# Envisioning Intelligent Information Technologies through the Prism of Web Intelligence

*Users and businesses alike get to turn their raw data into new science, technology, and money.*

## By Ning Zhong, Jiming Liu, and Yiyu Yao

Intelligent information technologies (iIT) focus on Web intelligence (WI), emphasizing the hybridization of techniques, along with multi-phase, distributed, and parallel processing. WI promises useful contributions to e-business intelligence and other Web-based technologies, including e-finance, e-science, e-learning, and e-service, and represents a significant benefit in IT development [9, 10, 12]. Although many IT techniques have been developed for intelligent information processing, even the most advanced are not yet mature enough to solve complex real-world problems. iIT can be regarded as the new generation of IT, encompassing the theories and applications of artificial intelligence (AI), pattern recognition, learning theory, data warehousing, data mining, knowledge discovery, grid computing, ubiquitous computing, autonomous agents, and multi-agent systems in the context of IT applications (such as e-commerce, business intelligence, social intelligence, knowledge grid, and knowledge community).

When investigating the future of IT, the computer science community should adopt an iIT perspective for several reasons. First, living in an information age, we are constantly developing new information media and technologies. Even personal information has become an important commodity. Enormous numbers of

new data records are generated every second of every day. It must be summarized and synthesized to support problem solving and decision making in business, science, government, and university organizations. The continued growth of related data collection efforts ensures that the fundamental problem addressed by iIT—how one understands and uses one's data—will continue to be critical.

In the recent sea change computer scientists have

> The W4 generation will enable us to gain practical wisdom simply from living, working, and playing, in addition to conventional information search and knowledge queries.

experienced in AI research, it is now more common to build on existing theories than propose new ones, base claims on rigorous theorems or hard experimental evidence rather than on intuition, and be relevant to real-world rather than toy applications. iIT is an interdisciplinary field. Solving real-world problems involves techniques developed not only in the AI community but in other related communities, including statistics, cognitive science, and neuroscience. iIT also facilitates development of human and social intelligence.

Data is the source of human knowledge. By analyzing and using it, software engineers are able to do two things:

- Turn it into new science and technology, helping discover new scientific laws, protect and develop the environment in which we all live, and facilitate development of new science and technology; and
- Turn it into money, using it to make strategic and tactical business decisions, win new customers, retain existing customers, and reduce the cost (and waste) of doing business.

A key iIT characteristic is the ability to combine AI, computational intelligence, WI, and intelligent agents in the design and implementation of intelligent Web-based information systems [10]. iIT development should thus be based on three main factors:

- New requirements for real-world applications (such as those in e-business, including e-commerce, e-finance, e-service, e-science, e-learning, e-government, and e-community);
- New architectures and methods for data repositories and data mining applications that satisfy these requirements; and
- New platforms and methods for large-scale distributed data sources that satisfy these requirements.

The key is being able to deal with the scalability and complexity of real-world problems. Doing so involves at least three major development methodologies: The first is hybridization, which optimally utilizes the advantages of existing methods (such as logic, including nonclassical logic, artificial neural networks, probabilistic and statistical reasoning, fuzzy sets, rough sets, and genetic algorithms).

Next is multi-phase process, or the methodology needed to solve complex real-world problems. Any future information system will have to integrate multiple subsystems. Processing information in them will not necessarily follow a well-defined linear sequence. Their operation involves multiple interrelated, perhaps iterative, phases. The dynamic organization of these phases is crucial to a system's viability.

The third is distributed and parallel processing. The Web's distributed, decentralized control means distributed and parallel processing is a must for the

next generation of IT. Grid computing makes this a relatively easy task. The full potential of distributed and parallel processing must be realized in new-generation information processing systems.

## IN DATA MINING

The three methodologies—hybridization, multiphase process, and distributed and parallel processing—have been cited in many studies of intelligent agents, WI, and data mining. Their application is illustrated in the field of data mining. Data mining may be viewed as an interdisciplinary field combining results from many other fields. In order to systematically deal with real-world data, a useful methodology involves constructing a hybrid system for data mining-related processes (such as data selection, feature extraction and reduction, knowledge discovery, and visualization).

