Automatic FAQ Chatbot Building Technology

Service Innovation Department   Hiroshi Fujimoto   Takanori Hashimoto   Yuriko Ozaki

Innovation Management Department   Takuya Kobayashi   Koji Yamazaki

FAQ chatbots that use AI to provide responses instead of a human are becoming more widespread. In the past, when building such FAQ chatbots, scenarios had to be built-in or incorporated using machine learning, so it was costly to generate each scenario and accuracy of the responses was lower than desired. NTT DOCOMO has developed a technology to generate chatbot response scenarios automatically from a set of FAQ question and answer texts. This enables the FAQ provider to build a chatbot easily, without the need to develop each scenario for the chatbot.

1. Introduction

It has become common practice for enterprise and administration call centers and other support departments to replace parts of their work with chatbots[*1] to reduce costs. Frequently Asked Questions (FAQs) repeatedly answer similar questions for different people, so they are particularly well suited to having a chatbot provide responses mechanically.

In the past, chatbots specializing in handling FAQs (hereinafter referred to as "FAQ chatbots") have generally been built based on either a scenario model, or a category model.

For a scenario-based chatbot, scenarios that branch the dialogue with the user are generated based on each of the user's queries, and responses defined for each branch are returned. For an FAQ, the chatbot continues the dialogue, even if the user's query is ambiguous, digging into the query and narrowing the possible responses until it can give an accurate response. However, as the number of

---

*1 Chatbot: A program that automatically conducts dialog with people with speech or text chat.

branches before reaching a response increases, the cost of generating the scenarios also increases, so this approach is not suitable for large FAQs containing hundreds or thousands of possible responses. Examples of services providing scenario-based chatbots include Repl-AI®*2 [1] and IBM Watson™*3 Conversation [2].

Category-based chatbots select the response that is best suited to the user's query from among prepared responses and present it to the user. Generally, the algorithm used to select an appropriate response is supervised machine learning*4. A single response is given to the user's query, so there is no need to prepare branching scenarios as with scenario-based chatbots, and it is relatively easy to apply this method to large-scale FAQs. However, it is not possible to clarify a user's query with on-going dialogue if it is ambiguous, so the accuracy of responses is generally lower than with scenario-based chatbots. One example of a service providing category-based chatbots is Microsoft Azure™*5 Bot Service [3].

NTT DOCOMO has developed an entirely new, automatic FAQ chatbot building technology to resolve issues with the scenario-based and category-based techniques described above. The proposed technology supports on-going dialogue through scenario branching and is able to generate scenarios for each branch automatically using a newly developed proprietary algorithm. Thus it is able to maintain the accuracy of a scenario-based chatbot, while being applicable to large-scale FAQs, like a category-based chatbot. Note that the QnA Maker API [4] from Microsoft is an existing technology for building FAQ chatbots automatically, but it is only able to build category-based chatbots, and to

the best of our knowledge, no other technology able to build scenario-based chatbots automatically currently exists.

NTT DOCOMO has also developed an automatic FAQ chatbot building system using the proposed technology. The system is a Web application that operates by calling an API implementing the proposed technology, and provides functionality including a GUI for building FAQ chatbots, visualization of user dialogue history, and links with existing chat platforms such as LINE*6 and Facebook Messenger®*7. The system has been providing services for corporate enterprises since March 2018.

This article describes the automatic FAQ chatbot building technology, and an evaluation of its performance. It also describes the automatic FAQ chatbot building system that we have developed.

2. FAQ Chatbot

2.1 Overview

An overview of the FAQ chatbot is shown in **Figure 1**. With this technology, intermediate data called a truth table is first generated from the FAQ data. A scenario-based FAQ chatbot is then built by registering the truth table in a natural-language dialogue platform [5] developed by NTT DOCOMO.

1) FAQ Chatbot Dialogue

An example of dialogue with the FAQ chatbot is shown in **Figure 2**.

The FAQ chatbot operates assuming that pairs of queries and responses, referred to as FAQ data, has been prepared ahead of time. Thus, by finding a query that matches the user's intent, a corresponding response can also be identified.

