Language-Agnostic Knowledge Graphs for Smarter Multilingual Chatbots

Alena **Vasilevich**[1], Michael **Wetzel**[1], Georg **Sedlbauer**[2] and Kerstin **Hubmer**[2]

[1]*Coreon GmbH, Rungestrasse 20, 10179 Berlin, Germany*

[2]*Vienna Business Agency, Mariahilfer Strasse 20, 1070 Vienna, Austria*

Abstract

CEFAT4Cities targets the development of multilingual cross-border e-Government services, facilitating the conversion of natural-language administrative procedures into machine-readable data. We showcase the integration of CEFAT4Cities results into SmartBot, a prototype of a multilingual chatbot, developed for the Vienna Business Agency (VBA) in scope of the project. SmartBot makes VBA's services discoverable in a user-friendly way, fine-targeting such topics as starting a new business and finding relevant grants among hundreds of funding opportunities. It is driven by multilingual AI that contains the results of CEFAT4Cities workflows, integrated into its domain knowledge along with multilingual domain-specific vocabularies, represented in a language-agnostic knowledge graph in Coreon. Thanks to the integrated multilingual knowledge system (MKS), SmartBot is able to infer connections between language-agnostic concepts and deal with terms, previously unseen by the bot's language model.

Keywords

Chatbots, knowledge management, terminology management, knowledge graphs

1. Introduction

Nowadays, tools and data provided by public sector and private organizations still tend to be institutionally fragmented. The fragmentation of European e-government fabric triggered the emergence of interoperability solutions, to unify and simplify interaction between cross-border and cross-sector services. EuroVoc[1] and ISA2[2] belong to such inter-operable solutions, fostering uniformity within technical, semantic, organizational, and legal layers across the EU [1, 2]. The existing Standards for Public Sector Information (PSI) provision supply instruments to describe e-Government services in a uniform way. Yet they remain mostly unexploited and often lack user-centric design, let alone multilingual functionality that would support the official linguistic diversity of the EU [2, 3]. CEFAT4Cities project (2020-2022)[3] targets this challenge of interaction between EU residents, businesses, and public services, aiming to speed up the adoption of multilingual cross-border eGovernment services. Its main objective is a software layer that

[1]https://op.europa.eu/en/web/eu-vocabularies
[2]https://ec.europa.eu/isa2/isa2_en
[3]https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/2019-eu-ia-0015
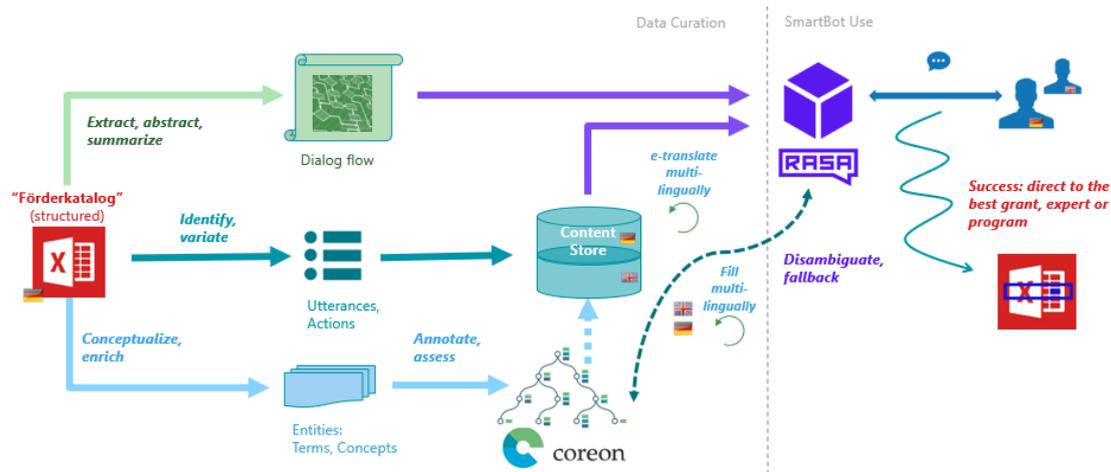
Figure 1: SmartBot's architecture.

facilitates the conversion of natural-language administrative procedures into machine-readable data (see [4] for details). Integrating its output into the existing EU resources, such as ISA2 and CPSV[4] that describe public services and associated life and business events [4], we created an open linked data repository, uniting concepts, relevant for businesses and citizens.

In this paper, we showcase how this resource is leveraged in a prototype of a real-life chatbot application, SmartBot, developed for the Vienna Business Agency (VBA)[5] in scope of the CE-FAT4Cities project. Lately, chatbots have started to emerge in various fields, featuring use-cases like information retrieval, service discoverability, customer service, and administrative workflows [5, 6, 7, 8]. Since dialogue is a natural way of interaction between humans, conversational agents designed to mimic this behaviour have potential to increase the efficiency of public services. In our case, SmartBot's goal is to automate and make VBA's services discoverable in a user-friendly way, targeting such topics as starting a new business in Vienna and helping users find relevant grants among hundreds of funding opportunities for companies of various scale.

