W

WEB INTELLIGENCE (WI)

INTRODUCTION

The study of Web intelligence (WI) was first introduced in several papers and books [see Refs. (1–19)]. Broadly speaking, WI is a new direction for scientific research and development that explores the fundamental roles as well as practical impacts of artificial intelligence (AI),[1] such as knowledge representation, planning, knowledge discovery and data mining, intelligent agents, and social network intelligence, as well as advanced information technology (IT), such as wireless networks; ubiquitous devices; social networks; and data/knowledge grids; and the next generation of Web-empowered products, systems, services, and activities.

On one hand, WI applies results from existing disciplines to a totally new domain. On the other hand, WI introduces new problems and challenges to the established disciplines. WI may be considered as an enhancement or an extension of AI and IT (4). The WI technologies revolutionize the way in which information is gathered, stored, processed, presented, shared, and used through electoronization, virtualization, globalization, standardization, personalization, and portals.

The challenges of Internet computing research and development in the next decade will be WI centric, focusing on how we can intelligently make the best use of the widely available Web connectivity. The new WI technologies will be determined precisely by *human needs in a post-industrial era*; namely (2):

- *information* empowerment,
- *knowledge* sharing,
- virtual *social* communities,
- *service* enrichment, and
- practical *wisdom* development.

We observed that one of the most promising paradigm shifts in the Web will be driven by the notion of *wisdom*, and developing the World Wide Wisdom Web (the *Wisdom Web*, or W4) will become a tangible goal for WI research (1,3,7). The new generation of the WWW will enable humans to gain wisdom of living, working, and playing in addition to information search and knowledge queries.

Great potential exits for WI to make useful contributions to e-business (including e-commerce and e-finance), e-science, e-learning, e-government, e-community, and so on. Many specific applications and systems have been proposed and studied. In particular, the e-business activity that involves the end user is undergoing a significant revolution (10). The ability to track users' browsing behavior down to individual mouse clicks has brought the vendor and end customer closer than ever before. It is now possible for a vendor to personalize his product message for individual customers at a massive scale, which is called *targeted marketing* (or direct marketing) (11–13). Web mining and Web usage analysis play an important role in e-business for customer relationship management (CRM) and targeted marketing. Web mining is the use of data mining techniques to discover automatically and extract information from Web documents and services (10,14,15). A challenge is to explore the connection between Web mining and the related agent paradigm, such as Web farming, that is the systematic refining of information resources on the Web for business intelligence (16).

This article investigates various ways to study WI and potential applications. The next section describes what is the Wisdom Web. The section after that discusses how to develop various Web-based portals, in particular, intelligent enterprise portals for e-business intelligence, by using WI technologies. Furthermore, based on the discussion, an intelligent Web-based business-centric schematic diagram of WI-related topics and conceptual levels of WI for developing the Wisdom Web are provided in this section, respectively. The section entitled Advanced topics for studying WI describes various ways for studying WI, which include the semantics in the Web and the Web as social networks, as well as proposes new approaches for developing semantic social networks. Based on the above preparation, the section on WI-Based Targeted Marketing shows how to offer advanced features that enable e-business intelligence such as targeted marketing, which is a new business model by an interactive one-to-one communication between marketer and customer, as well as deal with the scalability and complexity of the real world, efficiently and effectively, by using the knowledge grid middleware as a new infrastructure and platform. The final section provides concluding remarks.

THE WORLD WIDE WISDOM WEB (W4)

What is the Wisdom Web?

In the movie *Star Wars: Episode II*, an interesting scene is when Obi Wan Kenobi failed to locate any relevant information about a mysterious planet (where later he discovered the clone manufacturing ground), he turned to his friend for advice. His friend, who apparently knew more than the Jedi's academy knowledge banks combined, gave the following reply: Other people seek *knowledge*, but you my friend know *wisdom*.

The reply in the above scene also provides an answer to the question: What will be the next paradigm shift in the Web and the Internet? The next paradigm shift lies in the notion of *wisdom*. The goal of the new generation WI is to enable users to gain new *wisdom* of living, working, play-

[1]Here, the term of AI includes classic AI, computational intelligence, and soft computing.

ing, and learning, in addition to *information* search and *knowledge* queries. Here, the word of *wisdom*, according to the Webster Dictionary (Page: 1658) (17), implies the following meanings (emphasis added):

1. The quality of being wise; knowledge, and *the capacity to make due use of it*; *knowledge of the best ends and the best means*; *discernment and judgment*; discretion; sagacity; skill; dexterity.
2. The results of wise judgments; scientific or practical truth; acquired knowledge; erudition.

In the Web context, the manifestation of wisdom can best be illustrated with a minimalist Wisdom Web example.

When the Web Offers Practical Wisdom

Imagine that you are taking your first trip to the city of Montreal. You would like to find a really nice place to spend your evening. So, you walk into a Cyber Cafe on Sherbrook Street (the only street that you can recognize), and decide to get some *practical wisdom* from a public *Wisdom Web* outlet. You log in with a user name, "Spiderman," and ask:

> What is the best night life in Montreal during this season of the year?

The Wisdom Web *thinks* for about a second or two and then responds:

> Spiderman, the hockey games are on during this season of the year. Would you like to go?

You reply:

> Yes.

Then the Wisdom Web suggests:

> As far as I know, there are still some tickets left and you may purchase some at the Montreal Forum. It is easy to get there by taking Metro to the Atwater station.
>
> Now you decide that this could be an interesting evening for you...

One hour later, you arrive at the ticket office by Metro, but surprisingly find that the tickets left are all for the day after tomorrow when you will be traveling in Quebec City.

As you are a bit disappointed, you notice that there is a free Wisdom Web Kiosk right beside the ticket office. Well, that is convenient. So, without too much hesitation, you log on to the Wisdom Web, again as "Spiderman." The Wisdom Web still remembers your conversations an hour ago. As soon as it recognizes that you are "Spiderman," it says to you:

> Hello Spiderman, you were in such a hurry last time that I couldn't have a chance to tell you that all tickets available here are only for the day after tomorrow. They are quite expensive too...

