Insights from Conversational User Interface Projects

Stefan Schaffer
Norbert Reithinger
stefan.schaffer@dfki.de
norbert.reithinger@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Berlin, Germany

ABSTRACT

In this paper we highlight several difficulties and insights that we discovered during our practical work in the area of conversational user interfaces. We review four fields in which we conducted research and industries projects. We dig into issues like developing for various domains, disclose insights into "fun bot" development, and derive ideas for design guidelines for developing for domains with legal restrictions and for efficient interaction.

KEYWORDS

conversational user interfaces, chatbots, multimodal interaction

INTRODUCTION

DFKI's (German research center for artificial intelligence) Cognitive Assistants team in Berlin conducts research in several areas of HCI including novel interaction techniques and intelligent data processing. Since many years we realize and investigate human computer interfaces including multimodal and conversational systems [1, 2, 6].

When designing and implementing conversational user interfaces several trade-offs have to be made. For each conversational user interface to be realized key-factors have to be considered: domain, domain size, target user group, available input and output modalities as well as the technology used in the background. Therefore, in most of our dialog system projects we use the Multimodal Mobile Interaction and Rendering (MMIR) framework [3], a modular approach that allows for both the easy exchange of single components as well as the rapid creation of interactive prototypes. The open source[1] framework especially supports multimodal and mobile interaction while allowing for building

[1] https://github.com/mmig/mmir-starter-kit

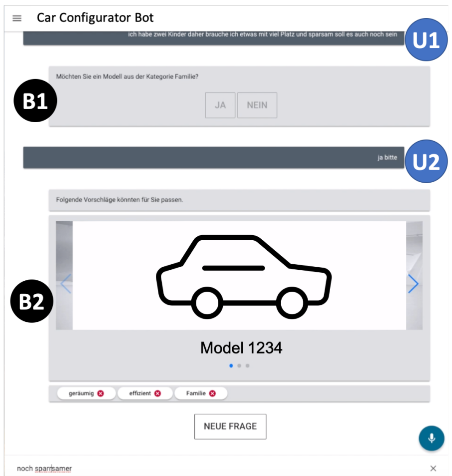

Figure 1: Car configurator bot. Due to the small font in the screenshot the user-bot conversation is depicted below.

- User (U1): "I have two **kids**, so I need something with **a lot of space** and it should be **economical** as well."
- Bot (B1): "Would you like a model from the **Family** category?"
  YES NO (Buttons)
- User (U2): YES (Button)
- Bot (B2): "The following options might fit your needs."

applications for multiple platforms. MMIR allows for an easy integration of different functionalities for building multimodal and conversational systems by providing components for automatic speech recognition, text-to-speech, build in grammar parsing, state machine based dialog management, and GUI interaction. It also provides easy access to mobile device's sensors and actuators.

In our work we utilized MMIR to realize several conversational user interfaces, some as stand alone applications and others as server client implementations. Thanks to the modular concept we also experiment with different background technologies for natural language understanding and dialog management.

In this paper we outline selected works from applied research and industry projects highlighting project specific difficulties and revealing insights that we gained from our practical work in the area of conversational user interfaces.

FLEXIBLE KNOWLEDGE REPRESENTATION OF A VAST DOMAIN

Huge domains can cause considerable efforts if you want to access the knowledge via a chatbot interface. In one of our chatbot projects we designed and implemented a conversational user interface for a car configurator bot for a big German car manufacturer. One challenge was that the huge amount of different equipment options for a car should be made accessible through a conversational interface. The system should also be able to automatically update the language processing capabilities if new equipment options are added or existing equipment options are changed or deleted. Existing equipment options could be accessed via a technical database. However, these equipment descriptions were lacking important semantic information that was needed to enable the foreseen business logic. A semi-automatic process was implemented in order to convert the equipment database into a taxonomy that augmented the equipment with additional information. The resulting equipment taxonomy included up to 3000 objects that were classified into the following super classes:

- `standard element`: a standard equipment, e.g. trailer hitch, panorama roof, CD player, etc. Further several semantic sub classes exist.
- `property element`: an element that can have a specific *property* (e.g. *red* lacquer, *black* leather upholstery, or attributes like *very* efficient).
- `concept`: a concept can be composed by clustering a number of elements. The concept *family* could be e.g. composed by putting together car elements that are particularly suitable for families (e.g. spacious, dark seat covers, child seat fastening)
- `technical value`: technical values require special consideration as they have to be interpreted depending on specific properties such as unit and scale format. Several semantic sub classes exist, e.g. to distinguish between number of seats, cubic capacity, maximum speed, etc.

