# SARA: Singapore's Automated Responsive Assistant, a multimodal dialogue system for touristic information

Andreea I. Niculescu, Ridong Jiang, Seokhwan Kim,
Kheng Hui Yeo, Luis F. D'Haro, Arthur Niswar, Rafael E. Banchs

Institute for Infocomm Research, 1 Fusionopolis Way, Singapore 138632
{andreea-n,rjiang,kims,yeokh,luisdhe,aniswar,rembanchs}@i2r.a-star.edu.sg
http://hlt.i2r.a-star.edu.sg/staff/

**Abstract.** In this paper we describe SARA, a multimodal dialogue system offering touristic assistance for visitors coming to Singapore. The system is implemented as an Android mobile phone application and provides information about local attractions, restaurants, sightseeing, direction and transportation services. SARA is able to detect the user's location on a map by using a GPS integrated module and accordingly can provide real-time orientation and direction help. To communicate with SARA users can use speech, text or scanned QR code. Input/output modalities for SARA include natural language in form of speech or text. A short video about the main features of our Android application can be seen at: http://vimeo.com/91620644. Currently, the system supports only English, but we are working towards a multi-lingual input/output support. For test purposes we also created a web version of SARA that can be tested for Chinese and English text input/output at: http://iris.i2r.a-star.edu.sg/StatTour/.

**Keywords:** multi-modal dialogue system, mobile application, tourist information system, natural language interaction

## 1    Introduction

The tourism industry is considered to be one of the biggest economic sectors generating an estimated eleven percent of the global domestic product [1]. Traditionally, tourists rely on using static information such as book guides, printed maps, and informative flyer material to locate points of interest. This way of getting to know a new place is considered to be useful although the source of information might be outdated.

Given the grow of the tourism sector and the prevalent usage of smart phones nowadays, the ICT industry has started to focus on creating tools to help travelers to orient and move smoothly in new environments [2]. Such tools include interactive maps, reservation and automatic check-in systems, personalized recommendation applications, travel guide assistants, etc.

## 1.1 Local context

According to the Singapore Tourism Board the number of tourists coming to Singapore is yearly increasing [3]. In 2013, the total number of tourists hit a new record: 15.5 million people came to visit the island, which represents the highest tourist rate in the past decade. This increasing amount of visitors requires additional information resources in terms of accommodation, board, food and beverage, transportation and touristic guidance.

In this context, SARA was built as a response to the growing demand for personal touristic assistance and offers a comfortable solution for those who want to explore the city by themselves and have no human guide around.

Currently, there are several other local touristic applications available at Google Play Android application store, such as *Your Singapore Guide*, *Your Singapore Navigation*, *Singapore Guide* etc., that provide information for tourists. However, these applications have a lower degree of interactivity, in the sense that they support neither speech input/output modalities nor question & answering (QA) style of interaction as SARA does, presenting the information as any other descriptive internet web page.

## 1.2 Related work

Generally, most of the applications concerning touristic guidance offer services in terms of either navigation or exploration. Applications meant for navigation include examples such as *Siri*, *Google Maps navigation*, *Sygic*, etc. On the other hand, applications offering exploration services, which present descriptive information for the user to read, include examples such as *Wikihood*, *Triposo*, *Your Singapore Guide*, etc. No doubt, these applications provide highly valuable information and help; but, for the users, navigation and exploration make more sense if performed simultaneously, i.e. without having to switch between applications [10]. Therefore, there is a need of creating a tool which can handle both of these categories.

Additionally, users need an application that does not distract them from their current task – which is walking on the street, looking around etc. In this sense, it is important to mention that most of the current applications offering navigation and exploration services use visual information displayed on the mobile phone screen. To address this problem many natural language systems were developed. Examples include pedestrian navigation systems [4] [5] [6] [7] [8] [9] [10], city guidance systems [11] [12] and QA systems [13].

Our system SARA complements the work mentioned above reuniting navigation and exploration in a single application that combines visual information with speech to create a more natural way of interaction. In the future, the system will also support multi-lingual services – at the moment still under development – which represents an additional feature that distinguishes SARA from all the other applications.

## 2    System architecture

SARA is based on a client-server architecture, as presented in figure 1: the client mobile application communicates with the server using a JSON Object based protocol.



