Chatbot for intelligent Internet content search support

Susana Pereira
Instituto Superior Técnico
Lisboa, Portugal
susanampereira@tecnico.ulisboa.pt

ABSTRACT

Over the last decades, the Internet has become the leading tool for communication and dissemination of information worldwide. The increase in the number of users has led to a colossal increase in the amount of information shared on the web, strongly driven by the popularization of social networks. Information has thus become more immediate, volatile, and increasingly difficult to track and validate.

In a constantly changing world, organizations must be able to track and predict future trends to be able to remain relevant and ensure their survival. In recent years there has also been a growing interest in chatbots, motivated by significant advances in the areas of Artificial Intelligence (AI) and Machine Learning.

The aim of this study is to design a chatbot that allows intelligent support for Internet content search. The use of AI solutions allows these conversational agents to simulate human conversations using Natural Language Processing, thus transforming them in-to valuable virtual assistants capable of performing a variety of tasks in an automated way.

KEYWORDS

Chatbots, Web Search, Natural Language Processing, Information Retrieval, Web Content Management

1 Introduction

In recent decades, the Internet has become one of the most important tools for communication and dissemination of information in the world. The number of Internet users, which in the first quarter of 2021 increased to 65.6% of the world's population [1], has led to a colossal growth in the amount of information shared on the web, strongly driven by the popularization of social networks [2].

The Internet has radically changed the way we act before ourselves and the world. Network communication enabled the creation of a global society in which the sharing of information and knowledge reached infinite proportions. In fact, the Internet is all we need to gain access to a whole world of knowledge that allows us to overcome almost every barrier in our daily life. As a tool for information research and knowledge transfer, it is becoming increasingly indispensable in modern life [3].

[4] identified the four main properties that characterize the web as an information repository: it is huge, dynamic, self-organized and hyperlinked. However, it is these same characteristics that pose some of the main challenges regarding information research.

All websites are composed of multiple web content, a term that can refer to both the type of information and the individual item used to convey this information [5]. There are several types of web content, the most common being presented in the form of text, image, audio, or video, which are managed through content management systems (CMS).

Nowadays, there is a huge amount and diversity of data available on the web, making it increasingly difficult to do research on specific subjects or find useful information on the Internet [6]. Content search is one of the most common activities performed on the web and search engines generally do a good job in searching for concrete keywords. However, they are often dependent on rankings and commercial support and may present biased responses. In addition, searching for terms or expressions without indication of context is still far from effective.

In a constantly changing world, organizations must be able to track and predict future trends to be able to stay relevant and ensure their survival [7]. To make business and decision-making processes faster and more efficient, it is necessary to develop advanced search solutions that allow filtering the information available on the web while contributing to the improvement of processes associated with content management.

However, despite many advances in the development of search engines for the web, especially on semantic web [8], it is not always easy to perform searches efficiently and users still experience difficulties in the process of searching and retrieving information [9], mainly due to the challenges arising from big data [10].

There are several factors that contribute to this difficulty, of which the following stand out:

1. Users may not be able to formulate their complex information needs in a single query [11].
2. Users may not be familiar with the domain where the search is based [12], the keywords needed to perform a query or the ways to reach it [13].
3. The ambiguity inherent in the understanding of the human language and the absence of context, especially in shorter amounts of text such as those used to perform queries[14].
4. Challenges related to content recommendation systems [15, 16] and search engine auto-complete features [17].
5. The existence of biased or misrepresented results that hinder the ability to assess their credibility [18].

On the other hand, there is an increasing interest in the use of AI tools such as chatbots to improve productivity and automate tasks. These systems are already being used at the enterprise level to help with tasks regarding customer support, sales, finance, marketing, human resources, amongst others.

Even though the quality of interactions between users and conversational agents is still a problem in performing more complex tasks, there is a clear interest in the use of chatbots, and the evaluation of these systems is an active research area in the AI community [19]. These intelligent conversational question-answer systems can help clarify an user's information needs by proactively asking them to clarify their questions [11] and thereby return more accurate results. Therefore, it would be very convenient to have low-cost and user-friendly tools that would assist on the task of searching for content on the Internet in situations where users do not master the lexicon and the context where the search fits.

1.1 Research problem

The development of a chatbot to support intelligent web search that could be easily implemented across multiple types of platforms can complement existing search methods and contribute to new research in this area. Thus, with this work it is intended to answer the following research question: **Is it feasible and possible to architect and implement a chatbot to support the search of information in the space contained in a given website?**

1.2 Goals

To obtain a response to the research problem identified, six goals have been set:

Goal 1: Investigate web content management problems.

Goal 2: Thoroughly investigate chatbot technologies and the different frameworks available.

Goal 3: Contextualize the research work with the concepts of natural language processing (NLP), deep learning (DL), machine learning (ML), information retrieval (IR) and other related concepts.

Goal 4: Design the proposed system architecture.

Goal 5: Design and train a conversation agent.

Goal 6: Test the prototype and evaluate the results.

2 Methodology

Design Science Research (DSR), or design research, is a paradigm widely used in information systems that consists of the development of innovative artifacts to solve specific problems within a given domain and thus contribute to the production of new scientific knowledge. The artifact is the central element of the DSR and should be well tested, understood and documented to ensure its pragmatic validity. The evaluation is a central phase of the process, in which the usefulness, quality and effectiveness of the artifact are demonstrated, using well-executed evaluation methods [20].