The elements of the taxonomy further allow for the direct integration of synonyms to enable a more flexible user input. In order to match `concepts` more easily so called *hints* are used. A *hint* for the concept family can e.g. be the word "baby" or "mother". Figure 1 exemplifies the usage of hints. In U1 a number of hints for the concept family can be found. The bot therefore suggests to take a look at models from the family category (B1). The user confirms (U2). As the result the bot presents a car model meeting the users requirements together with a promt and the current selected options, namely **spacious**, **efficient**, and **family** (B2).

To meet the requirement of automatic language processing update we developed an initialization procedure that requests the current version of the taxonomy and automatically converts the taxonomy into a grammar. During the chatbot usage the grammar is utilized to translate the user input into structured information that is processed by the business logic. Elements, concepts and technical values that are recognized can be utilized in order to request car models matching to the users requirements. The car models are generated by an additional service from the car manufacturer that makes use of the user input to calculate car models the match the user requirements.

The business logic foresees two basic dialog flows:

(1) `User request and result presentation`: all utterances are processed and if the utterance only contains standard elements, property elements or technical values the matching car models are generated as a result.
(2) `Concept match and concept proposal`: if the utterance matches a hint to a concept the chatbot asks the user if car models matching the respective concept should be displayed.

These dialog flows allow for an easy to use car configuration and enormously reduce the number of matching car variants for the user. As outlined this easy to use way of interacting with the domain was realized by putting immense effort into making the knowledge representation both detailed and flexible.

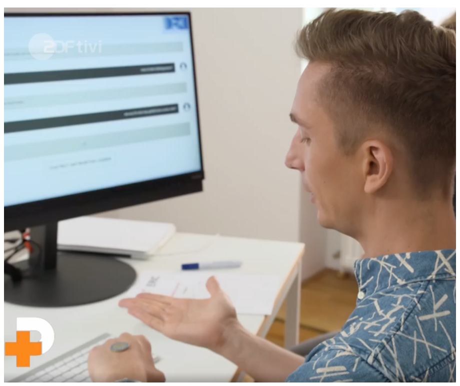

Figure 2: Entertainer Eric Mayer flipping a coin to decide weather he or the bot answers the next question.
Video: https://www.zdf.de/kinder/purplus/koennen-computer-denken-100.html

PROOF THE SPOOF: HOW WELL CAN CHILDREN DISTINGUISH BETWEEN HUMAN-HUMAN AND HUMAN-BOT INTERACTION

PUR+ is a German TV magazine for children moderated by the entertainer Eric Mayer. One episode of the magazine is about chatbots. We were asked to create a chatbot for this episode that should be able to mimic answers of Eric Mayer. The aim was to realize the so called *Eric bot*, a bot that can answer personal questions as if Eric Mayer himself would give the answers. We got a total of 185 questions answer pairs from Eric. From these Q/A pairs about 400 categories in AIML [9] were created. The bot was implemented as a browser-based web application using the MMIR framework. Keyboard or speech can be used alternatively to input information. As a fall back for questions were no answers were prepared an adapted ELIZA [8] component was used. The output prompts where adapted in

Figure 3: Prof. Wahlster and the Wolfgang Bot

such a way that the they better meet children speech and that ELIZA doesn't ask back any questions, as the goal was not to have a small talk conversation, but only to ask personal questions about Eric.

We conducted a chatbot experiment with students of an 8th grade high school in Berlin. The task was to ask questions via the chat interface and to assess whether the answer was given by Eric himself or by the chatbot. Eric was placed in another room. After each question Eric flipped a coin to determine whether he or the chatbot answered (Figure 2). The whole experiment was filmed and a report about the experiment was broadcast in the PUR+ magazine.

For time reasons, no further natural language processing could be integrated in addition to the AIML structures. Therefore the children's questions were entered by an experimenter to reduce typo and out-of-grammar errors that can quickly lead to chatbot malfunction.

