

# 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](mailto:zhong@maebashi-it.ac.jp)



# 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

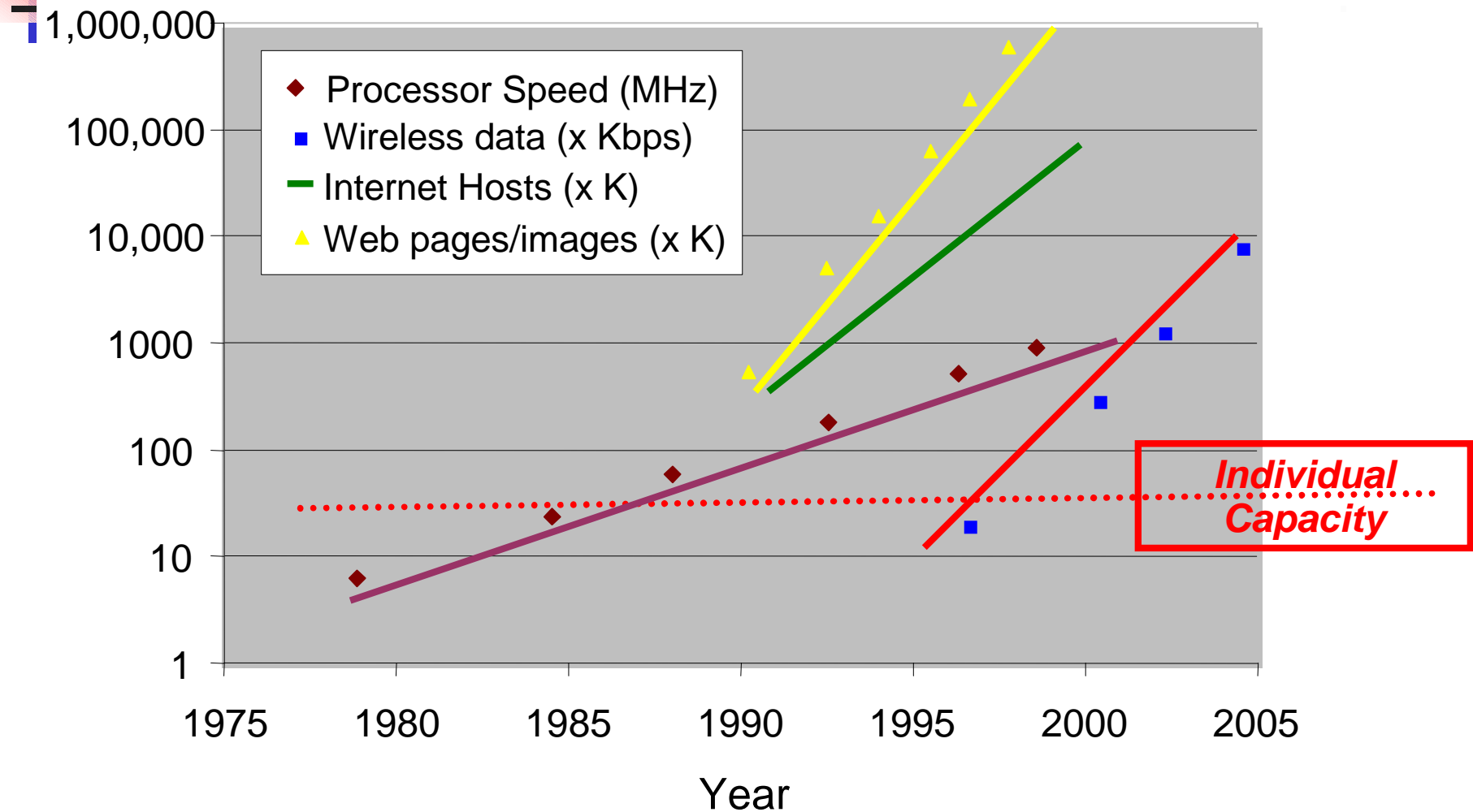


# Presentation Outline

---

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

# Moore's Law vs. Human Capacity





# Why WI?

---

- n Data production on the Web is at an exponential growth rate.
- n 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.
- n Only a few **academic** papers
  - ResearchIndex database
- n We need to narrow the gap between **industry** needs and **academic** research.



## *Web Intelligence* on the Web

---

Search Engine	Number of Hits (February 2001)	Number of Hits (February 2003)
Lycos	1,102,279	<b>7,163,922</b>
Google	1,080,000	<b>2,590,000</b>
AltaVista	1,271	<b>1,860,062</b>
Netscape	77	<b>1,900,000</b>
Yahoo	74	<b>2,510,000</b>



# Why WI?

---

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

- n Only a few **academic** papers

ResearchIndex database

- n We need to narrow the gap between **industry** needs and **academic** research.



# Industrial Interests in WI

---

- n [Web Intelligence](http://kis-lab.com/wi01/) [kis-lab.com/wi01/](http://kis-lab.com/wi01/)
- n [Web-Intelligence Home Page](http://www.web-intelligence.com/)  
[www.web-intelligence.com/](http://www.web-intelligence.com/)
- n [Intelligence on the Web](http://www.fas.org/irp/intelwww.html)  
[www.fas.org/irp/intelwww.html](http://www.fas.org/irp/intelwww.html)
- n [WIN: home](http://smarter.net/) **WEB INTELLIGENCE NETWORK,**  
[smarter.net/](http://smarter.net/)
- n [CatchTheWeb - Web Research, Web Intelligence Collaboration](http://www.catchtheweb.com/) [www.catchtheweb.com/](http://www.catchtheweb.com/)
- n [Infonoia: Web Intelligence In Your Hands](http://www.infonoia.com/myagent/en/baseframe.html)  
[www.infonoia.com/myagent/en/baseframe.html](http://www.infonoia.com/myagent/en/baseframe.html)



ImetriX<sup>®</sup>  
Your Web Intelligence<sup>®</sup>



imetrix.biz

 **fortelit**  
WEB INTELLIGENCE

*Where the future is always present*™

  
**BUSINESS OBJECTS**

 **weboptimiser**

**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 to use right now.

Web Intelligence starts with **search term analysis**. Firstly, search engines provide search term data to us what web users are asking for, and the respective volume and ranking of particular search terms which throw up some amazing insights.

[Continue below for further Web Intelligence](#)

**FREE INTRODUCTORY SERVICE** | **SERVICES PORTFOLIO** | **NEWS, VIEWS AND RESOURCES** | **ABOUT US** | **CONTACT US**

**WEB INTELLIGENCE** | **SEARCH OPTIMISATION** | **SUPPORT & MAINTENANCE**

**Search term analysis**  
**Content**

**WEB**  
intelligence

**Boldsearch**  
web intelligence

[Categories](#)  
[Most Popular](#)

Search

  
Web Intelligence Collaboration

[Advertise](#) | [About Us](#) | [Log-in](#) | [Contact Us](#) | [Privacy](#)

[Telania, Inc.](#) © 2003 All Rights Reserved Worldwide

[Make Boldsearch Your Homepage](#)



# Why WI?

---

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

- n Only a few **academic** papers

ResearchIndex database

- n We need to narrow the gap between **industry** needs and **academic** research.



# Observations

---

- n To accurately quantify success, websites must examine **visitor profiles** and **online behavior**, instead of the number of page views.
- n 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**.
- n *Web intelligence* must form the foundation for **personalization** and **one-to-one business**.



# Perspectives of WI

---

- n **WI** may be viewed as applying results from existing disciplines to a totally **new domain**.
- n **WI** introduces **new problems** and **challenges** to the established disciplines.
- n **WI** may be viewed as an **enhancement** or an **extension** of **IT** and **AI**.



