

A Machine Translation-Powered Chatbot for Public Administration

Dimitra Anastasiou⁺, Anders Ruge[&]
Radu Ion^{*}, Svetlana Segărceanu[#], George Suciu[#], Olivier Pedretti⁺,
Patrick Gratz⁺, Hoorieh Afkari⁺

⁺Luxembourg Institute of Science and Technology,

[&]SupWiz, Denmark,

^{*}Romanian Academy Institute for AI,

[#]R&D Department, BEIA, Romania

a.ruge@supwiz.com, radu@racai.ro,

{svetlana.segarceanu, george}@beia.ro

{dimitra.anastasiou, olivier.pedretti, patrick.gratz,

hoorieh.afkari}@list.lu

Abstract

This paper describes a multilingual chatbot developed for public administration within the ENRICH4ALL project. We argue for multilingual chatbots powered through machine translation (MT) and discuss the integration of the eTranslation service in a chatbot solution.

1 Introduction

In this paper, we introduce the Action ENRICH4ALL (E-goverNment [RI] CHatbot for ALL) which is about the development of a multilingual chatbot service to be deployed in public administration in Luxembourg, Denmark, and Romania. ENRICH4ALL is funded by the Connecting Europe Facility and its duration is from June 2021 to May 2023. The partners are Luxembourg Institute of Science and Technology, BEIA Consulting Romania, Romanian Academy Institute for AI and SupWiz, Denmark. In this paper, we refer to the benefits and challenges of e-government chatbots and to the integration of eTranslation with the chatbot platform.

2 Related Work

The benefits of having e-government chatbots are several: they can process service requests in huge

numbers, work 24/7, provide up-to-date information and consequently reduce operational costs. In some European countries, such as Denmark, Estonia, and Latvia, there are chatbots used in many public authorities, whereas in other countries, such as Romania or Luxembourg, there are not. Some of the challenges of using chatbots in public administration are the large number of relevant services, the complexity of administrative services, the context-dependent relevance of user questions, the differences in expert-language and user-language as well as the necessity of providing highly reliable answers for all questions (Lommatzsch, 2018). To these challenges, we should add the language diversity in Europe. The consequence of language diversity is that each EU country and each administration uses its own initiative to deploy a chatbot (often monolingual) resulting in a scenario where the interaction with e-government through virtual assistants is scarce and fragmented.

3 A multilingual chatbot in public administration

Particularly for administrative procedures, there are many requests from expatriates, who enter a new country. Application for residence, importing a car, starting-up a new business, and building a house are some of such requests. Public administration was also burdened with many questions related to the pandemic, which gave rise

to COVID-19 chatbots in Europe. We created (and actively develop) three datasets:

- COVID-19 (RO²)
- Construction permits (RO)
- Administrative questions (LTZ–FR–DE–EN)

The datasets are available at the project’s website and will soon be available at the European Language Grid. As for BERT language models, we use already existing ones for RO, FR, DE, EN, and we have developed and trained one for Luxembourgish³ to use for detecting question similarity and classification with user intent labels, but this is outside the scope of this paper.

3.1 BotStudio

BotStudio is the AI-powered chatbot developed by the Danish partner SupWiz, where the eTranslation API is now integrated. BotStudio can use fine-tuned BERT-based models built with HuggingFace APIs to appropriately map user intents to chat nodes in specific domains.

3.2 Integration of eTranslation

eTranslation⁴ is both a stand-alone MT tool and an API that can be integrated into various systems to facilitate multilingual services. The tool translates from and to 27 languages in different domains, including Russian, simplified Chinese, and recently Ukrainian. eTranslation is the neural MT tool provided by the European Commission to all EU bodies but also public services and SMEs across Europe. The latency of the service is low for small input texts, which makes it usable for real-time applications. Three arguments for using eTranslation compared to other translation services are: i) privacy is a priority; all data resides in Europe⁵; ii) it is free for SMEs; iii) it supports niche domains for formal language.

Figure 1 presents the eTranslation integration. One of the challenges is language identification. In our chatbot, we added a language identification service based on the PyPI `langdetect` package. For LTZ, a new language profile was added, while for DE, FR, EN, RO, and DA⁶, existing language profiles are used. For all languages, the language of the input question is automatically detected and

suggested at the top of a drop-down list containing all available languages. The questions are then translated into any of DE, FR, EN, RO, DA based on the domain of the user-entered question and on which dataset is being used. BotStudio finds the right answer in the QA database, eTranslation translates the answer back in the user’s selected language and BotStudio gives the output.

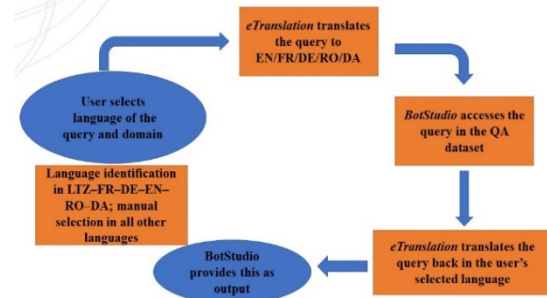


Figure 1. eTranslation integration workflow

4 Conclusion and Future Prospects

Our chatbot is an AI-based, MT-powered service, which proves available information to citizens 24/7 and reduces the administrative burden from public authorities. After the chatbot deployment, there will be additional data created and shared with the EC. Through data creation and training, eTranslation will be trained for other domains and maybe extended for LTZ, which is not supported in eTranslation. Generally, it would be interesting to integrate MT in open-domain conversational QA, e.g. ORConvQA (Qu et al., 2020).

Acknowledgement

The Action 2020-EU-IA-0088 has received funding from the EU’s 2014-2020 - CEF Telecom, under Grant Agreement No. INEA/CEF/ICT/A2020/2278547.

References

- Lommatzsch, A. 2018. A next generation chatbot-framework for the public administration. *International Conference on Innovations for Community Services*, Springer, Cham, 127-141.
- Qu, C. et al. 2020. Open-retrieval conversational question answering. *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 539-548.

² RO: Romanian, LTZ: Luxembourgish, FR: French, DE: German, EN: English.

³ <https://huggingface.co/raduion/bert-medium-luxembourgish>, 18.03.22

⁴ <https://joinup.ec.europa.eu/collection/connecting-europe-facility-cef/solution/cef-ettranslation/about>

⁵ See the privacy statement at <https://webgate.ec.europa.eu/etranslation/public/welcome.html>

⁶ <https://github.com/racai-ai/e4a-langdetect/commit/d29daa818fccdd9ce0d106dda1e7063bd48ee92#diff-d0066ed32bbca6cf4d1cfa89c16421197264b1bcb4cef39204fcf5cb4e5d291>