DESIDOC Journal of Library & Information Technology, Vol. 40, No. 6, Nov 2020, pp. 329-333, DOI : 10.14429/djlit.40.6.15611

Conceptualising a Library Chatbot using Open Source Conversational Artificial Intelligence

Mayukh Bagchi

Department of Information Engineering and Computer Science (DISI), University of Trento, Italy
E-mail: mayukh.bagchi@unitn.it

ABSTRACT

Conversational software, or chatbots in popular parlance, have been considered a potential game changer for people-centric institutions for some time now. In sync with the trend, various libraries and knowledge resource centres have also adopted chatbots within their technological fold, with an aim to provide improved services to patrons. The only bottleneck towards such implementation has been the dearth of open source conversational software platforms. The purpose of the paper is to conceptualise a novel library chatbot using a recently developed, artificial intelligence-powered open source conversational software platform named Rasa, and to propose its potential adoption by libraries. The paper introduces the essence of chatbot technology and their present-day application in libraries. The paper also illustrates the technical underpinnings of Rasa Stack that can be leveraged to develop a library chatbot, and reflects on the potential future research in this direction.

Keywords: Chatbot; Conversational software; Conversational AI; Natural language processing; Dialogue management; Rasa OpenStack; Human computer interaction; Information retrieval.

1. INTRODUCTION

The yearning of humans to converse with a computer system dates back to the inceptive days of computer science. Alan Turing's exposition on imitation games and learning machines, popularly known as turing test, provided the embryonic thrust to the possibility of computers having natural and intelligent conversations with humans[1]. Over the years, with the disruptive evolution of Information and Communication Technologies (ICT), computer based conversational agents, or chatbots in popular parlance, finally became a reality. In simple terms, chatbots are goal-based computer programs developed to imitate and effect intelligent conversational behaviour similar to humans, through the integration of an array of suitable computing technologies. Motivations behind developing such systems included real-time intelligent patron engagement and support, austerity in serving costs and the urge for increased revenues, amongst others. The generic characteristics envisioned and expected from chatbots like intent recognition, entity extraction, dialog automation and management, anthropomorphism and feedback-based correction were in sync with the motivational attributes mentioned before[2]. There have also been numerous parametric classifications of chatbots with respect to their domain of activity, methods of operation or the kind of services provided. The most significant among them, though, remains the categorisation into chatbots powered either by state machines or by artificial intelligence (AI). State machines (also known as finite automaton) are abstract machines storing a singular state amongst any finite number of states at a given time (essentially, operating through nested conditional programming constructs), and transitioning between states in response to inputs[3]. Artificial Intelligence (AI), on the other hand, implies implementation of concepts and algorithms from sub-arenas within AI (such as Machine Learning, Knowledge Representation etc.) to achieve a certain degree of learning capability and intelligence so as to achieve certain stated goal(s) with considerable autonomy. It is but natural to mention that the AI-powered chatbots would be much more inherently adept in handling unpredictable semantics than their state machine-based counterparts. Also, a long-standing matter of concern in the development of conversational software remains the dearth of effective and powerful open source software platforms. It is needless to mention the exorbitant cost of proprietary conversational AI technologies, in terms of purchase, support and human resource training. An open source conversational AI platform will have advantages like community support, customised training datasets and adaptive integration with backend systems, without any of the extra cost except technical training. Rasa Stack provides a powerful yet easy-to-use python-based open source software alternative to build contextual conversational agents[2].

The overarching thrust of the paper is to motivate libraries and allied knowledge-intensive institutions to adopt conversational software in general, and the technological finesse of Rasa in specific. The rest of the paper is organised as follows: Section 2.0 provides a prefatory glimpse of the state-of-the-art implementation of chatbots in various libraries, and related issues. Section 3.0 introduces the technical features of the Rasa open stack, and elucidates a possible conceptual architecture for a library chatbot utilizing it. Section 4.0

Received : 24 March 2020, Revised : 21 September 2020
Accepted : 27 October 2020, Online published : 03 December 2020

concludes the paper by identifying the possible library services where chatbots can be of assistance, and specifying the future research directions.

