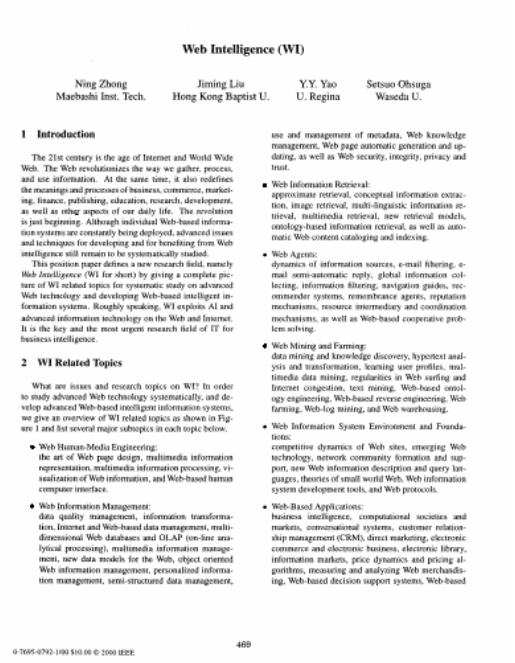
A Brief History of WI

- 1999: Collaborative research initiatives
  - Ning Zhong, Data Mining and Knowledge Systems
  - Jiming Liu, Intelligent agents and multi-agents
  - Yiyu Yao, Information retrieval and intelligent information systems

- Combined research efforts with common goal: create a new sub-discipline covering theories and techniques related to web information.

A Brief History of WI

- 2000: Publication of a two-page position paper on WI (Zhong, Liu, Yao, Ohsuga, COMPSAC 2000)



Web Intelligence (WI)

Ning Zhong, Maebashi Inst. Tech. · Jiming Liu, Hong Kong Baptist U. · Y.Y. Yao, U. Regina · Setsuo Ohsuga, Waseda U.

1 Introduction

The 21st century is the age of Internet and World Wide Web. The Web revolutionizes the way we gather, process, and use information. At the same time, it also redefines the meanings and processes of business, commerce, marketing, finance, publishing, education, research, development, as well as other aspects of our daily life. The revolution is just beginning. Although individual Web-based information systems are constantly being deployed, advanced issues and techniques for developing and for benefiting from Web intelligence still remain to be systematically studied.

This position paper defines a new research field, namely *Web Intelligence* (WI for short) by giving a complete picture of WI related topics for systematic study on advanced Web technology and developing Web-based intelligent information systems. Roughly speaking, WI exploits AI and advanced information technology on the Web and Internet. It is the key and the most urgent research field of IT for business intelligence.

2 WI Related Topics

What are issues and research topics on WI? In order to study advanced Web technology systematically, and develop advanced Web-based intelligent information systems, we give an overview of WI related topics as shown in Figure 1 and list several major subtopics in each topic below.

- Web Human-Media Engineering:
  the art of Web page design, multimedia information representation, multimedia information processing, visualization of Web information, and Web-based human computer interface.
- Web Information Management:
  data quality management, information transformation, Internet and Web-based data management, multi-dimensional Web databases and OLAP (on-line analytical processing), multimedia information management, new data models for the Web, object oriented Web information management, personalized information management, semi-structured data management, use and management of metadata, Web knowledge management, Web page automatic generation and updating, as well as Web security, integrity, privacy and trust.
- Web Information Retrieval:
  approximate retrieval, conceptual information extraction, image retrieval, multi-linguistic information retrieval, multimedia retrieval, new retrieval models, ontology-based information retrieval, as well as automatic Web content cataloging and indexing.
- Web Agents:
  dynamics of information sources, e-mail filtering, e-mail semi-automatic reply, global information collecting, information filtering, navigation guides, recommender systems, remembrance agents, reputation mechanisms, resource intermediary and coordination mechanisms, as well as Web-based cooperative problem solving.
- Web Mining and Farming:
  data mining and knowledge discovery, hypertext analysis and transformation, learning user profiles, multimedia data mining, regularities in Web surfing and Internet congestion, text mining, Web-based ontology engineering, Web-based reverse engineering, Web farming, Web-log mining, and Web warehousing.
- Web Information System Environment and Foundations:
  competitive dynamics of Web sites, emerging Web technology, network community formation and support, new Web information description and query languages, theories of small world Web, Web information system development tools, and Web protocols.
- Web-Based Applications:
  business intelligence, computational societies and markets, conversational systems, customer relationship management (CRM), direct marketing, electronic commerce and electronic business, electronic library, information markets, price dynamics and pricing algorithms, measuring and analyzing Web merchandising, Web-based decision support systems, Web-based

