Applying AI-Based Domain-Specific Semantic Search Function to Family Register Administration at the Municipal Office of the Osaka City Government

● Yoshimi Tsumori ● Tatsuki Oda ● Satoshi Yoshida ● Satoshi Hatakeyama
● Yutaka Matsuda

Points of contact for customers such as reception counters and call centers tend to have high staff turnover. Therefore, they must handle customer inquiries and leverage knowledge with competence without having the expertise of experienced staff. There are two ways to respond to the knowledge utilization needs: by means of FAQ and by leveraging structured knowledge information. Fujitsu has recently had an opportunity to offer the Osaka City Government in Japan its AI technology for domain-specific semantic search function. With this technology, necessary information can be identified from a large volume of data concerning family register administration and work manuals. Then it enables staff members without extensive expertise to efficiently handle their administrative work regarding family registers. It also eliminates the need to prepare the enormous learning data necessary to leverage FAQ. This paper explains the domain-specific semantic search function and describes the case in which it was introduced at the Osaka City Government.

1. Introduction

Knowledge management has been a common practice at enterprises for sharing and leveraging information in workplaces, for which they use systems such as groupware and data warehouse (DWH). These systems are useful to share information about the techniques of top-performing sales personnel, approaches to problem-solving based on the expert know-how of experienced on-site staff, and so on, allowing others to benefit by gaining insights into solutions. However, as the information is text-based, it ultimately requires human effort to read and comprehend these texts in order to judge the usefulness of the information.

Conventional knowledge management systems have been useful for gathering data, but not so far as mechanically extracting valuable information from the data and allowing people to take advantage of it.

Given this, there is a growing interest in enhanced big data processing and the utilization of AI technology. These may be effectively applied not only to knowledge management systems but also to other areas such as call centers, reception counters, and system maintenance, where work involves searches through FAQs and troubleshooting manuals.

In this paper, we first describe the AI-based domain-specific semantic search function, which Fujitsu has developed to apply AI to text processing technology. We then present a case in which this technology is applied to family register administration in collaboration with the Osaka City Government in Japan.

2. Challenges in knowledge utilization

It is often the case, particularly with business enterprises, that the introduction of knowledge management starts as a company-wide initiative. Their first task is to gather and record as much information as possible, which often results in the inclusion of poor quality information.

In early stages of the introduction of this system, it is therefore necessary to develop dictionary data to improve the accuracy of data searches. However, it is not easy to build an optimal dictionary. This issue often results in system users encountering difficulties in obtaining useful information and being likely to resort to the conventional method of asking their colleagues who have the information. As a result, it is not unusual that such a situation hinders the gathering and updating of knowledge, discouraging users to use the system

and creating a vicious circle.

Issues such as this are often due to the following technical problems inherent in conventional knowledge management technology.

1) Generating greater workloads to gather knowledge

Conventionally, the focus on ease of gathering and utilization of data meant texts such as reports and summaries generated too much preparation as they needed to fit pre-designed formats. Reformatting the information to specific formats, tagging it with property information, and specifying search targets generated a need to add descriptions of the text (outline, keywords, etc.), thereby increasing workloads.

2) Work to enhance convenience (development of dictionary data)

The development of dictionary data is an important factor in improving search accuracy. For example, a thesaurus is indispensable to enable a fuzzy search that accounts for variants of a certain expression, such as acronyms, in-house technical terms, and use of different alphabets such as English letters and katakana characters. The thesaurus data are defined so that, for example, a search of "engine" will retrieve all the following: "エンジン," "Engine," and "ENG."

Dictionary data constantly increase as there are new technologies and methods, requiring daily maintenance to ensure the data's usability.

To leverage knowledge management, it is crucial to provide information that is valuable from users' point of view. However, the considerable burden involved in maintaining that value is a serious problem.

3. Developed technology

Fujitsu utilizes machine learning to process a large amount of text data and generate vectors regarding word association. These vectors are further extended to the document level so that similarities between documents can be quantified. Fujitsu leverages this technology to offer an efficient system for executing searches at the keyword and sentence levels.

3.1 Labor-saving on registration to word dictionary

The system learns a large volume of text data, identifying attributes such as the order in which certain words appear in the texts and generating vectors for inter-word associations. Based on this, the words that follow a similar pattern of appearance in the documents are considered synonymous. Acronyms and in-house technical terms used in sentences can be treated in the same manner. This method is sometimes less accurate compared to a method of defining association with thesaurus data because the word association is judged based on the probability of usage patterns. However, its preparation is far less labor-intensive than developing dictionary data.

