

# **Web Intelligence (WI)**

*Research Challenges and Trends  
in the New Information Age*

(WI-2001 Keynote Talk)

Y.Y. Yao, Ning Zhong, Jiming Liu, Setsuo Ohsuga

# Outline

- Background information
- What is Web Intelligence (WI)?
- Why do we treat WI as a new sub-discipline of Computer Science?
- Topics of WI
- Current status
- Future directions

# Background (1999 -)

- Ning Zhong: Data (Web) Mining
- Jiming Liu: Intelligent Agents
- Setsuo Ohsuga: Intelligent Information Systems
- Y.Y. Yao: Information Retrieval
- Common and Combined Research Interests: Web Intelligence.

# What is WI?

**“Web Intelligence (WI) exploits Artificial Intelligence (AI) and advanced Information Technology (IT) on the Web and Internet.”**

- Combination of AI and IT.
- Applications of AI on the Web.
- Intelligent information systems on the Web.

# What is WI?

**“Web Intelligence (WI) concerns the design and implementation of intelligent systems on the new platform of the Web and Internet.”**

## IWIS

Intelligent **W**eb **I**nformation **S**ystems

# What is WI?

**“Web Intelligence (WI) can be characterized as the problems, theories, methodologies, and techniques studied by WI researchers.”**

- Everyone here will help define WI.
- Your input and participation will help define WI.
- WI will be judged by our achievements.



# Why WI?

- A fast growing interest in WI.

Statistics on Searching Web Intelligence on the Web  
(February 2001)

Industry 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.

# Why WI?

- The web provides a new and unique platform for computer applications.
  - Size
  - Complexity
- The Web increases the availability and accessibility of information to a much larger community than other computer applications.

# Why WI?

- The Web present new challenging and research problems.
- Existing theories and technologies need to be modified or enhanced.
- A new sub-discipline devoted to Web related research and applications might have a significant value.

# Perspectives of WI

- WI may be viewed as applying results from these 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 AI and IT.
- WI may become a sub-area of AI and IT or a child of a successful marriage of AI and IT.

# Perspectives of WI

- Studies on WI can benefit a great deal from the results, experience, success and lessons of AI and IT.
- A foundation of WI can be established by drawing results from many related disciplines.

# Perspectives of WI

- Mathematics: computation, logic, probability.
- Applied Mathematics and Statistics: algorithms, non-classical logics, decision theory, information theory, measurement theory, utility theory, theories of uncertainty, approximate reasoning.
- Psychology: cognitive psychology, cognitive science, human-machine interaction, user interface.

# Perspective of WI

- Linguistics: computational linguistics, natural language processing, machine translation.
- Information Technology: information science, databases, information retrieval systems, knowledge discovery and data mining, expert systems, knowledge-based systems, decision support systems, intelligent information agents.

# IWIS (Intelligent Web Information Systems)

- Four categories of AI systems (Russell and Norvig)  
designing philosophy of AI systems

ability,  
functionality  
of AI systems

System that thinks like humans	System that thinks rationally.
System that acts like humans.	System that acts rationally.

- IWIS can be similarly classified.

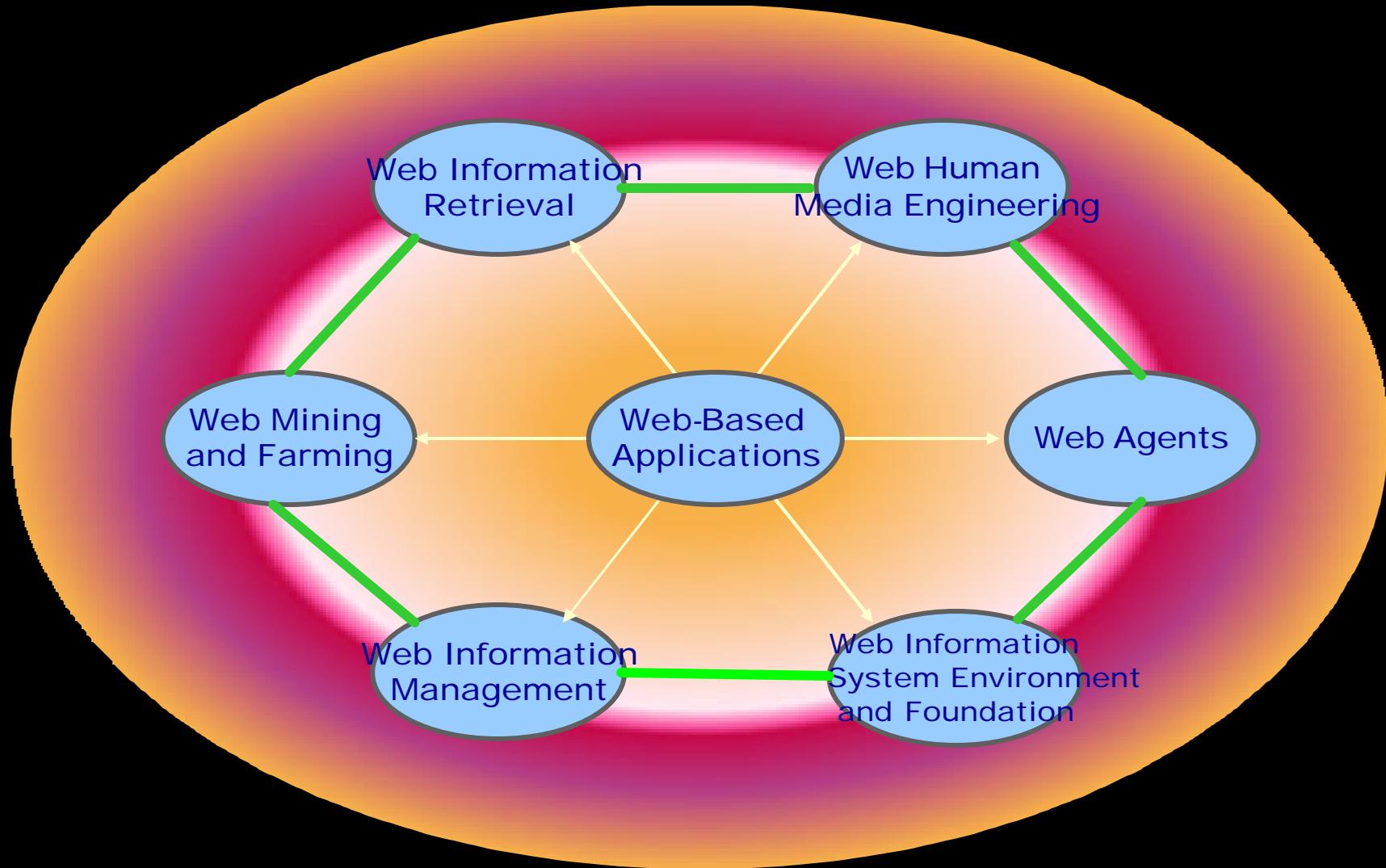
# IWIS (Intelligent Web Information Systems)