**Fig. 1.** SARA client-server architecture

### 2.1    Server side

On the server side, SARA uses a hybrid approach to natural language understanding (NLU) and dialogue topic tracking [14]. The components transform the recognition output passed from the native Android's speech recognizer into a semantic representation using rules and statistical models. The models were trained by using data collected for the Singapore touristic domain. A total of 40 hours of human-human dialogue data for English and Chinese was collected. The dialogues include sequences of questions asked by visitors and answers containing explanations provided by tour guides. The dialogues were manually annotated on three levels of semantics: words, utterances, and dialogue segments.

The dialogue manager (DM) implemented in SARA incorporates two different strategies: a rule-based approach using a set of manually defined heuristics for determining proper system actions to each input; and an example-based method using an index filled with input-response pairs collected from Wikipedia articles related to Singapore. For each user input the most similar example in the index is selected based on cosine similarities between term vectors with TF-IDF weights [15]. While the rule-based approach is mostly used for handling goal-oriented scenarios, the example-based approach focuses more on general question answering scenarios [16].

Finally, the natural language generation component uses a template-based approach [17] to generate an appropriate response to the user query. Once generated, the response is then passed to the native Android's text to speech engine.

SARA's main components are linked together by using APOLLO [18]. As shown in figure 2, APOLLO is a component-pluggable dialogue platform that allows the

interconnection and control of the different interdependent components used in the implementation of the system, such as:
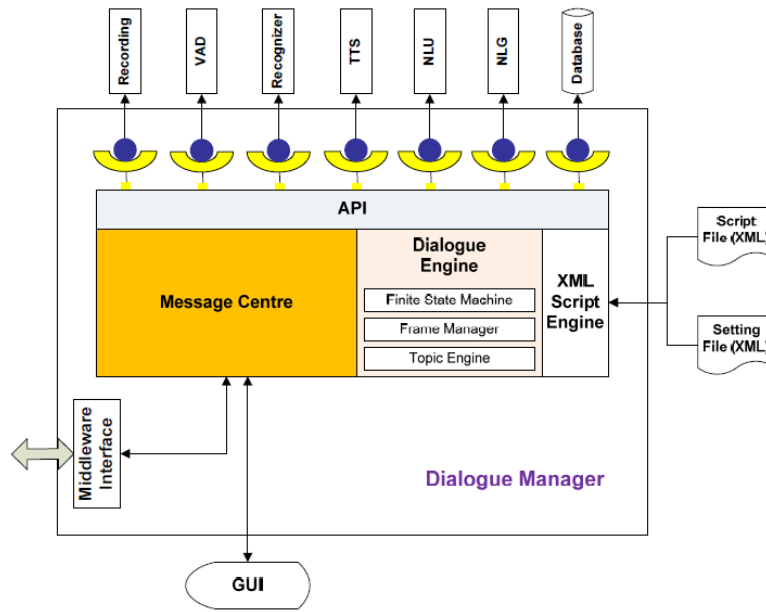


**Fig. 2.** APOLLO dialogue platform: basic architecture

- *Dialogue components*, which include the main components of dialogue engines, such as automatic speech recognition, natural language understanding, natural language generation and speech synthesis among others.

- *Input/output components*, which provide the means for integrating different input and output utilities including speech, text, image and video into the platform

- *Backend components*, which allow for the integration of different information sources, such as databases, web crawlers and browsers, rule and inference engines, as well as the different user profiles, short and long term memory contents stored by the system.

- *Task manager components*, which include all individual dialogue management engines that coexist within the platform in a concurrent manner

Apart from specifically designed plugins, the platform also allows socket communication using TCP-IP protocols. The component interconnection within the platform, as well as the information flow control can be programmed by using the platform's XML-based scripting meta-language.

4

## 2.2    Client side

On the client-side, SARA has 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 user registration and log-in. When the application is used for first time, the module creates a user profile containing the person's name and email. For sub-sequential logins the system retains the user's credentials, i.e. the user doesn't need to re-input the 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 user location – assuming the GPS is enabled – as well as other locations of interest retrieved by the system. The map is able to display routes as computed by the API and 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*. This module exploits the Android's native ASR and TTS resources to convert speech to text and text to speech.

