

Fall 2023

Implementing a Smart Campus Chatbot Using TF-IDF

Erica Lee
Bard College

Follow this and additional works at: https://digitalcommons.bard.edu/senproj_f2023



Part of the [Robotics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Lee, Erica, "Implementing a Smart Campus Chatbot Using TF-IDF" (2023). *Senior Projects Fall 2023*. 26.
https://digitalcommons.bard.edu/senproj_f2023/26

This Open Access is brought to you for free and open access by the Bard Undergraduate Senior Projects at Bard Digital Commons. It has been accepted for inclusion in Senior Projects Fall 2023 by an authorized administrator of Bard Digital Commons. For more information, please contact digitalcommons@bard.edu.

IMPLEMENTING A SMART CAMPUS CHATBOT USING TF-IDF

Senior Project Submitted to
The Division of Science, Math, and Computing
of Bard College

by
Erica Lee

Annandale-on-Hudson, New York
December 2023

Abstract

My goal for this project is to implement a smart campus chatbot that can answer questions about Bard College. As higher education continues to adapt to technological advancements, AI technology is becoming increasingly utilized. There are many different ways to create a retrieval-based chatbot, however, this chatbot uses Term Frequency - Inverse Document Frequency(TF-IDF). TF-IDF is a process of finding the highest similarity between questions and processed data by applying statistical weight to certain words. Chatbots in higher education contexts have been shown to increase engagement. Utilizing TF-IDF, I have created a Bard Campus chatbot to enhance the academic experience.

Dedication

I would like to dedicate my senior project to my parents.

Acknowledgments

I would like to acknowledge the contribution of my senior project advisor, Rose Sloan, who has guided me through the whole process. Without your guidance, this project would not have been possible. I would also like to thank Robert McGrail and Ethan Bloch. I am grateful for all the support I have received these past four years.

Table of Contents

Introduction.....	2
Capabilities, Benefits, and Challenges In Creating a Functional Chatbot.....	6
Background Information and Related Work	13
Methodology.....	17
Analysis of Result.....	23
Conclusion.....	28
Bibliography	

Introduction

As time progresses, Education often updates itself to account for changes in the world. Advancements in the divisions of science, math, arts, and more must be accounted for. Technological advancements are no different. The current education system is experiencing a shift as technology becomes a driving force in shaping the learning environment. Technology has greatly expanded access to education, as educational institutions worldwide actively incorporate digital tools and platforms. Massive amounts of information such as books, audio, images, videos, and more are available through the internet. Once viewed as a supplementary tool in education, technology has evolved into a fundamental catalyst for transformative change. Online learning platforms provide students with access to information, allowing them to engage with course content beyond traditional classroom boundaries. Furthermore, the roles of teachers and students have also been greatly influenced by the advancement of innovation being incorporated into the classroom. In the traditional methods of education, the teacher is the primary source of information and the students receive it. However, since technology has enabled access to information and educational opportunities, the teacher's role shifts to incorporate guidance as students take more responsibility for their learning. Through a more intuitive connection between inside and outside the classroom, educational institutions have begun redesigning learning spaces to enable this new model of education. Technology is a powerful tool that can transform education in many ways and we are seeing this firsthand in modern times.¹

In the same way that educational institutions have to adapt to the evolving plethora of new information, they also must adapt to the technologies at hand to provide the highest quality of teaching and learning possible. On both ends of the stick, the teacher gains the ability to teach in more interactive ways and the students gain more tools to aid in their learning process. In more current times, Artificial Intelligence is a growing avenue of higher education learning. As a tool, it has the potential to provide a more adaptive teaching and learning structure for users. Within the umbrella term of Artificial Intelligence, chatbots are becoming used more and more in different contexts; it may be able to play a larger role in a collegiate context.¹

A chatbot, also known as a chatterbox or Artificial Conversational Entity, is an AI program that mimics human conversations using techniques such as Natural Language Processing, image and video processing, and voice analysis. Specifically, a chatbot designed for a college management system employs AI algorithms to analyze user queries and responses. Users can easily interact with the chatbot by asking and receiving answers based on integrated artificial intelligence. Moving forward, chatbots can be helpful for a variety of reasons and different people. The implementation of a chatbot can provide answers and solutions to inquiries, relieving individuals from having to physically visit the college website for answers. It makes campus more accessible for students, faculty, and visitors who find themselves searching for locations, dates, and more. A chatbot is available to users regardless of the time or place and gives accurate information. In this project, we will create a Bard campus chatbot using TF-IDF to enhance the academic experience through increased accessibility and campus engagement.²

The Rationale for Developing a Smart Campus Chatbot

As technology continues to redefine how students consume information, there is a growing need for educational institutions to adapt. Modern learners, accustomed to instant access and personalized interactions, seek efficient and user-friendly services. A smart campus chatbot emerges as a strategic response to these evolving needs, providing a centralized, accessible, and intelligent interface for students to interact with campus resources and services.³ Even though this chatbot is designed to aid students in their interactions with Bard College, it can also be useful to visiting parents, speakers, and guests who are being accommodated on campus. Beyond the immediate viewers of the chatbot, Bard College hosts a beautiful thousand acres, including pathways to walk and enjoy a rural landscape that is open to the public. Visitors of all backgrounds and purposes may benefit from a smart and interactive campus chatbot.

