Technology Reports

Shabette-Concier for Raku-Raku Smartphone –Improvements to Voice Agent Service for Senior Users–

We have created a new version of "Shabette-Concier for Raku-Raku" for the Raku-Raku Smartphone (F-12D) to create an environment in which a wider range of users can use smartphone functions and services. In this update, we have revised the UI based on issues selected from user-evaluation testing results. We have also developed functions linking to a user-guide application and to the FAQ on the NTT DOCOMO Web site, to respond to questions regarding use of the terminal and various services.

Service and Solution Development Department *Wataru Uchida*
Midori Onogi

1. Introduction

Smartphones with User Interfaces (UI) that use large screens and touch panels, high-speed CPUs, and are equipped to use advanced wireless networks are spreading, and new services using these features are appearing one-after-another.

Highly literate users can handle these attractive new functions and services, but users that are not necessarily proactive about new functionality, including seniors, have also begun to use smartphones, and there is increasing need for functions and services that are easy to use for these users.

We have made improvements to "Shabette-Concier for Raku-Raku Ver.

1.0," the speech-agent service being provided for the Raku-Raku smartphone (F-12D), and began providing "Shabette-Concier for Raku-Raku Ver. 1.1" in January, 2013. In development of these improvements, we followed the two policies described below, in an attempt to create an environment enabling a wider range of users to use the features of advanced smartphones.

Policy (1): Enable more natural smartphone operation using speech operations. To accomplish this, change the UI to make it more intuitive.

Policy (2): Support resolving questions and issues regarding use of the smartphone. To accomplish this, add functions to answer questions regarding terminal functions and services.

For Policy (1), we conducted UI evaluation tests with elderly users and designed and implemented the UI to resolve selected issues from the results. For Policy (2), we implemented functions linking to the Frequently-Asked-Questions (FAQ) part of the customer support section of the NTT DOCOMO Web site [1], and to the user guide application [2].

In this article we describe these two initiatives.

2. Changes for Intuitive UI

To determine what aspects of Shabette-Concier operation are difficult to operate or understand, we conducted evaluation tests with senior users, in

cooperation with the ICT Design Center of the NTT Service Evolution Laboratory.

2.1 Test Overview

Six male and female subjects in their 50's and 60's with no experience using Shabette-Concier participated in testing. In the tests, subjects were given tasks related to various functions, which they attempted to perform using the version of the application before improvements. Examples include map searches to "Find destination XX," or using the email, weather or restaurant search functions.

2.2 Test Results

Three issues were selected from the test results.

Issue (1): Users did not understand the basic operation of the application.

Using voice commands to call functions or to perform searches requires pushing the microphone button, but we observed users attempting to give voice commands without doing so. This is likely because the users did not understand basic operation of the application.

Issue (2): Users could not think of what to say.

We observed cases in which, even though they knew what to do, they pushed the button before deciding what to say and got stuck in thought. This was likely because they did not have enough knowledge about Shabette-Concier functions or how to invoke them with voice commands.

Issue (3): Users did not understand what to do when multiple utterances were required.

Some functions, such as the Transit Guide, email, telephone and memos, provide a user interface that requires consecutive utterances. For example, to use email, the user says "Send an email." The screen transitions to the email creation screen, and the application prompts for speech to fill in the recipient and main text. However, some subjects made errors in their voice responses, such as speaking the text when prompted for the recipient. The Shabette-Concier UI was designed to operate with voice responses from the agent and user utterances, but follow-up is required if the user does not listen to the voice responses, or cannot immediately understand them.

2.3 Resolutions to Issues

The following UI changes were made to resolve the issues given in Section 2.2.

Change (1): Clarify operation by changing the screen designs

For the first issue, not knowing how to use the functions, we changed the screen design show clearly the sequence of pushing the microphone button, speaking and displaying the result, as shown in **Figure 1**. We increased the size of the microphone button on the top screen and made it flash to prompt the user to push it (Fig.

1(a)). We show a large dialog box on the speech recognition screen to indicate clearly that an utterance is required (Fig. 1(b)). We also adjusted the result screen to prevent user operation errors, removing the microphone button if further utterances are not required (Fig. 1(c)).

Change (2): Display utterance examples

For the second issue of not knowing what to say, we displayed typical operations, with examples of utterances to invoke them, on the screen as shown in **Figure 2**. If the microphone button is not pressed within a set time on the application top screen, the example utterances are displayed.

