

# Processamento de Linguagem Natural Aplicada na Construção de uma Agente Conversacional por meio do IBM Watson Assistant

Jeanluca Fernandes Pereira<sup>1</sup>, Merisandra Côrtes de Mattos Garcia<sup>1</sup>

Grupo de Pesquisa em Inteligência Artificial Aplicada – Universidade do Extremo Sul Catarinense (UNESC) - <sup>1</sup>Curso de Ciência da Computação

Criciúma – SC - Brasil

jeanlucafp@unesc.net, mem@unesc.net

**Abstract.** *Natural language processing systems combine techniques such as machine learning to improve your assertiveness against your databases. This research makes use of the IBM Watson Assistant for the development of a conversational agent for the first level service relevant to UNIEDU. After analyzing the agent through queries carried out by users, the usage data on the platform were collected and the performance of the tool was evaluated, reaching in the final model an accuracy of 80% of correct answers in the messages exchanged, the tool showed a satisfactory performance in relation to its database.*

**Resumo.** *Sistemas de processamento de linguagem natural combinam técnicas como aprendizado de máquina para melhorar sua assertividade de acordo com suas bases de dados. Esta pesquisa faz uso do IBM Watson Assistant para o desenvolvimento de um agente conversacional para o atendimento de primeiro nível pertinentes ao UNIEDU. Após a análise do agente mediante as consultas realizadas pelos usuários, foram coletados os dados de uso na plataforma e avaliada a performance da ferramenta, atingindo-se no modelo final uma precisão de 80% de acertos nas mensagens trocadas, apresentando um desempenho satisfatório em relação a sua base de dados.*

## 1. Introdução

A inteligência artificial pode ser definida como o estudo de agentes que recebem estímulos do ambiente e executam determinada ação, para isso o agente apresenta algumas capacidades como, aprendizado de máquina para que possa se adaptar as novas condições; raciocínio automatizado para a partir de informações salvas, ter novas conclusões; representação do conhecimento para que se recorde do que aconteceu e processamento de linguagem natural para se comunicar com os humanos [Russel e Norvig 2013].

O processamento de linguagem natural consiste em utilizar técnicas computacionais com o propósito de compreender e produzir conteúdo em linguagem humana. Assim, busca-se automatizar a análise da estrutura da linguagem e desenvolver tecnologias como tradução automática, reconhecimento de texto e voz, os quais são usados e adaptados em diferentes aplicações como, por exemplo, a Siri da Apple e o Google Tradutor [Hirschberg e Manning 2015].

Dentre os fatores que possibilitaram o desenvolvimento dessas aplicações tem-se o aumento da capacidade de processamento dos computadores, o *big data*, os algoritmos de aprendizado de máquina e uma vasta gama de estruturas de linguagens e interações humanas nas redes sociais [Hirschberg e Manning 2015].

A indústria de agentes conversacionais encontra-se em desenvolvimento, esta tecnologia interativa combinada com a inteligência artificial tem ocupado os chats online, tendo-se disponível diferentes serviços arquitetados em nuvem para o desenvolvimento e aprimoramento dos agentes conversacionais, são alguns deles, o IBM Watson Assistant, o Bot Framework da Microsoft, o AWS Lambda.

Serviços de compreensão de linguagem, como o IBM Watson Assistant, operam classificando uma frase de entrada em intenções e extraíndo entidades dela, essas intenções representam o objetivo da frase, as entidades consistem nos termos ou objetos que serão usados no objetivo [Etinger, 2018]. IBM Watson é uma plataforma de serviços cognitivos da IBM para negócios. A cognição consiste no processo que a mente humana utiliza para adquirir conhecimento a partir de informações recebidas.

A implementação de um agente conversacional pode ser dividida nas fases de configuração e de uso. A fase de configuração acontece quando o usuário acessa o agente e fornece algumas informações básicas para a configuração do ambiente. Ao término desta etapa ocorre a fase de uso que permite aos usuários dialogarem com o agente a fim de obter suas soluções, finalizando-se essas duas etapas pode ser realizada mais uma fase em que o agente recebe um feedback do seu trabalho, permitindo assim melhorias [Cyril 2018]. Em casos mais básicos, o sistema apenas reconhece as palavras-chave pesquisando em seu banco de dados, enquanto operações mais sofisticadas necessitam de processamento de linguagem natural, o que permite aos agentes conversacionais entenderem as intenções e as demandas mais complexas vindas do usuário, podendo responder melhor as perguntas frequentes [Molnar 2018].

Segundo Liddy (2001), no desenvolvimento de um agente conversacional deve ser usada a abordagem de etapas de linguagem que compreende as análises fonológica, léxica, sintática, semântica e pragmática. A análise fonológica interpreta os sons da fala por meio das palavras; a análise léxica interpreta o significado individual das palavras; a análise sintática analisa as palavras em uma frase, a fim de descobrir a estrutura gramatical, a sintaxe transmite o significado na maioria das linguagens, pois dependendo da ordem ou contexto a palavra tem significados diferentes; a análise semântica determina os possíveis significados na sentença; a análise pragmática identifica o discurso do ponto de vista social e cultural [Kumar 2011].

Os agentes conversacionais, também denominados de assistentes, *chatterbot*, *chatbot*, têm sido empregados por organizações e governos em websites e diversas aplicações, plataformas de mensagem instantânea, como WhatsApp, Telegram e

Messenger promovendo produtos, ideias e serviços [Szuts 2018]. A disponibilização dos serviços desses agentes em plataformas web e mobile proporcionam facilidades como no acesso, na inclusão digital, não tem a necessidade de pacotes de instalação, dependências externas, entre outras [Rahman 2014].

Aplicados em diferentes áreas, vários estudos como o de Verleger et al (2018), consistiram no desenvolvimento de agentes conversacionais voltados ao ensino superior. Também, diferentes grupos de ensino superior do país empregam agentes conversacionais para dialogarem com os estudantes sobre carreira e interesses profissionais a fim de indicarem profissões compatíveis com o seu perfil, para auxiliarem no relacionamento entre o estudante e a instituição, e para recursos como *e-learning*, biblioteca virtual, entre outros. Luckin et al (2016) afirma que os agentes conversacionais proporcionam aos estudantes liberdade para realizarem as perguntas, sem que se sintam constrangidos.