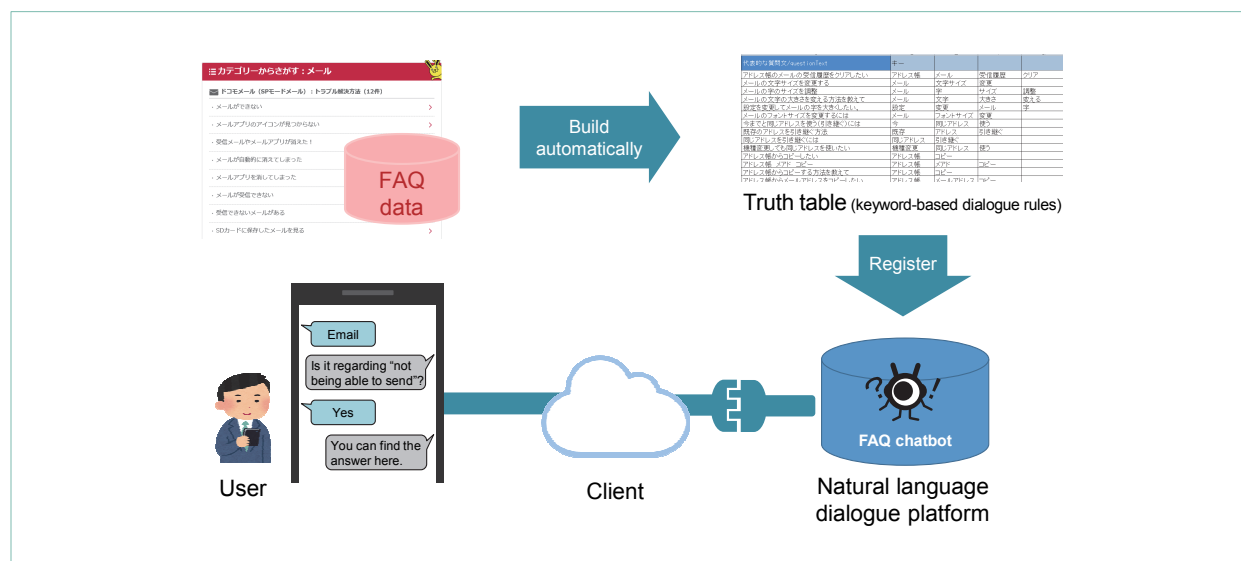Next, we discuss methods for correctly identi-

Figure 1　FAQ chatbot overview

Figure 2　Example interaction

fying queries in the FAQ chatbot. Often, a user's first input ("Initial query" in Fig. 2) is ambiguous, so it can be difficult to identify a query text from just the initial query. In such cases the user is asked additional questions ("Key question" in Fig. 2) to update the query information, and ultimately find an applicable query text.

2) Identifying Queries Based on the Truth Table

The FAQ chatbot searches for queries based on the truth table, which is used to generate dialogue

scenarios automatically, according to the user's query. It is composed of a list of query texts together with key phrases from the query texts, hereinafter referred to as "keys," as shown in **Figure 3**. Keys are phrases that a user is predicted to input when searching by keyword search for an applicable FAQ entry. The proposed system generates them automatically from the query texts. Since there are various expressions for a given key that users could potentially use, in addition to the truth table, the proposed technology also generates a variation dictionary, which stores expression variations for each key.

The FAQ chatbot searches the truth table based on the user's initial query to identify the most applicable query text. If it is difficult to identify a query text from the user's initial query, the system asks key questions to narrow down the possible query texts. Details of FAQ chatbot operation are described below.

2.2 Generating Truth Tables and Variation Dictionaries

To build an FAQ chatbot, the proposed technology generates a truth table and a variation dictionary, as described below.

1) Generating the Truth Table

The truth table is a collection of query texts broken down into keys, which are essential elements of each question (Fig. 3). The keys are not simply a collection of morphemes[*8], they are selected according to a particular algorithm in the proposed technology. Generating the truth table involves two processes: extracting keys using parts-of-speech rules, and discarding unnecessary keys based on a stop key list.

For example, the query, "I don't know my Wi-Fi®[*9] password" is separated into the keys, "Wi-Fi," "password," and "I don't know" (Q1 in Fig. 3). From the results of morpheme analysis, phrases that do not have meaning on their own, such as prepositions and helper verbs, are discarded, essentially selecting verbs and nouns as keys. However, if just single nouns and verbs are selected, special terms with multiple nouns, such as "docomo Wi-Fi," would be separated into the keys, "docomo" and "Wi-Fi." To resolve this sort of issue, we use combining rules to combine sequential nouns and numbers into a single key (Q2 in Fig. 3).

