Automatic Text Processing The Transformation Analysis And Retrieval Of Information By Computer Addison Wesley Series In Computer Science

An information retrieval (IR) system is designed to analyse, process and store sources of information and retrieve those that match a particular user's requirements. A bewildering range of techniques is now available to the information professional attempting to successfully retrieve information. It is recognized that today's information professionals need to concentrate their efforts on learning the techniques of computerized IR. However, it is this book's contention that it also benefits them to learn the theory, techniques and tools that constitute the traditional approaches to the organization and processing of information. In fact much of this knowledge may still be applicable in the storage and retrieval of electronic information in digital library environments. The fully revised third edition of this highly regarded textbook has been thoroughly updated to incorporate major changes in this rapidly expanding field since the second edition in 2004, and a complete new chapter on citation indexing has been added. Unique in its scope, the book covers the whole spectrum of information storage and retrieval, including: users of IR and IR options; database technology; bibliographic formats; cataloguing and metadata; subject analysis and representation; automatic indexing and file organization; vocabulary control; abstracts and indexing; searching and retrieval; user-centred models of IR and user interfaces; evaluation of IR systems and evaluation experiments; online and CD-ROM IR; multimedia IR; hypertext and mark-up languages; web IR; intelligent IR; natural language processing and its applications in IR; citation analysis and IR; IR in digital libraries; and trends in IR research. Illustrated with many examples and comprehensively referenced for an international audience, this is an indispensable textbook for students of library and information studies. It is also an invaluable aid for information practitioners wishing to brush up on their skills and keep up to date with the latest techniques.

Information Retrieval (IR) has concentrated on the development of information management systems to support user retrieval from large collections of homogeneous textual material. A variety of approaches have been tried and tested with varying degrees of success over many decades of research. Hypertext (HT) systems, on the other hand, provide a retrieval paradigm based on browsing through a structured information space, following pre-defined connections between information fragments until an information need is satisfied, or appears to be. Information Retrieval and Hypertext addresses the confluence of the areas of IR and HT and explores the work done to date in applying techniques from one area, to the other leading to the development of `hypertext information retrieval' (HIR) systems. An important aspect of the work in IR/HT and in any user-centred information system is the emergence of multimedia information and such multimedia information is treated as an integral information type in this text. The contributed chapters cover the development of integrated hypertext information retrieval models, and the application of IR and HT techniques in hypertext construction and the approaches that can be taken in searching HIR systems. These chapters are complemented by two overview chapters covering, respectively, information retrieval and hypertext research and developments. Information Retrieval and Hypertext is important as it is the first text to directly address the combined searching/browsing paradigm of information discovery which is becoming so important in modern computing environments. It will be of interest to researchers and professionals working in a range of areas related to information discovery.

This volume is a presentation of all methods of legal knowledge representation from the point of view of jurisprudence as well as computer science. A new method of automatic analysis of legal texts is presented in four case studies. Law is seen as an information system with legally formalised information processes. The achieved coverage of legal knowledge in information retrieval systems has to be followed by the next step: conceptual indexing and automatic analysis of texts. Existing approaches of automatic knowledge representations do not have a proper link to the legal language in information systems. The concept-based model for semi-automatic analysis of legal texts provides this necessary connection. The knowledge base of descriptors, context-sensitive rules and meta-rules formalises properly all important passages in the text corpora for automatic analysis. Statistics and self-organising maps give assistance in knowledge acquisition. The result of the analysis is organised with automatically generated hypertext links. Four case studies show the huge potential but also some drawbacks of this approach.