# 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



# What is WI?

---

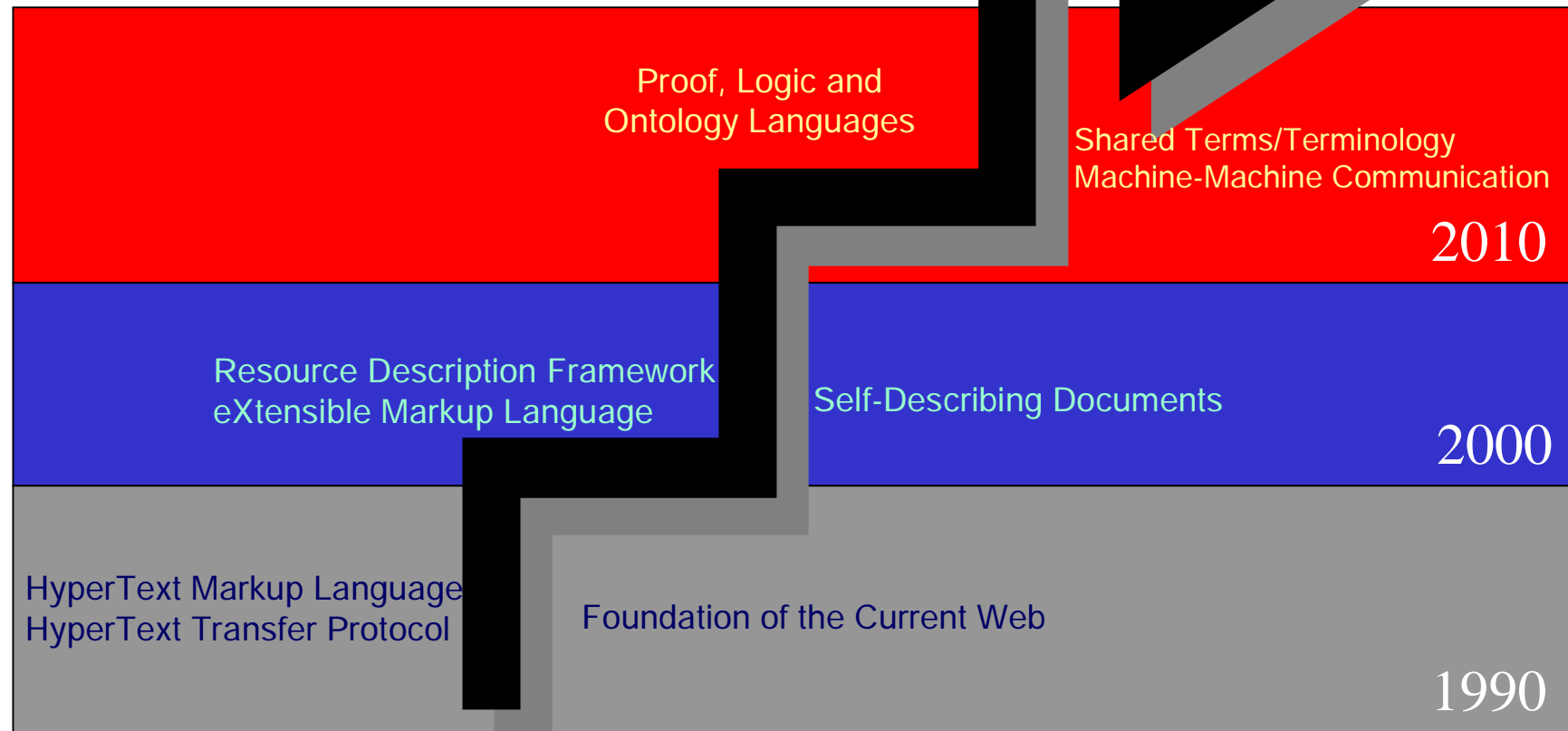
- 2 **Web Intelligence (WI)** exploits the fundamental and practical impact that *advanced Information Technology (IT)* and *innovative Artificial Intelligence (AI)* will have on the next generation Web:


$$WI = IT + AI$$

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

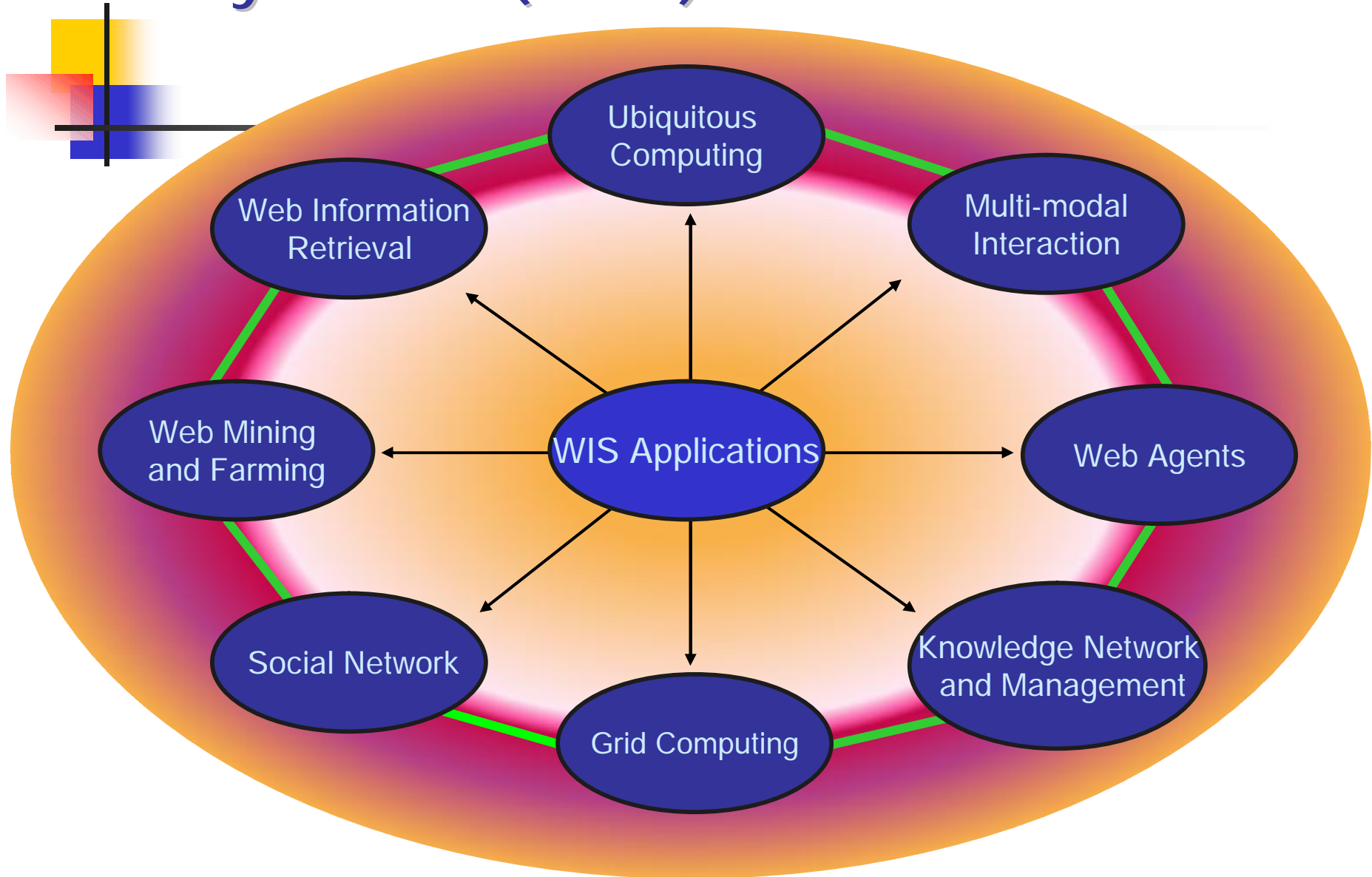
# The Evolving Web

*Intelligent Web*



Based on E.A. Feigenbaum, J. Hendler; WI-2001

# WI Systems (WIS)







# WI Systems (WIS)

---

- n WIS can be classified into four categories  
(based on Russell & Norvig's scheme)

designing philosophy of WIS

ability,  
functionality  
of WIS

System that <b>thinks</b> like <b>humans</b>	System that <b>thinks</b> <b>rationally</b>
System that <b>acts</b> like <b>humans</b>	System that <b>acts</b> <b>rationally</b>