There are also words that are not suitable as keys even though they are nouns or verbs. For

	Query	Key		
Q1	I don't know my Wi-Fi password.	Wi-Fi	Password	~~Don't know how~~
Q2	I want to connect to docomo Wi-Fi.	docomo Wi-Fi	Connect	Stop key
Q3	How do I send email?	Email	How to send	~~Want to know~~
Q4	Sending email	Email	Sending	Stop key
Q5	I can't delete photos.	Photos	Cannot delete	Delete_incomplete
Q6	I accidentally deleted a photo.	Photos	Deleted	Delete_complete

Figure 3　Truth table example

example, the query, "How do I send an email?" is separated into "email," "send," and "How do I," but "How do I" is used so often in a FAQ that it is not helpful as a key to distinguish among queries. To remove such unnecessary keys that the parts-of-speech rules have created, key removal rules called "stop keys" are used (Q1, Q3 in Fig. 3). The stop keys are a list created by using the FAQ data to select keys that are used often in the genre, such as "PC," "finance," or "tourism," together with synonyms for those keys.

In this way, only keys suitable for the FAQ chatbot are selected to generate the truth table.

2) Variation Dictionaries

Variation dictionaries are broadly categorized into those that absorb variation in terminology specific to each FAQ and those that absorb variation in more general phrases.

The former lists variations in representation of specialized terminology used in the prepared query texts, and are registered at the discretion of the FAQ chatbot designer. The latter are generated using the large volume of language resources in the natural language interaction platform, and apply to keys with variation in phrasing or usage, such as verbs and action nouns.

Variations in general phrases includes the differences between verbs and action nouns, as well as differences in conjugated forms of verbs. Depending on the query, verbs and action nouns, such as "transmit" and "transmission," mean the same thing, so one key is used. If there was no function to absorb this sort of variation, a query with "transmit the email" would not match a query with "transmission of the email." The variation dictionary absorbs such general variation based on a synonym

dictionary with approximately 1,400 verb/action nouns.

Variation due to verb conjugations are absorbed based on patterns of variation in usage for each verb, collected from analyzing approximately 400,000 FAQ data items. Specifically, all verb conjugation patterns are normalized using the expression-normalization function of the natural-language dialogue platform [5], reducing them to the three categories of positive, negative, and completed. These three categories are the most used in FAQs. Negative expressions like "I cannot delete a photo," and completed expressions like "I accidentally deleted a photo," in particular, affect the response that needs to be presented, so they are managed as separate categories (Q5, Q6 in Fig. 3).

3. FAQ Chatbot Behavior

The basic operation of the FAQ chatbot system is shown in **Figure 4**.

The FAQ chatbot repeats the following to refine a query.

- Extract keys from the text of the user's query and update the search key list ((1) to (3) in Fig. 4).
- Search the FAQ and decide on the response content using the search key list ((4) to (5) in Fig. 4.

Each of these behaviors is described in detail below.

1) Key Extraction and Search-key List Update

The FAQ chatbot uses the phrase matching function of the natural-language dialogue platform to extract the keys from the user's initial query
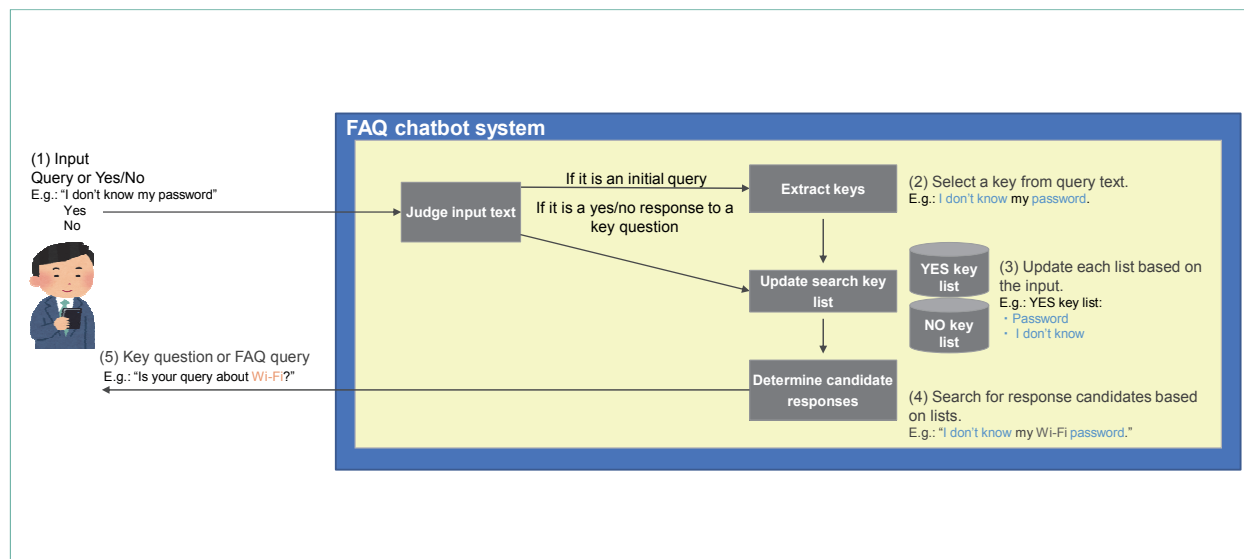
Figure 4　Basic FAQ chatbot system operation