The development of a smart campus chatbot is motivated by the need to improve the accessibility for the users. By offering a virtual assistant accessible from various devices and by operating beyond office hours, the users can have access to school information at any time and anywhere. They can get information on academics, events, and services at their convenience, reducing dependency on administrative staff. Also, while a live agent is limited to handling one or two operations at a time, a chatbot can operate without limitations. Having a chatbot can improve the response rate compared to relying solely on a human support team.⁴

Beyond the pragmatic benefits, a smart campus chatbot has the potential to foster a connected campus community. Serving as a virtual hub for information and assistance, the chatbot can go beyond its functional advantage to encourage a sense of belonging, collaboration,

and support among both students and staff. Once it is integrated into the daily lives of students and staff, the chatbot can become a virtual companion that makes the overall campus experience better.⁴ Especially at Bard, where the quiet surroundings and physical distances can lead to feelings of isolation or difficulty in reaching out first, the chatbot can connect students with lines of help at any time. A chatbot may be useful in de-escalating situations while fostering a more inclusive environment through the availability of information and responsiveness, making students feel included. As a ripple effect, Bard could experience further development towards a more cohesive community due to individuals being more engaged through being aware of campus activities, events, and services. On the further side of its informational role, the chatbot can serve as a resource for emotional support and guidance. By understanding individuals' needs, the chatbot can assist the overall well-being of students.

The goal of my senior project is to explore the world of chatbots and AI by investigating the various ways that one can be created. There are various options available for the creation of chatbots, such as Word2Vec, FastText, Doc2Vec, and so on, however, I utilize Term Frequency-Inverse Document Frequency (TF-IDF). My chatbot utilizes this numeric statistic while relying on web scraping. This process allows for the expansion of both my ability and knowledge pertaining to A.I. with the implementation of such in a higher education setting. We will explore the possibilities of a campus chatbot by discussing the various available chatbot resources and an analysis of historical framework with a breakdown of my Bard campus chatbot.

Capabilities, Benefits, and Challenges In Creating a Functional Chatbot

Capabilities

There are various capabilities that a chatbot built on TF-IDF must have, such as relativeness, retrieval capabilities, and contextual understanding. A chatbot must be able to take in a user's question, retrieve a correct answer, and ask relevant questions that may further enhance its ability to carry out its job, making the user happy. In further detail, the next paragraph will take a closer look into the fundamental capabilities that create a functional chatbot.

A chatbot must be able to respond to user questions that are related to its specific purpose. In my case, my chatbot must be able to respond to various questions related to campus life, academic programs, events, and other questions along those same lines. To do so, the chatbot must be designed in such a way that allows it to understand and interpret natural language inquiries that are asked. It must be intuitive to be able to understand questions that may be asked in different ways. Furthermore, following the questions, it should be able to respond with relevant and accurate information so that the user will have a helpful and impactful experience.²

In order to respond to the questions in a useful manner, the chatbot must have the ability to retrieve the information. I utilize web scraping in order to retrieve information from the Bard school website, so that the chatbot may be functional. This includes details about campus facilities, important dates, course offerings, and more. The chatbot must be trained to navigate

the scraped data from the website so that it may obtain information and respond in a clear, concise, and understandable manner.

There are common questions that a campus chatbot may be asked, such as cafeteria location and hours, information related to the gym, classes, and more. A chatbot must be trained to take in these common questions that students, families, or faculty may typically ask. For students specifically, this may look like questions related to campus rules, regulations, registration procedures, and admission requirements. The ability to handle common questions is essential to the function of this chatbot. Due to its limited nature (a chatbot designed specifically for the implementation of on-campus use), it must be able to take in these questions and understand them.

Along with the ability to handle common questions that may come forward, a chatbot must contain basic contextual understanding. This enables the chatbot to take in inquiries about various topics and then follow up. If a student asks questions about a certain event that is happening on campus, the chatbot must be able to provide information about the event, including related details such as the schedule or location. If another student asks about dining locations, the chatbot may be able to provide helpful information such as the hours of operation, or possibly an address if needed. Doing so only improves the relationship between the user, campus, and chatbot. Furthermore, through contextual understanding, it may also provide relevant web pages and links. This improves the likeability of the chatbot. Searching for certain information on the school website can be time-consuming sometimes. This may be due to various tabs combined with a large amount of information that is contained within a campus website. It would be most

helpful that the chatbot be able to guide students through and direct them to the appropriate places.

The process of web scraping may introduce different variations in the data retrieval process. Due to this possibility, the chatbot must be designed to understand when it can't find certain information based on a user's inquiry and provide clear responses on rephrasing questions or even alternative sources. This robust error-handling process points to an important reason why the chatbot even exists. The intention of a chatbot is to make life easier for the user. If the chatbot does not have an answer for a specific question, then it must point in the correct direction and aid the user in finding the solution that they are searching for. This could be done by providing alternative sources and solutions for the individual using the chatbot.