Ten Capabilities of the Wisdom Web

To make the above Wisdom Web scenario a reality, the following 10 fundamental capabilities have to be incorporated and standardized (2):

1. *Self-organizing servers*. The Wisdom Web will regulate automatically the functions and cooperations of related websites and application services available. A Wisdom Web server self-nominates automatically to other services its functional *roles* as well as corresponding spatial or temporal *constraints* and operational *settings*.
2. *Specialization*. A Wisdom Web server is an *agent* by itself, which is specialized in performing some roles in a certain service. The *association* of its roles with any service will be measured and updated dynamically, for instance, the association may be forgotten if it is not used for some time.
3. *Growth*. The population of *Wisdom Agents* will change dynamically, as new agents are self-reproduced by their parent agents to become more specialized or as aged agents and are deactivated.
4. *Autocatalysis*. As various roles of wisdom agents are created through specialization and are activated by the *Wisdom Search* requests, their associations with some services and among themselves must be aggregated autocatalytically. In this respect, the autocatalysis of associations is similar to the pheromone laying for positive feedback in an ant colony.
5. *Problem Solver Markup Language (PSML)*. PSML is necessary for wisdom agents to specify their roles and settings as well as relationships with any other services.
6. *Semantics*. The Wisdom Web needs to understand what are meant by "Montreal," "season," "year," and "night life," and what is the right judgment of "best," by understanding the granularities of their corresponding subjects and the whereabouts of their ontology definitions.
7. *Metaknowledge*. Besides semantic knowledge extracted and manipulated in the *Wisdom Search*, it is also essential for wisdom agents to incorporate a dynamically created source of metaknowledge that deals with the relationships between concepts and the spatial or temporal constraint knowledge in planning and executing services. It allows agents to self-resolve their conflict of interests.
8. *Planning*. In the above example, the goal is to find a function or an event that may sound attractive to a visitor. The constraint is that they must be happening during this season. Two associated subgoals are involved: To have an access to the recommended function or event, one needs a ticket. Furthermore, to go to get the ticket, one can travel by metro. In the Wisdom Web, ontology alone will *not* be sufficient.
9. *Personalization*. The Wisdom Web remembers the recent encounters and relates different episodes together, according to (1) "Spiderman," (2) time,

and (3) attainability of (sub)goals. In addition, it may identify other goals as well as courses of action for this user as their conversation continues.

10. *A sense of humor*. Although the Wisdom Web does not tell a funny story explicitly, it adds some punch lines to the situation or anxiety that "Spiderman" is presently in when he/she logs on for the second time, which will make "Spiderman" feel absurd.

Levels of WI

To develop a Wisdom Web to benefit from the information infrastructure that the Web has empowered, we have witnessed the fast development as well as applications of many WI techniques and technologies, which cover the following four conceptual levels at least:

1. *Internet-level communication, infrastructure, and security protocols*. The Web is regarded as a *computer-network system*. WI techniques for this level include Web data prefetching systems built upon Web surfing patterns to resolve the issue of Web latency. The intelligence of the Web prefetching comes from an adaptive learning process based on the observation and characterization of user surfing behavior (18,19).
2. *Interface-level multimedia presentation standards*. The Web is regarded as an *interface* for human–Internet interaction. WI techniques for this level are used to develop intelligent Web interfaces in which the capabilities of adaptive cross-language processing, personalized multimedia representation, and multimodal data processing are required.
3. *Knowledge-level information processing and management tools*. The Web is regarded as a *distributed data/knowledge base*. We need to develop semantic markup languages to represent the semantic contents of the Web available in machine-understandable formats for agent-based autonomic computing, such as searching, aggregation, classification, filtering, managing, mining, and discovery on the Web (20).
4. *Application-level ubiquitous computing and social intelligence environments*. The Web is regarded as a *basis for establishing social networks* that contain communities of people (or organizations or other social entities) connected by social relationships, such as friendship, coworking, or information exchange with common interests. They are Web-supported social networks or virtual communities. The study of WI concerns the important issues central to social network intelligence (social intelligence for short) (21). Furthermore, the multimedia contents on the Web are accessible not only from stationary platforms, but also increasingly from mobile platforms (22). Ubiquitous Web access and computing from various wireless devices needs adaptive personalization for which WI techniques are used to construct models of user interests by inferring implicitly from user behavior and actions (23,24).

In particular, the social intelligence approach presents excellent opportunities and challenges for the research and development of WI, as well as a Web-supported social network that needs to be supported by all levels of WI as mentioned above. This approach is based on the observation that the Web is now becoming an integral part of our society, and that scientists should be aware of it and take much care about handling social issues (25). Study in this area must receive as much attention as Web mining, Web agents, ontologies, and related topics.

Wisdom-Oriented Computing

Wisdom-oriented computing is a new computing paradigm aimed at providing not only a medium for seamless information exchange and knowledge sharing (20) but also a type of *man-made resources for sustainable knowledge creation, and scientific and social evolution*(2,3). The Wisdom Web, i.e., the Web that empowers wisdom-oriented computing, will reply on *grid-like service agencies* that self-organize, learn, and evolve their courses of actions to perform service tasks as well as their identities and interrelationships in Web communities. They will cooperate and compete among themselves to optimize their's as well as others, resources and utilities.

Self-organizing learning agents are computational entities that are capable of self-improving their performance in dynamically changing and unpredictable task environments. In Ref. (26), Liu has provided a comprehensive overview of several studies in the field of *autonomy oriented computing*, with in-depth discussions on self-organizing and with adaptive techniques for developing various embodiments of agent based systems, such as autonomous robots, collective vision and motion, autonomous animation, and search and segmentation agents. The core of those techniques is the notion of synthetic or emergent autonomy based on behavioral self-organization.

Developing the Wisdom Web will become a tangible goal for WI researchers and practitioners. The Wisdom Web will enable us to use the global connectivity optimally, as offered by the Web infrastructure, and most importantly, to gain the practical wisdoms of living, working, and playing, in addition to information search and knowledge queries.

To develop the new generation WI systems effectively, we need to define *benchmark* applications, i.e., a new *Turing Test*, that will capture and demonstrate the Wisdom Web capabilities (2).

Take the wisdom-oriented computing benchmark as an example. We can use a service task of compiling and generating a market report on an existing product or *a potential market report on a new product*. To get such service jobs done, an information agent on the Wisdom Web will mine and integrate available Web information, which will in turn be passed onto a market analysis agent. Market analysis will involve the quantitative simulations of customer behavior in a marketplace, instantaneously handled by other service agencies, involving a large number of semantic or computational grid agents (e.g. Ref. 27). Because the number of variables concerned may be in the order of hundreds or thousands, it can easily cost a single system years to generate one predication.