It is our great pleasure to welcome you to the 11th International Conference on Neural Information Processing (ICONIP 2004) to be held in Calcutta. ICONIP 2004 is organized jointly by the Indian Statistical Institute (ISI) and Jadavpur University (JU). We are con?dent that ICONIP 2004, like the previous conf- ences in this series,will providea forum for fruitful interactionandthe exchange of ideas between the participants coming from all parts of the globe. ICONIP 2004 covers all major facets of computational intelligence, but, of course, with a primary emphasis on neural networks. We are sure that this meeting will be enjoyable academically and otherwise. We are thankful to the track chairs and the reviewers for extending their support in various forms to make a sound technical program. Except for a few cases, where we could get only two review reports, each submitted paper was reviewed by at least three referees, and in some cases the revised versions were againcheckedbythereferees. Wehad470submissionsanditwasnotaneasytask for us to select papers for a four-day conference. Because of the limited duration of the conference, based on the review reports we selected only about 40% of the contributed papers. Consequently, it is possible that some good papers are left out. We again express our sincere thanks to all referees for accomplishing a great job. In addition to 186 contributed papers, the proceedings includes two plenary presentations, four invited talks and 18 papers in four special sessions. The proceedings is organized into 26 coherent topical groups.

The last decade has been one of dramatic progress in the field of Natural Language Processing (NLP). This hitherto largely academic discipline has found itself at the center of an information revolution ushered in by the Internet age, as demand for human-computer communication and information access has exploded. Emerging applications in computer-assisted information production and dissemination, automated understanding of news, understanding of spoken language, and processing of foreign languages have given impetus to research that resulted in a new generation of robust tools, systems, and commercial products. This volume focuses on the use of Natural Language Processing (NLP) in Information Retrieval, the technology that grew out of library research to become our best hope in dealing with today's information overload. The book gives a broad overview of the work being done at the junction of these two important fields, and suggests directions for future explorations. It is organized into two loosely structured parts: The first part, consisting of Chapters 1 through 7, discusses research systems and evaluations that represent major avenues where the impact of NLP technologies in information retrieval is being explored. The second part (Chapters 8 through 14) describes specific implementations and prototypes of information systems where NLP techniques are used or proposed to assist in accurate retrieval, text categorization, question answering, and in organizing the results for the user.

Audience: This book will be a valuable reference to researchers and practitioners in the fields of Natural Language Processing, Information Retrieval, and Computational Linguistics.

Christian Jacquemin shows how the power of natural language processing (NLP) can be used to advance text indexing and information retrieval (IR). In this book Christian Jacquemin shows how the power of natural language processing (NLP) can be used to advance text indexing and information retrieval (IR). Jacquemin's novel tool is FASTR, a parser that normalizes terms and recognizes term variants. Since there are more meanings in a language than there are words, FASTR uses a metagrammar composed of shallow linguistic transformations that describe the morphological, syntactic, semantic, and pragmatic variations of words and terms. The acquired parsed terms can then be applied for precise retrieval and assembly of information. The use of a corpus-based unification grammar to define, recognize, and combine term variants from their base forms allows for intelligent information access to, or "linguistic data tuning" of, heterogeneous texts. FASTR can be used to do automatic controlled indexing, to carry out content-based Web searches through conceptually related alternative query formulations, to abstract scientific and technical extracts, and even to translate and collect terms from multilingual material. Jacquemin provides a comprehensive account of the method and implementation of this innovative retrieval technique for text processing.

This book constitutes the refereed proceedings of the Second International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2001, held in Mexico City, Mexico in February 2001. The 38 revised full papers and 12 short papers presented together with three invited papers were carefully reviewed and selected from 72 submissions. The books is divided in topical sections on computational linguistic theories, semantics, anaphora and reference, disambiguation, translation, text generation, dictionaries and corpora, morphology, parsing techniques, text categorization, information retrieval, and structure identification and text mining.

"This book is an in-depth collection aimed at developers and scholars of research articles from the expanding field of digital libraries"--Provided by publisher.

