proceedings

MDPI

Proceedings

Designing an Open Source Virtual Assistant [†]

Anxo Pérez [*], **Paula Lopez-Otero** and **Javier Parapar**

Information Retrieval Lab, Centro de Investigación en Tecnoloxías da Información e as Comunicacións (CITIC), Universidade da Coruña, 15071 A Coruña, Spain; paula.lopez.otero@udc.es (P.L.-O.); javier.parapar@udc.es (J.P.)

[*] Correspondence: anxo.pvila@udc.es; Tel.: +34-881-01-1276

[†] Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Published: 21 August 2020

Abstract: A chatbot is a type of agent that allows people to interact with an information repository using natural language. Nowadays, chatbots have been incorporated in the form of conversational assistants on the most important mobile and desktop platforms. In this article, we present our design of an assistant developed with open-source and widely used components. Our proposal covers the process end-to-end, from information gathering and processing to visual and speech-based interaction. We have deployed a proof of concept over the website of our Computer Science Faculty.

Keywords: conversational assistant; chatbot; question answering; natural language processing; crawling; information retrieval

1. Introduction

Nowadays, conversational systems are part of our daily routines [1]. Tech giants are aware of their relevance, and they are incorporating these assistants to their platforms. Microsoft Cortana or Apple Siri are popularly used examples. Some companies, such as Google with the Assistant or Amazon with Alexa, are even manufacturing dedicated devices. Moreover, these services offer great opportunities for customer support [2]. Many companies' websites are gradually enabling conversational capacities to help users discover products and information [3,4]. On the other hand, voice commands have gained much attraction for user interaction [5]. There is a critical tendency to move towards audio controls over tactile interfaces [6]. The inclusion of voice capabilities was, therefore, a straightforward improvement for conversational agents. Apart from the business value of the technology, voice-enabled assistants are truly useful for people with functional diversity [7].

In this article, we propose an architecture for a conversational assistant. As we mentioned, our proposal covers the process end-to-end. We apply information retrieval, natural language processing, machine learning, and speech technologies to cover data acquisition to audio response and user questions.

2. Proposal

As mentioned previously, our architectural design involves all stages of the process. The system covers everything from information gathering and processing to visual- and speech-based interaction. For that, we have used models and techniques from different information processing fields. In this section, we will explain the process we have followed and the description of the technologies used in the development.

A web-crawler is in charge of the first step of the information-processing pipeline. In this phase, we retrieved the information from the domain webpages, and kept those documents up-to-date. For this task, we used Scrapy (https://scrapy.org/), a popular web scraper. It saves the data from the Internet and creates a repository containing the files to be indexed. The website (https://www.fic.udc.es/) contains documents both in Spanish and Galician. The second phase corresponds to text-processing, which includes sentence-splitting and indexing. We used ElasticSearch (https://www.elastic.co/), a distributed search engine based on Lucene, for both indexing and searching. We propose the use of the ElasticSearch identification component for tagging the documents. As we were building a conversational system, we indexed the data at both the document- and sentence-level. Indexing isolated sentences allowed us to answer many of the user's questions concisely and directly.

Our design accepts the user's input, both through writing and spoken queries. For processing voice queries, we used Kaldi (https://kaldi-asr.org/) for Automatic Speech Recognition (ASR). Finally, we provided a system response in text and audio format using Cotovia (http://gtm.uvigo.es/cotovia). Here, we again used language identification models to process user inputs and outputs correctly. In the case of voice interaction, we trained the automatic speech recognition language models with specific domain lexica [8]. On the voice response side, the system reproduces the responses, selecting the language accordingly to the user input. In this case, we used Cotovia pre-trained models to perform speech synthesis [9].

One crucial problem of spoken document retrieval is term misrecognition. This problem provokes the inability to process the information need correctly. ASR misrecognition produces term mismatch between user input and document content. We used efficient state-of-the-art retrieval models [10] based on n-gram decomposition in dealing with it. The system processes both searchable content and user input in that way to allow fast and robust query matching. These models achieve state-of-the-art effectiveness figures, while also being quite efficient [11].

For answering information needs, we designed a four-level cascade system. First, the system tries to classify the user's intent in some predefined structured tasks (e.g., the timetable for a subject or the date of an exam). If the input falls onto one of those categories, the answer is processed according to the defined pipelines. Second, if that was not the case, the system attempts to provide a direct answer to the specific user question. For that, we propose to use two approaches: best-sentence-matching and BERT-based question-answering [12]. Thirdly, there is no satisfactory direct answer, the system tries to provide the best document answer. Finally, if the system does not rank satisfory documents, it asks the user to reformulate the question.

Architecturally speaking, we are thus using a basic client-server application. The web client communicates with the backed-through rest services and WebSocket APIs . For the web interface, we used BotUI (https://github.com/botui/botui), a very intuitive Javascript library for conversational interfaces. The server contains the implementation of the different REST endpoints and WebSocket APIS for processing audio streams.

