

# Possibilities and Technical Challenges in Information and Knowledge Infrastructure

**Hideaki Takeda, Prof. Dr. Eng.**

Research into Artifacts, Center of Engineering (RACE), The University of Tokyo  
National Institute of Informatics (NII)

*takeda@nii.ac.jp*



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

## Theme in this talk

- Rapid growth of WWW is changing the traditional information distribution system in our society quickly
- What change happened?
- Which direction we are going?

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

# Information distribution system in society

- Ancient days (w/o written languages): oral culture
  - Very limited distribution and preservation of information
- Ancient days (w written languages): handwritten materials
  - Limited distribution and some amount of preservation of information
  - Small amount of information providers and information consumers
- With printing technology after Gutenberg
  - Good enough distribution and preservation of information
  - Small amount of information providers and large amount of information consumers
- Internet Era (WWW Era)
  - So much distribution and preservation of information
  - Large amount of information providers and consumers

## Web Now

- Page Amount: Over 20 billion pages
  - February 2006 (Yahoo!)
- Server Amount: Over 80 million servers (80,655,992)
  - April 2006 (Web Server Survey)

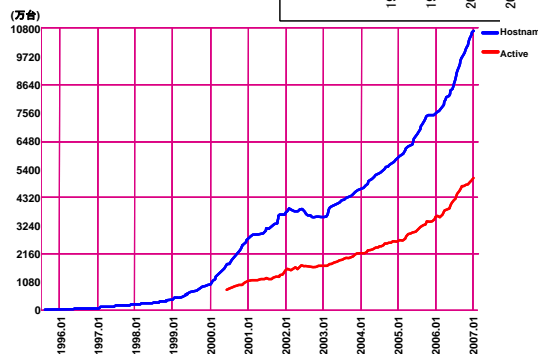
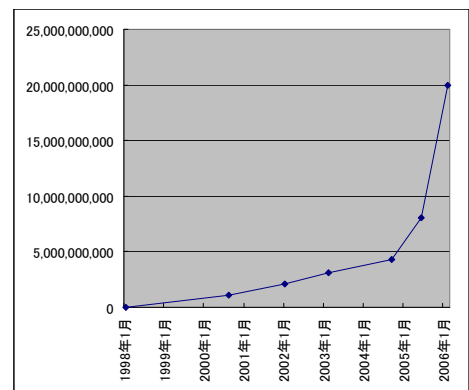


図1: Webサーバの台数の変化 (NetCraft, <http://www.netcraft.com>)

- Internet users: ca. 0.7 billion ( 14% of over 15 people)
  - March 2006
  - comScore <http://www.comscore.com/press/release.asp?press=849>

Top 15 Online Populations by Country, Among Visitors Age 15+* March 2006	
Unique Visitors (000)	
Source: comScore World Metrix	
	Unique Visitors (000)
<b>Worldwide Total</b>	<b>694,260</b>
United States	152,046
China	74,727
Japan	52,100
Germany	31,813
United Kingdom	30,190
South Korea	24,645
France	23,884
Canada	18,996
Italy	16,834
India	16,713
Brazil	13,186
Spain	12,452
Netherlands	10,969
Russia	10,833
Australia	9,735

## What is information distribution now?

- Focus on Information and Communication Activities (ICA) not on Information and Communication Technologies (ICT)

## From “old computing” to “new computing”

*“The old computing was about what computers could do; the new computing is about what users can do. Successful technologies are those that are in harmony with users’ needs. They must support relationships and activities that enrich the users’ experiences.”*

*Ben Shneiderman, Leonardo's Laptop: Human Needs and the New Computing Technologies, MIT Press, 2002*

- Paradigm shift is needed
  - *Technology-centered approach*
  - ***Human-centered approach***

