# **Special Feature: NGN Focus**

# Information Structuring Technology for Information Processing in the Ubiquitous Society

# Ryoichi Hayashi<sup>†</sup>, Masaichi Hasegawa, Shinji Morishita, Kengo Fujioka, Mitsuaki Tsunakawa, and Seiji Susaki

# Abstract

This article describes information structuring technology, which collects various types of user-originated information being distributed and stored on Internet services, indexes this massive amount of diverse information on the basis of individual behavior and history, and presents it according to a user's conditions with the aim of revealing new facts and promoting insights.

### 1. Introduction

The ever-expanding ubiquitous society has made it easier than ever for users to post photographs and comments on everyday life and events such as trips, athletics events, and shopping experiences on various kinds of Internet services without much conscious effort. Although some pieces of information may be interrelated, since such relationships are not stored anywhere, a user is forced to switch back and forth between services and rely on memory to find such associations. While conventional keyword-based search techniques present formulistic search results for specified keywords, to obtain pertinent results one must provide accurate search conditions. This is not necessarily easy; moreover, there are few means of finding items of value or interest among those search results.

In response to these problems, we are researching and developing information structuring technology for collecting information, assigning relationships between different pieces of information, and presenting those relationships according to certain conditions. This technology consists of an *information indexer*, which organizes information based on time,

† NTT Cyber Space Laboratories

Yokosuka-shi, 239-0847 Japan

place, and human relationships, and *base technology for social graphs*, which assigns relationships to information specific to human relationships.

# 2. Information structuring technology

# 2.1 Outline

Information structuring technology consists of three key processes: information collection, indexing, and presentation. The typical flow of actions is as follows (**Fig. 1**).

- (1) Various users post information (e.g., photographs and comments) about events and other day-to-day matters on various services, which store it.
- (2) The posted information is collected by a Web crawler or other means.
- (3) The information is indexed from the viewpoints of time, place, people, etc. These viewpoints are set in accordance with items (times, places, human relationships, etc.) that users tend to prefer when browsing through information.
- (4) Information is extracted, sorted, and displayed according to conditions and viewpoints specified by the user.
- (5) A user examines the displayed information, focuses on information that appears desirable or important, and switches viewpoint to extract and



Fig. 1. Concept of service based on information structuring technology.

sort new information.

(6) The user examines the newly displayed information and repeats (5) to get closer to the target or interesting information. This process is defined as *tracing*.

Information structuring technology lets users trace highly associative information that has associations between different pieces of information corresponding to certain conditions or to associations based on human relationships. It presents the user with highly reliable and relevant information while promoting the discovery of new information and the generation of insights.

#### 2.2 Information indexer

The information indexer does not limit itself to feature attributes belonging to information that is expected to be targeted. It also uses relationships between pieces of information, such as information about a traveling companion extracted from the user's personal history, so that information can be traced. This technology assigns relationships to information in a hierarchical manner based on subjective memory (e.g., information posted in schedulers and miniblogs (web logs)) and objective facts (information such as visited locations and companions obtained by analyzing data from a GPS (global positioning system) device, RF-ID (radio-frequency identification) tags, and other types of sensors). Indexing information in this way makes it possible to trace connected information systematically using key conditions such as place, time, people, and names.

When information is associated and traced in this way, the user can obtain information without the usual heavy reliance on memory. For example, consider the question: "What is the name of the Korean barbecue restaurant where Mr. Uesugi and I ate dinner during our trip to Yonezawa last month?" In this case, if a conventional search technique using the keywords "Yonezawa" and "Korean barbecue" were used, the user would have to search for relevant information from the huge number of search results returned, which would take a lot of effort. By contrast, the information indexer makes use of behaviorrelated facts included in the original question such as last month, Mr. Uesugi, trip, Yonezawa, and Korean



Fig. 2. Outline of base technology for social graphs.

barbecue. Thus, by tracing information indexed on the basis of such facts, the user can get the restaurant's name more efficiently.

### 2.3 Base technology for social graphs

As depicted in **Fig. 2**, base technology for social graphs can dynamically display a person's relationships with his or her friends in relation to various keywords (topics) in the form of a graph. This structure can also reveal the primary person (authority) for a certain topic on the periphery of the central character and the strength of that person's human relationships.

For example, the strengths of Mr. Kimura's relationships in Fig. 2 can be depicted as follows:

Kato, Yamada > Tanaka > Suzuki, Sato

For example, it turns out that Mr. Kato and Mr. Yamada have a common friend in Mr. Suzuki, which makes their friendship even stronger. Ms. Sato, meanwhile, has blocked Mr. Kato from accessing her, and since Mr. Suzuki is not a direct friend of Ms. Sato, their friendship is weak. Evaluating relationship strengths in this way and combining the results related to a certain topic lets one obtain the human relationships centered on Mr. Kimura in relation to various topics in the form of a graph (social graph).

Such social graphs can be used to trace information from human relationships and obtain more reliable information. They can help a user discover an authority on a certain topic and obtain useful information posted by that authority.

#### 3. Future outlook

We intend to continue our research with specific services in mind with the aim of establishing information structuring as a useful technology. By making efficient use of real information (weather and other environmental information, historical information about people and things, etc.) obtained from diverse services on the Next Generation Network, we expect to develop information structuring technology that can extract highly accurate situations related to our daily lives, making it applicable to new information searching services.



#### Ryoichi Hayashi

Senior Research Engineer, Open Source Software Computing Project, NTT Cyber Space Laboratories.

He received the Dr.Eng. degree from Tokyo Institute of Technology in 2001. He joined NTT Cyber Space Laboratories in 2001. He has been researching the integration of heterogeneous information sources. He is a member of the Japan Society of Mechanical Engineers.



#### Kengo Fujioka

Research Engineer, Open Source Software Computing Project, NTT Cyber Space Laboratories.

He received the M.S. degree in physics from the University of Tokyo in 1995. He joined NTT Communications and Information Processing Laboratories in 1995. He has been studying effective methods for handling huge volumes of data.



#### Masaichi Hasegawa

Systems Engineering Section, Solution Engineering Division, Corporate Business Headquarters, NTT EAST.

He joined NTT in 1991 and has been researching the integration of heterogeneous information sources.



#### Mitsuaki Tsunakawa

Senior Research Engineer, Supervisor, Open Source Software Computing Project, NTT Cyber Space Laboratories.

He received the B.S. degree in mathematics from Tsukuba University, Ibaraki, in 1990. He joined NTT Communications and Information Processing Laboratories in 1990. He has been researching database management system technology and the integration of heterogeneous information sources. He is currently engaged in R&D of large-scale distributed data management technologies. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan.



#### Shinji Morishita

Research Engineer, Open Source Software Computing Project, NTT Cyber Space Laboratories.

He received the B.S. degree in mathematics from Nagoya Institute of Technology, Aichi, in 1991. He joined NTT Communications and Information Processing Laboratories in 1991. He has been researching database management system technology and the integration of heterogeneous information sources.



#### Seiji Susaki

Director, Research and Development Center, NTT WEST.

He received the B.S. degree in mathematics from Nagoya University, Aichi, in 1988. He joined NTT Communications and Information Processing Laboratories in 1988. He has been researching natural language processing and information retrieval. He has held the current position since 2009.