Change (3): Use button wording to prompt users to additional utterances

For the third issue, of not understanding operation when consecutive utterances are required, we changed the wording on buttons operated by the user.

An example of the email recipient input screen is shown in **Figure 3**. As shown in Fig. 3(b), when entering the recipient, the wording on the microphone is "Say the recipient," clearly indicating what must be said.

3. Functionality to Answer Questions Regarding Functions and Services
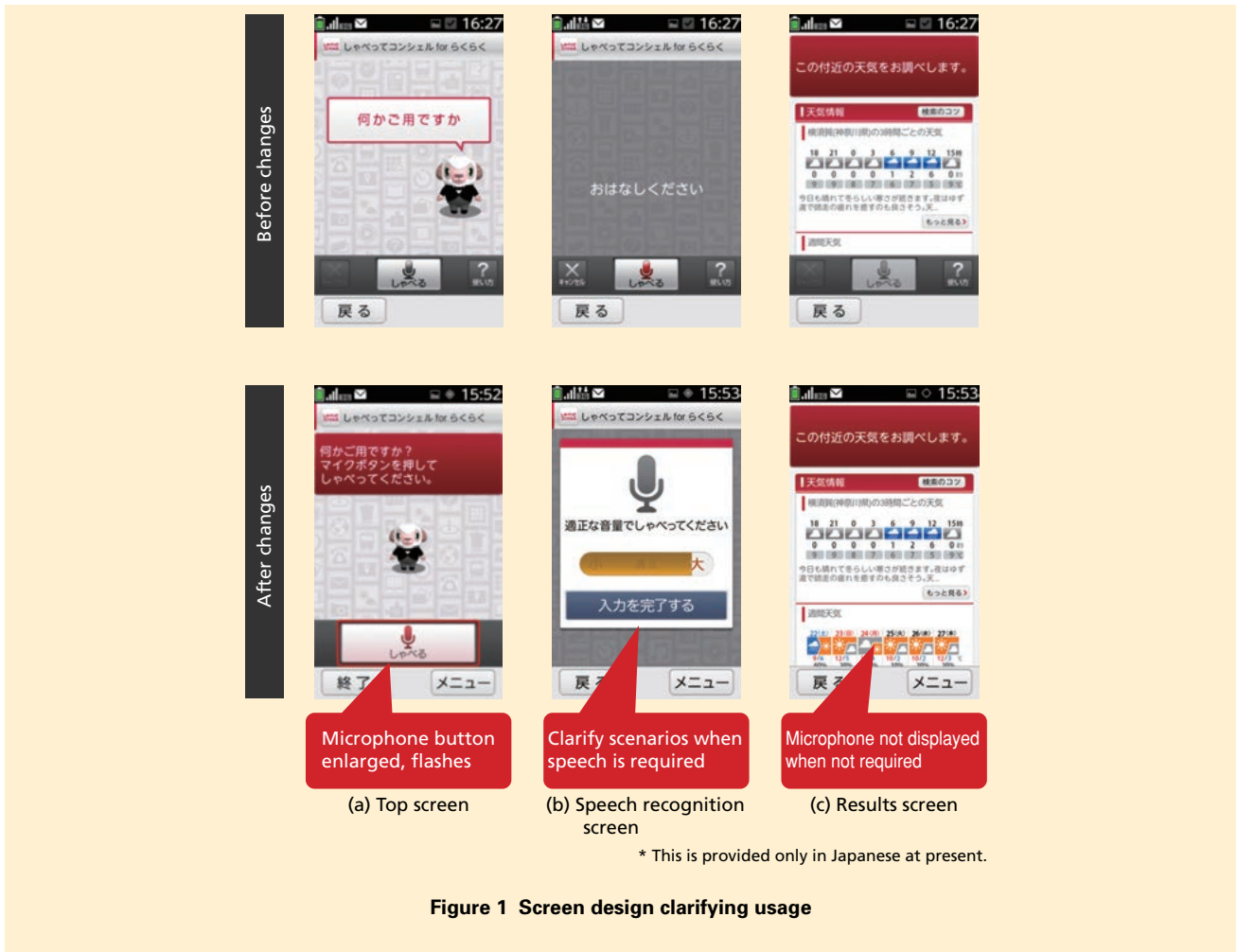
3.1 Linking with FAQ

There is FAQ section on the

(a) Top screen

(b) Speech recognition screen

(c) Results screen

Microphone button enlarged, flashes

Clarify scenarios when speech is required

Microphone not displayed when not required

* This is provided only in Japanese at present.