DEVELOPING INTELLIGENT PORTALS BY USING WI TECHNOLOGIES

What is a Portal?

A portal enables a company, an organization, or a community to create a *virtual organization* (or a virtual community) on the Web where key production/information steps are outsourced to partners and customers. In other words, a portal is a single gateway to personalized information needed to enable informed interdisciplinary research, services, and/or business activities. Developing intelligent portals is one of the most sophisticated applications on the Web.

Although specific features of various portals need to be considered, the common requirements of the portals for e-business, e-science, e-government, e-learning, among others, are such that

- they need a unique website (a single gateway) in which all of the contents related to the virtual organization can be accessed although such organization information is geographically distributed in multi-site, multi data repositories, and multi-institution, and
- they need to have easy access to expensive remote facilities, computing resources, and share information acquired from different subjects using different techniques and stored in dedicated knowledge-data bases.

Many organizations are implementing a corporate portal first and are then growing this solution into more of an *intelligent* B2B portal. By using a portal to tie in back-end enterprise systems, a company can manage the complex interactions of the virtual enterprise partners through all phases of the value and supply chain.

Here we would like to mention two typical types of enterprises, as examples,

- transnational corporations that have operations, subsidiaries, investments, or branches worldwide, and
- communities with many mid-sized/small-scale companies in a region,

that need such enterprise portals for supporting their e-business and e-commerce activities.

The Virtual Industry Park: An Example of Enterprise Portals

As an example for developing enterprise portals by using WI technologies, here we discuss how to construct an intelligent virtual industry park (VIP) that has been developing in our group. The VIP portal is a website in which all of the contents related to the small/medium-sized companies in Maebashi city, Japan can be accessed.

The construction process can be divided into three phases. We first constructed a basic system including the fundamental functions such as the interface for dynamically registering/updating enterprise information, the database for storing the enterprise information, automatic generation and modification of enterprise homepages, and the domain-specific, keyword-based search engine. When designing the basic system, we also started by analyzing customer performance: each customer what has bought, over time, total volumes, trends, and so on.

Although the basic system can work as a whole one, we now need to know not only past performance on the business front, but also how the customer or prospect enters our VIP portal to target products and to manage promotions and marketing campaigns. To the already demanding requirement to capture transaction data for additional analysis, we now also need to use the Web usage mining techniques to capture the clicks of the mouse that define where the visitor has been on our website. What pages has he or she visited? What is the semantic association between the pages he or she visited? Is the visitor familiar with the Web structure? Or is he or she a new user or a random one? Is the visitor a Web robot or other users? In search for the holy grail of "stickiness," we know that a prime factor is *personalization* for:

- making a dynamic recommendation to a Web user based on the user profile and usage behavior.
- automatic modification of a website's contents and organization.
- combining Web usage data with marketing data to give information about how visitors used a website for marketers.

Hence, we need to extend the basic VIP system by adding more advanced functions such as Web mining, an ontologies-based search engine, as well as automatic e-mail filtering and management.

Finally, a portal for e-business-intelligence can be implemented by adding e-business-related application functions such as targeted marketing and CRM, electronic data interchange, as well as security solution.

An Intelligent Enterprise Portal Centric Schematic Diagram of WI Technologies

From the example stated in the above subsection, we can see that developing an intelligent enterprise portal needs to apply results from existing disciplines of AI and IT to a

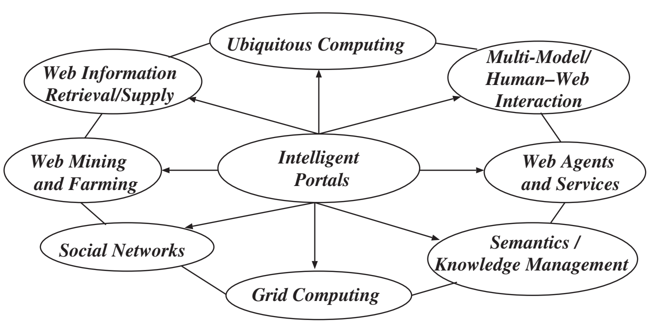

Figure 1. An intelligent enterprise portals centric schematic diagram of WI technologies.

totally new domain. On the other hand, the WI technologies are also expected to introduce new problems and challenges to the established disciplines on the new platform of the Web and the Internet. That is, WI is an enhancement or an extension of AI and IT.

To study advanced WI technologies systematically, and to develop advanced Web-based intelligent enterprise portals and information systems, we provide a schematic diagram of WI technologies from a Web-based, intelligent enterprise portals centric perspective in Fig. 1. In Fig. 1, directed lines denote that the development of intelligent enterprise portals needs to be supported by various WI related techniques, and undirected lines denote that the components of WI techniques are relevant each other.

Web Mining and Farming

The enterprise portal-based e-business activity that involves the end user is undergoing a significant revolution(10). The ability to track users' browsing behavior down to individual mouse clicks has brought the vendor and end customer closer than ever before. It is now possible for a vendor to personalize his product message for individual customers at a massive scale, which is called *targeted marketing* (or *direct marketing*) (11,13). Web mining and Web usage analysis play an important role in e-business for CRM and targeted marketing.

Web mining is the use of data mining techniques to discover and to extract information automatically from large Web data repositories such as Web documents and services (10,12,14,28). Web mining research is at the crossroads of research from several research communities, such as database, information retrieval, artificial intelligence, and especially the subareas of machine learning and natural language processing. Web mining can be divided into four classes of data available on the Web:

- *Web content*: the data that constitutes the Web pages and conveys information to the users, i.e., html, graphical, video, audio files of a Web page.
- *Web structure*: the data that formulates the hyper-link structure of a website and the Web, i.e., various HTML tags used to link one page to another and one website to another website.
- *Web usage*: the data that reflects the usages of Web resources, i.e., entries in Web browser's history and Internet temporary files, proxy server, and Web server logs.
- *Web user profile*: the data that provides demographic information about users of the website, i.e., users' registration data and customers' profile information.

Furthermore, web content, structure, and usage information, in many cases, are copresent in the same data file. For instance, the file names appeared in the log files and Web structure data contain useful content information. One may safely assume that a file named "WebLogMining.html" must contain information about web log mining. Similarly, the categories of web mining cannot be considered exclusive or isolated from each other. Web content mining sometimes must use Web structure data to classify a web page. In the same way, Web usage mining sometimes has to make use of Web content data and of Web structure information.