No estado de Santa Catarina tem-se um programa de bolsas universitárias denominado UNIEDU que possibilita a inclusão e permanência de estudantes da educação superior, contemplando-os com bolsas de estudo, pesquisa ou extensão. Os estudantes que concorrem a estas bolsas devem atender a requisitos estabelecidos no programa, os quais geram vários questionamentos ao setor institucional responsável. Considerando-se o exposto, esta pesquisa consistiu em disponibilizar a prototipação de um assistente virtual que atenda por meio de linguagem natural dúvidas referentes ao UNIEDU; descrever as técnicas de processamento de linguagem natural; conceber uma plataforma conversacional; empregar as técnicas de processamento de linguagem natural por meio do IBM Watson na construção de um agente conversacional; representar o conhecimento referente a dúvidas frequentes do UNIEDU.

Conforme Gartners (2016), no relatório *Hype Cycle for Emerging Technologies* em 2017, os avanços tecnológicos apontam para o crescimento de agentes conversacionais no período de cinco a dez anos, os quais serão responsáveis por mais de 85% dos centros de atendimento ao cliente, com agentes capazes de realizar atendimento personalizado com reconhecimento de fala, texto e análise comportamental.

No cenário desta pesquisa, que compreende escolas e universidades, observa-se o aumento nos investimentos em softwares inteligentes que são capazes de oferecer serviços de auxílio, ensino e aprendizagem mais personalizada e inclusiva [Luckin et al 2016].

## **2. Trabalhos relacionados**

Na literatura, existem trabalhos internacionais e nacionais relacionados a sistemas de processamento de linguagem natural aplicado na construção de um agente conversacional, dentre estes, pode-se citar a pesquisa publicada no ano de 2018 na IEEE *Frontiers in Education Conference*, com o objetivo de desenvolver uma interface inteligente de conversação para guiar os alunos em um curso introdutório de programação. Essa interface foi iniciada com uma base de conhecimento limitada na intenção de ser populada baseando-se nas interações com os alunos. O programa foi alimentado com alguns FAQs de fórum e wikis e durante as interações com os alunos caso o programa não tivesse uma resposta o aluno poderia fornecê-la e avaliá-la e uma notificação era enviada para o professor responsável pela turma, o *backend* do programa

utiliza o Microsoft QnA Maker para lidar com os processos de inteligência artificial e todas as outras interações são tratadas usando o PHP. O agente conversacional foi usado 48 vezes tendo em seus resultados uma média de 33.3% de respostas não encontradas, 16.6% de respostas respondidas com clareza e 18.8% de respostas que foram consideradas inúteis. Os autores concluíram que os agentes têm potencial de encorajar os estudantes a serem mais participativos e questionadores, além de oferecerem oportunidades de contextualização se comparado a pesquisas genéricas na internet [Verleger et al 2018].

DINA foi um trabalho publicado no ano de 2018 na *International Seminar on Application for Technology of Information and Communication*, os autores descrevem e propõem o desenvolvimento de um agente conversacional que ajuda estudantes em relação a processos administrativos na universidade. Segundo os autores a maioria das universidades utilizam sites corporativos para oferecer esses tipos de serviços, além de vários canais de comunicação como chat e telefone, sendo necessário tempo para responder a alta demanda. A base de conhecimento do agente foi construída utilizando recursos existentes de informação como FAQs e registro de sistema e o conhecimento foi representado na forma de regras. O fluxo de diálogo foi realizado por meio do DialogFlow que é uma plataforma de desenvolvimento de agente conversacional baseada em linguagem natural, sendo necessário criar entidades para declarar os vários tipos de perguntas [Santoso et al, 2018].

Agente conversacional Jill Watson, desenvolvido pela *Georgia Tech*, seu objetivo era se passar como um professor online durante o curso *Masters of Science in Computer Science*, segundo o autor os alunos não sabiam que o agente conversacional não era uma pessoa real, Jill Watson foi construído utilizando a plataforma do IBM Watson Assistant que era responsável pelas repostas aos questionamentos realizados pelos alunos no fórum do curso [Lipko 2016].

Diversas universidades estão começando a aderir ao uso de agentes conversacionais utilizando o IBM Watson Assistant, dentre elas a Universidade de *Deakin*, localizada em Melbourne, Austrália. Essa universidade criou um programa de atração e experiência estudantil chamado *Live the Future*, que em 2016 chamou a atenção da mídia internacional e gerou um aumento em suas matrículas atingindo a marca de 54.000 acadêmicos [Lacity et al 2017].

Silva e Matos (2018) propõem um estudo prático publicado pelo departamento de computação do Centro Universitário Adventista de São Paulo (UNASP) sobre implantação completa de um sistema inteligente baseado em um *chatbot* para atendimento em primeiro nível de usuários de uma organização corporativa. Inicialmente realizaram o levantamento bibliográfico de materiais e trabalhos correlatos voltados a tratamento de linguagem natural, a compreensão da plataforma IBM Watson Assistant para o desenvolvimento do *chatbot*, a representação da base de conhecimento e a validação dos resultados junto a corporação empregada no estudo de caso. Com os resultados dos testes iniciais concluíram que não era necessária uma base de dados muito grande, em função do algoritmo cognitivo do Watson generalizar relações.

### 3. Materiais e métodos

O desenvolvimento de um agente conversacional para esclarecimento de dúvidas referente ao UNIEDU consiste em uma pesquisa de natureza secundária, aplicada e de base tecnológica, segundo Wazlawick (2021) a pesquisa secundária busca entender dados de uma determinada área, considerando a sua evolução e o atual momento. A pesquisa quanto aos objetivos é descritiva, pois descreve os fatos como eles são e no que se refere aos procedimentos técnicos é uma pesquisa bibliográfica e observacional, visto que não ocorre a intervenção sistemática do pesquisador, mas sim a observação e conclusões a partir dos estudos realizados.

Nesta pesquisa é construído o agente conversacional, empregando as ferramentas de processamento de linguagem natural fornecidas pelo IBM Watson, utilizando-se do conhecimento adquirido referente ao UNIEDU.

Utilizando a técnica de processamento de linguagem natural do IBM Watson Assistant, é possível entender as entradas de texto informadas pelo usuário e fornecer uma resposta satisfatória. A arquitetura geral da pesquisa (figura 1) compreende uma interface de mensagens acessível, disponível em qualquer dispositivo com acesso à internet, realizando a conexão com os servidores responsáveis pelo algoritmo de processamento de linguagem natural e armazenamento de dados. Os dados referentes ao domínio de aplicação são empregados para o treinamento dos modelos pela plataforma IBM Watson Assistant, para então esta ser capaz de responder as perguntas enviadas por um aplicativo de mensagens, as quais são recebidas por meio de uma *Application Programming Interface (API)*.