3. Conclusions and Future Work

In this article, we presented our design of a conversational assistant based on open-source and well-known technologies. Even though we exemplified the design for a specific web domain for its implementation, the architecture introduced here can consume other information repositories, such as different enterprises data information or any databases.

There are many avenues for future work. We propose to improve the architecture with advanced Natural Language Generation (NLG) capacities. The fine-tuning of acoustic models for specific language variants is another interesting research address. When not depending on languages, such as Galician, we would favor the use of Tacotron as the TTS engine [13].

Funding: This work was supported by projects RTI2018-093336-B-C22 (MCIU & ERDF) and GPC ED431B 2019/03 (Xunta de Galicia & ERDF). Also, this work has received financial support from CITIC, *Centro de Investigación del Sistema universitario de Galicia*, which is financial supported by *Consellería de Educación, Universidade e Formación Profesional* of the *Xunta de Galicia* through the ERDF (80%) and *Secretaría Xeral de Universidades* (20%), (Ref ED431G 2019/01).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Masche, J.; Le, N.T. A Review of Technologies for Conversational Systems. In *International Conference on Computer Science, Applied Mathematics and Applications*; Springer: Cham, Switzerland, 2018; pp. 212–225, doi:10.1007/978-3-319-61911-8_19.
2. Brandtzaeg, P.B.; Følstad, A. Why People Use Chatbots. In *Internet Science*; Kompatsiaris, I., Cave, J., Satsiou, A., Carle, G., Passani, A., Kontopoulos, E., Diplaris, S., McMillan, D., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 377–392.
3. Chung, M.; Ko, E.; Joung, H.; Kim, S.J. Chatbot e-service and customer satisfaction regarding luxury brands. *J. Bus. Res.* **2018**. doi:10.1016/j.jbusres.2018.10.004.
4. Majumder, A.; Pande, A.; Vonteru, K.; Gangwar, A.; Maji, S.; Bhatia, P.; Goyal, P. An Approach Based on Category-Sensitive Retrieval. In *European Conference on Information Retrieval Automated Assistance in E-commerce*; Springer: Cham, Switzerland, 2018; pp. 604–610, doi:10.1007/978-3-319-76941-7_51.
5. Bhalla, A. An exploratory study understanding the appropriated use of voice-based Search and Assistants. In *IndiaHCI'18: Proceedings of the 9th Indian Conference on Human Computer Interaction*; Association for Computing Machinery: New York, NY, USA, 2018; pp. 90–94, doi:10.1145/3297121.3297136.
6. Porcheron, M.; Fischer, J.E.; Reeves, S.; Sharples, S. Voice Interfaces in Everyday Life. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–12, doi:10.1145/3173574.3174214.
7. Brewer, R.N.; Findlater, L.; Kaye, J.J.; Lasecki, W.; Munteanu, C.; Weber, A. Accessible Voice Interfaces. In Proceedings of the Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '18, New York, NY, USA, 3–7 November 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 441–446, doi:10.1145/3272973.3273006.
8. Povey, D.; Ghoshal, A.; Boulianne, G.; Burget, L.; Glembek, O.; Goel, N.; Hannemann, M.; Motlíček, P.; Qian, Y.; Schwarz, P.; et al. The Kaldi speech recognition toolkit. In Proceedings of the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Waikoloa, HI, USA, 11–15 December 2011.
9. Banga, E.R.; Mateo, C.G.; Pazó, F.J.M.; González, M.G.; Magariños, C. Cotovía: An open source TTS for Galician and Spanish. In Proceedings of the "IberSPEECH 2012:" "VII Jornadas en Tecnología del Habla" " and III Iberian SLTech Workshop", Madrid, Spain, 21–23 November 2012.
10. Lopez-Otero, P.; Parapar, J.; Barreiro, A. Statistical language models for query-by-example spoken document retrieval. *Multim. Tools Appl.* **2020**, *79*, 7927–7949, doi:10.1007/s11042-019-08522-z.
11. Lopez-Otero, P.; Parapar, J.; Barreiro, A. Efficient query-by-example spoken document retrieval combining phone multigram representation and dynamic time warping. *Inf. Process. Manag.* **2019**, *56*, 43–60. doi:10.1016/j.ipm.2018.09.002.
12. Carrino, C.P.; Costa-jussà, M.R.; Fonollosa, J.A.R. Automatic Spanish Translation of the SQuAD Dataset for Multilingual Question Answering. *arXiv* **2019**, arXiv:1912.05200.
13. Wang, Y.; Skerry-Ryan, R.; Stanton, D.; Wu, Y.; Weiss, R.J.; Jaitly, N.; Yang, Z.; Xiao, Y.; Chen, Z.; Bengio, S.; et al. Tacotron: Towards End-to-End Speech Synthesis. *arXiv* **2017**, arXiv:1703.10135.