Towards Web Intelligence (WI)

(AWIC’03 Keynote Talk)

Ning Zhong
Maebashi Institute of Technology

http://www.kis-lab.com/zhong
zhong@maebashi-it.ac.jp
1. **Why** is WI Research Significant?
2. **What** is the Scope of WI Research?
3. Ultimate **Goal** of WI Research
4. **Four Levels** of WI Support
5. **An Agenda** for WI Research
1. Why is WI Research Significant?
2. What is the Scope of WI Research?
3. Ultimate Goal of WI Research
4. Four Levels of WI Support
5. An Agenda for WI Research
Moore’s Law vs. Human Capacity

- Processor Speed (MHz)
- Wireless data (x Kbps)
- Internet Hosts (x K)
- Web pages/images (x K)

Individual Capacity

Year
Why WI?

- Data production on the Web is at an exponential growth rate.

- A fast growing interest in WI
  
  Statistics on Searching Web Intelligence on the Web (February 2001 & 2003)
  
  Industrial interest in WI (Google), the majority of the top 40 pages are industry related.

- Only a few academic papers

- ResearchIndex database

- We need to narrow the gap between industry needs and academic research.
### Web Intelligence on the Web

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>Number of Hits (February 2001)</th>
<th>Number of Hits (February 2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lycos</td>
<td>1,102,279</td>
<td>7,163,922</td>
</tr>
<tr>
<td>Google</td>
<td>1,080,000</td>
<td>2,590,000</td>
</tr>
<tr>
<td>AltaVista</td>
<td>1,271</td>
<td>1,860,062</td>
</tr>
<tr>
<td>Netscape</td>
<td>77</td>
<td>1,900,000</td>
</tr>
<tr>
<td>Yahoo</td>
<td>74</td>
<td>2,510,000</td>
</tr>
</tbody>
</table>
Why WI?

- Data production on the Web is at an exponential growth rate.
- A fast growing interest in WI
- Statistics on Searching Web Intelligence on the Web (February 2001 & 2003)
  **Industrial interest in WI** (Google), the majority of the top 40 pages are *industry related*.
- Only a few academic papers
- We need to narrow the gap between *industry* needs and academic research.
Industrial Interests in WI

- **Web Intelligence**: kis-lab.com/wi01/
- **Web-Intelligence Home Page**: www.web-intelligence.com/
- **Intelligence on the Web**: www.fas.org/irp/intelwww.html
- **WIN: home WEB INTELLIGENCE NETWORK**, smarter.net/
- **CatchTheWeb - Web Research, Web Intelligence Collaboration**: www.catchtheweb.com/
- **Infonoia: Web Intelligence In Your Hands**: www.infonoia.com/myagent/en/baseframe.html
Web Intelligence
Turning search data into Net Equity

Web Intelligence, the collection and analysis of data about your web site, your current and prospective customers, competitors and partners, is a must for any search engine optimisation campaign. Our Web Intelligence service (which is available either as an integral part of a programme or as a separate service) gives you a wealth of fascinating data you can put right now.

Web Intelligence starts with search term analysis. Firstly, search engines provide search term data on what web users are asking for, and the relative volume and ranking of particular search terms within their search engines. We include below for further information:

Search term analysis

Search

Categories
Most Popular

Advertise | About Us | Log-In | Contact Us | Privacy

Telania, Inc. © 2003 All Rights Reserved Worldwide

Make Boldsearch Your Homepage
Why WI?

- Data production on the Web is at an exponential growth rate.
- A fast growing interest in WI
- Statistics on Searching Web Intelligence on the Web (February 2001 & 2003)
- Industrial interest in WI (Google), the majority of the top 40 pages are industry related.
- Only a few academic papers
- ResearchIndex database
- We need to narrow the gap between industry needs and academic research.
Observations

- To accurately quantify success, websites must examine visitor profiles and online behavior, instead of the number of page views.

- To support business goals, companies need a robust Web intelligence model capable of accommodating new strategies and technologies. The cornerstone of Web intelligence is the ability to uniquely identify and segment users.