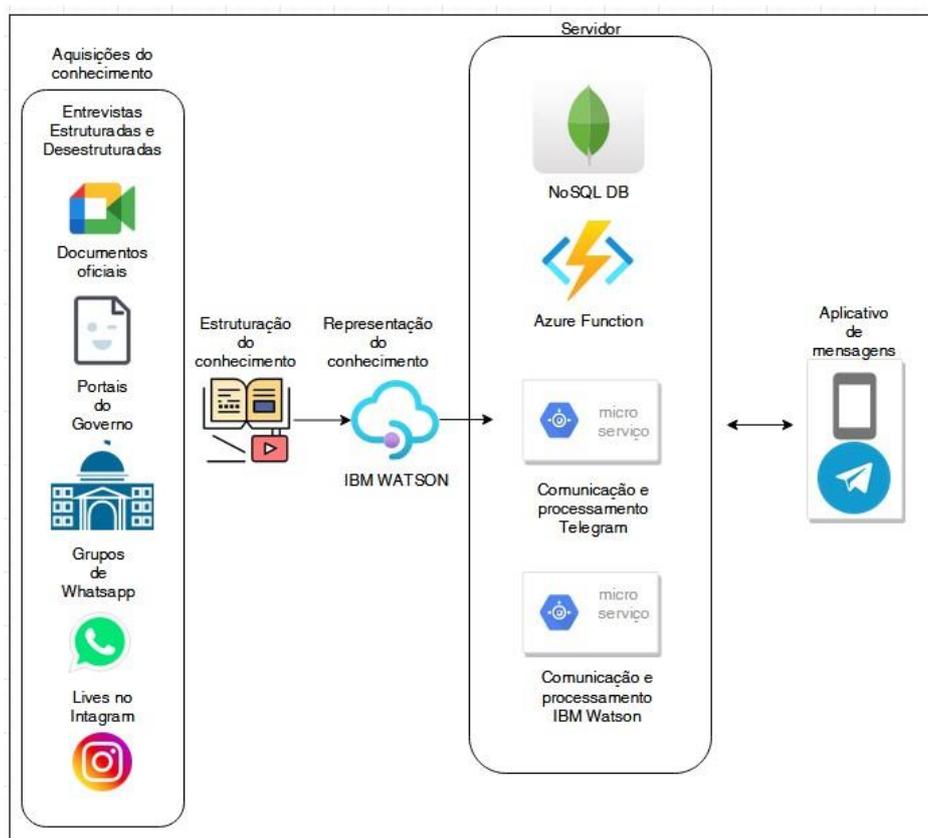


Figura 1. Arquitetura geral da pesquisa

O desenvolvimento desta pesquisa compreendeu as etapas de aquisição do conhecimento, em que foram obtidos os conhecimentos necessários para criação dos fluxos de conversas; desenvolvimento do protótipo de agente conversacional por meio do IBM Watson que compreendeu a criação do fluxo de conversão, empregabilidade do processamento de linguagem natural do IBM Watson Assistant, construção dos microsserviços responsáveis pela comunicação e interpretação dos serviços do IBM Watson, assim como a concepção do banco de dados e dos servidores utilizados pelos microsserviços. Finalmente, realizou-se análise do agente conversacional mediante as consultas realizadas pelos usuários.

### **3.1. UNIEDU**

Programa de Bolsas Universitárias de Santa Catarina (UNIEDU) consiste em um programa de bolsas de estudo destinado a estudantes de graduação e pós-graduação que residem no estado de Santa Catarina. O programa foi desenvolvido para aumentar o acesso ao ensino superior, oferecendo oportunidades a aqueles que não possuem condições financeiras para continuar os estudos. As bolsas de estudos que são oferecidas variam de 25% a 100% de acordo com o índice de carência reportado pelo estudante durante o processo seletivo (Santa Catarina 2021).

Diferente de outros programas do governo, o UNIEDU é renovado semestralmente, para que o beneficiado permaneça no programa é necessário que o estudante participe novamente do processo de seleção.

O programa oferece três tipos de bolsa, podendo ser de estudo, de pesquisa ou de extensão. Os estudantes que são contemplados com a bolsa de estudo devem realizar 20 horas semestrais de programas e projetos sociais que são elaborados pela instituição de ensino. A bolsa de pesquisa, fornece ao estudante auxílio financeiro no valor de um salário mínimo vigente, para ser um beneficiário é necessário realizar uma pesquisa de iniciação científica com o apoio de um professor orientador. A bolsa de extensão também fornece auxílio financeiro no valor de um salário mínimo, a fim de o estudante realize atividades extensionistas, sob a orientação de um professor, realizando o compartilhamento mútuo de conhecimentos com a comunidade [Santa Catarina, 2021]

#### **3.1.1 Aquisição do conhecimento**

A aquisição do conhecimento necessária para a construção do protótipo foi realizada por meio de entrevistas desestruturadas junto a especialista do domínio de aplicação que é a Gerente de Atenção ao Estudante da UNESC. A partir desta entrevista, definiu-se o nicho de conhecimento a ser contemplado pela aplicação que consistiu nos questionamentos frequentemente realizados pelos estudantes, bem como as informações gerais sobre o UNIEDU, como no que consiste, formas de participação, benefícios, entre outros.

A busca pelo conhecimento foi realizada nos canais de atendimento do UNIEDU na UNESC, nas entrevistas estruturadas com a especialista no domínio de aplicação e os atendentes da Gerência de Atenção aos Estudantes da UNESC, nas lives no Instagram

sobre o UNIEDU, nos grupos de WhatsApp e documentos oficiais do UNIEDU. Assim, as informações foram obtidas por meio de canais oficiais da UNESCO, do Governo do Estado de Santa Catarina e do UNIEDU.

### **3.1.2 Organização e representação do conhecimento adquirido**

No treinamento de máquina empregou-se o conjunto de dados sobre o UNIEDU, constituído de 46 intenções de atos, sendo que 32 destas são específicas para o contexto do UNIEDU, as demais referem-se as ações importantes para um agente conversacional, como por exemplo, compreender os feedbacks positivos e negativos, as confirmações, as negações, as saudações, as suas capacidades e os seus objetivos. Cada uma destas intenções de atos possui uma média de 10 exemplos de sentenças.