3.2 Utilization of internal documents

Internal documents that are created through routine work, such as daily reports and user inquiry histories, are used as learning data. For example, creating FAQs requires sets of question-answer pairs. However, some questions have more than one answer, and these answers must be organized and gathered as knowledge. Where new pieces of knowledge are generated, they must be added to the knowledge base to keep it up to date, and this creates more maintenance work.

The technology we have developed is able to learn information about the generation of new knowledge from the reports and inquiry history data. This necessitates the processing of data and forms to feed into AI, but there is no need to recreate data again for knowledge accumulation. There is also an advantage in that machine learning handles the documents that are prepared internally and learns company-specific terms in the process.

3.3 AI-based domain-specific semantic search function

Take the word "engine" for example; it refers to automotive engines in a general context, but in computer science, it refers to computer programs that run certain commands, as in an AI engine or search engine. It is therefore important to understand that the same word may have different meanings depending on the domain in which it is used.

Domain-specific thesauruses will be useful in this context. "Domain-specific" here means that the system learns and comprehends words and their meanings specific to areas of specialization by learning the documents in these areas.

Fujitsu's AI-based domain-specific semantic search function uses AI to learn texts in a specific area of specialization to quantify the levels of similarity between words

and documents automatically. It facilitates document searches without the need to create domain-specific thesauruses. By leveraging this technology, it is possible to provide a knowledge search system that significantly reduces the burdens of preparing training data and dictionary data as well as maintenance workloads that were conventionally problems.

Furthermore, it combines the core AI technology with a UI that is adaptable to the information required by specific clients so that the system is easy for them to use in the context of their specific work requirements.

3.4 Context-specific UI in combination with AI engine

The results induced through AI may not necessarily be the needed information. When we run a search on the Internet, we often look through the results from the top, one by one, until we find the piece of information that answers our query. For kinds of work that involves AI-enabled searches, it is necessary to provide a system that accounts for unique attributes and particular requirements of the work.

For example, in order to quickly identify the information needed among the outcomes of a search, a method that narrows as it drills down into information is effective. To have an overview of the information, it is necessary to present the results not in a list form but in other visual representations like a network node image to allow for further searches.

By combining the core AI technology with UI to enhance convenience, the system will become truly usable.

4. Evaluation and case of practical application

The administrative office of the Osaka City Government called for a bid to construct and provide AI-based services to support administrative work on a long-term contract. In response to this, Fujitsu proposed our AI-based domain-specific semantic search function, which was adopted. In this section, we describe the initiatives taken at the Osaka City Government and outline the AI-based family register administration service, which was adopted by the municipal government.

4.1 Challenges and Initiatives at the Osaka City Government

Ward offices in Osaka City found it necessary to leverage knowledge such as the know-how of experienced staff members in family register administration because of the shortening of staff transfer cycles and the increase in short-term employment. Changes in social values as well as the increasing cases of handling family registers that involve non-Japanese persons necessitate expert knowledge, and it was necessary to reduce the time needed to search for the knowledge.

To address these challenges, we ran a pilot operation of family register administration services at ward offices in Osaka City using AI-based domain-specific search functions. This initiative aims to provide knowledge support for staff members irrespective of the length of their experience in the work, reduce the time for handling tasks, and improve accuracy and efficiency as well as civil service quality.

4.2 Features of AI-based service to support staff work

We first aimed to develop a service that is easy for staff to use. This requires a feature that helps them when they receive family register applications in practice by filtering information and showing options necessary to handle the cases.

For example, less experienced staffs sometimes do not understand certain technical terms or have a clear idea as to what to look for. Even in these cases, they can simply select provided menu options to use preset keywords and search for the documents that apply to the case acceptance. This feature will enable a service with support like that of experienced staff with a wealth of expert knowledge.

4.3 Characteristic features of AI-based service

1) Learning data management

The system can register texts submitted to a search as learning data. This feature allows learning data to be compiled even after the system is put in operation, enhancing the response accuracy.

2) Supplementary learning data management

Supplementary learning data are used to develop a knowledge structure outside the learning data. These data include not only the guidelines and FAQ

documents supplied by the Osaka municipal office but also such external sources as newspaper articles and Wikipedia entries to be learned. This parallel learning strengthens the links between technical terms found in the texts in the database and general vocabulary not included in them. For example, the system can learn that the terms "international marriage" and "extranational marriage" are similar in meaning.

3) User dictionary management

This feature allows machine learning to learn from the sentences in documents at a word level to develop a structured body of knowledge. By registering the technical terms specific to family registers, it can correctly recognize individual words and improve the accuracy of answers. Take "judicial divorce" for example; the Japanese term for this comprises two parts—lawsuit and divorce. Without the learning process, the system would recognize these parts separately. But with the learning, the term "contested divorce" can be properly identified.

