Highly Accurate Similar Case Retrieval System for Call Centers Using Two-word Linked Expressions

Yukako Kitagawa† *and Motonori Tamaki*

Abstract

In call centers, responding quickly and accurately to inquiries from customers regarding services and problems is important from the perspective of customer satisfaction. NTT Cyber Solutions Laboratories is developing a highly accurate system for retrieving similar cases using two-word linked expressions. The goal of the system is to effectively utilize cases from past inquiries.

1. Importance of case retrieval systems

As IP services proliferate and information technology innovations evolve more with each passing day, there is an increasing demand for call centers at the companies providing these services and products to offer fast, accurate responses to ensure customer satisfaction. However, it is becoming increasingly difficult to respond with the required speed and accuracy based only on the efforts of the call center staff (hereafter called agents) to acquire new skills, because this task demands a wide range of knowledge regarding IP services, and because new products and services are being developed and marketed rapidly. A report from Purdue University in the United States suggests that 60% of inquiries to call centers are repeats of previous cases [1]. This suggests that call centers need a case retrieval system that can efficiently reutilize cases from past inquiries to make up for the limitations of agent skills and reduce the time required for investigating current inquiries and forming appropriate responses.

2. Issues of conventional case retrieval technologies

Figure 1 shows a concept drawing for a typical case retrieval system at a call center. In this diagram, a "case" is defined as a combination of two elements: i) the "inquiry" created based on the customer's inquiry and ii) the "response" to that inquiry. Currently, the technologies most commonly used in case retrieval systems are "diagnostic technologies using expert systems" and "full-text retrieval technologies". The features and issues related to these two existing technologies are discussed below.

2.1 Diagnostic technologies using expert systems

Diagnostic technologies using expert systems are case retrieval technologies that can find relevant cases by having the agent respond to a series of questions posed by the system. Their advantage is that relevant cases can be located simply by giving responses, so accurate retrieval can be done quickly. However, the cases must be classified and stored in advance, so when a new case comes up, somebody must decide which part of the system it applies to and store it there.

Furthermore, if a case comes up that is outside the scope of the system, it is necessary to reevaluate the system itself. For this reason, this technology is effective in fields like medical diagnosis, where systems only change on a medium to long-term basis, but it is not suitable for fields where many new technologies and services are created in the short term, as in telecommunications.

2.2 Full-text retrieval technologies

Full-text retrieval technologies are also referred to as "keyword searches" and "word searches". When using full-text retrieval technologies in case retrieval, questions or responses are used as text for search

† NTT Cyber Solutions Laboratories
Musashino-shi, 180-8585 Japan
E-mail: kitagawa.yukako@lab.ntt.co.jp

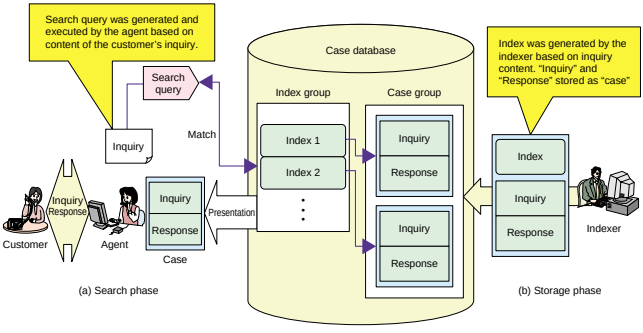

Fig. 1. Case retrieval systems.

commands, and an index is automatically generated for each text based on the frequency with which the words in that text appear.

For this reason, there is no need for special tasks when cases are stored, as in diagnostic expert technologies. Also, with this search method, the agent only needs to input combinations of words expected to be in the text, so ease of use is also an advantage.

Because the index is generated based on the frequency with which the words in that text appear, the index does not necessarily reflect the original intent or meaning of the text. Also, because the search is executed simply by inputting words, the search results often include a certain quantity of text that has no relationship whatsoever to the intended search, so it takes time to find the desired text.