Data mining usually involves multiple steps, including data preparation, preprocessing, search for hypothesis generation, pattern formation, knowledge evaluation, representation, refinement, and management. It may also be iterative until the mined result is satisfactory for the user's purpose [2, 11]. As different types and sizes (gigabytes or even terabytes) of data are accumulated on multiple sites in large organizations, a particular user may need to access many data sources. The system must therefore support distributed mining, combining partial results into a meaningful whole.

## HOW WI REPRESENTS iIT

The study of WI, introduced in [5, 9, 10, 12] explores the fundamental roles and practical effect of AI[1] (such as knowledge representation, planning, knowledge discovery and data mining, intelligent agents, and social network intelligence), as well as advanced IT (such as wireless networks, ubiquitous devices, social networks, and data/knowledge grids) on next-generation Web-based products, systems, services, and activities. On the one hand, WI applies results from existing disciplines to a totally new domain. On the other, WI introduces new problems and challenges to the established disciplines. WI may be viewed as an enhancement and/or extension of AI and IT.

Internet computing research and development in the next decade will be WI-centric, focusing on how to best use widely available Web connectivity. The new WI technologies will aim to satisfy five main post-industrial human needs [5]:

- Information empowerment;
- Knowledge sharing;
- Virtual social communities;
- Service enrichment; and
- Practical wisdom development.

One promising paradigm shift on the Web will be driven by the notion of wisdom, and developing the World Wide Wisdom Web (the Wisdom Web, or W4) will be a tangible goal for WI research [5]. The W4 generation will enable us to gain practical wisdom simply from living, working, and playing, in addition to conventional information search and knowledge queries.

WI has great potential as a key iIT in intelligent enterprise portals for e-business intelligence, enabling organizations to create a virtual enterprise where production steps are outsourced to multiple partners. Many organizations today implement a corporate portal first, then grow it into more of an intelligent B2B portal. By using a portal to link back-end enterprise systems, an organization can manage the complex interactions of its virtual enterprise partners through all phases of the value and supply chain.

Developing intelligent enterprise portals involves a deep understanding of both centralized and distributed information structures on the Web. Information and knowledge on the Web are either globally distributed via the multilayer infrastructure of Web protocols or located locally, centralized on an intelligent portal providing Web services. However, neither approach is perfect. As pointed out in [1], the intelligent portal approach limits the uniformity of data formats and access, while the global semantic Web approach faces limitations involving combinational complexity.

Addressing these issues involves developing and using a Web-based problem-solving system for portal-centralized, query-answering intelligent Web services and decision making [9]. The core of such a system is the Problem Solver Markup Language (PSML)—designed to represent multiple information for Web-based problem solving—and PSML-based distributed Web inference engines [5]. When developing intelligent portals based on WI technologies, PSML must provide the following support functions:

- Expressive power and functional support for complex adaptive, distributed problem solving;
- Automatic reasoning on the Web by incorporating globally distributed content and metaknowledge automatically collected from the Semantic Web and from social networks with local databases and knowledge bases;