A challenge is to explore the connection between Web mining and the related agent paradigm such as Web farming that is the systematic refining of information resources on the Web for business intelligence (16). Web farming extends Web mining into an evolving breed of information analysis in a whole process of Web-based information management including seeding, breeding, gathering, harvesting, refining, and so on.

ADVANCED TOPICS FOR STUDYING WI

With respect to different levels of WI as mentioned in the section entitled "Levels of WI," the Web can be studied in several ways.

Studying the Semantics in the Web

One of the fundamental WI issues is to study the *semantics* in the Web, called the semantic Web, that is, modeling semantics of Web information to

- allow more of the Web content (not just form) to become machine readable and processible.
- allow for recognition of the semantic context in which Web materials are used.
- allow for the reconciliation of terminological differences between diverse user communities.

Thus, information will be machine-processible in ways that support intelligent network services such as information brokers and search agents (20,29).

Main Components of the Semantic Web. The semantic Web is a step toward intelligence of the Web. It is based on languages that make more semantic content of the page available in machine-readable formats for agent-based computing. The main components of semantic Web techniques include:

- a unifying data model such as RDF (Resource Description Framework).
- languages with defined semantics, built on RDF, such as OWL.
- ontologies of standardized terminology to mark up Web resources, used by semantically rich, service-level descriptions (such as OWL-S, the OWL-based Web Service Ontology), and to support tools that assist the generation and processing of semantic markup.

Ontologies and agent technology can play a crucial role in Web intelligence by enabling Web-based knowledge processing, sharing, and reuse between applications. Generally defined as shared formal conceptualizations of particular domains, ontologies provide a common understanding of topics that can be communicated between people and agent-based systems.

An ontology is a formal, explicit specification of a shared conceptualization (30). It provides a vocabulary of terms and relations to model the domain and specifies how you view the target world. An ontology can be very high-level, consisting of concepts that organize the upper parts of a knowledge base, or it can be domain-specific such as a chemical ontology. We here suggest three categories of ontologies: *domain-specific*, *task*, and *universal*.

A domain-specific ontology describes a well-defined technical or business domain.

A task ontology might either be domain-specific, or might be a set of ontologies with respect to several domains (or their reconstruction for that task), in which relations between ontologies are described for meeting the requirement of that task.

A universal ontology describes knowledge at higher levels of generality. It is a more general-purpose ontology (or called a common ontology) that is generated from several domain-specific ontologies. It can serve as a bridge for communication among several domains or tasks.

Roles of Ontologies. Generally speaking, a domain-specific (or task) ontology forms the heart of any knowledge information system for that domain (or task). Ontologies provide a way of capturing a shared understanding of terms that can be used by human and programs to aid in information exchange. Ontologies have been gaining popularity as a method of providing a specification of a controlled vocabulary. Although simple knowledge representation such as Yahoo's taxonomy provides notions of generality and term relations, classic ontologies attempt to capture precise meanings of terms. To specify meanings, an ontology language must be used.

Ontologies will play a major role in supporting information exchange processes in various areas. The roles of ontologies for WI include:

- communication between Web communities.
- agent communication based on semantics.
- knowledge-based Web retrieval.
- understanding Web contents in a semantic way.
- social network and Web community discovery.

More specifically, new requirements for any exchange format on the Web are:

- *Universal expressive power*. A Web-based exchange format must be able to express any form of data.
- *Syntactic interoperability*. Applications must be able to read the data and get a representation that can be exploited.
- *Semantic interoperability*. One important requirement for an exchange format is that data must be understandable. It is about defining mappings between terms within the data, which requires content analysis.

The semantic Web requires interoperability standards that address not only the syntactic form of documents, but also the semantic content. Ontologies serve as metadata schemes for the semantic Web, providing a controlled vocabulary of concepts, each with explicitly defined and machine-processible semantics.

A semantic Web also lets agents use all (meta) data on all Web pages, allowing it to gain knowledge from one site and apply it to logical mappings on other sites for ontology-based Web retrieval and e-business intelligence. For instance, ontologies can be used in e-commerce to enable machine-based communication between buyers and sellers, vertical integration of markets, and description reuse between different marketplaces. Web-search agents use ontologies to find pages with words that are different syntactically but similar semantically.

Although ontology engineering has been studied over the last decade, few (semi) automatic methods for comprehensive ontology construction have been developed. Manual ontology construction remains a tedious, cumbersome task that can easily result in a bottleneck for WI. Learning and construction of domain-specific ontology from Web contents is an important task in both text mining and WI (31–34).

Studying the Web as Social Networks

The study of the Web as a network has resulted in a better understanding of the sociology of Web content creation; it has improved the search engines on the Web dramatically and has created more effective algorithms for community mining and for knowledge management.

We can view the Web as a directed network in which each node is a static web page to another. Thus, the Web can be studied as a *graph* that connects a set of people (or organizations or other social entities) connected by a set of social relationships, such as friendship, coworking or information exchange with common interests (21,35,36).

Social Network Analysis. The main questions about the Web graph include:

- How big is the graph?
- Can we browse from any page to any other?
- Can we exploit the structure of the Web?
- What does the Web graph reveal about social dynamics?
- How to discover and manage the Web communities?

Modern social network theory is built on the work of Stanley Milgram (37). Milgram found so-called the *small-world phenomenon*, that is, typical paths took only six hops to arrive. Ravi Kumar et al. (35) observed there is a strong structural similarity between the Web as a network and social networks. The small-world phenomenon constitutes a basic property of the Web, which is not only interesting, but also useful.

Current estimates suggest that the Web graph has several billion nodes (pages of content) and an average degree of about 7. A recurrent observation on the Web graph is the prevalence of power laws: The degree of nodes are distributed according to inverse polynomial distribution (18,19,38–40).

The Web captures automatically a rich interplay between hundreds of millions of people and billions of pages of content. In essence, these interactions embody a *social network* involving people, the pages they create and view, and even the Web pages themselves. These relationships have a bearing on the way in which we create, share, and manage knowledge and information. It is our hope that exploiting these similarities will lead to progress in knowledge management and business intelligence.