3. Case retrieval system using two-word linked expressions

3.1 What are "two-word linked expressions"?

NTT Cyber Solutions Laboratories has proposed two-word linked expressions for use in searching for cases quickly and accurately [2], [3]. The term "two-word linked expression" refers to two words extracted from the inquiry content, based on the modifier-modificand relationship. It expresses the main theme of the text and the link between them (Fig. 2).

For example, consider the following inquiry: "Immediately after changing my ISP for the FLET'S ADSL service, I was unable to connect to the Internet." The agent decides what the main theme of this sentence is, and extracts the words "connect" and "Internet". Next, he links the two words based on the modifier-modificand relationship to form a two-word linked expression "Internet – connect". The agent inputs this two-word linked expression into the system and executes the search. The results are achieved by matching the search query using the two-word linked expression with the indexes of stored cases in a case database to find corresponding cases. For this reason, even when cases are stored in the case database, indexes are created and stored using two-word linked expressions from the inquiry content.

3.2 Features of two-word linked expressions

By using two-word linked expressions when storing or searching for cases, we can express the intent or meaning of the case in the index, which could not be done with indexes for full-text retrieval technologies.

For example, consider the following set of three phrases: "Use of computers in English education"; "Use of English in computer education"; and "Use of computers and English in education". Each expresses a combination of the words English, education, computer, and use, but it is difficult to grasp the intent or meaning of the phrases through the combinations of the words alone.

If we extract words from the phrases based on a

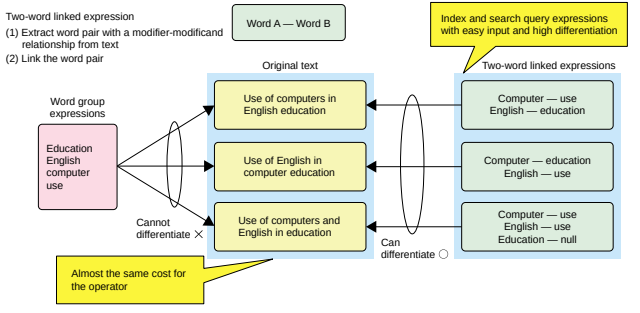

Fig. 2. Two-word linked expressions.

modifier-modificand relationship and express the theme as a two-word link, however, then we can express the themes in three separate categories: i) "computer – use" + "English – education", ii) "English – use" + "computer – education", iii) "computer – use" + "English – use" + "education – null"

With word combinations alone, all these cases would be expressed in the same way, and it would not be possible to clearly indicate the meaning or intent of the text. With two-word linked expressions, however, even though the same number of words are used, it is possible to express the meaning or intent of the text specifically using different combinations of words, thus enabling the three texts to be differentiated. In particular, in similar case retrieval for call centers, many texts are likely to include the same words, such as products, services, and devices name, even though they refer to different cases. Therefore, the search results from a full-text retrieval system often include a certain quantity of text that has no relationship whatsoever to the intended search, so it takes time to find the desired text.

In other words, by using two-word linked expressions, it is possible to narrow down the text based on the intended search—a function that was impossible with conventional full-text retrieval technologies using indexes based on word combinations—thus raising the search performance.

Another feature of this method is that the user can create the search query simply by linking two words extracted from the inquiry at the time of the search, so the work involved is essentially the same as for search tasks using word input for the full-text retrieval.

3.3 Main part and supplementary part expressions

We studied different types of inquiries based on a variety of case databases, such as the FAQs (frequently asked questions) often found on the Internet. As a result of our studies, we found that these inquiries can generally be divided into three main types: "What", "How", and "Symptom".