Figure 1 Screen design clarifying usage

* This is provided only in Japanese at present.

Figure 2 Display screen for speech example

NTT DOCOMO Web site, providing answers to some of the questions commonly received from users. We have various initiatives to provide customer support for users, but the FAQ gathers a
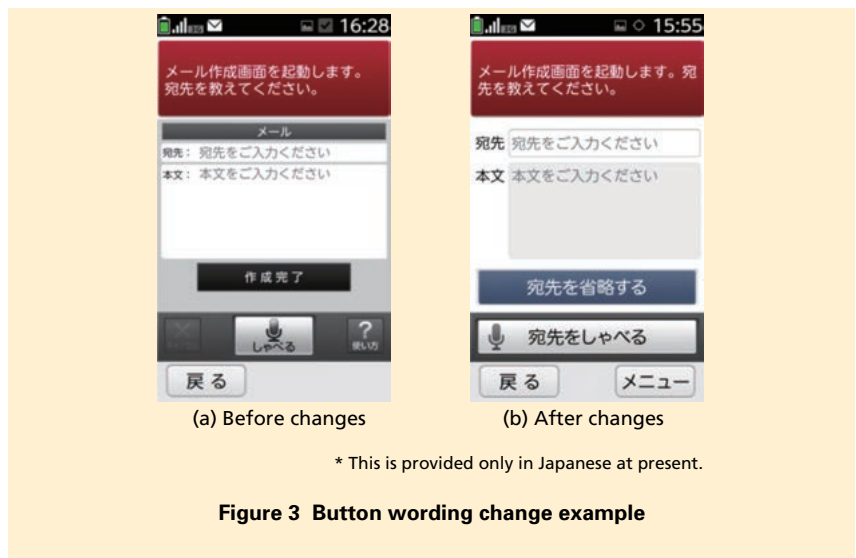

(a) Before changes

(b) After changes

* This is provided only in Japanese at present.

Figure 3 Button wording change example

variety of information in one place, such as fees and how to receive services, in the form of a Web site. This enables users to obtain answers immediately, so we have implemented links to the FAQ to support users in resolving any doubts or issues related to our services.

A screen shot of linking to the FAQ is shown in **Figure 4**. On the top screen (Fig. 4(a)), if a question regarding mobile phone services is entered, such as "How do I suspend services?," an FAQ search result page is retrieved and displayed on the screen (Fig. 4(b)).

The operational flow is shown in **Figure 5**. It is largely the same as for earlier versions of Shabette-Concier [3].

When a service-related question is asked, the information is transmitted to the Shabette-Concier server on the cloud (Fig. 5(1)). The Shabette-Concier server has speech recognition and semantic analysis functions, and the speech recognition function converts the speech to text (Fig. 5(2)). The semantic analysis function determines the intention of the text and decides on the next task (terminal function or ser-
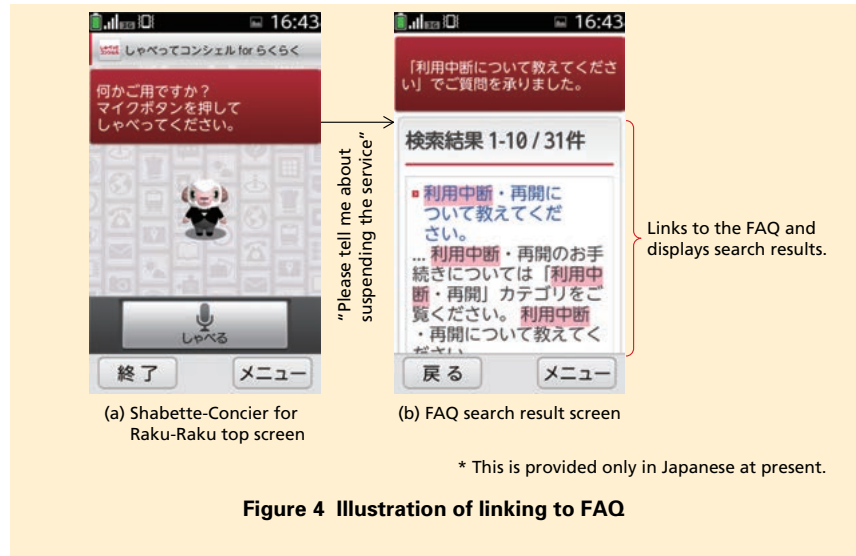

(a) Shabette-Concier for Raku-Raku top screen

(b) FAQ search result screen

"Please tell me about suspending the service"

Links to the FAQ and displays search results.