[21] present a mental model that summarizes the main steps to carry out research and evaluation through the DSR methodology (DSRM). This is a model commonly accepted in information system research [22] and includes the following six sequential activities: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication The iterative process inherent to the DSR allows the improvement of the system by identifying risks and uncertainties in each iteration, which results in fewer errors and leads to the creation of more efficient solutions.

The Figure 1 presents the DSR methodology applied to this project.

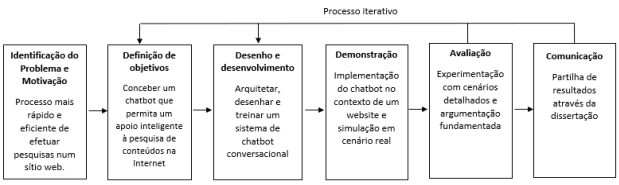

Figure 1 – Design Science Research Methodology (Adapted from [21])

3. Literature review

In this chapter, the literature is reviewed with the objective of presenting a theoretical framework and perceiving the state of the art of the topics that will be addressed during this research study.

3.1 Content Management

In [23] content is defined as information produced through the editorial process and ultimately destined to human consumption through publication. This definition points to the basic dichotomy of content management: the difference between management and distribution.

All websites are composed of multiple web content, a term that may refer to both the type of information and the individual item used to transmit that information [5]. Users visit the web pages with one or more goals in mind and therefore content plays a very important role in building a website [24]. There are several types of web content divided essentially into two categories: text content and multimedia content.

CMS enable the creation, management, distribution, and publication of content, including corporate content. They ensure that the right information goes to the right user, organize, and give access to all types of digital content, contain information about files and link to other files, facilitating their location and access, and can encompass all content created within an organization [25]

3.1.1 Information Retrieval

Information Retrieval is the process of representing, storing, organizing, and accessing information [26].

Within the context of information retrieval systems (IRS), a distinction should be made between data recovery and IR, the first being used to determine the documents containing the keywords searched.

An SRI typically searches for structured or semi-structured data collections, such as web pages documents, images, and videos, and is typically used when a collection reaches a size that is no longer supported by traditional techniques [27].

3.1.2 Web content search

The two most frequently used methods to find information available on the Internet are browsing and searching. Browsing consists of the process of following a path of hyperlinks, which may or may not be organized in directories, until the desired information is reached. On the other hand, web search is performed through a system that tries to find matches between the user's words and the documents available in web.

A search engine is a software system designed to find information stored in computer systems from user-indicated keywords, reducing the search time required to find the information. They appeared soon after the beginning of the Internet and allow a user to search for any type of information on the web, presenting results in a more organized, fast, and efficient way.

3.1.3 Current challenges in web search

The digital age and exponential growth of the Internet in recent years have led companies and organizations to store more and more information on web, often in an unorganized way. The huge volume of data makes it difficult to find specific issues, making it clear that new ways of accessing information are evident [28, 29].

Despite being increasingly integrated into operating systems, software applications and other interactive environments that extend beyond the traditional web browser [30], web search is an extremely demanding process [31]. When searching a website, users are interested in quickly getting the right answers to their queries, so they don't waste their time browsing multiple pages through traditional navigation menus or searching for keywords in optimized search engines to return the best results based on marketing strategies.

The growing popularity of smart devices, personal assistants, and customer relationship management (CRM) systems has encouraged the scientific community to develop new methodologies for virtual assistants, including chatbots [31].

3.2 Artificial intelligence

Artificial intelligence is one of the most recent areas in science and engineering and has played a very important role in the era of digital transformation.

AI incorporates various tools, techniques, and algorithms, including ML, DL, neural networks, and NLP that, along with increased data, constant connectivity, and high-performance computing, can add new levels of efficiency and sophistication to future technologies [32].

3.2.1 Machine Learning

Machine learning is a multidisciplinary study area whose main objective is to make predictions through computers [33].

There are two broad categories that vary according to the level of human intervention in data preparation and labeling: supervised learning and unsupervised learning. There are also reinforcement techniques, which learn and identify patterns with the aim of reacting to an environment [34].

3.2.2 Deep Learning

Deep learning is a branch of machine learning that is based on the basic principle that there are several layers of processing with complex structures and where each step receives as input the result of the previous step, thus creating an idea of depth [35].

DL techniques have made it possible to achieve important advances in problem solving that AI has not yet been able to address, particularly regarding natural language understanding

(NLU), tasks such as topic classification, sentiment analysis, answer to questions and machine translation. These techniques can be applied in many fields of science, as well as at the enterprise and governmental level [36].

3.3 Intelligent personal assistant tools

Intelligent personal assistant tools are one of the main applications of artificial intelligence. They are software agents capable of performing tasks or services based on commands or questions, which can be provided by text or voice.

By using AI techniques such as NLP, voice recognition, machine translation or knowledge representation [37], digital assistants typically have a high level of interactivity and intelligence, which allows them to increase their performance while executing tasks [38].

3.3.1 – Chatbots

A chatbot is a system designed to simulate conversations with humans, especially over the Internet, capable of performing a wide variety of tasks with varying levels of complexity. They are a typical example of AI systems and promise to revolutionize the field of human-computer interaction.

Conversation agents have evolved from simple, pattern-aware programs to very complex systems that use AI technologies. They use a combination of NLP, NLU, natural language generation (NLG), and ML techniques, allowing them to be more flexible in maintaining a conversation [29] and substantially simplify man--computer interaction, thanks to their natural language paradigm [39]. These features contribute to increased user satisfaction, as they help to find information in a more comfortable way than other less sophisticated and more time-consuming interfaces [28].

The number of chatbot-related technologies is already quite significant and is growing every day.