- *Avatar display module*. This module uses image functionalities to display an avatar on the phone screen (see figure 3). The avatar responds to instructions and commands provided by the system performing different activities, such as thinking, smiling, searching, answering, asking questions, etc. The avatar is accompanied by 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 the application knows. For non-system generated QR codes, i.e. 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 any link and start navigating in the Internet.

- *Client-server communication 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 a JSON Object based protocol to interchange variable information between the server and the client.

## 3    User interface design

As shown in figure 3, the application has two main screens: a login screen and a dashboard/home screen. The dashboard screen contains a user input field, an avatar display and four navigation option buttons related to:
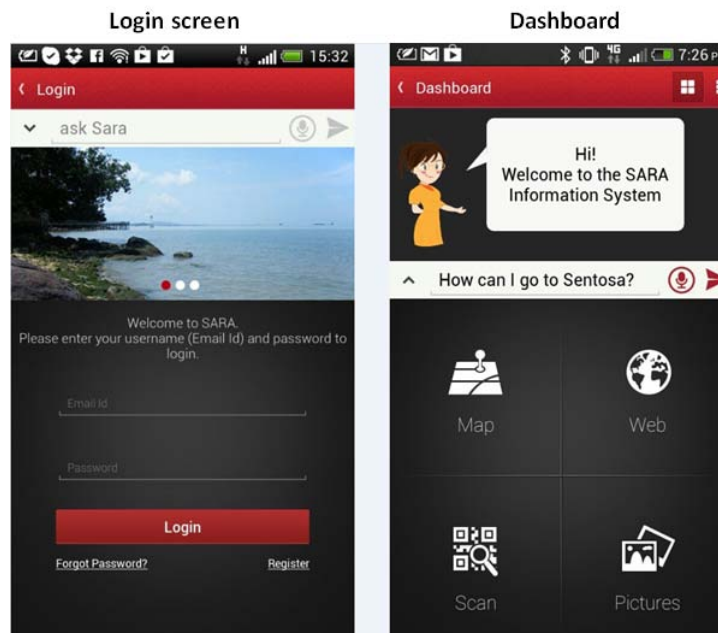


**Fig. 3.** SARA mobile app: main screens

- An interactive map in which users can see locations of interest and get directions.

- A web browser that, when available, opens a webpage relevant to the query.

- A picture gallery, where pictures related to the query are stored and displayed.

- A scanner that offers users with the possibility of scanning either generic or specific system generated QR codes.

By tapping on the buttons the user can navigate at any moment into one of the four specific function screens presented in figure 4.
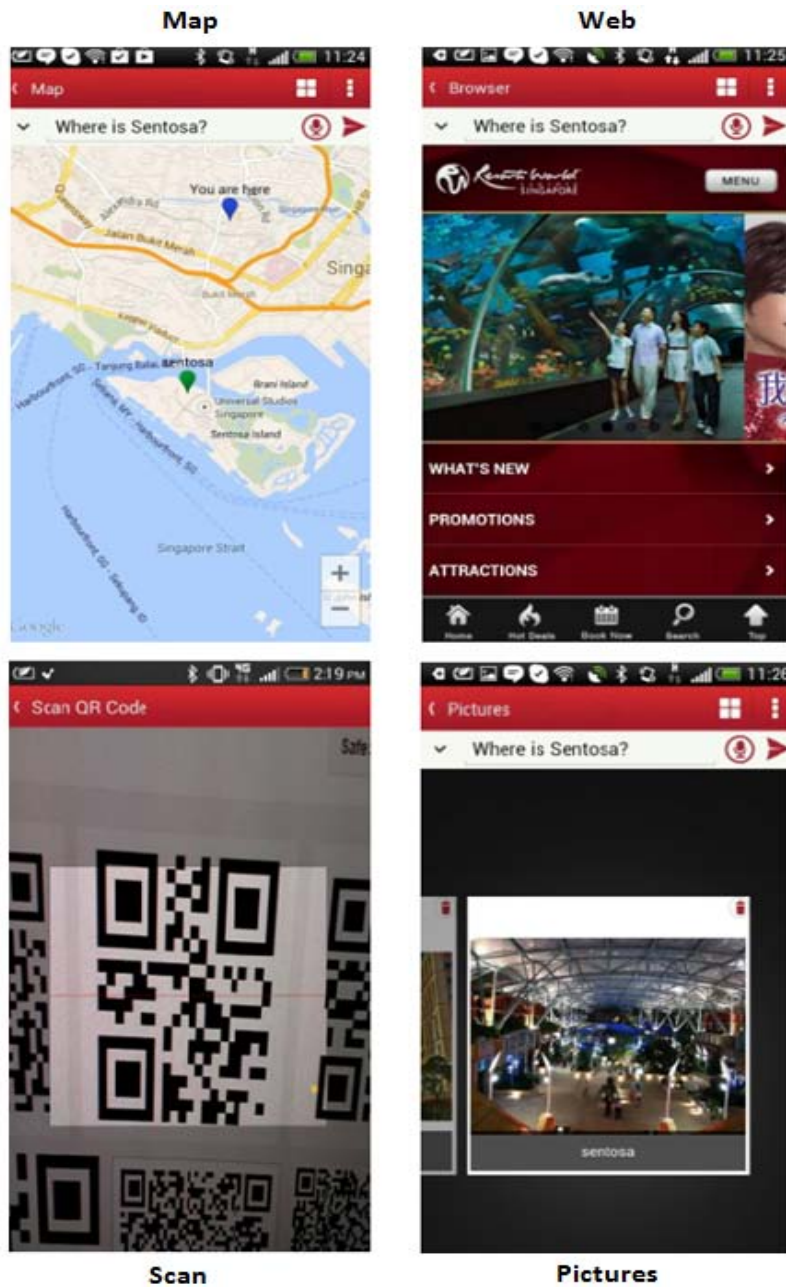