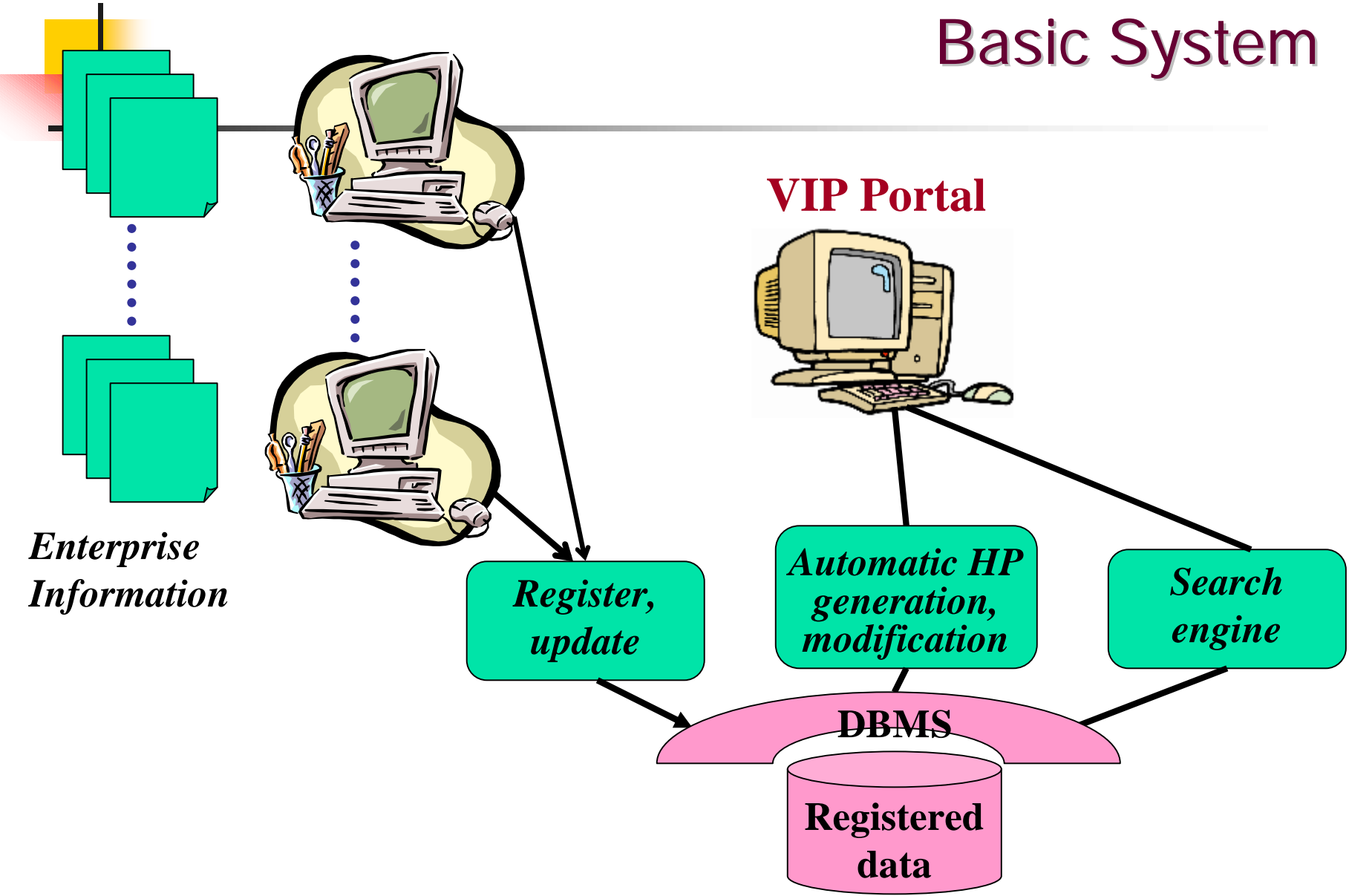


**An Example of WIS:**

---

**Virtual Industry Park (VIP)**

# Basic System





## Advanced Questions

---

- 2 How the customer or prospect enters our VIP portal in order to target products and manage promotions and marketing campaigns?
- 2 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.



## Advanced Questions (2)

---

- 2 What pages has he or she visited?
- 2 What is the semantic association between the pages he or she visited?
- 2 Is the visitor familiar with the Web structure?  
Or is he or she a new user or a random one?
- 2 Is the visitor a Web robot or other users?
- 2 .....

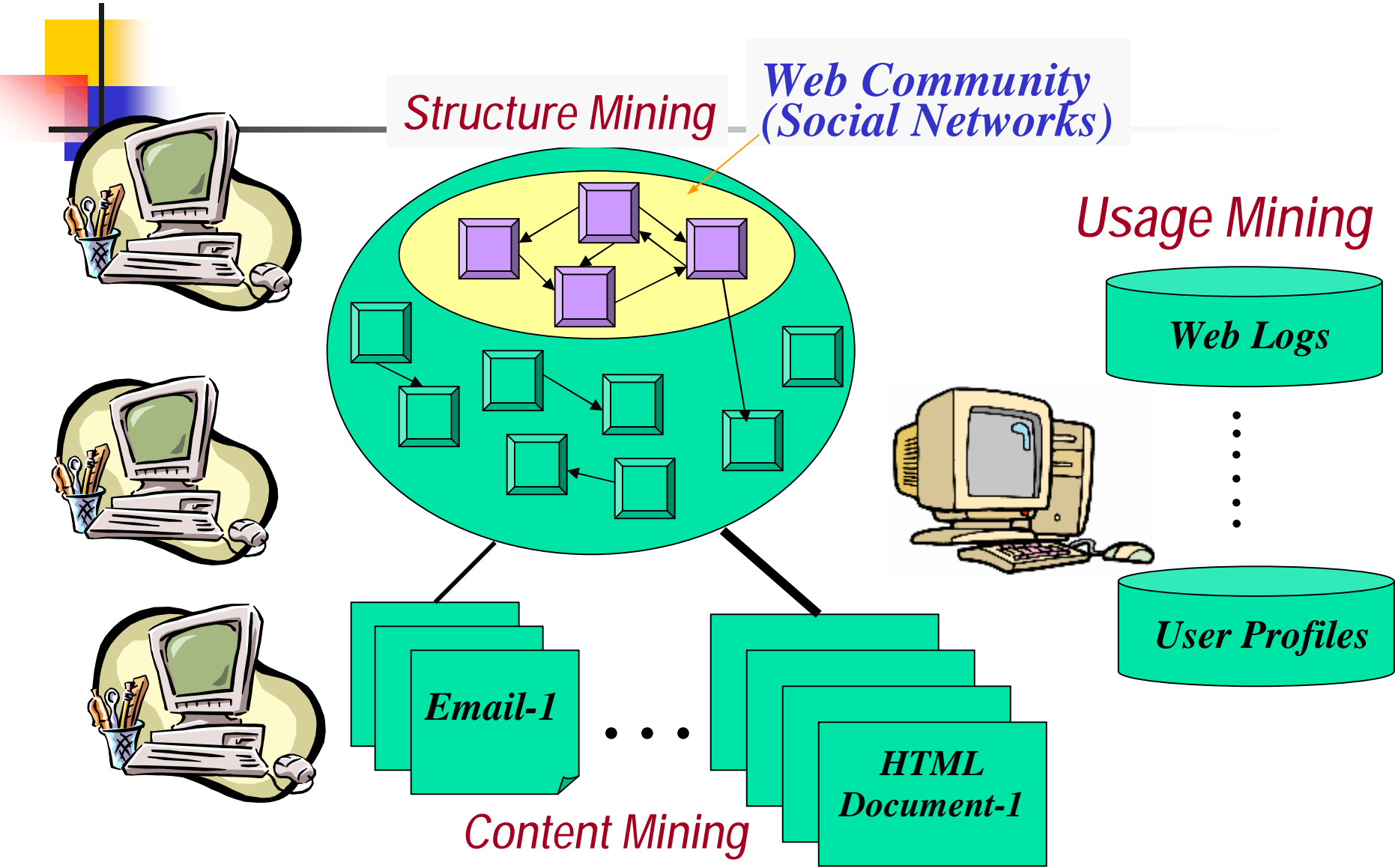


# Modeling User Groups

---

- **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.

