Chatbot to Support Frequently Asked Questions from Students in Higher Education Institutions

José Rodrigues Neto[1], Péricles B.C. Miranda[1], Rafael Ferreira Mello[1], André C.A. Nascimento[1]

[1]Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco (UFRPE)

Resumo. *Os chatbots são ferramentas que utilizam inteligência artificial para simular uma conversação humana. Eles podem ser utilizados para diferentes aplicações, como fornecer atendimento dentro de um e-commerce, preenchimento de formulários ou responder a FAQs (Frequently Asked Questions). Este trabalho propõe o desenvolvimento de um chatbot para auxiliar os estudantes de uma universidade pública brasileira na busca de informações relacionadas a dúvidas em disciplinas, processos administrativos, e dúvidas gerais sobre seu curso ou universidade. O sistema desenvolvido consegue entregar uma precisão alta na classificação da intenção da pergunta e ter resposta ao usuário em uma ampla margem de tópicos diferentes.*

Abstract. *Chatbots are tools that use artificial intelligence to simulate human conversation. They can be used for different applications, such as providing customer service within an e-commerce store, filling out forms or answering FAQs (Frequently Asked Questions). This work proposes the development of a chatbot to help students from a Brazilian public university in the search for information related to doubts in subjects, administrative processes, and general doubts about their course or university. The developed system is able to deliver a high accuracy in the classification of the intention of the question and to have an answer to the user in a wide range of different topics.*

1. Introdução

O ingresso no ensino superior causa uma série de mudanças na vida de um estudante. Pelo fato do ambiente universitário se tratar de um contexto inédito em sua vida, o estudante pode se sentir perdido em meio a tantos fatores de estresse. Essa situação pode se tornar mais grave com a falta de orientação acerca da burocracia e processos que muitas vezes se afunila a nível de departamento, e que costuma ser um dos maiores empecilhos na adaptação do jovem a universidade, como explica (Teixeira et al. 2008). Assim como mudanças de forma periódica em horários, requisitos e docentes das disciplinas acabam por requerer um constante esforço do aluno na busca de informações em sites e emails da coordenação. As instituições de ensino podem utilizar as Tecnologias de Informação e Comunicação (TIC) para diminuir o impacto desse problema na vida acadêmica do aluno.

Por outro lado, os chatbots vêm evoluindo e conquistando um espaço há muito defasado: o atendimento. Aplicações de chatbot vêm sendo difundidas em diferentes áreas do conhecimento, seja para atendimento em lojas online (Bhawiyuga et al. 2017), assistentes médicos (Srivastava and Singh 2020) ou ferramenta de educação assistiva

(da Silveira et al. 2019). Um chatbot é capaz de reconhecer uma certa gama de sentenças presentes em sua base de dados e buscar respostas de forma instantânea para uma grande quantidade de usuários. Este tipo de aplicação possibilita a modernização do processo de fornecimento de informações, provendo agilidade na resolução de dúvidas e em possíveis problemas, tal como realizar processos de preenchimento de documentos de forma automatizada. Lidando com o atual estado-da-arte, como podemos agregar mais à experiência de um sistema de respostas automáticas lidando especialmente com o âmbito acadêmico? Basear-se neste tipo de questionamento nos impede de seguir a ampla disponibilidade de chatbots para atendimento ao cliente em *marketplaces* e trazer benefícios relacionados ao nosso contexto.

Diante deste contexto, este trabalho tem a proposta de desenvolver uma aplicação de chatbot para auxiliar os alunos de uma universidade pública brasileira a ter acesso a informações acerca de disciplinas, processos administrativos ou dúvidas gerais de forma simples e direta. Esta aplicação deve abranger informações acerca dos processos burocráticos da universidade e estar apto a responder às dúvidas mais comuns que alunos de curso de graduação podem ter em seus primeiros semestres de curso, assim como a manutenção de informações atualizadas a cada período de disciplinas. Entre essas dúvidas, são mais recorrentes as relacionadas com requisitos de disciplina, processos de estágio, atividades complementares e dúvidas gerais como a explicação de siglas usadas comumente por alunos da universidade, por exemplo, RU (Restaurante Universitário), VA (Verificação de Aprendizagem), entre outros. Os resultados indicam que o chatbot proposto consegue alcançar uma precisão de aproximadamente 90.9% na qualidade da resposta enviada.

