SARA – Singapore's Automated Responsive Assistant for the Touristic Domain

Andreea I. Niculescu, Rafael E Banchs, Ridong Jiang, Seokhwan Kim, Kheng Hui Yeo, Arthur Niswar

Institute for Infocomm Research (I^2R)

{andreea-n, rembanchs, rjiang, kims, yeokh, aniswar}@i2r.a-star.edu.sg

Abstract

In this paper we describe SARA - Singapore's Automated Responsive Assistance – an Android mobile phone application for touristic information in Singapore. The application provides information about local attractions, restaurants, sightseeing direction and transportation services. Using a GPS integrated module SARA is able to detect the user's location on a map providing orientation and direction help. Users can use speech, text or scanned QR code to interact with SARA. Input/output modalities for SARA include natural language in form of text or speech. A short video about the main features of our Android application can be seen here: http://vimeo.com/91620644. Currently, SARA supports only English, but we are working towards a multi-lingual support. For test purposes we created a web version of SARA that can be tested for Chinese and English text input at: http://iris.i2r.a-star.edu.sg/StatTour/.

Index Terms: multi-modal dialogue system, mobile application, tourist information system, natural language interaction

1. Introduction

According to the Singapore Tourism Board the number of tourist visiting Singapore is each year constantly increasing [1]. In 2013 Singapore reached a record number of 15.5 million visitors, which represents the highest tourist rate in the past decade. The increasing number of tourist requires additional resources not only in terms of accommodation, board and transportation, but also in terms of touristic guidance and sightseeing information. SARA is built as a response to this increasing demand for touristic information offering a comfortable solution for those who want to explore the city by themselves and have no human guide around.

Currently, there are other commercial applications in Singapore, such as "Your Singapore Guide" and "Your Singapore Navigation" that provide guidance and information for tourists. However, these applications have a lower degree of interactivity, in the sense that they don't support speech input/output or question & answer (QA) style of interaction as SARA does.

2. Interface design and user interaction

As shown in figure 1 the application has two main screens: a login screen and a dash board screen containing a user input field, an avatar and four navigation options: 1) an interactive map display (with zoom in, zoom out , scroll, move and rotate functions) where the user can see how to get from his current location to the target location; 2) a web browser that opens a page where information related to the user query is presented; 3) a picture collection also related to the user query; 4) a QR scan option that offers users the possibility to scan a QR code and to display the decoded result in the browser (see figure 1).

The interaction style with the system is multimodal: users can talk or type the question and receive information from SARA in spoken, written and graphical form.

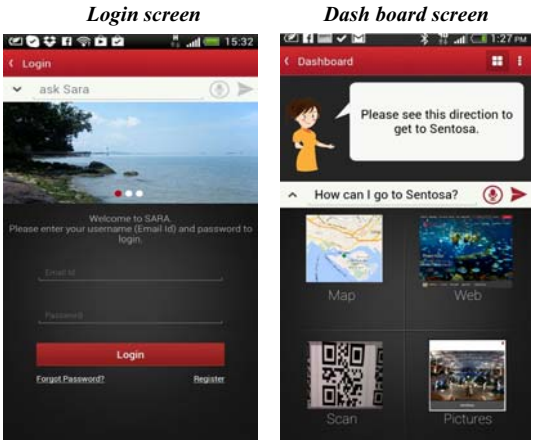

Figure 1: *SARA main screens*

Users can request different type of information, such as how to get to a particular place in Singapore, detailed information about sightseeing, currency exchange rate, weather, restaurant recommendations, museum, ticket prices, opening times, hotels addresses and telephone numbers. Additionally, users can request the system to make reservations, call for taxis and send SMS. Telephone numbers are to be provided by the server.

3. System architecture

SARA is based on client-server architecture, as presented in figure 2. The client communicates with the server using JSON object protocol.

Figure 2: *SARA client server architecture*

3.1. Server side

On the server side SARA uses a hybrid approach to **natural language understanding** (NLU). The component transforms the recognition output passed from the native Android's **speech recognizer** into a semantic representation using rules and statistical models which were built based on data collected for Singapore touristic domain. A total of 40 hours of human-human dialogue data for English and Chinese was collected. The dialogues usually included sequences of requests from the users and provision of information by the tourist guides, as

well as suggestions and recommendations. Dialogues were manually annotated on three levels of semantics: word, utterance, and dialogue segment levels.

