

PROGRAMA DE BOLSAS ITAÚ (PBI)

Edital 1º Semestre de 2021

Modalidade de bolsa	Iniciação Científica
Título do projeto:	<i>Chatbot Q&A multi-agente</i>
Tópico abordado no projeto:	<i>Chatbots</i>
Nome do aluno:	Enzo Bustos Da Silva
Depto e Unidade do Aluno:	PCS - POLI
Ano de Ingresso na USP:	2019
Nome do Orientador:	Anna Helena Reali Costa
Depto do Orientador:	PCS
E-mail do Aluno:	enzobustos@usp.br
E-mail do Orientador:	anna.reali@usp.br
Link para o CV Lattes do Aluno:	CV Lattes - Enzo Bustos
Link para o CV Lattes do Orientador:	CV Lattes - Anna Reali

Chatbot Q&A multi-agente

Enzo Bustos da Silva (aluno de IC)
Anna Helena Reali Costa (orientadora)

Resumo

O objetivo deste trabalho de Iniciação Científica é tomar contato com arquiteturas de agentes conversacionais e desenvolver um *chatbot* informativo de Q&A utilizando uma arquitetura multi-agente. Para motivação, como a ONU marcou o ano de 2021 como o início da década dos oceanos e o território marítimo brasileiro é uma verdadeira riqueza nacional que abrange uma posição de destaque em diversos tópicos, como economia, meio ambiente, pesquisa científica e soberania nacional, este tópico foi escolhido para o domínio de aplicação do *chatbot*. Este território recebe o nome de Amazônia Azul por ser comparável em extensão à Amazônia florestal. Como esse tópico é desconhecido por uma vasta parte da população, isso motivou esse trabalho, com o qual dados publicamente disponíveis serão utilizados para o desenvolvimento dos conteúdos das conversas do *chatbot*.

Palavras-Chave: *Chatbots*, sistemas multi-agente, Q&A

1 Introdução

Atualmente, os simuladores de conversas naturais feitos por humanos são um tipo de *software* que está cada vez mais comum no cotidiano das pessoas por conta de sua vasta aplicabilidade no mercado e facilidade de manuseio pelo usuário. Esses sistemas de diálogo são comumente conhecidos por *chatbots* e, apesar de contarem com diversas terminologias para sua classificação, podem ser agrupados em três grupos majoritários[3][4]:

- **Sociais:** cujo objetivo consiste em propiciar interação e entretenimento ao usuário. Sendo assim, esses *chatbots* são projetados para manter conversas e interagir de diversas formas com os indivíduos por períodos de longa duração. São classificados como sociais os programas cujo funcionamento é semelhante ao “Robô Ed” e a “XiaoIce”[8].
- **Orientados a tarefas:** são programados para auxiliar o usuário, da maneira mais eficiente possível, na resolução de uma tarefa específica. Dessa forma, os diálogos seguem uma estrutura definida e têm duração limitada. Esse tipo de *chatbot* é muito utilizado em serviços de atendimento online, como quando há um assistente virtual acompanhando o usuário durante, por exemplo, uma compra de ingressos de cinema.

- **Q&A:** são majoritariamente aplicados para sanar dúvidas recorrentes dos usuários através de respostas elaboradas autonomamente a partir de coletâneas de textos fornecidos a priori ou da web. São encontrados com frequência nas redes sociais de lojas e empresas que recebem muitas perguntas de vários clientes. Eles são muito úteis, pois além de automatizar o serviço, ainda conseguem funcionar vinte e quatro horas por dia, sete dias por semana, sem queda de produtividade ou qualidade do serviço. Esse tipo de *chatbot* é amplamente utilizado para facilitar a consulta do usuário às *FAQs* da empresa[5], disponibilizando respostas de forma rápida e humanizada a perguntas muito frequentes como “Qual o horário de funcionamento da loja?”.

Como nosso objetivo é criar um *chatbot* informativo, o modelo Q&A é mais apropriado. A temática escolhida foi a região da Amazônia Azul, uma vez que a Organização das Nações Unidas (ONU) declarou que de 2021 a 2030 como a Década da Ciência Oceânica para o Desenvolvimento Sustentável e, sabendo que o Brasil é dono de uma enorme costa com 7.367 km[2], tão grande que é chamada de Amazônia Azul para chamar a atenção às imensas riquezas da área oceânica sob jurisdição brasileira, equivalentes a 3.539.919 km² [7].

Dessa forma, o objetivo deste trabalho de Iniciação Científica será desenvolver um *chatbot* multi-agentes para que este promova a disseminação de informações e conhecimentos para o público geral acerca dessa área brasileira.

2 Metodologia

A arquitetura escolhida é a de multi-agentes, usando ferramentas disponíveis no *Watson Assistant*, devido à facilidade de criação oferecida por estas ferramentas, permitindo não somente um primeiro contato com *chatbots* mas também desenvolvendo conhecimentos e habilidades do aluno nesta importante área de pesquisa.