"What" inquiries relate to the content of products and services, as in "What is ...?" In these inquiries, the case can be expressed in one two-word link expression. In "How" inquiries, the customer wants to find a way to achieve a certain goal, so the expression takes the form of "How can I ...?" "Symptom" inquiries generally take the form of "I can't ...", describing a failure, a malfunction, or another type of problem, and asking for a way to deal with it. For these "How" and "Symptom" inquiries, it is necessary to provide a response that is appropriate to the customer's usage environment. Thus, in addition to the customer's request and the symptoms to be resolved, the case description should also include information about the environment related to the devices and services that the customer is using. Then, search accuracy should increase if the target of the search also includes this environment-related information.

The index or search query related to the inquiry

content therefore comprises a pair of two-word linked expressions: a main part indicating the theme of the inquiry and a supplementary part expressing environment-related information, such as the devices and services that the customer is using (Fig. 3). Using this configuration, the search targets only cases in which both the main and supplementary parts match exactly, and cases with different environments can be eliminated from the search results. If no perfectly matching cases are found, then the search can be executed targeting only the main part and disregarding the supplementary part; in this way, although the environment may be different, the agent can locate reference information that will be useful in forming a response. It is also possible to search for cases in which only the supplementary part matches. This type of search could be effective for handling a vague customer report because it might be possible to identify details of an inquiry from the environment-related information.

3.4 Multi-indexing

In both indexing and retrieval, terms in two-word linked expressions often fluctuate even within a given text. For example, consider the following phrase "the use of computers in English education". If you ask people to express this as a two-word linked expression, some might use "English – education" while others might use "computer – use". This variation of terms in indexes and retrieval queries means that queries often fail to retrieve relevant cases. This problem represents sub-optimal performance for the case-retrieval system.

To compensate for the variation of terms in indexes and retrieval queries, we propose using multiple index terms for two-word linked expressions as index terms for each text. This method, called the multi-index method, can find a relevant case by using any of several two-word linked expressions; that is, it provides improved search performance.

3.5 Index addition algorithm

As described in 3.4, the variation of terms in indexes and retrieval queries means that cases that should be found often are not. In the section, we propose a form of re-retrieval through query expansion, in which the new index term is simultaneously added to the retrieved text. When an agent cannot retrieve any cases using existing two-word linked expressions from the case database, the following algorithm is applied to re-retrieve similar cases and add new index terms to the retrieved text.

- Step 1. Query expansion

  The original two-word linked expression is bro-

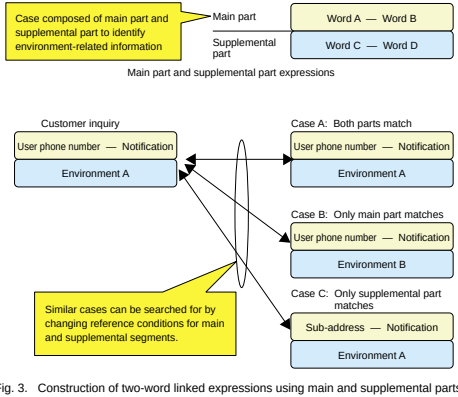

Fig. 3. Construction of two-word linked expressions using main and supplemental parts.

ken up into the individual words, which are then used to retrieve similar cases by full-text search.

- Step 2. Result evaluation
  The agent evaluates whether each retrieved case is relevant.
- Step 3. Index addition
  A re-retrieved case that fits the agent's needs, that is, a case identified as relevant, is indexed with the original two-word linked expression as a new index term.

This algorithm can find cases that cannot be retrieved by using two-word linked expressions. Furthermore, new two-word linked expressions can simultaneously be added.

4. Retrieval performance using two-word linked expressions

To evaluate the performance of searches using two-word linked expressions, we extracted 136 cases related to communication networks, and conducted comparison tests of full-text retrieval and searches using two-word linked expressions. The subjects were 42 male and female adults. Each subject was asked to conduct searches using both of these retrieval methods.

