

Conversational interfaces and the Web

Gonzalo Ripa

LIFIA, Facultad de Informática, UNLP

`gripa@lifia.info.unlp.edu.ar`

Abstract. Conversational User Interfaces (CUI), both in voice interaction and chatbot ways, have become an important and popular way of interacting with current applications. These kinds of interfaces are nowadays spreading around the Web for multiple purposes, although the vast majority of Web sites do not provide them yet. There is still a huge gap between what Web applications allow users to do via Graphical User Interfaces (GUI), and what CUI offers. Also, there are unanswered questions about CUI design and its usability impacts on web browsing. Although other approaches for CUIs creation, the proposed approach transfers the responsibility of creating the conversational interfaces to the users of the Web browser, instead of remaining an implementation decision taken by the Web site owners. In this sense, an End-User Development (EUD) environment was designed to allow end users to define conversational interfaces by themselves, based on annotations of UI elements from Web Sites. In the case of Chatbots, plugged into the target Web sites, and in the case of Voice Interfaces (VUI), creating applications for smart speakers' devices based on third-party Web sites content. The underlying ideas and the reference architecture could be used to create tools that allow fast prototyping of CUIs, which could be important to conduct exploratory studies about the impact of using them in Web applications.

Keywords. Conversational interfaces, chatbots, voice user interfaces, end-user development, Web contents.

1 Introduction

Conversational interfaces are becoming a relevant way to access content and services that are normally delivered by Web applications and GUIs. Despite the progress in Model Driven Engineering [1] and Multi-Modal User Interfaces [2], most of the Web sites are not developed with conversational technologies in mind. In particular, the use of VUIs may occur in different contexts: many mobile applications are now able to react to voice commands, and also new uses of voice-based conversation arose, for instance, in the area of Web accessibility based on voice commands. Going to chatbots, those that are embedded in the Website UI may be created ad-hoc [1], or by some middleware based on the annotation of the HTML document [3]. Although these are valuable contributions, it would be ideal that conversational interfaces could

be created automatically from the website GUI. Recent papers [4] [2] have coined the idea of “conversational Web browsing” i.e. “a dialog-based, natural language interaction with websites”. Although it seems a promising idea we are still far from materializing this concept. The creation of a conversational interface still depends on application developers, and their specification is not trivial and the cost could be high for many Web site owners. Before focusing on an automatic “conversational transcoding” to create conversational interfaces from GUI, it is required to better understand the different impacts of this interaction in the context of conversational Web browsing and particularly attend to the difference in this interaction when it is done in computers and mobile devices, specially in the case of Chatbots, where the role can change drastically.

2 State Of The Art

Despite the existence of valuable uses of voice interaction, like some approaches based on improving Web accessibility [5], or another one related to performing Web actions with the voice [6], more complex conversational interfaces are still missing to access Web contents and services. Beyond receiving a voice/text command and giving a response based on some specific Web content, the conversational interface may propose an interactive way to deal with that response, explore more from there, etc. These kinds of actions can be achieved by offering conversational interfaces between applications and users. Recently, some new research has arisen in this regard. Prior works have defined the concept of conversational Web interaction [4] and conversational Web browsing [7]. The former proposes a conceptual framework to generate Chatbots for conversational interaction with websites through annotation techniques. In the latter, the authors show an implementation of their prior architecture. It is worth mentioning that both articles are based on an architecture that depends on HTML annotations that must be defined by the application developers. In contrast, here is proposed an annotation based approach that allows end users and developers to add annotations to any Web site on the client side without requiring programming skills and without depending on Web sites owner’s annotations. Another interesting work [8] focuses on ways to build natural language interfaces for individual web APIs so that different virtual assistants don’t have to implement their interfaces, they only need to integrate the proposed framework. The nature of previous approaches makes them very different since the approach proposed here involves the design of the conversational interface as a separated model created by users, while for example, the approach proposed by Baez et al. [4] focuses on a conceptual architecture for automatizing the creation of these interfaces. In this regard, the approach proposed allows combining Web content from multiple sources, because it separates Web content access and return from the conversational interface design. This is a remarkable difference from several mentioned approaches that use a complete Web site as the input of an automatic process and implies Web sites’ developers’ intervention.