O conhecimento está representado em nós de domínio, para cada nó tem-se um conhecimento específico que é fornecido como resposta a uma determinada solicitação. No apêndice A, são apresentadas as 44 intenções de atos, sendo as 32 primeiras referentes ao UNIEDU e as demais voltadas a um agente conversacional independente do domínio de aplicação. Esta base de conhecimento está salva em arquivos do tipo.csv (separado por ponto e vírgula), sendo que cada intenção é um arquivo.

Na figura 2 tem-se a representação dos exemplos de sentenças sobre o UNIEDU para o nó de intenção *#ComprovanteBemFamiliarCasaGanhada*, que considera casos de herança e de doação realizada para os estudantes ou para o seu grupo familiar.

← | #ComprovanteBemFamiliarCasaGanhada

#ComprovanteBemFamiliarCasaGanhada

Name your intent to match a customer's question or goal

Description (optional)

Add a description to this intent

User example

Type a user example here

Add unique examples of what the user might say. (Pro tip: Add at least 5 unique examples to help Watson understand)

Add example

User examples (7) ↑

- ganhei um terreno de herança
- ganhei uma casa de herança
- meus pais ganharam um casa
- meus pais ganharam um terreno
- meus pais ganharam uma casa
- Meus pais ganharam uma casa/terreno/apartamento de herança
- Meus pais ganharam uma casa/terreno/apartamento de herança. Como faço

Figura 2. Exemplos para intenção de #ComprovanteBemFamiliarCasaGanhada

Na figura 3 são apresentados os exemplos para o nó de intenção voltado ao funcionamento do agente conversacional #General\_Agent\_To\_Agent.

← | #General\_Connect\_to\_Agent

User example

Type a user example here

Add unique examples of what the user might say. (Pro tip: Add at least 5 unique examples to help Watson understand)

Add example

- Isso é muito frustrante, eu quero falar com uma pessoa de verdade real.
- me passe para um atendente humano
- Ok deixe-me tentar outra vez. Eu estou tentando falar com um ser humano para resolver diversos problemas faz tempão e dado que o meu telefone t
- Pode me provar que é um representante?
- Por favor, você pode entrar em contato comigo com um agente humano?
- Posso ter informações de contato para falar com um agente ao vivo.
- Preciso contatar a um atendente
- Quero falar com alguém de outra área.
- quero falar com um humano
- Você pode me conectar com uma pessoa real?
- Você pode por favor me transferir para alguém que possa me ajudar, eu mandei um e-mail para você e você não leu direito e simplesmente saiu.

Figura 3. Exemplos para intenção de #General\_Connect\_to\_Agent

### 3.2. Processamento de linguagem natural com IBM Watson Assistant

O algoritmo de processamento de linguagem natural para o agente conversacional desenvolvido foi por meio do IBM Watson Assistant, versão 2, disponível na IBM Cloud<sup>1</sup>, que foi projetado para ajudar a escalar o uso da inteligência artificial nas organizações. As capacidades do Assistant foram desenhadas para oportunizar a automação da inteligência artificial, oferecer uma maior precisão em processamento de linguagem natural e ampliar a confiança nos resultados derivados de predições da inteligência artificial [IBM 2021].

O IBM Watson Assistant emprega técnicas de aprendizado de máquina e aprendizado profundo para responder as perguntas do usuário final com maior precisão e utilizando conjuntos de dados relativamente pequenos. A inteligência artificial no núcleo do Watson Assistant é projetada para identificar corretamente as permutações de intenções nas interações do mundo real. Dentre as habilidades do IBM Watson Assistant incluem-se o Discovery, Reading *Comprehension*, FAQ *Extraction* e o Modelo de classificação de intenção do IBM Watson Assistant.

<sup>1</sup> Disponível em <https://us-south.assistant.watson.cloud.ibm.com/crn:v1:bluemix:public:conversation:us-south:a%2Fef199eb900e848efb24c858a382c3e46:c50aa5e1-86af-4bfc-b101-4b2c661e73b4::/home>

O IBM Watson Discovery inclui o suporte a vários idiomas, como também continuamente, a IBM adiciona suporte a outras linguagens para auxiliar as organizações a construir soluções de processamento de linguagem natural.

O recurso de compreensão de leitura, *Reading Comprehension*, retorna um fato específico ou uma resposta curta contida em um texto, para isso recupera vários parágrafos candidatos do conjunto de documentos fornecidos, procura uma resposta para a pergunta em questão e retorna as respostas correspondentes. Este recurso aplica a compreensão contextual para entender as consultas e aproveitar os modelos de linguagem para extrair respostas específicas do documento consultado, o usuário recebe uma pontuação de confiança que indica o quão confiante o sistema está em cada resposta.

O recurso FAQ *Extraction* é uma técnica de processamento de linguagem natural também do IBM *Research* que automatiza e extrai pares de perguntas de documentos de FAQ, tendo sido projetada para funcionar na habilidade de pesquisa do Watson Assistant, que busca respostas para as perguntas dos usuários finais na documentação. Essa funcionalidade aumenta a possibilidade dos usuários finais encontrarem as respostas de que precisam ao interagir com agentes virtuais acionados por inteligência artificial.

O Modelo de classificação de intenção do IBM Watson Assistant é responsável por compreender a intenção do usuário final, permitindo aos administradores do sistema treinarem mais rapidamente o modelo, fornecendo resultados mais precisos e utilizando menos dados em comparação com outras plataformas. As intenções referem-se aos objetivos expressos na entrada de um usuário, ao reconhecê-los o serviço Watson Assistant pode escolher o fluxo de diálogo correto para a resposta e quais as intenções de negócios o agente conversacional deve manipular para os usuários, ensinando ao Watson sobre eles. A cada intenção deve-se fornecer pelo menos cinco exemplos de elocuições que são habitualmente empregadas pelos usuários para indicarem os seus objetivos. A tecnologia de inteligência artificial pode entender a intenção presente nas frases e suas infinitas variações de palavras e erros ortográficos. Mesmo para uma intenção mais básica, é necessário um processamento complexo de linguagem natural e técnicas de classificação [Loveys 2020].

No desenvolvimento da presente pesquisa empregaram-se as habilidades de Discovery, FAQ extraction e o modelo de classificação de intenção do IBM Watson Assistant.