# ART(Activity and Relationship Table)

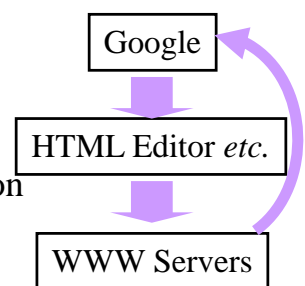
	<b>COLLECT</b> (Information)	<b>RELATE</b> (Communication)	<b>CREATE</b> (Innovation)	<b>DONATE</b> (Dissemination)
<b>Self</b>				
<b>Family and friends</b> (2-50 intimates)				
<b>Colleagues and neighbors</b> (50-5,000 regular encounters)				
<b>Citizens and markets</b> (5,000+)				

Ben Shneiderman, Leonardo's Laptop: Human Needs and the New Computing Technologies, MIT Press, 2002

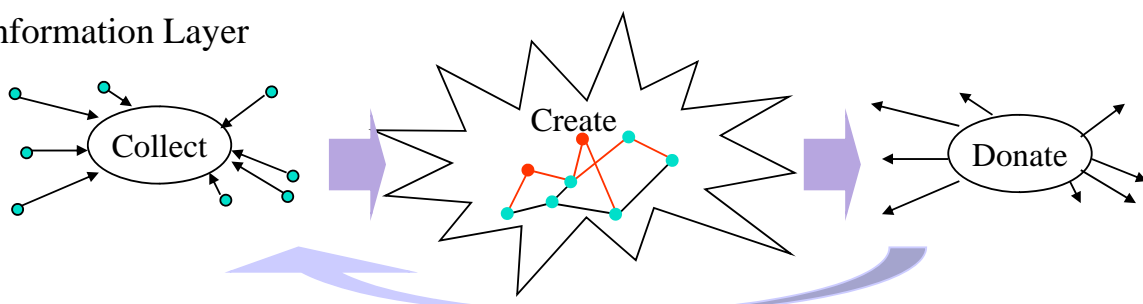
Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo} 

## Information Activities

- A cycle of information exploitation
  - **Collect**
    - ◆ Find and retrieve information
  - **Create**
    - ◆ Process (classify, extract, combine, mix, ...) information
    - ◆ Generate new information
  - **Donate**
    - ◆ Publish and distribute information



Information Layer



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo} 

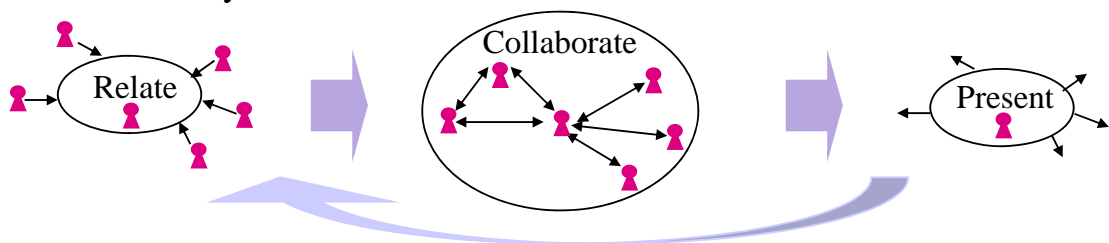
## Three reasons for WWW

- Freedom for joining and inter-linking
  - Same as Internet
- Economics
  - Same as Internet
- Easiness
  - Different from Internet

## Communication Activities

- A cycle of human relationship exploitation
  - **Relate**
    - ◆ Find and contact people
  - **Collaborate**
    - ◆ Work with other people (organized work, teamwork, cooperation, ...)
  - **Present**
    - ◆ Identify and contribute ourselves to communities