2. LITERATURE REVIEW

Libraries have always been at the forefront of recognizing, absorbing and effectuating technology into their folds as compared to other academic and memory institutions. Some have even considered adopting artificial conversational entities for purposes like reference interviews, in an attempt to re-attract users to utilise their services[4]. In fact, the University of Nebraska-Lincoln library's Artificial Intelligence Mark-up Language (AIML) based Pixel chatbot (essentially, based on pattern matching) was among the first such chatbots to go live in the USA[5]. There has also been a serious deliberation among Canadian libraries about incorporating chatbots and aligned Human Computer Interaction (HCI) techniques in facilitating dedicated library services such as web site navigations, digital reference interviews and virtual story narrations[6]. A collaborative project shared amongst select Swiss public libraries, enterprises and information science student groups developed the first Swiss library chatbot – Kornelia (also the first public library chatbot in the world)[7]. A library chatbot prototype has been developed at the University of Technology Sydney (UTS) based on a comprehensive explanatory study of the role of librarians in ensuring the friendly and trustworthy conversational design of a library chatbot[8]. The possible utilisation of chatbots in special libraries such as law libraries have also been discussed and researched upon[9]. The role of chatbots as instructional and learning technologies employed in different institutional arenas has also been enthusiastically debated by the scientific community[10].

Generally speaking, there are several insights from the studies above[2,4,5,7,8] which necessitates the adoption and implementation of AI powered open source conversational software platforms like Rasa in libraries, and reinforces its superiority and novelty. Firstly, almost all of the existing library chatbots have been developed on platforms which are essentially either tuned for small-scale use (lacks scalability) or designed for exclusive research purposes. None of the chatbots cited above can satisfactorily handle complex, contextual dialogue management scenarios (which requires a strong embedded grounding of Machine Learning techniques) and composite Application Programming Interface (API) interactions. Professional documentation and ease of learning are the other critical features which many of these chatbot development platforms lack. Finally, most of the chatbot development software tools above are not completely platform independent.

3. TECHNICAL ARCHITECTURE

The methodological architecture proposed in the paper utilises the Rasa Stack[2] (Fig. 1) as the principle family of open source conversational AI technologies in conceptualizing the library chatbot. It is further refined and enhanced by other allied technological paradigms within the broad scope of AI. The architectural framework has been discussed in its entirety in the following subsections with relevant mappings to potential usage within library landscapes, wherever appropriate.

3.1 About Rasa Stack

The Rasa Stack is an integrated, modular machine learning framework comprising highly scalable, python-based open source libraries for developing semantically enriched artificial conversational entities, i.e. chatbots. The Rasa Stack is comprised of the following two main independent components:

- *Rasa NLU* is the Rasa Stack's go-to library for Natural Language Understanding (NLU). Combining statistical techniques from Natural Language Processing (NLP) and Machine Learning (ML), it attempts to teach the chatbot

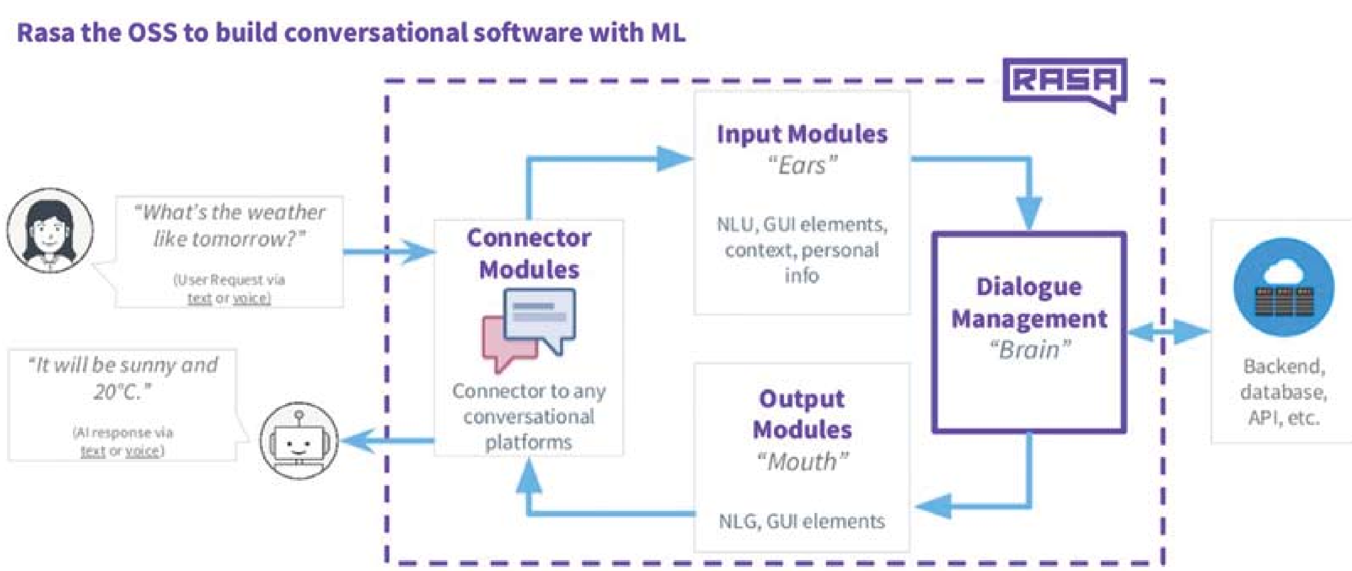

Figure 1. Rasa stack (Source: https://mc.ai/conversational-ai-chatbot-using-rasa-nlu-rasa-core-how-dialogue- handling-with-rasa-core-can-use/).

