

## “Know Who” System Using iMage (Information Mixable Graph Explorer)

*Kenji Otomo<sup>†</sup>, Takahiko Murayama, Masami Seno,  
and Yuichi Kokubo*

### Abstract

Our next-generation Web information sharing engine called iMage (information mixable graph explorer) supports the discovery of new knowledge or information by integrating and analyzing resource description framework (RDF) data extracted from various data sources. We applied it to the “Know Who” system in a joint experiment and got excellent results.

### 1. Expanding the Semantic Web

The Semantic Web, which is advocated by Tim Berners-Lee, is a project that aims to improve the convenience of the World Wide Web (WWW) by using metadata (e.g., creator, date, and keywords) or meaning given to web pages. At first, people were skeptical about the Semantic Web becoming popular because of its constraint—somebody must add metadata to web pages. However, recently a lot of metadata has been added automatically to contents in social networking services (SNSs) or blog (weblog) services, and users are starting to add tags to their contents. Therefore, the basis for the spread of Semantic Web is ready now. For example, automatic article distribution by RSS (rich site summary) is a kind of Semantic Web service. We are developing iMage (information mixable graph explorer) [1], which extracts and discovers new knowledge or information by using metadata used in the Semantic Web.

### 2. iMage

The resource description framework (RDF) [2] is the W3C (World Wide Web Consortium) standard for

encoding metadata on the Semantic Web. RDF specifies many rules for adding metadata and one of its important feature is that data can be expressed in the form of a graph if it is written in RDF. The fundamental unit of an RDF description is called a *triple*. For example, a sentence “the author of the paper is X” can be expressed as a graph with two nodes, “X” and “paper”, connected by an arc labeled “author”. Connecting all these descriptions, metadata can be expressed as a large graph. We carefully examine the structure of this graph and make the assumption that distinctive patterns within this large graph are important and informative.

Although there have been many trials on retrieving informative knowledge or information by processing and analyzing content (information) itself using natural language processing, we think that we can retrieve informative knowledge just by focusing on distinctive patterns in a graph structure without such highly complex processing.

A technological overview of iMage is shown in **Fig. 1**. In this example, metadata of two resources, a database of researchers and data from an SNS, is expressed as a large graph. From this graph, we can discover a skilled person in a particular technology or an active person in an SNS about a particular technology.

<sup>†</sup> NTT Information Sharing Platform Laboratories  
Musashino-shi, 180-8585 Japan  
Contact: csm@lab.ntt.co.jp