Communication Layer

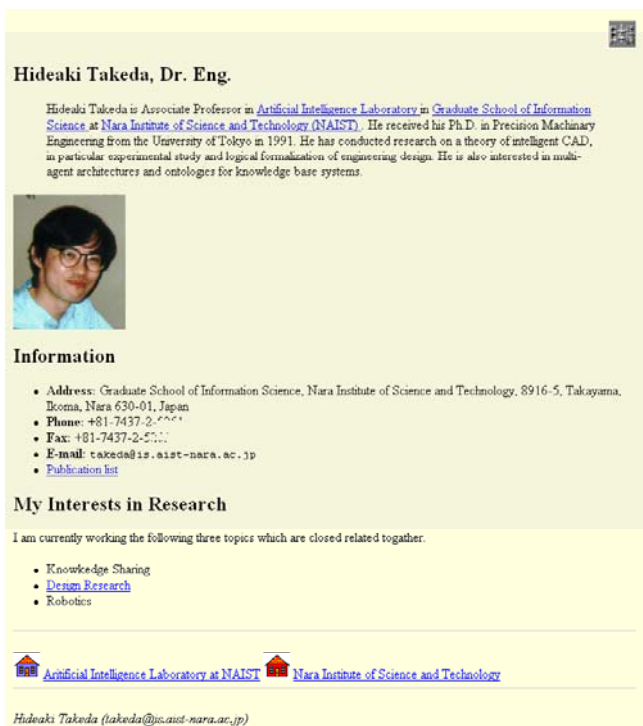


# Web as Communication Media

- WWW is mainly considered as media for information activities.
- But ...

# Web as Communication Media

- When WWW was created, many researchers created their “homepages”
  - “Homepage”
    - ◆ Data and information on research
    - ◆ Self-introduction
      - Profile
      - “What’s New”
      - Friends
  - Group Page
    - ◆ Data and information
    - ◆ Group members
    - ◆ “What’s New”
    - ◆ ...



The screenshot shows a personal homepage for Hideaki Takeda, Dr. Eng. The page has a light yellow background. At the top right is a small logo. Below it, the name "Hideaki Takeda, Dr. Eng." is displayed. A short biography follows, mentioning his position as Associate Professor in the Artificial Intelligence Laboratory at NAIST, his Ph.D. from the University of Tokyo in 1991, and his research interests in intelligent CAD, experimental study, and ontologies. A small portrait photo of him is shown. Below the photo is an "Information" section with contact details: address, phone, fax, and email. A "My Interests in Research" section lists three topics: Knowledge Sharing, Design Research, and Robotics. At the bottom, there are logos for the Artificial Intelligence Laboratory at NAIST and Nara Institute of Science and Technology, along with the email address takeda@jis.aist-nara.ac.jp.

# Web as Communication Media

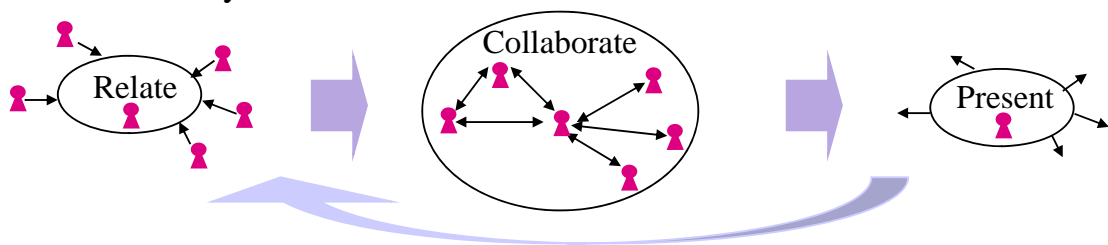
- When WWW was created, many researchers created their “homepages”
  - “Homepage”
    - ◆ Data and information on research
    - ◆ Self-introduction
      - Profile
      - “What’s New”
      - Friends
  - Group Page
    - ◆ Data and information
    - ◆ Group members
    - ◆ “What’s New”
    - ◆ ...