**Fig.4.** SARA mobile app: specific function screens

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, i.e. the system answers the user's question verbally accompanying the speech with maps, images or web browser information.

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 hours, hotels addresses, telephone numbers, etc. Additionally, users can also request the system to place a telephone call or send a text message (SMS). The telephone numbers are to be provided by the server.

The maps, images or web browser information are automatically loaded and presented on the output screen once the answer is fetched from the database. The user has the option to check additional information related to his query by going back to the main screen and tapping on the corresponding navigation button – for example, if the user asks how to go to Sentosa island, starting from his current location, the system explains directions showing a map; alternatively, the user can go to the specific function screens to check pictures or web information related to Sentosa island.

An interesting strategy is currently under development to handle situations when no database records match a particular question. The system informs the user it has no answer and asks if the user can help to teach something about the topic requested. The user has the option to rephrase the question or to input new information that can be later used to populate the database. In this way, two important objectives can be achieved: first, through rephrasing the impact of out-of-vocabulary words can be eliminated and thus, an answer can be found; and second, the database can be easily enlarged with new entries, i.e. without having to perform additional data collections that are usually costly and time consuming.

## 4 System evaluation

In this section we describe a user evaluation we conducted on SARA. For this evaluation five different use cases were designed, each one containing several specific tasks. These use cases are presented in table 1. Each participant was requested to complete three scenarios. Since the first three scenarios (Marina Bay, Sentosa and Orchard Road) are conceptually very similar, each participant was asked to select one out of these and additionally to complete also the last two scenarios from the table.

A total of 10 participants from our research lab participated in the evaluation. After completing all three scenario tasks, the participants were asked to fill in an evaluation questionnaire. The questionnaire was developed in accordance with the recommendations of the ISO/IEC 9126 quality standard model and, accordingly, it focuses on system "usability", "reliability" and "functionality" [19]. For each of the questions presented in the questionnaire, each participant was requested to provide one of the following five categorical scores: strongly agree (SA), agree (A), neutral (N), disagree (D), or strongly disagree (SD).

| User scenario | Specific Objectives |
|---|---|
| Marina Bay | Find a convenient MRT station to reach the Marina Bay area |
| | Find a suitable place for having lunch |
| | Ask for places of interest in the Marina Bay area |
| | Select one place of interest and ask questions about it (including location) |
| Sentosa | Find and select transportation options to Sentosa |
| | Ask for places of interest in Sentosa |
| | Select one place of interest and ask questions about it (including location) |
| | Find a suitable place for having dinner |
| Orchard Road | Find a convenient MRT station to reach Orchard Road |
| | Ask for shopping centres in Orchard Road and select one to go |
| | Ask for a hotel conveniently located near the shopping centre you selected |
| | Find a suitable place for having dinner |
| Moving around Singapore | Ask for directions on how to go to the zoo |
| | Ask for information related to the zoo |
| | Find a suitable place for having lunch in some central Singapore area |
| | Find directions on how to reach that area |
| | Ask information about museums in Singapore |
| General Information | Ask about the type of weather in Singapore |
| | Ask information about the languages spoken in Singapore |
| | Find out information about the currency and exchange rates in Singapore |
| | Find out when was Singapore founded and who was the founder |
| | Ask about places of interest to visit in Singapore |
| | Select two places of interest and find about their locations |