2. Trabalhos Relacionados

Esta seção tem como principal objetivo listar aplicações de *Chatbots* no contexto educacional. Vários *Chatbots* já foram desenvolvidos para a área da educação, entre eles, podemos destacar o CVChatbot (Dehon et al. 2018). Os autores identificaram uma deficiência entre a comunicação dos professores com os alunos que utilizam o Ambiente Virtual de Aprendizagem (AVA). Esse problema é geralmente ocasionado pela falta de hábito dos alunos em verificar sua caixa de entrada do e-mail institucional ou a falta de verificação das notificações das salas virtuais na própria plataforma. Através de uma integração com o Moodle, o CVChatbot é capaz de aprimorar o processo de comunicação entre os alunos e os professores, enviando mensagens aos alunos através do Facebook Messenger sempre que houver uma nova atualização nas salas virtuais em que ele está matriculado. Através de um questionário de avaliação realizado pelos autores, foi constatado que aproximadamente 72% dos alunos entrevistados consideram o uso do CVChatbot útil no processo de comunicação entre professores e alunos.

Os chatbots também podem ser utilizados para facilitar o processo de aprendizagem. Em (Mikic-Fonte et al. 2018) é apresentado um projeto de chatbot que auxilia os estudantes na disciplina de Arquitetura de Computadores, respondendo às perguntas mais frequentes realizadas pelos alunos. O chatbot fornece uma interface que permite os professores cadastrarem, modificarem e excluírem perguntas e respostas do banco de dados de questões, facilitando a manutenção dos conhecimentos do chatbot. Esse chatbot retira do professor a cansativa tarefa de responder sempre as mesmas perguntas dos alunos e permite que o estudante tenha autonomia durante seu aprendizado por não precisar aguar-

dar uma resposta do professor quando houver alguma dúvida. Os autores não realizaram uma avaliação detalhada sobre a eficiência ou adesão ao chatbot.

Outro exemplo de chatbot educacional é o UNIBOT (Patel et al. 2019). O UNIBOT consegue responder várias perguntas relacionadas aos estudantes e ao funcionamento da instituição de ensino. O UNIBOT utiliza expressões em consultas SQL para encontrar uma resposta apropriada. Porém, essa técnica só fornece respostas relevantes ao usuário caso a pergunta do usuário e a pergunta cadastrada no banco de dados contenham as mesmas palavras, não reconhecendo palavras sinônimas. Essa condição faz com que esse chatbot não obtenha um resultado satisfatório. Esse chatbot foi desenvolvido de modo a permitir que, com poucas alterações, qualquer instituição possa implementar o chatbot em seu site institucional. O projeto é *web-based* e oferece uma interface gráfica simples e similar a um aplicativo de mensagens.

Este trabalho propõe a criação de um chatbot que utiliza técnicas de Processamento de Linguagem Natural (PLN) para auxiliar na resolução de possíveis dúvidas que um aluno pode ter relacionados a processos burocráticos da universidade, dúvidas gerais sobre o curso, requisitos das disciplinas, informações sobre siglas usadas frequentemente na faculdade e localizações de edifícios importantes da universidade. O uso de PLN permite compreender a língua natural e extrair o sentido das perguntas, permitindo uma maior assertividade nas respostas fornecidas.

