In Search of the Wisdom Web

Web intelligence will produce the new tools and infrastructure components necessary to create a Web that serves its users wisely.

Ning Zhong
Maebashi Institute of Technology

Jiming Liu
Hong Kong Baptist University

Yiyu Yao
University of Regina

Web intelligence offers a new direction for scientific research and development, pushing technology toward manipulating the meaning of data and creating a distributed intelligence that can actually get things done. WI explores the fundamental and practical impact that artificial intelligence and advanced information technology will have on the next generation of Web-empowered systems, services, and environments.

In an era dominated by the World Wide Web, Grid computing, intelligent-agent technology, and ubiquitous social computing, WI represents information technology's next challenge.

THE INTELLIGENT WEB

The Web significantly affects both academic research and everyday life, revolutionizing how we gather, store, process, present, share, and use information. It offers great opportunities and challenges in many areas, including business, commerce, marketing, finance, publishing, education, and research and development.

Despite current technological advances, we still cannot predict what the Web's next paradigm shift will be. However, we propose that this change will transform the Web into an *intelligent* entity—hence, the term *Web intelligence*.[1-3]

The next-generation Web will go beyond improved information search and knowledge queries and will help people achieve better ways of living, working, playing, and learning. To fulfill its potential, the intelligent Web's design and development must incorporate and integrate several fundamental capabilities.

Autonomic Web support

The intelligent Web functions essentially as an autonomic entity. The Web automatically regulates the functions and cooperation of related Web sites and available application services.

Making the Web autonomic requires that we meet several challenges:

- *Reflexive server propagation.* An intelligent Web server must be able to automatically self-delegate its functional roles to other services, along with its corresponding spatial or temporal constraints and operational settings.
- *Specialization.* An intelligent Web server must itself be an agent specialized in performing some roles in a certain service. The association of its roles with any service will be measured and updated dynamically.
- *Growth.* The intelligent Web agent population will change dynamically as agents self-repro-

0018-9162/02/$17.00 © 2002 IEEE

We expect the Web to extend both the knowledge and the intelligence of artificial assistants.

duce to become more specialized or as agents deactivate.

- *Autocatalysis.* Catalysis is a reaction between two or more entities precipitated by a separate agent. As search requests activate intelligent Web agents, they will evolve through specialization to fill various roles, generating associations with some services and among themselves that will aggregate autocatalytically.

Representation

Intelligent Web agents can use the Problem Solver Markup Language (PSML) to specify their roles, settings, and relationships with any other services. The intelligent Web must also have the ability to process and understand natural language. It must understand and correctly judge the meaning of concepts expressed in words, such as "good," "best," and "season." Further, the intelligent Web must grasp the granularities of these terms' corresponding subjects and the location of their ontology definitions.

Self-direction and learning

In addition to the semantic knowledge that an intelligent search can extract and manipulate, intelligent Web agents must also incorporate a dynamically created source of *metaknowledge* that deals with the relationships between concepts and the spatial or temporal constraint knowledge that planning and executing services use. This allows the agents to self-resolve their conflicts.

To solve specific problems, intelligent Web agents must be able to *plan*. The planning process uses goals and associated subgoals, as well as constraints. In the intelligent Web, ontologies alone will not be sufficient.

Personalization

The intelligent Web can personalize interactions by remembering a particular user's recent encounters and relating the topics and sites that a user accesses during different online sessions. It may further identify other goals and courses of action as a user's interactions broaden and deepen, providing ever more data upon which to base its recommendations. As part of its personalized approach to user services, the intelligent Web will interact with the user when executing these tasks.

In summary, semantics contributes a vital aspect to the intelligent Web. We expect the Web to extend not only the *knowledge* of artificial assistants, but also their *intelligence*.

WI'S FOUR LEVELS

We can study Web intelligence on at least four conceptual levels, ranging from the lower, hardware-centered level to the higher, application-centered level. This framework builds upon the fast development and application of various Web technologies.