The *broader* social network is a self-organizing structure of users, information, and communities of expertise (21,23). Such social networks can play a crucial role in implementing next-generation enterprise portals with functions such as data mining and knowledge management for discovery, analysis, and management of social network knowledge.

The social network is placed at the top of a four-level WI infrastructure as described in the section on “levels of WI” and is supported by functions, provided in all Levels of WI, including security, prefetching, adaptive cross-language processing, personalized multimedia representation, semantic searching, aggregation, classification, filtering, managing, mining, and discovery.

Semantic Social Networks for Intelligent Enterprise Portals. One of the most sophisticated applications on the Web today is *enterprise information portals* operating with state-of-the-art markup languages to search, retrieve, and repackage data. The enterprise portals are being developed into an even more powerful center based on component-based applications called Web Services (21,23).

WI researchers must study both centralized and distributed information structures. Information on the Web can be either globally distributed throughout the Web within multilayer over the infrastructure of Web protocols, or located locally, centralized on an intelligent portal providing Web services (i.e., the intelligent service provider) that is integrated to its own cluster of specialized intelligent applications. However, each approach has a serious flaw. As pointed out by Alesso and Smith (23), the intelligent portal approach limits uniformity and access, whereas the global semantic Web approach faces combinatory complexity limitations.

A way to solve the above issue is to develop and use the Problem Solver Markup Language (PSML), for collecting globally distributed contents and knowledge from Web-supported, semantic social networks and incorporating them with locally operational knowledge/databases in an enterprise or community for local centralized, adaptable Web intelligent services.

The core of PSML is distributed inference engines that can perform automatic reasoning on the Web by incorporating contents and meta-knowledge autonomically collected and transformed from the semantic Web with locally operational knowledge-data bases. A feasible way as the first step to implement such a PSML is to use existing Prolog-like logic language with agent technologies. In our current experiments, KAUS is used for representation of local information sources and for inference and reasoning. KAUS is a knowledge management system developed in our group that involves data/knowledge bases on the basis of an extended first-order predicate logic and relational data model (41,42). KAUS enables representation of knowledge and data in the first-order logic with data structure in multi-level and can be easily used for inference and reasoning as well as transforming and managing both knowledge and data.

By using this information transformation approach, the dynamic, global information sources on the Web can be combined with the local information sources in an enterprise portal for decision making and e-business intelligence.

Soft Computing for WI

Another challenging problem in WI is how to deal with uncertainty of information on the wired and wireless Web. Adapting existing soft computing solutions, when appropriate for WI applications, must incorporate a robust notion of learning that will scale to the Web, adapt to individual user requirements, and personalize interfaces. Ongoing efforts exist to integrate logic (including nonclassical logic), artificial neural networks, probabilistic and statistical reasoning, fuzzy sets, rough sets, granular computing, genetic algorithm, and other methodologies in the soft computing paradigm, to construct a hybrid approach/system for Web intelligence.

WI-BASED TARGETED MARKETING

An enterprise portal for business intelligence needs the function of WI-based targeted marketing, which is integrated with WI related capabilities such as Web mining, the ontologies-based search engine, personalized recommendation, as well as automatic e-mail filtering and management (8).

Targeted marketing aims at obtaining and maintaining direct relationships between suppliers and buyers within one or more product/market combinations. Targeted marketing becomes more and more popular because of the increased competition and the cost problem.

Furthermore, the scope of targeted marketing can be expanded from considering only how products are distributed, to include enhancing the relationships between an organization and its customers (43) because the strategic importance of long-term relationships with customers. In other words, once customers are acquired, customer retention becomes the target. Retention through customer satisfaction and loyalty can be improved greatly by acquiring and exploiting knowledge about these customers and their needs. Such targeted marketing is called “targeted relationship marketing” or “CRM” (44).

The Market Value Function (MVF) Model

In addition to WI related capabilities, targeted marketing is an important area of applications for data mining and for data warehousing (4,45). Although standard data mining methods may be applied for the purpose of targeted marketing, many specific algorithms need to be developed and applied for direct marketer to make decisions effectively.

Let us consider now a typical problem of targeted marketing. Suppose a health club needs to expand its operation by attracting more members. Assume that each existing member is described by a finite set of attributes. It is natural to examine existing members to identify their common features. Information about the health club may be sent to nonmembers who share the same features of members or similar to members. Other examples include promotion of special types of phone services and marketing of different classes of credit cards. In this case, we explore the relationships (similarities) between people (objects) based on their attribute values. The underlying assumption is that *similar type of people tend to make similar decisions and to choose similar services*. Techniques for mining association rules may not be applicable directly to this type of targeted marketing. One may produce too many or too few rules. The selection of a good set of rules may not be an easy task. Furthermore, the use of the derived rules may produce too many or too few potential new members.

To address this issue, we proposed a new model for targeted marketing by focusing on the issues of knowledge representation and computation of market values (4,12). More specifically, we assume that each object is represented by its values on a finite set of attributes. Also, we assume that market values of objects can be computed using a linear market value function. Thus, we may consider the proposed model to be a *linear* model, which is related to, but is different from, the linear model for information retrieval.

Let U be a finite universe of objects. Elements of U may be customers or products we are interested in market oriented decision making. The universe U is divided into three pair-wise disjoint classes, i.e., $U = P \cup N \cup D$. The sets P, N, and D are called *positive*, *negative*, and *don't know* instances, respectively. Take the earlier health club example, P is the set of current members, N is the set of people who had refused to join the club previously, and D is the set of the rest. The set N may be empty. A targeted marketing problem may be defined as finding elements from D, and possibly from N, that are similar to elements in P, and possibly dissimilar to elements in N. In other words, we want to identify elements from D and N that are more likely to become new members of P. We are interested in finding a market value function so that elements of D can be ranked accordingly.

Information about objects in a finite universe is given by an information table (46,47). The rows of the table correspond to objects of the universe, the columns correspond to attributes, and each cell is the value of an object with respect to an attribute. Formally, an information table is a quadruple:

$$S = (U, At, \{V_a | a \in At\}, \{I_a | a \in At\})$$

where U is a finite nonemptyset of objects, At is a finite nonempty set of attributes, V_a is a nonempty set of values for $a \in At$, $I_a : U \rightarrow V_a$ is an information function for $a \in At$. Each information function I_a is a total function that maps an object of U to exactly one value in V_a. An information table represents all available information and knowledge. Objects are only perceived, observed, or measured by using a finite number of properties (46).