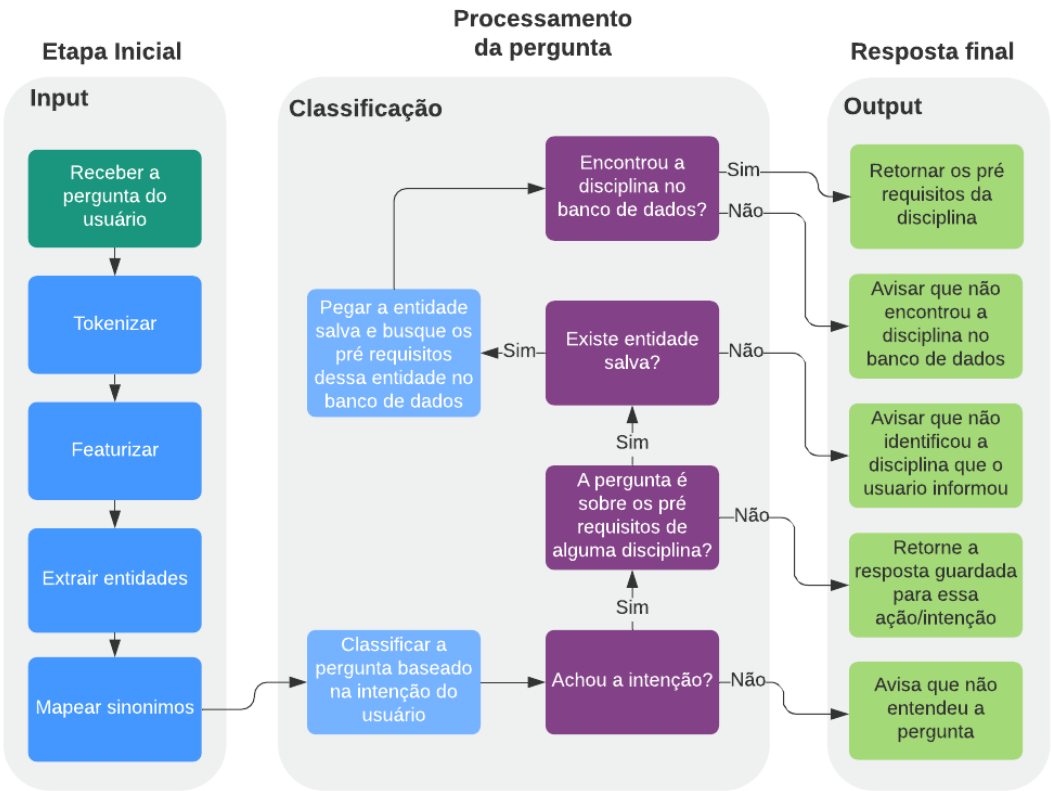

Figura 1. Fluxograma do processamento do sistema de chatbot.

3. Desenvolvimento do chatbot proposto

O chatbot foi desenvolvido utilizando a biblioteca RASA[1], cujo objetivo é facilitar a construção de sistemas deste tipo, oferecendo diversos algoritmos e um ambiente compacto para a sua criação, o qual se utiliza de algoritmos para associar uma pergunta do usuário a uma equivalente armazenada no banco de dados. Este processo envolve passos como pré-processamento do texto, substituição de sinônimos na frase enviada, extração de entidades (palavras que remetem a uma única opção dentre todas as possíveis, por exemplo, Cálculo é uma disciplina dentre todas do curso, logo numa frase ela é a entidade) na sentença, classificação de intenção da pergunta do usuário (a que tópico ela está relacionada) e só então utilizar a base de dados para obter as informações necessárias.

O fluxograma apresentado na Figura 1 descreve o procedimento interno do sistema para que diante da questão realizada pelo usuário, seja conferida a resposta mais adequada. O processamento feito em cada um desses passos tem o intuito de permitir que o usuário possa realizar a mesma pergunta de formas diferentes e o sistema entenda e encaminhe para a mesma resposta. O benefício disto é evitar situações que o banco de dados possua a resposta para um questionamento, mas o sistema não entenda a pergunta pois foi feita com uma variação na escrita, mudança na estrutura da sentença ou com algum erro gramatical ou de digitação. As etapas do fluxograma serão detalhadas a seguir.