This chatbot will not be perfect, however, processes must be included so that it can be expanded upon and improved. There must be updates to the chatbot's knowledge based upon changes to the school website and locations of web scraping. There will be improvements to the use and ingenuity of the chatbot as well. Working to make the chatbot more accessible and intelligent continues even after the implementation. Updates should be made to improve the overall system. Therefore, we must monitor user interactions, collect and analyze the feedback, and then use the information to enhance the user experiences.

Benefits and Challenges of Implementing Chatbots at Bard specifically

Many benefits may come from implementing a chatbot at Bard. There is overall enhanced user engagement and interactions that come from chatbots, and at Bard, the same would hold. In an

educational setting, immediate responses to queries provide opportunities for engagement and interactions to be increased. This provides support for a dynamic and responsive environment on campus, facilitating engagement between Bard students and faculty with the college.

Bard is a large campus that hosts a plethora of buildings, events, and general activities. With improved accessibility to information and resources, students can be up to date on courses, schedules, and campus services. There could be time-sensitive inquiries on-campus spaces that often showcase events such as MPR and Fisher Center. These different campus venues host an array of diverse events that show off artistic inspiration and cultural pride. This ability enhances the student experience and reduces the time spent searching for resources, which allows students to showcase their passions to as many people as possible. With the demographics and history of events in mind, a Bard chatbot may be able to go beyond basic assistance and provide a more personalized experience. Through offering tailored responses for individual needs, the relationship between students with campus life would improve tenfold. With adaptive resources, it will be able to cater to the diverse needs of students and faculty.

A bard chatbot may be able to streamline administrative processes on campus, which sometimes may prove to be a hassle. At an institution like Bard, many processes exist to get specific things done, and smart technology can aid in simplifying these steps. A chatbot optimizes efficiency. Tasks that include admissions, enrollment, and general inquiries can now be automated, making the greater administrative staff free to tackle other serious issues that arise. This improves the overall livelihood of the college by offloading some of the less difficult tasks and providing more faculty with the tools necessary to get things done.

Overall, Bard College only stands to benefit from the incorporation of a chatbot. With a chatbot, the support system goes beyond its previous boundaries and becomes a more intimate part of a student's academic journey. Not only do the students and faculty gain from having a chatbot, but there will be residual benefits for visitors who are unfamiliar with the campus. On parent's weekend, a chatbot could save time for all that come to partake in the festivities and see the beautiful campus their children will enjoy for the rest of their academic lives. Even guest speakers who come to give certain talks may need help arriving at certain destinations and learning about local amenities. A chatbot could provide a valuable service to many.

Challenges and Limitations

The incorporation of smart technology, specifically a chatbot, will include a variety of benefits and challenges at the same time. One of the more prominent challenges that come with technological advancements like AI includes consistent maintenance and regular updates. These ensure continuous functionality, which is important for a campus chatbot to have. To ensure that the chatbot is effective and efficient, it will require regular monitoring to take in user feedback and incorporate new developments. It is a guarantee that A.I. technology will continue to advance and adapt, therefore these worldly updates must be incorporated to maintain relevance. This will ensure a happy user experience, no matter what the inquiry may be.

In the context of Bard's rural campus, one challenge could potentially be integration. Compatibility, coordination with different departments, and data synchronization must be taken into account. There exists an intricate ecosystem of faculties and departments within Bard that

require careful consideration. There must be facilitation of conversation between various departments to guarantee a cohesive and dynamic user experience. While an AI system such as a chatbot can aid in releasing certain faculty from having to do certain tasks, thus freeing them to do other work, it still will require human interaction between people to enhance efficiency and effectiveness on Bard's more rural campus. This means a balance must exist between the chatbot and the people who maintain it.

Strategies for Mitigating Challenges

To mitigate challenges that may arise with the incorporation of a chatbot, there must be practices created that maintain and evolve the chatbot's functionality. The practices should not limit the chatbot to only sustainability but rather ride the current to a more integrated tomorrow. This includes performance monitoring, but with careful consideration, then implementation of user feedback. Software updates are critical. It may be beneficial to collaborate with different AI developers while staying abreast of technological advancements. However, with Bard College's innate creativity, it may also prove useful to collaborate with students and use the next generation of Bard students to continue advancing the software. This may enable students to utilize their fresh perspectives alongside existing and active contributors to evolve the software. Through such a collaborative effort partaking in the A.I. revolution, would foster a new sense of pride and innovation in the Bard community.

Furthermore, integration strategies and careful planning with existing systems will mitigate the challenges of communication between departments. To address certain integration

challenges, there must be compatibility assessments involving IT experts during the process.

There must also be clear communication channels between the departments to facilitate a smooth transition into a more technologically advanced college campus. By creating a collaborative framework that involves both technological experts and Bard students, the college can navigate these challenges in a way that positions itself at the forefront of innovative campus technologies.