0-7695-0792-1/00 $10.00 © 2000 IEEE

Figure 1. A schematic diagram of WI related topics

distributed information systems, Web-based electronic data interchange (EDI), Web marketing, and Web publishing.

3 WI Related Case Studies

WI presents an excellent opportunity as well as challenge for the research and development of new generation of information processing technology, as well as for exploiting business intelligence. Specifically, e-commerce activity that involves the end user is undergoing a significant revolution [8]. The ability to track users' browsing behavior down to individual mouse clicks has brought the vendor and end customer closer than ever before. It is now possible for a vendor to personalize his product message for individual customers at a massive scale. This is called *Targeted Marketing*.

Hackathorn proposed Web farming that is the systematic refining of information resources on the Web for business intelligence [3].

Adriaans et al. systematically investigated the data on the Web and the features of semistructured data [1].

Zhong, Yao et al. proposed a way of mining peculiar data and peculiarity rules that can be used for Web-log mining [11]. They proposed ways for targeted marketing by mining classification rules and market value functions [15, 10]. They are also working on text mining on the Web including automatic construction of ontology, e-mail filtering systems, and Web-based e-business systems [14, 16].

Liu et al. are working on e-commerce agents [7]. Liu and Zhong are also working on Web agents and KDDA (Knowledge Discovery and Data Mining Agents) [5, 6].

4 WI Conferences

We initiated a new high-quality, high-impact biennial conference series, namely the Asia-Pacific Conference on Web Intelligence (WI). The first meeting in this new series, WI'2001, will be held in Maebashi City, Japan, October 23-26, 2001 (http://kis.maebashi-it.ac.jp/wi2001).

WI'2001 is an international forum for researchers and practitioners to present the state-of-the-art in the development of Web intelligence, to examine performance characteristics of various approaches in Web-based intelligent information technology, and to cross-fertilize ideas on the development of Web-based intelligent information systems among different domains. By idea-sharing and discussions on the underlying foundations and the enabling technologies of Web intelligence, WI'2001 is expected to stimulate the future development of new models, new methodologies, and new tools for building a variety of embodiments of Web-based intelligent information systems.

References

[1] Adriaans, P., Zantinge, D. *Data Mining*. Addison-Wesley (1996).

[2] Agrawal, R., Imielinski, T., Swami, A. "Mining Association Rules between Sets of Items in Large Databases". *Proc. of the ACM SIGMOD Conference on Management of Data* (1993) 207-216.

[3] Chen, M., Han, J., Yu, P.S. "Data Mining: An Overview from a Database Perspective". *IEEE Trans. on Knowledge and Data Engineering* 8 (1996) 866-883.

[4] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press (1996).

[5] Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J. "Knowledge Discovery in Databases: An Overview". In [4] 1-27.

[6] Han, J., Cai, Y., Cercone, N. "Data-Driven Discovery of Quantitative Rules in Relational Databases". *IEEE Trans. on Knowledge and Data Engineering* 5 (1993) 29-40.

[7] Imielinski, T., Virmani, A., Abdulghani, A. "DataMine: Application Programming Interface and Query Language for Database Mining". *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining* (1996) 256-262.

[8] Lin, T.Y., Cercone, N. (eds.) *Rough Sets and Data Mining: Analysis of Imprecise Data*. Kluwer Academic Publishers (1997).

[9] Lin, T.Y., Liu, Q. "Rough Approximate Operators: Axiomatic Rough Set Theory". In [8] 256-260.

[10] Lin, T.Y., Yao, Y.Y., Louie, E. "Value Added Association Rules". *Proc. of the Pacific-Asia Conference on Knowledge Discovery and Data Mining* (2002) 328-333.