3.3.2 – Natural Language Processing

Natural Language Processing is a transdisciplinary study area that brings together computer science, linguistics, statistics, and artificial intelligence and aims to process and analyze natural languages automatically [40]

Understanding and using human languages is used to develop techniques that allow computers to recognize and manipulate natural expressions to perform certain tasks. Most techniques are based on ML. NLU is a fundamental part of PLN tasks and can be used, for example, to implement user interfaces such as chatbots.

3.4 Differences between search engines and chatbots for web search

In recent years we have seen a great growth in the research and use of chatbots with the aim of reducing the cognitive burden needed to obtain basic information from a website [41, 42]. IR using a chatbot is especially complex due to its interaction with the natural language, and especially when the domain is restricted [43]. In addition, in information search scenarios, these systems still have very limited conversational capabilities [44] and serve primarily as proxies of existing web search engines [45].

Conventional search sites (or search engines) use a one-sided method that lists the results for an incoming search word. On the other hand, today's chatbots aim to find and provide the services and data that people need at the right time in conversation form, without running separate apps or websites [46].

Today's web search engines are not conversational systems because they do not respond with natural language, even if the user enters queries using natural language [47]. Unlike search engines, IRS and recommendation systems, search-designed chat systems are characterized by a sequence of interactions between the user and the system to meet the user's information needs, performing tasks such as searching for files, scheduling reminders, surfing the net or even chit chatting [48].

Another important difference is that in the search tool there is a link between any word and the document where it is located, while in the conversation system, a link is usually provided only to the most significant words [49].

3.6 Opportunities and challenges in developing chatbots for web information search

Based on the literature, it is concluded that the use of chatbots to support intelligent content search on websites has been successfully studied and implemented by some researchers. There are, however, a few opportunities and challenges to be considered in the development of these tools, as it can be seen on the table 1.

Table 1: Opportunities and challenges

Opportunities	Challenges
There has been an increase in interest in the use of chatbots equipped with Artificial Intelligence to perform actions of content search and information retrieval on the web.	Users like to use chatbots to get quick and accurate answers. However, they consider them limited and become frustrated when the system cannot immediately understand them [50].
The results of the study by [51] reveal that users prefer chatbots that provide a "human-like" natural language conversational capacity.	The lack of effectiveness in the use of chatbots may be related to the lack of standards for chatbot evaluation [52].

The motivational factor for the use of chatbots most frequently reported is "productivity" [53].	It would be interesting to compare the ease of use of the chatbot with a conventional search engine [31].
Personalized recommendations are becoming increasingly popular to allow people to get products and services [54]	It is necessary to rethink the user interfaces associated with chatbots [41].

4. Proposed solution

This section consists of the third phase of the DSR methodology and includes the detailed explanation of the proposed solution designed to solve the research problem identified.

4.1 User and task analysis

Based on the literature review, we found that using chatbots for web search is seen as an interesting and potentially idea. To better understand the target audience, a questionnaire was conducted to potential users of the proposed chatbot system, with the aim of assessing whether the features initially thought for the prototype would be appropriate, assess the level of knowledge they had about chatbots and ascertain the level of interest and benefits that a chatbot system to support intelligent web search could offer against other tools. This analysis has also identified the tasks that users need to perform and how they do so today, using existing systems.

The questionnaire, elaborated using the Google Forms tool, consists of 25 multiple choice questions and was designed so that it could be answered by any type of user, regardless of their age or training, having the only condition of being a Portuguese language speaker. It was published on several online channels, namely in social networks and in groups of students and enthusiasts in the areas of AI and Data Science. In total, 29 valid answers were counted. Of the total number of respondents, 62.1% identified themselves as female and 37.9% as male, more than half were aged between 25 and 64 years, and the vast majority had educational qualifications at the level of higher education.

Although most users had indicated that they have medium to good knowledge of the use of information computer technologies (ICT), there are still a considerable number of users who have never used chatbot systems, and those who have already used it indicated that they do not uses them regularly (Figure 2). This shows that, on the one hand, these systems are not yet widely disseminated to users less adept at ICT and, on the other hand, that they are not yet seen as an alternative to existing systems.

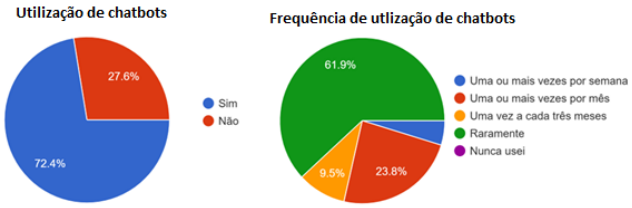

Figure 2 –Chatbots usage and frequency

The main reason users indicated for using a chatbot instead of other options available was the desire to try the chatbot. This confirms that chatbots are still seen primarily as a novelty, but a novelty that users are willing to experience due to other reasons presented, such as ease of use and speed in getting the answer to a question. In addition, 38.1% said they had already had a positive experience with chatbot.

Even though they have never used a chatbot, most users (96.6%) indicated that it would be interesting to use this type of system as a support tool for searching content on the Internet and 88,5% consider that the fact that chatbots only return results based on the context of the conversation is an advantage over traditional search engines. It has also been found that users are willing to share some personal data with a chatbot system, which shows that they would be interested in taking advantage of more advanced features such as conversation and personalized suggestions.

All respondents also showed interest in using chatbot for searching for content on a given site, but most would use it only as a support tool to other forms of research (Figure 3).