# Web Mining





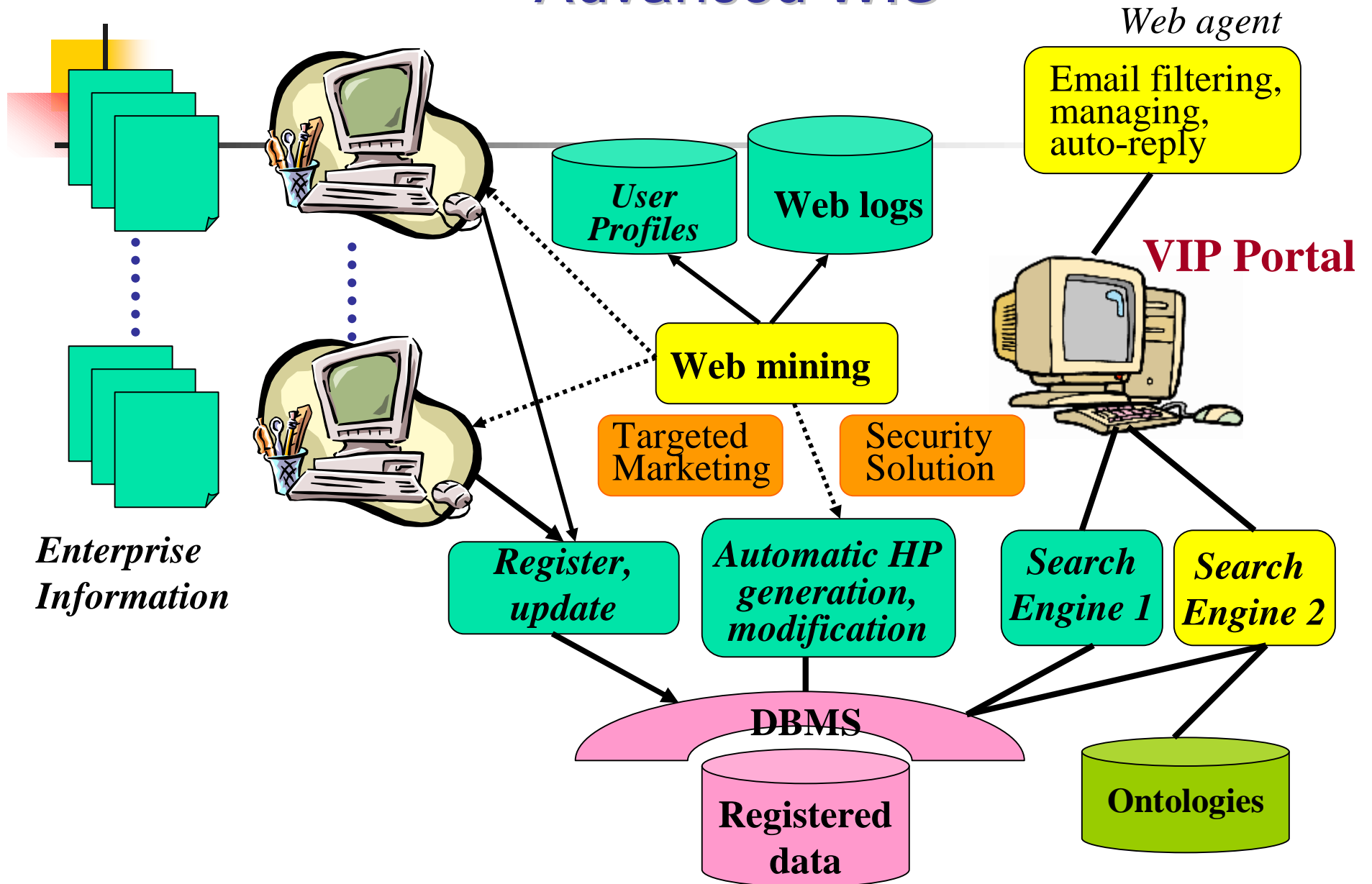
## Personalization: A Prime Factor for

---

- ≈ 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





# Web Based Targeted Marketing

---

- **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

---

- 2 We argue that the next paradigm shift in the Web is towards **wisdom**.
- 2 Developing the **Wisdom Web** will become a tangible goal for WI research.
- 2 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

---

- 2 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**.



# Fundamental Capabilities of the Wisdom Web (2)

---

- 2 2. Problem Solver Markup Language (PSML): Specifying **roles**, **settings**, and **relationships** with any other services.
- 2 3. Semantics: Providing **right judgment of concepts**, such as “best” and “season” and **granularities** of their corresponding subjects.
- 2 4. Meta-knowledge: Dealing with spatial or temporal constraints/conflicts in **planning and executing services**.



# Fundamental Capabilities of the Wisdom Web (3)

---

- ≥ 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-4

Application-level ubiquitous computing  
and social intelligence utilities

Level-3

Knowledge-level information processing  
and management tools

Level-2

Interface-level multi-media presentation  
standards

Level-1

Internet-level communication, infrastructure,  
and security protocols





## Levels of WI (2)

---

- 2 The study of WI concerns the important issues central to **social network intelligence**.
- 2 Furthermore, the multimedia contents on the Web are not only accessible from stationary, but increasingly mobile platforms.
- 2 *Ubiquitous Web access and computing* from various wireless devices needs adaptive personalization.



# 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



# 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

---

- ² 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**.
- ² A “semantic” language that ties the information on a page to machine readable semantics (**ontology**).

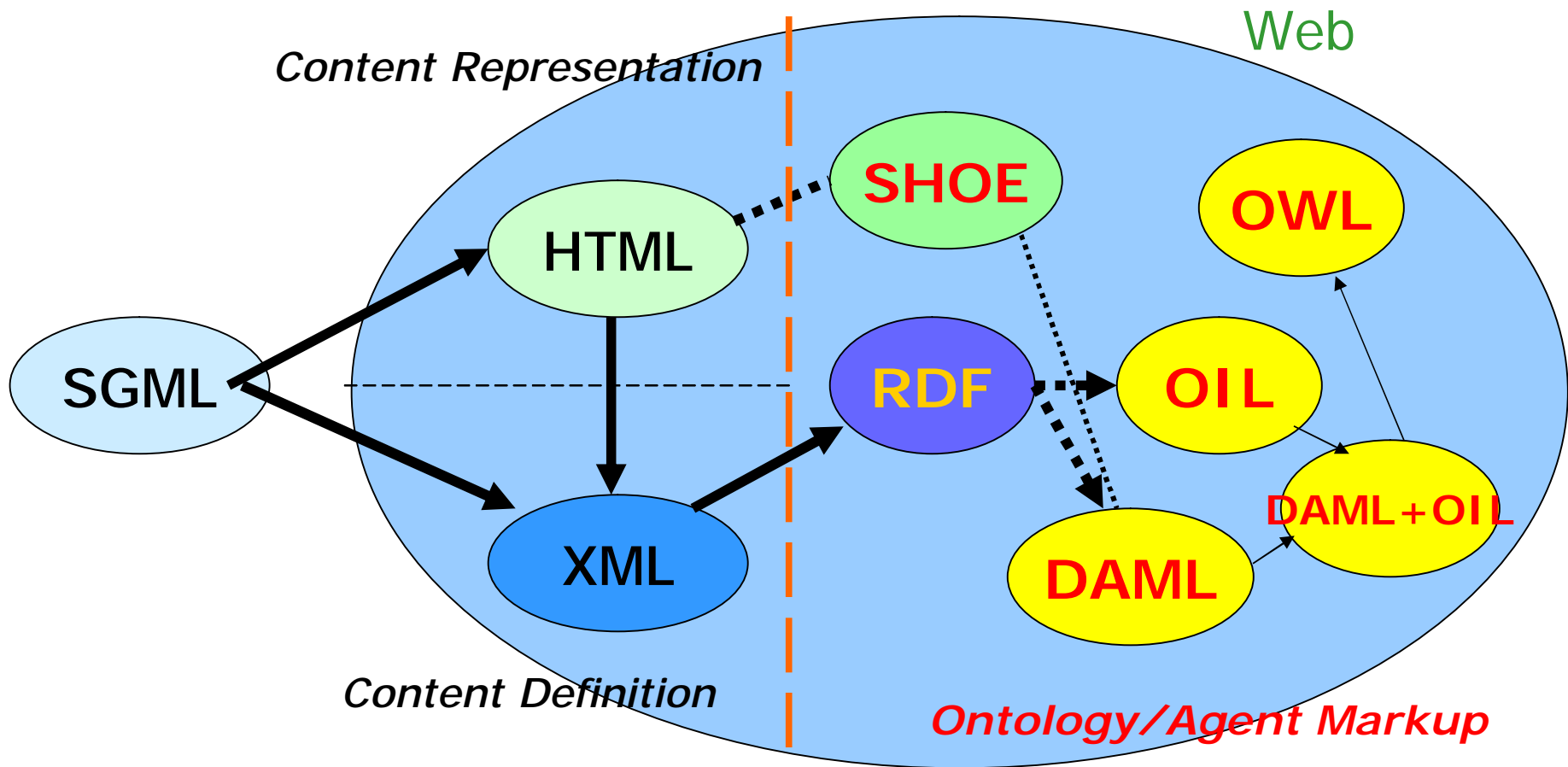


# Components of Semantic Web Techniques

---

- 2 A unifying **data model** such as RDF.
- 2 **Languages** with defined semantics, built on RDF, such as OWL (DAML+OIL).
- 2 **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