Just for Communication

# Communication Activities

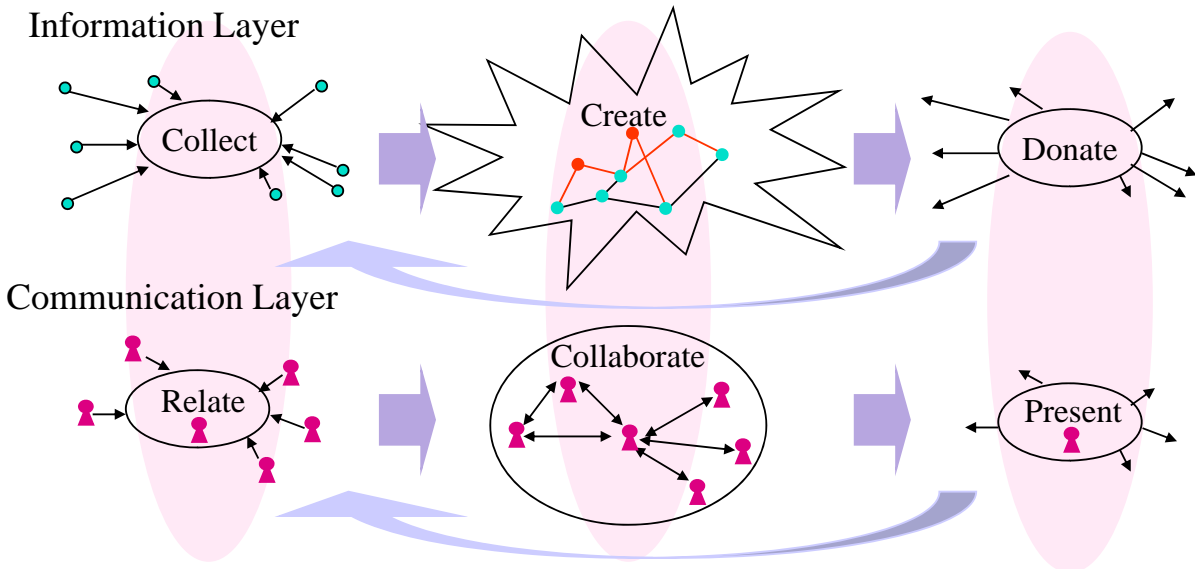
- A cycle of human relationship exploitation
  - *Relate*
    - ◆ Find and contact people
  - *Collaborate*
    - ◆ Work with other people (organized work, teamwork, cooperation, ...)
  - *Present*
    - ◆ Identify and contribute ourselves to communities

Communication Layer



# Information and Communication Activities

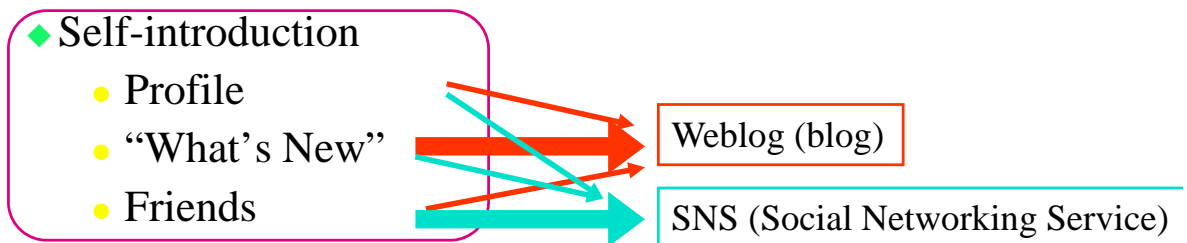
- Two layers for our activities
  - Information layer** only concerns explicitly represented and processed information.
  - Communication layer** concerns potential information. Potential information can be revealed through communication among people.



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}

# Web as Communication Media

- When WWW was created, many researchers created their “homepages”
  - “Homepage”
    - Data and information on research



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



## Direction for ICA

- More focus on the communication layer
  - Social network
    - ◆ Social Network Analysis
    - ◆ Network Science
    - ◆ Social Network Mining
    - ◆ Social Network-based Services
  - Communities