Na versão 2 do Watson Assistant, emprega-se a técnica de AutoML que tenta vários algoritmos e combinações de recursos e parâmetros para encontrar os melhores resultados para um determinado conjunto de dados, sem a intervenção humana. O AutoML requer poder computacional e tempo de aprendizagem, sendo complementado com técnicas de meta-aprendizagem que aceleram e otimizam a detecção de intenções por meio de diferentes algoritmos de aprendizado de máquina que funcionam em vários conjuntos de dados [Karimkhany 2020].

O algoritmo de detecção de intenções na versão 2 do IBM Watson Assistant é mais preciso quando comparado ao de outras soluções comerciais e de código aberto, tendo-se 79% de precisão do modelo gerado por esta versão do IBM Watson e de 76,3% na versão anterior. Assim, o agente conversacional do IBM Watson Assistant

consegue responder mais frequentemente as solicitações do usuário, sem o envolvimento de um agente humano. Em dezembro de 2020, foi publicado um artigo técnico que comparou o desempenho técnico do IBM Watson Assistant com plataformas semelhantes como do Google, Microsoft e RASA, bem como BERT, um projeto de código aberto patrocinado pelo Google. De acordo com os resultados do benchmark, o Watson Assistant é 5,6 pontos percentuais mais preciso do que o Google Dialogflow e 14,7 pontos percentuais mais preciso do que o Microsoft LUIS [Haode Qi et al 2020].

### 3.2.1 Criação do fluxo de conversação

A criação do fluxo de conversação trata do encaminhamento da conversa conforme a intenção do usuário. A partir da aquisição de conhecimento do negócio tem-se a criação dos arquivos CSV, a realização de *upload* destes para o IBM Watson Assistant e a construção das intenções. A cada nova intenção tem-se o refinamento do modelo de inteligência artificial por meio do seu treinamento.

Para cada intenção foi criado um nó no fluxo que apresenta seus nós filhos, conforme a complexidade do diálogo. Dessa forma, constrói-se uma lista não ordenada de nós. Para as intenções de dúvidas frequentes foram criados dois nós, um que compreende a intenção e seu nó filho que responde de acordo.

Apesar do treinamento, ainda acontece do agente conversacional não entender algumas frases. A fim de resolver esse problema utilizou-se uma técnica recorrente em *chatbots* que é a de sugestões. Caso a porcentagem de certeza sob a intenção de uma sentença não seja satisfatória para o fornecimento de uma resposta correta, sugere-se quatro intenções de contexto similar ao que foi solicitado. Assim, mesmo que o *chatbot* não consiga entender o usuário, este consegue utilizá-lo por meio da navegação no fluxo de sugestões.

### 3.3 Servidor de aplicação

O servidor da aplicação consiste em três microsserviços que foram desenvolvidos na linguagem Javascript por meio da ferramenta Node.js em sua versão 14.15.1. Esta ferramenta é multiplataforma, de código aberto e executa códigos JavaScript no servidor, baseando-se no V8 que é um interpretador de JavaScript do Google. O ambiente de programação utilizado foi o Visual Studio Code e o sistema operacional Windows 10.

A arquitetura do servidor foi projetada para suportar diversas plataformas de mensagens existentes. Os microsserviços foram criados utilizando o paradigma de computação nativa em nuvem que é uma abordagem no desenvolvimento de software que emprega a computação em nuvem para criar e executar aplicativos escaláveis em ambientes modernos e dinâmicos, os quais são executados em *containers* Docker que são orquestrados pelo Kubernetes [Liu F et al. 2011].

Incluídos nesses microsserviços estão o serviço de comunicação e processamento com aplicativo de mensagem, serviço de comunicação e processamento com o IBM Watson, serviço de recebimento de mensagens do IBM Watson, assim como o banco de dados NoSQL MongoDB.

### 3.3.1 Microsserviço de comunicação e processamento com aplicativo de mensagem

A interface de comunicação com o usuário se dá por meio do Telegram, que é um aplicativo de troca de mensagens similar, disponível gratuitamente para Android, IOS, Windows, Linux e MAC. A escolha desse aplicativo deu-se em função da sua alta disponibilidade, fácil acesso, licença de uso gratuita e pelo acesso livre a sua API de comunicação. Dessa forma, o desenvolvedor pode criar seus projetos de forma independente na plataforma. A figura 4 mostra a interação da aplicação com o usuário-interface.

Este serviço é responsável pela comunicação direta com a *Application Programming Interface* (API) do Telegram, por meio da biblioteca *telegraf* na versão 3.17.3, disponibilizada pelo gerenciador de pacotes de bibliotecas *Node Package Manager* (NPM) na versão 6.14.8.

Essa aplicação controla e distingue cada usuário que inicia uma nova conversa, mantendo sua privacidade garantindo que suas mensagens sejam únicas. Após esse processo o microsserviço interage com o serviço de comunicação e processamento do IBM Watson, utilizando-se do protocolo de rede *Transmission Control Protocol* (TCP) para obter a resposta da entrada fornecida pelo usuário. Assim que a solicitação é processada esse serviço interpreta a resposta e envia para o usuário solicitante finalizando sua etapa no processo.

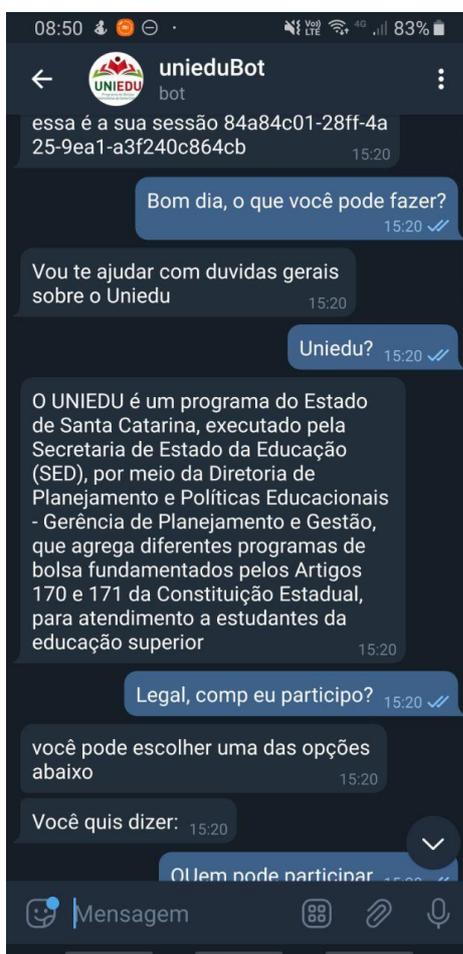


Figura 4. Interação da aplicação com o usuário-interface

### 3.3.2 Microserviço de comunicação e processamento com o IBM Watson

O microserviço de comunicação e processamento com o IBM Watson é responsável por receber uma entrada de texto, armazená-la no banco de dados MongoDB, enviá-la para o serviço do IBM Watson Assistant, interpretar a resposta recebida e devolvê-la para o solicitante.

