

Preface

There has been a large increase in the amount of information that is stored in and available from online databases and the World Wide Web. This information abundance has made the task of locating relevant information more complex. Such complexity drives the need for intelligent systems for searching and for information retrieval.

The information needed by a user is usually scattered in a large number of databases. Intelligent agents are currently used to improve the search for and retrieval of information from online databases and the World Wide Web. Research and development work in the area of intelligent agents and web technologies is growing rapidly. This is due to the many successful applications of these new techniques in very diverse problems. The increased number of patents and the diverse range of products developed using intelligent agents is evidence of this fact.

Most papers on the application of intelligent agents for web data mining and information retrieval are scattered around the world in different journals and conference proceedings. As such, journals and conference publications tend to focus on a very special and narrow topic. This book includes critical reviews of the state-of-the-art for the theory and application of intelligent agents for web data mining and information retrieval. This volume aims to fill the gap in the current literature.

The book consists of openly-solicited and invited chapters, written by international researchers in the field of intelligent agents and its applications for data mining and information retrieval. All chapters have been through a peer review process by at least two recognized reviewers and the editor. Our goal is to provide a book that covers the theoretical side, as well as the practical side, of intelligent agents. The book is organized in such a way that it can

be used by researchers at the undergraduate and post-graduate levels. It can also be used as a reference of the state-of-the-art for cutting edge researchers.

The book consists of 18 chapters covering research areas such as: new methodologies for searching distributed text databases; computational intelligence techniques and intelligent agents for web data mining; multi-agent collaborative knowledge production; case-based reasoning and rule-based parsing and pattern matching for web data mining; multilingual concept-based web content mining; customization, personalization and user profiling; text processing and classification; textual document warehousing; web data repository; knowledge extraction and classification; multi-agent social coordination; agent-mediated user profiling; multi-agent systems for electronic catalog retrieval; concept matching and web searching; taxonomy-based fuzzy information filtering; web navigation using sub-graph and visualization; and networking e-learning hosts using mobile agents. In particular, the chapters cover the following:

In Chapter I, “Necessary Constraints for Database Selection in a Distributed Text Database Environment,” *Yang and Zhang* discuss that, in order to understand the various aspects of a database, is essential to choose appropriate text databases to search with respect to a given user query. The analysis of different selection cases and different types of DTDs can help develop an effective and efficient database selection method. In this chapter, the authors have identified various potential selection cases in DTDs and have classified the types of DTDs. Based on these results, they analyze the relationships between selection cases and types of DTDs, and give the necessary constraints of database selection methods in different selection cases.

Chapter II, “Computational Intelligence Techniques Driven Intelligent Agents for Web Data Mining and Information Retrieval” by *Mohammadian and Jentzsch*, looks at how the World Wide Web has added an abundance of data and information to the complexity of information disseminators and users alike. With this complexity has come the problem of locating useful and relevant information. Such complexity drives the need for improved and intelligent search and retrieval engines. To improve the results returned by the searches, intelligent agents and other technology have the potential, when used with existing search and retrieval engines, to provide a more comprehensive search with an improved performance. This research provides the building blocks for integrating intelligent agents with current search engines. It shows how an intelligent system can be constructed to assist in better information filtering, gathering and retrieval.

Chapter III, “A Multi-Agent Approach to Collaborative Knowledge Production” by *Dodero, Díaz and Aedo*, discusses how knowledge creation or

production in a distributed knowledge management system is a collaborative task that needs to be coordinated. The authors introduce a multi-agent architecture for collaborative knowledge production tasks, where knowledge-producing agents are arranged into knowledge domains or marts, and where a distributed interaction protocol is used to consolidate knowledge that is produced in a mart. Knowledge consolidated in a given mart can, in turn, be negotiated in higher-level foreign marts. As an evaluation scenario, the proposed architecture and protocol are applied to coordinate the creation of learning objects by a distributed group of instructional designers.

Chapter IV, “Customized Recommendation Mechanism Based on Web Data Mining and Case-Based Reasoning” by *Kim*, researches the blending of Artificial Intelligence (AI) techniques with the business process. In this research, the author suggests a web-based, customized hybrid recommendation mechanism using Case-based Reasoning (CBR) and web data mining. In this case, the author uses CBR as a supplementary AI tool, and the results show that the CBR and web data mining-based hybrid recommendation mechanism could reflect both association knowledge and purchase information about our former customers.