- *Internet-level communication, infrastructure, and security protocols.* At its core, the Web is a computer-network system. WI techniques for this level include Web data prefetching systems built upon Web surfing patterns to resolve latency issues. The intelligence of the Web's prefetching routines comes from an adaptive learning process based on observations of user surfing behavior.
- *Interface-level multimedia presentation standards.* The Web functions as an interface for human-Internet interaction. At this level, the Web interfaces require adaptive cross-language-processing, personalized-multimedia-representation, and multimodal-data-processing capabilities.
- *Knowledge-level information processing and management tools.* The Web serves as a distributed data and knowledge base. Accessing and manipulating this information requires semantic markup languages to represent the Web's contents in machine-understandable formats. Agent-based autonomic computing functions such as searching, aggregation, classification, filtering, managing, mining, and discovery can then use this data.[4]
- *Application-level ubiquitous computing and social intelligence environments.* The Web can form the basis for establishing social networks that contain communities of people, organizations, or other social entities. Social relationships—such as friendship, coworking, or exchanging information about common interests—connect these entities. The study of WI thus encompasses issues central to social network intelligence.

Users access the Web's multimedia content from stationary desktop computers and increasingly from mobile platforms as well.[5] Ubiquitous Web access and computing from various wireless devices requires even greater adaptive personalization. WI should suit these needs well by providing techniques for use in constructing interest models derived from implicit inferences based on user behavior.[6]

SUPPORTING SOCIAL INTELLIGENCE NEEDS

WI researchers must study both centralized and distributed information structures. The Web's information and knowledge can be organized in two ways:

- globally distributed within multilayers over the Web's protocol infrastructure, or
- locally centralized on an intelligent portal that provides Web services joined to its own cluster of specialized intelligent applications.

Each approach, however, has a serious flaw: The global semantic Web approach faces combinatory complexity limitations, while the intelligent-portal approach limits uniformity and access.[6]

One way to solve this problem is to use PSML for collecting globally distributed content and knowledge from Web-supported social networks. The collected data can then be incorporated with locally operational knowledge-databases to provide a local enterprise or community with centralized, adaptable, Web-intelligent services.

Social networks create a self-organizing structure of users, information, and expert communities.[7] Such networks can play a crucial role in combining next-generation enterprise portals and Web search engines with functions such as data mining and knowledge management to discover, analyze, and manage social-network knowledge.

The social network resides atop the four-level WI framework and benefits from support functions drawn from all levels of Web intelligence, including security, prefetching, adaptive cross-language processing, personalized multimedia representation, semantic searching, aggregation, classification, filtering, managing, mining, and discovery.

WI RESEARCH AGENDA

WI research incorporates knowledge from existing disciplines, such as artificial intelligence and information technology, in a totally new domain. At the same time, WI research also enriches these established disciplines as it introduces new topics and challenges. Figure 1 shows a schematic diagram of interrelated research topics from a Web-based, business-centric perspective.

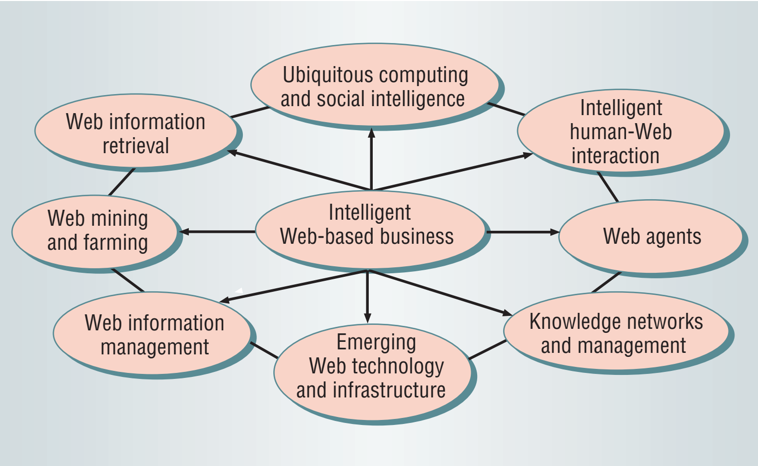

Figure 1. Topics related to Web intelligence. Diagram shows interrelated research topics from an intelligent Web-based, business perspective.