3.1. Base de perguntas

Para o treinamento do chatbot proposto neste trabalho, foram extraídas 29 questões das áreas de "Perguntas Frequentes" dos sites do curso de Ciência da Computação de uma universidade pública brasileira. Estas foram divididas manualmente em tópicos por assuntos relacionados, e para cada pergunta foram formuladas, em média 9 perguntas semelhantes, com o intuito de enriquecer a base de dados de treinamento e avaliar a similaridade das respostas. Ao fim deste processo a base continha 476 perguntas. A Tabela 1 apresenta informações quantitativas do banco de perguntas.

Tabela 1. Quantidade de perguntas disponíveis em cada classe.

Classe	Perguntas iniciais	Perguntas finais
Atividades complementares	8	71
Disciplinas	10	281
Trilhas	2	25
Estágio	3	36
Outros	6	63
Total	29	476

Para o tratamento de pré-requisitos de disciplinas, foi realizado um processamento no documento disponibilizado pela universidade contendo as principais informações de todas as 77 disciplinas ofertadas no período. As informação extraídas foram: carga horária, pré requisitos, obrigatoriedade ou não da disciplina, ementa, o departamento que a oferta e número de créditos ofertados.

[1]https://pypi.org/project/rasa/

3.2. Etapa inicial: Pré-processamento

Um passo importante antes de calcular similaridade ou classificar o texto e pré-processar os dados recebidos, para reduzir ruído, e selecionar as características do texto que são mais relevantes. O fluxo de pré-processamento consiste em tokenização, extração de características, extração de entidades e mapeamento de sinônimos. A importância da utilização de técnicas de pré-processamento para melhores resultados é comum em trabalhos de classificação de texto, como (Srividhya and Anitha 2010), (Krouska et al. 2016) e (Toman et al. 2006), que analisam a importância do pré-processamento de texto antes de aplicar os algoritmos de classificação.

A tokenização no início do pré-processamento é feita para tratarmos o dado recebido a nível de palavra, não mais de sentença. O dado é dividido em vocábulos chamados *tokens* e através dele podemos aplicar as demais etapas do pré-processamento. Como a extração de características, um processo que reúne as palavras extraídas pelo processo anterior em um 'saco de palavras' ou *'bag of words'*, criando um multiconjunto que associa a palavra e a quantidade de vezes que essa foi utilizada na oração, é aplicada para que a palavra seja reconhecida em decorrência de sua quantidade independente de sua posição. Para essa etapa, este trabalho utilizou o tf-idf (Chowdhary 2020).

A extração de entidades é utilizada para selecionar determinadas entidades que especificam ou delimitam o escopo do restante da frase (Nadeau and Sekine 2007). Por exemplo, na frase "Preciso saber os pré requisitos para cursar Cálculo 2", a expressão "Cálculo 2" seria a entidade, pois é a disciplina a qual se quer realizar a intenção de buscar os pré requisitos. Dessa maneira, pode-se aplicar essa técnica para extrair nomes de disciplinas, lugares e siglas de modo a que a resposta seja formada de forma personalizada para aquela entidade, ao contrário de um retorno generalizado.

O mapeamento de sinônimos tem seu uso voltado para relacionar palavras escritas na sentença com algum possível sinônimo visando manter uma linguagem padronizada (Naser-Karajah et al. 2021). Então, assim, é possível diminuir o número de frases necessárias para o treinamento e uso do chatbot, uma vez que apenas é necessário mapear palavras sinônimas, não sendo necessário a escrita da frase na forma exata a utilizada no treino, além de gerar uma comodidade na hora da escrita da pergunta já que siglas podem ser usadas, por exemplo, para a descoberta de disciplinas, e.g. 'IA' é automaticamente mapeado para 'Inteligência Artificial'.

3.3. Processamento da pergunta

Esta etapa é responsável por analisar a pergunta enviada pelos alunos para identificar a melhor resposta que o chatbot poderia dar para atender a necessidade levantada. Para isso, algumas etapas são necessárias. As próximas seções explicam cada uma destas etapas.