Chapter V, “Rule-Based Parsing for Web Data Extraction” by *Camacho, Aler* and *Cuadrado*, discusses that, in order to build robust and adaptable web systems, it is necessary to provide a standard representation for the information (i.e., using languages like XML and ontologies to represent the semantics of the stored knowledge). However, this is actually a research field and, usually, most of the web sources do not provide their information in a structured way. This chapter analyzes a new approach that allows for the building of robust and adaptable web systems through a multi-agent approach. Several problems, such as how to retrieve, extract and manage the stored information from web sources, are analyzed from an agent perspective.

Chapter VI, “Multilingual Web Content Mining: A User-Oriented Approach” by *Chau* and *Yeh*, presents a novel user-oriented, concept-based approach to multilingual web content mining using self-organizing maps. The multilingual linguistic knowledge required for multilingual web content mining is made available by encoding all multilingual concept-term relationships using a multilingual concept space. With this linguistic knowledge base, a concept-based multilingual text classifier is developed to reveal the conceptual content of multilingual web documents and to form concept categories of multilingual web documents on a concept-based browsing interface. To personalize multilingual web content mining, a concept-based user profile is generated from a user’s bookmark file to highlight the user’s topics of information interests on

the browsing interface. As such, both explorative browsing and user-oriented, concept-focused information filtering in a multilingual web are facilitated.

Chapter VII, “A Textual Warehouse Approach: A Web Data Repository” by *Khrouf* and *Soulé-Dupuy*, establishes that an enterprise memory must be able to be used as a basis for the processes of scientific or technical developments. It has been proven that information useful to these processes is not solely in the operational bases of companies, but is also in textual information and exchanged documents. For that reason, the authors propose the design and implementation of a documentary memory through business document warehouses, whose main characteristic is to allow the storage, retrieval, interrogation and analysis of information extracted from disseminated sources and, in particular, from the Web.

Chapter VIII, “Text Processing by Binary Neural Networks” by *Beran* and *Macek*, describes the rather less traditional technique of text processing. The technique is based on the binary neural network Correlation Matrix Memory. The authors propose the use of a neural network for text searching tasks. Two methods of coding input words are described and tested; problems using this approach for text processing are then discussed.

In the world of artificial intelligence, the extraction of knowledge has been a very useful tool for many different purposes, and it has been tried with many different techniques. In Chapter IX, “Extracting Knowledge from Databases and ANNs with Genetic Programming: Iris Flower Classification Problem” by *Rivero*, *Rabuñal*, *Dorado*, *Pazos* and *Pedreira*, the authors show how Genetic Programming (GP) can be used to solve a classification problem from a database. They also show how to adapt this tool in two different ways: to improve its performance and to make possible the detection of errors. Results show that the technique developed in this chapter opens a new area for research in the field, extracting knowledge from more complicated structures, such as neural networks.

Chapter X, “Social Coordination with Architecture for Ubiquitous Agents — CONSORTS” by *Kurumatani*, proposes a social coordination mechanism that is realized with CONSORTS, a new kind of multi-agent architecture for ubiquitous agents. The author defines social coordination as mass users’ decision making in their daily lives, such as the mutual concession of spatial-temporal resources achieved by the automatic negotiation of software agents, rather than by the verbal and explicit communication directly done by human users. The functionality of social coordination is realized in the agent architecture where three kinds of agents work cooperatively, i.e., a personal agent that serves as a proxy for the user, a social coordinator as the service agent,

and a spatio-temporal reasoner. The author also summarizes some basic mechanisms of social coordination functionality, including stochastic distribution and market mechanism.

In Chapter XI, “Agent-Mediated Knowledge Acquisition for User Profiling” by *Andreevskaya*, *Abi-Aad* and *Radhakrishnan*, the authors discuss how, in the past few years, Internet shopping has been growing rapidly. Most companies now offer web service for online purchases and delivery in addition to their traditional sales and services. For consumers, this means that they face more complexity in using these online services. This complexity, which arises due to factors such as information overloading or a lack of relevant information, reduces the usability of e-commerce sites. In this study, the authors address reasons why consumers abandon a web site during personal shopping.