[11] Liu, B., Hsu, W., Ma, Y. "Integrating Classification and Association Rule Mining". *Proc. of the 4th International Conference on Knowledge Discovery and Data Mining* (1998) 80-86.

[12] Mitchell, T.M. *Machine Learning*. McGraw-Hill (1997).

[13] Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers (1991).

[14] Piatetsky-Shapiro, G., Matheus, C.J. "The Interestingness of Deviations". *Proc. of the AAAI-94 Workshop on Knowledge Discovery in Databases* (1994) 25-36.

[15] Silberschatz, A., Tuzhilin, A. "What Makes Patterns Interesting in Knowledge Discovery Systems". *IEEE Trans. on Knowledge and Data Engineering* 8 (1996) 970-974.

[16] Yao, Y.Y. "Information Tables with Neighborhood Semantics". In [9] 371-384.

[17] Yao, Y.Y., Zhong, N. "Granular Computing using Information Tables". Lin, T.Y., Yao, Y.Y., and Zadeh, L.A. (eds.) *Data Mining, Rough Sets and Granular Computing*. Physica-Verlag (2002).

[18] Zhong, N., Yao, Y.Y., Ohshima, M. "Peculiarity Oriented Multi-Database Mining". *IEEE Trans. on Knowledge and Data Engineering* 15 (2003) 952-960.

[19] Zhong, N., Yao, Y.Y., Ohsuga, S. "Peculiarity Oriented Multi-Database Mining". *Proc. of the 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases* (1999) 136-146.

[20] Zhong, N., Skowron, A., Ohsuga, S. (eds.) *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*. Springer-Verlag (1999).

[21] Zhong, N. and Zhou, L. (eds.) *Methodologies for Knowledge Discovery and Data Mining*. Springer-Verlag (1999).

[22] Zhong, N., Yao, Y.Y., and Kakemoto, Y. "Automatic Construction of Ontology from Text Databases". *Proc. of the 2nd International Conference on Data Mining*. Kluwer Academic Publishers (2000).

[23] Zhong, N., Dong, J.Z., and Ohsuga, S. "Rough Set Based Rule Discovery and Its Applications". *Cybernetics and Systems* 32 (2001) 909-930.

[24] Zhong, N., Liu, C., and Ohsuga, S. "Dynamically Organizing KDD Processes". *International Journal of Pattern Recognition and Artificial Intelligence* 15 (2001) 451-473.

[25] Zhong, N., Yao, Y.Y., and Ohshima, M. "Peculiarity Oriented Multi-Database Mining". *IEEE Trans. on Knowledge and Data Engineering* 15 (2003) 952-960.

[26] Zhong, N., Yao, Y.Y., and Ohshima, M. "Peculiarity Oriented Multi-Database Mining". *Proc. of the 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases* (1999) 136-146.

[27] Zhong, N., Skowron, A., and Ohsuga, S. (eds.) *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*. Springer-Verlag (1999).

[28] Zhong, N. and Zhou, L. (eds.) *Methodologies for Knowledge Discovery and Data Mining*. Springer-Verlag (1999).

[29] Zhong, N., Yao, Y.Y., and Kakemoto, Y. "Automatic Construction of Ontology from Text Databases". *Proc. of the 2nd International Conference on Data Mining*. Kluwer Academic Publishers (2000).

[30] Zhong, N., Dong, J.Z., and Ohsuga, S. "Rough Set Based Rule Discovery and Its Applications". *Cybernetics and Systems* 32 (2001) 909-930.

[31] Zhong, N., Liu, C., and Ohsuga, S. "Dynamically Organizing KDD Processes". *International Journal of Pattern Recognition and Artificial Intelligence* 15 (2001) 451-473.