* This is provided only in Japanese at present.

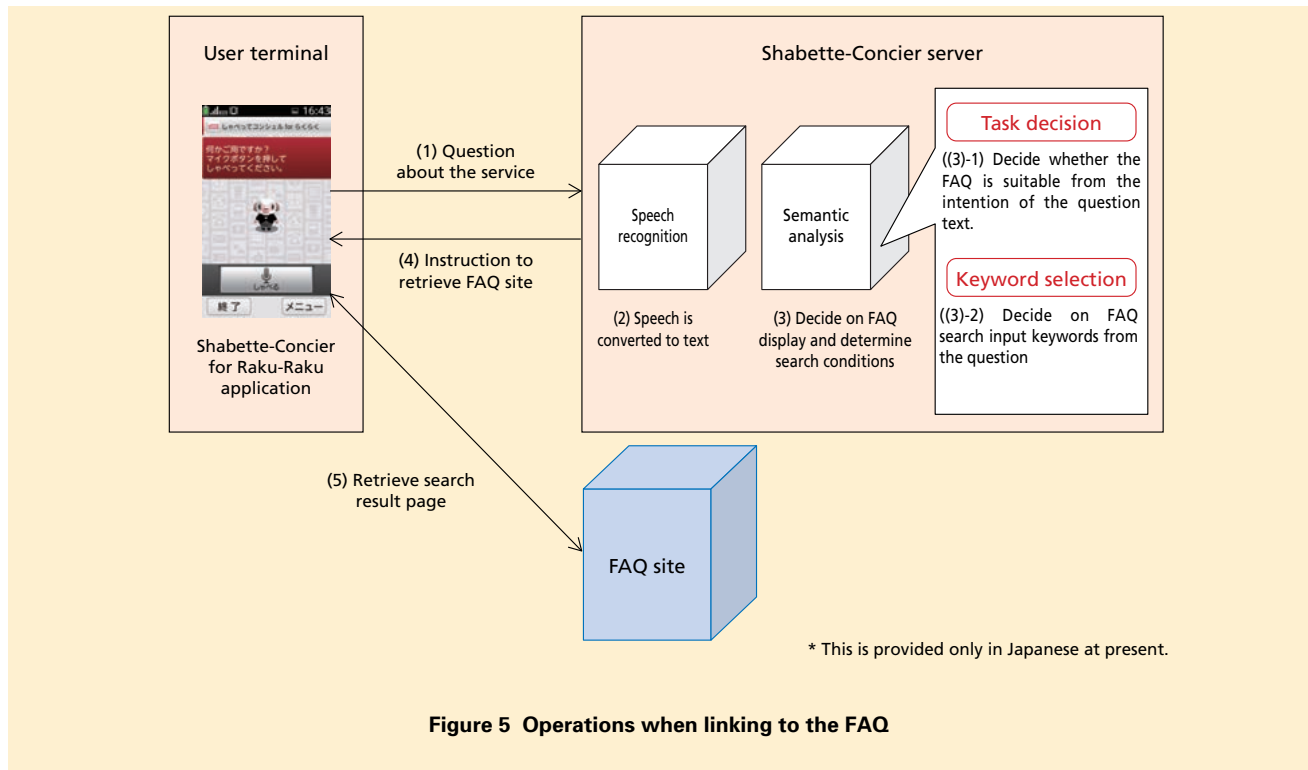**Figure 4 Illustration of linking to FAQ**

Figure 5 Operations when linking to the FAQ

vice) and the search conditions (Fig. 5(3)). An operating instruction is passed to the terminal application (Fig. 5(4)), and the FAQ search result is retrieved (Fig. 5(5)).

Semantic analysis consists of two functions: task decision and keyword extraction. Task decision uses a machine learning[*1] method called multi-class classification[*2], to decide on a suitable service or function corresponding to the user's request, based on the user's utterance converted to text. With the current development, we modified the multi-class classification operation to select the FAQ as the applicable task when questions related to services such as rates and procedures as well as tasks that we have covered earlier, such as transit information (Fig. 5(3)-1).