text and matches them with the set of keys for each query in the truth table. In doing so, the user is not expected to be aware of which keys are actually in the truth-table queries, but the variation dictionary is used to absorb some of the variation in key expressions.

　The search key list is then updated based on the keys obtained from the user's utterance. The FAQ chatbot keeps two search key lists: a "YES key list" of keys associated with a positive intention, and a "NO key list" of keys associated with a negative intention. A query is identified by searching the truth table with these lists. The YES key list includes keys from the user's initial query and updates from key questions that the user has answered in the affirmative. The NO key list includes updates from key questions that the user answered in the negative.

2)　FAQ Search Using the Search Key Lists and Deciding the Response
　The FAQ chatbot uses the two search key lists

described above to search for queries stored in the truth table. Specifically, it searches for queries that have keys registered in the YES key list and do not have keys registered in the NO key list. Queries may have multiple matches, and such cases are categorized as follows.

- Complete match: A query for which all keys in the truth table entry match with the YES key list, and none match with the NO key list.
- Incomplete match: A query for which only some of the keys in the truth table entry match the YES key list, but not enough for a complete match.

　If all matches for the query are incomplete, the FAQ chatbot replies with a question regarding a search key that was not matched (This is called a key question. For example, the chat bot replies "Are you asking about Wi-Fi?" to see if the key is relevant to the user's intention). If the user answers

affirmatively to a key question, the key is added to the YES key list, otherwise it is added to the NO key list. One or the other list is updated and the FAQ chatbot repeats the search for a query in the truth table.

If a complete match is found, the system asks the user to confirm the query. This is called an FAQ query. For example, something like "Does this query express what you mean?" (followed by the selected query), confirms that it matches the user's intentions. If the user answers in the affirmative, the dialogue is finished, and the answer to the confirmed query is returned. If the user answers negatively, the next FAQ or key question is presented.

4. Performance Evaluation

4.1 Evaluation Conditions

We evaluated the performance of the FAQ chatbot with real users, and discuss the results of evaluating the proposed technology below. The FAQ chatbot was built using FAQ data generated from the manual of a particular internal system, including 306 answer texts.

We measured the accuracy of responses and the average number of interactions to reach a response. Response accuracy was the proportion of initial queries for which the FAQ chatbot was ultimately able to present a correct response, and the average number of interactions was the average number of interactions needed to reach an answer.

After performing the first evaluation, we tuned the FAQ data, made additions to the variation dictionary and stop key list, and performed a second evaluation. The first time, we used the FAQ data as-is, applied the proposed technology and conducted the evaluation, but for the second time, we applied some tuning and then re-evaluated to see how much the accuracy could be improved.

Specific details of the tuning are described next. The FAQ data used for the first evaluation was not adequately complete building the FAQ chatbot, and contained issues such as duplicate queries or answers, and errors in the responses themselves. This resulted in cases where the FAQ chatbot was unable to extract suitable keys, or where it was unable to present a correct answer, even when a query text was identified using key questions. Accordingly, we ensured that there were no duplicate queries for any of the answers, and tuned the queries to be more concise, so that keys could be extracted correctly. We also did not use a variation dictionary or stop key list for the first evaluation, so variations in the keys from user queries were not absorbed and unnecessary keys were selected, resulting in degraded performance. We added key variations to the variation dictionary for these errors, and added appropriate stop keys to deal with these issues.