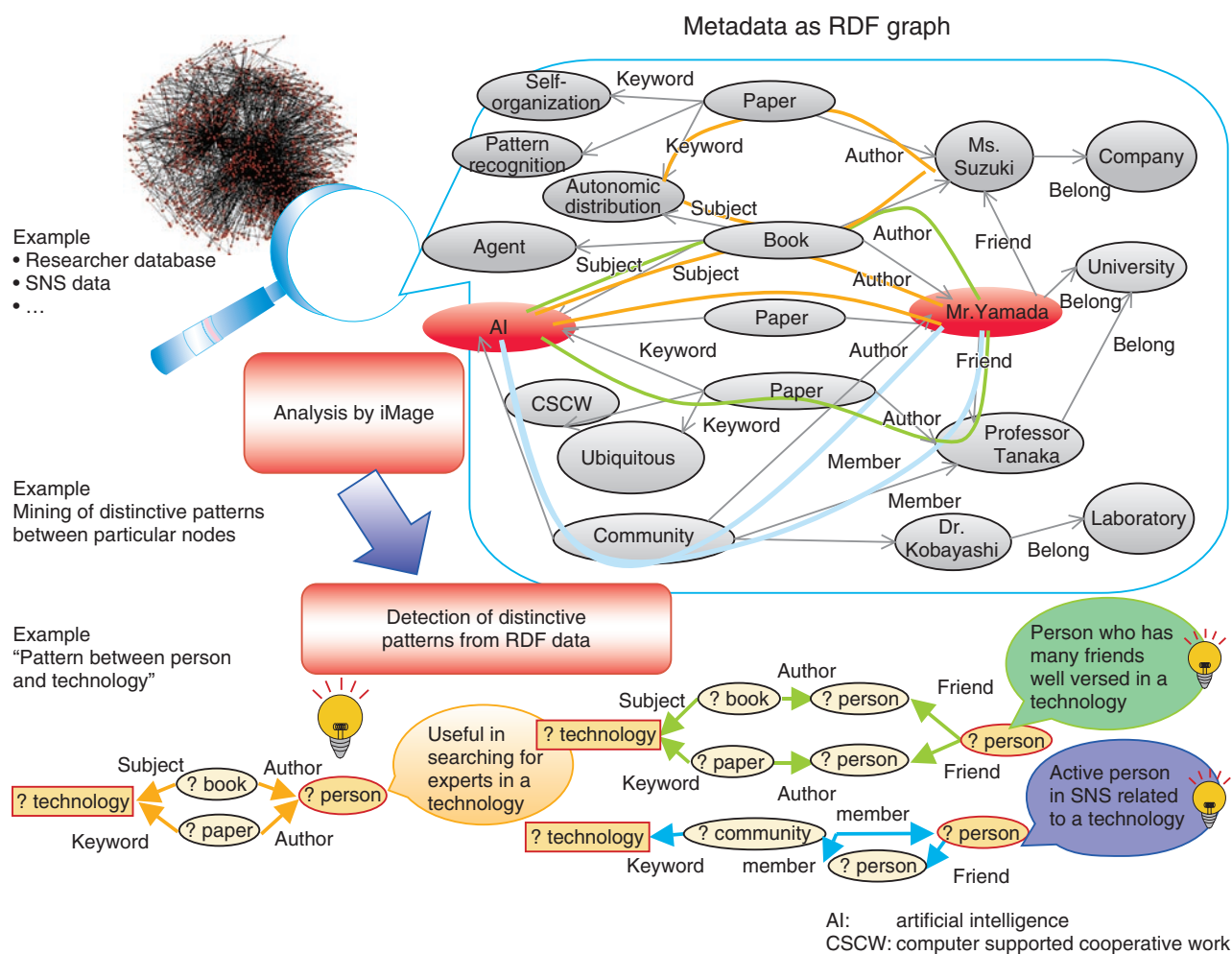


Fig. 1. Overview of iMage.

### 3. Application to the Know Who system

Because iMage can discover important relationships about people, we think that it is applicable to the Know Who system. By using the data integration functions of iMage, we can utilize a company's existing data for new applications such as the Know Who system.

#### 3.1 Joint experiment on Know Who system for sales persons

In a joint experiment between NTT Information Sharing Platform Laboratories and NTT Software, we developed a Know Who system for sales persons by using data in NTT Software and verified its usefulness by having sales persons use the service in practice.

#### 3.1.1 Need for Know Who system for sales persons

In a company that performs system integration, such as NTT Software, there are many cases where sales persons want to know whether there are any experts on a certain technology in their company. Since many sales persons are not well informed about such matters, we developed a system called the "Technology-case matching service" that can assist in finding appropriate persons or divisions to contact. This service enables sales persons to find the right persons quickly and prevent a mismatch between a project and the people assigned to it.

#### 3.1.2 Data sources

We use two main kinds of data for our Know Who system.

