, American Association for Artificial Intelligence  
445 Burgess Drive  
Menlo Park, CA 94025

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

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 / xiii

*Lucja M. Iwańska and Stuart C. Shapiro*

Contributors / xix

## *Section One*

### Formal, Computational, Representations and Inference Methods Based on Natural Language

Section Introduction / 3

*Lucja M. Iwańska and Stuart C. Shapiro*

#### *Chapter One*

#### Natural Language Is a Powerful Knowledge Representation System: The UNO Model / 7

*Lucja M. Iwańska*

This chapter presents a novel view of natural language as a powerful, general-purpose knowledge representation system particularly suitable for handling knowledge in not well formalized domains. It discusses a number of highly desirable representational and inferential characteristics unique to natural language. They include great expressiveness combined with computational tractability, rich structure, exploitation of logical contradiction and logical redundancy, inherent underspecificity and context-dependency, and facilitating learning. A formal computational model of natural language that closely simulates these characteristics is presented. Many examples illustrate the computational mechanism of representing and reasoning with meaning of natural language and different types of knowledge, including taxonomic, probabilistic, temporal and some aspects of spatial knowledge. A large-scale implementation of the model is also discussed. The system allows one to automatically acquire and utilize knowledge from large-scale corpora of textual documents.

#### *Chapter Two*

#### Natural Language Syntax and First Order Inference / 65

*David A. McAllester and Robert Givan*

The authors have argued elsewhere that first order inference can be made more efficient by using nonstandard syntax for first order logic. In this chapter, they define a syntax for first order logic based on the structure of natural language under Montague semantics. They show that, for a certain fairly expressive fragment of this language, satisfiability is polynomial time decidable. The polynomial time decision procedure can be used as a subroutine in general purpose inference systems and seems to be more powerful than analogous procedures based on either classical or taxonomic syntax.

*Chapter Three***Issues in the Representation of Real Texts: The Design of KRISP / 77***David D. McDonald*

Attempting to understand a journalist's real texts puts special demands on the formalism that a language comprehension system uses to represent what it has understood. KRISP is an experimental representation system that was designed to address these demands: high-speed operation, the ability to accommodate a very wide range of grammatical constructions, and reversibility. Its fundamental motivations stem from investigations over the course of the last decades to try and establish just why it was that other representational systems were always turning out to be awkward when used as the source for the generation of fluent prose. Formally, KRISP is essentially an object-oriented repackaging of a typed lambda calculus, paying particular attention to the representation of partially saturated relations and providing first class objects to represent the binding of a variable to a value. As a representational system, it embodies a theory of how information is structured in a natural language text. This is a hypothesis about the nature of the objects that texts denote and the principles of semantic interpretation that map them to a model comprised of such objects. It is intended to run in close coordination with the SPARSER natural language understanding system.

*Chapter Four***Episodic Logic Meets Little Red Riding Hood—A Comprehensive Natural Representation for Language Understanding / 111***Lenhart K. Schubert and Chung Hee Hwang*

The authors describe a comprehensive framework for narrative understanding based on episodic logic. This situational logic was developed and implemented as a semantic representation and commonsense knowledge representation that would serve the full range of interpretive and inferential needs of general natural language understanding. The most distinctive feature of episodic logic is its natural languagelike expressiveness. It allows for generalized quantifiers, lambda abstraction, sentence and predicate modifiers, sentence and predicate reification, intensional predicates (corresponding to wanting, believing, making, etc.), unreliable generalizations, and perhaps most importantly, explicit situational variables linked to arbitrary formulas that describe them. These allow episodes to be explicitly related in terms of part-whole, temporal and causal relations. Episodic logical form is easily computed from surface syntax and lends itself to effective inference.

*Chapter Five***SNePS: A Logic for Natural Language Understanding and Commonsense Reasoning / 175***Stuart C. Shapiro*

The use of logic for knowledge representation and reasoning systems is controversial. There are, indeed, several ways that standard first order predicate logic is inappropriate for modeling natural language understanding and commonsense reasoning. However, a more appropriate logic can be designed. This chapter presents several aspects of such a logic.