- Web intelligence must form the foundation for personalization and one-to-one business.
Perspectives of WI

- **WI** may be viewed as applying results from existing disciplines to a totally **new domain**.

- **WI** introduces **new problems** and **challenges** to the established disciplines.

- **WI** may be viewed as an **enhancement** or an **extension** of **IT** and **AI**.
1. Why is WI Research Significant?
2. **What is the Scope of WI Research?**
3. Ultimate Goal of WI Research
4. Four Levels of WI Support
5. An Agenda for WI Research
What is WI?

Web Intelligence (WI) exploits the fundamental and practical impact that advanced Information Technology (IT) and innovative Artificial Intelligence (AI) will have on the next generation Web:

- Integration of IT with AI/SC
- Applications of AI on the Web
- Intelligent systems, services, and environments on the Web

WI = IT + AI
The Evolving Web

Intelligent Web

Foundation of the Current Web

Resource Description Framework
HyperText Markup Language
HyperText Transfer Protocol

Proof, Logic and Ontology Languages

Self-Describing Documents

Shared Terms/Terminology
Machine-Machine Communication

1990
2000
2010

Based on E.A. Feigenbaum, J. Hendler; WI-2001
WI Systems (WIS)

- Ubiquitous Computing
- Multi-modal Interaction
- Web Information Retrieval
- Web Mining and Farming
- Social Network
- Grid Computing
- Web Agents
- Knowledge Network and Management
- WIS Applications
WIS can be classified into four categories (based on Russell & Norvig’s scheme):

1. System that **thinks** like humans
2. System that **acts** like humans
3. System that **thinks** rationally
4. System that **acts** rationally
Basic System

Enterprise Information

Register, update

Automatic HP generation, modification

Search engine

VIP Portal

DBMS

Registered data
Advanced Questions

- How the customer or prospect enters our VIP portal in order to target products and manage promotions and marketing campaigns?
- To the already demanding requirement to capture transaction data for further analysis, we now also need to use the Web 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?
......
Modeling User Groups

- **Recurrent user** is familiar with the Web structure, and can find the useful information right away.
- **Rational user** is new to the website, and knows clearly what he/she wants and selects a direction based on the information of hyperlinks.
- **Random user** has no strong intention to get something, and just wanders among pages.
Personalization: A Prime Factor 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.
Advanced WIS

Enterprise Information

Web mining
- Web logs
- User Profiles
- Email filtering, managing, auto-reply
- VIP Portal
- Enterprise Information

Register, update
- Automatic HP generation, modification
- Targeted Marketing
- Security Solution
- Search Engine 1
- Search Engine 2

DBMS
- Registered data
- Ontologies
E-BI needs Web based targeted marketing, which is integrated with other functions of WI such as Web mining, personalized recommendation, and e-mail filtering.

Multiple data sources that are obtained from multiple customer touch points, 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-faced view of their customers.

Extending Web mining to Web farming that is treated more like a large agricultural business including planting and harvest.
Presentation Outline

1. Why is WI Research Significant?
2. What is the Scope of WI Research?
3. Ultimate Goal of WI Research
4. Four Levels of WI Support
5. An Agenda for WI Research
Goal of WI Research

- We argue that the next paradigm shift in the Web is towards wisdom.
- Developing the Wisdom Web will become a tangible goal for WI research.
- The new generation of the WWW will enable humans to gain wisdom of living, working, playing, and learning, in addition to information search and knowledge queries.
**Fundamental Capabilities of the Wisdom Web**

1. The **autonomic Web support**: Self-regulating the functions and cooperation of available application services.

**Research challenges:**

1. **Reflexive server propagation**: How to automatically self-delegate its functional roles to other services, along with its spatial or temporal constraints and operational settings.

2. **Specialization**: How to become specialized in performing some roles in a certain service.

3. **Growth**: How to self-reproduce.

4. **Autocatalysis**: How to evolve through specialization to fill various roles, generating associations with some services and among themselves that will aggregate autocatalytically.
2. **Problem Solver Markup Language (PSML):** Specifying roles, settings, and relationships with any other services.

3. **Semantics:** Providing right judgment of concepts, such as “best” and “season” and granularities of their corresponding subjects.