The results of the experiment are currently analyzed. We are particularly interested in the children's strategies to find out if the answer came from the human or the chatbot. A total of 120 questions were asked. 26 times Eric was thought to be the chatbot (false negative) and 14 times the chatbot was thought to be Eric (false positive). Further insights of the experiment will be published soon.

FUN GOES FAST: THE WOLFGANG BOT

For the 30 years anniversary of DFKI, we wanted to present the then longtime CEO Prof. Wolfgang Wahlster, a fun robot that was able to mimic a real CEO. We envisioned a cardboard "robot" able to interact with fun facts extracted from press interviews with Prof. Wahlster. The "robot" itself consists of two cardboard boxes, one for the torso and one for the head. The head turns automatically to the loudest sound in the environment, mimicking attention, based on simple sound sensors. A tablet on the front of the torso runs the speech-based chatbot (Figure 3).

In the first version the tablet-based chatbot was based on the PUR+ bot. Four interviews with about 120 Q/A-pairs with Prof. Wahlster were scrapped and transformed into AIML-structures. Additionally, we extended PUR+-bots ELIZA component to add some idiosyncrasies. The tablet-based fronted uses DFKI's MMIR framework to add voice I/O with Googles recognizer and synthesis. In addition to buttons on the tablet, the system could be controlled with Flic buttons, attached to the robot. The chatbot's software in its first version needed about 1 person week to be fully operable and was a big success at the 30 year's jubilee.

However, AIML poses a significant bottleneck for robust interaction. Unless you do not carefully design your AIML input patterns, the interaction experience is not overwhelming using speech in a noisy environment, even with the additional ELIZA mode. Version 2 of the chatbot was designed with the aim to run locally on the tablet using DFKI's MMIR framework, without using AIML. Input processing is based on n-gram matching of the recognized string to the questions, resulting in a much higher robustness of NL understanding. Additionally, facts from Prof. Wahlster's homepage and Wikipedia page were added, as well as a simple voice based calculator and a weather forecast. An intern realized the chatbot within two weeks after making himself comfortable with all the necessary tools. The robustness of and consequentially the fun of interacting with Wolfgang Bot in its second version is definitely higher. The App works on any phone or tablet with Android 5 or higher

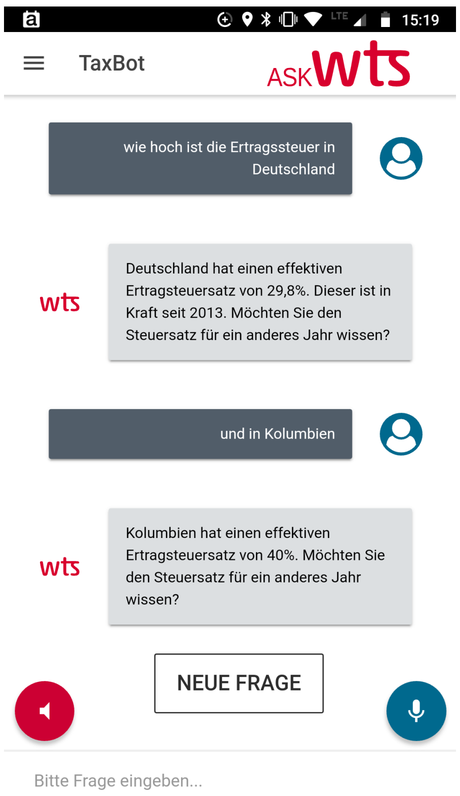

Figure 4: WTS tax chatbot allowing for guided multi-step dialogs.
Video: https://youtu.be/fBL-ei6hyfI

TAKE YOUR USER BY THE HAND: GUIDED MULTI-STEP DIALOGS FOR COMPLIANCE WITH LEGAL RESTRICTIONS

WTS (www.wts.com) is a global tax consulting company with affiliated legal advice. International clients must consider fiscal law from many countries. Consultants must be able to quickly access general tax information as well as solve multi-national cases. The WTS TaxBot was developed as a showcase for quick information access for specialised cases and exemplifies Q/A tasks and guided processes (Figure 4). On the one hand it provides quick information about profit tax rates worldwide, solving the problem of frequent verbal request addressed to specialists. In addition, we implemented multi-step dialogs to estimate cross-border tax rates for certain processes and events.