A comunicação com o IBM Watson Assistant ocorre por meio da biblioteca *ibm-watson*, versão 5.2.1, disponível no gerenciador de pacotes do Node. A interpretação de resposta recebida pelo serviço da IBM foi realizada segundo a documentação do Assistant na versão 2<sup>2</sup>.

### 3.3.3 Banco de dados NoSQL MongoDB

O banco de dados desta pesquisa foi desenvolvido e estruturado no MongoDB na sua versão 3.6.1 e controlado pela interface MongoDB Compass versão 1.23. A escolha pelo MongoDB deu-se em função de ser um software de código aberto e da sua estrutura NoSQL, o que permite uma melhor distribuição e rapidez na organização e análise de grande quantidade de dados, sendo recomendado para projetos desse contexto.

A integração entre os serviços Node.js e o banco de dados MongoDB foi realizada por meio da biblioteca *mongodb*, versão 3.4.1, disponibilizada pelo gerenciador de pacotes de bibliotecas *Node Package Manager* (NPM) na versão 6.14.8.

### 3.3.4 Serviço de recebimento de mensagens do IBM Watson

Serviço *serverless* hospedado no portal Azure da Microsoft, essa escolha ocorreu pela necessidade de ter alta disponibilidade em nuvem, já que sua função é receber eventos enviados pelo IBM Watson Assistant e armazená-los no banco de dados NoSQL para realização de análises futuras. O serviço é gerenciado pela Microsoft, não tendo-se preocupações com o servidor, tais como queda de energia, problemas com hardware ou de segurança.

### 3.3.5 Análise do agente conversacional mediante as consultas realizadas

A presente pesquisa se caracterizou como um estudo observacional transversal, pois conforme Jung (2004) os dados da análise foram coletados em um intervalo de tempo curto realizando-se um recorte momentâneo do comportamento do agente conversacional mediante as consultas realizadas.

O agente conversacional foi disponibilizado em nuvem, sem possibilidade de identificação das pessoas que realizaram as consultas, na manhã de 05 de junho de 2021, momento em que se teve o maior pico de solicitação de consultas ao *chatbot* foram feitas as análises.

---

<sup>2</sup> Disponível em <https://cloud.ibm.com/apidocs/assistant/assistant-v2>

A coleta de dados foi realizada por meio da análise dos *logs* de conversa registrados pelo agente conversacional que consiste nos *logs* de conversas registradas entre o *chatbot* e o solicitante, com o objetivo de avaliar a eficiência dos diálogos, a partir desses dados foram analisadas oito métricas para análise de agentes conversacionais, as quais são: a quantidade total de usuários solicitantes, a média de abertura de solicitação por usuário, a quantidade total de solicitações, a quantidade total de mensagens trocadas, o valor médio de mensagens trocadas por solicitação, a porcentagem de erros contra acertos, a quantidade de mensagens com baixa compreensão e as intenções mais requisitadas.

A quantidade total de usuários solicitantes compreende o número de usuários que acessaram a plataforma do agente conversacional e iniciaram pelo menos um diálogo no período de tempo supracitado. A média de abertura de solicitação por usuário trata da média de solicitações de abertura de chamado por um usuário.

A quantidade total de solicitações consiste no número total de solicitações entre usuários ativos e o agente que ocorreram durante o período selecionado. Uma única conversa consiste em mensagens que um usuário ativo envia para o agente e as mensagens que este envia para responder. Se o agente inicia a conversa e, em seguida o usuário fechar o aplicativo sem responder, essa conversa não será incluída na contagem total.

A quantidade total de mensagens trocadas refere-se ao número de mensagens trocadas por todos os usuários naquele intervalo de tempo. Enquanto o valor médio de mensagens trocadas por solicitação é o número de mensagens recebidas durante o período de tempo dividido pelo total de solicitações no período, consistindo em quantas mensagens um usuário troca em média com o agente conversacional.

A porcentagem de erros contra acertos é calculada com base no número total de conversas por dia, dentre essas quantas foram cobertas, o que significa que as intenções no diálogo entenderam as solicitações do usuário e foram capazes de resolvê-las e quantas não foram cobertas, portanto a entrada não correspondeu a uma intenção no diálogo.

A quantidade de mensagens com baixa compreensão, *fall back rates*, é o número de mensagens individuais com entendimento fraco. Essas mensagens não são classificadas por uma intenção e não contém nenhum contexto conhecido, são as mensagens que o agente informa que não está preparado para a demanda. A revisão de mensagens não reconhecidas pode ajudar a identificar possíveis problemas de diálogo.

As *top* intenções requisitadas, dentre todas as intenções criadas, realiza o *ranking* das intenções mais solicitadas pelos usuários.

Mediante as análises sob as métricas, realizou-se o aprimoramento dos exemplos nas intenções, porque houve repetições nas consultas. Dentre as mensagens trocadas que não tiveram mais de 80% de confiança sob a resposta, foram classificadas de acordo com o contexto de negócio correto. Já as que tiveram baixa compreensão e não faziam parte do domínio do UNIEDU foram marcadas como irrelevantes.

#### 4. Resultados e discussão

Os resultados obtidos nesta pesquisa originaram-se da análise do agente conversacional mediante as consultas realizadas pelos usuários, seguindo-se as métricas explicitadas anteriormente.

No período do estudo transversal ocorreram 27 solicitações de conversas, no que se refere a métrica de Quantidade total de solicitações, cada conversa conclui-se com uma média de 5,85 mensagens com o agente conversacional. Estas conversas foram realizadas por 22 usuários desconhecidos e diferentes, sendo que cada usuário solicitou uma nova conversa mais de uma vez com o *chatbot*, totalizando-se 158 mensagens trocadas (figura 5).