- Project planning documents
  - Software utility information
- NTT software handles all projects in the form of

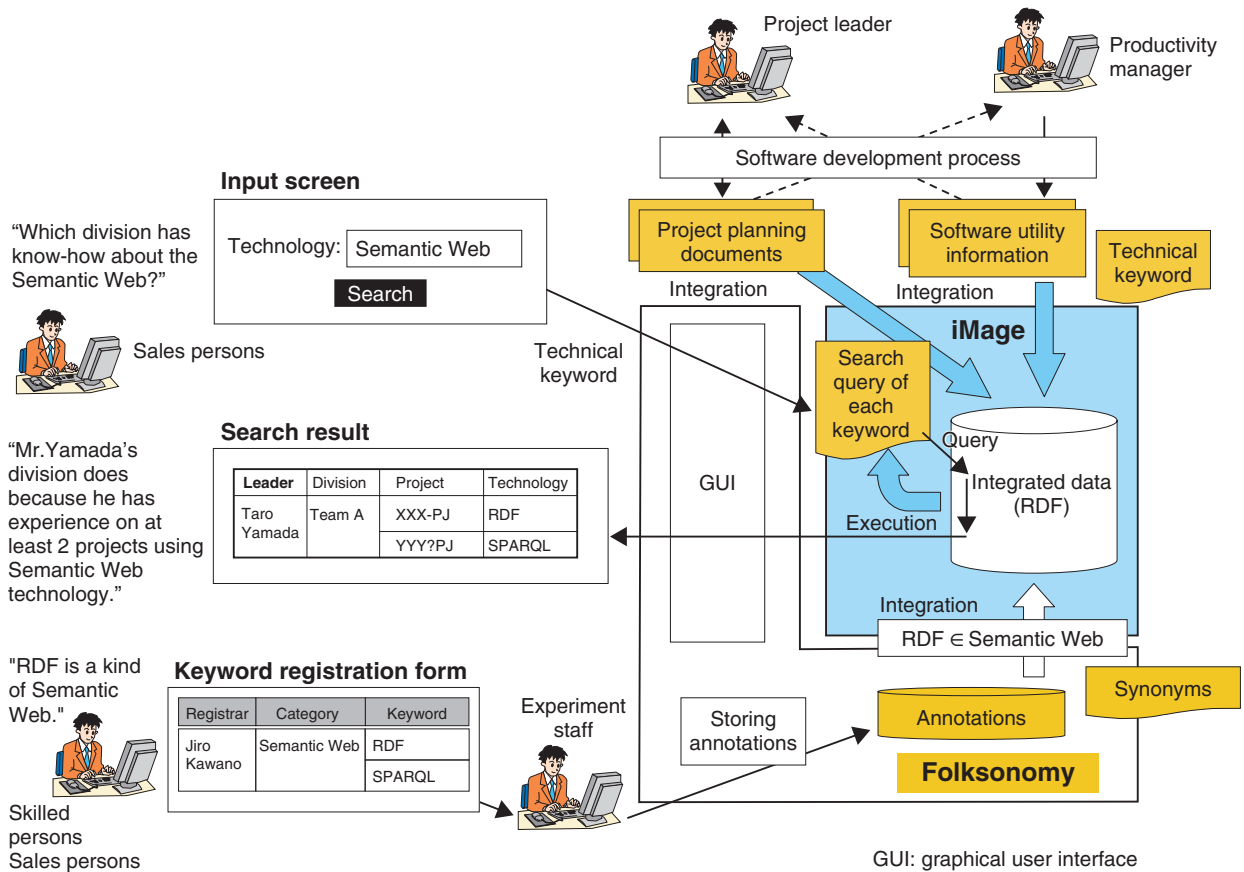


Fig. 2. Structure of human resources matching system.

project planning documents. A project planning document gives the type of project, the name of the division in charge, the name of the project leader, the development period, the process of development, the project members and their skills, and so on, so it is suitable for the Know Who system, which can search for persons or divisions by using technical keywords. In this experiment, we used about 8000 items of data over the past four years related to about 1000 persons. In our Know Who system, the selection of search keywords is very important. Technical keywords are extracted by the system administrators and the domain experts, and we also use software utility information to extract technical keywords. Software utility information is expected to include important and proven product names for system integrators.

**3.1.3 System architecture**

The architecture of our Know Who system is shown in Fig. 2. iMage analyzes the structure of RDF data and extracts distinctive patterns, which are used as queries when a search is executed. To enable the Know Who system to be operated by users, we had to

develop the graphical user interface shown in Fig. 2 because iMage provides only the libraries used in searches.