In practical operation, the cost of this sort of tuning will also need to be considered. To do so, we checked the actual amount of work involved in tuning for the second evaluation. Note that for the first and second evaluations, we used different people conducting the evaluation, to ensure that the same questions were not asked for both evaluations.

4.2 Evaluation Results

The accuracy of answers, and the average number of interactions are shown in **Table 1**.

Table 1　Evaluation results

Evaluation	Response accuracy	Average no. of interactions
First	77%	4.3
Second	91%	1.8

We first discuss the results of the first evaluation. The table shows that the FAQ chatbot was able to respond to user queries with a high accuracy of 77%. In this case, the FAQ chatbot was built using FAQ data with absolutely no tuning. This demonstrates that the proposed technology is able to build a chatbot that can answer queries with nearly 80% accuracy immediately. However, the number of interactions before reaching an answer was 4.3, which is high. This was due particularly to having too few stop keys, which resulted in extraction of unnecessary keys, and subsequently, the generation of wasteful key questions.

Next, we discuss the results of the second evaluation. After tuning, the evaluation resulted in very high response accuracy of 91%. The most effective effort was expansion of the variation dictionary. This corrected errors that prevented the correct response from being presented because the keys in the query text were not in the initial query, so it was not found. Tuning the FAQ data also helped in extracting appropriate keys from the query texts and associating query texts with the correct answers, so accuracy of answers increased. By expanding the stop key list, the average number of interactions was also improved significantly, to 1.8.

The above shows that even higher accuracy can be achieved by tuning the input data.

Lastly, we discuss the amount of work required to perform the tuning. Tuning involved inspection work, analyzing the results from the first evaluation and extracting the causes of degradation in the accuracy; generating and updating FAQ data according to the inspection work; and expanding the variation dictionary and stop key lists according to the inspection work. These tasks required 4.4, 3.3, and 0.7 person-days respectively, totaling a relatively short time of approximately 8 person-days. Note that this resulted in addition of 295 entries to the variation dictionary and 53 entries to the stop key list.

5. Automatic FAQ Chatbot Building System

5.1　Overview

The FAQ chatbot provides an API for data management. The automatic FAQ chatbot building system is a Web application with an FAQ chatbot building function that uses this API. With conventional scenario-based FAQ chatbots, the developer of the chatbot service would need to design chatbot dialogs from the FAQ data to build the chatbot. With this system the developer only needs to upload the FAQ data using a Web browser. The system automatically designs the dialog content and builds the chatbot, so the chatbot can be built in much less time.

5.2 PDCA Cycle to Improve Chatbot Response Accuracy

When considering providing a real service, on-going improvement in the accuracy of FAQ chatbot responses is a major issue. With our system, building and improvements to the FAQ chatbot can be done at the same time. Improving accuracy of responses is done by repeating the process shown in **Figure 5**. First, the developer downloads the dialog history of the FAQ chatbot and users on the system. Next, the queries that the FAQ chatbot was unable to answer are found, and answers to them are added to the FAQ data. This updated data is then be uploaded through a Web browser, thereby completing update of the FAQ

chatbot. Repetition of this update process increases the accuracy of FAQ chatbot's responses. As with building the chatbot, there is no need to design dialogs when performing this type of update. The work is done mechanically, so these improvements can be done quickly.

