Towards leveraging conversational agents for instructors and learners to find and access learning resources

1st Hendrik Steinbeck
Hasso Plattner Institute
Potsdam, Germany
hendrik.steinbeck@hpi.de

2nd Theresa Elfriede Isa Zobel
Hasso Plattner Institute
Potsdam, Germany
theresa.zobel@hpi.de

3rd Dr. Christoph Meinel
Hasso Plattner Institute
Potsdam, Germany
christoph.meinel@hpi.de

***Abstract*—Learning material and resources are a necessary input for educational courses and the process of academic knowledge-gain. In this paper, we outline a proposal for a cloud-based tool to find and access learning resources via a conversational agent. As the literature shows, chatbots are a suitable tool for content discovery and provide a low-entry utility, as well as showing previous implementation projects in the past. As the evolution of the web progressed, more media-types, especially videos for learning emerged. The primary goal of our proposed architecture is combining one search query from the user to trigger multiple crawlers and discoverability 'agents' to shorten search time and support the learning experience. This will be especially useful for users with less technical knowledge and in time-constraint searching scenarios. As a work-in-progress, the proposed implementation will be part of a undergraduate capstone project following a design-thinking engineering process.**

***Index Terms*—Discoverability, learning material recommendation, chatbot for learning**

I. Introduction

Famous quotes and narratives regarding education revolve around the availability of material and its accessibility ("learning as a resource that multiples, if used"). As intuitive as these statements are agreed upon, they not only neglect the necessary steps to find and use learning material, but even more, they take the existence of learning resources for granted. Different universities and lectures use different materials and course-work. From here, two scenarios are possible: A student searches for additional material online, most likely, because the available material is not sufficient. Consequently, the second scenario is required: A professor or lecturer from another institute uploaded their material publicly. Whether this resource is a pdf, a website or a video does not matter at this point. From here, the general problem from a student and lecturer point of view is the discoverability of content and the adjacent overview of available learning resources. In a first step, a service is needed to ensure content is exposed and 'findable'. In a second step, a consuming application is necessary, aggregating and providing the found material. Such a application could come in various shapes: A dashboard, a smartphone-app, a digital assistant or other forms. In this study, we focus on a text-based chatbot application, because we argue that this kind of approach needs a low computational thinking skill: A lecturer or learner with the desire to learn something about a certain topic can be guided through a chatbot interface, drill-down relevant search results and leverage several (heterogeneous) sources, without adding complexity of different search queries and various browser tabs.

In this context, we focus in the given paper on the outline of a work-in-progress architecture of a discovery system for learning resources by leveraging chatbots (conversational agents - CA). The paper will outline the current state of chatbots that are used in the realm of digital education, shows potential angles for a discovery system and the current draft of the technical architecture. The overall research questions of this technical implementation is therefore:

RQ1: How can conversational agents help finding and accessing learning resources?
RQ2: How can an architecture incorporate a growing and heterogeneous list of sources and platforms?

The first research question will be covered by an overview of the existing theoretical and practical concepts that were applied so far, while the second question delivers our main contribution to the field by outlining the proposed architecture: Closing content gaps and discoverability issues for students and lecturers. With our concept and intended architecture, we close the gap between existing educational content that is "ready" to be used by learners and provide a less-technical audience with the access of said material. The following sections will narrow down the problem at hand, the existing research on the topic and our solution approach.

II. Foundation

As for now, we will focus on videos, since their usage in Massive Open Online Courses (MOOCs) is somewhat dominant on various platforms. These platforms have the

2021 World Engineering Education Forum/Global Engineering Deans Council (WEEF/GEDC) | 978-1-6654-2488-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/WEEF/GEDC53299.2021.9657307

downside of a silo-effect: One platform hosts videos and provides a search function, but is limited to its own content. Our CA targets these individual silos and aggregates video specific content for the user. Here we see three advantages in contrast to traditional search engines: First, less-technical skilled users still use search engines manually, potentially stopping their search after (only) looking on YouTube. A one-click query search across multiple platforms would enable these users. Second, comparing material manually is time-consuming for lecturers and tutors. We argue that finding appropriate content and evaluating it increases the cognitive load. Simplifying the search could ease this effect. Third, search-engines have a much broader result queue (e.g., media-type-wise - mixing pdf, videos and html sites) and work on a "winner takes it all"-mentality: Smaller 'silos' or domain-specific repositories usually do not show up on the first page of search engines[1].