O algoritmo de processamento de linguagem natural fornecido pelo IBM Watson Assistant obteve uma performance de 79,7% de acerto sob as mensagens trocadas, tendo-se 32 mensagens sem entendimento, o que corresponde a 20,3%. Destas 32 mensagens, 22 foram classificadas como irrelevantes, pois o nível de compreensão delas foi demasiadamente baixo, fazendo com que o agente conversacional não empregasse as sugestões de intenções com contexto semelhantes. Analisando essas mensagens, identificou-se que se tratavam de expressões relacionadas a outros domínios de aplicação referente a questões mais filosóficas, as quais o *chatbot* não apresenta capacidade de resposta.

Das mensagens trocadas, dentre as *top* intenções mais requisitadas incluem-se, #Confirmacao, #Saudacao, #OQueEUiedu, #ComoSeCadastrar, #DocumentosEnsinoMedio e #General\_Agent\_Capabilities, dentre todas as conversas analisadas no período do estudo transversal a intenção que solicita um atendente humano foi requisitada apenas uma vez. Com os dados oriundos das interações com o agente conversacional o modelo foi refinado, atingindo-se uma precisão para o agente conversacional de 80.1%.

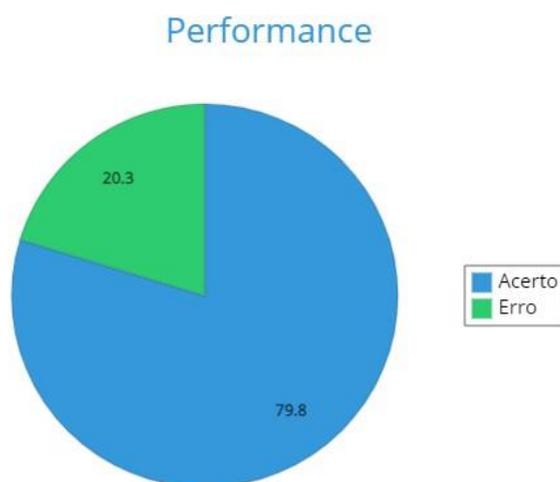


Figura 5. Performance do agente

Os valores obtidos nas métricas empregadas mediante as análises realizadas indicaram bons resultados para utilização do IBM Watson Assistant na construção de um agente conversacional. Semelhante ao estudo de Silva e Matos (2018), as etapas iniciais da construção do agente constituíram-se de obter uma base de conhecimento limitada, a partir desta o agente pode começar a ser utilizado, refinando-se o modelo durante o processo, os autores utilizam-se de mais metodologias para análises do agente conversacional mediante as consultas realizadas pelos usuários, como a avaliação do agente após a conversa, dessa forma a conclusão que o agente foi realmente útil é mais objetiva.

O estudo de Verleger (2018) utilizou-se de uma plataforma diferente para a construção do agente conversacional, sua amostra foi semelhante com esta pesquisa, o agente construído foi utilizado 48 vezes com os seguintes resultados, a média de respostas não encontradas foi de 33.3%, média de respostas respondidas com clareza 16.6%, média de respostas consideradas inúteis 18.8%, perante a esses dados a plataforma do IBM Watson Assistant provou-se ter um desempenho superior.

Em ambos os estudos as conclusões foram além da performance obtida pelo agente conversacional, os autores concluíram que os agentes têm potencial de encorajar os usuários a serem mais participativos, contribuir para um fluxo de atendimento mais limpo, além de oferecer oportunidades de contextualização sobre o domínio abordado pelo agente se comparado com pesquisas genéricas na internet.

## 5. Conclusão

Nesta pesquisa desenvolveu-se um agente conversacional por meio de processamento de linguagem natural utilizando as ferramentas do IBM Watson Assistant. A performance do algoritmo de processamento de linguagem natural foi avaliada por meio da análise as consultas realizadas no período do estudo transversal. O *chatbot* foi desenvolvido integrando um aplicativo de mensagens (Telegram) com APIs que se dividem em microsserviços e fazem o processamento e interpretação dos serviços do IBM Watson Assistant.

O IBM Watson Assistant se mostrou viável para manipulação e estruturação do conhecimento adquirido, tendo o fato de que existem várias possibilidades (caminhos) de como o usuário poderia iniciar a interação com o *chatbot*, mesmo com uma base de dados relativamente pequena e com as funcionalidades de detecção de intenção e sugestão com base no contexto, o agente conversacional conseguiu realizar os atendimentos com uma taxa baixa de solicitação de um atendente humano. A ferramenta também permite um ambiente de melhoria contínua, pois quanto mais é utilizada pode-se realizar refinamentos no modelo e aprimorar a sua precisão, bem como outras métricas como avaliação do atendimento e *chatbot rates*.

Quanto aos resultados obtidos é possível identificar que grande parte dos atendimentos realizados pelo *chatbot* são relacionados a dúvidas de primeiro nível como, onde encontro o link do site para cadastro, quais são os prazos dentre outras. Considerando que a empresa necessita de funcionários dedicados para o atendimento de e-mails, atendimentos telefônicos ou presenciais, e esta concentração de atendimentos, embora facilite o controle e estatística de chamados de um cliente, contribui muito com um gargalo no processo de interação entre as pessoas, podendo afetar em muito

procedimentos comerciais/acadêmicos. Neste contexto a implantação de uma ferramenta de atendimento automatizado como um agente conversacional que realize um primeiro nível de atendimento, auxiliaria e poderia diminuir as demandas de atendimento humano, liberando os colaboradores do setor para a realização de outras tarefas.

Em pesquisas futuras recomenda-se o aprimoramento dessa ferramenta de *chatbot*, empregando outras técnicas de inteligência artificial como reconhecimento de imagens, com finalidade de enviar fotos de documentos e a ferramenta ser capaz de identificar e apresentar os erros contidos, reconhecer a fala e converter para texto, aumentando a acessibilidade e as possibilidades de interação com o *chatbot*. Também, a integração com redes sociais e sites administrativos, respondendo a busca que o usuário fez no navegador.