Researchers have achieved much in the field of Web intelligence, but several important issues provide the basis for the ongoing WI research agenda.

Web mining

With the rapid growth of information stored on the Web, researchers can use data mining to explore many different resources. Web mining applies data mining techniques to large Web data repositories such as documents, logs, and services.

Web mining research thus straddles the crossroads of research undertaken by the database, information retrieval, machine learning, and natural-language-processing communities, among others. We expect Web mining to play an increasingly important role in future intelligent Web information services.

Automatic ontology construction

Ontology is a key enabling technology for developing the semantic Web. It plays a crucial role in Web-based knowledge processing and management, such as Web community communication, semantic-based agent communication, knowledge-based Web information retrieval, Web content understanding, and Web community discovery.

Comprehensive ontology construction and learning, an active field of study for at least a decade, still requires more automatic methods. The tedious, cumbersome task of building an ontology manually remains a bottleneck that severely slows the development of semantics-based Web intelligence.

PSML and Web inference engine

Distributed inference engines form PSML's core. These engines can perform automatic reasoning on the Web by incorporating autonomically collected and transformed content and metaknowledge into locally operational knowledge and databases. A feasible way to implement PSML is to use an existing Prolog-like logic language supplemented with agents that perform dynamic-content updates, metaknowledge collection, and transformation tasks.

The intelligent Web seeks to provide resources that encourage sustainable knowledge creation and scientific and social evolution.

Social network intelligence

The social intelligence approach to Web computing presents new opportunities for WI research and development. As the Web becomes an integral part of our society, WI can and should support Web-based social networks at all levels. Study in this area must receive as much attention as Web mining, Web agents, ontologies, and related topics.

Web-based computing

The intelligent Web seeks to provide not only a medium for seamless information exchange and knowledge sharing, but also the sort of human-crafted resources that encourage sustainable knowledge creation and scientific and social evolution. The intelligent Web will rely on Grid-like service agencies that self-organize, learn, and evolve their courses of action to perform service tasks and transform their identities and interrelationships in communities. These services will also cooperate and compete among themselves to optimize their resources and utilities and those of others.

Benchmark applications

To effectively develop and evaluate systems and applications that address WI research issues, we must consider benchmark applications that will demonstrate these capabilities.

Suppose we want to conduct a Web-based search to compile the data and generate a market report for an existing product or a potential new product. To perform these tasks, an information agent will mine and integrate available Web information, which will in turn be passed to a market analysis agent.

The analysis will involve the quantitative simulation of customer behavior in a marketplace, instantaneously handled by other service agencies involving a large number of Grid agents. Given that the number of variables can number in the hundreds or thousands, generating one prediction can easily require significant computer resources.

ABOUT THIS ISSUE

The nine articles in this issue represent the current state of research and development in WI-related areas, as well as WI's theoretical and application aspects.

Social networks

The first three articles address social network intelligence. To provide a better understanding of the sociology of Web content creation, Ravi Kumar and colleagues present a study of the Web as a network of social networks. In "The Web and Social Networks," they observe that the Web's latent structure will be important for future algorithms such as higher-precision search engines and more effective Web mining.

Toyoaki Nishida describes a social intelligence design for WI. "Social Intelligence Design for the Web," Nishida's survey of major issues affecting social intelligence design, demonstrates why we must successfully embed WI technologies in human society and amalgamate them with everyday life.

Toru Ishida presents a new direction in the development of social intelligence: agent technology. In "*Q*: A Scenario Description Language for Interactive Agents," Ishida details the specifications for *Q*, a language for developing interaction scenarios between agents and users to ensure the social acceptance of interactive agents in Web applications.

