Semantic Computing at AIST

2006-05-31
HASIDA Koiti
ITRI, AIST
Semantic Gap

- People and computers don't share meaning and value.
  - We don't understand computers.
  - Computers don't understand us.
- So they cannot collaborate well.
We Don't Understand Computers. (Computers Don't Understand Themselves, either.)

- I installed Service Pack 2 into my PC running Windows XP. Since then I cannot connect to wireless LAN. Why?
- I cannot remove a strange line in MS Word.
- We cannot coordinate workflow systems with each other in our intranet.
Computers Don't Understand Us.

- I cannot find the information I want. The search engine returns a lot of irrelevant information and little relevant information.
- Web sites are very hard to keep easy to use.
- Performance improved by banning intra-corporate e-mails.
Semantic Computing

- **Glassbox Computer**
  - design and operation of computer systems through semantics shared with people
  - semantic model of data and process

- **Straightforward provision of services meaningful to people**

- **People can understand, compose, and improve software.**

- **Emergent total optimization by accumulation of improvements by many users**
Ontology of Documents
Ontology of Patent Claim

The `claim' class subsumes the `Jepson-type claim' class.

Each `claim' class instance has one or more `constituent' properties with `technology' class instances as values.

The `claim' class instance has one or more `constituent' properties with `technology' class instances as values.
Presupposes Jepson-type claim 0

**Mass Spectroscope (0)**

- **Constituent**: Ion Source (1)
- **Constituent**: Mass Analyzer (2)
- **Constituent**: Ion-Electron Converter (4)
- **Constituent**: Electron Detector (3)
- **Constituent**: Voltage Controller (12)
- **Constituent**: Slit (10)

**Extract Ion A from (1)**

- **Purpose**: (2) Separates A
- **Purpose**: (2) Extracts B
- **Purpose**: (4) Converts B to Electron C
- **Purpose**: (3) Detects C and Extracts as Electric Signal

**Constraints**

- **Vs** = **V0** - **k1**
- **Vc** = **V0** - **k2**

**V0** = Ion-Extraction Voltage on (1)

**Vs** = Voltage on (10)

**Vc** = Converter Voltage on (4)

**k1** and **k2** are Constants
displaying, on a display unit, a list of labels $L$ in which are present a node $z \in F(x)$ and a node $y$ of which a link $y-z$ is contained in the database $D$ and of which the label $y$ is $L$, for each of the nodes $x$ of a search question $Q$
Clarified Semantic Structure

each node $x$ in retrieval query $Q$

display the list of $L$ on the display unit

quantify

intension

$z \in F(x)$.

Database $D$ contains link $y$-$z$. The label of $y$ is $L$.

内包

z∈F(x)。

データベース$D$がリンク$y$-$z$を含む。$y$のラベルが$L$である。
Semantic Authoring

S. Handschuh (Karlsruhe U.)
Coarse-Grain Graphical Content

- Result of semantic authoring
- Easy for people to understand and compose
  - explicit logical structure
  - no intersentential order

I was hungry. → cause → I had a snack. → cause → I had had a lunch. → cause → I became full. → concession
Fine-Grain Graphical Content

- automatic analysis of coarse-grain graphical content
- retrieval, translation, summarization, etc.
- too minute for people to handle
Semantic Authoring is Easier than Text Composition (1/2)

I had a snack.

cause

I had a lunch.

cause

I was hungry.

cause

I became full.
Semantic Authoring is Easier than Text Composition (2/2)

- A text synonymous with the graph in the previous page:
  
  I had had a lunch. But I was hungry, and so I had a snack. Then I became full.

- This relation is hard to reflect in the text.

  I had had a lunch but I was hungry. So I had a snack. Then I became full.
Traditional Idea Processor

- No standardized relations
  - Only the author or participants of brain storming can understand.
  - hard to share and reuse
- Cost of text composition
  - big apparent cost  — limited spread

Semantic Authoring

- Standardization of relations
  - ISO/TC37/SC4/TDG3
  - retrieval, summarization, translation, etc.
- Automatic text generation
  - small cost  — wide spread
Discussion-Supporting Groupware

How to eliminate illegal bike-parking?

Prepare more bike-parking lots.

Remove illegally-parked bikes immediately.

We have to keep them for six months.

We don't have enough space to keep them.

That is not profitable.

That is not profitable.

Prepare more bike-parking lots.

Remove illegally-parked bikes immediately.

We have to keep them for six months.

