SMART CITIES AND EGOVERNANCE

Innovative Mind Mapping System Connecting to Smart City IoT Networks

With the contribution of

CEFAT**4Cities**   CROSS**LANG**

Challenge & Context

To ease interaction with a city's administrative services, the creation of a chatbot is an easy and popular option, with many open source platforms currently available (Rasa, built for multidisciplinary enterprise teams, for example, can be an excellent starting point). Filling the bot with the right content, however, is a whole different story. Since the aim of the bot is to do the searching – not the users or citizens themselves, the bot needs an accurate map that can predict exactly what a user is looking for, together with the relevant next steps to take or suggest. Normally, it takes a dedicated team of experienced editors to create such a mind map by collecting content, extracting all relevant information, and creating engaging user stories from this data, formulated using appropriate keywords, questions/answers pairs, and pointers to online content. On average, this whole process takes a couple of weeks to complete, even for a very limited use case. Imagine this for multiple use cases, in all EU languages administering a metropolitan area with citizens originating from all over the world.

Solution

To support cities in creating mind maps of their public services – or semantic networks as we call them with a more technical term – the CEFAT4Cities project[1], where FIWARE Foundation is one of the partners, has created a processing pipeline that ingests public services legacy data (from e.g. websites, administrative forms, existing applications) in multiple EU languages (Croatian, Dutch, English, French, German, Italian, and Norwegian) and transforms this data into a network of connected services that can be used across applications and

---

[1] The main objective of the CEFAT4Cities Action is to develop a "Smart cities natural language context", providing multilingual interoperability of the Context Broker DSI and making public Smart City services multilingual, with pilots in Vienna and Brussels.

languages. **By connecting the pipeline to the FIWARE Context Broker[2], the mind map is made available to any app or sensor within the Smart City IoT network.**

Why is this useful for a city? Taking as an example a chatbot for expats: any app developer can now collect all required information straight from the Smart City IoT network. This includes procedures to facilitate a citizen's onboarding, such as the registration of a business entity, the search for child day care or the practice schedule of the local football team. What if a city wants to boost tourism? By publishing opening hours of museums onto the Smart City network, **developers can easily integrate this information and combine it with other information they can find through the FIWARE Context Broker**, for example which pieces of art are currently on exhibition. By making all public services data available as Open Linked Data, app developers can combine this information easily with other city information and create software that helps users in a more productive way.

The CEFAT4Cities project was implemented by CrossLang[3], Coreon[4], the Brussels Business Agency (BECI)[5], the Vienna Business Agency (Wirtschaftsagentur Wien), and the FIWARE Foundation[6]. It was sponsored by the Connecting Europe Facility (CEF Telecom – Horizon2020) and received support from the CEF eTranslation Digital Service Infrastructure for the automated translation of public services data.

---

[2] The Context Broker Generic Enabler is the core and mandatory component of any "Powered by FIWARE" platform or solution. It enables the management of context information in a highly decentralised and large-scale manner. The Orion-LD Context Broker Generic Enabler is a NGSI-LD Broker, which supports the NGSI-LD and the NGSI-v2 APIs. With regard to Orion-LD and Scorpio, the 1.4.1 release came out in Q2/2021, 1.5.1 is planned to be released before the end of 2021. With version 1.0.0 (planned for Q4/2021), each release aligns with NGSI-LD 1.3.1 as defined by ETSI. Further releases are implementing new features defined by ETSI as the living specification evolves. FIWARE release 8.0 fully sets NGSI-LD at the core of Context Brokers.

[3] CrossLang is an independent privately owned consulting and systems integration company based in Ghent, Belgium. Helping clients adapt to market change and increase their wallets, getting leaner, faster and more capable in the process.

[4] Coreon GmbH is based in Berlin and Athens.Coreon is also a solution. Coreon is a knowledge system that enables humans to model, manage, and visualise knowledge in data streams and objects.

[5] Beci (Brussels Enterprises Commerce & Industry) is the result of a partnership between the Chamber of Commerce and the Union des Entreprises de Bruxelles. Beci represents two thirds of the employment in Brussels and more than 35,000 companies.

[6] The Vienna Business Agency supports local and international companies in all phases of their business development, providing advice on all corporate issues and helping expats get up and running in Vienna.