4) Search engine

This feature employs the AI-based domain-specific semantic search function to reference the structured knowledge data registered in the learning model and search for the information associated with the keywords. The search results will list the most similar or relevant document first in terms of the keywords so that staff can quickly reach the required document in the list.

5) Ranking management

The search result page has a set of evaluation buttons to make feedback possible on whether the result is useful or not. The feedback provided by users as they click these buttons will help to compile the information deemed useful. The built-in ranking management feature uses the feedback data to display search results with higher feedback ranking closer to the top of the search result list to improve staff productivity.

6) Search result feedback

This feature compiles the search history and user feedback. The data is used to review the learning data for machine learning and user dictionary data, followed by the re-learning of the structured knowledge model to improve the response accuracy of the AI-based service.

4.4 AI-based service system outline

Figure 1 shows the system configuration of the AI-based service.

1) The learning model is created by developing structured knowledge data based on family register literature, specialist dictionary data, and other data that staff frequently use. This model is deployed on Zinrai Platform Service.
2) The system references the learning model via the application programming interface (API) for domain-specific semantic search function using the application on the web application server, then returns the search results to the staff terminal.

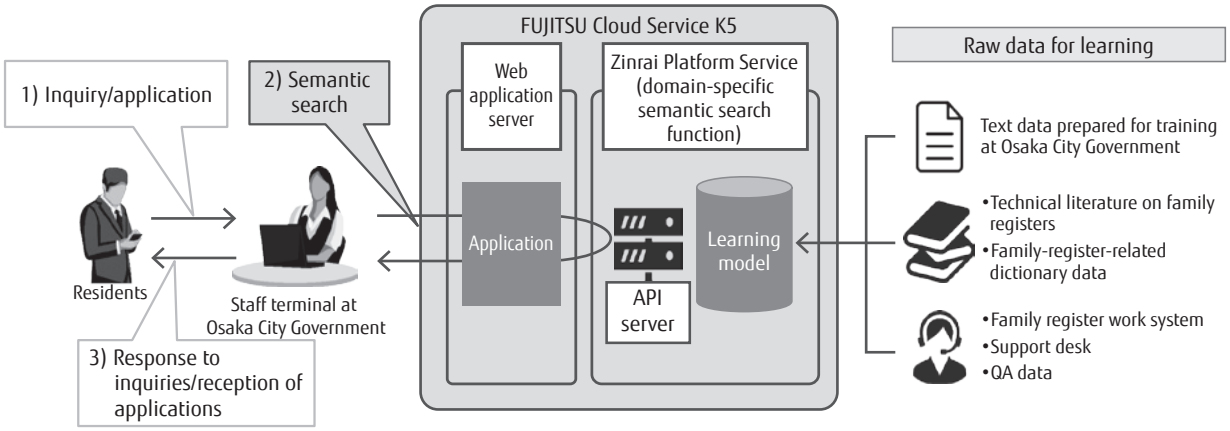

Figure 1
System configuration.

4.5 Reducing the burden of creating learning data

Reception points such as call centers often employ dialog-type AI, which requires sets of learning data based on a large volume of conversational data and/or questions tagged with multiple answers. However, such data have not been developed for family register administration. Preparing those data for this work would require many person-hours. Our AI-based domain-specific semantic search function does not need to rearrange raw data into a QA format. Thus, it can significantly reduce the user workload for creating learning data. With these features in place, we succeeded in quickly providing the AI-based service for family register administration to the Osaka City Government. The continued learning also helped to reduce the burden involved in data preparation, minimizing the mid/long-term effort to operate the AI-based service.

5. Conclusions

This paper described the AI-based domain-specific semantic search function and presented a case in which this technology is applied to family register administration in collaboration with the Osaka City Government.

The AI-based service for staff work support presented in this paper has been operating as a pilot case at two ward offices of the Osaka City Government (Higashi-yodogawa and Naniwa Wards). The AI-based domain-specific semantic search function is also employed elsewhere as corporate internal search engines. Thus, we firmly believe that this technology is widely applicable in a business context as well as in the domain of public administration and services.

Yoshimi Tsumori
Fujitsu Ltd.
Mr. Tsumori is currently engaged in AI-based service integration.

Tatsuki Oda
Fujitsu Research Institute
Mr. Oda is currently engaged in analytics consultancy.

Satoshi Yoshida
Fujitsu Ltd.
Mr. Yoshida is currently engaged in the development of administrative systems for local authorities.

Satoshi Hatakeyama
Fujitsu Ltd.
Mr. Hatakeyama is currently engaged in the development of administrative systems for local authorities.

Yutaka Matsuda
Fujitsu Ltd.
Mr. Matsuda is currently engaged in the sales of administrative systems for local authorities.