Knowledge networking and management

The next three articles deal with intelligent knowledge networking and management. "Toward Behavioral Intelligence in the Semantic Web" by Joanna J. Bryson and colleagues examines the particulars of semantic Web intelligence and agent intelligence. The authors propose that regarding the semantic Web's content as *behavioral* intelligence can provide a new perspective on the future potential for this storehouse of information.

In "Ontology-Based Knowledge Management," Dieter Fensel and colleagues describe the On-To-Knowledge Project in which a consortium of European-based commercial and academic partners are building an ontology-based knowledge management tool suite. These tools will help developers process the many heterogeneous, distributed, and semistructured documents typically found in large-company intranets and on the Internet.

Michele Missikoff and colleagues present a learning approach that partially automates ontology engineering. In "Integrated Approach to Web Ontology Learning and Engineering," they describe how using this technique can make automatic ontology learning and construction a timely and attractive alternative compared to the tedious process of building an ontology manually.

Web intelligence techniques

The final three articles describe techniques for improving the intelligent Web. In "Data Mining for Web Intelligence," Jiawei Han and Kevin C-C.

Chang explore the concept of Web data mining, identifying some challenging issues and discussing how we can use data mining techniques to best achieve Web intelligence.

"From Computational Intelligence to Web Intelligence" by Nick Cercone and colleagues describes existing computational intelligence solutions, such as natural language interfaces and machine translation techniques, and explores how their application might affect WI.

Finally, in this month's Web Technologies column, pp. 107-108, "Using Knowledge Anchors to Reduce Cognitive Overheads" by Stephen Ransom and Xindong Wu presents an effective approach that specifically addresses the algorithmic-complexity aspects of WI and digital libraries.

We hope this special issue will serve as a useful introduction to some of the fundamental issues that WI researchers are facing as they pursue their efforts in the design and implementation of the intelligent Web. ■

Acknowledgment

We wish to express our gratitude to the authors of all the submitted papers for their contributions.

References

1. J. Liu et al., "The Wisdom Web: New Challenges for Web Intelligence (WI)," to be published in *J. Intelligent Information Sys*, vol. 20, no. 1, 2003.
2. Y. Yao et al., "Web Intelligence (WI): Research Challenges and Trends in the New Information Age," *Web Intelligence: Research and Development,* Lecture Notes in Artificial Intelligence 2198 (LNAI 2198), N. Zhong et al., eds., Springer-Verlag, Heidelberg, 2001, pp. 1-17.
3. N. Zhong et al., "Web Intelligence (WI)," *Proc. 24th Int'l Computer Software and Applications Conf.* (COMPSAC 2000), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 469-470.
4. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, no. 5, 2001, pp. 34-43.
5. M. Weiser, "The Future of Ubiquitous Computing on Campus," *Comm. ACM,* vol. 41, no. 1, 2001, pp. 41-42.
6. H.P. Alesso and C.F. Smith, *The Intelligent Wireless Web*, Addison-Wesley, Reading, Mass., 2002.
7. P. Raghavan, "Social Networks: From the Web to the Enterprise," *IEEE Internet Computing*, vol. 6, no. 1, 2002, pp. 91-94.

Ning Zhong *is a professor of computer science in the Department of Information Engineering, Maebashi Institute of Technology, Maebashi City, Japan. His research interests include Web intelligence, data mining, and knowledge discovery. He received a PhD in the Interdisciplinary Course on Advanced Science and Technology from the University of Tokyo. He is a member of the IEEE, the ACM, and the AAAI. Contact him at zhong@maebashi-it.ac.jp.*

Jiming Liu *is an associate professor in the Department of Computer Science, Hong Kong Baptist University. His research interests include autonomy-oriented computing, Web intelligence, and multiagent systems. Liu received a PhD in electrical engineering from McGill University, Canada. He is a member of the ACM and a senior member of the IEEE. Contact him at jiming@comp.hkbu.edu.hk.*

Yiyu Yao *is a professor of computer science in the Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada. His research interests include information retrieval, Web intelligence, and granular computing. He received a PhD in computer science from the University of Regina. He is a member of the IEEE and the ACM. Contact him at yyao@cs.uregina.ca.*