4. **Meta-knowledge:** Dealing with spatial or temporal constraints/conflicts in planning and executing services.
5. Planning: Planning with goals and associated sub-goals.

In the Wisdom Web, ontology alone will not be sufficient.

6. Personalization: Understanding recent encounters and relating different episodes together.

7. A sense of humor: Interacting with a user on a personal level.
Presentation Outline

1. Why is WI Research Significant?
2. What is the Scope of WI Research?
3. Ultimate Goal of WI Research
4. Four Levels of WI Support
5. An Agenda for WI Research
Four Levels of WI Support

**Level-1**: Internet-level communication, infrastructure, and security protocols

**Level-2**: Interface-level multi-media presentation standards

**Level-3**: Knowledge-level information processing and management tools

**Level-4**: Application-level ubiquitous computing and social intelligence utilities
The study of WI concerns the important issues central to social network intelligence.

Furthermore, the multimedia contents on the Web are not only accessible from stationary, but increasingly mobile platforms.

Ubiquitous Web access and computing from various wireless devices needs adaptive personalization.
1. Why is WI Research Significant?
2. What is the Scope of WI Research?
3. Ultimate Goal of WI Research
4. Four Levels of WI Support
5. An Agenda for WI Research
An Agenda

- Semantic Web mining and automatic construction of ontologies
- Social network intelligence
- PSML and Web inference engine
- Wisdom Web based computing
An Agenda

- Semantic Web mining and automatic construction of ontologies
- Social network intelligence
- PSML and Web inference engine
- Wisdom Web based computing
The Semantic Web

The Semantic Web is based on languages that make more of the semantic content of the page available in machine-readable formats for agent-based computing.

A “semantic” language that ties the information on a page to machine readable semantics (ontology).
Components of Semantic Web Techniques

- A unifying data model such as RDF.

- Languages with defined semantics, built on RDF, such as OWL (DAML+OIL).

- Ontologies of standardized terminology for marking up Web resources, used by semantically rich service-level descriptions, and tools that assist the generation and processing of semantic markup.
Ontology Markup Languages

Content Definition

Content Representation

Web

Ontology/Agent Markup

- SGML
- HTML
- XML
- SHOE
- RDF
- OIL
- OWL
- DAML
- DAML+OIL
Observation

- Ontologies and agent technology can play a crucial role in WI by enabling Web-based knowledge processing, sharing, and reuse between applications (corresponding to Level-3 of WI).
Categories of Ontologies

- A **domain-specific** ontology describes a well-defined technical or business domain.
- A **task** ontology might be either domain-specific or reconstructed from a set of domain-specific ontologies for meeting the requirement of a task.
- A **universal** ontology describes knowledge at higher levels of generality. It can be generated from several domain-specific/task ontologies, as a bridge for communication among domains/tasks/communities.
A Domain-Specific Ontology on Web Intelligence

- Web Information Retrieval
  - Web Human-Media Engineering
    - Art of Web Page Design
    - Multimedia Information Representation
    - Web-Based Human Computer Interface
  - Approximate Retrieval
  - Ontology-Based Information Retrieval
  - Automatic Web Content Cataloging and Indexing

- Web Information System Environment and Fundations
  - Competitive Dynamics of Web Sites
    - Emerging Web Technology
    - Network Community Formation and Support

- Web-Based Applications
  - Business Intelligence
    - Web-Based Decision Support Systems
    - Web Publishing

- Web Information Management
  - Data Quality Management
  - Internet and Web-Based Data Management
  - Web Knowledge Management

- Web Mining and Farming
  - Data Mining and Knowledge Discovery
    - Text Mining
    - Web-Based Ontology Engineering
  - Dynamics of Information Sources
    - E-mail Filtering
    - Web-Based Cooperative Problem Solving

- Web Agents
A Domain-Specific Ontology on IAT
A Task Ontology Generated from the WI and IAT Ontologies
Automatic Construction and management of Domain-Specific Ontologies

A Challenging Topic in WI
An Agenda

- Web mining and automatic construction of ontologies
- Social network intelligence
- PSML and Web inference engine
- Wisdom Web based computing
The Web as a Graph

We can view the Web as a directed network that connects **people** (organizations or social entities).