This book presents recent developments in automatic text analysis. Providing an overview of linguistic modeling, it collects contributions of authors from a multidisciplinary area that focus on the topic of automatic text analysis from different perspectives. It includes chapters on cognitive modeling and visual systems modeling, and contributes to the computational linguistic and information theoretical grounding of automatic text analysis.

th CICLing 2009 markedthe 10 anniversary of the Annual Conference on Intel- gent Text Processing and Computational Linguistics. The CICLing conferences provide a wide-scope forum for the discussion of the art and craft of natural language processing research as well as the best practices in its applications. This volume contains ?ve invited papers and the regular papers accepted for oral presentation at the conference. The papers accepted for poster presentation were published in a special issue of another journal (see the website for more information). Since 2001, the proceedings of CICLing conferences have been published in Springer's Lecture Notes in Computer Science series, as volumes 2004, 2276, 2588, 2945, 3406, 3878, 4394, and 4919. This volume has been structured into 12 sections: – Trends and Opportunities – Linguistic Knowledge Representation Formalisms – Corpus Analysis and Lexical Resources – Extraction of Lexical Knowledge – Morphology and Parsing – Semantics – Word Sense Disambiguation – Machine Translation and Multilinguism – Information Extraction and Text Mining – Information Retrieval and Text Comparison – Text Summarization – Applications to the Humanities A total of 167 papers by 392 authors from 40 countries were submitted for evaluation by the International Program Committee, see Tables 1 and 2. This volume contains revised versions of 44 papers, by 120 authors, selected for oral presentation; the acceptance rate was 26. 3%.

NLDB 2005, the 10th International Conference on Applications of Natural L- guage to Information Systems, was held on June 15–17, 2005 at the University of Alicante, Spain. Since the ?rst NLDB conference in 1995 the main goal has been to provide a forum to discuss and disseminate research on the integration of natural language resources in information system engineering. The development and convergence of computing, telecommunications and information systems has already led to a revolution in the way that we work, communicate with each other, buy goods and use services, and even in the way that weentertainandeducate ourselves.The revolutioncontinues,andoneof its results is that large volumes of information will increasingly be held in a form which is more natural for users than the data presentation formats typical of computer systems of the past. Natural language processing (NLP) is crucial in solving these problems, and language technologies will make an indispensable contribution to the success of information systems. We hope that NLDB 2005 was a modest contribution to this goal. NLDB 2005 contributed to advancing the goals and the high international standing of these conferences, largely due to its Program Committee, composed of renowned researchers in the ?eld of natural language processing and inf- mation system engineering. Papers were reviewed by three reviewers from the Program Committee. This clearly contributed to the signi?cant number of - pers submitted(95).Twenty-ninewereacceptedasregularpapers,while18were accepted as short papers.

With the growing use of information technology and the recent advances in web systems, the amount of data available to users has increased exponentially. Thus, there is a critical need to understand the content of the data. As a result, data-mining has become a popular research topic in recent years for the treatment of the "data rich and information poor" syndrome. In this carefully edited volume a theoretical foundation as well as important new directions for data-mining research are presented. It brings together a set of well respected data mining theoreticians and researchers with practical data mining experiences. The presented theories will give data mining practitioners a scientific perspective in data mining and thus provide more insight into their problems, and the provided new data mining topics can be expected to stimulate further research in these important directions.

Extracting content from text continues to be an important research problem for information processing and management. Approaches to capture the semantics of text-based document collections may be based on Bayesian models, probability theory, vector space models, statistical models, or even graph theory. As the volume of digitized textual media continues to grow, so does the need for designing robust, scalable indexing and search strategies (software) to meet a variety of

user needs. Knowledge extraction or creation from text requires systematic yet reliable processing that can be codified and adapted for changing needs and environments. This book will draw upon experts in both academia and industry to recommend practical approaches to the purification, indexing, and mining of textual information. It will address document identification, clustering and categorizing documents, cleaning text, and visualizing semantic models of text.

This book constitutes the thoroughly refereed proceedings of the Second International Joint Conference on Natural Language Processing, IJCNLP 2005, held in Jeju Island, Korea in October 2005. The 88 revised full papers presented in this volume were carefully reviewed and selected from 289 submissions. The papers are organized in topical sections on information retrieval, corpus-based parsing, Web mining, rule-based parsing, disambiguation, text mining, document analysis, ontology and thesaurus, relation extraction, text classification, transliteration, machine translation, question answering, morphological analysis, text summarization, named entity recognition, linguistic resources and tools, discourse analysis, semantic analysis NLP applications, tagging, language models, spoken language, and terminology mining.