2. Chatbot's architecture

Figure 1 displays SmartBot's architecture. The prototype is powered by Rasa Open Source[6], a framework for building conversational AI assistants. Domain knowledge as well as VBA-specific vocabularies are organised as a language-agnostic knowledge graph and curated in Coreon[7]. Rasa Open Source is modular by design: it consists of two primary components, Natural Language Understanding (NLU) and dialogue management (Rasa Core), and allows easy integration with other systems. NLU component is responsible for *understanding* the input received from the user: it handles intent classification and entity identification in users'

---

[4]https://ec.europa.eu/isa2/solutions/core-public-service-vocabulary-application-profile-cpsv-ap_en
[5]https://viennabusinessagency.at/
[6]https://github.com/RasaHQ/
[7]https://www.coreon.com/

utterances. The dialogue management component predicts the next action in a conversation based on the context. Rasa SDK handles all of our custom code: it is organised as custom actions that search databases, make API calls, trigger a handover of the conversation to a human, etc. Rasa Open Source is therefore adjustable to developers' needs, featuring straightforward integration and data control [9].

On top of Rasa, our architecture features an integration with Coreon Multilingual Knowledge System (MKS) [10, 11]. MKS is a a semantic knowledge repository, comprised of concepts linked via relations. Following the semantic web standards, it caters for visual discovery, access, drafting, and re-usability of any assets, organised in language-agnostic knowledge graphs. Since the linking is performed at the concept level, we can abstract from language-specific terms and model structured knowledge for phenomena that reflect the non-deterministic nature of the human language, such as word sense ambiguity, synonymy, homonymy, and multilingualism. Linking per concept ensures smooth maintenance of relations without additional data clutter: relation edges are independent from labels and terms and other metadata. It thus helps exchange information among acting systems and ensures that its precise meaning is understood and preserved among all parties, in any language.

3. Leveraging language-agnostic knowledge graphs

A big part of any chatbot's implementation is associated with domain data. In our case, a smooth cooperation in knowledge transfer is facilitated by MKS: VBA domain experts used it to model their domain knowledge, populating and curating it as a graph (see *Data Curation* side in Figure 1). The repository also incorporates the interoperability layer and public service-themed multilingual vocabularies. Aside from easy knowledge drafting, there are 4 concrete challenges that are tackled by the incorporation of language-agnostic knowledge graphs in virtual assistants: i) multilingualism; ii) language-independent entity management; iii) enabling semantic search; iv) dealing with homonymy and unseen terms.

In the European context, multilingualism is a big asset, yet it also brings along a conceptual challenge: the *kind* of multilingualism served tends to heavily influence the architecture and scalability of a solution. We decided to go with individual NLU models per language, i.e. keeping them language-specific, while making dialogue management – *Stories* – universal, adding an extra layer of abstraction to maintain consistency in bot's behaviour across languages. It implies that the core model should not have a single language-specific string among the training data, but rather an abstraction for the representation of entities, like language-independent IDs. We abstracted from entity maintenance in distinct languages, replacing language-specific terms in the NLU training data with their unique Coreon concept IDs. Maintaining entities in each language separately would be tedious and not consistent, particularly since the VBA domain knowledge is not static. Also, agnostic entities are crucial for keeping the Core module language-agnostic, abstracted from entity names in a specific language. Once VBA decides to expand SmartBot's language capabilities with a new language, this method of *universal entities* will ensure smooth model development and minimization of the labeling effort.
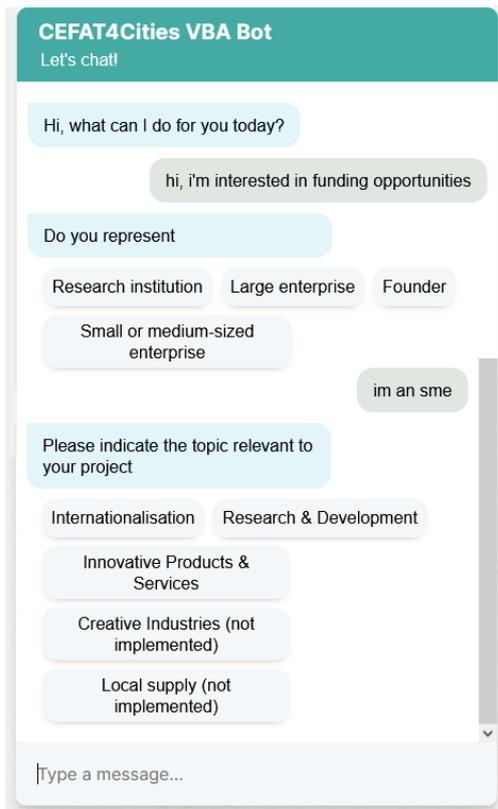
Figure 2: A demo dialogue snippet.