3.3.1. Classificação de intenção

No chatbot proposto, intenções são as diferentes categorias das perguntas relacionada as dúvidas dos alunos, e que geralmente associadas com uma resposta comum. Desta forma, vários tipos de perguntas podem ser mapeadas e agrupadas em tópicos similares, o que restringe as perguntas que serão selecionadas para o cálculo da similaridade com a pergunta recebida pelo chatbot. No banco de dados utilizado, as perguntas foram separadas

em alguns tópicos que possuíam respostas semelhantes, dessa forma cada pergunta recebida pode ser classificada entre as diferentes intenções.

Questionamentos acerca de: pré requisitos de disciplinas, trilhas de disciplinas, atividades complementares, estágio, professores e calendário acadêmico costumam ser comuns aos ambientes acadêmicos. Estes tópicos foram transformados nas intenções deste chatbot. Ou seja, a pergunta, antes de ser relacionada a uma resposta, foi associada com um conjunto de perguntas semelhantes para a escolha de seu tópico. Também foram adicionados tópicos de "bate-papo", para pedidos corriqueiros que não se encaixam na seção de dúvidas, por exemplo saudações, despedidas, piadas, dentre outros.

Para a classificação das intenções foi utilizado o atual estado-da-arte, o DIET (*Dual Intent Entity Transformer*) (Bunk et al. 2020), uma arquitetura multi-tarefa que pode ser utilizada tanto para classificação de intenção, quanto para reconhecimento de entidades. O DIET une todas as possíveis mensagens do conjunto intenções do treino em um espaço vetorial semântico, visando maximizar a semelhança vetorial entre as mesmas intenções e maximizar a diferença entre intenções distintas.

3.3.2. Busca de respostas

Após classificada a intenção, o chatbot realiza a busca de resposta, para isso ele se baseia no algoritmo de recuperação de informação, o ResponseSelector, que varre o banco de dados desenvolvido que possui exemplos de conversações passadas para assim, baseado na intenção atual, o bot enviar a resposta adequada para cada intenção.

O chatbot também leva em consideração respostas anteriores para se basear na próxima interação, por exemplo um "Sim" por si só não significa nada, mas um "Sim" quando a última interação foi "Você quer ouvir mais piadas?", significa que o usuário quer que o chatbot realize novamente o ato de contar piadas, assim como um "Sim" uma vez que houve um "Eu te ajudei?" como ultima sentença significa que o aluno está retornando um feedback positivo acerca do funcionamento do chatbot.

3.4. Avaliação do chatbot proposto

Para realizar uma experimentação detalhada e verificar a eficácia de diferentes algoritmos e suas combinações, foram gerados 9 arquivos de configurações com diferentes tipos de algoritmos, visando a escolha do arquivo de configuração que tenha uma eficácia satisfatória com um tempo de processamento e uso de memória reduzidos, a fim de reduzir o tempo de treino e obter uma maior agilidade no envio das respostas para o aluno.

Para a experimentação, foram avaliado dois grupos de algoritmos, um com gerados 4 variantes e outro com 5 variantes. Esses grandes grupos tem como principal diferença a existência de um pacote pré-treinado com palavras recorrentes da língua portuguesa. A extração desse pacote de palavras em chatbot foi feito pelo Spacy[2], uma biblioteca focada em processamento de linguagem natural. O grupo com o prefixo "Spacy"possui esse pacote de palavras, diferentemente do grupo "No Pre", que não o possui.

[2]https://spacy.io/

A separação desses dois grandes grupos se dá à escolha de diferentes algoritmos para a extração de características. A biblioteca Rasa, usada para esse trabalho, disponibiliza 4 principais *featurizers* sendo eles: *CountVectorsFeaturizer*, *RegexFeaturizer*, *LexicalSyntacticFeaturizer* e *SpacyFeaturizer*.

O *CountVectorsFeaturizer*, converte a mensagem enviada pelo usuário em uma *bag of words*, um vetor que contém cada palavra usada na mensagem associado a quantidade de vezes que ela foi utilizada. Já o *RegexFeaturizer* executa uma conversão na mensagem enviada para uma expressão regular, uma forma concisa e flexível de identificar cadeias de caracteres de interesse. Diferentemente o *LexicalSyntacticFeaturizer* cria recursos sintáticos e léxicos baseados na mensagem enviada. Por fim, o *SpacyFeaturizer*, que só pode ser utilizado quando se possui um pacote de palavras pré processadas do Spacy, cria um vetor de representação da mensagem enviada pelo usuário e sua resposta.