This book presents revised versions of the lectures given at the 8th ELSNET European Summer School on Language and Speech Communication held on the Island of Chios, Greece, in summer 2000. Besides an introductory survey, the book presents lectures on data analysis for multimedia libraries, pronunciation modeling for large vocabulary speech recognition, statistical language modeling, very large scale information retrieval, reduction of information variation in text, and a concluding chapter on open questions in research for linguistics in information access. The book gives newcomers to language and speech communication a clear overview of the main technologies and problems in the area. Researchers and professionals active in the area will appreciate the book as a concise review of the technologies used in text- and speech-triggered information access.

This book constitutes the refereed proceedings of the 17th International Conference on Applications of Natural Language to Information Systems, held in Groningen, The Netherlands, in June 2012. The 12 full papers, 24 short papers and 16 poster papers presented in this volume together with a full-paper length invited talks were carefully reviewed and selected from 90 submissions. The rapidly evolving state-of-the-art in NLP and the shifting interest to appcliations targeting document and data collections available on the Web, including an increasing amount of user generated content, is reflected in the contributions to this book. Topics covered are information retrieval, text classification and clustering, summarization, normalization of user generated content, "forensic" NLP, ontologies and natural language, sentiment analysis, question answering and information extraction, terminology and named entity recognition, and NLP tools development.

The refereed proceedings of the 6th International Workshop on Computational Processing of the Portuguese Language, PROPOR 2003, held in Faro, Portugal, in June 2003. The 24 revised full papers and 17 revised short papers presented were carefully reviewed and selected from 64 submissions. The papers are organized in topical sections on speech analysis and recognition; speech synthesis; pragmatics, discourse, semantics, syntax, and the lexicon; tools, resources, and applications; dialogue systems; summarization and information extraction; and evaluation.

Electronic Multimedia Publishing brings together in one place important contributions and up-to-date research results in this fast moving area. Electronic Mulitmedia Publishing serves as an excellent reference, providing insight into some of the most challenging research issues in the field.

This book constitutes the refereed proceedings of the Third European Conference on Research and Advanced Technology for Digital Libaries, ECDL'99, held in Paris, France in September 1999. The 26 revised full papers presented were carefully reviewed and selected from a total of 124 submissions. The book is divided in topical sections on image categorization and access, audio and video in digital libraries, information retrieval, user adaptation, knowledge sharing, cross language issues, case studies, and modelling, accessability and connectedness.

The first edition of ELL (1993, Ron Asher, Editor) was hailed as "the field's standard reference work for a generation". Now the all-new second edition matches ELL's comprehensiveness and high quality, expanded for a new generation, while being the first encyclopedia to really exploit the multimedia potential of linguistics. * The most authoritative, up-to-date, comprehensive, and international reference source in its field * An entirely new work, with new editors, new authors, new topics and newly commissioned articles with a handful of classic articles * The first Encyclopedia to exploit the multimedia potential of linguistics through the online edition * Ground-breaking and International in scope and approach * Alphabetically arranged with extensive cross-referencing * Available in print and online, priced separately. The online version will include updates as subjects develop ELL2 includes: * c. 7,500,000 words * c. 11,000 pages * c. 3,000 articles * c. 1,500 figures: 130 halftones and 150 colour * Supplementary audio, video and text files online * c. 3,500 glossary definitions * c. 39,000 references * Extensive list of commonly used abbreviations * List of languages of the world (including information on no. of speakers, language family, etc.) * Approximately 700 biographical entries (now includes contemporary linguists) * 200 language maps in print and online Also available online via ScienceDirect – featuring extensive browsing, searching, and internal cross-referencing between articles in the work, plus dynamic linking to journal articles and abstract databases, making navigation flexible and easy. For more information, pricing options and availability visit www.info.sciencedirect.com. The first Encyclopedia to exploit the multimedia potential of linguistics Ground-breaking in scope - wider than any predecessor An invaluable resource for researchers, academics, students and professionals in the fields of: linguistics, anthropology, education, psychology, language acquisition, language pathology, cognitive science, sociology, the law, the media, medicine & computer science. The most authoritative, up-to-date, comprehensive, and international reference source in its field

