ENDA: Insights into Building a Chatbot for Open Government Data

Fritz Meiners*, Fabian Kirstein**

*Fraunhofer FOKUS, Berlin, Germany, fritz.meiners@fokus.fraunhofer.de
**Fraunhofer FOKUS, Berlin, Germany, and Weizenbaum Institute for the Networked Society, Berlin, Germany, fabian.kirstein@fokus.fraunhofer.de

Abstract: The frictionless access to Open Data via portals and traditional search paradigms currently lacks usability. In order to tackle this problem, we developed a prototype of a chatbot for Open Government Data called ENDA, which is based on the ChatScript framework and the Linked Data specification for public sector datasets DCAT-AP. User requests are mapped to corresponding SPARQL queries using pattern-matching techniques. The initial requirements were derived via a Wizard-of-Oz study involving potential users. During evaluation against the European Data Portal it was revealed that existing limitations hinder the development of a production-grade chatbot.

Keywords: Chatbot, Linked Open Data, DCAT-AP

1. Introduction

Today Open Data is prevalent in many different domains, published by nonprofit organizations, companies, and authorities from the public sectors alike. Open Data portals in the European Union are encouraged to publish their data using the DCAT-Application profile for data portals in Europe (DCAT-AP)[1], a linked data ontology specifically designed for "describing public sector datasets in Europe". However, (Janssen, Charalabidis, Zuiderwijk, 2012) conclude, that (meta) data not being found is one of the "impediments that influence the open data process from the perspective of open data users". The hypothesis of our work is that this problem could be approached by employing chatbots as a means of interaction. That creates the illusion that users are communicating with a human, when in fact they are not. Instead, algorithms interpret the input of the users and consequently try to reply in a meaningful way. The foundation of our work is a Wizard-Of-Oz (WOO) study, which has been conducted to get a better understanding of the way users will interact with the chatbot. In a WOO experiment, users are asked to interact with a given system. However, instead of the system a human produces the output.

---

[1] https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/

2. Design and Implementation

Based on the insights gained from the WOO experiment a dialogue flow and system was designed. The service offered by ENDA is split into the following three tasks: (1) user interaction, (2) dialog management (NLP), and (3) construction and handling of SPARQL queries. The first two tasks were implemented using ChatScript[2], a pattern-matching based framework for developing chatbots. Once a user's intent and correspondig entities have been detected the SPARQL middleware maps the extracted keywords to the applicable fields specified by DCAT-AP. The retrieved datasets are then passed back in a human readable way. The system is accessible via a web frontend.

3. Findings & Outlook

A chatbot depicts a very user-centric application, since the interaction scheme is much more in line with human conversational patterns than traditional user interfaces. Therefore, a complete evaluation of any chatbot should include a structured usability test with real users. Some research was conducted in this field (Kuligowska, 2015). However, our experiences from the practical implementation in comparision to the user-driven requirement elicitation did not justify such an evaluation. Several surrounding conditions impeded (for now) an implementation of an applicable Open Government Data chatbot. We have derived three major recommendations from our findings, which can act as guidelines for the development of production-grade Linked Open Data chatbot applications:

1) The quality, integrity and completeness of the metadata correlates with the potential abilities of the chatbot.
2) The interface for the retrieval of metadata has to offer sufficient performance and rich query features.
3) The design and implementation of a meaningful and communicative dialog flow requires substantial resources and domain knowledge.

Our work has shown that the popular DCAP-AP standard and mature frameworks like ChatScript are a solid foundation for developing novel approaches to access Open Government Data. However, adoption demands an in-depth examination of data quality and correct application of standards. Future work will focus on covering more input phrases and extending the dialog flow. Hardening the bot against low quality metadata and providing suggestions for users on limiting the result set could also be considered valuable improvements. Finally, user studies will have to be conducted.

References

Janssen, M., Charalabidis, Y., Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government.

---

[2] https://github.com/ChatScript/ChatScript

Kuligowska, K. (2015). Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents.

About the Authors

Fritz Meiners

Fritz Meiners M.Sc. is a researcher and software developer at the Fraunhofer Institute for Open Communication Systems. He graduated from the Humboldt University of Berlin in December 2019. He is currently engaged in the domains of Open Data and Open Government, as well as Urban Mobility and Smart Cities. Accordingly, he participated in related projects like the European Data Portal, Data quality guidelines for the publication of datasets in the EU Open Data Portal, and URBANITE.

Fabian Kirstein

Fabian Kirstein M.Sc. is a researcher and software developer at the Fraunhofer Institute for Open Communication Systems. He graduated from the HTW Berlin in Applied Computer Science and his work focuses on the area of Open Data, Open Science, interactive web platforms, service-oriented architectures and dezentralised data management, like Blockchain technology. In those domains he participated in several national and international research and industry projects, as the Open Data Portal of the city of Hamburg, the Policy Compass project, the European Data Portal and the Industrial Data Space.