How it works

To create the semantic network of public services, CEFAT4Cities partners start from a few abstract templates that describe what a public service looks like (who can submit a form to get access to which service, providing which type of proof?) and what the interacting entities look like (are we dealing with an organisation or a citizen?). These abstract templates consist of nodes and links (hence the term "semantic network") and are provided by the European Interoperability Framework which governs data standards to ensure that data can be used across as many applications as possible.

Next, these templates are used as extraction filters to transform unstructured human natural language (occurring on websites, online forms, etc.) into machine-readable semantic networks which can be utilised in any software application. In short, thousands of pages of raw text are transformed into structural representations that can be used by machines, such as chatbots.

The process runs as follows: data is collected automatically from websites, then only those pages containing public service information are selected. Next, paragraphs describing administrative procedures are extracted and syntactically analysed to identify nodes occuring in the template. Finally, relations between the extracted nodes are identified (the most challenging part of the process) and the information is delivered in a standardised Open Linked Data format.

Note that the process designed can deal with all sorts of raw text in various EU languages.

To achieve all this, multilingual Artificial intelligence techniques such as automated classification are used, topic modelling, clustering, syntactic parsing, machine translation, unsupervised bilingual language induction, shallow parsing, paraphrasing, and question/answer pair generation.

Through a dedicated data schema, the Open Linked Data format used to publish results, is compatible with the FIWARE Context Broker. Any follow-up effort or
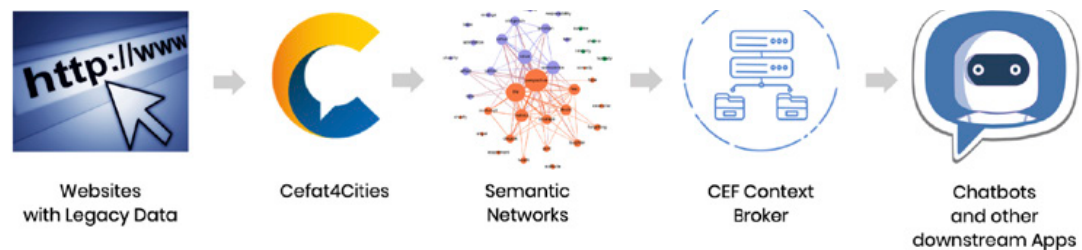
Figure 1 - Architecture

downstream software application can use this schema to publish or subscribe to the public services content created.

When developing the solution, several challenging issues were discovered. As described earlier, discovering links between nodes (connecting for example an administrative procedure and all the evidence a citizen must provide to fulfil it) proved to be a non-trivial task. A unique solution had to be built, combining syntactic parsing and classification, since no out-of-the-box components existed to do this. Throughout the pipeline, a balance was needed between using monolingual AI models and multilingual AI models using translated data, since many linguistic AI models only exist for a couple of languages.

Finally, often the language itself was problematic. Current AI models excel at "recognising" the meaning of a word when it appears within a larger body of text, but when words occur isolated (for example in a title or a table) recognition and translation become more difficult. In addition, there is also the typical "call-to-action language" used on websites. This sort of language (for example, "Need more info on the necessary formalities to start as a self-employed professional?") typically packages pieces of declarative information as questions, throwing the systems question/answer pair extraction off balance.

Benefits & Impact

The CEFAT4Cities project is currently coming to the end, but it has already impacted the way people think about public service data in two major European Cities: **The Brussels and Vienna Business agencies have successfully built a demonstrator chatbot with Open Linked Data generated by the CEFAT4Cities pipeline – AND they realise that the data can be shared and used for other purposes**.

Looking at the development costs of a multilingual AI-powered chatbot (i.e., a chatbot that "knows" your business domain, and not only knows how to greet and answer FAQs) **figures as high as €120,000 and a development time of about 12 months circulate around the Internet**. For this budget, one can get around 1000s of intents (questions users want to be answered) and 10,000s of example sentences to train a chatbot with. As a comparison, **when the CEFAT4Cities system was run on only 50 city web pages, it was able to automatically extract 10 intents and generate 21,000 questions in 7 languages in under 30 GPU minutes**, without human intervention. In addition to this, the system links questions and answers directly to web content though a NoSQL[7] index while it can scan city websites in real-time and update the generated data accordingly.