In more vivid terms, the following scenarios would benefit from our CA-based discoverability tool:

- Jamie, a 22 year old undergraduate student in the field of political science needs an additional explanation regarding the median voter theorem. He stopped searching after finding no detailed explanation on YouTube.
- Dr. Mendoza, a post-doc in the field of digital-health, needs an animated video of the steps of a weakly supervised predictive algorithm for brain tumors. He hoped to find some open-source material instead of creating something "quick and dirty" on his own.
- As a tutor, Jill wants to provide some extra-credit assignment for the Huffman tree without designing an exercise from zero.

A. Problems

From there, more problems arise in the day-to-day tasks of students and lecturers: Some learning material is outdated, tackle a subject only on a shallow basis or is in the wrong language. In Germany, multiple projects have tried to develop specific search engines and platforms for open educational resources (OER) with the aim to increase discoverability and distribution of OER. We argue that building the n-th platform is not sufficient, because the mechanisms of crawler and bot-agents could deliver more results than yet another 'silo solution'. As for the silos and origin sources of the material, our CA and crawler is as dependent to access public repositories as the human user is. Internal or login-based systems might be off-limits. As a result for the implementation, the search results could be divided between "deep" and shallow results. Most MOOC providers and platforms have a public course description that can be used for analysis and distinction.

An underlying issue is the competency (some call this the digital literacy) to find and use discovery tools. For instance, searching via google, more advanced operators like filetype: pdf, site:example.org or "logic gates" video.mp4 could be leveraged to in-/excluded and drill down results to specific file types. While this is a common approach for people with technical knowledge, the average user will unlikely start a 'structured query search' for a learning assignment at hand. With a user who has low digital literacy skills, the actual learning constraint is rather a meta-problem. A low-barrier solution via CA could shift this, focusing on the actual learning process instead of a gathering part. As described above, instructors and professors would benefit from such a tool as well. Although they are more likely to have these skills due to their scientific education, time and lecture preparation is a big concern. Various topics - including our '*logic gates*' - are taught around the world in different settings and digital courses. The concept of 'curated content' can be applied as well, meaning that a lecturer is recommending (video) content to students. As a win-win situation, the learners can use relevant content that has been "vetted" and approved, while lecturers save time and resources due the usage of existing material.

A technical challenge are the heterogeneous sources, meta-data and queries. To search for 'logic gates' YouTube would use the variable '*search_query*', while cousera calls that variable only '*query*'. Our own MOOC platform (link anonymized) uses an even shorter variable ('*q=logic gates*'). One step further, the XML-tree of the result page needs to be analyzed as well, because we do not simply want to redirect to the YouTube search result, but aggregate various of these heterogeneous sources. Once this source-specific behavior is implemented, the search queries have the advantage of being language agnostic: If could use person '*logic gates*' or its German equivalence.
In the next section the potential application areas of chatbots are introduced.

B. Chatbots

Chatbots are conversational systems that interact with a user in natural language using AI methods [4]. Reference [9] has outlined three layers (perspective, dimensions, characteristics) and delivered a taxonomy of chatbots - including the usage for e-learning as a utility-based system. Interactions can be conducted for various reasons, including answering questions, performing tasks or accessing data. In the process, chatbots can access and process external information via, for example, APIs. Reference [3] classifies chatbots into two categories: Task-orientated and non-task-orientated dialogue systems. Task-oriented chatbots like Alexa or Siri execute a specific task for a user, for example, switching off lights or getting

[1]For instance, searching "logic gates" on Google.com and filtering via "video results" yields 9 YouTube videos and one blog article with an embedded video.

directions to a destination. Non-task-orientated chatbots have the purpose of giving advice, answering questions as well as entertaining the users. Reference [11] provides four additional categories:

- Goal-based chatbots that aim to achieve a primary goal (task) with only a little information from the user.
- Knowledge-based chatbots can answer questions based on their specific knowledge bases.
- Service-based chatbots are categorized by their specific facility. For example a customer can access and request changes for a hotel or flight booking.
- Response Generated-based chatbots use a dialogue-manager to select a response in natural language based on the input. Typically these types of chatbot are used for answering frequently asked questions.

Chatbots have many advantages for a user as well as a platform operator. A user can talk or speak to the chatbot at any time in natural language and receive answers. For a platform operator, the provision of a chatbot is quite simple and inexpensive thanks to various frameworks. In addition, money can be saved on human customer support. Advances in AI methods and thus the capabilities of the chatbot allow for more complex or emotional conversations.