# Observation

---

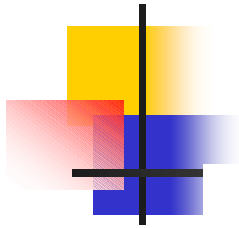
- 2 *Ontologies* and *agent technology* can play a crucial role in WI by enabling **Web-based knowledge processing, sharing, and reuse** between applications (corresponding to **Level-3 of WI**).



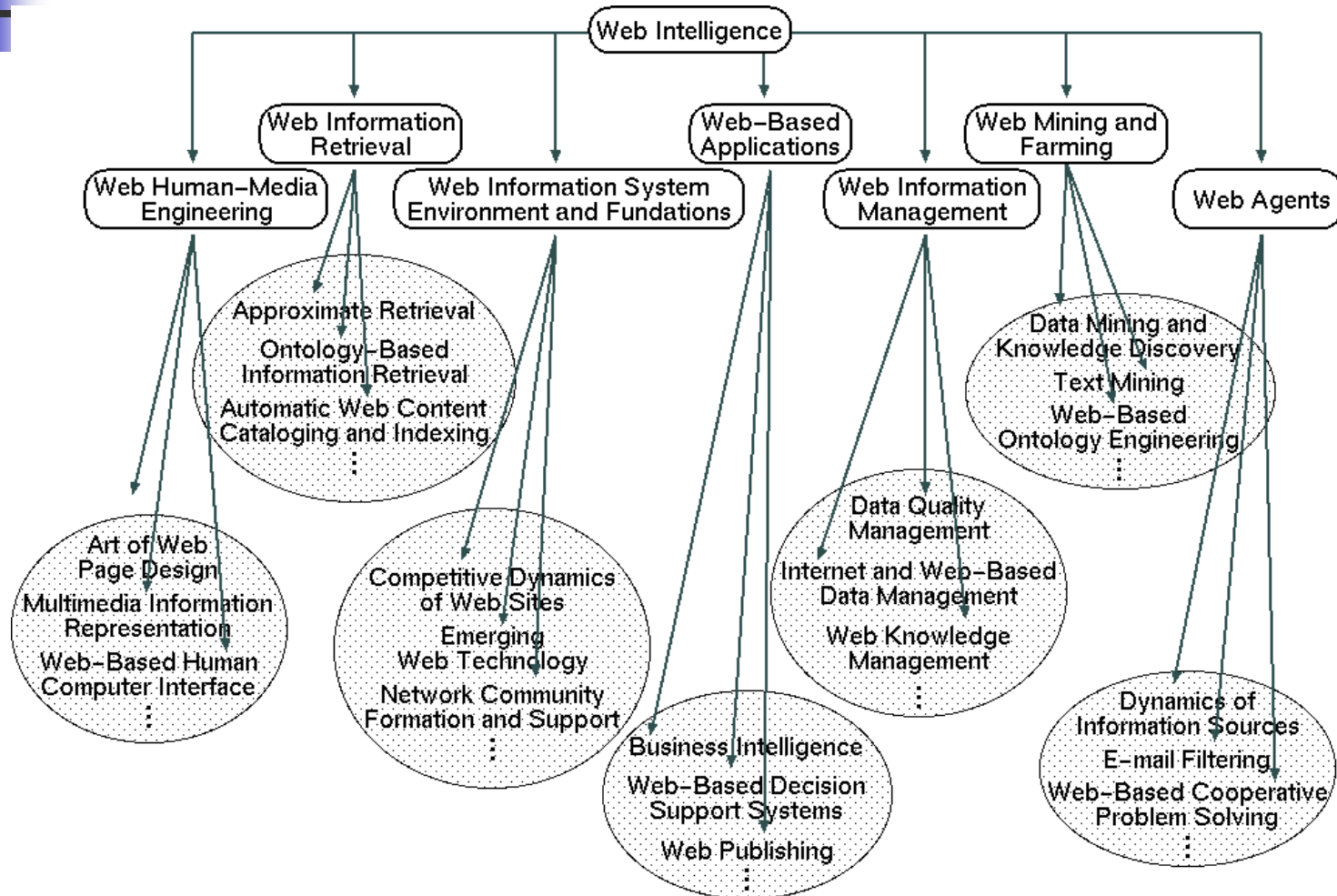
# Categories of Ontologies

---

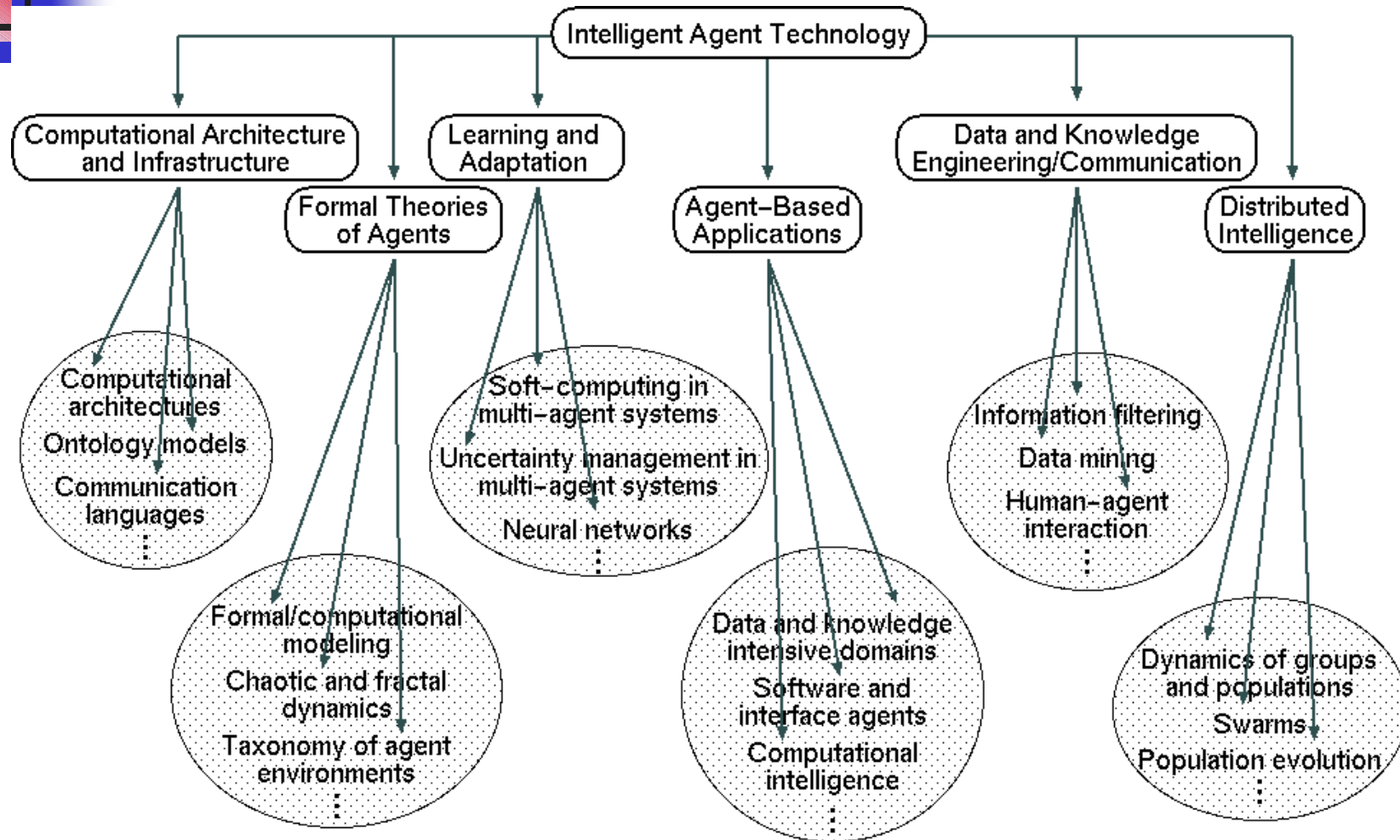
- A **domain-specific** ontology describes a well-defined technical or business domain.
- A **task** ontology might be either domain-specific or reconstructed from a set of domain-specific ontologies for meeting the requirement of a task.
- A **universal** ontology describes knowledge at higher levels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



