Conversation-linked Navigation System—Provides Conversation-related Information during a Telephone Conversation

Takeya Mukaigaito†, Shinya Takada, Daigoro Yokozeki, Rie Sakai, Miki Sakai, Takahiko Murayama, and Katsuya Arai

Abstract

This article introduces a conversation-linked navigation system that uses keywords occurring during a telephone conversation to guide users to relevant content on the Internet in a seamless manner in real time.

1. Interconnecting the telephone and Internet

Japan now has more than 70 million Internet users [1], and Internet applications such as e-mail, Web browsing, and blogs are becoming established methods for communication, information gathering, and information posting. Internet-based business-to-consumer e-commerce reached 5.6 trillion yen in 2004, and the importance of the Internet has not escaped the attention of corporate enterprises [2]. The ongoing conversion of the telephone system to Internet protocol (IP) should lead to telephone services with an even greater affinity to Internet services. For example, Google is now providing a beta service called Click-to-Call [3] that enables a user to make a call directly to the advertiser by clicking on an Internet advertisement, thereby achieving a link from the Internet to the telephone. On the other hand, a link from the telephone to the Internet can only be achieved at present by manually inputting a URL (uniform resource locator) into a Web browser. We consider that guiding telephone users to Internet services in a seamless manner can create value, so we are researching and developing a conversation-linked navigation system [4]. The concept is schematically illustrated in **Fig. 1**.

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan
E-mail: mukaigaito.takeya@lab.ntt.co.jp

2. Conversation-linked navigation system

Our conversation-linked navigation system provides you with information related to a telephone conversation you are conducting in real time. The current prototype works with IP telephony based on personal computers. This information comes from the Internet or from your own computer. For example, when you are talking about a certain scenic spot or service, its official homepage and/or related keywords might be displayed (**Fig. 2**). Likewise, if you are talking about upcoming events, your calendar can be displayed so that you can check your schedule while conversing by telephone. This conversation-linked navigation system is expected to provide the following benefits to users and companies that provide services on the Internet.

(1) Benefits for users

The automatic, realtime presentation of information related to conversation provides the following benefits to users:

- Immediate access to detailed and accurate information associated with products, keywords, and other items occurring in a conversation
- Expanded and enhanced conversation based on presented information

(2) Benefits for companies

The direct presentation of conversation-related information to users enables companies that provide their own Web site to acquire telephone users as customers and to enjoy the following benefits:

- Users that mention topics related to the compa-

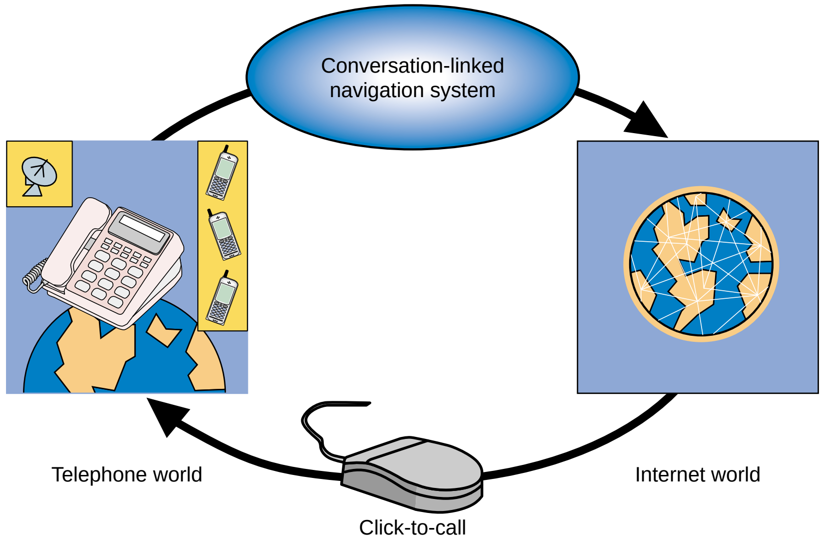

Fig. 1. Conversation-linked navigation system.