**3.1.4 Distinctive patterns**

First, iMage analyzes the structure of RDF data expressed as a graph. Then, it extracts distinctive patterns from it. Finally, it retrieves knowledge or information by using the patterns as queries. Some examples of distinctive patterns automatically extracted by iMage and then selected by system administrators are shown in Fig. 3. Ordinary search engines only list persons that match the given technical keyword. However, iMage also categorizes them persons based on distinctive patterns. For example, the persons in this result are categorized as a supervisor (pattern 1), project leader (pattern 2) or technical staff member (pattern 3) of a project related to the keyword, as shown in Fig. 3.

**3.1.5 Innovations to improve effectiveness**

We can measure the effectiveness of the Know Who system by checking whether all of the experts for the given technical keyword are in the search result. To

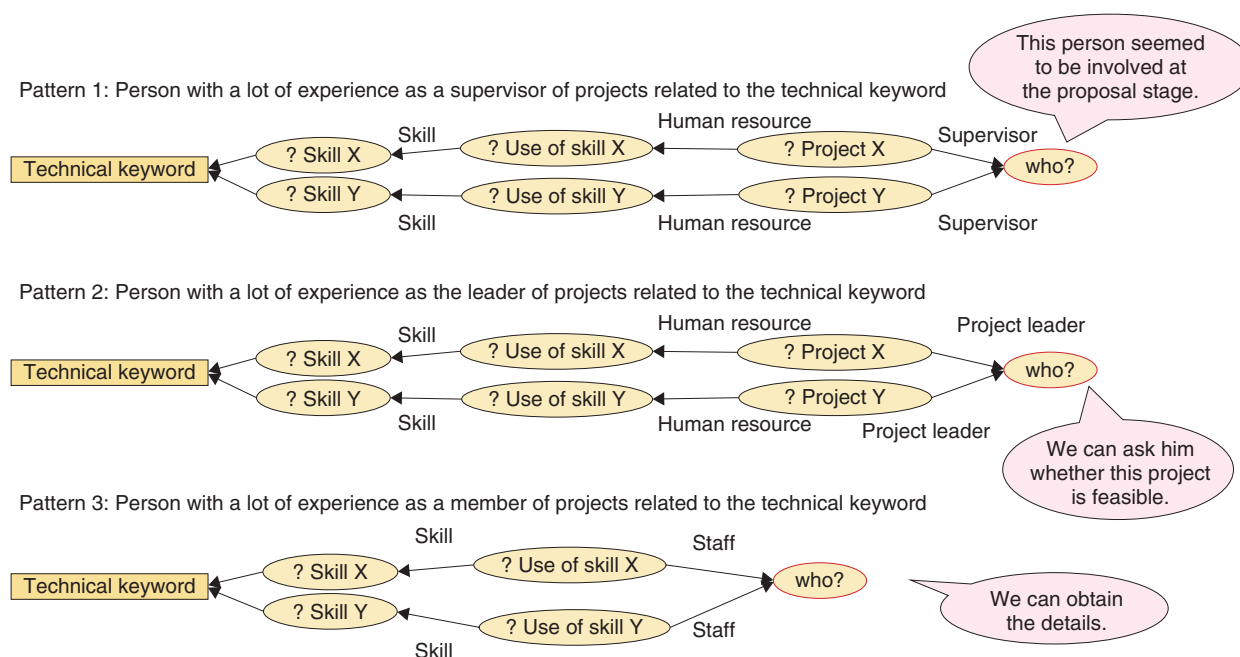


Fig. 3. Distinctive patterns used in Know Who system.

improve the effectiveness, we introduced two innovations to the system.

- Support for synonyms by using folksonomy [3]
- Support for different spellings

Synonym support using folksonomy means that when the given keyword is too specific to find experts in the given field, the search field is expanded by using synonyms on a higher level of abstraction. In the keyword registration form in Fig. 2, RDF and SPARQL were chosen as synonyms of Semantic Web. Therefore, when no one was retrieved using the keyword SPARQL, the system also tried searching with the keyword RDF. We tried to improve the system's effectiveness by providing a function that lets users add synonyms by themselves, just like the mechanism of folksonomy.