Chatbots are popular and applicable in many different areas such customer support, emotional and social support, education (as presented in the state-of-art analysis [18]), as well as information and requests [1]. Famous examples are Eliza [17], the first chatbot developed in 1966 by Joseph Weizenbaum and Cleverbot [15], a chatbot that teaches undergraduates the Turing test. The increasing popularity and high number of chatbots is also related to the various platforms on which even non-technical users can build a chatbot. These include chatbots on Telegram [10], Facebook Messenger [8], [15] and WhatsApp.

III. Related Research

This section presents and discusses existing research and projects in the area of access to and discovery of educational resources specifically for teachers and learners. Additionally, dialogue-based systems in educational context and for obtaining digital content are introduced.

Reference [12] presents a case study conducted with non-high technology mathematics and science teachers to find out how digital educational resources for instructional purposes are discovered and accessed. The study showed that teachers used different search strategies depending on the age group and considered them to be correct and current. The resources found were mostly used as teaching material without further processing. Finally, the authors discuss a digital library system for digital educational resources, which will provide teachers with tools to find, process and reuse resources.

Reference [13] is a versatile tool based on Linked Data Design Issues that helps users find and use digital educational resources from OpenCourseWare repositories. Linked Data aims to help connect and provide greater access to individual silos of educational resources. The interface also offers the possibility to display the resources visually and interactively.

As already mentioned, chatbots can be used in many different areas, including education. Jill Watson [6] describes a chatbot application in an online educational context. The chatbot acts as a virtual teaching assistant to support learners with answers to frequently asked questions, regularly update the students via a forum and improves the interaction between learners and teachers. TruthBot [7] helps users to find news and information on a specific topic, as well as to evaluate and check them for accuracy. Another research project introduces a chatbot for students to get timely help with questions and suggestions for educational resources. This will also reduce the workload of teachers who often have limited availability and do not have or know about every educational resource [16]. A chatbot that makes recommendations for scientific resources is OnTimeRecommend [14], which suggests relevant publications and domain-specific topics to the scientific community through a RESTful web service.

IV. Solution Approach

A. Concept

In the following section an idea is presented which supports non-technical users, especially teachers and learners by discovering and accessing digital educational resources (videos) with the help of a non-task-orientated knowledge-based chatbot. The user can ask the dialogue-based system questions or request recommendations for a topic. This ensures simple and fast communication in a dialogue-based form as the chatbot is always available and provides suggestions and answers in no time. Additionally, the users might not really know that they are speaking with a chatbot instead of a real person. Consequently, the support and recommendations seem more personalized and the user feels more satisfied and pleased with the results. Furthermore, the communication with the chatbot can be in written or spoken form.

Various scenarios are covered with the system. It should be possible for users to obtain different learning materials, especially videos, from different platforms with simple search phrases. This means the user provides the chatbot with a topic or several keywords on a topic and receives a list of useful materials. This use case is presented as an example in Figure (1) where a teacher asks for resources about "logic gates". As a result, the chatbot queries various websites in its internal logic via crawlers or APIs and gives the user the three resources that best match the keyword. If the users are not satisfied with the results, they can ask the chatbot for more resources. The chatbot will then provide more results but also ask the users why they were not satisfied with the first answers. This feedback is evaluated manually and is important for the continuous development of the chatbot to provide more appropriate results in the future. With the help of this system, the personas presented in Section II can obtain simplified access to otherwise unknown or difficult-to-access material, so that knowledge exchange and knowledge discovery are specifically promoted. Giving practical examples, our student can leverage more platforms, finding a specific blog-page about the median voter theorem, while the lecturer, Dr. Mendoza used the CC-BY-4.0-tag to get appropriate OER material regarding the visualization of the algorithm.