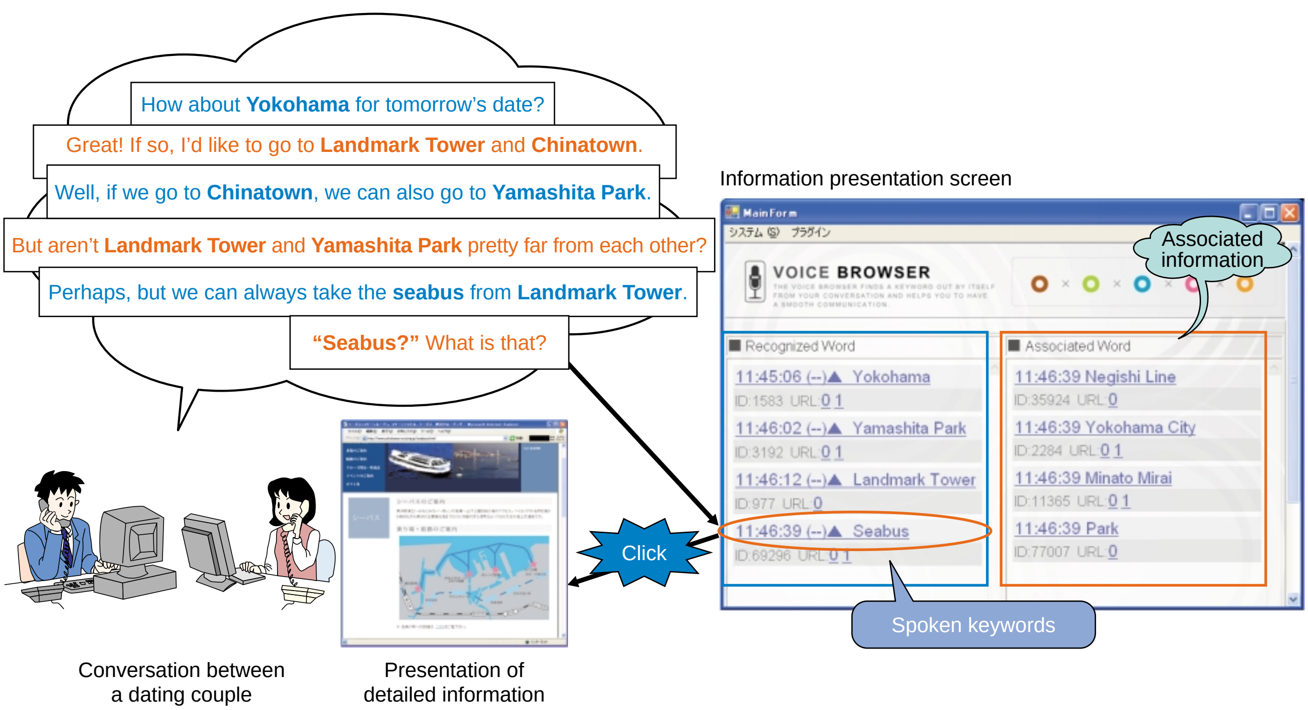

Fig. 2. System operation.

nies can be guided to specific Web sites, which will expand the reach of those providers and add value to those sites.

- Users can be guided to e-commerce sites, whose sales will thus increase.

The following sections describe the technology and mechanism behind this conversation-linked navigation system.

3. Information provision during a conversation

One of main features of this system is that it focuses on proper nouns and common nouns that are considered to have important meaning in a conversation and provides detailed information and associated information about those keywords in real time when they are spoken in a conversation.

(1) Detailed information

This system provides detailed information related to a keyword spoken during a conversation. If such a keyword is a proper noun, the official homepage related to it is presented. If it is a common noun without a homepage, then the results obtained by submitting the keyword to a search engine are provided.

(2) Associated information

This system provides new keywords different from but associated with the keywords spoken during a conversation. For example, if you are talking about a certain athlete, the system can provide information about another athlete that is related in some way. You can also view detailed information by keywords presented as associated information.

4. Information-extraction technology

The following describes knowledge information and the process of extracting relevant information from it, which are basic elements of this system (**Fig. 3**).