The keyword selection function configures the search conditions required to obtain an appropriate result for the various services and functions selected by the task decision process. We added a function that configures the URL and the search keywords for the FAQ. We extract words appearing in the user's utterance for search keywords, including the names of services, and words, such as "Charges," that are expected to appear with these service names (Fig. 5(3)-2).

3.2 Linking to the User Guide

The User Guide is an application that allows users to check how to use or configure their terminal. NTT DOCOMO

provides a variety of information related to using and configuring terminals on our Web site. The User Guide provides the most comprehensive information that can be consulted. Search functions are provided, so the required information can be displayed instantly. We implemented links to the user guide in order to support users in resolving their issues and concerns related to terminal functions and configuration. An illustration of linking to the User Guide is shown in **Figure 6**. When a question is asked regarding use or configuration of functions, such as "Configuring silent mode," the User Guide is launched with search keywords (Fig. 6(b)). The User guide is searched within the application internally using the keywords, and if there is one result, the page is displayed directly (Fig 6(b-1)). If there are multiple results, they are displayed in a list (Fig. 6(b-2)).

The operational flow is shown in **Figure 7**. As with linking to the FAQ, we added to the functions for task decision and keyword selection, enabling the User Guide application to be launched.

For task decision, we modified the classification model to select the User Guide as the applicable function when questions related to terminal use were asked.

Next, we describe keyword selection. Natural utterances usually contain multiple words that will be effective as search keywords, and there are various words that can be used to express the same meaning. When all of these words are used to search the contents of the User Guide, often no search results are returned, so users find it difficult to use. We implemented keyword selection to select and reject keywords from the utterance, and to transition to the User
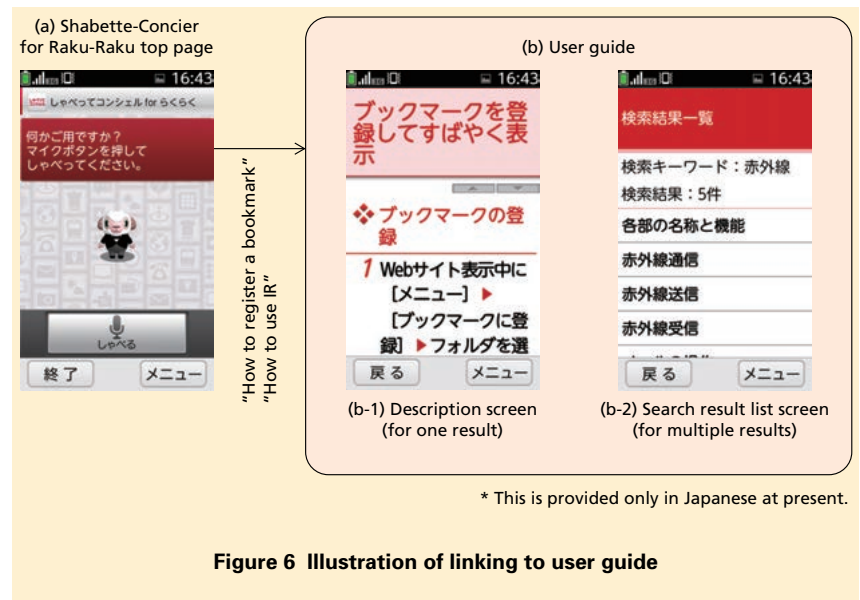

(a) Shabette-Concier for Raku-Raku top page

"How to register a bookmark"
"How to use IR"

(b) User guide

(b-1) Description screen (for one result)

(b-2) Search result list screen (for multiple results)

* This is provided only in Japanese at present.