---

[1]Here the term AI includes all aspects of iIT.
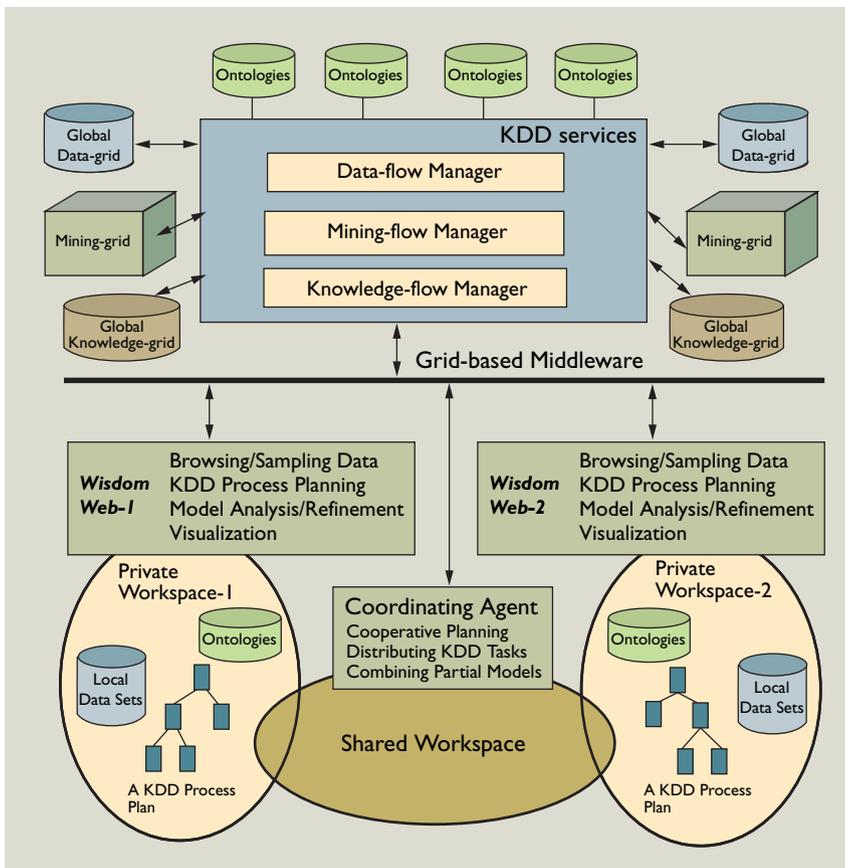
farming, the ontology-based search engine/question-answering system, personalized recommendation, and automatic email filtering and management) [9, 10]. Being able to track users' browsing behavior down to individual mouse clicks has brought vendors and their end customers closer than ever. It is now possible for vendors to personalize their product message for individual customers on a massive scale on the Web. Web farming extends Web mining into information analysis for Web-based information, including seeding, breeding, gathering, harvesting, and refining [4].

Customer data can be obtained from multiple customer touchpoints. In response, multiple data sources, including the Web, wireless communication and devices, call center, and brick-and-mortar store data, should be integrated into a single data warehouse to provide a view of customers, including their personal preferences, interests, and expectations. A multi-strategy, multi-agent data mining framework is required for the related analysis [11].

The main reason for developing a multi-agent data mining system is that various data mining agents must be able to cooperate in the multi-step data mining process, performing multi-aspect analysis, as well as multi-level conceptual abstraction and learning. Another reason is that a data mining task is decomposed into sub-tasks. They can be solved using one or more data mining agents distributed over different computers. The decomposition problem leads developers to the challenge of distributed cooperative system design.

A new infrastructure and platform is also needed as middleware to enable Web-based direct marketing for multi-aspect analysis from multiple data sources. One way to perform this direct marketing is to create a grid-based, organized society of data mining agents, or data mining grid, on the grid computing platform (such as the Globus toolkit) [3]. Various data mining agents are used for multi-aspect data analysis and targeted marketing tasks; they are organized into a grid with multi-layer components (such as data grid, mining grid, and

• Representation and organization of multiple data/knowledge sources for distributed Web inference and reasoning;
• Combined reasoning methods; and
• Personalized models of user behavior, dynamically representing and managing it.

One way to begin to implement a PSML is to use a Prolog-like logic language with agent technologies. In our 2004 experiments we used a knowledge acquisition and utilization system (KAUS) to represent local information sources, as well as for inference and reasoning. KAUS is a knowledge management system involving databases and knowledge based on an extended first-order predicate logic and data model [6]. KAUS enables representation of knowledge and data in the first-order logic with data structure for inference and reasoning, as well as for transforming and managing knowledge and data.