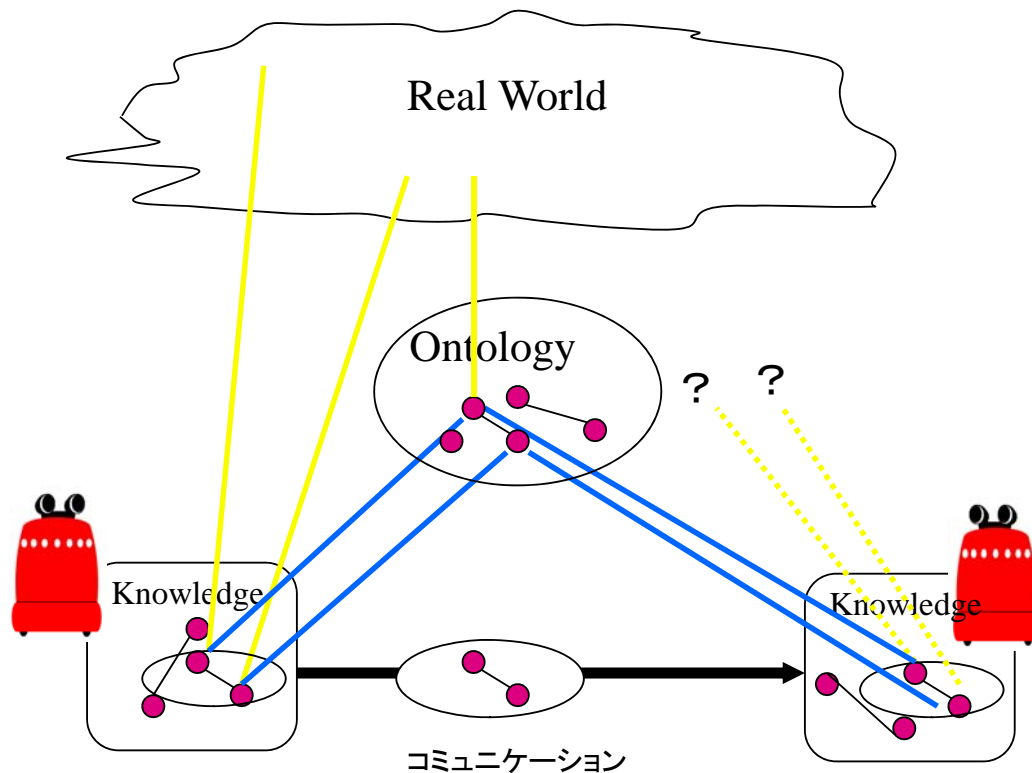
## Shifting from information to knowledge

- Ontology for knowledge sharing

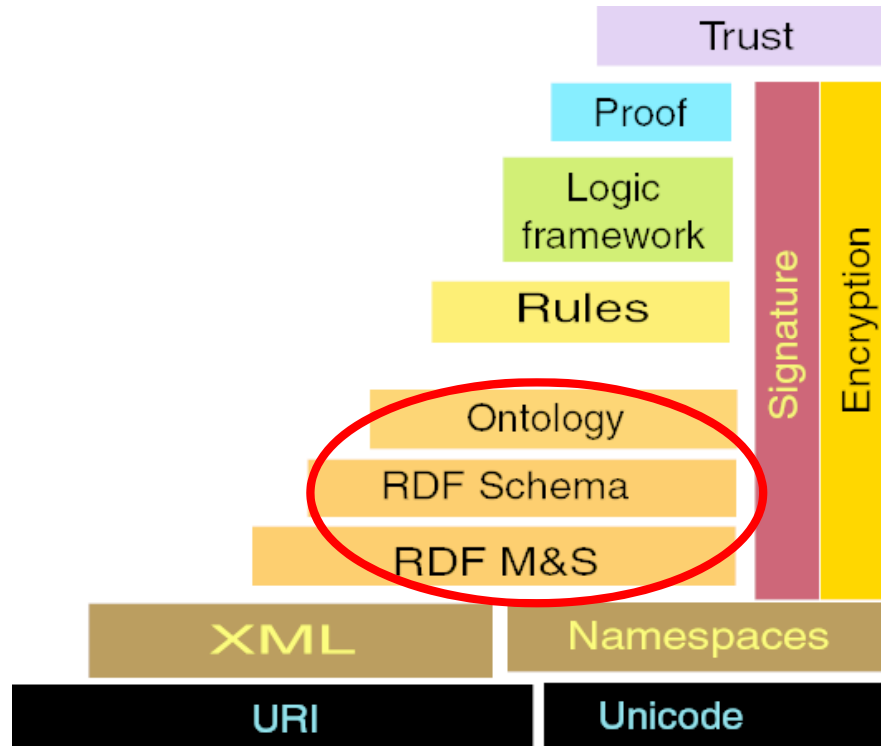
## Why knowledge sharing is needed?