Background Information and related work

The first chatbot, ELIZA, was developed in 1960 by an MIT Professor, Joseph Weizenbau, to mimic psychotherapists from only 200 lines of code. From there forward, various chatbots, such as Siri, Google Now, and Amazon Alexa, were developed and over time have become more advanced. At the beginning of 2016, there was the first wave of artificial intelligence technology in the different designs that were coming forward. Popular messaging and social network platforms such as Facebook sought to utilize chatbots as an intuitive way for customers to do certain daily activities. The introduction of chatbots into society has created a new surge for more conversational chatbots that aid in providing accurate information for society. Now, chatbots and certain AI technologies are being implemented in a multitude of locations, including college campuses. In 2022, the birth of chat GPT sparked new conversions over the implementation of AI in higher education. The revolution is not chatGPT, but rather AI, however, chatGPT is one of the first and biggest installments of AI and has become its conduit. In recent times, we have seen the use of chat GPT worldwide in different settings. Students are using the capabilities of chat GPT to carry out a wide variety of tasks. As it is trained in such a large amount of data, it can produce realistic and coherent answers. Due to its NLP capabilities, it has become a very versatile and efficient tool for students around the world.⁵

Many higher education institutions utilize chatbots to reinforce connections between students and their campuses. California State University, Northridge, for instance, implemented a college chatbot, CSUNny, to reduce the rate of freshmen who have committed to a college not showing up to class in the fall and to improve the graduation rate. CSUNny was first introduced

in 2018 and when tested on students, it demonstrated higher retention rates and increased likelihood of graduation. The implementation of CSUNny proved successful, with students demonstrating a greater willingness to ask college-related questions to the chatbot than to humans. Students prefer to use chatbots since they think that the chatbot does not judge them for not knowing certain information and it aligns with their reluctance to make phone calls to the administration. The developments of CSUNny showcase the emerging capabilities that chatbots have through their exemplification of fostering engagement, retention, and academic success.⁶

Similar success can be found by Georgia State University, one of the first institutions to use a chatbot with the goal of helping incoming students navigate the campus, register for classes, and get started in college. Georgia State University's use of the chatbot "Pounce" successfully reduced the number of students not showing up in class by 22% and has continued to evolve since then. It evolved to support academic activities, reminding students of exams and assignment deadlines. Students using Pounce showcased higher pass rates and improved grades. In the case of CSUN and Georgia State University, the chatbots are potentially transforming admissions and academic support.⁷

Chatbots & Dialogue Systems

There are two categories of chatbots. One category is rule-based systems, where chatbots operate based on predefined rules and patterns that they are trained on. These chatbots rely on explicit instructions, making them suitable for well-structured use cases with limited adaptability.

These rules cover a range of possible user queries and intents, and the chatbot responds based on matching these predefined rules. The other category is self-learning chatbots, which use AI or machine learning algorithms to reply to the user's queries using Natural Language Processing (NLP).⁸

Rule-based chatbots, also known as decision-tree chatbots, use pre-defined questions to guide users into solutions the chatbot is familiar with and has answers for. It leads users through a series of questions resembling a flowchart, allowing them to select the most appropriate option to proceed. It keeps offering a follow-up option until the user and the chatbot reach the decision together. Rule-based chatbots map out conversations, in anticipation of what the user might ask and how the chatbot should respond. Building and fitting this chatbot into existing systems is easy and highly secure as it cannot operate autonomously. The constraint lies in its limited ability to learn independently, relying on its predetermined rule sets. It cannot answer any questions outside of defined rules and does not learn through interactions.⁸

The second category of chatbots is Self-learning chatbots. Self-learning chatbots have the ability to understand the genuine intent behind user queries, offering effective responses. They can be further categorized into retrieval-based and generative models. Retrieval-based chatbots operate on directed flows or graphs and they are trained to rank the optimal response from a predefined set of responses. These responses can be manually entered or derived from a knowledge base. Retrieval-based chatbots are the most common type of chatbot in recent times. They often work great for specific tasks such as customer support or getting feedback. On the other hand, generative chatbots do not have set responses; they are trained on extensive prior conversations to generate responses. They are used for casual conversations, always ready to

respond back to the users. Although they consistently produce responses for conversation flows, generative chatbots can generate answers that might seem random or have incorrect grammar and syntax. ⁸

My chatbot is a retrieval-based chatbot. It operates by examining user queries, distinguishing the primary intentions, and tailoring its responses accordingly. In order to ensure that the interactions are purposeful, the chatbot must understand the nuances and conversational context of each query. The response given by the chatbot is sure to be relevant, facilitating intentionally helpful dialogue. The intent refers to the intention or purpose of the user in the conversational flow and the entity is a data point or value that a chatbot can extract from a conversational flow. The intent is an action that the user wants to perform and the entity is a keyword that the chatbot extracts from the user's queries. For instance, if the user is trying to get information about the operation hours of the library, the chatbot identifies the intent as operation hours and the relevant entity as the library. This intent and entity enables the chatbot to provide precise information to the users. ¹¹

Methodology

Prior to implementing the chatbot, I made 5 questions that the smart college chatbot could potentially handle:

- 1) How can I apply for financial aid or scholarships?
- 2) How does the housing assignment process work for freshmen?
- 3) What extracurricular activities or clubs are available for students?
- 4) What health services are provided on campus for students?
- 5) How does the college support students in their career development?