A market value function (MVF) is a real-valued function from the universe to the set of real numbers, $r : U \rightarrow \Re$. In the context of information retrieval, the values of r represent the potential usefulness or relevance of documents with respect to a query. According to the values of r, documents are ranked. For the targeted marketing problem, a market value function ranks objects according to their potential market values. For the health club example, a market value function ranks people according to their likelihood of becoming a member of the health club. The likelihood may be estimated based on its similarity to a typical member of P.

We studied the simplest form of market value functions, i.e., the linear discriminant functions. Let $u_a : V_a \rightarrow \Re$ be a utility function defined on V_a for an attribute $a \in At$. The utility $u_a(\cdot)$ may be positive, negative, or zero. For $v \in V_a$, if $u_a(v) > 0$ and $I_a(x) = v$, i.e., $u_a(I_a(x)) > 0$, then attribute a has a positive contribution to the overall market value of x. If $u_a(I_a(x)) < 0$, then a has a negative contribution. If $u_a(I_a(x)) = 0$, then a has no contribution. The pool of contributions from all attributes is computed by a linear market value function of the following form:

$$r(x) = \sum_{a \in At} w_a u_a(I_a(x)) \tag{1}$$

where w_a is the weight of attribute a. Similarly, the weight w_a may be positive, negative, of zero. Attributes with larger weights (absolute value) are more important, and attributes with weights close to zero are not important. The overall market value of x is a weighted combination of utilities of all attributes. By using a linear market value function, we have implicitly assumed that contributions made by individual attributes are independent. Such an assumption is known as utility independence assumption commonly. Implications of utility independence assumption can be found in literature of multi-criteria decision making (48).

The market value model proposes a linear model to solve the target selection problem of targeted marketing by drawing and extending result from information retrieval (4,12). It is assumed that each object is represented by values of a finite set of attributes. A market value function is a linear combination of utility functions on attribute values, which depends on two parts: *utility function* and *attribute weighting*.

The market value function has some advantages. First, it can rank individuals according to their market value instead of classifying; second, the market value functions is interpretable; and last, the system of the market value function can perform without expertise.

Multi-Aspect Analysis in Multiple Data Sources

Generally speaking, customer data can be obtained from multiple customer touchpoints. In response, multiple data sources that are obtained from multiple customer touchpoints, including the Web, wireless, call centers, and brick-and-mortar store data, need to be integrated into a distributed data warehouse that provides a multi faceted view of their customers, their preferences, interests, and expectations for multi aspect analysis. Hence, a multi strategy and multi agent data mining framework is required (6,49).

One of main reasons for developing a multi agent data mining system is that we cannot expect to develop a single data mining algorithm that can be used to solve all targeted marketing problems because of the complexity of real-world applications. Hence, various data mining agents need to be used cooperatively in the multi step data mining process for performing multi aspect analysis as well as multi level conceptual abstraction and learning.

The other reason for developing a multi agent data mining system is that when performing multi aspect analysis for complex targeted marketing problems, a data mining task needs to be decomposed into subtasks. Thus, these sub tasks can be solved by using one or more data mining agents that are distributed over different computers and multi data repositories on the Internet. The decomposition problem leads us to the problem of distributed cooperative system design.

In the VIP stated in the section on the virtaul Industry Park for instance, mainly three kinds of data sources are considered, namely, customer database, products database, and Web farming database. Furthermore, in addition to the MVF based data mining method (12) mentioned in the section on the MVF model, we have developed various data mining methods, such as the GDT-RS inductive learning system for discovering classification rules (50), the LOI (learning with ordered information) for discovering important features (51,52), as well as the POM (peculiarity oriented mining) for finding peculiarity data/rules (53), to deal with each of such data sources, separately, for various services oriented multi aspect data analysis.

However, when we try to integrate the three kinds of data sources together into the advanced VIP system, we must know how to interact with each of those sources to extract the useful pieces of information, which then have to be combined for building the expected answer to the initial request. Hence, the core question is how to manage, represent, integrate, and use the information coming from huge, distributed, multiple-data sources.

Here, we would like to emphasize that how to manage, analyze, and use the information intelligently from different data sources is a problem that not exists only in the e-business field, but also in e-science, e-learning, e-government, as well as all WI systems and services (54,55). The development of enterprise portals and e-business intelligence is a good example for trying to solve such problem.

Building a Data Mining Grid

To implement an enterprise portal (e.g., the VIP discussed previously) for Web-based targeted marketing and business intelligence, a new infrastructure and platform as the *middleware* is required to deal with large, distributed data sources for multi aspect analysis. One methodology is to create a grid-based, organized society of data mining agents, called a *Data Mining Grid* on the grid computing platform (e.g., the Globus toolkit) (27,55–58). A data mining and must do the following:

- Develop various data mining agents, as mentioned in the section on the MVP model, for various services, oriented multiaspect data analysis;
- Organize the data mining agents into a multi layer grid, such as a data-grid, mining-grid, or knowledge-grid, under the Open Grid Services Architecture that aligns firmly with service-oriented architecture and Web services and understands the user's questions, transforms them to data mining issues, discovers the resources and information about the issues, and obtains a composite answer or solution.
- Use a conceptual model with three level workflows, namely data flow, mining flow, and knowledge flow, with respect to the data grid, the mining grid, and the knowledge grid, respectively, for managing the grid of data mining agents for multi aspect analysis in distributed, multiple-data sources and for organizing the dynamic, status-based business processes.

  That is, the data mining grid is made of many smaller components that are called *data mining agents*. Each agent by itself can only do one simple thing. Yet when we join these agents in a *grid*, this implements more complex targeted marketing and business intelligence tasks.

  Furthermore, ontologies are also used for description and for integration of multi data source and grid-based data mining agents in data mining process planning (6,7,28), which will provide the following:

- a formal, explicit specification for integrated use of multiple data sources in a semantic way.
- a conceptual representation about the sorts and properties of data/knowledge and data mining agents, as well as relations between data/knowledge and data mining agents.
- a vocabulary of terms and relations to model the domain, and specifying how to view the data sources and how to use data mining agents.
- a common understanding of multiple data sources that can be communicated between grid-based data mining agents.