4.1 Knowledge information

Knowledge information is a set of keywords and relationships among them. The conversation-linked navigation system makes use of knowledge information as basic information to be provided to the user during a telephone call. The conversation-linked navigation system extracts detailed information and associated information using this knowledge information.

For detailed information about keywords to be provided, each keyword must have a URL that points to detailed information as an attribute value. A keyword also has specific information that indicates whether there is another keyword with the same pronunciation or spelling but a different meaning (i.e., a homonym). Also, for associated information about keywords to be provided, each keyword must have link information pointing to other keywords that are semantically related to form a network of keywords in which one acts as a node. In other words, keywords that are relevant in meaning to a certain original keyword are placed close to the original keyword, while keywords that express different concepts are placed farther away from the original keyword on the network.

Various methods for setting these links between keywords can be considered, but we are currently using a method of setting links by semantic meaning because we think that semantic meaning is the most important attribute for links. Specifically, a keyword that appears in a sentence that describes another keyword is defined as having a semantic relationship

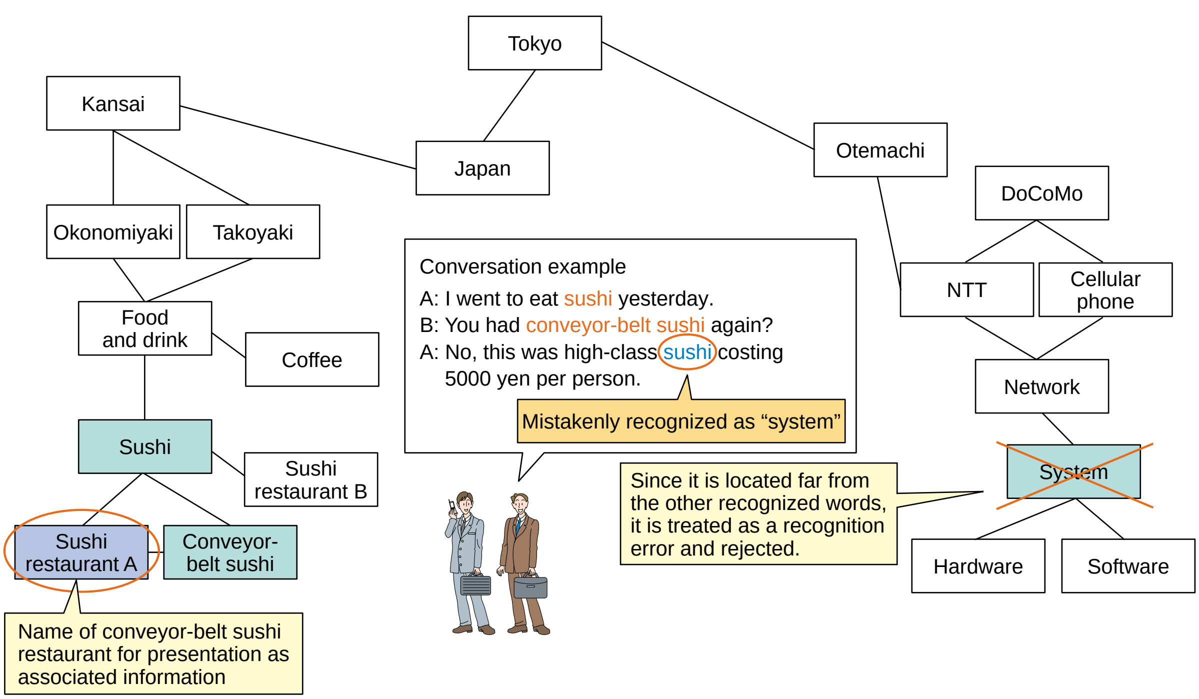

Fig. 3. Method of finding keywords.

with that keyword, and a link that connects those keywords is generated.