We don't have enough space to keep them.
Collaborative Semantic Authoring

- Traditional Groupware
  - IBIS, Coordinator, Open Meeting, etc.
  - improved efficiency and quality of discussion
    - reduced redundancy
    - simultaneous utterances
    - better coverage of important points
    - deeper discussion
  - weakness • • • usable only for group work

- Collaborative SA
  - seamless unification of individual SA as a major usual task and group work
  - the above merits
    + advanced retrieval, summarization, etc.
From e-Mails to Collaborative SA

- **Perspicuous semantic structure** develops as communication proceeds.
- No spams.
- TODOs
  - coordination with DB
  - access control of users/groups
  - notification (like RSS feed)
ISO/TC37/SC4/TDG3
Semantic Content Representation
ISO/TC37

Terminology and Other Language Resources

- SC1: Principles and Methods
- SC2: Terminography and Lexicography
- SC3: Computer Applications for Terminology
  - ISO12620: Data Categories
- SC4: Language Resources Management
ISO/TC37/SC4

Language Resources Management

- Chair: Laurent Romary
- Secretariat: Key-Sun Choi

- WG1: Basic descriptors and mechanisms for language resources (Laurent Romary)
- WG2: Representation schemes (Kiyong Lee)
  - Multimodal meaning representation scheme
- WG3: Multilingual text representation
- WG4: Lexical resources/database (Nicoletta Calzolari)
- WG5: Workflow of LR management
ISO/TC37/SC4/Ad Hoc TDGs

- **TDG1:** Metadata (Peter Wittenburg)
- **TDG2:** Morphosyntax (Gil Francopoulo)
- **TDG3:** Semantic Content Representation (Koiti Hasida)
  - Discourse relations (Koiti Hasida)
  - Dialogue acts (Harry Bunt)
  - Referential structures and links (Laurent Romary)
  - Logico-semantic relations (Scott Farrar)
  - Temporal entities and relations (Kiyong Lee)
  - Semantic roles and argument structure (Thierry Declerck)
  - More?
Expected Products

- Not ISs (International Standards) in ISO’s official sense
- But Standard Registries of Data Categories
  - discourse relations, dialogue acts, etc.
Scope of TDG3

- Semantics, Abstracting Syntax Away
  - Semantic DCs usable with various annotation schemes
    - We’re not writing annotation manuals.
  - We don’t care syntax-semantics mapping, syntactic markup and markables, etc.

- Deliverables
  - Concrete Data Category Registries
    - semantic types of function words/morphemes and their taxonomy
      - not full dictionaries or encyclopedias
  - Documents on These DCs
Back on Semantic Computing
Upgrading Semantic Levels

- window system
- operating system
- file system
- semantic authoring
- semantic platform
- RDF database
Constraint-Based Collaboration
Constraint for Collaboration

Ms. A

Mr. B

author

boss

author

docA

derive

docB

derive

docC

part

part

part

paraA1

derive

paraB1

derive

paraC1

author

author
Constraint-Respecting Structure
Workflow Constraint

- **X should be older than Y.**
  - Y’s author is notified of X’s revision.

- **Complex interaction**
  - lots of flows back and forth
OSS: Open-Semantics Software

- Designed in User’s Semantics
  - not just open-source!
- Constraint Programming
  - abstraction/simplification away from process/interaction
  - ultimate data-orientation
- Ontology
  - vocabulary for user’s meaning and value
Semantic Enterprise System

Design and Operation by Business Semantics

- Incremental and emergent total optimization
  (a la Enterprise Architecture)
  - accumulation of improvements by users
  - integration of business operation, regulation, and computer system
- Knowledge circulation in and across organizations
  - AIST-SOA
Knowledge-Circulating Society
Knowledge Circulation

- social sharing, reuse, and extended reproduction of knowledge
- participation of everybody in every situation

general public

users
- producers
- consumers
- mediators

service

shared DB

document

provision of knowledge

acquisition of knowledge
Knowledge Circulation in Research (So Far)

- Knowledge-Circulation period > 2 years
- Papers are hard to read/write.
New Mode of Research

- Collaborative creation of huge graphical content
- Publication of sentences, not papers
- Fast knowledge circulation
  - In a week?
- Evaluation better than IF and CI
- Network analysis
- Visualization
- Retrieval, translation, summarization
e-Knowledge Government

- Limitation of representative democracy
  - increasing diversity and complexity of political issues

- Involvement of all the citizens
  - collection and analysis of public opinions and knowledge
  - policy making and consensus building

- Given effective discussion by all the people:
  - no need for representative/indirect democracy
  - compositional democracy ・・・ KAWAKITA Jiro
  - deliberative democracy

- IT-based support
  - retrieval, summarization, translation, etc.
  - Weblog not sufficient
    - no systematic support to formation of long inference chains