The main problem was to collect and structure the very specific tax domain knowledge needed, both the rates and also the tax rate calculation. For the dialog design of the multi-step dialogs we had to create user stories and interview experts, since no dialog data are available in this special domain. Also, the tax rate computation is not publicly available. The dialog flow has to respect the dimensions correctness vs. naturalness vs. flexibility (in the order of importance).

Using MMIR, we realized a domain specific NLP stack that addresses the special tax lingo. Since correctness is of utmost importance the example flow in the dialog manager (using a finite state machine) is very conservative and fixed to the domain to provide correct information. The TaxBot as a technology demonstrator was presented at a public event sponsored by WTS in October 2017.

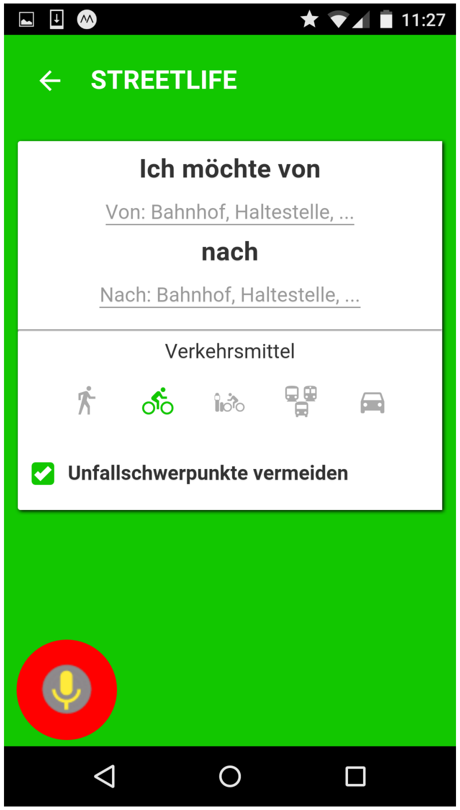

Figure 5: The STREETLIFE mobility app. The conversational view directly suggests the user a possible utterance ("I want to go from *origin* to *destination*.") which can be understood by the system
Video: https://youtu.be/oN3TbmUIxnM

SEMI-CONVERSATIONAL INTERACTION AND DESIGN FOR EFFICIENCY

Our previous research showed that users tend to prefer input modalities with higher efficiency [7]. It was e.g. shown that users switch from GUI to VUI usage if effort in terms of interaction steps can be saved. Based on these results we decided to make use of semi-conversational user interface in several projects (e.g. in [5]). By switching from a classical GUI to a conversational style of interacting with the user the efficiency of the interaction can be increased. When conceptualizing new interfaces we often consider this benefit of multimodal interaction as a design principle.

As an example Figure 5 shows the conversational view of the STREETLIFE app [5]. The app is a mobility application supporting the user to find itineraries from A to B. By tapping the microphone button the classical GUI (not shown) converts into the conversational view allowing the user to input origin and destination information via speech. In parallel to spoken input the graphical elements in the conversational view can be used to customize the modes of transport and to activate an "avoid accident hot-spots" feature. The STREETLIFE app further has a smartwatch extension using speech as the main input modality.

CONCLUSION AND FUTURE WORK

Our conversational user interface projects showcase a number of existing difficulties that may arise during implementation in various domains. Furthermore, we illustrate possible design guidelines for the successful implementation of conversational interfaces. Using the example of a car configurator bot, we show that large domains can cause greater effort during implementation, especially if the taxonomy used should be flexible changeable. On the other hand the Eric Bot and the Wolfgang Bot showcase that Q/A bots can be realized comparably fast with an acceptable quality - in our case considering the perceived experience of children and the perceived fun-factor of DFKI employees and guests. We were not totally happy with AIML and came up with alternative solutions. If compliance with legal restrictions is required, a guided process using state machine-based dialog management is still a reliable and practicable approach. Putting a focus on the efficiency of HCI conversational interaction can be beneficial and a paradigm of semi-conversational interaction could be a valuable alternative for many conventional man-machine interfaces.