For the two-word linked expressions, two types of searches were tested: searches in which both the main and supplemental parts matched perfectly (A) and ones in which only the main part matched (B). For the full-text retrieval also, two types of searches were tested: using a Boolean AND search for all words input (C), and using a Boolean OR search for all words input (D).

In comparing the two methods, we used "recall ratio" and "precision ratio" indexes to express retrieval performance. The recall ratio expresses the ratio of the amount of information that was actually retrieved to the amount of information stored in the retrieval system that matches the search request. This index ranges from 0 to 1, with 1.0 indicating that all of the matching information in the retrieval system was successfully retrieved. A result of less than 1 indicates that there is still stored information remaining in the system that could not be retrieved.

The precision ratio index shows how much of the retrieved information matches the search request. This index also ranges from 0 to 1, with 1.0 indicating that all of the information retrieved matches the search request. A result of less than 1 indicates that some of the retrieved information does not match the search purpose.

The results of the tests are as follows (Fig. 4): the precision ratio for full-text retrieval (C) was 0.49. The search using two-word linked expressions (A) had a precision ratio of 1.0, a perfect score. There were 136 cases in all, so the performance will likely decrease as the number of cases increases, but these results indicate that our method has great potential compared with full-text retrieval.

In terms of the recall ratio, while the full-text retrieval (D) scored 0.94, the two-word linked expressions (B) scored 0.64, indicating that there were some cases that could not be retrieved. In the future, we intend to increase the search performance by using

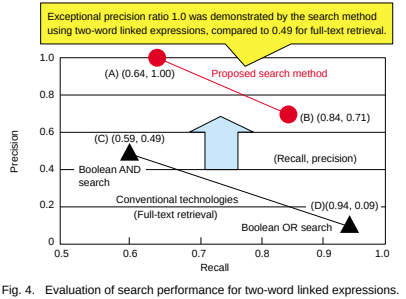

Fig. 4. Evaluation of search performance for two-word linked expressions.

full-text retrieval technologies to complement our two-word linked expressions.

5. Future developments

Having demonstrated that two-word linked expressions achieve higher retrieval performance than full-text retrieval, we plan to continue studying ways of improving the recall ratio and other aspects of the retrieval performance to make a high-precision similarity retrieval system using this technology. We also feel that it is necessary to study ways of improving the overall system usability, such as the ease of inputting search equations and storing cases. To do this, we will develop prototype systems that handle the full range of case retrievals that occur in call centers.

References

[1] Perdue University, "Research Highlights," Call Center, No. 7, p. 12, 2001.
[2] Y. Kitagawa, T. Kuwabara, M. Tamaki, T. Kokubun: "A proposal for inquiry indexing methods using two-word linked expressions," IEICE Conference, p. 106, D8-9, 2002 (in Japanese).
[3] Y. Kitagawa, T. Kokubun, T. Kuwabara, M. Tamaki: "A proposal for search methods using two-word linked expressions," IPSJ Research Report, NL-152, 2002 (in Japanese).

Yukako Kitagawa
Research Engineer, Service Platform Development Project, NTT Cyber Solutions Laboratories.
She received the B.S. degree in administrative studies from the Prefectural University of Kumamoto, Kumamoto in 1998 and the M.S. degree in mathematics and computer science from Kumamoto University, Kumamoto in 2001. In 2001, she joined NTT Access Network Service Systems Laboratories, Chiba, Japan. In 2002, she joined NTT Cyber Solutions Laboratories, Tokyo, Japan. She is a member of the Information Processing Society of Japan.

Motonori Tamaki
Senior Research Engineer, Service Platform Development Project, NTT Cyber Solutions Laboratories.
He received the B.S. and M.S. degrees in information science from the University of the Ryukyu, Okinawa in 1990 and 1992, respectively. In 1992, he joined NTT Field System R&D Center, Ibaraki, Japan. In 2002, he joined NTT Cyber Solutions Laboratories, Tokyo, Japan. He is a member of the Institute of Electronics, Information and Communication Engineers.