*Section Two***Knowledge Representation and Acquisition for Large-Scale, General-Purpose Natural Language Processing Systems****Section Introduction / 193***Lucja M. Iwańska and Stuart C. Shapiro*

*Chapter Six***A Multi-Level Approach to Interlingual Machine Translation:  
Defining the Interface between Representational Languages / 207***Bonnie J. Dorr and Clare R. Voss*

This chapter describes a multi-level design, i.e., a nonuniform approach to interlingual machine translation, in which distinct representational languages are used for different types of knowledge. The authors demonstrate that a linguistically-motivated “division of labor” across multiple representation levels has not complicated, but rather has readily facilitated, the identification and construction of systematic relations at the interface between each level. They describe a model of interpretation and representation of natural language sentences that has been implemented as part of an interlingual machine-translation system called PRINCITRAN.

*Chapter Seven***Uniform Natural (Language) Spatio-Temporal Logic: Reasoning about  
Absolute and Relative Space and Time / 249***Lucja M. Iwańska*

In this chapter, uniform spatio-temporal reasoner is presented. Its representational, inferential and computational characteristics closely resemble natural language. It is demonstrated that important inferences about time and space can be captured by a general representation and reasoning mechanism inherent in natural language many aspects of which are closely mimicked by the proposed computational model of natural language. As a result, a uniform representation and inference, and therefore a simple architecture, for temporal, spatial and other reasoning is accomplished. It is also shown that computing logical, context-independent and some nonmonotonic, context-dependent inferences for temporal, spatial and other objects is analogous.

Real-life examples illustrate the representational and reasoning capabilities of the proposed natural language style spatio-temporal reasoner, including many previously unaccounted aspects of temporal and spatial information conveyed by common English temporal expressions, reasoning with information from arbitrary Boolean temporal expressions involving explicit negation, disjunction and conjunction, handling temporal quantifiers, handling infinite number of temporal and spatial relations, handling both absolute and relative temporal and spatial information, and handling nonnumeric qualitative temporal and spatial information.

*Chapter Eight***Mixed Depth Representations for Dialog Processing / 283***Susan W. McRoy, Syed S. Ali, and Susan M. Haller*

The authors describe their work on developing a general purpose tutoring system that will allow students to practice their decision-making skills in a number of domains. The tutoring system, B2, supports mixed-initiative natural language interaction. The natural language processing and knowledge representation components are also general purpose—which leads to a tradeoff between the limitations of superficial processing and syntactic representations and the difficulty of deeper methods and conceptual representations. Their solution uses a mixed-depth representation, one that encodes syntactic and conceptual information in the same structure. As a result, they can use the same representation framework to produce a detailed representation of requests and to produce a partial representation of questions. Moreover, the representations use the same knowledge representation framework that is used to reason about discourse processing and domain information—so that the system can reason with (and about) the utterances, if necessary. This work is the first (and to our knowledge, the only) implementation of mixed-depth representations for dialog processing.

*Chapter Nine*

**Enriching the WordNet Taxonomy with  
Contextual Knowledge Acquired from Text / 301**

*Sanda M. Harabagiu and Dan I. Moldovan*

This chapter presents a possible solution for the problem of integrating contextual knowledge in the WordNet database. Contextual structures are derived from three sources: (1) minimal contexts—in the form of semantic net transformations of WordNet glosses; (2) dynamic contexts rendered by webs of lexico-semantic paths revealing textual implied information and (3) static contexts—represented by patterns of concepts and semantic links. The relevance of these structures is shown on a three-tiered benchmark, comprising word-sense disambiguation, coreference resolution, and acquisition of domain patterns for information extraction.

*Chapter Ten*

**Fully Automatic Acquisition of Taxonomic Knowledge from Large  
Corpora of Texts: Limited-Syntax Knowledge Representation System  
based on Natural Language / 335**

*Łucja M. Iwańska, Naveen Mata, and Kellyn Kruger*

This chapter presents a new method for fully automatic knowledge acquisition from large corpora of unseen texts. The approach exploits simple, efficiently, fast and reliably extractable, parsable and in-depth interpretable constructs of natural language specialized to convey taxonomic knowledge. It allows one to acquire large quantities of high quality general-purpose knowledge and practically eliminates costly and error-prone human pre&post-processing. Examples of the system-acquired concepts are discussed.