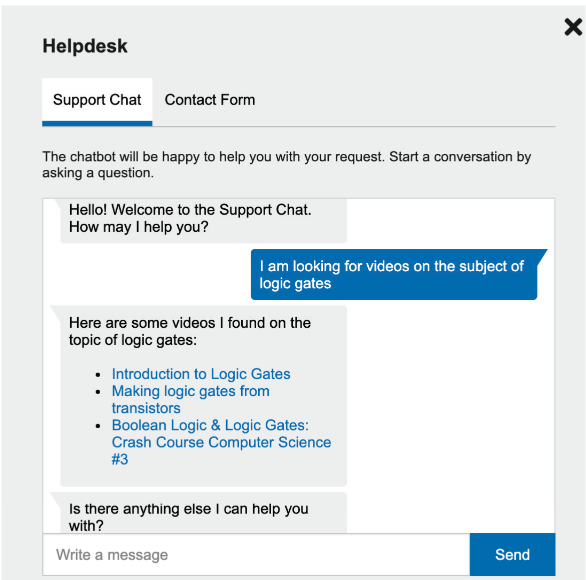

Fig. 1. Example conversation: user asks the chatbot for video material to the topic ”logic gates”

In addition, content gaps can be identified via a query to the chatbot. If the chatbot finds few or no resources on a topic or keyword, this is a potential for lecture or content creators. They can then produce targeted material, especially videos or even MOOCs, on these topics and thus close knowledge gaps and consequently earn money. Another use case is the creation of analyses. For example, operators of MOOC platforms can use this system to show topics with high and low popularity and relevance. These analyses can be used to promote MOOCs in a targeted manner and again to identify gaps in content and topics. This procedure was preceded by a framework comparison in a graduate thesis, resulting in the decision to use opens-source branch of the Rasa-Framework[2]. In this current stage, the research team focus on the phases *ideation*, *prototyping* and testing. As our target groups are instructors and learners alike, both user bases will be incorporated in tests and feedback cycles. From there, personas can be derived and consulted once an MVP exists. In a final evaluation phase, a second publication is planned to report effectiveness and actual usage of the system. This is especially true for the outlined MOOC usage, due to the in-house platform openhpi[3]. Proven and stable increments of the CA can also be tested in a broader online userbase.

B. Architecture

The architecture of the system consists of two major building blocks as illustrated in Figure (2). On the one hand, the chatbot framework has a corresponding interface through which the user can communicate with the chatbot. On the other hand, crawlers are needed, which are addressed by the chatbot and then search various websites and platforms for certain keywords. However, it is also possible for the chatbot to address some platforms directly via an API, if one is already available. Part of the search and result options should be media-type, length, upload date and potential copyright tags (or OER tags). The platforms or websites requested include among others Vimeo, YouTube or MOOC platforms such as edX, VHB, imoox or openhpi. The list of sources is not static and can be extended by other relevant platforms. Rasa will be used as the chatbot framework. This is an open-source framework that has the advantage of being hosted on premise. Therefore, concerns about security and privacy may be low or even non-existent.

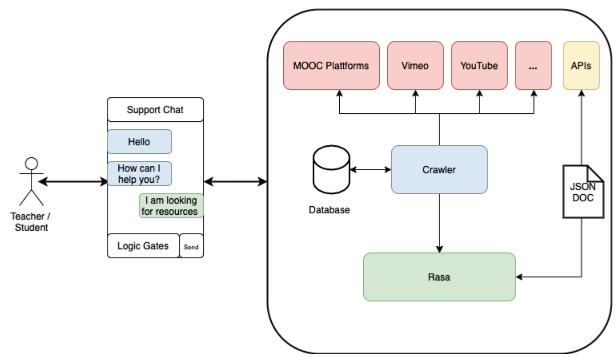

Fig. 2. Architecture consisting of chatbot framework Rasa with the user interface, the APIs and crawlers

V. Conclusion

The given paper discussed the necessity to discover and access heterogeneous sources in order to obtain learning resources and videos in particular. With the proposed implementation draft, we want to bridge the digital divide between the already enabled and tech-savvy users and the less technical skilled students and lecturers. As outlined, this would benefit various scenarios and personas, as well as providing a benefit for a platform, due to better discoverability and content exposure. From a technical standpoint, the research questions can be answered by an architecture that leverages platform specific crawlers to find fitting (video) content and a service within that architecture that normalizes and processes the results. From there, a chatbot can be the consuming application to present the results to a user. In order to call this a successful

[2]https://rasa.com/open-source/

[3]https://openhpi.com

implementation, which will be part of an undergraduate capstone project, the actual feedback by users is relevant. As a result, a design-thinking implementation process is scheduled, focusing on the described users (students, tutors, lecturers) and their needs. Here, a primary limitation unfolds, as we do not evaluate the quality of the material, but merely focus on a qualitative presentation of available learning resources. As previous projects (see Section III) had implemented similar search-engines of specific projects, some combined it with a rating system based on user-input (e.g. merlot.org, [2]). This feature is not prioritized in our current state, as chatbots leverage feedback mechanisms directly, rather than linking it to a specific result or item. With a working prototype, future research and analysis need to be conducted regarding the efficacy of the chatbot's recommendations. Potential metrics could be a perceived usefulness, a shorter search time or an increased learning effect. Another research step is incorporating the above described user feedback as a traditional loop-back function in chatbots, or the consideration to include user recommendation regarding a topic.

