

AIDA-Bot: A Conversational Agent to Explore Scholarly Knowledge Graphs

Antonello Meloni¹, Simone Angioni¹, Angelo Salatino², Francesco Osborne²,
Diego Reforgiato Recupero¹, and Enrico Motta²

¹ Department of Mathematics and Computer Science, University of Cagliari (Italy)
{antonello.meloni, simone.angioni, diego.reforgiato}@unica.it
² Knowledge Media Institute, The Open University, Milton Keynes (UK)
{angelo.salatino, francesco.osborne, enrico.motta}@open.ac.uk

Abstract. Chatbots have become increasingly popular in the last years among both businesses and consumers. These conversational agents typically use natural language understanding technologies for reformulating questions as queries to a knowledge base and answer the user according to the resulting information. This demo paper proposes a general software architecture for chatbots that can be employed for exploring large-scale knowledge graphs of scholarly data. It also introduces AIDA-Bot, a working prototype that implements this architecture and can answer queries about scientific publications, research topics, researchers, universities, and conferences. In order to show the flexibility of the proposed solution, we implemented two versions of AIDA-Bot which can run, respectively, on Alexa devices and web browsers.

Keywords: Scholarly Data · Knowledge Graphs · ChatBots · Virtual Assistant

1 Introduction

Chatbots have become increasingly popular in the last years among both businesses and consumers. These conversational agents are able to assist users by providing responses to questions and carrying out a variety of tasks such as customer care, ordering items, booking tickets, giving driving directions, and so on. They typically use natural language understanding technologies for reformulating questions as queries to a knowledge base and answer the user according to the resulting information [1]. Knowledge Graphs (KGs) are becoming a standard solution for representing these kinds of structured data. Therefore, recent chatbots need to interact with KGs and allow the users to ask a range of relevant questions for navigating the information they describe.

In this context, scientific research represents a fascinating domain to explore. Today, we have several large-scale knowledge graphs which describe research



Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

papers and other research outcomes such as OpenCitations, Scopus, Semantic Scholar, Aminer, CORE, ORKG, and others. Some recent solutions such as the Open Research Knowledge Graph [4] and the Artificial Intelligent Knowledge Graph [3] even offer a structured representation of research entities such as approaches, tasks, material, and metrics. An intelligent conversational agent able to exploit this great mass of scientific knowledge could effectively support researchers, students, developers, and other users interested in scientific research.

As first step in this direction, this demo paper proposes a general software architecture for chatbots that can be employed for exploring scholarly data. We also introduce AIDA-Bot, a working prototype that implements this architecture in order to answer queries about scientific publications, research topics, researchers, universities, and conferences. AIDA-Bot exploits the Academia/Industry DynAmics (AIDA) Knowledge Graph³ [2], which includes 21M publications annotated with the research topics in the Computer Science Ontology (CSO) [6]⁴. In order to show the flexibility of the proposed architecture, we implemented two versions of AIDA-Bot. The first one relies on the Amazon Alexa Skill Kit and can be run on Alexa devices⁵. The second one uses standard web technologies (Javascript, JQuery, Web SpeechAPI) and can thus be run from a web browser. This demo is available at <https://w3id.org/aida/bot>.

In summary, the contributions of this demo paper are the following: 1) we propose a general architecture for building chatbots in the scholarly domain; 2) we formalize and develop a set of questions that can be asked to the chatbot using a knowledge graph of scholarly metadata; 3) we implement two versions of AIDA-Bot for browsers and Alexa devices.

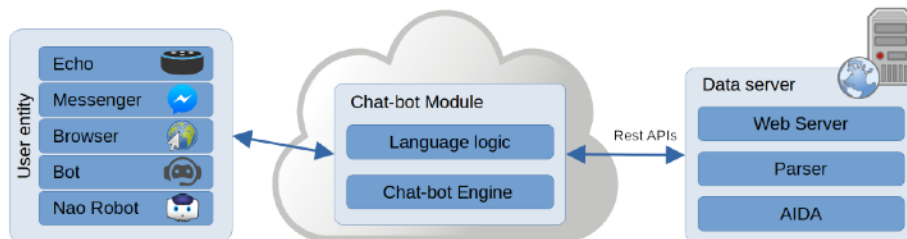


Fig. 1. System Architecture.

2 The Chatbot Architecture

The general system architecture is shown in Figure 1. It is composed by three parts. The first one is the *User Entity*, which can be a voice assistant device (e.g., Amazon Echo), an Instant Messaging application (e.g., Facebook Messenger, Telegram), a browser, a chatbot, a robotic platform (e.g., NAO [5], Pepper) or

³ Academia/Industry DynAmics Knowledge Graph - <http://w3id.org/aida/>

⁴ Computer Science Ontology - <http://w3id.org/cso/>

⁵ Video showing how to install AIDA-Bot on Alexa https://youtu.be/7ANn_u-zX1Q

any other device where the user directly interacts by sending voice (talking to the device) or text (writing to the device).

The second part is the *Chat-bot Module*, which includes the Chat-bot Engine and the Language Logic module. The Chat-bot Engine takes care of managing events and communications between the user and the Data Server. Specifically, it handles the conversion from speech to text and from text to speech and verifies that all the required information to formulate a SPARQL query to the Data Server have been acquired. The Language Logic module translates the results of the query, returned by the Data Server, into natural language.

The third part is the *Data server*, which includes the Web Server, the Parser, and the Knowledge Graph (e.g., AIDA). The Parser extracts from the user question all the parameters necessary for determining the query type and generating the query. It also interacts with the Knowledge Graph for verifying the entities extracted from the text (e.g., *machine learning*) and determining their types (e.g., *topic*).