CONCLUDING REMARKS

WI has been recognized as one of the most important as well as the fastest-growing IT research fields in the era of the World Wide Web, knowledge Web, grid computing, intelligent agent technology, and ubiquitous social computing. WI technologies will continue to produce the new tools and the infrastructure components necessary for creating intelligent enterprise portals that can serve users *wisely*.

To meet the strong demands for participation and the growing interests in WI, the Web Intelligence Consortium (WIC) was formed in spring 2002. The WIC (http://wi-consortium.org/) is an international non-profit organization dedicated to advancing world-wide scientific research and industrial development in the field of WI. It promotes collaborations among world wide WI research centers and organizational members, technology showcases at WI related conferences and workshops, WIC official book and journal publications, WIC newsletters, and WIC official releases of new industrial solutions and standards.

In addition to major WI related conferences/workshops, such as IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, and numerous special issues in international journals/magazines, such as IEEE Computer, a WI-focused scientific journal, *Web Intelligence and Agent Systems: An International Journal* (refer to the WIC homepage), has been providing a standard international forum for disseminating results of advanced research and development in the field of WI.

The interest in WI is growing very fast. We would like to invite everyone, who are interested in the WI related research and development activities, to join the WI community. Your input and participation will determine the future of WI.

ACKNOWLEDGMENTS

We are very grateful to those who have joined or supported the WI community, members of the WIC advisory board, WIC technical committee, and WIC research centres, as well as keynote/invited speakers of WI-IAT conferences, in particular, J. Bradshaw, W. Buntine, N. Cercone, P. Doherty, B. B. Faltings, E.A. Feigenbaum, G. Gottlob, J. Hendler, N. Jennings, W.L. Johnson, C. Kesselman, P. Langley, H. Lieberman, V. Lesser, J. McCarthy, T. M. Mitchell, S. Ohsuga, P. Raghavan, Z. W. Ras, P. Schuster, A. Skowron, K. Sycara, B. Wah, M. Wooldridge, X. Wu, P. S. Yu, and L. A. Zadeh. We thank them for their strong support.

REFERENCES

1. J. Liu, N. Zhong, Y. Y. Yao, Z. W. Ras, The wisdom web: new challenges for web intelligence (WI), *J. Intell. Inform. Sys.*, **20**(1): 5–9, 2003.
2. J. Liu, Web intelligence (WI): what makes wisdom web? *Proc. Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03),* 2003, pp. 1596–1601.
3. J. Liu, New challenges in the world wide wisdom web (W4) research, in N. Zhong, et al. (eds.), *Foundations of Intelligent Systems*, LNAI 2871, Springer, 2003, pp. 1–6.
4. Y. Y. Yao, N. Zhong, J. Liu, and S. Ohsuga, Web intelligence (WI): research challenges and trends in the new information age, in N. Zhong, et al. (eds.), *Web Intelligence: Research and Development*, LNAI 2198, Springer, 2001, pp. 1–17.
5. N. Zhong, J. Liu, Y. Y. Yao, and S. Ohsuga, Web intelligence (WI), *Proc. 24th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC 2000)*, Piscataway, NJ: IEEE CS Press, 2000, pp. 469–470.
6. N. Zhong, Y. Y. Yao, J. Liu, and S. Ohsuga (eds.), *Web Intelligence: Research and Development*, LNAI 2198, New York: Springer, 2001.
7. N. Zhong, J. Liu, and Y. Y. Yao, In search of the wisdom web,. *IEEE Computer*, **35**(11): 27–31, 2002.
8. N. Zhong, J. Liu, and Y.Y. Yao, (eds.), *Web Intelligence*, New York: Springer, 2003.
9. N. Zhong, J. Liu, and Y. Y. Yao, Envisioning intelligent information technologies (iIT) from the stand-point of web intelligence (WI), *Commun. ACM*, **50**(3): 89–94, 2007.
10. J. Srivastava, R. Cooley, M. Deshpande, P. Tan, Web usage mining: discovery and applications of usage patterns from web data, *SIGKDD Explorations, Newsletter of SIGKDD*, **1**: 12–23, 2000.
11. A. R. Simon, S. L. Shaffer, *Data Warehousing and Business Intelligence for e-Commerce*, San Francisco, CA: Morgan Kaufmann, 2001.
12. Y. Y. Yao, N. Zhong, J. Huang, C. Ou, and C. Liu, Using market value functions for targeted marketing data mining, *Interna. J. Pattern Recogn. Artificial Intell.*, **16**(8): 1117–1131, 2002.
13. N. Zhong, J. Liu, and Y. Y. Yao, Web intelligence (WI): a new paradigm for developing the wisdom web and social network intelligence, in N. Zhong, et al. (eds.), *Web Intelligence*, New York: Springer, 2003, pp. 1–16.
14. R. Kosala and H. Blockeel, Web mining research: a survey, *ACM SIGKDD Explor. News.*, **2**: 1–15, 2000.
15. Z. Lu, Y. Y. Yao, N. Zhong, Web log mining, in N. Zhong, et al. (eds.), *Web Intelligence*, New York: Springer, 2003, pp. 172–194.
16. R. D. Hackathorn, *Web Farming for the Data Warehouse*, San Francisco, CA: Morgan Kaufmann, 2000.
17. N. Porter (ed.), *Webster's Revised Unabridged Dictionary*, G&C. Merriam Co, 1913.
18. J. Liu, S. Zhang, Y. Ye, Agent-based characterization of web regularities, in N. Zhong, et al. (eds.), *Web Intelligence*, New York: Springer, 2003, pp. 19–36.
19. J. Liu, S. Zhang, J. Yang, Characterizing web usage regularities with information foraging agents, *IEEE Trans. Know. Data Engin.* **16**(4): 2004.
20. T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, *Scientific Am.* **284**: 34–43, 2001.
21. P. Raghavan, Social networks: from the web to the enterprise, *IEEE Internet Computing*, **6**(1): 91–94, 2002.
22. M. Weiser, The future of ubiquitous computing on campus, *CACM*, **41**(1): 41–42, 1998.
23. H. P. Alesso, C. F. Smith, *The Intelligent Wireless Web*, Reading, MA: Addison-Wesley, 2002.
24. D. Billsus, et al., Adaptive interfaces for ubiquitous web access, *CACM*, **45**: 34–38, 2002.
25. T. Nishida, Social intelligence design for the web, *IEEE Computer*, **35**(11): 37–41, 2002.
26. J. Liu, *Autonomous Agents and Multi-Agent Systems: Explorations in Learning, Self-Organization and Adaptive Computation*, Singapore: World Scientific, 2001.
27. F. Berman, From teragrid to knowledge grid, *CACM*, **44**: 27–28, 2001.
28. N. Zhong, Knowledge discovery and data mining, *The Encyclopedia of Microcomputers*, **27**(suppl. 6): 235–285, 2001.
29. S. Decker, P. Mitra, and S. Melnik, Framework for the semantic web: an RDF tutorial, *IEEE Internet Comp.*, **4**(6): 68–73, 2000.
30. D. Fensel, *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, New York: Springer, 2001.
31. Y. Li and N. Zhong, Mining Ontology for Automatically Acquiring Web User Information Needs, *IEEE Trans. Know. Data Engineer.*, **18**(4): 554–568, 2006.
32. A. Maedche and S. Staab, Ontology learning for the semantic web, *IEEE Intell. Sys.*, **16**(2): 72–79, 2001.