Based upon important questions that incoming students may have and overall campus health-related resources, these questions could be a good starting point for a college chatbot's abilities. This will be used as a metric for the potential that this Bard campus chatbot may have in advancing the accessibility and growth of the college. The following paragraphs will illustrate the process followed to create this campus chatbot.

The smart campus chatbot that I am implementing is a retrieval-based chatbot that can answer user questions. First I have to collect data through web scraping techniques. Web scraping is a technique to retrieve unstructured data from a website and store it in a structured way. This is the initial phase of constructing the chatbot as it allows for the extraction of relevant information from the official website of Bard College. Since the chatbot should be able to give

answers that are the most up-to-date and accurate information, utilizing web scraping can ensure access to that information. Useful data is extracted mainly from the student handbook, as all the important information (academic programs, campus facilities, and general information about Bard College) can be found there. This approach of the data gathering phase is effective in retrieving structured data such as paragraphs and text which could then be processed and used to train the chatbot. Web scraping the official website of Bard College provides a diverse, contextually relevant, and comprehensive dataset as it can gather data from a wide range of topics that users can potentially inquire about.¹²

To successfully scrape the website, I first browse the website and learn about how the data is structured for display by inspecting the site using developer tools. The tool used to extract the data by web scraping involves utilizing Python libraries such as BeautifulSoup and Requests. Using Python's request library, the website's HTML code can be obtained from the Python script. Upon successfully scrapping the raw HTML content, preprocessing the data by cleaning out unnecessary HTML and CSS elements is needed to extract raw text that is accessible and necessary. To do so, Beautiful Soup, a Python library for parsing structured data, is used to obtain only the informative content without any HTML markup and CSS artifacts.¹²

Now that my chatbot has the data to be trained on, it has to learn how to recognize questions and answer them in an efficient manner. In the realm of artificial intelligence, natural language processing (NLP) is an important machine learning technology that enables computer programs to interpret, manipulate, and comprehend human language. However, human language can be complicated. There are many different ways to express one concept and the ambiguities in human speech, such as slang, dialects, and social context, make it hard to write software that

accurately determines the intended meaning of text or voice data. These are the reasons that the development of NLP applications is inherently challenging. However, to address these challenges, the Natural Language Toolkit (NLTK), a Python library, offers assistance for the difficulties that arise when interpreting human speech. NLTK provides a range of built-in tokenizers, such as letter, path, word, keyword, class, N-gram, and more. Among these, the word-punkt tokenizer is used most commonly as it effectively segments sentences at blank spaces. The NLTK tokenizers are recognized for their accuracy, speed, and efficiency.¹³

Understandably, NLP is a difficult step in creating a functioning chatbot that can recognize the nuances of human language. Outside of slang and dialects, the query description can change from one person to another. One user may ask a question in a clear and concise manner while another user may hint at what they need without directly stating it. Since each user is different from one another, the chatbot is required to find what exact information the user seeks to know and find a correct response for the corresponding user query. The chatbot system firstly removes the stop words from the user input, if they are present in the queries asked by the user. Stop words are simply words that occur commonly in a language. In English, examples would include articles, conjunctions, and prepositions such as the, an, for, and that. These are words that may be important for verbal communication but do not provide any meaning for search engines, chatbots, etc. Therefore, the chatbot must select meaningful tokens from the data and eliminate these kinds of words so that it may operate efficiently and carry out its job. After removing the stop words from the user queries, the tokenization and lemmatization process are done. Tokenization is a process of taking a set of text or text and breaking it up into its individual

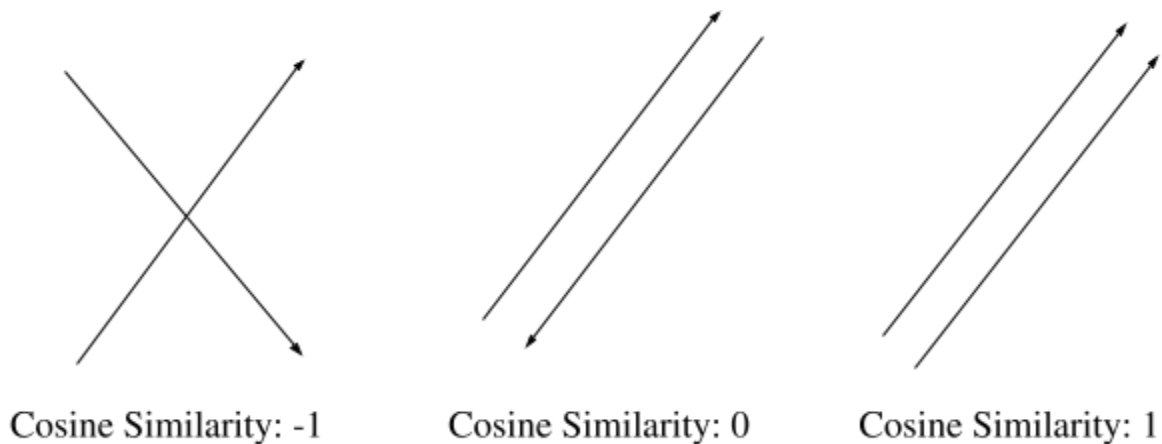
words or sentences. Lemmatization is the process of gathering the different inflected forms of a word so they can be dissected as a solitary item and is a variation of stemming.