Using an information-transformation approach helps developers combine the Web's dynamic, global information sources with local information sources in an enterprise portal for decision making and e-business intelligence.

Targeted telemarketing (also called direct marketing) is a new marketing strategy for e-business intelligence [8] that integrates Web-based direct marketing with other WI functions (such as Web mining and

knowledge grid) under the Open Grid Services Architecture, which responds to user queries by transforming them into data mining methodologies and discovering resources and information about them.

Computer scientists are able to use a conceptual model with three levels of dynamic workflows—data flow, mining flow, and knowledge flow—corresponding to the Grid with three layers—data grid, mining grid, and knowledge grid—in order to manage data mining agents for multi-aspect analysis in distributed, multiple data sources. The workflows are also useful for dynamically organizing status-based business processes, using ontologies to describe and integrate multi-data-source and grid-based data mining agents in data mining process planning [11]. They must also provide the following:

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

The figure here outlines an e-business portal's use of an agent-based multi-database mining grid architecture on the Wisdom Web. The system's architecture includes two main components: the multi-layer grid and the Wisdom Web. The multi-agent-based data mining grid architecture requires at least four types of meta agents:

*Assistant.* To help e-business users perform various work activities (such as browsing and sampling data and planning data mining process) and analyze/refine models on the Wisdom Web;
*Interacting.* To help users in their cooperative work activities (such as communication, negotiation, coordination, and mediation);
*Mobile.* To help move to global data grid services and execute within data grids; and
*System.* To administer multiple data mining agents to register and manage many components, monitor events and the status of the workspaces and agent meeting places, and collect relevant measurement following predefined metrics.

In such a multi-agent architecture, agents are cre-

ated by and perform on behalf of users or other agents. They aim to achieve modest goals, reflecting the characteristics of autonomy, interaction, reactivity to environment, and proactive functions. The main components of the architecture in the figure interoperate in the following ways:

*Establish workspaces.* In large data mining processes, groups of people work as teams. Individual e-businesspeople have their own private workspace,

> WI has great potential as a key iIT in intelligent enterprise portals for e-business intelligence, enabling organizations to create a virtual enterprise where production steps are outsourced to multiple partners.

while the group has a shared workspace. The people (or their agents) manage, control, and schedule work, accessing global databases and distributing data analysis tasks to the mining grid based on some resource allocation policy;
*Create agents.* Data mining agents support the e-business tasks and the data mining process; interacting agents support communications; mobile agents support data mining tasks within data grids; and system agents by default support the components in the architecture;
*AgentMeetingPlaces.* AMPs support communication among agent groups; some system agents (such as

creation, deletion, and bookkeeping) are created by default to manage the AMPs;

*Repositories.* Repositories are global, local to one person, to a group of people, or distributed; local databases and model repositories are accessed from associated workspaces; global data grids are accessed only from controlling workspaces; and mobile agents travel to global data grids and execute there; and

*Process models.* Existing data mining process models are allowed into a workspace; for example, the planning and replanning techniques described in [11] can be applied.
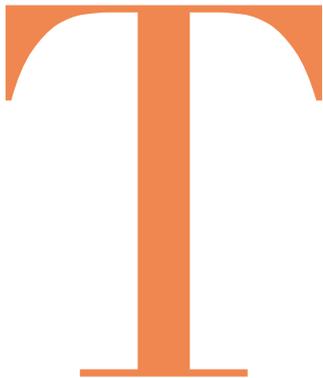
The idea of e-business intelligence illustrates why it is so important for developers to study and use WI technologies systematically to deal with the scalability and complexity of real-world problems. Using WI technologies to intelligently manage, analyze, and use information from distributed data sources is a problem not only in e-business but in e-science, e-learning, e-government, and all WI systems and services. Developing enterprise portals and e-business intelligence is a good example of how software engineers might try to deliver such functions.