4.2 Information extraction

This section describes the method for extracting detailed information and associated information about a keyword spoken during a telephone conversation using the above knowledge information. The conversation-linked navigation system uses a speech recognition engine to determine what the user is saying on the telephone. Note that a keyword output by the engine may be a homonym. For example, the engine might output "Robin" when the correct word is "Robyn" because they have the same pronunciation, and the engine cannot distinguish "Washington" as the name of the capital of the USA from "Washington" as the name of the first president of the USA because the meaning depends on the context. In addition, the application of current speech recognition engines to general, unrestricted conversation can result in frequent recognition errors. For this reason, the identification of homonyms and the removal of speech recognition errors are performed as preprocessing before the extraction of detailed information and associated information.

To identify homonyms and remove speech recognition errors, as well as to extract associated information, the system uses link information between keywords in knowledge information. When you are talking on the telephone, you are more likely to utter keywords that are similar in meaning. Thus, when a homonym exists among the keywords output by the speech recognition engine, the system examines other words spoken at this time and judges which words closely linked to a certain keyword occur most frequently in this conversation in order to identify the intended keyword.

Next, to remove speech recognition errors, the system excludes those keywords that are semantically distant from the other keywords output by the speech recognition engine. In other words, the system treats a keyword deemed to be located far away on the knowledge-information network as a recognition error.

Keywords extracted after homonym identification and speech recognition error removal are now judged to be actual keywords spoken by the user, and ones that lie on the shortest path between any pair of actual keywords can be output as associated information. If the actual keywords chosen as a pair are adjacent to each other on the knowledge-information network, then keywords adjacent to both of them can be extracted as associated information.

In a network constructed of keywords acting as nodes, the relationship between inter-keyword distance and semantic similarity will differ according to the total number of keywords and the number of links connecting them. This relationship is consequently established on an empirical basis using, for example, the average distance between the keywords conducted by the knowledge-information network. The distance to be used as a reference when extracting information can therefore be given as a parameter.

The system extracts detailed information through direct use of the URLs set in both the keywords deemed to be correct and the keywords extracted as associated information in the above information-extraction process.

5. System configuration

A block diagram of the conversation-linked navigation system is shown in **Fig. 4**. The system consists of the information-extraction function described in the previous section, IP-phone and instant-messenger applications, a speech recognition engine, an information-display function, and an information-sharing function. The system inputs the speech output from the microphone of the user's IP-phone into the speech recognition engine and passes the engine's output to the information-extraction function. Note here that keywords recognized at one's own terminal are also transmitted to the other party's terminal using the information-sharing function. This is because treating the other party's spoken words as a target for information extraction in addition to one's own speech means that information will be provided for the whole conversation. For example, if one user says "movie" and the other says "Shinjuku", then information about movie theaters in the Shinjuku district of Tokyo can be provided based on the speech of both users.

Also, considering that the speech recognition engines in use today have a higher recognition rate for a specific speaker than for general speakers, this system simply sends the results of speech recognition performed at each user's terminal to the other user's terminal, thereby solving the problem of insufficient speech recognition rate. It is also possible to enable or disable the transmission of speech-recognition results at one's own terminal as required for privacy.

The system forwards recognized keywords and associated information output by the information-extraction function to the information-display func-

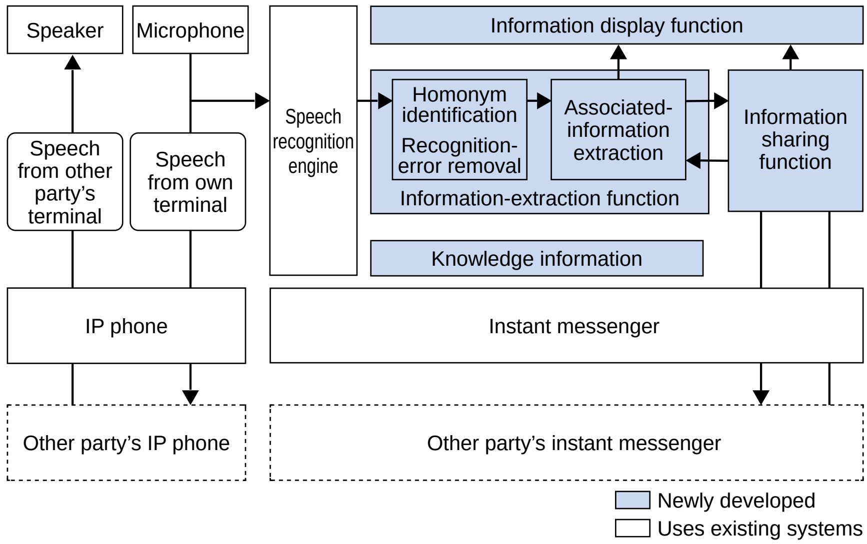

Fig. 4. Basic design of the conversation-linked navigation system.