Figure 3: Usage of chatbots for content search as an autonomous tool or as a support tool

Regarding other features that users would like to see included, the most popular were: personalized conversations (65.5%), possibility to obtain summaries of content generated by artificial intelligence without having to access the original website (62.1%), ability to easily share content with others (55.2%), sentiment analysis and emotional intelligence (44.8%) and possibility to directly recognize and translate the results presented in other languages (41.4%). This shows that users are interested in adding new features to search that are not yet present in traditional search

engines. On the other hand, 27.6% of respondents indicated that they would to be able to transfer the conversation to a real-time assistant, suggesting that for some users chatbots may not yet be autonomous enough to complete this task.

Despite some of the caveats already identified, most respondents (72.4%) consider that chatbots have the potential to replace traditional search engines and become the main content search tool on the Internet, considering it a technology with a promising future.

4.2 Requirements

Before proceeding with the design of the prototype, it is important to determine the necessary requirements for its development and the functionalities that should be included in the chatbot. This way, a set of functional and non-functional requirements have been defined that should be considered throughout their life cycle (Table 2).

Table 2 – System requirements

Functional requirements	Non-functional requirements
The chatbot must be able to start a conversation and interact with the user	Select a chatbot creation platform that allows deployments across multiple channel types
The *chatbot* should be able to understand natural language and be able to identify the user's intent, regardless of how the user built the prayer	Select a channel that allows you to integrate the chatbot simply and quickly
The chatbot should be able to help users find what they're looking for quickly	The chatbot should be able to be used by as many users as possible
The main purpose of the chatbot should be to return news	Responses to user questions should be quick
The casual conversation feature will be a secondary way to get user preferences	The system should always be available
The replies returned should be consistent with the context	Chatbot *must meet the usability criteria for WCAG 2.0 web content*

4.3 Use cases

This section specifies the use cases thar describe the actions that the user will perform on the system on the Table 3.

Table 3 – Use Cases

Use Case	Description
UC1: Get news	The purpose of this use case is to enable news to be obtained based on the preferences indicated by the user. The system welcomes the user. The user starts a conversation with *the chatbot.* The user indicates that they do not want to register. The system asks for the user's preferences The user indicates their preferences. The system returns a news story related to the user's preferences. find no news that matches the search, returns no news, and asks the user for new data to perform a new search. User preferences are recorded in the database and the user can continue to use the system or end the conversation This use case is intended to register the user.
UC2: User registry	The system welcomes the user, and the user initiates a chatbot conversation. The user indicates that it is not yet registered. The system requests the user's name and email and confirms the registration. The user's data is written to the database and the user can continue to use the system.
UC3: User llogin	This use case is intended to register the user. The system welcomes the user. The user starts a conversation with *the chatbot.* The user indicates that it is already registered. The system asks the user's email to authenticate. The system confirms that the user is registered.
UC4: Chit chat	This use case is intended to allow a casual conversation between the system and the user to obtain more data for the search. The user must be previously registered and make the system welcomes the user. The user starts a conversation with the *chatbot.* The user indicates that it is already registered. The system authenticates the user. If the system cannot find that user's record, the user can try to re-authenticate, re-register, or proceed to the search without registering. The system initiates casual conversation. The user participates in the casual conversation. User preferences are written to the database, and the user can continue to use the system or end the conversation.

4.4 Activity diagram

To better understand the interaction process to be performed, an activity diagram was elaborated, illustrated by the image in Figure 5, which represents the synthesis of the flow of interactions between the user and the chatbot agent.

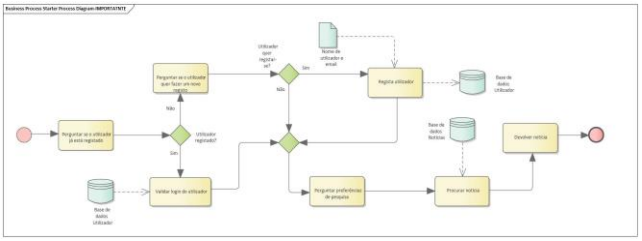

Figure 4 – Chatbot activity diagram

4.5 Architecture

The choice of the type of architecture for a chatbot system depends on several factors, such as the audience it is intended for,

the resources you need to access, and the communication models you want to deploy.

There are many platforms that enable the development of chatbots. After an initial analysis of some of the available options, and considering the requirements previously presented, Google Dialogflow was the environment chosen for developing the prototype.

Dialogflow is a complete NLP platform that includes NLU and NLG modules and a dialogue manager responsible for the flow of the conversation between a user and the chatbot. It also has an inline code editor that permits the configuration a webhook service to perform actions and retrieve information from external databases.

Google Sheets was chosen to create a simple database with knowledge base information that contains chatbot responses and stores user data.

The final architecture, which can be seen in Figure 5, is complete with the User Interface, used to place requests and receive replies from the chatbot.

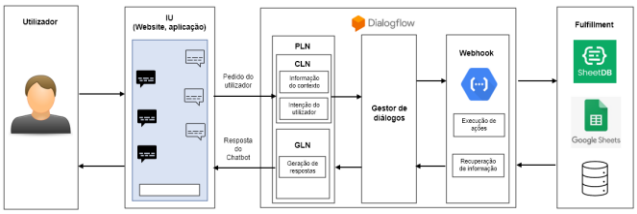

Figure 5 – Architecture of the proposed chatbot system

5 Prototype development

The artifact developed in the context of this work is a chatbot system that aims to help users to search in a space contained within a particular website. In this sense, two daily update news sites were chosen, one on cars and the other on motorcycles, from which six sections were selected to extract the contents to be stored in the system knowledge base.