Developing the Wisdom Web represents a tangible goal for WI research [5, 10]. The paradigm of Wisdom Web-based computing aims to provide not only a medium for seamless information exchange and knowledge sharing but a type of artificial resource for sustainable knowledge creation, and scientific and social evolution. The Wisdom Web is likely to rely on grid-like service agencies that self-organize, learn, and evolve their actions in order to perform service tasks, as well as maintain their identities and relationships in communities. They will also cooperate and compete among themselves in order to optimize their own, as well as others', resources and utilities.

A notable research challenge in Wisdom Web-based computing is how to develop and demonstrate a ubiquitous agent community, that is, an intelligent infrastructure that enables agents to look ahead, then plan and deliver what users want [5]. It works like personal agency; for instance, it can help a user manage tedious daily routine activities (such as processing email, placing orders, organizing meetings, and downloading news).

## CONCLUSION

iIT represents a paradigm shift in information processing, driven by WI, the Wisdom Web, grid computing, intelligent agents, autonomy-oriented computing, and other technological forces. WI is one of the most important of these forces, as well as a fast-growing iIT research field in its own right. iIT research could yield the new tools and infrastructure components necessary for creating intelligent portals throughout the Web. **C**

### REFERENCES
1. Alesso, H. and Smith, C. *The Intelligent Wireless Web.* Addison-Wesley, Reading, MA, 2002.
2. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., Eds. *Advances in Knowledge Discovery and Data Mining.* MIT Press, Cambridge, MA, 1996.
3. Foster, I. and Kesselman, C. *The Grid 2: Blueprint for a New Computing Infrastructure.* Morgan Kaufmann, San Francisco, 2004.
4. Hackathorn, R. *Web Farming for the Data Warehouse.* Morgan Kaufmann, San Francisco, 2000.
5. Liu, J. Web intelligence: What makes Wisdom Web? In *Proceedings of the 18th International Joint Conference on Artificial Intelligence* (Acapulco, Mexico, Aug. 9–15). Morgan Kaufmann, San Francisco, 2003, 1596–1601.
6. Ohsuga, S. Framework of knowledge-based systems: Multiple meta-level architecture for representing problems and problem-solving processes. *Knowledge Based Systems 3,* 4 (1990), 204–214.
7. Raghavan, P. Social networks: From the Web to the enterprise. *IEEE Internet Computing 6,* 1 (2002), 91–94.
8. Yao, Y., Zhong, N., Huang, J., Ou, C., and Liu, C. Using market value functions for targeted marketing data mining. *International Journal of Pattern Recognition and Artificial Intelligence 16,* 8 (2002), 1117–1131.
9. Zhong, N., Liu, J., and Yao, Y., Eds. *Web Intelligence.* Springer, 2003.
10. Zhong, N., Liu, J., and Yao, Y. In search of the Wisdom Web. *IEEE Computer 35,* 11 (2002), 27–31.
11. Zhong, N., Liu, C., and Ohsuga, S. Dynamically organizing KDD processes. *International Journal of Pattern Recognition and Artificial Intelligence 15,* 3 (2001), 451–473.
12. Zhong, N., Yao, Y., Liu, J., and Ohsuga, S., Eds. *Web Intelligence: Research and Development, LNAI 2198.* Springer, 2001.

**NING ZHONG** (zhong@maebashi-it.ac.jp) is head of the Knowledge Information Systems Laboratory and a professor in the Department of Systems and Information Engineering, Graduate School, Maebashi Institute of Technology, Japan. He is also an adjunct professor in the International WIC Institute at Beijing University of Technology, Beijing, China.

**JIMING LIU** (jiming@comp.hkbu.edu.hk) is a professor and head of the Department of Computer Science at Hong Kong Baptist University, Hong Kong, China. He is also an adjunct professor in the International WIC Institute at Beijing University of Technology, Beijing, China.

**YIYU YAO** (yyao@cs.uregina.ca) is a professor in the Department of Computer Science at the University of Regina, Regina, Saskatchewan, Canada. He is also an adjunct professor in the International WIC Institute at Beijing University of Technology, Beijing, China.