The next step in the process is what allows for the chatbot to return accurate answers. The preprocessed data obtained from web scraping is utilized to create a numerical representation suitable for training machine learning models. This involves a key concept known as TF-IDF, Term Frequency-Inverse Document Frequency (a method often used in natural language processing for representing text). TF-IDF is used by the chatbot in its attempt to understand what is being asked and find the answer to that question through a relative comparison of words in the question with words in its data. It tries to find data that is most similar to the question in order to give back an accurate and helpful answer. TF-IDF is commonly used in information retrieval and text mining. TF-IDF represents the product of TF(Term Frequency), which refers to how many times a given term appears in a document, and IDF(Inverse Document Frequency), which measures the weight of the word in the document. Words with a higher TF-IDF score are considered more significant in the context of that sentence, paragraph, or document.¹⁴

When utilizing TF-IDF, The terms that appear frequently in the document are less important than the terms that appear rarely. Once text is quantified using methods like TF-IDF, statistical approaches can be applied to determine how important certain words are within a document or to perform tasks such as keyword extraction, determining the ranking of search results in search engines, and calculating the similarity between documents in a text corpus. It is commonly used for tasks such as calculating document similarity, determining the importance of search results in search systems, and assessing the importance of specific words within a document. In summation, TF-IDF is a method that assigns weights to each word based on that

word's frequency and its inverse document frequency. To produce TF-IDF, I utilize scikit-learn, a Python package designed to facilitate machine learning and AI algorithms. Scikit-learn has a built-in `TfidfVectorizer` class that produces the TF-IDF matrices.¹⁸

Now that the words have assigned weights to them through TF-IDF, I need to compute similarity scores in order for the chatbot to find the relevant answers. To compute a similarity score, I used cosine similarity scores to calculate the numeric quantity that performs the similarity between the two words. Cosine similarity refers to the similarity between two vectors based on the cosine of the angle between them. When the directions of two vectors are identical, it has a value of 1. If they form a 90° angle, the value is 0. If they point in opposite directions (180°), the value is -1. In essence, cosine similarity ranges from -1 to 1, and a value closer to 1 indicates higher similarity. Intuitively, it measures how similar the directions of two vectors are.¹⁶



The cosine similarity between two vectors A and B is expressed by the following formula:

$$similarity = \cos(\Theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

The cosine similarity between the user's input vector and the vectors representing the dataset is calculated through the formula. The chatbot could then identify the most relevant response from its knowledge base. The answer given back is based on the similarity between the terms in the questions. Therefore, when a person inquires about the hours of operation for dining service, the chatbot analyzes the important words and finds the most similar answers. It seeks out the name of the dining place (Kline or DTR) and the hours of operation in its database then returns with an answer based on the similarity it finds. The cosine similarity is what allows my chatbot to be helpful and interactive for the questions it gets asked.

In the chatbot implementation, specific questions were manually crafted to address particular types of user queries. This approach was used to handle questions that may not be effectively addressed through TF-IDF similarity matching alone. For questions related to the college's presence on social media, a response was created manually to produce direct links to Bard College's social media profiles. Lab hours at the Reem Kayden Center(RKC) are also addressed manually since lab hours are specific and may change based on the location. This was done to ensure accuracy in responding to questions.

Analysis of Result

As previously mentioned, the following 5 questions would be used as a metric of potential for the chatbot:

1. How can I apply for financial aid or scholarships?
2. How does the housing assignment process work for freshmen?
3. What extracurricular activities or clubs are available for students?
4. What health services are provided on campus for students?
5. How does the college support students in their career development?

The chatbot gave the following responses to the questions above:

1. *“Increases in need-based scholarship may be obtained only upon application to the Financial Aid Review Committee, which meets three times per year (June, August, and December). This process also applies to students who did not apply for or receive aid when they entered Bard College, and now find they need to apply for need-based scholarships. Information on how to apply to the Review Committee may be obtained by contacting the Office of Financial Aid.”*
2. *“Student Housing Information All first-year and sophomore students are required to live on campus. First-year students are assigned housing based on the timely submission of housing profiles found on the Admissions Office website. Although it is often difficult to fulfill specific requests, every effort is made by the Office of Residence Life & Housing to do so. Students are typically notified of their placements by mid-July. Students are generally housed for Language and Thinking and Citizen Science in their permanent room assignments...”*

3. *“The Office of Student Activities (OSA) offers resources to all student clubs and supports Bard’s academic mission by contributing to the social, cultural, and intellectual life of the College. OSA also advocates for students by addressing their interests and requests concerning student life. This office serves as a resource for the entire Bard community by organizing large-scale events, overseeing the planning and registration process for student-run events, advising Student Government and student clubs, and collaborating with other offices, programs, and community members to enhance campus life, interfaith and intercultural relations, and more...”*
4. *“Robbins Annex, phone: 7433E-mail: Website: www.bard.edu/healthservicesMonday–Friday 9:00 a.m. – 5:00 p.m. by appointmentBard Health Service provides outpatient care to currently registered students during their academic year to promote optimal physical, emotional, intellectual, and social well-being through primary health care...”*
5. *“Career Development OfficeCampus Center 201, phone: 7539E-mail: Websites: www.bard.edu/cdo and bard.joinhandshake.comThe Bard College Career Development Office (CDO), located in Bertelsmann Campus Center 201, welcomes all students with internship, job, and career-related questions....”*