The proposed system requires external integration to get real-time information. Although Google Sheets has its own API, SheetDB was the chosen approach to convert the spreadsheets into a JSON API. This solution allows RESTful requests and similar functionalities to a relational database.

Dialogflow allows two fulfillment options: via webhook and integrated code editor (inline editor). For this work we used the integrated code editor, which is precisely intended for the development of prototypes. The inline code editor is compatible only with Node.js and uses Dialogflow fulfillment library by default which, although no longer maintained, remains a valid option for less complex projects.

5.2 – Drawing and training of the conversation agent

The first step in the development of a chatbot in Dialogflow is the creation of an agent, an NLP module capable of translating texts introduced by a user during a conversation to structured data, to be understood by applications and services.

There are two intents that are automatically created for each new agent: a welcome intent, activated when a user starts a conversation with a greeting, and a default response, triggered when the agent does not recognize a user expression. For this prototype, 22 new intentions and 5 entities were created.

The intentions establish how to map the expressions entered by the user to a corresponding response, to be returned by the system. This mapping is done with the help of training phrases, which are defined within each intent, and which correspond to examples of possible expressions introduced by the user. With these examples, the agent can evaluate which training phrases are closest to the phrases entered by the user, to return the most appropriate intent.

Intentions and entities are essential to build answers, but to draw a conversation you need an element that links the different intentions. Dialogflow contexts are similar to the context of the natural language and allow you to control the flow of a conversation. In this way, you can configure input and output contexts in each intent, and whenever there is a match in one or more input contexts, all output contexts configured for that intent are enabled.

5.3 Implementation

Dialogflow allows you to integrate agents into various platforms and is fully compatible with Google Assistant, Slack, Facebook Messenger, Telegram, LINE and Dialogflow's own integrations (Web Demo and Dialogflow Messenger).

After some experiences with some of these channels, the integration chosen was the Dialogflow Messenger integration. This option provides a customizable chat dialog for the agent that can be embedded in a website and, like the integration options in instant messaging services, allows you to create dynamic replies using fulfillment. The chat dialog is implemented as a dialog window that can be opened and closed by the user. When opened, the chat dialog appears above its contents in the lower-right corner of the screen.

Once the integration was chosen, a website was created on which the prototype was implemented, and which also includes all the supporting documentation necessary for the use of the chatbot.

The design of the interface considered the simplicity and minimalist aspect. Dialogflow Messenger allows for various types of advanced response, namely in card forma. To make it easier to enter text on smaller screens such as smartphones, it was decided, whenever possible, to use buttons to guide the conversation. The colors were chosen to create a contrast between chatbot speeches

and user instructions. Figure 6 shows some examples of the chatbot implemented on a website.

6 Evaluation

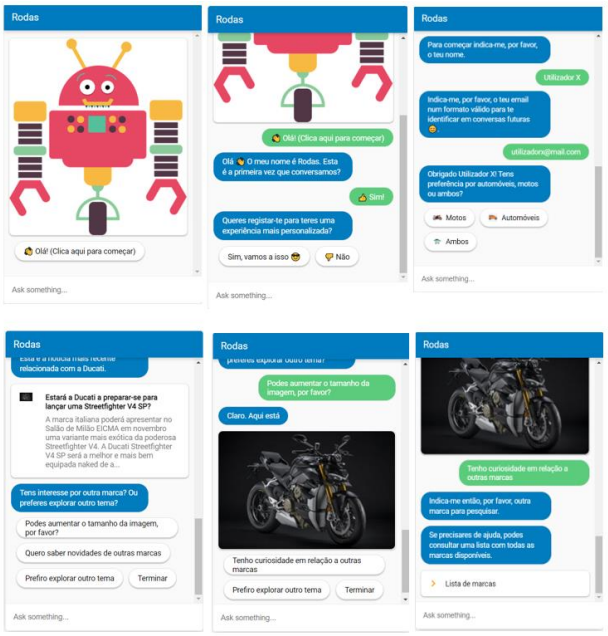

Figure 6 – Views of the chatbot implemented on a website

Although, to date, no standard has been established for the evaluation of chatbots [55], the guarantee of the good development of chatbot, as with any other software, depends on an appropriate evaluation.

Among several possible methods and metrics, in this study we opted for a test session with users, with the objective of assessing their degree of satisfaction in the use of the proposed system. In this way, it will be possible to assess the feasibility of implementing a chatbot system for intelligent content search in the contained space of a given website, through the empirical evaluation of how users perform the proposed tasks.

These tests were carried out in a controlled environment, where data was collected regarding how a user interacts with the system and the problems they are up against. The system was evaluated in terms of effectiveness, efficiency, and achievement of the objectives.

The prototype was tested by a total of 9 volunteers, which resulted in a varied sample of users, with different characteristics. None of the individuals had contact with the system before the start of the tests. A test protocol was followed, and volunteers were asked to complete an initial questionnaire, followed by three tasks in the chatbot and, in the end, the completion of a questionnaire where they were asked to evaluate the functionalities of the tested system and the degree of satisfaction.

The sample consists of nine individuals of both genders, aged over 25 years, with educational qualifications at the level of secondary education or higher education and different levels of knowledge about chatbots. As regards gender, 6 of the volunteers are female, which is equivalent to 66.7%, and three users are male, accounting for 33.3% of all respondents. Of the total number of respondents, 77.8% had educational qualifications at the level of higher education and 22.2% at secondary education level. When asked about the previous use of chatbots, 33.3% of respondents admit to having never used this type of system, while 66.7% said they had used it previously.

Volunteers were asked to perform three tasks in the chatbot to measure the effectiveness and efficiency of the system, as well as the degree of satisfaction in the use of the system.