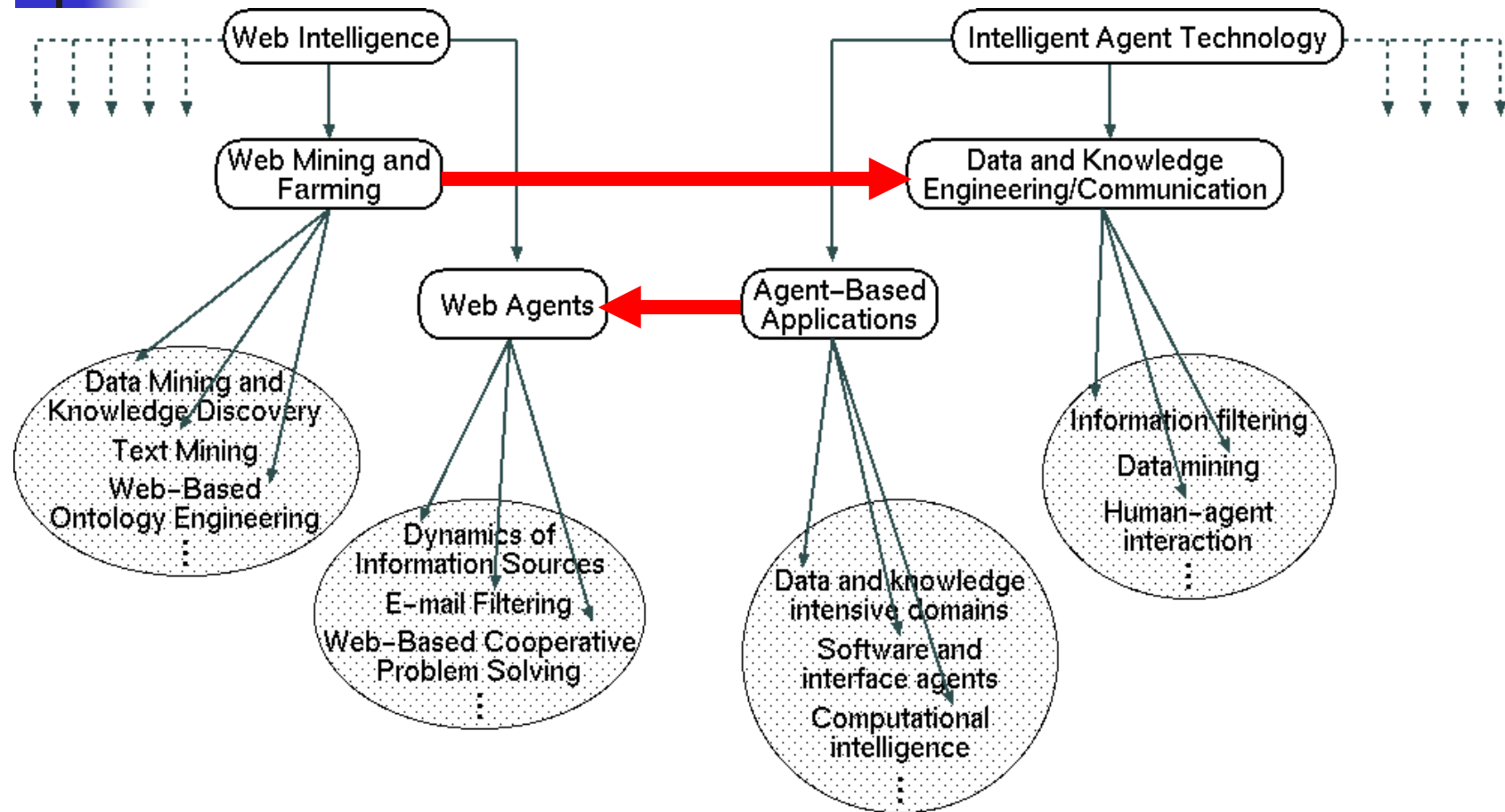
# A Domain-Specific Ontology on WI



# A Domain-Specific Ontology on IAT



# A Task Ontology Generated from the WI and IAT Ontologies

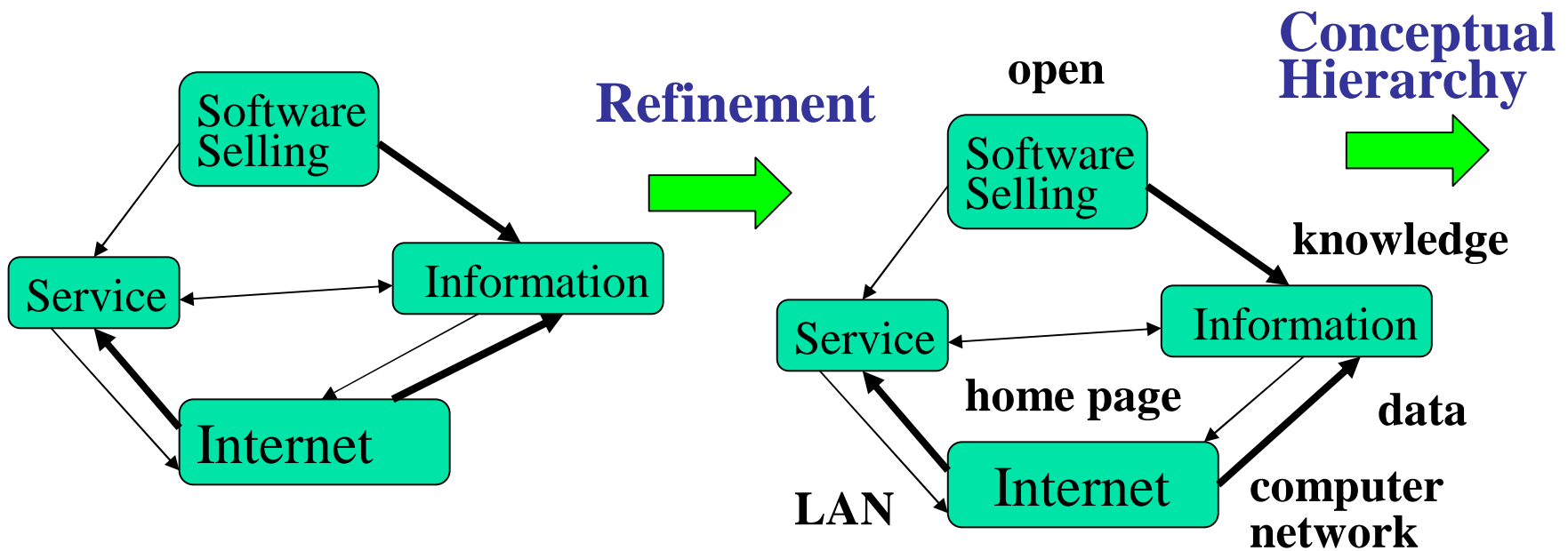
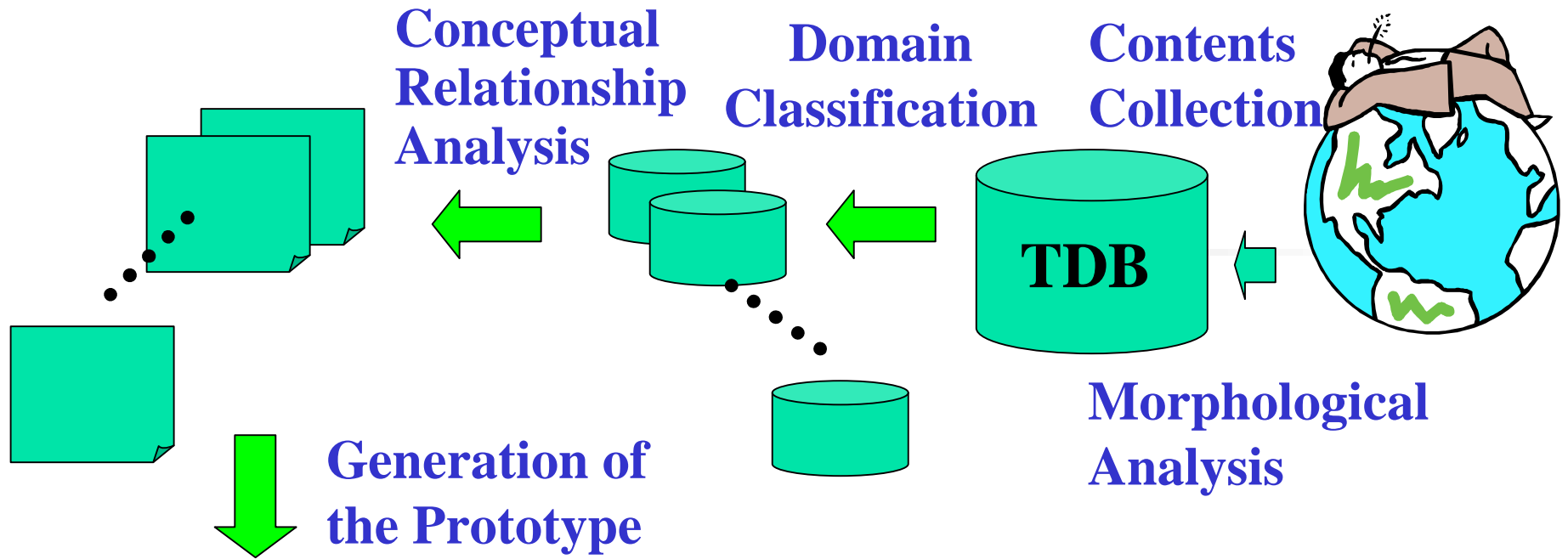