- A full understanding of an intelligent system involves explanations at various levels.
- Conceptual formulation and mathematical modeling.
- Physical system design and implementation.
- Many levels of development can be defined.

# Topics of WI

- Prefix labeling: adding “Web” as a prefix to an existing topic.
  - Web digital library
  - Web information retrieval
  - Web agents
- Postfix labeling: adding “Web” as a postfix to an existing topic.
  - digital library on the Web
  - information retrieval on the Web

# Topics of WI



# Web Mining

- Find patterns, knowledge, etc. from Web usage files or Web documents.
  - Web log mining
  - Web structure mining
  - Web content mining
- Applications:
  - E-commerce
  - Targeted marketing

# Web Mining

- From the explicit physical links between Web documents and structures of Web documents, extract implicit logical connections between documents or segments of documents.
  - citation of papers
  - clustering of Web documents
  - identifying people with common interests
  - identifying emerging research topics and communities on the Web.
- Generating logical, personal views of the Web.

# Web Information Retrieval

- Search engines.
- Multi-lingual retrieval.
- Search agents.
- Text analysis and text mining.
- Ontology.
- Semantic web and markup languages.

# Granular Retrieval Model

- A rough description or granulated view of the Web documents can be obtained by ignoring some unimportant details.
- Three languages should be supported in the granular retrieval model
  - a “granulation” language allows a user to create a granulated view.
  - a “navigation” language allows a user to change views
  - a “retrieval” language allows the user to perform an actual search.

# Intelligent Web Agents

- Personalized Multimodal Interface.
- Push and Pull.
- Pattern Discovery and Self-Organization.
- Information Gateway.
- Reward.
- Matchmaking.
- Decision support.
- Delegation.
- Collaborative Work Support.

# WI: Current status

- **The 2001 International Conference on Web Intelligence.**
- **IEEE COMPUTER -- Special Issue on WEB INTELLIGENCE (to appear in October 2002).**
- Special issues on Web Intelligence on several leading journals.
- A hard-covered edited book on Web Intelligence.

# WI: Future

- **It is difficult to predict the future without uncertainty.**
- **The interest in WI is growing very fast.**
- **WI may serve as a sub-discipline of computer science on its own rights.**
- **WI is attracting and will attract the best researchers in the field.**
- **Your input and participation will determine the future of WI.**

# **Thank you!**

- **Please contact us if you need further information or have suggestions on WI.**
- **Please spread the word regarding any WI information.**
- **Let us work together on WI.**

# Size

- October 2001, Google can search 1,610,476,000 pages.
- February 2001, Google was searching 1,346,966,000 pages.

- Steve Lawrence and C. Lee Giles,

NEC Research Institute (1998)

<http://www.neci.nec.com/~lawrence/websize.html>

An estimated lower bound on the size of the indexable Web is 320 million pages.     [back](#)

# Complexity

- Connectivity and diversity of Web documents.
- A heterogeneous collection of structured, unstructured, semi-structured, inter-related, and distributed Web documents consisting of text, images and sounds.
- Theories, methodologies and technologies of Web based information systems.
- Many different types of search engines and agents.

[back](#)

# Statistics on Searching Web Intelligence on the Web

Search Engine	Number of Hits
Lycos	1,102,279
Google	1,080,000
Excite	223,825
AltaVista	1,271
Netscape	77
Yahoo	74
LookSmart	62

[back](#)

# Industry interest in WI

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

[back](#)

# ResearchIndex database

- Effective Personalization Based on Association Rule.. - Bamshad Mobasher Honghua  
Center for **Web Intelligence**, School of Computer Science, International Conference on Information and Knowledge Management (CIKM 2001).
- FOCI: A Personalized Web Intelligence System - Ah-Hwee Tan Hwee-Leng  
*Proceedings of IJCAI workshop on Intelligent Techniques for Web Personalisation, Seattle, pp. 14-19, August 2001.*
- Integrating E-Commerce and Data Mining: Architecture and.. - Suhail Ansari Ron  
"Leaders will use metrics to fuel personalization" and that "firms need **web intelligence**, not log analysis." [back](#)