4. Resultados

Durante o desenvolvimento do chatbot foi utilizada as medidas F1_score, uma média harmônica entre a precisão e cobertura, para realizar a validação e experimentação da abordagem proposta. Para avaliar a capacidade de generalização dos nossos modelos foi utilizada a técnica de validação cruzada K-Fold (Efron 1983), onde foram utilizados 5 *folders*. A Figura 2 apresenta a performance, medida na métrica de F1_score, de cada configuração de grupos e algoritmos utilizados para análise do desempenho da classificação de intenções. No eixo horizontal o número de exemplos que foi utilizado para o treino é exibido, no vertical pode-se ver o resultado de F1_score médio das 5 execuções separando os dados em treino e teste através de um sorteio aleatório.

Constatou-se também que utilizar um pacote de palavras pré treinadas não impacta significativamente na classificação da intenção, aumenta o uso de memória e impacta na velocidade do processamento da pergunta pelo chatbot, assim como o uso dos algoritmos *LexicalSyntacticFeaturizer* e *RegexFeaturizer* que pouco acrescem no resultado da classificação e terminam por estender o tempo do processamento da mensagem. Dado esses resultados, foi escolhido o não uso de palavras pré treinadas e o uso apenas do algoritmo *CountVectorsFeaturizer* para a featurização das perguntas. Por fim, as figuras 3, 4, 5 e 6 apresentam exemplos de interações de alunos com chatbot incluindo informações como intenção previstas, sua confiança e escolha de ação de resposta.

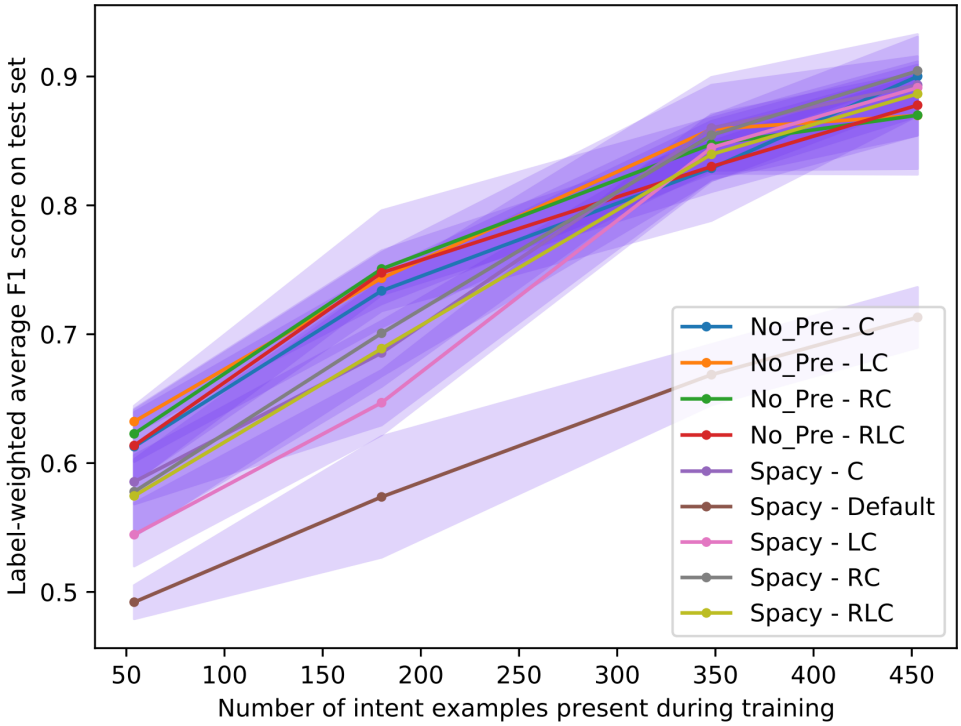

Figura 2. Comparação de diferentes arquivo de configurações do chatbot.