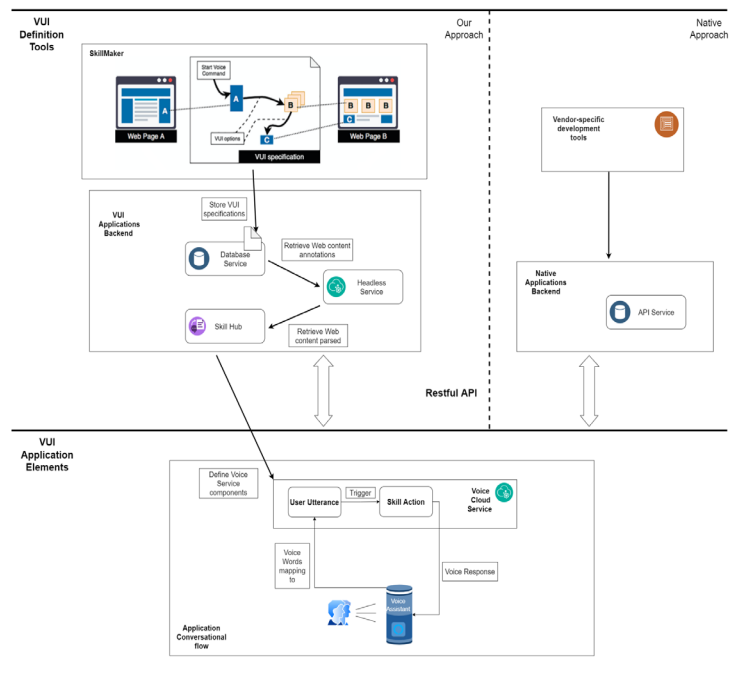
Related to the end-user development aspect of the approach, chatbot development has been tackled from this perspective, particularly for data exploration [9]. However, that work does not support voice user interfaces, and also it does not focus on Web

interaction.

3 Problem Statement And Contributions

Most Web sites and content that exist nowadays can be consumed almost exclusively through a GUI way and were not conceived to be accessed using the voice or through some other conversational interface like could be a Chatbot. What Web applications allow users to perform via Graphical User Interfaces (GUI) is limited, and CUIs could break these limits. The proposed approach, which was implemented as an end-user development environment, allows users to add conversational interfaces to any existing Web site on the client side. In this way, we move the responsibility of creating these interfaces from the Web site's owners to end users. Then, these are automatically weaved on the target Website when it is loaded in the Web browser (in the case of Chatbots) or are automatically available to be consumed in some voice-based device (in the case of VUIs). The main idea is to create conversational interfaces based on existing Web GUIs by defining a set of web element annotations. Imagine users interacting with a chatbot that offers a conversational interaction inside an e-commerce domain: end users could want to get a list of products filtered by some criteria. They must create information models using the EUD tool of this approach identifying and selecting each part of the web elements that conforms to a product for sale, i.e. annotating web contents. Figure 1 shows the overview of the approach from the point of view of VUIs creation but is analogous to chatbot conversations creation too.

Fig. 1: Overview of the idea



The most important is to understand that the EUD tool includes a GUI part (SkillMaker) where end users select and annotate parts of web sites to create VUI specifications. These specifications will be saved into a database service and processed by other back end services (VUI App Backend) which communicates via Restful API with the VUI applications running in some voice-based device: in this case, the VUI applications are created by users, unlike the native approach where these applications are created by owners.

The approach doesn't focus on AI aspects related to the detection of users' requests, but useful interfaces are generated based on simple rule-based conversations, as pointed out in [8]. Although the prototype is focused on the definition of simple conversational interfaces by end users, the underlying ideas and the reference architecture could serve as a basis to create tools for fast prototyping and implementation of these interfaces, which could be important to conduct exploratory studies about the impact of their use in Web applications and also a way to collect data for the creation of automatic transcoding of GUIs into conversational interfaces.

4 Research Methodology

The research is based fundamentally on the analysis of the associated literature (a systematic review was done) and carrying out at least two experiments during the process:

- Experimental test with users for the final product (VUI application) generated by using EUD tools.
- Experimental test with users evaluating the web content annotation process, the conversational interface (Chat bot) definition process carried out using EUD tools, and the utility of the final product (Chatbot) generated for Web and mobile environments.

For the first experiment were collected primary data combining several techniques like surveys, providing a questionnaire of 10 questions multiple-choice based on SUS (System Usability Scale) standard to measure the usability; as well the observation of the subjects performance during experiment execution, to detect difficulties that users can face.

5 Evaluation Plan

Considering that performing experiments can be expensive (we need many users and different domain profiles) but with interesting results, some challenges appear for evaluate the proposed approach with end users:

- Flexibility with which a user can annotate Web site contents.
- The usefulness of conversational interfaces generated as interaction alternatives for a Web site (with the limitations that the conversation model used has).

- The utility of conversational interfaces according to the Website domain. i.e. Are academic sites just as useful as e-commerce (a large database with products) or sites where there is very deep navigation?
- Potential usability improvements for Web sites with weak designs.
- In the end, it will be evaluated:
 - When using a conversational interface to interact with Web sites is interesting.
 - What features should have the conversational interface to be useful? I.e. is it valid to show a carousel inside a chatbot to navigate objects that it returns or is it better to show them as a list?