Our future work will again include the design and development of conversational user interfaces. We will road-test the usage of chatbots in the museum [4]. Further possible projects in the pipeline include the usage of chatbots for eGovernment services as well as a conversational user interface to perform cognitive behavioural therapy.

ACKNOWLEDGMENTS

This research was partially supported by the DAYSTREAM project of the research initiative "mFund", which is funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI) of the Federal Republic of Germany under funding number 19F2031A. We thank Janni for his interest in conversational interaction, the whole PUR+ team for the great TV experience, our team of student workers and interns for their great work at DFKI, and WTS for their interest in chatbots.

REFERENCES

[1] Norbert Reithinger, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löckelt, Jochen Müller, Norbert Pfleger, Peter Poller, Michael Streit, and Valentin Tschernomas. 2003. SmartKom - Adaptive and Flexible Multimodal Access to Multiple Applications. In *Proceedings of Fifth International Conference on Multimodal Interfaces. International Conference on Multimodal Interfaces (ICMI-2003), Fifth, November 5-7, Vancouver, British Columbia, Canada*. ACM.

[2] Norbert Reithinger, Simon Bergweiler, Ralf Engel, Gerd Herzog, Norbert Pfleger, Massimo Romanelli, and Daniel Sonntag. 2005. A look under the hood: design and development of the first SmartWeb system demonstrator. In *ICMI '05: Proceedings of the 7th international conference on Multimodal interfaces. International Conference on Multimodal Interfaces (ICMI), New York*. ACM Press, New York, NY, USA, 159–166.

[3] Aaron Ruß. 2013. MMIR Framework: Multimodal Mobile Interaction and Rendering. In *INFORMATIK 2013: Informatik angepasst an Mensch, Organisation und Umwelt. Workshop Modellbasierte Entwicklung von Benutzungsschnittstellen (MoBe-13), befindet sich INFORMATIK 2013, September 16-20, Koblenz, Germany (Lecture Notes in Informatics (LNI))*, Matthias Horbach (Ed.), Vol. P-220. Gesellschaft für Informatik (GI), Köllen Druck+Verlag GmbH, 2702–2713.

[4] Stefan Schaffer, Oliver Gustke, Julia Oldemeier, and Norbert Reithinger. 2018. Towards Chatbots in the museum. In *Proceedings of the 2nd Workshop on Mobile Access to Cultural Heritage. Workshop on Mobile Access to Cultural Heritage (mobileCH-2018), 2nd Workshop on Mobile Access to Cultural Heritage, befindet sich Mobile HCI 2018, September 3-6, Barcelona, Spain*, Sylvain Castagnos, Tsvi Kuflik, Ioanna Lykourentzou, and Manolis Wallace (Eds.), Vol. Vol-2176. CEUR Workshop Proceedings.

[5] Stefan Schaffer and Norbert Reithinger. 2014. Intermodal personalized Travel Assistance and Routing Interface. In *Mensch Computer 2014 - Tagungsband. Mensch & Computer, 14. Fachübergreifende Konferenz für Interaktive und Kooperative Medien - Interaktiv unterwegs - Freiräume gestalten, August 31-September 3, München, Germany*, Andreas Butz, Michael Koch, and Johann Schlichter (Eds.). De Gruyter Oldenbourg, 343–346.

[6] Stefan Schaffer and Norbert Reithinger. 2016. Benefit, Design And Evaluation Of Multimodal Interaction. In *Proceedings of the 2016 DSLI Workshop. ACM International Conference on Human Factors in Computing Systems (CHI), Designing Speech and Multimodal Interactions for Mobile, Wearable, and Pervasive Applications*. online.

[7] Stefan Schaffer, Robert Schleicher, and Sebastian Möller. 2015. Modeling input modality choice in mobile graphical and speech interfaces. *International Journal of Human-Computer Studies* 75 (2015), 21–34.

[8] Joseph Weizenbaum. 1966. ELIZA - a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (1966), 36–45.

[9] Hong Zhang, Rajiv Kishore, Raj Sharman, and Ram Ramesh. 2007. Agile Integration Modeling Language (AIML): A conceptual modeling grammar for agile integrative business information systems. *Decision Support Systems* 44, 1 (2007), 266–284.