# Automatic Construction and management of Domain-Specific Ontologies

---

A Challenging Topic in WI





# An Agenda

---

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





# The Web as a Graph

---

- 2 We can view the Web as a directed network that connects **people** (organizations or social entities).
  
- 2 Research Questions:
  - 2 How big is the graph? (outdegree and indegree)
  - 2 Can we browse from any page to any other? (clicks)
  - 2 Can we exploit the structure of the Web? (searching and mining)
  - 2 How to discover and manage the Web communities?
  - 2 What does the Web graph reveal about social dynamics?
  - 2 How different is browsing from a "random walk"?



# The Web as a Graph (2)

---

- 2 Exploit structures of the Web for
  - 2 Crawl strategies
  - 2 Search
  - 2 Mining communities
- 2 Classification/organization
- 2 **Web anthropology**
  - 2 Prediction, discovery of structures
  - 2 Sociological understanding
  - 2 **The Web world is now becoming an integral part of our society.**



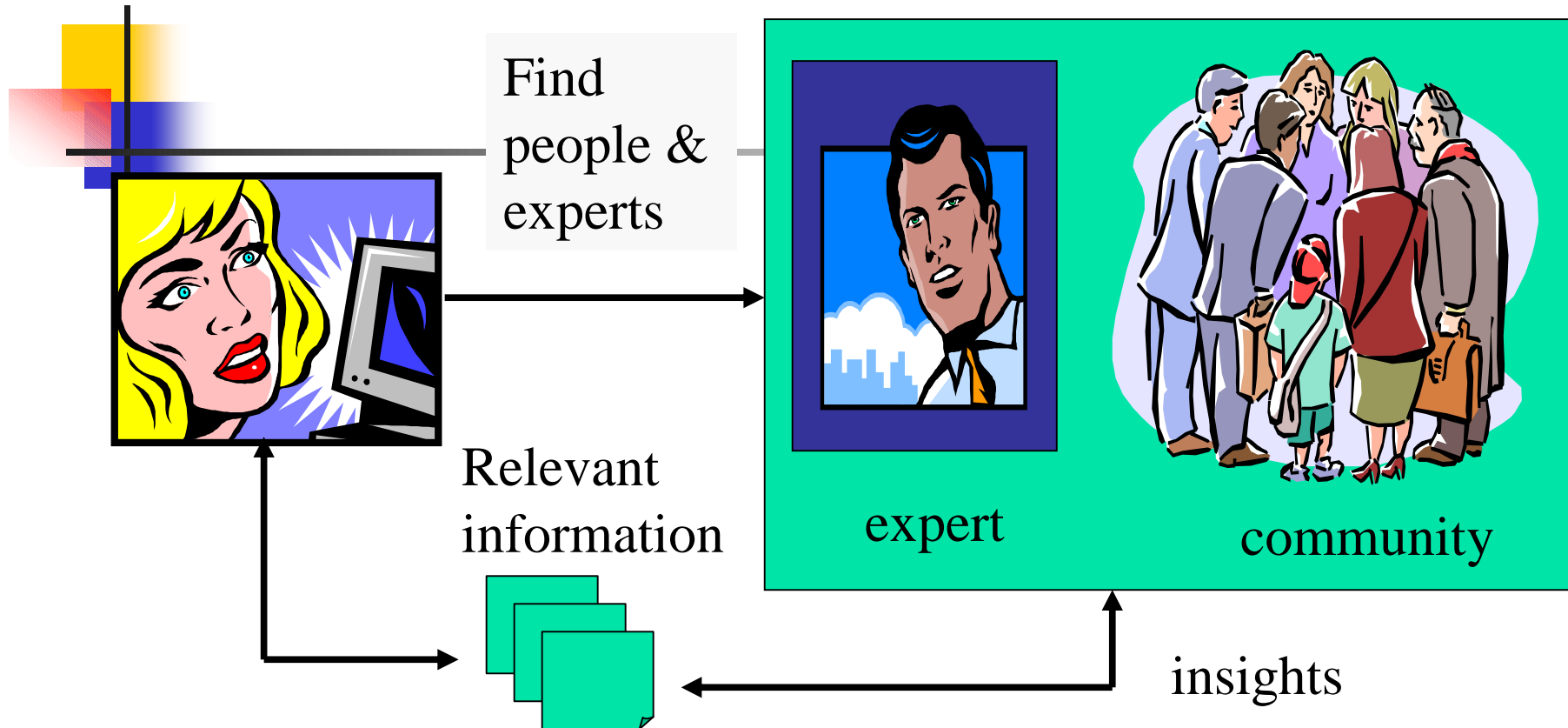
# What/Why Social Network?

---

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



Social Network theory is now significantly influencing **search engine** and **portal development** for the Web and the enterprise.