Figure 6 Illustration of linking to user guide

Guide only if there was at least one page with similar content to display, even if no content in the User Guide completely answers the question. The process of keyword selection is shown in **Figure 8**. To estimate the number of search results for content in the User Guide, a list indicating the correspondence between each page and the number of search keywords on it is maintained on the Shabette-Concier server. The set of keyword candidates selected from the utterance is compared with the index of the User Guide to estimate the number of search results. If the number of results is zero, one keyword is eliminated and the number of search results is estimated again with the remaining keywords. This is repeated until there are one or more search results, and the final remaining set of keywords is used
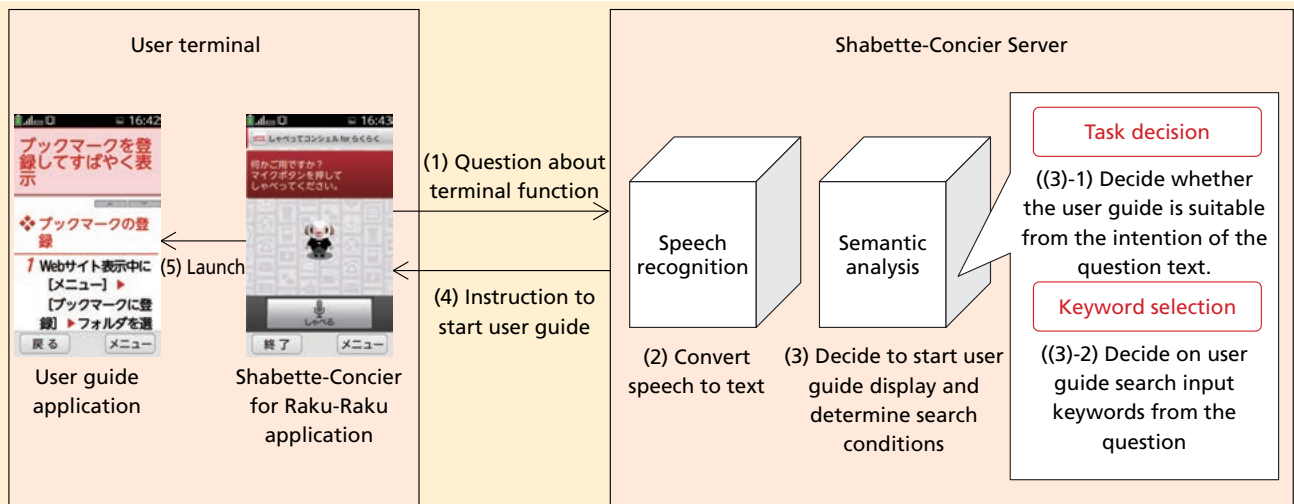
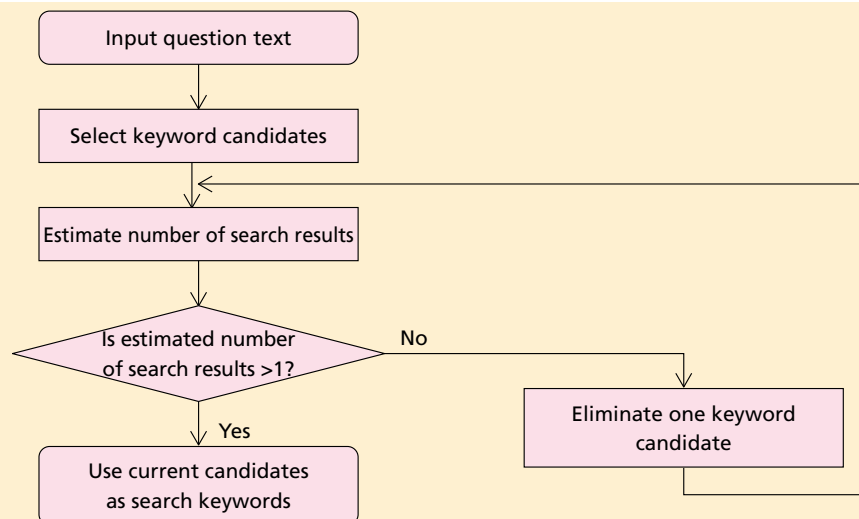Figure 7 Operation when linking to the user guide

Figure 8 Keyword selection process

when launching the guide.

4. Conclusion

We have described a new version of Shabette-Concier for the Raku-Raku Smartphone (F-12D), aimed at helping a wider range of users to use smartphone functions and services.

In the future we will continue initiatives to help users use smartphones more easily, including surveys of smartphone use after release of this product.

REFERENCES

[1] NTT DOCOMO: "Frequently Asked Questions (FAQ)."
http://faq.nttdocomo.co.jp/faq/

[2] NTT DOCOMO: "User's Guides | Customer Support | NTT DOCOMO."
http://www.nttdocomo.co.jp/english/support/trouble/manual/download/guide/index.html

[3] T. Yoshimura: "Shabette-Concier Service realized by Natural Language Processing," IPSJ SIG Notes, 2012-SLP-93(4), pp. 1-6, 2012 (in Japanese).