The spelling variant support treats alternative spellings of a word as equivalent including alphanumeric words and Japanese hiragana, katakana, and kanji. To handle this, we prepared a dictionary of representative words. We solved the problem by normalizing the words in RDF data and the search keywords to the words in the dictionary.

### 3.2 System evaluation

We evaluated the system by conducting a survey in the form of a questionnaire. The profiles of the subjects are shown in Fig. 4. We asked the subjects to use

the system and then checked the results by asking the system administrators. The results of the questionnaire are shown in Fig. 5. The search with the underlined bold-faced keywords delivered the most effective result because all of the persons in the result were experts including some newly discovered ones who were previously unknown to the subject.

We can conclude that the system is effective at supporting the jobs of system integrators for the following reasons: In 80% of the search results, experts were found. Moreover, many underlined bold-faced keywords were input by users D and E, who are sales persons, i.e., our target system users, who are not familiar with technology.

### 4. Conclusion and future plans

Through a joint experiment, we confirmed that our iMAGE technology is applicable to and effective for the field of the Know Who system. We also found that we need to enhance the scalability of the system to handle larger volumes of data than those used in this experiment, for example, data covering a longer term or data for large-scale enterprises. While developing an efficient data collection mechanism and improving the search accuracy, we are moving forward toward commercialization of the system through implementations in various domains.

<p><b>Mr. A (Senior consultant)</b></p> <ul style="list-style-type: none"> <li>- NTT Group career: about 15 years</li> <li>- NTT Software career: about 15 years</li> <li>- Sales career: about 1 year</li> <li>- Usual searching method: Talking to sales person for relevant solution, talking to project leaders or sales staff, information network of a supervisor, human network</li> </ul>	<p><b>Ms. B (Senior engineer)</b></p> <ul style="list-style-type: none"> <li>- NTT Group career: about 7 years</li> <li>- NTT Software career: about 7 years</li> <li>- Sales career: about 6 years</li> <li>- Usual searching method: Talking to familiar experts. Browsing organization</li> </ul>	<p><b>Ms. C (Senior engineer)</b></p> <ul style="list-style-type: none"> <li>- NTT Group career: about 26 years</li> <li>- NTT Software career: about 4 years</li> <li>- Sales career: about 3 years</li> <li>- Usual searching method: Human network</li> </ul>
<p><b>Mr. D (Former GM of sales)</b></p> <ul style="list-style-type: none"> <li>- NTT Group career: about 34 years</li> <li>- NTT Software career: about 20 years</li> <li>- Sales career: about 5 years (sales support)</li> <li>- Usual method for search: Human network</li> </ul>	<p><b>Mr. E (GM of sales)</b></p> <ul style="list-style-type: none"> <li>- NTT Group career: about 30 years</li> <li>- NTT Software career: about 15 years</li> <li>- Sales career: about 3 years</li> <li>- Usual searching method: Records of meetings</li> </ul>	

GM: general manager

Fig. 4. Profiles of test subjects.