Dessa forma, será criado um *bot* de *interface*, chamado de *Agent Bot* que será capaz de reconhecer a intenção do usuário e em seguida encaminhar para o *bot* especialista apropriado para tratamento e geração de respostas. Um *bot* especialista é aquele treinado para responder perguntas sobre um determinado tema relacionado a ele, como por exemplo Clima ou Biodiversidade.

Cada um desses *bots* especialistas é equivalente a uma *skill* dentro do *Watson Assistant*, que seria um domínio do *chatbot* como um todo. Essa arquitetura ainda é eficiente se o usuário quiser transitar entre os diferentes *bots*; isso seria plausível no nosso estudo por exemplo ao perguntar uma informação geral e depois procurar detalhes sobre essas informações nos *bots* especialistas. Na arquitetura multi-agentes, cada agente é um *bot* especialista. A agregação de tópicos de especialidade fica, assim, mais fácil, permitindo uma construção em partes do *chatbot*.

Um esquema dessa estrutura pode ser visto na Figura 1.

A dinâmica do processo é então definida em passos:

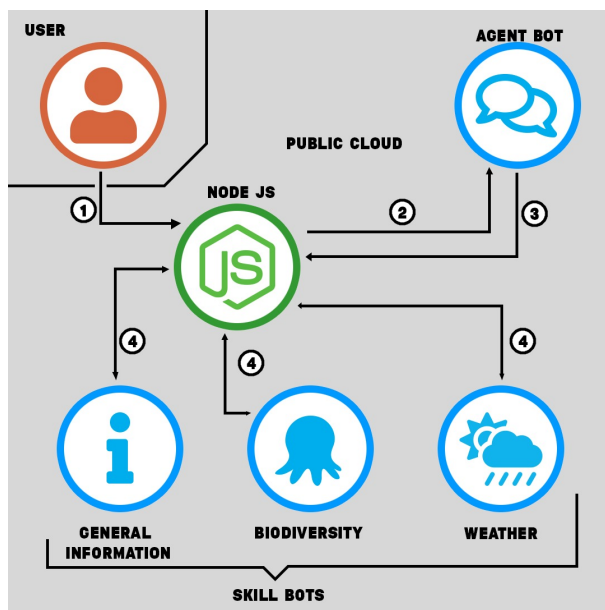


Figura 1: Esquema da arquitetura multi-agentes para o *chatbot*.

1. O usuário acessa a aplicação e envia uma mensagem, constituída por uma pergunta.
2. O aplicativo Node.js redireciona a mensagem para o *Agent Bot*.
3. O *Agent Bot* determina a intenção do usuário e repassa para aquele *Skill Bot* responsável por diálogos na intenção detectada.
4. A conversa do usuário então se passa com esse *Skill Bot*, especialista em um determinado tema no contexto da Amazônia Azul. Ao término da conversa, um novo ciclo se inicia.

A referência [1] dá mais informações básicas sobre o processo, a arquitetura e as ferramentas a serem utilizadas. Um modelo de referência para a pesquisa aqui desenvolvida está detalhada em [6], sobre um outro tema de interesse (Covid-19).

3 Relevância e Potenciais Aplicações

As ideias principais deste projeto são a de propiciar uma nova abordagem para algoritmos de agentes conversacionais, expandindo a capacidade de resposta sobre assuntos específicos. Dessa forma, visa-se contribuir com o desenvolvimento de inovações em *chatbots* com uma real competência em interagir de forma proveitosa com seus usuários. Esses *chatbots* são importantes intermediários para

dar assistência adequada às dúvidas e questões dos usuários. Existe a expectativa que este trabalho propicie publicações em conferências ou periódicos da área, além de compartilhar progressos e conhecimento com os demais membros e pesquisadores vinculados ao C²D.

4 Materiais e Infraestrutura

Esse trabalho será baseado na linguagem de programação *Python* e suas principais bibliotecas de machine learning, bem como na ferramenta *Watson Assistant*, técnicas de *Web Scrapping* e uso de *APIs* pertinentes para gerar as informações necessárias para o projeto.

Vale ressaltar que estas ferramentas são de acesso livre e não incorrem em custos adicionais de projeto. Pretende-se usar máquinas e servidores do C²D para as etapas de projeto, pois é sabido que a busca de hiperparâmetros adequados para o sistema Q&A exigem alto poder computacional. Se necessário, na etapa final do projeto será usada uma infraestrutura de nuvem. O bolsista deverá trabalhar no C²D e participar de atividades com os demais membros do centro, com o objetivo de fomentar novas colaborações.