The usability measures and criteria used in this assessment are as follows:

- *Effectiveness:* Measured through the number of satisfactory responses and the number of users who have successfully completed the proposed tasks.
- *Efficiency:* Measured by analyzing the operation of the system.
- *Satisfaction:* Measured through the experience of use, degree of difficulty in using the system and recommendation of chatbot to other people.

6.1.3 Results

After performing the three proposed tasks, volunteers were asked to fill out a questionnaire based on their interaction with the prototype. The answers obtained allowed to evaluate the functionalities of the system in terms of effectiveness and efficiency, as well as their degree of satisfaction.

Effectiveness is a metric that allows you to measure the relationship between the tasks performed by users and the achievement of satisfactory results to their searches. In this test, most volunteers (88.9%) were able to complete all the proposed tasks, with only one user indicating that they were able to complete only one task.

In turn, efficiency is the metric that allows you to verify that the system worked as expected. This metric is measured by investigating the existence of errors that made the system work smoothly. There were also two users who encountered a system failure and were forced to resume their interaction with chatbot. Despite these setbacks, no user considered that the chatbot had not been efficient in performing tasks.

The degree of satisfaction is one of the measures of the usability test. We measure this metric through questions related to the experience of interaction with chatbot, degree of difficulty in using the system and whether they would be receptive to recommending the chatbot to others. No user negatively classified

their interaction with the chatbot and most considered it to be easy to use.

The results show that users had a positive experience with the chatbot, which can be proven by the fact that most users say that they would recommend the chatbot tested to others, as can be seen in Figure 7.

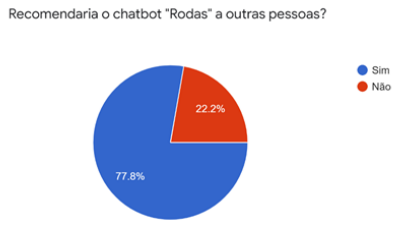

Figure 7 – Number of persons who would recommend the chatbot to other persons

Conclusion

This work was designed with the purpose of verifying whether it would be possible and feasible to architect and implement a chatbot to support the search for information in the space contained in a given website. For this, the DSR methodology was used.

After a brief introduction and the specification of the problem, exhaustive research was carried out on content management, chatbots and information research on the Internet, as well as other related technologies, to verify the state of the art and assess the feasibility and interest in the use of these systems.

Based on the literature review, it was found that the usage of chatbots as a web information research tool is seen as an interesting and potentially idea, both by industry and the scientific community. However, it has been found that information retrieval through these systems is especially complex due to natural language and they still have very limited conversational capabilities.

To carry out the proposal of the chatbot conceptual model, we started by analyzing users and tasks, based on a questionnaire. This analysis has also identified the tasks that users need to perform and how they do so today, using existing systems. It has been found that these systems are not yet used very often, but that there is interest in the use of these tools for the search of content on the Internet, mainly due to their AI capabilities. Chatbots are seen by users as a technology with a promising future and with the potential to become the leading content search tool on the Internet.

7.1 Future work

In terms of future work, there are many lines of development that can be followed:

- Compare the use of search engines with chatbot in terms of effectiveness and ease of use.
- Rethink the chatbot UI and the way content is displayed to make it more accessible from smaller screens like smartphones. The automatic creation of abstracts generated by artificial intelligence can be a way to decrease the amount of information presented and can also facilitate the reading and comprehension of the text.
- Explore the integration of sentiment analysis and emotional intelligence features to try to identify users' emotions
- Expand the model to other use cases, namely within the areas of information management, communication, and business intelligence.

REFERENCES

[1] The Internet World Stats. 2021. *World Internet Users Statistics and 2021 World Population Stats*. https://www.internetworldstats.com/stats.htm

[2] Julia Heidemann, Mathias Klier and Florian Probst. 2012. Online social networks: A survey of a global phenomenon. *Computer Networks, 56*(18), (December 2021), 3866–3878. https://doi.org/10.1016/j.comnet.2012.08.009

[4] Amy N. Langville and Carl D. Meyer. 2012. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press.

[5] Alexandru Tatar, Marcelo Dias de Amorim, Serge Fdida and Panayotis Antoniadis. 2014. A survey on predicting the popularity of web content. *Journal of Internet Services and Applications, 5*(1), 1–20. https://doi.org/10.1186/s13174-014-0008-y

[6] Michael Chau and Hsinchun Chenn. 2003. Personalized and Focused Web Spiders. In: *Zhong N., Liu J., Yao Y. (eds) Web Intelligence*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-05320-1_10

[7] Kenneth Laudon and Jane Laudon. 2017. *Management Information Systems: Managing the Digital Firm* (15th ed.). Pearson.