TheAsiaInformationRetrievalSymposium(AIRS)wasestablishedbytheAsian information retrieval community after the successful series of Information - trieval with Asian Languages (IRAL) workshops held in six di?erent locations in Asia,

starting from 1996. While the IRAL workshops had their focus on inf- mation retrieval problems involving Asian languages, AIRS covers a wider scope of applications, systems, technologies and theory aspects of information retrieval in text, audio, image, video and multimedia data. This extension of the scope re?ects and fosters increasing research activities in information retrieval in this region and the growing need for collaborations across subdisciplines. We are very pleased to report that we saw a sharp increase in the number of submissions and their quality, compared to the IRAL workshops. We received 106papersfromninecountriesinAsiaandNorthAmerica,fromwhich28papers (26%) were presented in oral sessions and 38 papers in poster sessions (36%). It was a great challenge for the Program Committee to select the best among the excellent papers. The low acceptance rates witness the success of this year's conference. After a long discussion between the AIRS 2004 Steering Committee and Springer, the publisher agreed to publish our proceedings in the Lecture Notes in Computer Science (LNCS) series, which is SCI-indexed. We feel that this strongly attests to the excellent quality of the papers.

This book teaches the principles of natural language processing and covers linguistics issues. It also details the language-processing functions involved, including part-of-speech tagging using rules and stochastic techniques. A key feature of the book is the author's hands-on approach throughout, with extensive exercises, sample code in Prolog and Perl, and a detailed introduction to Prolog. The book is suitable for researchers and students of natural language processing and computational linguistics.

Held in Gaithersburg, MD, Nov. 4-6, 1992. Evaluates new technologies in information retrieval. Numerous graphs, tables and charts.

This book constitutes the refereed proceedings of the Second East European Symposium on Advances in Databases and Information systems, ADBIS '98, held in Poznan, Poland in September 1998. The 25 revised full papers presented were selected from a total of 90 submissions and six extended abstracts within a special section. "East meets West". The papers are organized in topical sections on query langues, optimization, collaborative systems, schema integration, storage and version management, object systems, knowledge discovery and the Web, and systems design.

This book introduces text analytics as a valuable method for deriving insights from text data. Unlike other text analytics publications, Practical Text Analytics: Maximizing the Value of Text Data makes technical concepts accessible to those without extensive experience in the field. Using text analytics, organizations can derive insights from content such as emails, documents, and social media. Practical Text Analytics is divided into five parts. The first part introduces text analytics, discusses the relationship with content analysis, and provides a general overview of text mining methodology. In the second part, the authors discuss the practice of text analytics, including data preparation and the overall planning process. The third part covers text analytics techniques such as cluster analysis, topic models, and machine learning. In the fourth part of the book, readers learn about techniques used to communicate insights from text analysis, including data storytelling. The final part of Practical Text Analytics offers examples of the application of software programs for text analytics, enabling readers to mine their own text data to uncover information.

Approximate reasoning is a key motivation in fuzzy sets and possibility theory. This volume provides a coherent view of this field, and its impact on database research and information retrieval. First, the semantic foundations of approximate reasoning are presented. Special emphasis is given to the representation of fuzzy rules and specialized types of approximate reasoning. Then syntactic aspects of approximate reasoning are surveyed and the algebraic underpinnings of fuzzy consequence relations are presented and explained. The second part of the book is devoted to inductive and neuro-fuzzy methods for learning fuzzy rules. It also contains new material on the application of possibility theory to data fusion. The last part of the book surveys the growing literature on fuzzy information systems. Each chapter contains extensive bibliographical material. Fuzzy Sets in Approximate Reasoning and Information Systems is a major source of information for research scholars and graduate students in computer science and artificial intelligence, interested in human information processing.