As Internet technologies develop rapidly, companies are shifting their business activities to e-business on the Internet. Worldwide competition among corporations accelerates the reorganization of corporate sections and partner groups, resulting in a break from the conventional steady business relationships. Chapter XII, “Development of Agent-Based Electronic Catalog Retrieval System” by *Nagano*, *Tahara*, *Hasegawa* and *Ohsuga*, represents the development of an electronic catalog retrieval system using a multi-agent framework, *Bee-gent*TM, in order to exchange catalog data between existing catalog servers. The proposed system agentifies electronic catalog servers implemented by distinct software vendors, and a mediation mobile agent migrates among the servers to retrieve electronic catalog data and bring them back to the departure server.

Chapter XIII, “Using Dynamically Acquired Background Knowledge for Information Extraction and Intelligent Search” by *El-Beltagy*, *Rafea* and *Abdelhamid*, presents a simple framework for extracting information found in publications or documents that are issued in large volumes and which cover similar concepts or issues within a given domain. The general aim of the work described is to present a model for automatically augmenting segments of these documents with metadata, using dynamically acquired background domain knowledge in order to help users easily locate information within these documents through a structured front end. To realize this goal, both document structure and dynamically acquired background knowledge are utilized

Web search engines are one of the most popular services to facilitate users in locating useful information on the Web. Although many studies have been carried out to estimate the size and overlap of the general web search engines, it may not benefit the ordinary web searching users; they care more

about the overlap of the search results on concrete queries, but not the overlap of the total index database. In Chapter XIV, “A Study on Web Searching: Overlap and Distance of the Search Engine Results” by *Zhu, Deng, Fang* and *Zheng*, the authors present experimental results on the comparison of the overlap of top search results from AlltheWeb, Google, AltaVista and Wisenut on the 58 most popular queries, as well as on the distance of the overlapped results.

Chapter XV, “Taxonomy Based Fuzzy Filtering of Search Results” by *Vrettos* and *Stafylopatis*, proposes that the use of topic taxonomies is part of a filtering language. Given any taxonomy, the authors train classifiers for every topic of it so the user is able to formulate logical rules combining the available topics, (e.g., Topic1 AND Topic2 OR Topic3), in order to filter related documents in a stream of documents. The authors present a framework that is concerned with the operators that provide the best filtering performance as regards the user.

In Chapter XVI, “Generating and Adjusting Web Sub-Graph Displays for Web Navigation” by *Lai, Huang* and *Zhang*, the authors relate that a graph can be used for web navigation, considering that the whole of cyberspace can be regarded as one huge graph. To explore this huge graph, it is critical to find an effective method of tracking a sequence of subsets (web sub-graphs) of the huge graph, based on the user’s focus. This chapter introduces a method for generating and adjusting web sub-graph displays in the process of web navigation.

Chapter XVII, “An Algorithm of Pattern Match Being Fit for Mining Association Rules” by *Shi* and *Zhang*, discusses the frequent amounts of pattern match that exist in the process of evaluating the support count of candidates, which is one of the main factors influencing the efficiency of mining for association rules. In this chapter, an efficient algorithm for pattern match being fit for mining association rules is presented by analyzing its characters.

Chapter XVIII, “Networking E-Learning Hosts Using Mobile Agent” by *Quah, Chen* and *Leow*, discusses how, with the rapid evolution of the Internet, information overload is becoming a common phenomenon, and why it is necessary to have a tool to help users extract useful information from the Internet. A similar problem is being faced by e-learning applications. At present, commercialized e-learning systems lack information search tools to help users search for the course information, and few of them have explored the power of mobile agent. Mobile agent is a suitable tool, particularly for Internet information retrieval. This chapter presents a mobile agent-based e-learning tool which can help the e-learning user search for course materials on the Web. A proto-

type system of cluster-nodes has been implemented, and experiment results are presented.

It is hoped that the case studies, tools and techniques described in the book will assist in expanding the horizons of intelligent agents and will help disseminate knowledge to the research and the practice communities.