[8] Anita Kumari & Jawahar Thakur. 2019. Semantic web search engines: A comparative survey. *International Journal of Scientific Research in Computer Science Engineering and Information Technology, 5*(1). https://doi.org/10.32628/CSEIT195115

[9] Edgar Marx, Ali Khalili and Andre Valdestilhas. 2017. Semantic Search User Interface Patterns: An Introduction. In *Human Computer Interaction (HCI-Europe)*. WSCG, Plzen, Czech Republic. http://hdl.handle.net/11025/29622

[10] Binxing Fang, Yan Jia, Li Xiaoyong, Aiping Li and Xindong Wu. 2017. Big search in cyberspace. *IEEE Trans-actions on Knowledge and Data Engineering, 29*(9), 1793-1805. https://doi.org/10.1109/TKDE.2017.2699675

[11] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani and William Bruce Croft. 2019. Asking clarifying questions in open-domain information seeking conversations. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 475-484. https://doi.org/10.1145/3331184.3331265

[12] Wei-Guang Teng, Wei-Hsun Wen and Yue-Cen Liu. 2012. From experience to expertise: Digesting cumulative information for informational web search. *Journal of Information Science and Engineering, 28*(1):161-176. https://www.researchgate.net/publication/220587940

[13] Pavlos Fafalios, Helge Holzmann, Vaibhav Kasturia and Wolfgang Nedji. 2017. Building and querying semantic layers for web archives. *ACM/IEEE Joint Conference on Digital Libraries (JCDL),* 1-10. https://doi.org/10.1109/JCDL.2017.7991555

[14] Ganggao Zhu and Carlos A. Iglesias. 2018. Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Systems with Applications, 101*, 8-24. https://doi.org/10.1016/j.eswa.2018.02.011

[15] Waleed Azmy and Ossama Emam (2013). Content-based recommendation system using search engine. *The Egyptian Society of Language Engineering*, ESOLE'13, Cairo, 13. https://www.researchgate.net/publication/267631372

[16] Marwa Hussien Mohamed, Mohamed Helmy Khafagy and Mohamed Hasan Ibrahim. 2019. Recommender systems challenges and solutions survey. *International Conference on Innovative Trends in Computer Engineering (ITCE),* Aswan, Egypt, 149-155. https://doi.org/10.1109/ITCE.2019.8646645

[17] Timothy J. Hazem, Alexandra Olteanu, Gabriella Kazai, Fernando Diaz and Michael Golebiewski. 2020. On the social and technical challenges of web

search autosuggestion moderation. arXiv [cs.CY]. https://arxiv.org/abs/2007.05039

[18] Markus Kattenbeck and David Elsweiler. 2019. Understanding credibility judgements for web search snippets. *Aslib Journal of Information Management. 71*(3), 368-391. https://doi.org/10.1108/AJIM-07-2018-0181

[19] Elayne Ruane, Théo Faure, Ross Smith, Dan Bean, Julie Carson-Berndsen and Anthony Ventresque. 2018. BoTest. *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*. https://doi.org/10.1145/3180308.3180373

[20] Alan R. Hevner, Salvatore T. March, Jinsoo Park and Sudha Ram. 2004. Design science in information systems research. *MIS Quarterly,28*(1), 75-105. https://doi.org/10.2307/25148625

[21] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger and Samir Chatterjee. 2007. A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/mis0742-1222240302

[22] Alan Hevnern and Samir Chatterjee. 2010. Design science research in information systems. *Integrated Series in Information Systems (ISIS), 22*, 9-22. Boston, MA: Springer. https://doi.org/10.1007/978-1-4419-5653-8

[23] Deane Barker. 2016. *Web Content Management: Systems, Features, and Best Practices* (1st ed.). O'Reilly Media.

[24] Jesse James Garrett. 2010. *The Elements of User Experience: User-Centered Design for the Web and Beyond* (2nd Edition) (Voices That Matter) (2nd ed.). New Riders.

[25] Jove Jankulovski and Pece Mitrevski. 2017. Cloud Computing and Content Management Systems: A Case Study in Macedonian Education. *International Journal on Cloud Computing: Services and Architecture, 7*(5), 01–13. https://doi.org/10.5121/ijccsa.2017.7501

[26] Marcos Baez, Florian Daniel, Fabio Casati. 2020. Conversational Web Interaction: Proposal of a Dialog-Based Natural Language Interaction Paradigm for the Web. In: Følstad A. et al. (eds) *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science, vol 11970*. Springer, Cham. https://doi.org/10.1007/978-3-030-39540-7_7

[27] Mark Sanderson and W. Bruce Croft. 2012. The History of Information Retrieval Research. In *Proceedings of the IEEE, vol. 100*, no. Special Centennial Issue, pp. 1444-1451, https://doi.org/10.1109/JPROC.2012.2189916

[28] Daniel Carlander-Reuterfelt, Álvaro Carrera, Carlos A. Iglesias, Óscar Araque, Juan F. Sánchez-Rada and Sergio Muñoz. 2020. JAICOB: A Data Science Chatbot. *IEEE Access, 8*, 180672-180680. https://doi.org/10.1109/ACCESS.2020.3024795

[29] Eduardo M. Eisman, Víctor López and Juan Luis Castro. 2012.. A framework for designing closed domain virtual assistants. *Expert Systems with Applications, 39*(3), 3135–3144. https://doi.org/10.1016/j.eswa.2011.08.177

[30] Matt Boyd and Nick Wilson. 2018.. Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. *PLOS ONE, 13*(3), e0194811. https://doi.org/10.1371/journal.pone.0194811

[31] Marie-Claire Jenkins, Richard Churchill, Stephen Cox and Dan Smith. 2007. Analysis of User Interaction with Service Oriented Chatbot Systems. In: Jacko J.A. (eds) *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments. HCI 2007. Lecture Notes in Computer Science, vol 4552*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-73110-8_9

[32] Amr Kayid. 2020. The role of Artificial Intelligence in future technology. *International Journal of Innovative Research in Advanced Engineering (IJIRAE), Issue 04*, Volume 5, 146-148. https://doi.org/10.26562/IJIRAE.2018.APAE10086

[33] Jafar Alzubi, Anand Nayyar and Akshi Kumar. 2018. Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series, 1142,* 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

[34] Mohamed Alloghanim, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, Ahmed J. Aljaaf. 2020. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In: Berry M., Mohamed A., Yap B. (eds) *Supervised and Unsupervised Learning for Data Science. Unsupervised and Semi-Supervised Learning.* Springer, Cham. https://doi.org/10.1007/978-3-030-22475-2_1

[35] Ian Goodfellow, Yoshua Bengio and Aaron. Courville. 2017. *Deep learning*. The MIT Press.