References

[1] P. Brandtzaeg and A. Følstad, "Why People Use Chatbots," Nov. 2017. doi: 10.1007/978-3-319-70284-1_30.

[2] R. Cafolla, "Project MERLOT: Bringing Peer Review to Web-Based Educational Resources," Journal of Technology and Teacher Education, vol. 14, no. 2, pp. 313–323, Apr. 2006.

[3] H. Chen, X. Liu, D. Yin, and J. Tang, "A Survey on Dialogue Systems: Recent Advances and New Frontiers," SIGKDD Explor. Newsl., vol. 19, no. 2, pp. 25–35, Nov. 2017, doi: 10.1145/3166054.3166058.

[4] A. Følstad and P. B. Brandtzæg, "Chatbots and the new world of HCI," interactions, vol. 24, no. 4, pp. 38–42, Jun. 2017, doi: 10.1145/3085558.

[5] R. W. Gehl, "Teaching to the Turing Test with Cleverbot," Transformations: The Journal of Inclusive Scholarship and Pedagogy, vol. 24, no. 1–2, pp. 56–66, 2014.

[6] A. K. Goel and L. Polepeddi, "Jill Watson: A Virtual Teaching Assistant for Online Education," p. 21.

[7] A. Gupta et al., "TruthBot: An Automated Conversational Tool for Intent Learning, Curated Information Presenting, and Fake News Alerting," arXiv:2102.00509 , Jan. 2021, Accessed: Jul. 10, 2021. . Available: http://arxiv.org/abs/2102.00509

[8] C. Holotescu, "MOOCBuddy: a chatbot for personalized learning with MOOCs," p. 4, 2016.

[9] A. Janssen, J. Passlick, D. Rodríguez Cardona, and M. H. Breitner, "Virtual Assistance in Any Context: A Taxonomy of Design Elements for Domain-Specific Chatbots," Bus Inf Syst Eng, vol. 62, no. 3, pp. 211–225, Jun. 2020, doi: 10.1007/s12599-020-00644-1.

[10] M. Muslih et al., "Developing Smart Workspace Based IOT with Artificial Intelligence Using Telegram Chatbot," in 2018 International Conference on Computing, Engineering, and Design (ICCED), Sep. 2018, pp. 230–234. doi: 10.1109/ICCED.2018.00052.

[11] M. Nuruzzaman and O. K. Hussain, "A Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks," in 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), Xi'an, Oct. 2018, pp. 54–61. doi: 10.1109/ICEBE.2018.00019.

[12] M. M. Recker, J. Dorward, and L. M. Nelson, "Discovery and Use of Online Learning Resources: Case Study Findings," Journal of Educational Technology & Society, vol. 7, no. 2, pp. 93–104, 2004.

[13] Piedra, N., López, J., Chicaiza, J., & Caro, E. T. (2014, April). Serendipity a platform to discover and visualize Open OER Data from OpenCourseWare repositories. In Proc. Open Educ. Multicultural World (OCWC).

[14] S. S. Sivarathri et al., "Chatbot Guided Domain-science Knowledge Discovery in a Science Gateway Application," p. 4.

[15] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," Computers & Education, vol. 151, p. 103862, Jul. 2020, doi: 10.1016/j.compedu.2020.103862.

[16] K. Souali, O. Rahmaoui, M. Ouzzif, and I. El Haddioui, "Recommending Moodle Resources Using Chatbots," in 2019 15th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Nov. 2019, pp. 677–680. doi: 10.1109/SITIS.2019.00110.

[17] J. Weizenbaum, "ELIZA" a computer program for the study of natural language communication between man and machine," Commun. ACM, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.

[18] R. Winkler and M. Soellner, "Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis," Proceedings, vol. 2018, no. 1, p. 15903, Aug. 2018, doi: 10.5465/AMBPP.2018.15903abstract.