to understand the message of the user (for example, the information needs of a patron expressed through a digital reference interview in an academic library). The library has the potential to train on real-time data generated from users interacting with the chatbot. The two pivotal processes carried out by the library on the user-generated data are intent classification and entity extraction.

- *Rasa Core* is an ML-based library devoted to dialog management within the Rasa Stack. It attempts to predict the best-suited subsequent action that the chatbot has to take (like suggesting a list of reference sources matched against the information need of a library patron), based on a combination of parameters like the message as understood by the Rasa Core, the constantly refining training data, and, the state of the conversation including its historical precedents. It has robust Natural Language Generation (NLG) capabilities.

The generic workflow of the Rasa Stack in a chatbot developed on its modular technological premise involves (conceptually mapped to potential library use-case scenarios):

- Connector module, which are essentially Conversational User Interface (CUI) platforms (custom-developed or third party). It acts as a conduit for input of messages and output of responses with respect to the patron – chatbot interactions, just like the entire lifecycle of a reference interview involving patron(s) and librarian(s).
- Input module, where Rasa NLU comes into play in combination with the training data, attempts to extract structured information in the form of intents and entities from the message of the patron, just like when a reference librarian tries to analyse and decode the exact information need of a patron as expressed in a reference interview.
- Dialog Management module employs the Rasa Core library on top of the extracted structured information to frame an action (can be iterative), which can be a response based on NLG, a database connection attempt or an API call. This is akin to, for example, the comprehensive resource search and retrieval that a librarian proceeds with to satisfactorily cover the information needs which a patron expresses.
- Output module, to convey the response of the chatbot to the user through the connector module. It can leverage Rasa Core's template-based response functionality or it can also feature an open source third-party NLG plug-in. This is analogous to the list of reference resources compiled and delivered by a librarian to a user in satisfaction of his/her information need.

Finally, some of the distinctive features of the Rasa Stack in comparison with other hosted solutions in conversational software are:

- It runs locally and has no binding whatsoever with respect to any specific platform. It can either be hosted on a suitable cloud-based platform or any preferred server. As a result, there are no associated network overheads and the quality of service is also maintained
- Since Rasa Stack is not hosted, it empowers the administrator of the chatbot to be in charge of any operational data and avoid proprietary lock-ins
- Being a modular open source software stack, its source code can be tuned, trained and modelled as per the requirements of the scope of the chatbot, and it also has active online community support.

3.2 Knowledge Space

Knowledge Space is the arena of knowledge activity around which the functioning of the chatbot revolves. It is also known as knowledge domain in the parlance of formal sciences. It can be construed as a sphere of discourse where the chatbot deliberates over data variously representing intents, entities, attributes, relationships and responses, reined in by a set of common minimum constraints. General chatbots are the ones which are not devoted to any particular knowledge arena and can attempt to strike a conversation surrounding any theme in general, whereas, domain- specific chatbots are specially conceptualised to understand and respond to intents and entities in a particular knowledge space, and, can fail to provide an appropriate response to a message outside of its scope. In the same breadth, chatbots conceptualised for public or academic libraries are more general in nature as compared to chatbots for special libraries designed to support the information requirements of scholars, say, in a research institute devoted to international affairs research.