tion for displaying the keywords to a user. Clicking on the displayed keyword displays detailed information associated with the keyword in a browser window, i.e., displays the content pointed to by the URL connected to the keyword. At this time, URLs displayed on one's own terminal are also transmitted to the other party's terminal using the information-sharing function. This enables both parties to view the same Web page while carrying on a conversation.

6. Conclusion and future outlook

Our conversation-linked navigation system can provide detailed information related to keywords occurring in a conversation and associated information related to those keywords in real time. This system is currently being researched and developed at NTT Information Sharing Platform Laboratories. We plan to continue researching and developing this system focusing on expanding the knowledge information and incorporating new functions such as ones for personalization and sharing. We also wish to expand conversation-linked navigation services to fields for which the features of this system are especially suitable.

References

[1] Internet Association Japan (editorial supervision), "Internet White Paper 2005," Impress, pp. 29-48, 2005 (in Japanese).
[2] http://www.meti.go.jp/press/20050628001/e-commerce-set.pdf
[3] http://www.google.com/help/faq_clicktocall.html
[4] T. Mukaigaito, S. Takada, D. Yokozeki, R. Sakai, T. Murayama, and K. Arai, "Conversation-linked Navigation System," FIT2005 (4th Forum on Information Technology), pp. 393-394, 2005 (in Japanese).

Takeya Mukaigaito

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

He received the B.E. degree in engineering science from the University of Tsukuba, Ibaraki, in 1989. In 1989, he joined NTT Communication and Information Processing Laboratories, Kanagawa. He is a member of the Information Processing Society of Japan (IPSJ).

Shinya Takada

Network Application Platform Project, NTT Information Sharing Platform Laboratories.

He received the B.S. and M.S. degrees in physics from the University of Tokyo, Tokyo, in 1996 and 1998, respectively. In 1998, he joined NTT Software Laboratories, Tokyo. His current research area is billing-related systems, especially usage-related aspects.

Daigoro Yokozeki

Software Architecture Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in electronics, information and communication engineering from Waseda University, Tokyo, in 1996 and 1998, respectively. In 1998, he joined NTT Software Laboratories, Tokyo. He is engaged in server-side middleware technology such as ebXML, Web Services, and transaction processing system in NTT Information Sharing Platform Laboratories.

Rie Sakai

Software Architecture Project, NTT Information Sharing Platform Laboratories.

She received the B.S. and M.S. degrees in electrical engineering from Purdue University, USA, in 1994 and 1996, respectively. She joined NTT Multimedia Platform Laboratories, Tokyo, in 1996 and engaged in research on Net-Libra (networked digital library). Since 2005, she has been working on this topic at NTT Information Sharing Platform Laboratories.

Miki Sakai

Software Architecture Project, NTT Information Sharing Platform Laboratories.

She received the B.S. and M.S. degrees in mathematics from Tokyo University of Science, Tokyo, in 2001 and 2003, respectively. In 2003, she joined NTT Information Sharing Platform Laboratories.

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. In 1986, he joined NTT Communication and Information Processing Laboratories, Yokosuka. 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, the University of Electro-Communications, Japan. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Japan Society of Artificial Intelligence.

Katsuya Arai

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

He received the B.E. and M.E. degrees in electronic engineering from the University of Electro-Communications, Tokyo, in 1988. His research interests include computer-human interaction, CSCW, ubiquitous computing and Web technologies. He is a member of the IEEE Computer Society, the Association for Computing Machinery, and IPSJ.