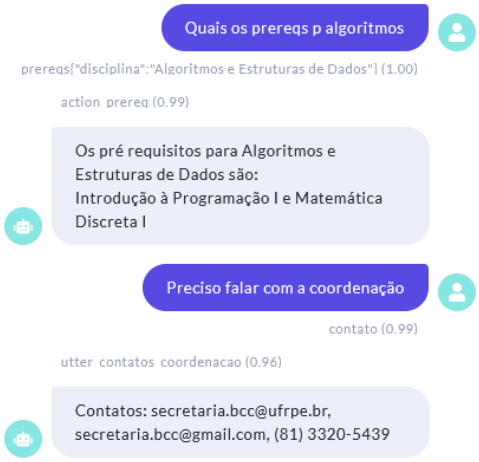

Figura 3. Demonstração de resposta do chatbot.

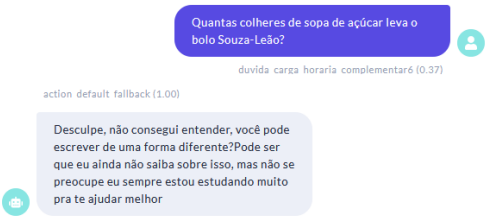

Figura 4. Demonstração da mensagem de erro.

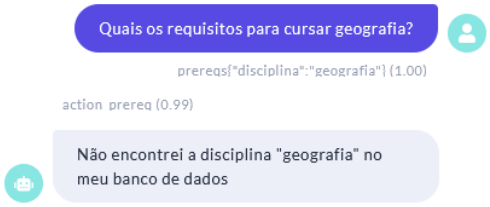

Figura 5. Demonstração de resposta do chatbot sobre disciplina inexistente no banco de dados.

Figura 6. Exemplo de bate-papo ou *chitchat*

5. Discussão

A classificação de intenção restringe o número de dados a ser analisado e comparado, tratando o problema inicial da amplitude de tópicos que o sistema deve abordar. Classificando a pergunta em um assunto em comum com outras perguntas da base de dados, o nosso modelo de classificação por similaridade lida apenas com os dados de perguntas relacionadas.

O pré-processamento aplicado obteve sucesso na extração do conteúdo das perguntas contribuindo para a alta precisão na comparação das sentenças. Pode-se analisar este ponto com a alta taxa de acerto no nosso cálculo de similaridade. Por causa disso, o sistema pode reconhecer diferentes maneiras de fazer a mesma pergunta e mapear estas em uma resposta equivalente, abrangendo mais perguntas que podem ser feitas independente do uso de sinônimos.

Além das perguntas sobre as informações da universidade, o nosso sistema também trata as perguntas relacionadas às disciplinas, utilizando os métodos de processamento de texto para retornar ao usuário as informações das disciplinas presentes na nossa base de dados mesmo que o usuário não digite o seu nome rigidamente igual, podendo assim optar por uma sigla ou nome encurtado. Isso possibilita ao usuário obter com mais facilidade os pré requisitos da disciplina que deseja, um ponto que os trabalhos relacionados não abordaram.

6. Conclusões e trabalhos futuros

Com o avanço da tecnologia e nas pesquisas sobre inteligência artificial, especialmente na área de Processamento de Linguagem Natural, se tornou possível a criação de chatbots de conversação cada vez mais próximo da linguagem humana. Os chatbots fornecem alto nível de eficiência no atendimento, suporte se houver alta demanda de atendimentos, e diminuição no custo operacional.

Este trabalho apresentou um chatbot para auxiliar os novos estudantes acerca dos processos administrativos da universidade, dúvidas recorrentes do curso e conhecimento de pré requisitos de disciplinas. O chatbot está apto a responder as perguntas mais frequentes que um aluno pode ter em seus primeiros meses na universidade. Além do benefício ao estudante, as coordenações dos cursos também são beneficiadas ao evitar a tarefa repetitiva de responder sempre as mesmas perguntas todo início de semestre.