[36] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, G. 2015. Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature1453953

[37] Stuart Russell and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach* (Pearson Series in Artifical Intelligence) (3rd ed.). Pearson.

[38] Alexander Maedche, Stefan Morana, Silvia Schacht, Dirk Werth and Julian Krumeich. (2016). Advanced User Assistance Systems. *Business & Information Systems Engineering, 58*(5), 367–370. https://doi.org/10.1007/s12599-016-0444-2

[39] Florian Daniel, Maristella Matera, Vittorio Zaccaria and Alessandro Dell'Orto. 2018. Toward truly personal chatbots. *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services*, 31–36. https://doi.org/10.1145/3195555.3195563

[40] Diksha Khurana, Aditya Koli, Kiran Khatter and Sukhdev Singh. 2017. Natural Language Processing: State of The Art, Current Trends and Challenges. https://arxiv.org/abs/1708.05148

[41] Saurabh Srivastava and Tadinada Venkata Prabhakar. 2020. Intent Sets. *Proceedings of the 2020 12th International Conference on Computer and Automation Engineering*, 194–199. https://doi.org/10.1145/3384613.3384639

[42] Aditya Ankur Yadav, Ishan Garg and Pratistha Mathur. 2019. PACT - Programming Assistant ChaTbot. *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT),* 1. https://doi.org/10.1109/icct46177.2019.8969070

[43] Daniela López De Luise, Andrés Pascal, Claudia Alvarez, Marcos Tournoud, Carlos Pankrac and Juan Manuel Santa Cruz. 2021. Information Retrieval in Restricted Domain for ChatterBots. In: Balas V., Jain L., Balas M., Shahbazova S. (eds) *Soft Computing Applications. SOFA 2018. Advances in Intelligent Systems and Computing, vol 1222*. Springer, Cham. https://doi.org/10.1007/978-3-030-52190-5_23

[44] Boris Galitsky. 2019. Learning Discourse-Level Structures for Question Answering. In: *Developing Enterprise Chatbots*. Springer, Cham. https://doi.org/10.1007/978-3-030-04299-8_7

[45] Alexandra Vtyurina, Denis Savenkov, Eugene Agichtein and Charles Clarke. 2017. Exploring Conversational Search With Humans, Assistants, and Wizards. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. https://doi.org/10.1145/3027063.3053175

[46] Seong-Soo Jeong and Yeong-Seok Seo. Improving response capability of chatbot using twitter. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-019-01347-6

[47] Ning Sa and Xiaojun Yuan. 2020. Challenges in conversational search: Improving the system capabilities and guiding the search process. *WMSCI 2020 - 24th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings. Volume 3, 2020*, Pages 37-42. 24th World Multi-Conference on Systemics, Cybernetics and Informatics, WMSCI 2020 https://www.iiis.org/CDs2020/CD2020Summer/PapersS3.htm#/

[48] Boris Galitsky. 2021. A Virtual Social Promotion Chatbot with Persuasion and Rhetorical Coordination. In: *Artificial Intelligence for Customer Relationship Management. Human–Computer Interaction Series*. Springer, Cham. https://doi.org/10.1007/978-3-030-61641-0_4

[49] Bayan Abu Shawar and Eric Atwell. 2007. Different measurement metrics to evaluate a chatbot system. *HLT-NAACL 2007.* https://aclanthology.org/W07-0313/

[50] Amélie Beriault-Poirier, Sandrine Prom Tep, Sylvain Sénécal. 2019. Putting Chatbots to the Test: Does the User Experience Score Higher with Chatbots Than Websites? In: Ahram T., Karwowski W., Taiar R. (eds) *Human Systems Engineering and Design. IHSED 2018. Advances in Intelligent Systems and Computing, vol 876*. Springer, Cham. https://doi.org/10.1007/978-3-030-02053-8_32

[51]. Mohit Jain, Pratyush Kumar, Ramachandra Kota and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. *Proceedings of the 2018 Designing Interactive Systems Conference,* pp 895-906.. https://doi.org/10.1145/3196709.3196735

[52] Jacky Casas, Marc-Olivier Tricot, Omar Abou Khaled, Elena Mugellini and Philippe Cudré-Mauroux. 2020. Trends & Methods in Chatbot Evaluation. *Companion Publication of the 2020 International Conference on Multimodal Interaction,* 280–286. https://doi.org/10.1145/3395035.3425319

[53] Petter Bae Brandtzaeg, & Asbjørn Følstad. 2017. Why People Use Chatbots. *Internet Science,* 377–392. https://doi.org/10.1007/978-3-319-70284-1_30

[54] Boris Galitsky. 2019. A Content Management System for Chatbots. In: *Developing Enterprise Chatbots*. Springer, Cham. https://doi.org/10.1007/978-3-030-04299-8_9

[55] Wari Maroengsit, Thanarath Piyakulpinyo, Korawat Phonyiam, Suporn Pongnumkul, Pimwadee Chaovalit and Thanaruk Theeramunkong. 2019. A Survey on Evaluation Methods for Chatbots. *Proceedings of the 2019 7th International Conference on Information and Education Technology - ICIET 2019*, 111–119. https://doi.org/10.1145/3323771.3323824