33. M. Missikoff, R. Navigli, P. Velardi, Integrated approach to web ontology learning and engineering, *IEEE Computer*, **35**(11): 60–63, 2002.
34. N. Zhong, Representation and construction of ontologies for web intelligence, *Internat. J. Foundations Comp. Sci.*, **13**(4): 555–570, 2002.
35. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, The web and social networks, *IEEE Computer*, **35** (11): 32–36, 2002.
36. W. Li, N. Zhong, J. Liu, Y. Y. Yao, C. Liu, Perspective of Applying the Global E-mail Network, *Proc. 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06)*, IEEE Computer Society Press, pp. 117–120, 2006.
37. S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge, MA: Cambridge University Press, 1994.
38. R. Albert, H. Jeong, A. L. Barabasi, Diameter of the world-wide web, *Nature*, **410**: 130–131, 1999.
39. B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, R. M. Lukose, Strong regularities in world wide web surfing, *Science*, **280**: 96–97, 1997.
40. B. A. Huberman, L. A. Adamic, Growth dynamics of the world-wide web, *Nature,* **410**: 131, 1999.
41. S. Ohsuga, Framework of knowledge based systems - multiple meta-level architecture for representing problems and problem solving processes, *Knowledge Based Sys.*, **3**(4): 204–214, 1990.
42. H. Yamauchi and S. Ohsuga, Loose coupling of KAUS with existing RDBMSs, *Data & Knowledge Engineering*, **5**(4): 227–251, 1990.
43. W. Klosgen and J. M. Zytkow, *Handbook of Data Mining and Knowledge Discovery*, Oxford: Oxford University Press, 2002.
44. R. Stone, *Successful Direct Marketing Methods*, 6th ed., Lincolnwood, IL: NTC Business Books , 1996.
45. P. Van Der Putten, Data mining in direct marketing databases, in W. Baets (ed)., *Complexity and Management: A Collection of Essays*, Singapore: World Scientific, 1999.
46. Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning about Data*, Dordrecht: Kluwer, 1991.
47. Y.Y. Yao and N. Zhong, Granular computing using Iinformation tables, in T. Y. Lin, Y. Y. Yao, and L. A. Zadeh (eds.), *Data Mining, Rough Sets and Granular Computing*, Berlin: Physica-Verlag, 2002, pp. 102–124.
48. P. C. Fishburn, Seven independence concepts and continuous multiattribute utility functions, *J. Math. Psycho.*, **11**: 294–327, 1974.
49. N. Zhong, C. Liu, and S. Ohsuga, Dynamically organizing KDD process, *Internat. J. Pattern Recog. Artif. Intell.*, **15**(3): 451–473, 2001.
50. N. Zhong, J. Z. Dong, C. Liu, and S. Ohsuga, A hybrid model for rule discovery in data, *Knowledge Based Sys.* **14**(7): 397–412, 2001.
51. Y. Sai, Y. Y. Yao, and N. Zhong, Data analysis and mining in ordered information tables, *Proc. 2001 IEEE International Conference on Data Mining (ICDM'01)*, Piscataway, NJ: IEEE Computer Society Press, 2001, pp. 497–504.
52. N. Zhong, Y. Y. Yao, J. Z. Dong, and S. Ohsuga, Gastric cancer data mining with ordered information, in J. J. Alpigini, et al. (eds.), *Rough Sets and Current Trends in Computing*, LNAI 2475, New York: Springer, 2002, pp. 467–478.
53. N. Zhong, Y. Y. Yao, and M. Ohshima, Peculiarity oriented multi-database mining, *IEEE Trans. Knowl. Data Engineer.*, **15**(4): 952–960, 2003.
54. J. Hu and N. Zhong, Organizing multiple data sources for developing intelligent e-Business Portals, *Data Mining Know. Dis.*, **12** (2–3): 127–150, 2006.
55. M. Cannataro, and D. Talia, The knowledge grid, *CACM*, **46**: 89–93, 2003.: N. Zhong, J. Hu, S. Motomura, J. L. Wu, and C. Liu, Building a data mining grid for multiple human brain data analysis, *Computat. Intell.*, **21**(2): 177–196, 2005.
56. I. Foster and C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*, San Francisco, CA: Morgan Kaufmann, 1999.
57. I. Foster and C. Kesselman, *The Grid 2: Blueprint for a New Computing Infrastructure*, San Francisco, CA: Morgan Kaufmann, 2004.
58. J. Nabrzyski, J. M. Schopf, J. Weglarz, *Grid Resource Management*, Dordrecht: Kluwer, 2004.

FURTHER READING

A. Congiusta, A. Pugliese, D. Talia, and P. Trunfio, Designing Grid Services for distributed knowledge discovery, *Web Intell. Agent Sys,* **1**(2): 91–104, 2003.

J. A. Hendler and E. A. Feigenbaum, Knowledge is power: the semantic web vision, in N. Zhong, et al. (eds.), *Web Intelligence: Research and Development*, LNAI 2198, Springer, 2001, 18–29.

N. Zhong and J. Liu (eds.), *Intelligent Technologies for Information Analysis*, New York: Springer, 2004.

NING ZHONG
Maebashi Institute
of Technology
Maebashi City, Japan

JIMING LIU
University of Windsor
Windsor, Ontario, Canada

YIYU YAO
University of Regina
Regina, Saskatchewan, Canada