A Brief History of WI

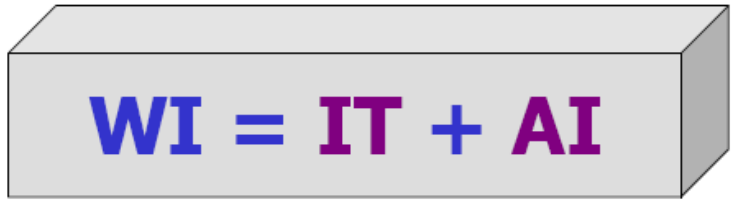
- 2001: First Asia-Pacific Conference on Web Intelligence
- 2002: Publication of first special issue on WI in IEEE Computer
- 2002: Web Intelligence Consortium
- 2003: First edited book on WI
- 2005: The international WIC Institute

Motivation

- The sheer size of Web
  - Difficulties in the storage, management, and efficient and effective retrieval
- Complexity of Web
  - Heterogeneous collection of structured, unstructured, semi-structured, interrelated, and distributed Web documents
  - Consist texts, images and sounds

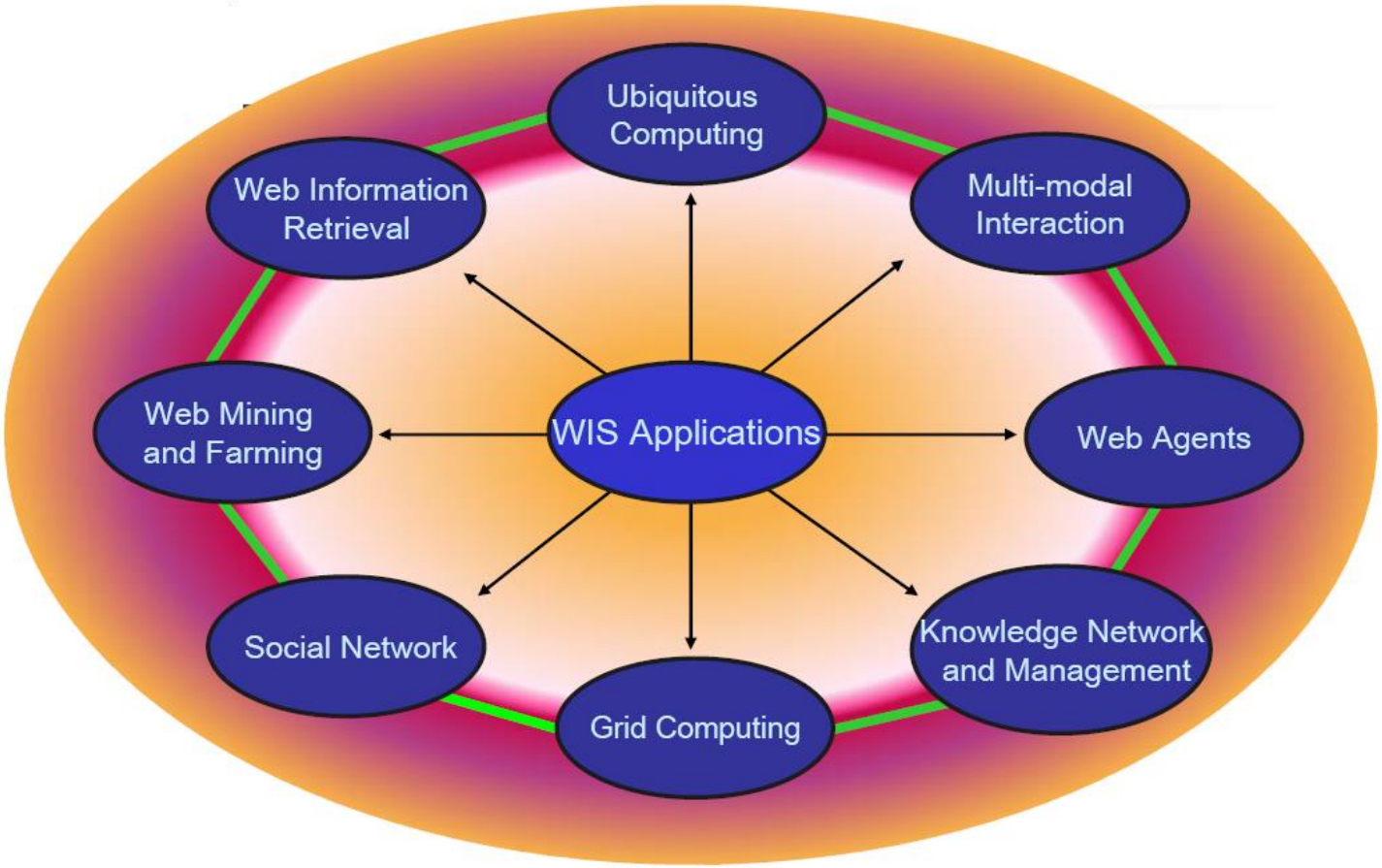
What is Web Intelligence

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



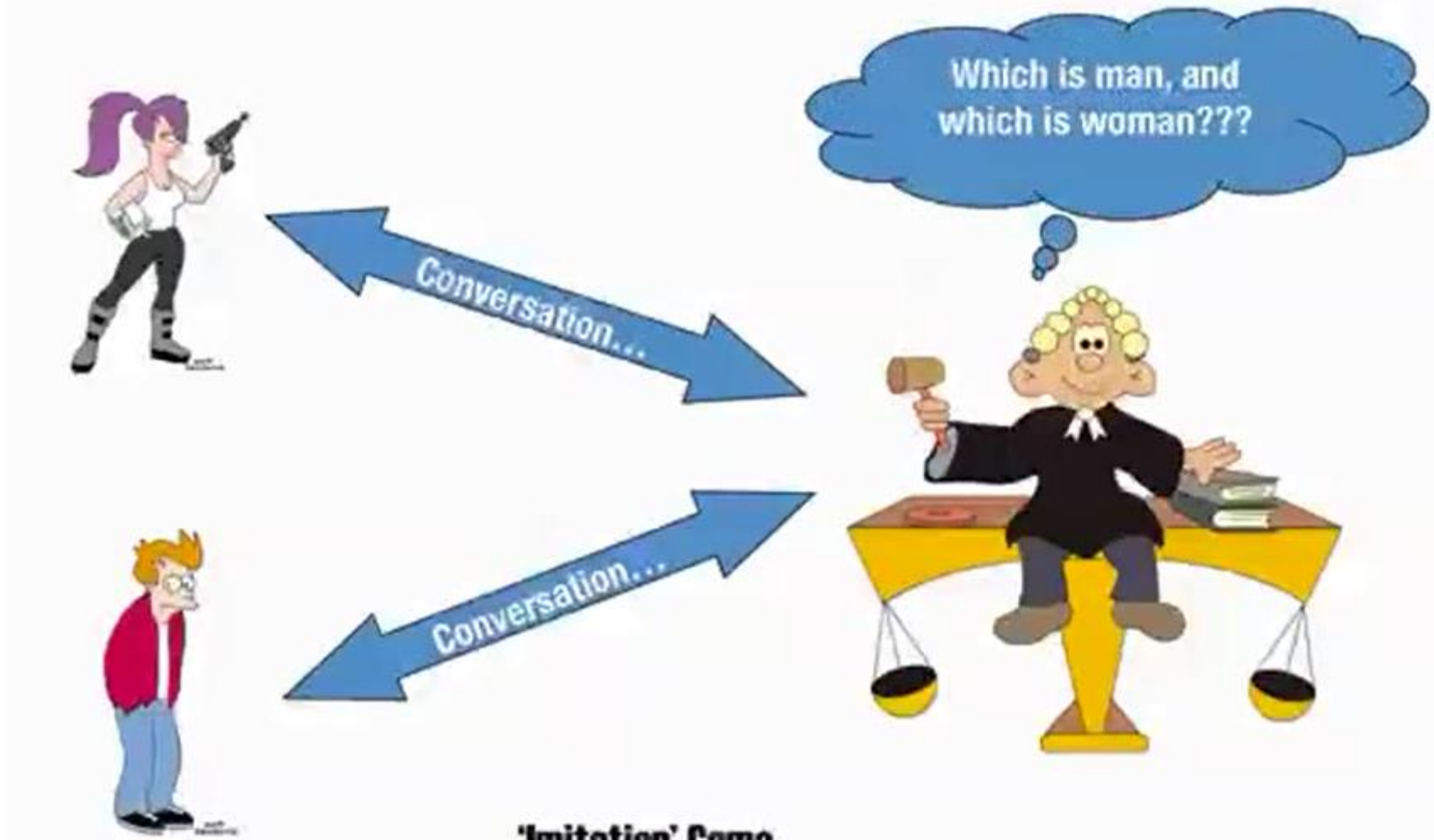
  - Integration of IT \
  - Applications of AI on the Web