The **dialogue manager** (DM) implemented in SARA also incorporates two different strategies: one is a rule-based approach using a set of manually defined heuristics for determining proper system actions to each input. The other strategy is an example-based method using an index which contains input-response pairs collected from Singapore-related articles in Wikipedia. For each input from a user, the most similar example in this index is selected based on cosine similarities between term vectors with TF-IDF weights [2]. While the rule-based approach is mostly used for handling goal-oriented scenarios, the example-based approach focuses more on general question answering scenarios.

Finally, the **natural language generation** component uses a template-based approach to generate an appropriate response to user's query and passed it to the native Android's **text to speech engine**.

To interconnect all these components we use the APOLLO framework [3], a component-pluggable **dialogue platform** that allows the interconnection and control of the different independent components used in the implementation of the system, such as dialogue, input/output, backend and task manager components**.** Apart from specifically designed plugins, the platform also allows for socket communication using TCP-IP protocols. The interconnection among the different components within the platform and the control of the information flow can be programmed by using the APOLLO's XML-based scripting meta-language.

3.2. Client side

On the client-side, SARA has a total of nine software modules that enable different system functionalities. Below, we describe briefly each of these modules:

- **User Registration and Management** - this module is responsible for the user registration and log-in. When the application is used for the first time, the module creates a user profile containing the person's name and email. For subsequent logins the system retains user's credentials, i.e. the user doesn't need to re-input password and email address. To update or modify his/her profile the user can access a set-up page directly from the application.
- **Map Display and Interactivity Module -** this module is responsible for displaying the current location of the user - assuming the GPS is enabled - as well as the locations of interest retrieved by the system. The map is able to display routes as computed by the API and also to link map locations to URL addresses provided by the server.
- **GPS Tracking Module** - this module uses all on-phone necessary resources to determine the exact location of the user. Position coordinates are communicated to the served-based information services for geo-localization purposes. If the user chooses not to enable GPS/Location tracking module, the system provides relevant answers without taking the geographic context into consideration.
- **On-phone Speech Capabilities Module** - this module exploits Android's native ASR and TTS resources to convert speech to text and text to speech for the different input/output modules involved in the interaction.
- **Avatar Display Module** - this module uses image functionalities to display an avatar on the phone screen (see figure 1). The avatar responds to instructions and commands provided by the system performing different activities (thinking, smiling, searching, answering, asking questions, etc.). The avatar also displays a text-bubble in which textual and hyper-textual information is provided to the user.
- **QR Code Reader Module** - this module is responsible for scanning and reading QR codes. System generated QR codes are identified by a specific prefix that is known by the application. For non-system generated QR codes (any other QR code that does not belong to the SARA platform), the mobile application is able to decode and execute the corresponding default action (display the text, browse the corresponding URL, etc.)
- **Integration with Phone Call and SMS -** this module uses on-phone capabilities to support basic functionalities, such as placing a phone call, or sending an SMS.
- **Internet Browsing Module** - this module uses Android resources to display the contents of URLs provided by the server-based information system or the QR code reader. It has browsing capabilities in the sense that once an URL has been displayed, the user can click on the link and start navigating in the internet.
- **Communication with Server Module -** this module is responsible for sending and receiving all necessary information between the client (mobile application) and the server-based information system (APOLLO). It uses the JSON Object protocol to pass variable information between the server and the client.

4. Conclusions and future work

In this paper we presented SARA, an automatic touristic information system for the city of Singapore. The system is currently implemented as an Android mobile phone application and as web-based system.

In the future, we are planning to extend the system with multi-lingual capabilities - at the moment the mobile application uses only English while the web system can be used for both English and Chinese. Also, we are planning to create an iOS and Windows mobile version of the application, improve the current user interface, enlarge the data base coverage with additional touristic information, and integrate with our digital receptionist and restaurant recommendation systems.

Additionally, we plan to conduct a formal system evaluation by engaging actual users to interact with the system and use it for accomplishing some tasks while moving around Singapore. The result of this study will help to improve future versions of SARA regarding both performance and usability.

5. References

[1] http://www.stb.gov.sg/

[2] Banchs, R. E., & Li, H. IRIS: a chat-oriented dialogue system based on the vector space model. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 37-42). Association for Computational Linguistics, 2012

[3] Jiang, R.D., Tan, Y.K., Limbu, D.K. and Li, H., "Component pluggable dialogue framework and its application to social robots ", in *Proc. Int'l Workshop on Spoken Language Dialog Systems*, 2012