3.3 Natural Language Understanding

The technological foundation of the proposed architecture's natural language understanding capabilities rests on Rasa NLU[2]. The main focus of this stage of the architecture remains extraction of structured information from unstructured text input (mostly, in the form of user's messages)- encapsulated through intents and entities. Intent refers to the intention that the user is expressing through his/her message in natural language, and entities are the extracted pieces of structured information which would contribute in helping the chatbot to chart out well-reasoned actions. For example, a dedicated patron of a public library might approach its chatbot (which can appear somewhere in the public library's homepage) for updating his/her email id. To emulate a near empathetic conversation, the chatbot expresses "Hello! How can I help you?" to which the patron responds: "Please update my email id. It is abc@example.com". The role of the NLU apparatus at this stage would be to identify the intention of the message as 'email change' and to further extract the new email of the patron as a structured entity for future actions (for example, like online delivery of the daily literary digest). Intent classification in Rasa NLU is usually done through a step-by-step machine learning mechanism. The input message text is first tokenised into a bag of words, and the tokens are assigned word vector representations through any suitable language - specific pre-trained word vector library. Further, an average of the word vectors is taken to synthesise them into a sentence representation mathematically, which are then fed into a multi-class classifier like Support Vector Machine (SVM) which labels the user input with a specific intent as per the confidence score quantified by the predictive model (for example, something like confidence (email_change) = 98 per cent, and confidence (email_delete) = 2 % in our case). Besides, Rasa has also been integrated with an approach named "supervised word vectors", which can uncover multiple intents of an input message through the optimisation of a similarity-detecting mathematical function. The uniqueness of the approach is that it doesn't need any language-specific pre-trained word vector library, thus lending it a domain-specific multilingual character. The entity extraction workflow also works simultaneously with the intent classification. It tokenises the input message and assigns grammatical parts of speech to each token. These are further made to pass through a chunker which aids in training the model to unsheathe multi-worded entities (like addresses or name of books) and semantically annotates these chunks using Named-Entity Recognition (NER), with the final resultant as labelled extracted entities (for example, Rabindranath Tagore : Person, abc@example.com: Email). The role of classifiers as regards the entity extraction process in Rasa NLU is also very interesting. It employs a directed structured prediction approach using Conditional Random Fields (CRF)[11] instead of multi-step approaches like utilizing a binary classifier route. CRFs take into account the context of a message using the neighbouring tokens, and, sequentially labels the succession of tokens in the form of entities in one-go. Further, Rasa also offers capabilities to predict intents and entities together in-a-go using a well-informed pipeline employing the above technologies.

3.4 Dialogue Management

In addition to NLU, dialogue management also forms the crux of Rasa's technological stack[2]. It comprises Rasa Core along with its NLG features, actualizing rationalised actions and human-like conversations by the chatbot. Rasa's approach is completely based on machine learning (essentially, reinforcement learning), wherein, instead of rule-based systems, the back-end model is trained and refined on real-scenario conversational data (for example, conversational data from the chatbot's interaction with library patrons in the past 15 days) which can be effectively utilised in approximating such conversations. In general, structured information extracted during the NLU phase like intents and entities are taken as inputs in this stage and combined with indicators reflecting the present state as well as the past actions (such as API calls, database look-ups, knowledge base search etc.) taken by the chatbot, to predict the next best possible action (such as updating the profile of a patron). In the Rasa Core[2], this is principally done through deep learning artefact like Recurrent Neural Networks (RNNs), which takes into account contextual state information through its exhibition of dynamic behaviour, and updates the state of the conversation through suitable action. There is also an optional information security embedding in the Rasa stack known by the name of action mask, which when effectuated over and above RNNs, can help provide layered authentication control over different levels of users. If incorporated within the technical architecture, it also compulsorily includes a re-normalisation layer so as to effect proper action which can further be an API call or a response in the natural language, and, the iterations continue as required. Rasa Core is also distinct in terms of its learning and refinement capabilities with respect to feedback provided by the users. Its Interactive Learning functionality enables the chatbot to detect disapprovals from users (extracted from feedback about the chatbot service from library patrons), retrain its model and refine its responses to a much greater extent- all while being in live conversation with the library patron. Finally, there must be an ambient Conversational User Interface (CUI) at the front-end to facilitate conversation with patrons, designed implementing the best principles and practices from HCI. Researchers at Rasa have also developed the "Recurrent Embedding Dialogue Policy (REDP)" framework which ingrains states and actions from dialogue management into a single representative vector space which facilitates learning and reusing dialogues across domains[12] (thus lending the AI powered conversational engine a domain-agnostic character to be utilised and re-utilised across diverse knowledge-intensive institutions).

4. CONCLUSIONS AND FUTURE WORK

In the context of practical implications in library and information services, it would be relevant to note that though open source AI-based conversational software requires more in-depth research in order to perfectly learn and emulate human conversations, it nonetheless has boundless powers to induce a tectonic shift in the way libraries function and interact with their patrons. It is also pertinent to note that for widespread implementation of such technology, extensive technical training for specialised library personnel, and a significant budgetary allocation is required. As already established through the use of predominantly rule-based conversational software, chatbots can have wide-ranging applications in different facets of library services like:

- virtual reference interviews and services, wherein the developed library chatbot can understand the patron's information need via automatic query processing, and return appropriate, authoritative information sources as results
- 24*7 patron support which will enable information resource centers to be connected full-time and in real-time with patrons
- community information services, another important dimension of information services where conversational AI based library chatbots can play a significant role. For example, it can aid in automatic dissemination of authoritative information and downplaying of misinformation in case of risk management during pandemics (like the COVID-19 pandemic[13])
- library resource suggestions through its novel technique of intent elicitation based on Natural Language Understanding, wherein the patron's intent (mood/emotion/query) is mapped with his/her previous information seeking and understanding pattern, and accordingly library resources are suggested.