Seria possível realizar avanços na inteligência do chatbot caso fosse realizado um estudo de melhor número de épocas para os algoritmos utilizados para pré processar e classificar o *input* do usuário. Além disso pode-se atualizar o RASA para a nova versão (2.x), a qual já possui novos algoritmos e novos recursos que devem permitir implementações de melhorias, para isso também seria necessário a conversão dos arquivos utilizados no projeto para o novo padrão adotado.

Além do mais a implementação de preenchimento de documentos e formulários via chatbot, assim como o retorno de mais informações acerca da disciplina como seu código, ementa, área ou departamento são áreas que podem ser exploradas para gerar mais comodidade ao aluno, assim como para a coordenação. De mesma maneira, é possível criar a possibilidade de obtenção de comprovante de vínculo com a universidade via chatbot. Semelhantemente, adicionar uma base de siglas utilizadas na universidade (como RU, DEFIS, VA, etc.), bem como a localização dos principais edifícios da universidade de modo a que venha facilitar a inserção do novato à vida acadêmica.

Referências

[Bhawiyuga et al. 2017] Bhawiyuga, A., Fauzi, M. A., Pramukantoro, E. S., and Yahya, W. (2017). Design of e-commerce chat robot for automatically answering customer question. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 159–162.

[Bunk et al. 2020] Bunk, T., Varshneya, D., Vlasov, V., and Nichol, A. (2020). Diet: Lightweight language understanding for dialogue systems.

[Chowdhary 2020] Chowdhary, K. (2020). Natural language processing. *Fundamentals of artificial intelligence*, pages 603–649.

[da Silveira et al. 2019] da Silveira, C., da Silva, A. R., Herpich, F., and Tarouco, L. M. R. (2019). Uso de agente conversacional como recurso de aprendizagem sócio-educacional. *RENOTE-Revista Novas Tecnologias na Educação*, 17(3).

[Dehon et al. 2018] Dehon, P., Silva, A., Inocêncio, A., Castro, C., Costa, H., and Júnior, P. (2018). Cvchatbot: Um chatbot para o aplicativo facebook messenger integrado ao ava moodle. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 29(1):1623.

[Efron 1983] Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78(382):316–331.

[Krouska et al. 2016] Krouska, A., Troussas, C., and Virvou, M. (2016). The effect of preprocessing techniques on twitter sentiment analysis. In *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–5. IEEE.

[Mikic-Fonte et al. 2018] Mikic-Fonte, F. A., Llamas-Nistal, M., and Caeiro-Rodríguez, M. (2018). Using a chatterbot as a faq assistant in a course about computers architecture. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–4.

[Nadeau and Sekine 2007] Nadeau, D. and Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

[Naser-Karajah et al. 2021] Naser-Karajah, E., Arman, N., and Jarrar, M. (2021). Current trends and approaches in synonyms extraction: Potential adaptation to arabic. In *2021 International Conference on Information Technology (ICIT)*, pages 428–434. IEEE.

[Patel et al. 2019] Patel, N. P., Parikh, D. R., Patel, D. A., and Patel, R. R. (2019). Ai and web-based human-like interactive university chatbot (unibot). In *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 148–150.

[Srivastava and Singh 2020] Srivastava, P. and Singh, N. (2020). Automatized medical chatbot (medibot). In *2020 International Conference on Power Electronics IoT Applications in Renewable Energy and its Control (PARC)*, pages 351–354.

[Srividhya and Anitha 2010] Srividhya, V. and Anitha, R. (2010). Evaluating preprocessing techniques in text categorization. *International journal of computer science and application*, 47(11):49–51.

[Teixeira et al. 2008] Teixeira, M. A. P., Dias, A. C. G., Wottrich, S. H., and Oliveira, A. M. (2008). Adaptação à universidade em jovens calouros. *Psicologia escolar e educacional*, 12(1):185–202.

[Toman et al. 2006] Toman, M., Tesar, R., and Jezek, K. (2006). Influence of word normalization on text classification. *Proceedings of InSciT*, 4:354–358.