Each session begins with a welcome message and an invitation to enter a question. The user input is sent from the *User entity* to the *Chat-bot Module* and handled by the Chat-bot Engine which sends it via the Web Server to the Parser. The Parser returns all the parameters extracted from the user question. If one or more required parameters are missing, the Chat-bot Engine asks them to the user. It then generates a SPARQL query for retrieving all the relevant information and sends it to the Data Server, which in turn queries the Knowledge Graph and returns the results. Finally, the Chat-bot Engine uses the Language Logic module for converting the results to natural language and sends them back to the *User Entity*.

3 AIDA-Bot

AIDA-Bot allows users to query information from the AIDA knowledge graph [2], which includes the metadata of 21M publications in the field of Computer Science. AIDA was generated using an automatic pipeline that is fed with data from Microsoft Academic Graph, Dimensions, DBpedia, the Computer Science Ontology [6], and the Global Research Identifier Database. It is publicly available under CC BY 4.0 and can be downloaded as a dump or queried via a triplestore from <http://w3id.org/aida/sparql>.

The users can ask questions about six classes of items (*authors, papers, conferences, organizations, citations, and topics*) and their instances (e.g., *Yoshua Bengio, University of Oxford, ISWC*). The current version of AIDA-Bot supports three types of queries: *count, list, and describe*.

The *count* query allows the user to count the number of items of a certain class *C* within a context specified by a certain instance *I*. For instance, “*count papers on machine learning*” or “*count authors in ISWC*”. The *list* query allows the user to list the top *N* items of a class *C* according to an order *O* within a context specified by an instance *I*. Some examples are “*list the top 3 authors in ISWC ordered by publication number*” or “*list the top 5 papers in Semantic*”

Web order by citations in the last five years". Finally, the *describe* query returns the description of a certain instance *I* according to a template used for its class. For instance, the user can say "*describe ISWC*" and the system will return a number of pre-computed analytics on ISWC, including standard academic metrics (e.g., citations, h5-index), as well as the main research areas, universities, companies, and countries involved in the conference. Currently, the describe query is only available for authors, conferences and organizations, but we are working on expanding it to all the classes in AIDA.

The engine automatically recognizes the most common syntactic variations of these questions. If the user forgets to specify a required element, the chatbot will ask for it. The user can also simply use the query keyword (e.g., *count*) and the chatbot will guide them through the process by asking for all the required parameters. The instances are automatically resolved and assigned to their class. If an instance is ambiguous (e.g., *ISWC*), the chatbot will ask the user to choose one of the alternatives (e.g., *International Semantic Web Conference*, *International Symposium on Wearable Computers*).

As previously mentioned, we developed two different solutions to interact with AIDA-Bot. The first one is based on the Alexa Skill Kit⁶ and can be run on Alexa devices. This solution uses the Node.js technology for the Chat-bot module. The full code is available at <https://github.com/infovillasimius/aida>. The second solution can be run on a standard web browser and is based on Javascript and JQuery. If the browser is compatible with the Web Speech API (e.g., Google Chrome, Safari, Opera), it also provides voice support.

4 Conclusion

In this demo paper we presented an architecture for chatbots that takes advantage of large-scale knowledge graphs of scholarly data and AIDA-Bot, a conversational agent that allows users to query the AIDA knowledge graph. The architecture is general as each module can be replaced or extended and used to query other scholarly knowledge graphs. It can also support several different devices to interact with the conversational agent. AIDA-Bot is currently available on the web and Alexa and we plan to further develop it in the following months. We are now working on evaluating AIDA-Bot with different kinds of users and collecting the feedback of the research community. We also intend to further develop its functionalities and use it as a base to develop a complete virtual assistant for researchers.

References

1. Ait-Mlouk, A., Jiang, L.: Kbot: A knowledge graph based chatbot for natural language understanding over linked data. *IEEE Access* **8**, 149220–149230 (2020). <https://doi.org/10.1109/ACCESS.2020.3016142>

⁶ <https://developer.amazon.com/en-US/alexa/alexa-skills-kit>

2. Angioni, S., Salatino, A.A., Osborne, F., Recupero, D.R., Motta, E.: Integrating knowledge graphs for analysing academia and industry dynamics. In: ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium. pp. 219–225. Springer (2020)
3. Dessì, D., Osborne, F., Recupero, D.R., Buscaldi, D., Motta, E., Sack, H.: Ai-kg: an automatically generated knowledge graph of artificial intelligence. In: International Semantic Web Conference. pp. 127–143. Springer (2020)
4. Jaradeh, M.Y., Oelen, A., Farfar, K.E., Prinz, M., D’Souza, J., Kismihók, G., Stocker, M., Auer, S.: Open research knowledge graph: next generation infrastructure for semantic scholarly knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture. pp. 243–246 (2019)
5. Recupero, D.R., Spiga, F.: Knowledge acquisition from parsing natural language expressions for humanoid robot action commands. *Inf. Process. Manag.* **57**(6), 102094 (2020). <https://doi.org/10.1016/j.ipm.2019.102094>, <https://doi.org/10.1016/j.ipm.2019.102094>
6. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Birukou, A., Osborne, F., Motta, E.: The computer science ontology: A comprehensive automatically-generated taxonomy of research areas. *Data Intelligence* **2**(3), 379–416 (2020). <https://doi.org/10.1162/dint-a-00055>