The core goal of SmartBot is to serve the user relevant grant recommendations based on previously provided input (see Figure 2 for a conversation snippet). This implies that the bot will have to fetch records with relevant VBA grants. To achieve this, we match information drawn from the user's input that influences the funding outcome (e.g., intents, entities, and their types extracted by Rasa NLU). Since grant information was also imported into MKS and each grant entry linked with relevant entity types, we can leverage these relations between concepts in the repository. With this functionality, SmartBot is able to fetch relevant funding entries even when terms extracted from the user's input are not explicitly appearing among VBA funding entries: the bot navigates parental and associative relations of the extracted entity and infers if there are any semantically close or connected concepts, linked with specific funding entries. Ultimately, we cover this scenario: given a VBA grant for SMEs focusing on environmental protection, a user X, searching for grants for small businesses doing roof planting/vertical gardening, and a user Y, looking for funding to support a startup that calculates CO2 footprint for businesses, would both land at the aforementioned grant.

Unseen terms and homonymy is tackled by the KG in the same fashion. If users choose to use terminology previously unknown to the model, SmartBot will first try to get its meaning using the connector to Coreon rather than taking a standard fallback. If a German user enquires about the amount of money they can get from VBA; they refer to money as *Kohle*, a slang term homonymous to *Kohle*, "coal", a fossil. The NLU model does not know this term, so the bot makes an API request, searches for it in MKS, and finds two hits in two distinct concepts. The first one belongs to **CO2** concept in a branch dealing with resource-saving and sustainability. The second one is found among synonyms for *Geldmittel*, denoting financial **funds** and has a more generic parent **Geld**, "money". Since quite a few *terms* of the concept **Geldmittel** are known to the NLU model and the context of the conversation is corresponding, the meaning of *Kohle* is disambiguated for the chatbot; subsequently, SmartBot informs the user about the amount of money they can qualify for.

Chatbots are becoming a turning point for rationalizing of business processes. Here we investigated technical feasibility and described the implementation of the prototype that can support VBA and serve the needs of Vienna residents, catering to the interaction in the language of their choice and understanding the intents of their requests.

Combining Rasa Open Source with reusable multilingual KG data, we delivered the intelligent chatbot solution, robust, extendable, and modular – a steady reference point for similar activities

to facilitate provision of PSI. Accommodating the chatbot interaction to the user's needs, VBA SmartBot automatically overcomes the language gap, contributing to the elevation of local public services to the European scale and red tape reduction.

Acknowledgments

References

[1] K. Bovalis, V. Peristeras, M. Abecasis, R.-M. Abril-Jimenez, M. A. Rodriguez, C. Gattegno, A. Karalopoulos, I. Sagias, S. Szekacs, S. Wigard, Promoting interoperability in europe's e-government, Computer 47 (2014) 25–33.

[2] E. Tambouris, Using chatbots and semantics to exploit public sector information, EGOV-CeDEM-ePart 2018 (2018) 125–132.

[3] E.-J. Mulder, D. Snijders, Playing the telephone game in a multilevel polity: On the implementation of e-government services for business in the eu, Government Information Quarterly (2020) 101526–101534.

[4] J. Van den Bogaert, A. Defauw, S. Szoc, F. Everaert, K. Van Winckel, A. Kramchaninova, A. Bardadym, T. Vanallemeersch, Cefat4cities, a natural language layer for the isa2 core public service vocabulary, in: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, 2020, pp. 483–484.

[5] H. Mehr, H. Ash, D. Fellow, Artificial intelligence for citizen services and government, Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., no. August (2017) 1–12.

[6] A. Stamatis, A. Gerontas, E. Tambouris, On using chatbots and cpsv-ap for public service provision, EGOV-CeDEM-ePart 2019 (2019) 133–139.

[7] S. M. Adnan, A. Hamdan, B. Alareeni, Artificial intelligence for public sector: chatbots as a customer service representative, in: International Conference on Business and Technology, Springer, 2020, pp. 164–173.

[8] C. Koch, B. Linnik, F. Pelzel, E. Sultanow, S. Welter, S. Cox, A reference architecture for on-premises chatbots in banks and public institutions, in: INFORMATIK 2021, Gesellschaft für Informatik, Bonn, 2021, pp. 1265–1281. doi:10.18420/informatik2021-106.

[9] D. Braun, A. Hernandez Mendez, F. Matthes, M. Langen, Evaluating natural language understanding services for conversational question answering systems, in: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, Association for Computational Linguistics, Saarbrücken, Germany, 2017, pp. 174–185. doi:10.18653/v1/W17-5522.

[10] M. Wetzel, Multilinguale taxonomien mit coreon. wissens- und sprachmanagement in einer lösung, Rechte, Rendite, Ressourcen. Wirtschaftliche Aspekte des Terminologiemanagements 14 (2014) 41–51.

[11] W. Ziegler, Metadaten für intelligenten content, Intelligente Information: Schriften zur Technischen Kommunikation 22 (2017) 51–66.