**Table 1.** Five use cases for Singapore touristic agent evaluation

Based on the interaction results for each considered scenario we computed the objective completion rate, as the percentage of the specific objectives completed with respect to the total number of objectives in the task. Figure 4 summarizes the objective completion rates for the first scenarios (Marina Bay, Sentosa and Orchard Road), the second scenario (Moving around Singapore) and the third scenario (General Information).

As seen in figure 4, scenario 3 (General Information) achieved the best performance with an objective completion rate of 60%. In contrast, scenario 2 (Moving around Singapore) achieved the worst performance in terms of objective completion rate: 33%. Here, the most common errors reported by the users – which were later confirmed by the dialogue session logs – were related to the proper identification of

venues and venue directions. This information is of fundamental importance for fine tuning and improving the system performance for future release.
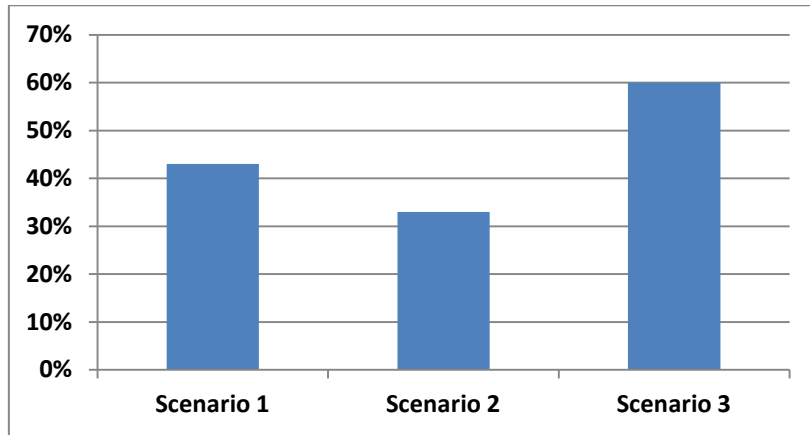


**Fig. 4.** Objective completion rates for the evaluated scenarios.

The results of the subjective evaluation on the user experience are presented in figure 5. As seen from the figure, the highest percentages were achieved for user "agreeing" (SA+A) on statements concerning the usability, reliability and functionality of the system. On the other hand, it is also important to notice that there are visible "disagreement" statements (D+SD) percentage bars concerning the usability and reliability of the system. These negative results were caused in particular by scenarios when venues and directions were not properly identified by the system.
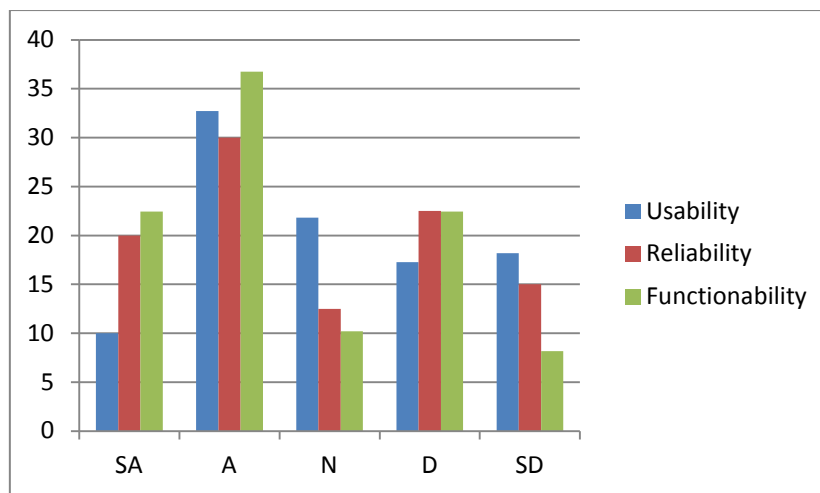


**Fig. 5.** Results of the subjective evaluation on user experience

# 5 Conclusions and future work

In this paper we presented SARA, an automatic touristic information system for the city of Singapore, which combines general QAs about Singapore's touristic spots and task-oriented dialogues for hospitality and transportation services. The system is currently implemented as Android application and as web-based system.