6 Preliminary Or Intermediate Results

The first experiment mentioned in the “Research methodology” section was already carried out. The main goal was to assess whether or not there is a significant user experience difference between applications generated by the approach EUD tool, and native applications (based on Alexa or Google Assistant SDKs). The experiment aims at answering the following research question: *is the user experience equivalent for Native apps and apps generated by the EUD tool?*

For this experiment, it was decided to define and mimic existing apps from the Alexa store using the EUD tool. In this way, we were able to verify whether or not the apps made with the approach are similar to a native one that meets the same requirements. In particular, 2 use cases: weather and news domains. It was designed as a completely randomized experiment where end users (20 Argentinean subjects, 10 females and 10 males, spread between 16-78 yo and with different education levels from different formation fields) with no previous experience with Alexa service, were asked to perform a set of tasks on one of the application versions (native and EUD tool generated). For the analysis of the research question mentioned above, the System Usability Scale (SUS) was used to capture the perceived usability with both approaches, where the resultant means are pretty similar from 87,00 (native) to 88,50 (approach), which denotes not much difference in the SUS score. As a preliminary result of the evaluation carried out, applications generated using the approach provide a similar user experience to native solutions.

7 Conclusions And Learned Lessons

Conversational user interfaces have been a strong research area for a long time. Despite the process in which Web applications owners started to migrate or support this kind of interaction for new devices and interaction contexts, there is still a huge gap between what users can already do with these interfaces and common access to Web content. In fact, many Web applications are still not supporting this new kind of interaction. Beyond this, although this gap could disappear in the near future, other issues such as personalizing or adapting this kind of interface will surface, as

happened in the past with other kinds of interaction. As it was mentioned before, the approach proposed was evaluated by comparing the applications generated with our approach and equivalent native applications. In this way, we showed that the applications obtained with the approach provide a similar user experience to native solutions.

Although so far the tool was not evaluated to provide evidence of the effort required for creating applications, I strongly believe that using the EUD tool is very convenient if we compare the effort required to create native applications, moreover considering the learning curve for both approaches. In this sense, it is planned to study if the EUD tool may also be used for creating another kind of conversational interface such as Web chatbots. This adapting stage is already in development. Besides, it is planned to perform experiments to better understand how end users may use the environment and measure if end users actually are able to create their applications.

8 Ph.D. Stage

Middle Stage

References

1. Marco Brambilla, Jordi Cabot, and Manuel Wimmer. *Model-Driven Software Engineering in Practice*. Synthesis Lectures on Software Engineering. Morgan & Clay pool Publishers, 2012.
2. Nadia Elouali, José Rouillard, Xavier Le Pallec, and Jean Claude Tarby. Multimodal interaction: a survey from model driven engineering and mobile perspectives. *J. Multimodal User Interfaces*, 7(4):351–370, 2013.
3. Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle M. Lottridge. Understanding the long-term use of smart speaker assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):91:1–91:24, 2018.
4. Marcos Báez, Florian Daniel, and Fabio Casati. Conversational web interaction: Proposal of a dialog-based natural language interaction paradigm for the web. In *Chat bot Research and Design - Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19-20, 2019, Revised Selected Papers*, volume 11970 of *Lecture Notes in Computer Science*, pages 94–110. Springer, 2019.
5. Yevgen Borodin, Faisal Ahmed, Muhammad Asiful Islam, Yury Puzis, Valentyn Melnyk, Song Feng, I. V. Ramakrishnan, and Glenn Dausch. Hearsay: a new generation context-driven multi-modal assistive web browser. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pages 1233–1236. ACM, 2010.
6. Julia Cambre, Alex C. Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Y. Tsai, Chinmay Kulkarni, and Jofish Kaye. Firefox voice: An open and extensible voice assistant built upon the web. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pages 250:1–250:18. ACM, 2021.
7. Pietro Chittò, Marcos Báez, Florian Daniel, and Boualem Benatallah. Automatic generation of chatbots for conversational web browsing. In *Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings*, volume 12400 of *Lecture Notes in Computer Science*, pages 239–249. Springer, 2020.

8. Yu Su, Ahmed Hassan Awadallah, Madian Khabsa, Patrick Pantel, Michael Gamon, and Mark J. Encarnación. Building natural language interfaces to web apis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 177–186. ACM, 2017.
9. Ludovica Piro, Giuseppe Desolda, Maristella Matera, Rosa Lanzilotti, Sara Mosca, and Emanuele Pucci. An interactive paradigm for the end-user development of chatbots for data exploration. In *Human-Computer Interaction - INTERACT 2021 - 18th IFIP TC 13 International Conference, Bari, Italy, August 30 - September 3, 2021, Proceedings, Part IV*, volume 12935 of *Lecture Notes in Computer Science*, pages 177–186. Springer, 2021.