Web Intelligence System

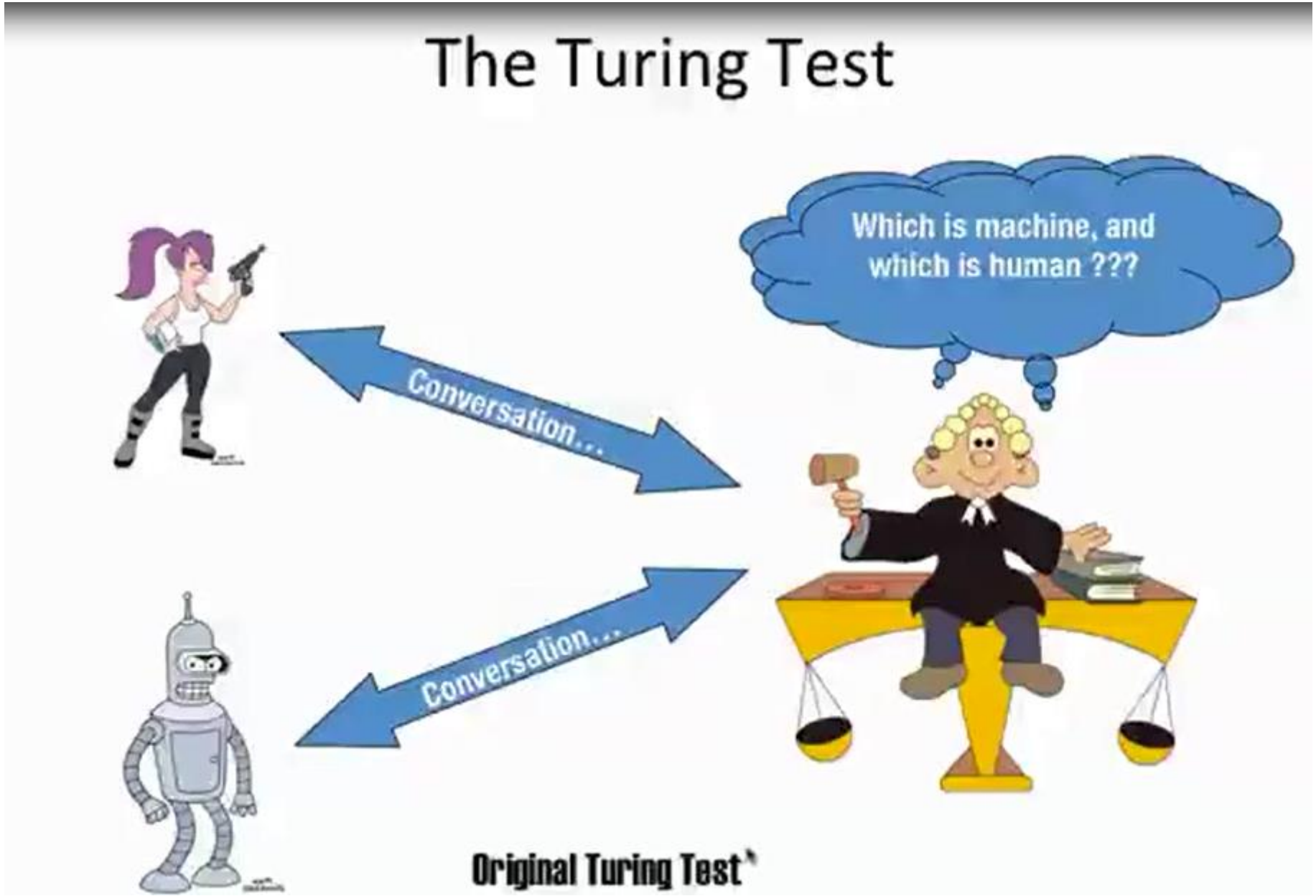


Based on Zhong`s AWIC03 keynote talk

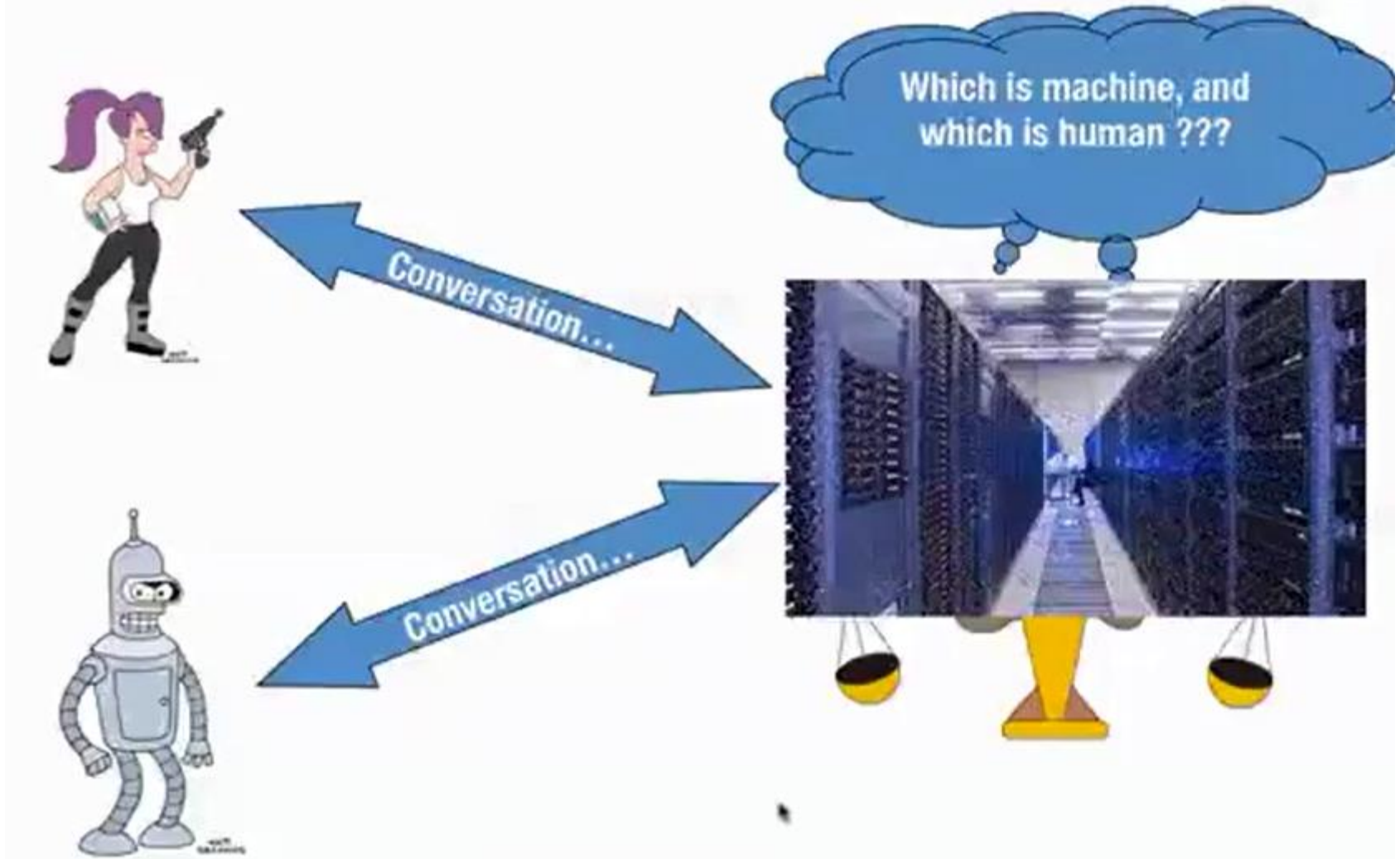
The Turing Test



The Turing Test



The *Reverse* Turing Test



The *Reverse* Turing Test

Conversation...

Which is machine, and which is human ???

Conversation...

overlooks inquiry

The *Reverse* Turing Test

Conversation...

Like / Dislike
Shopper / Surfer
Rich / Poor
Ethnicity

Conversation...

Which is man, and which is woman ???

Happening all the time!