Research Questions:
- How big is the graph? (outdegree and indegree)
- Can we browse from any page to any other? (clicks)
- Can we exploit the structure of the Web? (searching and mining)
- How to discover and manage the Web communities?
- What does the Web graph reveal about social dynamics?
- How different is browsing from a “random walk”?
The Web as a Graph (2)

- Exploit structures of the Web for
  - Crawl strategies
  - Search
  - Mining communities

- Classification/organization

- Web anthropology
  - Prediction, discovery of structures
  - Sociological understanding
  - The Web world is now becoming an integral part of our society.
What/Why Social Network?

- A Social Network comprises a set of people with a pattern of interactions among them.

- The broader Social Network includes not only people but also information — plays a crucial role in knowledge management.

The Web supported Social Network is a self-organizing structure of users, information, and communities of expertise.
Social Network theory is now significantly influencing search engine and portal development for the Web and the enterprise.
An Agenda

- Web mining and automatic construction of ontologies
- Social network intelligence
- PSML and Web inference engine
- Wisdom Web based computing
Web information/knowledge could be:

- either **globally, distributed** throughout the Web,
- or **locally, centralized** on an intelligent portal providing Web services (i.e. the intelligent service provider) that is joined to its own cluster of specialized intelligent applications.
How to Integrate Global and Local Information?

- To develop and use PSML (Problem Solver Markup Language), for
  - collecting *globally* distributed content and knowledge from Web-supported social networks, and
  - incorporating it with *locally* operational knowledge-data bases in an enterprise or community for local centralized, adaptable Web intelligent services or decision-making.
Automatic Transformation

Global info-sources

Social Network Mining

Local Info-sources

Content Mining

Inference engine

Prolog-like format

KB

DB

PSML

DAML+OIL

meta
An Agenda

- Web mining and automatic construction of ontologies
- Social network intelligence
- PSML and Web inference engine
- Wisdom Web based computing
Wisdom Web Based Computing

To provide
- not only a medium for information/knowledge exchange/sharing
- but also a type of man-made resources for sustainable knowledge creation and scientific and social evolution.

The Wisdom Web will reply on grid-like agencies:
- self-organize, learn, and evolve their courses of actions in order to perform service tasks as well as their identities and interrelationships in communities, and
- cooperate and compete among themselves in order to optimize their as well as others’ resources and utilities.
Creating Data Mining Grids

- A new platform as the **middleware** is required to deal with multiple very-large data sources for multi-aspect analysis in portals for business intelligence.
- Creating a grid-based, organized society of data (Web) mining agents, called **Data Mining Grid**:
  - To develop various data mining **agents** for different targeted marketing tasks;
  - To organize the agents into a **grid** with multi-layer under the Web as a **middleware**;
  - To use the grid for multi-aspect analysis in distributed, multiple data sources;
  - To manage the grid by a multi-level control authority.
To Learn More about WI ...

Books

The IEEE/WIC International Conference on Web Intelligence

IEEE CS TCCI

Journals

wi-consortium.org
The WIC (http://wi-consortium.org) is an international non-profit organization dedicated to promoting world-wide scientific research and industrial development in the era of Web and agent intelligence.

The WIC specializes in the development and promotion of new WI-related research and technologies through collaborations with WI research centers throughout the world and organization/individual members, technology showcases at WI conferences and workshops, WIC official book and journal publications, the WIC newsletter, and WIC official releases of new industrial solutions and standards.
Acknowledgements

- Profs. Jiming Liu, Yiyu Yao, Edward A. Feigenbaum, Setsuo Ohsuga, Benjamin Wah, Philip Yu, Lotfi A. Zadeh, and Xindong Wu, etc.

- WIC Technical Committee

- WIC Research Centers in Australia, Beijing, Canada, India, Japan, and Spain, among others.
References & Further Reading

- N. Zhong, J. Liu, Y.Y. Yao (eds.): IEEE Computer Special Issue on Web Intelligence, 35 (11) (November 2002)
References & Further Reading (2)

Thank You !