Admittedly, the generated data still needs human validation, but considering the rate at which the CEFAT4Cities system outputs data and takes over the heavy lifting from humans (manually researching the business domain, clustering topics, creating the mental model, extracting intents, compiling and annotating the data sets, extracting questions and answers, translating ,encoding the chatbot data, …), there is plenty of time saved that can be used for fine-tuning the produced data sets.

---

[7] NoSQL database provides a mechanism for storage and retrieval of data that is modelled in means other than the tabular relations used in relational databases.' but could only find a simple definition for database and not index.

The system currently exists as a prototype for the semantic modelling of public services in Croatian, Dutch, English, French, German, Italian, and Norwegian, with both the number of domains and languages expected to increase in the future.

From the onset of the project, the aim was to help smaller cities, as they have less means to build their own semantic network of public services, let alone to do this in a multilingual way. **Looking at the first results, it is believed this ambition can be achieved, provided that a sufficient amount of evangelisation is carried out**. The FIWARE community is definitely helping with this. Achieving this goal would greatly benefit smaller cities, as it will allow them to implement multilingual e-Government solutions at a much faster pace and contribute to the free movement of EU citizens in general.

Added value through FIWARE

The CEFAT4Cities project was supported by the **FIWARE Foundation** through its cooperation as a project partner. **With its philosophy of open APIs and the possibility to distribute Open Linked Data across a city network, the FIWARE Context Broker technology proved to be a perfect match for the CEFAT4Cities project**.

In addition, the interoperability and the interconnectivity **offered by the FIWARE Context Broker ecosystem, enabled the creation of powerful applications that can combine data in unexpected ways. For example, all metadata provided through the European Data Portal is immediately available through the FIWARE Context Broker. This opens opportunities for public service providers to link their data with related data that may be of interest for their citizens**. For example, adding cycling infrastructure data to the public services data set may provide citizens with ways of reaching the town hall in a carbon neutral manner, or it may inspire app developers to find the safest itinerary to reach the nearest park, library, language school or tennis court.

The reach of the FIWARE community network is immense, through its affiliations with Smart City organisations at European, regional, and national levels, connecting to potential stakeholders across the European Union was highly facilitated.

Next steps

The CEFAT4Cities project started with the ambition to take an inventory of public service catalogues across Europe and, if necessary, build them back up from scratch if they were not available, using whichever legacy data in whichever EU language available. As the project progressed recognizing the potential of such data and the possibilities when interconnecting it with existing data sets became increasingly evident.

The AI-driven chatbots developed, were initially conceived as educational tools to showcase the potential of semantically networked public services, but as it turned out, they constituted a business case in their own right.

CEFAT4Cities intends to disseminate the project's findings on as many Smart City platforms as possible, encouraging cities to take up this solution for the benefit of their citizens.

References

- CEF Context Broker Webinar – Start your Digital Transformation Journey
- CEFAT4Cities Project page
- Multilingual services for smart cities – European Association for Machine Translation 2020
- Van den Bogaert, J., Defauw, A., Szoc, S., Everaert, F., Van Winckel, K., Kramchaninova, A., … & Vanallemeersch, T. (2020, November). **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** (pp. 483-484).

Author & Contributors

Joachim Van den Bogaert
CTO at CrossLang
Contact @ joachim@crosslang.com

CrossLang – www.crosslang.com

Categories

Domains (s)	Smart Cities, Smart Tourism, eGovernance, Public Services
User (s)	Brussels Business Agency, Vienna Business Agency, European Cities, Public Administration, Government
Key words	Interoperability, Multilingual AI, Ontologies, Semantic Modelling, Open Data

Contact us

Having any questions? Want to contribute with another Impact Story?
Please contact **Tonia Sapia** @ tonia.sapia@fiware.org

Want to see more Impact Stories? Please visit www.fiware.org/impact_stories

FIWARE
Open APIs for Open Minds

SMART CITIES AND EGOVERNANCE

Innovative Mind Mapping System Connecting to Smart City IoT Networks

FIWARE
MARKET
PLACE
FIWARE

FIWARE

Be certified and featured
in the FIWARE Marketplace.

Never miss an update or a new
Impact Story. Join our Newsletter!

GO TO THE MARKETPLACE

SUBSCRIBE

Find Us On

Twitter   Facebook   LinkedIn   YouTube   Github