With the rising importance of multilingualism in language industries, brought about by global markets and world-wide information exchange, parallel corpora, i.e. corpora of texts accompanied by their translation, have become key resources in the development of natural language processing tools. The applications based upon parallel corpora are numerous and growing in number: multilingual lexicography and terminology, machine and human translation, cross-language information retrieval, language learning, etc. The book's chapters have been commissioned from major figures in the field of parallel corpus building and exploitation, with the aim of showing the state of the art in parallel text alignment and use ten to fifteen years after the first parallel-text alignment techniques were developed. Within the book, the following broad themes are addressed: (i) techniques for the alignment of parallel texts at various levels such as sentence, clause, and word; (ii) the use of parallel texts in fields as diverse as translation, lexicography, and information retrieval; (iii) available corpus resources and the evaluation of alignment methods. The book will be of interest to researchers and advanced students of computational linguistics, terminology, lexicography and translation, both in academia and industry.

This book constitutes the refereed proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning, IDEAL 2002, held in Manchester, UK in August 2002. The 89 revised papers presented were carefully reviewed and selected from more than 150 submissions. The book offers topical sections on data mining, knowledge engineering, text and document processing, internet applications, agent technology, autonomous mining, financial engineering, bioinformatics, learning systems, and pattern recognition.

This book is based on the workshop on New Approaches to Learning for Natural Language Processing, held in conjunction with the International Joint Conference on Artificial Intelligence, IJCAI'95, in Montreal, Canada in August 1995. Most of the 32 papers included in the book are revised selected workshop presentations; some papers were individually solicited from members of the workshop program committee to give the book an overall completeness. Also included, and written with the novice reader in mind, is a comprehensive introductory survey by the volume editors. The volume presents the state of the art in the most promising current approaches to learning for NLP and is thus compulsory reading for researchers in the field or for anyone applying the new techniques to challenging real-world NLP problems.

This book constitutes the thoroughly refereed proceedings of the 7th International Workshop on Computational Processing of the Portuguese Language, PROPOR 2006. The 20 revised full papers and 17 revised short papers presented here are organized in topical sections on automatic summarization, resources, translation, named entity recognition, tools and frameworks, systems and models, information extraction, speech processing, lexicon, morpho-syntactic studies, and Web, corpus and evaluation.

"This book provides relevant theoretical frameworks and the latest empirical research findings in biomedicine information retrieval as it pertains to linguistic granularity"--Provided by publisher.

Held in Gaithersburg, MD, August November 2-4, 1994. The conference was co-sponsored by the National Inst. of Standards and Technology (NIST) and the Advanced Research Projects Agency (ARPA) and was attended by 150 people involved in the 32 participating

groups. Evaluates new technologies in text retrieval. Includes 34 papers: indexing structures, fragmentation schemes, probabilistic retrieval, latent semantic indexing, interactive document retrieval, and much more. Numerous graphs, tables and charts.

Automatic Indexing and Abstracting of Document Texts summarizes the latest techniques of automatic indexing and abstracting, and the results of their application. It also places the techniques in the context of the study of text, manual indexing and abstracting, and the use of the indexing descriptions and abstracts in systems that select documents or information from large collections. Important sections of the book consider the development of new techniques for indexing and abstracting. The techniques involve the following: using text grammars, learning of the themes of the texts including the identification of representative sentences or paragraphs by means of adequate cluster algorithms, and learning of classification patterns of texts. In addition, the book is an attempt to illuminate new avenues for future research. Automatic Indexing and Abstracting of Document Texts is an excellent reference for researchers and professionals working in the field of content management and information retrieval.

"This book presents various applications of the growing perspective of agent technologies as they apply the web engineering"--Provided by publisher.

Statistical approaches to processing natural language text have become dominant in recent years. This foundational text is the first comprehensive introduction to statistical natural language processing (NLP) to appear. The book contains all the theory and algorithms needed for building NLP tools. It provides broad but rigorous coverage of mathematical and linguistic foundations, as well as detailed discussion of statistical methods, allowing students and researchers to construct their own implementations. The book covers collocation finding, word sense disambiguation, probabilistic parsing, information retrieval, and other applications.