<p><b>Mr. A (Senior consultant)</b></p> <ul style="list-style-type: none"> <li>- <u>Solution 1</u></li> <li>- <u>Solution 2</u></li> <li>- <b><i>Solution 3</i></b></li> <li>- <u>Solution 4</u></li> <li>- <b><i>Solution 5</i></b></li> <li>- Solution 6</li> <li>- Solution 7</li> <li>- <b><i>Task name 1</i></b></li> <li>- <b><i>Product name 1</i></b></li> </ul>	<p><b>Ms. B (Senior engineer)</b></p> <ul style="list-style-type: none"> <li>- <u>Solution 8</u></li> <li>- <u>Product name 2</u></li> <li>- <i>Product name 3</i></li> <li>- <u>Product name 4</u></li> <li>- <i>Task name 2</i></li> <li>- <b><i>Elemental technology 1</i></b></li> <li>- <u>Elemental technology 2</u></li> </ul>	<p><b>Ms. C (Senior engineer)</b></p> <ul style="list-style-type: none"> <li>- <b><i>Elemental technology 3</i></b></li> <li>- <i>Solution 9</i></li> <li>- <b><i>Product name 5</i></b></li> <li>- <b><i>Elemental technology 4</i></b></li> <li>- <b><i>Elemental technology 5</i></b></li> <li>- <b><i>Task name 3</i></b></li> <li>- <b><i>Elemental technology 6</i></b></li> </ul>
<p><b>Mr. D (Former sales chief)</b></p> <ul style="list-style-type: none"> <li>- <b><i>Component technology 7</i></b></li> <li>- <u>Product name 6</u></li> <li>- <b><i>Task name 4</i></b></li> <li>- <b><i>Elemental technology 8</i></b></li> <li>- <b><i>Elemental technology 9</i></b></li> <li>- <b><i>Task name 5</i></b></li> </ul>	<p><b>Mr. E (Sales chief)</b></p> <ul style="list-style-type: none"> <li>- <b><i>Task name 6</i></b></li> <li>- <b><i>Task name 7</i></b></li> <li>- <b><i>Task name 8</i></b></li> <li>- <u>Product name 8</u></li> <li>- <u>Solution 10</u></li> </ul>	

Keywords are anonymized with a sequential number for each category: solution, component technology, product name, task name.

- (A) **Underlined bold-faced keywords:** Only experts were returned and some of them were previously unknown to the subject.
- (B) **Italic bold-faced keywords:** Both experts and non-experts were returned and some of the experts were previously unknown to the subject.
- (C) **Underlined keywords:** Only experts were detected and all of them were known to the subject.
- (D) **Italic keywords:** Both experts and non-experts were returned and all of experts were previously known to the subject.
- (E) **Red keywords:** No experts were detected.
- (F) **Gray shaded keywords:** We could not categorize them because there were too many results, etc.

Fig. 5. Results of questionnaire.

---

## References

- [1] H. Sato, K. Iiduka, I. Pramudiono, K. Otomo, and T. Murayama, "iMage: Serendipities retrieved from re-construction of information," Semantic Web Conference 2007, Tokyo, Japan (in Japanese).
- [2] <http://www.w3.org/TR/rdf-concepts/>
- [3] <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>



### Kenji Otomo

Research Engineer, Software Architecture Project, NTT Information Sharing Platform Laboratories.

He received the B.S. and M.S. degrees in computer science from Keio University, Kanagawa, in 1988 and 1990, respectively. He joined NTT Software Laboratories in 1990. His research interests include Web services and the Semantic Web. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and the Information Processing Society of Japan



### Masami Seno

Senior Engineer, Technology and Product Department, Strategic Technology Center, NTT Software.

He received the B.S. and M.S. degrees in mathematics from Waseda University, Tokyo, in 1984 and 1986, respectively. He joined NTT Software in 1987. He has been engaged in the development of new products using Web services or the Semantic Web.



### Takahiko Murayama

Senior Research Engineer, Supervisor, Software Architecture Project, NTT Information Sharing Platform Laboratories.

He received the B.E. degree in communication engineering and the M.E. degree in information engineering from Tohoku University, Miyagi, in 1984 and 1986, respectively. He joined NTT Communication and Information Processing Laboratories in 1986. His research interests include Web services and the Semantic Web. Currently, he is also a visiting associate professor at the Graduate School of Information Systems at the University of Electro-Communications. He is a member of IEICE.



### Yuichi Kokubo

Chief Engineer, Technology and Product Department, Strategic Technology Center, NTT Software.

He received the B.S. degree in mathematics from Nihon University, Tokyo, in 1997. He joined NTT Software in 1997. He has been engaged in the development of new products using Web services or the Semantic Web.

---