5 Atividades e Cronograma

Este projeto foi concebido para ser realizado no período compreendido por um ano. As atividades compreendidas neste período serão listadas abaixo e o cronograma se encontra na tabela 1:

Etapa 1 : revisão bibliográfica, com estudo da literatura de *chatbots* Q&A.

Etapa 2 : definição da arquitetura a ser construída.

Etapa 3 : construção e programação do modelo.

Etapa 4 : testes preliminares com base de dados existentes e melhoria da arquitetura.

Etapa 5 : Coleta de dados e criação da base de dados sobre a Amazônia Azul.

Etapa 6 : Testes exaustivos.

Etapa 7 : Organização e documentação do projeto para uso futuro.

Etapa 8 : Escrita de artigo para divulgação científica.

Tabela 1: Cronograma

Atividades	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês 10	Mês 11	Mês 12
Etapa 1												
Etapa 2												
Etapa 3												
Etapa 4												
Etapa 5												
Etapa 6												
Etapa 7												
Etapa 8												

Referências

- [1] Watson assistant multi bot agent. <https://github.com/IBM/watson-assistant-multi-bot-agent>. Acessado em: 09/01/2021.
- [2] Belmiro M. Castro, Frederico P. Brandini, Marcelo Dottori, and João F. Fortes. A amazônia azul: recursos e preservação. *Revista USP*, (113):7, June 2017.
- [3] Asbjørn Følstad, Marita Skjuve, and Petter Bae Brandtzaeg. *Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design*. Springer International Publishing, 2019.
- [4] Jianfeng Gao, Michel Galley, and Lihong Li. Neural approaches to conversational AI. *Foundations and Trends® in Information Retrieval*, 13(2-3):127–298, 2019.
- [5] Bhavika R. Ranoliya, Nidhi Raghuwanshi, and Sanjay Singh. Chatbot for university related FAQs. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, September 2017.
- [6] R. K. W. Tsuzuki and C. N. Asai. Desenvolvimento de um chatbot informacional sobre a covid-19. page 69, 2020.
- [7] Marianne Wiesebron. Amazonia azul: Pensando a defesa do território marítimo brasileiro. *Rev. Brasileira de Estratégia e Relações Internacionais*, 2(3):107–131, 2013.
- [8] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93, 2020.



Júpiter - Sistema de Gestão Acadêmica da Pró-Reitoria de Graduação

HISTÓRICO ESCOLAR COMPLETO

Quantidade de reingressos: 0

Unidade: Escola Politécnica
Aluno: **11261531/1 - Enzo Bustos da Silva**
Ingresso: Vestibular 2 Lista - 01/02/2019
Curso: 3032/3170 - Engenharia
Habilitação: Engenharia Elétrica - Ênfase em Computação

Dados pessoais

Data de nascimento: 16/05/2000
Naturalidade: Não informado
Cédula de identidade: RG 552555526 São Paulo
Nacionalidade: Não informado

Forma de ingresso : Processo seletivo - Vestibular
Data de ingresso : Fev/2019
Classificação na carreira: 247

Endereço da Unidade:
Professor Luciano Gualberto 380 Travessa 3
CEP: 05508-010 São Paulo-SP



Júpiter - Sistema de Gestão Acadêmica da Pró-Reitoria de Graduação

HISTÓRICO ESCOLAR COMPLETO

Quantidade de reingressos: 0

Unidade: Escola Politécnica
Aluno: 11261531/1 - Enzo Bustos da Silva
Ingresso: Vestibular 2 Lista - 01/02/2019
Curso: 3032/3170 - Engenharia
Habilitação: Engenharia Elétrica - Ênfase em Computação

Sigla	Nome da Disciplina	Creditos			Atividade			FREQ	NOTA
		AU	TR	CH	CE	CP	ATPA		
2019 1º. Semestre									
4323101	Física I	3		45				100	7.3 A
MAC2166	Introdução à Computação	4		60				100	8.8 A
MAT2453	Cálculo Diferencial e Integral I	6		90				95	8.4 A
MAT3457	Álgebra Linear I	4		60				100	6.3 A
PCC3100	Representação Gráfica para Projeto	3	1	75				100	7.1 A
PQI3102	Introdução à Engenharia Química	3	1	75				90	7.5 A
QFL2129	Química Inorgânica	4		60				100	8.0 A
2019 2º. Semestre									
4323102	Física II	2		30				90	6.3 A
MAT2454	Cálculo Diferencial e Integral II	4		60				100	7.1 A
MAT3458	Álgebra Linear II	4		60				98	6.0 A
PME3100	Mecânica I	6		90				100	6.5 A
2020 1º. Semestre									
0303200	Probabilidade	2		30				100	6.6 A
4323201	Física Experimental A	2		30				100	9.2 A
4323203	Física III	4		60				80	7.5 A
MAT2455	Cálculo Diferencial e Integral III	4		60				100	7.5 A
PCS3111	Laboratório de Programação Orientada a Objetos para Engenharia Elétrica	3		45				100	8.8 A
PCS3115	Sistemas Digitais I								T
PEF3208	Fundamentos de Mecânica das Estruturas								T
PSI3211	Circuitos Elétricos I								T
PSI3212	Laboratório de Circuitos Elétricos								T
2020 2º. Semestre									
4323202	Física Experimental B	2		30				100	7.2 A
4323204	Física IV	4		60				80	8.3 A
MAT2456	Cálculo Diferencial e Integral IV	4		60				100	7.2 A
PCS3110	Algoritmos e Estruturas de Dados para Engenharia Elétrica	4		60				75	7.9 A
PMR3508	Aprendizado de Máquina e Reconhecimento de Padrões	4		60				97	6.9 A
PRO3200	Estatística	4		60				100	5.9 A