# An Agenda

---

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



# Existing Web Information Structures

---

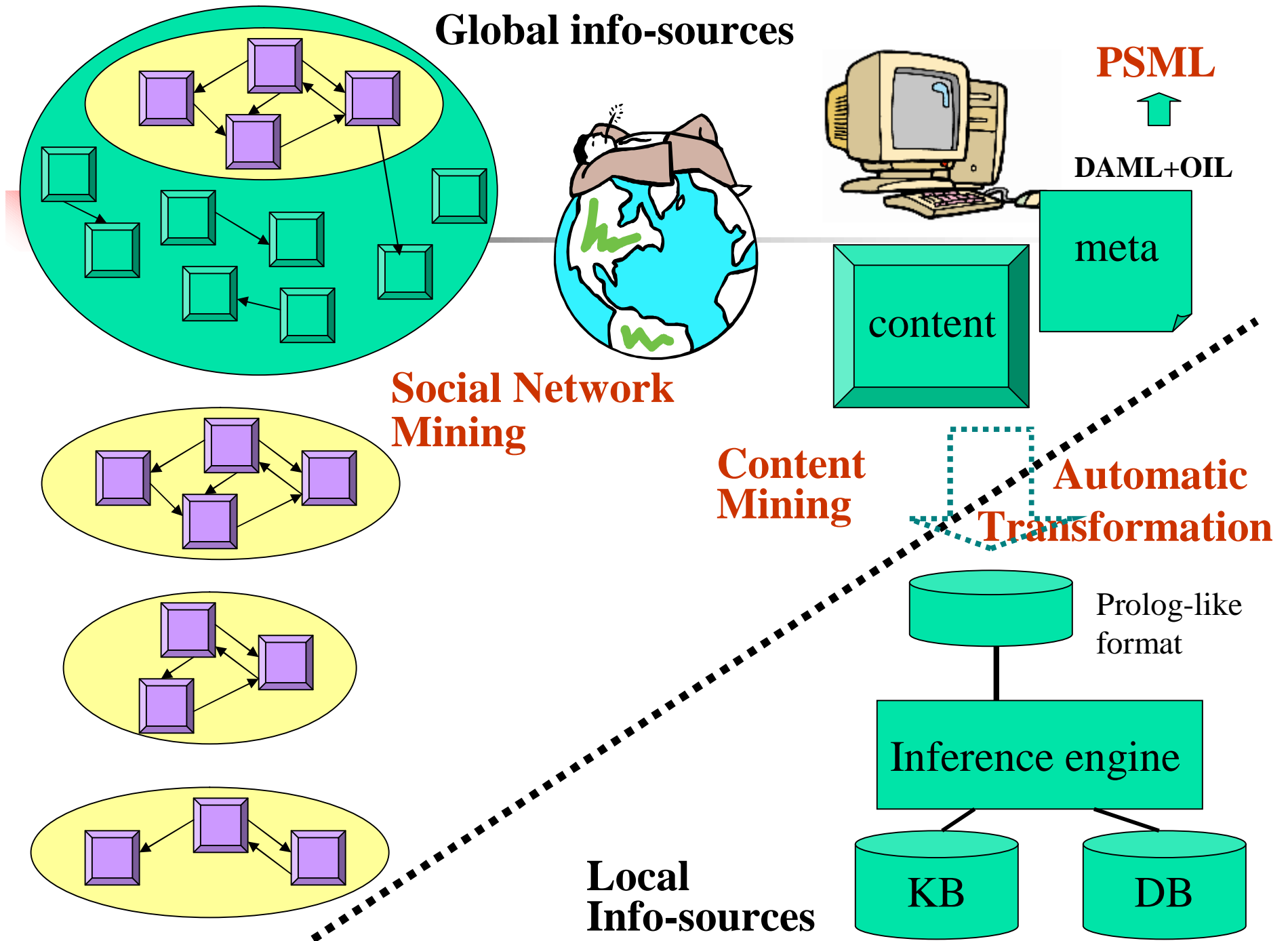
- 2 Web information/knowledge could be:
  - 2 either **globally, distributed** throughout the Web,
  - 2 or **locally, centralized** on an intelligent portal providing Web services (i.e. the intelligent service provider) that is joined to its own cluster of specialized intelligent applications.



## How to Integrate Global and Local Information?

---

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







# An Agenda

---

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



# Wisdom Web Based Computing

---

- 2 To provide
  - 2 not only a medium for information/knowledge exchange/sharing
  - 2 but also a type of man-made resources for sustainable knowledge creation and scientific and social evolution.
- 2 The **Wisdom Web** will rely on **grid-like agencies**:
  - 2 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
  - 2 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 ...

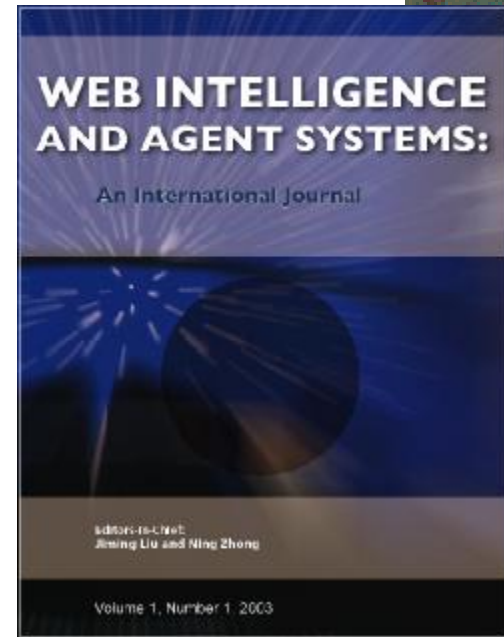
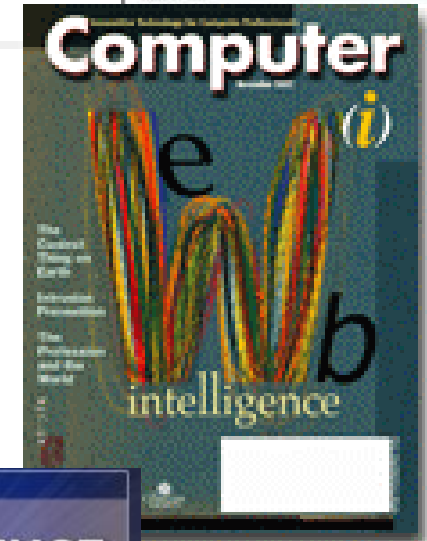
## Books



[wi-consortium.org](http://wi-consortium.org)

n The IEEE/WIC International Conference on Web Intelligence

n IEEE CS TCCI



## Journals



# Web Intelligence Consortium (WIC)

---

- 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

---

- n Profs. Jiming Liu, Yiyu Yao, Edward A. Feigenbaum, Setsuo Ohsuga, Benjamin Wah, Philip Yu, Lotfi A. Zadeh, and Xindong Wu, etc.
- n WIC Technical Committee
- n WIC Research Centers in Australia, Beijing, Canada, India, Japan, and Spain, among others.



# References & Further Reading

---

- N. Zhong, J. Liu, Y.Y. Yao (eds.): Web Intelligence, Springer, March 2003
- J. Liu, N. Zhong, Y.Y. Yao, Z. Ras: “The Wisdom Web: New Challenges for Web Intelligence (WI)”, Journal of Intelligent Information Systems, 20 (1) 2003
- N. Zhong, J. Liu, Y.Y. Yao (eds.): IEEE Computer Special Issue on Web Intelligence, 35 (11) (November 2002)
- N. Zhong, J. Liu, Y.Y. Yao: “In Search of The Wisdom Web”, IEEE Computer, 35 (11) (2002) 27-31
- N. Zhong, Y.Y. Yao, J. Liu, S. Ohsuga (eds.): Web Intelligence: Research and Development, LNAI 2198, Springer (2001)
- N. Zhong, A. Nakamaru, M. Ohshima, J. Wu, H. Mizuhara “Peculiarity Oriented Mining in Multiple Human Brain Data”, Proc. Ideal 2003, LNCS Springer (2003)
- N. Zhong, Y.Y. Yao, M. Ohshima “Peculiarity Oriented Multi-Database Mining”, IEEE TKDE, 15 (4) (2003)



## References & Further Reading (2)

---

- 2 N. Zhong “Toward Web Intelligence”, Proc. AWIC’03 LNAI Springer (2003)
- 2 Y.Y. Yao, N. Zhong, J. Liu, S. Ohsuga: “Web Intelligence (WI): Research Challenges and Trends in the New Information Age”, in N. Zhong et al (eds) LNAI 2198 (2001) 1-17
- 2 T. Berners-Lee, J. Hendler, O. Lassila, “The Semantic Web”, Scientific American, 284 (5) (2001) 34-43
- 2 H.P. Alesso and C.F. Smith, The Intelligent Wireless Web, Addison-Wesley, Reading, Mass. (2000)
- 2 R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins “The Web and Social Networks”, IEEE Computer Special Issue on Web Intelligence, 35 (11) (November 2002) 32-36
- 2 A. Congiusta, A. Pugliese, D. Talia, P. Trunfio “Designing Grid Services for Distributed Knowledge Discovery”, Web Intelligence and Agent Systems, an international journal, IOS Press, 1 (2) (2003)
- 2 N. Zhong, C. Liu “Dynamically Organizing KDD Processes”, IJPRAI, World Scientific (2001)





*Thank You !*