The chatbot demonstrated its ability to provide accurate and relevant responses to user queries. The success of the chatbot in its ability to deliver appropriate answers is due to the utilization of the TF-IDF technique. This approach enables the chatbot to detect keywords within a user’s query and enhances the chatbot to provide relevant and contextually appropriate

responses. In detail, when the chatbot receives a user input, it utilizes the TF-IDF vectors derived from the training corpus to calculate the cosine similarity between the input and the data text that the chatbot provides responses. As previously stated, it selects the response with the highest similarity score, ensuring that the provided answer is contextually relevant to the user's query, suggesting its ability to retrieve information. Another strength of the chatbot is its ability to handle diverse topics, ranging from financial aid processes to campus facilities. Overall, the performance of the chatbot turned out to be successful and this ability shows its potential expansion in scalability and adaptability.

However, the possible response that the chatbot can give is limited, suggesting several areas for improvement. Although the TF-IDF methodology identifies the keywords, it can't handle the semantic meaning of the text perfectly. For example, the chatbot lacks the ability to distinguish between the difference between "Stevenson Library" and "Stevenson Gym". Even though TF-IDF is effective for capturing keywords, it falls short in discerning small differences. The ambiguity in distinguishing between locations like a library and a gym could be rooted in the shared vocabulary or common phrases associated with both places. TF-IDF depends on the structure of the document corpus, therefore if both "Stevenson Library" and "Stevenson Gym" contain similar patterns in their descriptions or contexts, then it will be challenging for TF-IDF to differentiate between the two locations. Both places may have similar weights in regards to how relevant they are, making it difficult to statistically distinguish the two. If the features that inform TF-IDF are not emphasized enough in the text data, then the response by the chatbot may be sub-optimal. In order to account for this deficit, the chatbot must be trained to understand more advanced natural language processing techniques and possibly semantic models. Contextual

embeddings or word embeddings can better represent the distinction between terms with high similarities by capturing semantic relationships. The chatbot's logic will now be further advanced than before and include a custom entity recognition for certain locations. Beyond the Stevenson Library and Gym, this could also be helpful for the Fisher Studio Arts building and the Fisher Center for Performing Arts. Through incorporating different methods, the system will be able to understand and differentiate entities based on predefined rules. Having such a deeper intuition and customization incorporates domain-specific knowledge, moving beyond TF-IDF. The chatbot's performance in handling complex queries and multi-turn conversations represents another area for enhancement. Using a more advanced natural language understanding (NLU) mechanism can also make the chatbot understand the nuance of the user input better. NLP processes text from structure, grammar, typo, and point of view but NLU is the one that helps the machine to infer the intent behind the language text. Thus, combining NLP and NLU will enable the chatbot to be aware of the conversation's context while simultaneously being able to guide the users on the conversation topic, extracting meaning through a deeper level of interpretation. Implementing a dialogue management system or a context-aware mechanism could contribute to more coherent and contextually aware interactions. One of the goals that are included in the creation of a functional chatbot is smooth conversational skills. The ability for the chatbot to converse back and forth in a way that makes the user feel comfortable. In order for a chatbot to provide a certain level of intimacy, it must be able to infer and process information that includes a greater display of colloquial language. By extending the unigram TF-IDF model to include bigrams or trigrams, the model can capture more contextual information. The vectorizer can be adjusted to include n-grams during training.²⁰

Conclusion

In this project, the focus has been on the implementation of a smart college chatbot that can be used in the educational system. There has been a growing global landscape, specifically in the American higher education system, that utilizes smart technology in an efficient manner which enhances the academic experience. In the case of Bard, there exists a predominately rural campus that engages with the world in a very unique way. The incorporation of smart campus technology, namely a chatbot, stands only to enhance the Bard campus experience. It provides opportunities for engagement within the bard community, improves accessibility, and streamlines administrative processes. On a larger scale, the implementation may push Bard forward to the front of the AI revolution currently happening.

The creation of the actual chatbot creates various challenges. The response possibilities are currently limited and it also has trouble making minor distinctions. While the TF-IDF system still requires some deeper fine-tuning, the overall project displays the potential that a campus chatbot may have. This project has demonstrated successful examples of what a chatbot is capable of. Beyond what it has accomplished, there are a few things that can be improved. One very important feature that my chatbot currently lacks is a “User-Friendly Interface”. It is important to take into consideration the audience and what they are looking for in a chatbot. The ultimate goal is to create a user-friendly interface that facilitates a closer relationship with Bard and its students, families, and faculties. The way that the chatbot is presented to the audience is an important aspect of its functionality. My chatbot also should have contextual understanding and the ability to use concise follow-up questions to ensure a satisfied user experience. This

creates a more engaging and conversational experience with the chatbot. It also aids in making the AI system contain more natural feelings comparable to speaking to faculty in person.