*Chapter Eleven*

**A Computational Theory of Vocabulary Acquisition / 347**

*William J. Rapaport and Karen Ehrlich*

As part of an interdisciplinary project to develop a computational cognitive model of a reader of narrative text, the authors are developing a computational theory of how natural-language-understanding systems can automatically acquire new vocabulary by determining from context the meaning of words that are unknown, misunderstood, or used in a new sense. "Context" includes surrounding text, grammatical information, and background knowledge, but no external sources. The authors' thesis is that the meaning of such a word can be determined from context, can be revised upon further encounters with the word, converges to a dictionarylike definition if enough context has been provided and there have been enough exposures to the word, and eventually settles down to a steady state that is always subject to revision upon further encounters with the word. The system is being implemented in the SNePS knowledge-representation and reasoning system.

## Appendices

*Appendix A*

**Propositional, First-Order, and Higher-Order Logics / 379**

*Stuart C. Shapiro*

*Appendix B*

**Relations, Lattices, Algebras, Generalized Quantifier:  
Definitions and Theorems / 397**

*Łucja M. Iwańska*

*Appendix C*

Representational and Inferential Challenges of  
Natural Language : Examples and Data / 403

*Lucja M. Iwańska*

Bibliography / 415

Index / 441

# 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	<i>Minds and Machines</i> 3(1) Kluwer Academic Publishers	1993
Knowledge representation and inference for natural language processing	<i>Expert Systems</i> 9(1) JAI Press	1996
Knowledge representation for natural language processing in implemented systems	<i>Natural Language Engineering</i> 3(2) Cambridge University Press	1997
Context in natural language processing	<i>Computational Intelligence</i> 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  
[tiger.cs.uwm.edu/~syali/](http://tiger.cs.uwm.edu/~syali/)

Karen Ehrlich

Computer Science Department  
SUNY College at Fredonia  
[www.cs.fredonia.edu/~ehrich/](http://www.cs.fredonia.edu/~ehrich/)

Bonnie J. Dorr

Department of Computer Science  
University of Maryland, College Park  
[www.umiacs.umd.edu/~bonnie/](http://www.umiacs.umd.edu/~bonnie/)

Robert Givan

Department of Electrical  
& Computer Engineering  
Purdue University  
[dynamo.ecn.purdue.edu/~givan/](http://dynamo.ecn.purdue.edu/~givan/)

Susan M. Haller

Computer Science Department  
University of Wisconsin—Parkside  
[cs.uwp.edu/staff/haller/](http://cs.uwp.edu/staff/haller/)

Sanda Harabagiu

Department of Computer Science  
and Engineering  
Southern Methodist University  
[www.seas.smu.edu/~sanda/](http://www.seas.smu.edu/~sanda/)

Chung Hee Hwang

Microelectronics & Computer Technology  
Corporation (MCC)  
[hwang@mcc.com](mailto:hwang@mcc.com)

Łucja M. Iwańska

Department of Computer Science  
Wayne State University  
[www.cs.wayne.edu/profs/iwanska.html](http://www.cs.wayne.edu/profs/iwanska.html)

Kellyn Kruger

Department of Computer Science  
Wayne State University  
[kkk@cs.wayne.edu](mailto:kkk@cs.wayne.edu)

Naveen Mata

Department of Computer Science  
Wayne State University  
[osm@cs.wayne.edu](mailto:osm@cs.wayne.edu)

David McAllester

AT&T Labs Research  
[www.research.att.com/~dmac/](http://www.research.att.com/~dmac/)

David D. McDonald

Gensym Corporation  
[davidmcdonald@alum.mit.edu](mailto:davidmcdonald@alum.mit.edu)

Susan W. McRoy

Department of Electrical Engineering and  
Computer Science  
University of Wisconsin—Milwaukee  
[www.cs.uwm.edu/~mcroy/](http://www.cs.uwm.edu/~mcroy/)

Dan Moldovan

Department of Computer Science  
and Engineering  
Southern Methodist University  
[www.seas.smu.edu/~moldovan/](http://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/](http://www.cse.buffalo.edu/~rapaport/)

Lenhart Schubert

Department of Computer Science  
University of Rochester  
[www.cs.rochester.edu/u/schubert/](http://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/](http://www.cse.buffalo.edu/~shapiro/)

Clare R. Voss

Information Science and Technology Di-  
rectorate, Army Research Laboratory,  
Adelphi, MD  
[voss@arl.mil](mailto:voss@arl.mil)

# Section One