The workshop on Applications of Natural Language to Information Systems (NLDB)hassince1995providedaforumforacademicandindustrialresearchers and practitioners to discuss the application of natural language to both the development and use of software applications.

Theuseofnaturallanguageinrelationtosoftwarehascontributedtoimpr- ing the development of software from the viewpoints of both the developers and the users. Developers bene?t from improvements in conceptual modeling, so- ware validation, natural language program speci?cations, and many other areas. Users bene?t from increased usability of applications through natural language query interfaces, semantic webs, text summarizations, etc. The integration of natural language and information systems has been a - search objective for a long time now. Today, the goal of good integration seems not so far-fetched. This is due mainly to the rapid progress of research in natural language and to the development of new and powerful technologies. The in- gration of natural language and information systems has become a convergent point towards which many researchers from several research areas are focussing.

The International Conference on Intelligent Computing (ICIC) was set up as an annual forum dedicated to emerging and challenging topics in the various aspects of advances in computational intelligence fields, such as artificial intelligence, machine learning, bioinformatics, and computational biology, etc. The goal of this conference was to bring together researchers from academia and industry as well as practitioners to share ideas, problems and solutions related to the multifaceted aspects of intelligent computing. This book constitutes the proceedings of the International Conference on Intelligent Computing (ICIC 2005), held in Hefei, Anhui, China, during August 23–26, 2005. ICIC 2005 received over 2000 submissions from authors in 39 countries and regions. Based on rigorous peer reviews, the Program Committee selected 563 high-quality papers for presentation at ICIC 2005; of these, 215 papers were published in this book organized into 9 categories, and the other 348 papers were published in five international journals. The organizers of ICIC 2005 made great efforts to ensure the success of this conference. We here thank the members of the ICIC 2005 Advisory Committee for their guidance and advice, the members of the Program Committee and the referees for reviewing the papers, and the members of the Publication Committee for checking and compiling the papers. We would also like to thank the publisher, Springer, for their support in publishing the proceedings in the Lecture Notes in Computer Science series. Particularly, we would like to thank all the authors for contributing their papers.

This book constitutes the refereed proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2002, held in Mexico City, Mexico in February 2002. The 44 revised papers presented together with four invited papers were carefully reviewed and selected from a total of 67 submissions. The papers are organized in topical sections on semantics, word sense disambiguation, amaphora, syntax and parsing, part of speech tagging, lexicon and corpus, text generation, morphology, speech, spelling, information extraction and information retrieval, summarization, text mining, and text classification and categorization, document processing, and demo descriptions.

Automatic Text ProcessingThe Transformation, Analysis, and Retrieval of Information by ComputerReading, Mass. : Addison-WesleyParallel Text ProcessingAlignment and Use of Translation CorporaSpringer Science & Business Media The amounts of information that are ?ooding people both at the workplace and in private life have increased dramatically in the past ten years. The number of paper documents doubles every four years, and the amount of information stored on all data carriers every six years. New knowledge, however, increases at a considerably lower rate. Possibilities for automatic content recognition in various media and for the processing of documents are therefore becoming more important every day. Especially in economic terms, the e?cient handling of information, i.e., ?- ing the right information at the right time, is an invaluable resource for any enterprise, but it is particularly important for small- and medium-sized ent- prises. The market for document management systems, which in Europe had a volume of approximately 5 billion euros in 2000, will increase considerably over the next few years. The BMBF recognized this development at an early stage. As early as in 1995, it pooled national capabilities in this ?eld in order to support research on the automatic processing of information within the framework of a large collaborative project (READ) involving both industrial companies and research

centres. Evaluation of the results led to the conclusion that research work had been successful, and, in a second phase, funding was provided for the colla- rative follow-up project Adaptive READ from 1999 to 2003. The completion of thesetwoimportantlong-termresearchprojectshascontributedsubstantiallyto improving the possibilities of content recognition and processing of handwritten, printed and electronic documents.