Créditos obtidos: aula :80 trabalho : 2 total: 82

Créditos matriculados: aula :80 trabalho : 2 total: 82

Carga Horária Total: 1260 h

Média ponderada com reprovações: 7.4

Média ponderada: 7.4

Os dados abaixo foram processados em : 05/08/2020

Atenção: Valores para seleção das optativas : 3.5054

Média normalizada : 5.8643

Média ponderada de seu curso : 6.0750



Júpiter - Sistema de Gestão Acadêmica da Pró-Reitoria de Graduação

HISTÓRICO ESCOLAR COMPLETO

Quantidade de reingressos: 0

Unidade: Escola Politécnica
Aluno: **11261531/1 - Enzo Bustos da Silva**
Ingresso: Vestibular 2 Lista - 01/02/2019
Curso: 3032/3170 - Engenharia
Habilitação: Engenharia Elétrica - Ênfase em Computação

Totais de Créditos Acumulados por Tipo de Obrigatoriedade:

	Aula	Trabalho
Obrigatórias	76	2
Opt.Eletivas	0	0
Opt.Livres	4	0

- As notas variarão de zero a dez, podendo ser aproximadas até a primeira casa decimal (Regimento Geral, artigo 83).
- Será aprovado, com direito aos créditos correspondentes, o aluno que obtiver nota final igual ou superior a cinco e tenha, no mínimo, setenta por cento de frequência na disciplina (Regimento Geral, artigo 84).
- Um crédito aula corresponde a 15 horas de carga horária semestral, e o trabalho a trinta.
- Este Histórico Escolar é completo, mostrando eventuais reprovações e/ou trancamentos de matrícula.

Legenda:

AU = Crédito Aula	TR = Crédito Trabalho	CH = Carga Horária
MA = Matriculado	T = Trancado	CE = Carga Horária de Estágio
A = Aprovado	RN = Reprovado por Nota	AE = Aproveitamento de Estudo
RA = Reprovado por Nota e Frequência	DI = Dispensado	RF = Reprovado por Frequência
CP = Carga Horária práticas componentes curriculares	DS = Dispensado por prova de suficiência (Res. CoG 4844/01)	
ATPA = Atividades Teórico-Práticas de Aprofundamento	P = Pendente	
I = Inscrito	IR = Inscrição Reservada	IT = Inscrição em Turma Lotada
IP = Inscrição em optativa Preterida		IL = Inscrição em Lista de Espera

Média Ponderada das disciplinas em que o aluno obteve aprovação (não inclui notas de AE e DS).

Observações

Curso reconhecido pela Portaria CEE/GP nº 186 de 06/08/2020, D.O.E. de 07/08/2020.



Enzo Bustos da Silva

Endereço para acessar este CV: <http://lattes.cnpq.br/2213125912686237>

ID Lattes: **2213125912686237**

Última atualização do currículo em 06/01/2021

Possui ensino-medio-segundo-graupelo Colégio Agostiniano Mendel(2017). Tem experiência na área de Engenharia Elétrica, com ênfase em Computação. **(Texto gerado automaticamente pela aplicação CVLattes)**

Identificação

Nome

Enzo Bustos da Silva 

Nome em citações bibliográficas

SILVA, E. B.

Lattes iD

 <http://lattes.cnpq.br/2213125912686237>

Endereço

Formação acadêmica/titulação

2019

Graduação em andamento em Engenharia Elétrica - Ênfase em Computação.
Universidade de São Paulo, USP, Brasil.

2015 - 2017

Ensino Médio (2º grau).
Colégio Agostiniano Mendel, CAM, Brasil.

Áreas de atuação

1.

Grande área: Engenharias / Área: Engenharia Elétrica / Subárea: Computação.

Idiomas

Inglês

Compreende Bem, Fala Razoavelmente, Lê Bem, Escreve Razoavelmente.

Alemão

Compreende Pouco, Fala Pouco, Lê Pouco, Escreve Pouco.

Produções

Produção bibliográfica