Providing relevant information with questions makes for an intuitive interaction.

Having explored the landscape of retrieval-based chatbots and implementing one, while learning the mechanisms that allow for it to function, I would like to next tackle creating a generative chatbot. Using a Recurrent Neural Network (RNN), I believe the shift would create a more dynamic and conversational AI system. The advancement to a new chatbot system may provide a more personalized and immersive interaction. The pursuit of a functional campus chatbot for Bard will continue to highlight technological advancement as the forefront of priorities and progress the campus toward a stronger communal space.¹⁷

Bibliography

[1] How Has Technology Changed Education | Purdue Online. Purdue University Online.

[accessed 2023 Dec 11].

<https://online.purdue.edu/blog/education/how-has-technology-changed-education>

[2] What is a Chatbot? [accessed 2023 Dec 11].

<https://www.oracle.com/chatbots/what-is-a-chatbot/>

[3] Why does every university need a smart college chatbot? Engati. [accessed 2023 Dec 11].

<https://www.engati.com/blog/college-inquiry-chatbot-system-for-automation-in-education-sector>

[4] Zahedi DS. The impact of technology on education: A closer look. The Times of India.

[accessed 2023 Dec 11].

<https://timesofindia.indiatimes.com/blogs/voices/the-impact-of-technology-on-education-a-closer-look/>

[5] journalist AVTAV is a, Education editor at EF on H. How Universities Can Use AI Chatbots to Connect with Students and Drive Success. Technology Solutions That Drive Education.

[accessed 2023 Dec 11].

<https://edtechmagazine.com/higher/article/2023/02/how-universities-can-use-ai-chatbots-connect-students-and-drive-success>

[6] CSUN Is Using A Bot To Get Freshmen To Show Up For Class. LAist. 2018 Jul 30 [accessed 2023 Dec 11].

<https://laist.com/news/csuns-new-chatbot-nags-students-to-show-up-for-class-and-theyre-listening>

[7] Classroom Chatbot Improves Student Performance, Study Says. Georgia State News Hub. 2022 [accessed 2023 Dec 11].

<https://news.gsu.edu/2022/03/21/classroom-chatbot-improves-student-performance-study-says/>

[8] Jurafsky D, Martin JH. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. 2nd ed. Upper Saddle River, N.J: Pearson Prentice Hall; 2009. (Prentice Hall series in artificial intelligence).

[9] Rule-Based vs. AI Chatbot: Which One is Better? – AirDroid. [accessed 2023 Dec 11].

<https://www.airdroid.com/ai-insights/rule-based-vs-ai-chatbot/>

[10] Rule-Based Chatbots vs. AI Chatbots: Key Differences. [accessed 2023 Dec 11].

<https://www.hubtype.com/blog/rule-based-chatbots-vs-ai-chatbots>

[11] Python R. Beautiful Soup: Build a Web Scraper With Python – Real Python. [accessed 2023 Dec 11]. <https://realpython.com/beautiful-soup-web-scraper-python/>

[12] What is Natural Language Processing? | IBM. [accessed 2023 Dec 11].

<https://www.ibm.com/topics/natural-language-processing>

[13] TF-IDF — Term Frequency-Inverse Document Frequency – LearnDataSci. [accessed 2023 Dec 11].

[https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/#:~:text=Using%20scikit%2Dlearn-,What%20is%20TF%2DIDF%3F,%2C%20relative%20to%20a%20corpus\).](https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/#:~:text=Using%20scikit%2Dlearn-,What%20is%20TF%2DIDF%3F,%2C%20relative%20to%20a%20corpus).)

[14] Understanding TF-IDF for Machine Learning. Capital One. [accessed 2023 Dec 11].

<https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/>

[15] Tutorials DN. TF-IDF and Cosine Similarity in Machine Learning. Dot Net Tutorials.

[accessed 2023 Dec 11].

<https://dotnettutorials.net/lesson/tf-idf-and-cosinesimilarity-in-machine-learning/>

[16] Martinelli A. Personality for Your Chatbot with Recurrent Neural Networks. Medium. 2017 Mar 30 [accessed 2023 Dec 11].

<https://towardsdatascience.com/personality-for-your-chatbot-with-recurrent-neural-networks-2038f7f34636>

[17] Chauhan NS. How to build a Chatbot. The AI dream. 2020 Sep 21 [accessed 2023 Dec 11].

<https://www.theaidream.com/post/how-to-build-a-chatbot>

[18] Stormon A. 5 Chatbot Challenges and How to Overcome Them. Medium. 2016 Oct 19

[accessed 2023 Dec 11].

<https://chatbotsmagazine.com/5-chatbot-challenges-and-how-to-overcome-them-caccc3a26d7c>

[19] Scott J. What's in a word? Medium. 2021 Mar 29 [accessed 2023 Dec 11].

<https://towardsdatascience.com/whats-in-a-word-da7373a8ccb>

[20] NLP vs NLU: What's The Difference? BMC Blogs. [accessed 2023 Dec 11].

<https://www.bmc.com/blogs/nlu-vs-nlp-natural-language-understanding-processing/>