5.3 Linking with Existing Chatbot Platforms

The ability to build and update FAQ chatbots mechanically and in a short time is an advantage, but responding with text input to repeated questions places a burden on users. To reduce this burden, the system is able to display buttons, as is done by other existing chat platforms. Specifically,
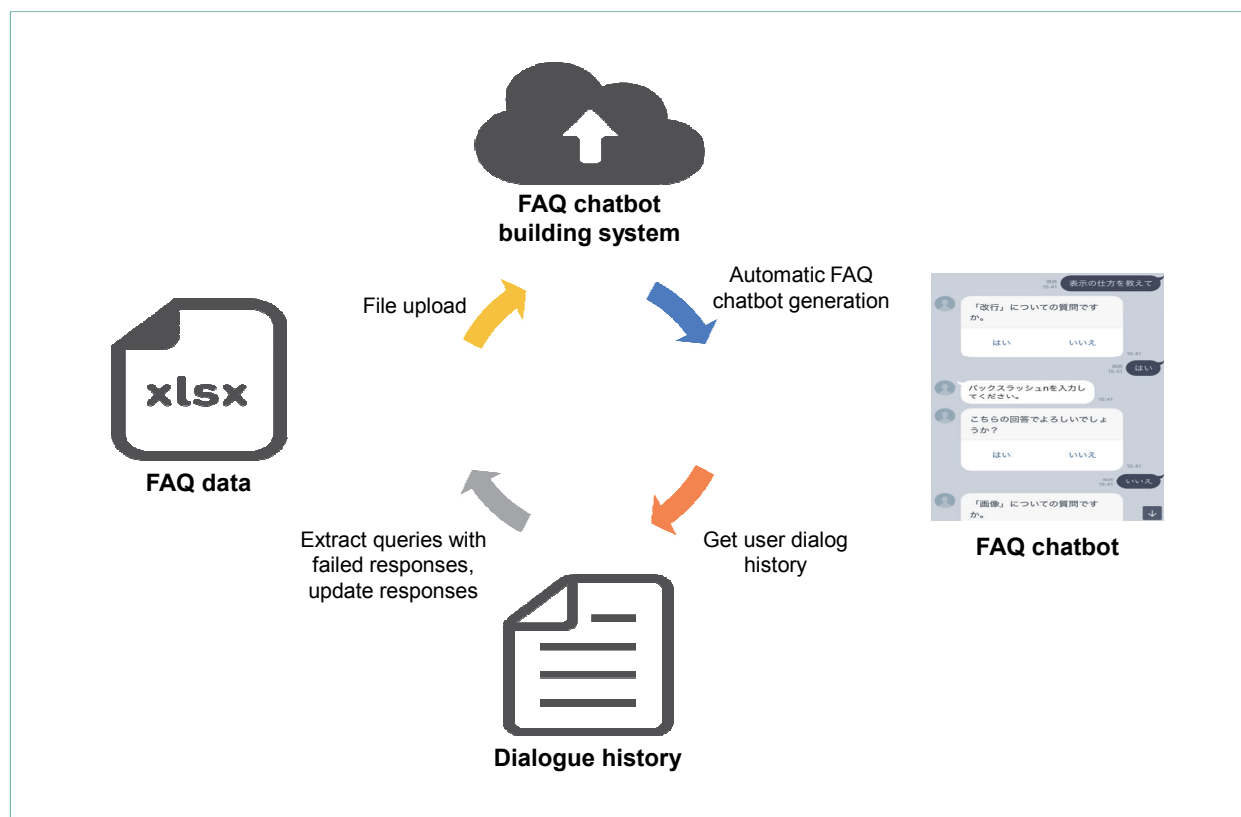

Figure 5   PDCA cycle to improve quality

when the FAQ chatbot asks the user a question, the user can respond by simply pressing a Yes or No button as they make their way to the answer.

6. Conclusion

This article has described a technology to build FAQ chatbots automatically. The proposed technology automatically builds a chatbot from FAQ data, supports processing of large volumes of FAQ data, and is able to respond to user queries very accurately through the use of automatically-generated dialogue scenarios. We evaluated the system using real FAQ data, and the FAQ chatbot achieved accuracy exceeding 90% with only eight person-days of tuning. We also gave an overview of our automatic FAQ chatbot building system, which is based on the proposed technology.

In the future, we intend to further improve accuracy and performance of the proposed technology and system, based on feedback we obtain through commercial services.

REFERENCES

[1] Repl-AI Web site.
https://repl-ai.jp/

[2] IBM: "Conversation | IBM Watson Developer Cloud."
https://www.ibm.com/watson/jp-ja/developercloud/conversation.html

[3] Microsoft: "Azure Bot Service - Chatbot | Microsoft Azure."
https://azure.microsoft.com/ja-jp/services/bot-service/

[4] Microsoft: "QnA Maker API | Microsoft Azure."
https://azure.microsoft.com/ja-jp/services/cognitive-services/qna-maker/

[5] K. Onishi et al.: "Natural-language Dialogue Platform for Development of Voice-interactive Services," NTT DOCOMO Technical Journal, Vol.17, No.3, pp.4–12, Jan. 2016.