Our first evaluation round showed that our system needs several improvements before it can reach the level of a successful commercial application. The information collected during the evaluation provided valuable feed-back on system performance and user experience. As such, our next goal is to work towards the improvements suggested by our test participants.

Additionally, we are planning to extend the system with multilingual capabilities – at the moment, the mobile application supports only English, while the web system can be used for both English and Chinese.

Another improvement planned concerns with error handling strategies for misspelled venues, either for incorrectly typed user inputs or out-of-vocabulary misrecognized speech inputs. In the future, the system will be able to provide alternative options in cases when, for example, users misspell the name of a venue.

We also plan to create an iOS and Windows mobile version of the application, improve the usability of current user interface, enlarge the database coverage with additional touristic information, and integrate it with our digital receptionist and restaurant recommendation systems.

Finally, we plan to conduct a second system evaluation by engaging real users to interact with the system. The result of this evaluation study will ultimately help us to improve future versions of SARA regarding both, performance and user experience.

## References

1. Kabassi, K. Personalizing recommendations for tourists. Telematics and Informatics, 27(1):51-66, February 2010
2. Wium, M. Design and Evaluation of a Personalized Mobile Tourist Application. Master Thesis. Norwegian University of Science and Technology, 2010
3. Singapore Tourist Board, http://www.stb.gov.sg/
4. Malaka R. and A. Zipf. Deep Map – challenging IT research in the framework of a tourist information system. In Information and Communication Technologies in Tourism 2000, pages 15–27. Springer, 2000

5. Raubal, M. and S. Winter. Enriching way finding instructions with local landmarks. In Second International Conference GIScience. Springer, USA, 2002

6. Dale, R., Geldof, S. and Prost, J. CORAL: Using Natural Language Generation for Navigational Assistance. In Proceedings of ACSC2003, Australia, 2003

7. Bartie, P. and W. Mackaness. D3.1.2 - The Space-Book City Model. Technical report, The SPACEBOOKProject (FP7/2011-2014 grant agreement no.270019), 2013

8. Shroder, C.J., Mackaness, W. and Gittings, B. Giving the Right Route Directions: The Requirements for Pedestrian Navigation Systems. Transactions in GIS, pages 419–438, 2011

9. Dethlefs, N. and H. Cuayahuitl. Hierarchical Reinforcement Learning and Hidden Markov Models for Task-Oriented Natural Language Generation. In Proc. of ACL, 2011

10. Janarthanam, S., Lemon, O., Bartie, P., Dalmas, T., Dickinson, A., Liu, X., Mackaness, W. and Webber, B. Evaluating a city exploration dialogue system combining question-answering and pedestrian navigation. In Proc. ACL 2013.

11. Ko, J., Murase, F., Mitamura, T., Nyberg, E., Tateishi, M., Akahori, I. and Hataoka, N. CAMMIA: A Context-Aware Spoken Dialog System for Mobile Environments. In IEEE ASRU Workshop, 2005

12. Kashioka, H., Misu, T., Mizukami, E., Shiga, Y., Kayama, K., Hori, C. and Kawai, H. Multimodal Dialog System for Kyoto Sightseeing Guide. In: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2011

13. Webb, N. and B. Webber. Special Issue on Interactive Question Answering: Introduction. Natural Language Engineering, 15(1):1–8, 2009

14. Kim, S., Banchs, R.E. and Li, H. Wikipedia-based Kernels for Dialogue Topic Tracking, in Proceedings of ICASSP, 2014

15. Banchs, R. E. and 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

16. Xue, X., Jeon, J. and Croft, W. Retrieval models for question and answer archives. In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 475-482, 2008

17. Becker, T. Practical Template-Based Natural Language Generation with TAG. In: Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6), pp. 101–104, 2002

18. 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

19. ISO/IEC 9126 quality standard model http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39752