- WWW explicates needs for common background knowledge among people
  - In fixed information distribution, all people are expected to share background knowledge
  - In flexible information distribution such as WWW, it is not guaranteed.
- Ontology:  
*An ontology is an explicit specification of a **conceptualization***  
[Gruber]

## Ontology and Agent Communication



# Architecture for the Semantic Web



Tim Berners-Lee <http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo} 

## The Aim of The Semantic Web

- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

*[The Semantic Web](#), Scientific American, May 2001, Tim Berners-Lee, James Hendler and Ora Lassila*

- The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for **automation**, **integration** and **reuse of data across various applications**.

<http://www.w3.org/2001/sw/>

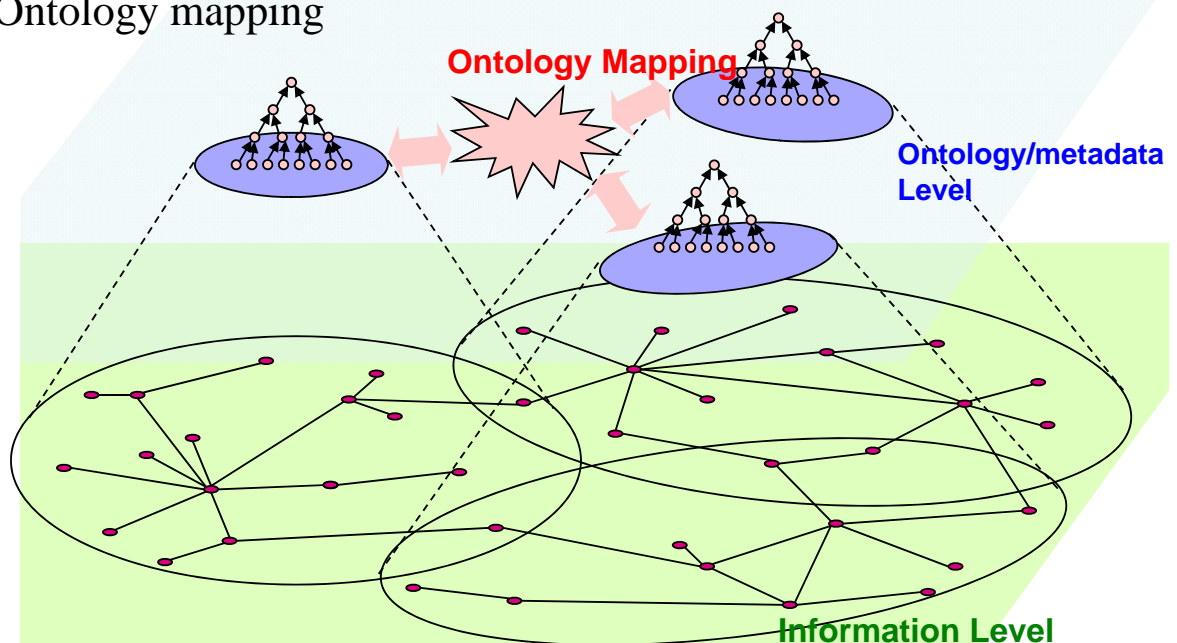
Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo} 

# Challenges for Realizing Semantic Web

- Representation of Ontology-based Web
  - OWL (Web Ontology Language)
- Definition of Ontologies
  - Top-level ontology
  - Domain ontologies
- Manipulation of multiple ontologies
  - Even a single target can be represented in different ontologies
  - Communities often develop own ontologies
  - Mapping or mediating among different ontologies are needed

## Community-based information distribution and sharing

- Information distribution within communities
- Information distribution cross communities
  - Ontology mapping



# Next Information distribution system: Summary

- Information and communication activities
  - vs. Information and communication technologies
- Communication level: Social network and community
  - vs. information level
- Ontology level
  - vs. Information level
- Distributed ontologies
  - vs. a unified ontology