Future independent research in this direction should stress on how conversational AI should handle negation, multilingual entity extraction or Out-Of-Vocabulary (OOV) words, which are indeed amongst the toughest problems in NLU. It will also be interesting to see as a future work, the integration of library chatbots with semantic knowledge management systems, which remains the ideal information service framework compliant with end-to-end semantic integration, coupled with AI technologies. The idea would be to develop a practically feasible semantic knowledge management system in the context of library and information services, which automates domain-specific, context-focused, language and diversity aware processing and resolution of information needs. A library chatbot developed on the technical premises of open source conversational AI, integrated with semantic technologies like knowledge graphs and semantic reasoning-inferencing engines in the back end remains a viable option for such a framework.

REFERENCES

1. Turing, A.M. Computing machinery and intelligence. *Mind,* 1950, **49**, 433-460. https://www.csee.umbc.edu/courses/471/papers/turing.pdf (accessed on 19 March 2020).
2. Bocklisch, T.; Faulkner, J.; Pawlowski, N. & Nichol, A. Rasa: Open source language understanding and dialogue management. *In* NIPS Conversational AI workshop, CA, USA, 2017. https://arxiv.org/pdf/1712.05181.pdf (accessed on 19 March 2020).
3. Carroll, J. & Long, D. Theory of finite automata with an introduction to formal languages. Prentice Hall, Englewood Cliffs, 1989.
4. Vincze, J. Virtual reference librarians (Chatbots). *Libr.Hi. Tech. News,* 2017, **34**(4), 5-8. doi: 10.1108/LHTN-03-2017-0016.
5. Allison, D.A. Chatbots in the library: is it time? Faculty Publications, UNL Libraries, 280, 2011. doi: 10.1108/07378831211213238.
6. Rubin, V.L.; Chen, Y. & Thorimbert, L.M. Artificially intelligent conversational agents in libraries. *Libr. Hi. Tech.*, 2010, **28**(4), 496-522. doi: 10.1108/07378831011096196.
7. Dohrmann, E.; Hügi, J.; Scheurer, N.; Trummer, A.; Schneider, R., & Eggenberg, C. Bridging the virtual and the physical space: Kornelia–a chatbot for public libraries. *In* BOBCATSSS, Italy, 2010.
8. Susan Mckie, I.A. & Narayan, B. Enhancing the academic library experience with chatbots: An exploration of research and implications for practice. *J. Aust. Libr. Inf. Assoc.*, 2019, **68**(3), 268-277. doi:10.1080/24750158.2019.1611694.
9. Talley, N.B. Imagining the use of intelligent agents and artificial intelligence in academic law libraries. *Law Libr. J.,* 2016, **108**(3), 383–401.
10. Kane, D. A. The role of chatbots in teaching and learning. In E-learning and the academic library : Essays on innovative initiatives, McFarland, UC Irvine: UCI Libraries' Chatbot Files (ANTswers), 2016. doi: 10.7280/D1PO75.
11. Lafferty, J.; McCallum, A. & Pereira, F.C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers (accessed on 19 March 2020).
12. Vladimir, V.; Drissner-Schmid, A. & Nichol, A. Few-shot generalisation across dialogue tasks. *arXiv preprint arXiv:1811.11707*, 2018. https://arxiv.org/abs/1811.11707 (accessed on 19 March 2020).
13. Hui, D.S.; I, Azhar E; Madani, T.A.; Ntoumi, F; Kock, R; Dar, O., *et al.* The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health – the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis*., 2020, **91**, 264–66. doi: 10.1016/j.ijid.2020.01.009.

CONTRIBUTOR

Mr Mayukh Bagchi is currently a research student of Computer Science at the Department of Information Engineering and Computer Science (DISI), University of Trento, Italy. He holds an MS degree in Library and Information Science from the Documentation Research and Training Centre (DRTC), Indian Statistical Institute, Bangalore. His research interests include knowledge management, knowledge organisation, knowledge representation and data science.

He is responsible for end-to-end conceptualisation, design and authoring of the present study.