## 6. Referências

- Agus Santoso, Heru e colab (2018). Dinus Intelligent Assistance (DINA)
- Borin, M. O uso de Ambientes Virtuais de Aprendizagem com Agentes Pedagógicos Emocionais para capacitação em organizações. Programa de Pós- Graduação em Sistemas e Processos Industriais/ Universidade de Santa Cruz do Sul, 2010. 120p. Dissertação de Mestrado.
- Creative Technology for Human Life, iSemantic 2018, p. 417–423, 2018.
- Chatbots Journal. Disponível em: <<https://chatbotsjournal.com/>>
- Domingues, Maria J. C. de S. Mídia e Aprendizagem: Um estudo comparativo entre Hipertexto e Chatterbot. 2003. 112f. Tese (Programa de Pós Graduação em Engenharia da Produção) – UFSC, Florianópolis, 2003. Gartner, Inc. (NYSE: IT) <http://www.gartner.com/newsroom/id/3412017>.
- IBM. IBM. 2021. Disponível em: <[https://www.ibm.com/br-pt/watson?mhsrc=ibmsearch\\_a&mhq=Watson](https://www.ibm.com/br-pt/watson?mhsrc=ibmsearch_a&mhq=Watson)>.
- Haode QI, A.S. Benchmarking Intent Detection for Task-Oriented Dialog Systems. 7 dec 2020.
- Santa Catarina. <Http://www.uniedu.sed.sc.gov.br/>. [N.a.]. Disponível em: <<http://www.uniedu.sed.sc.gov.br/index.php/uniedu-principal>>.
- Hirschberg, J. e Manning, C. D. Advances in natural language processing. Science, v. 349, n. 6245, p. 261–266, 2015 Tradução.
- Jung, Carlos. (2004). Metodologia para Pesquisa & Desenvolvimento: Aplicada a Novas Tecnologias, Produtos e Processos.
- Karimkhany, K. Watson Assistant improves intent detection accuracy, leads against Kumar, Ela. Natural language processing. Tradução . [S.l.]: I.K. International Publishing House, 2011.

- Lacity, M. et al. Reimagining the University at Deakin: An IBM Watson Automation Journey Research on Business Services Automation. Deakin, Australia: Deakin University, 2017. p. 1-26.
- Lemon, Oliver. Conversational Interfaces. Data-Driven Methods for Adaptive Spoken Dialogue Systems, p. 1–4, 2012
- Liddy, Elizabeth D. Natural language processing. In Encyclopedia of Library Information Science, 2nd Ed. NY. Marcel Decker, Inc. 2001
- Liu, F et al. NIST Cloud Computing Reference Architecture. National Institute of Standards and Technology, 7 set 2011.
- Lipko, H. Meet Jill Watson: Georgia Tech's first AI teaching assistant. 2016. Disponível em: <<https://pe.gatech.edu/blog/meet-jill-watson-georgia-techs-first-ai-teaching-assistant>>. Acesso em: 18 jul. 2018.
- Loveys, B. Innovations in natural language processing from IBM to help enterprises better understand the language of their business. TÍTULO. 9 dez. 2020. Disponível em: <<https://www.ibm.com/blogs/watson/2020/12/innovations-in-natural-language-processing-from-ibm-to-help-enterprises-better-understand-the-language-of-their-business/>>.
- Lucchesi, Ivana L. et al. Avaliação de um ChatBot no Contexto Educacional: Um Relato de Experiência com Metis. RENOTE - Revista Novas Tecnologias na Educação. v. 16, n. 1, 2018.
- Luckin, Rose et al. Intelligence Unleashed An argument for AI in Education. Open ideas at Pearson, 2016.
- Luger, George F. Inteligência artificial: estruturas e estratégias para a resolução de problemas complexos. 4. ed Porto Alegre: Bookman.
- Molnar, Gyorgy; Szuts, Zoltan. The Role of Chatbots in Formal Education. SISY 2018 - IEEE 16th International Symposium on Intelligent Systems and Informatics, Proceedings, p. 197–201, 2018.
- Ortiz, Charles L. The Road to Natural Conversational Speech Interfaces. IEEE Internet Computing, v. 18, n. 2, p. 74–78, 2014
- Rozga, Szymon. Introducing the Microsoft Bot Framework. Practical Bot Development, p. 167–241, 2018
- Rahman, A. M.; AL Mamun, Abdullah; Islam, Alma. Programming challenges of chatbot: Current and future prospective. 5th IEEE Region 10 Humanitarian Technology Conference 2017, R10-HTC 2017, v. 2018- Janua, p. 75–78, 2018.
- RUSSEL, Stuart J.; NORVING, Peter. Inteligência Artificial. 3. ed. Rio de Janeiro: Prentice Hall, 2013. 1016 p.

SILVA, A; MATTOS, R. IBM Watson como Ambiente para Desenvolvimento e Execução de um Chatbot – Um Estudo de Caso Aplicado ao Processo de Atendimento ao Usuário, São Paulo SP, Brasil, 2018.

MOLNAR, Gyorgy; SZUTS, Zoltan. The Role of Chatbots in Formal Education. SISY 2018 - IEEE 16th International Symposium on Intelligent Systems and Informatics, Proceedings, p. 197–201, 2018.

VERLEGER, Matthew; PEMBRIDGE, James. A Pilot Study Integrating an AI-driven Chatbot in an Introductory Programming Course. Proceedings - Frontiers in Education Conference, FIE, v. 2018- Octob, p. 1–4, 2019.

WEIGELT, Sebastian e TICHY, Walter F. Poster: ProNat: An Agent-Based System Design for Programming in Spoken Natural Language. 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, 2015.

## Apêndice A – Colcoar um nome igual título

Intenção	Descrição	Qtd. Exemplos
1-#ComoSaberSeFuiClassificado	Como saber se fui classificado	6
2-#ComoSeCadastrar	Como se cadastrar	8
3-#ComprovanteBemFamiliarCarro	Comprovante do grupo familiar – informações de veículo	5
4-#ComprovanteBemFamiliarCasaGanha	Comprovante do grupo familiar- informações de imóvel recebido	8
5-#ComprovanteBemFamiliarCasaPropria	Comprovante do grupo familiar- informações de imóvel próprio	5
6-#ComprovanteBemFamiliarObrigatorio	Comprovante do grupo familiar- documentos obrigatórios	6
7-#ComprovanteBemFamiliarCertidaoImovel	Comprovante do grupo familiar – certidões do imóvel	4
8-#ComprovanteBemFamiliarOndeEncontrar	Comprovante do grupo familiar – onde encontrar os documentos	5
9-#ComprovanteBemFamiliarSemNada	Comprovante do grupo familiar – não possui os documentos	5
10-#ComprovanteBemFamiliarSemRenda	Comprovante do grupo familiar- não possui renda	6
11-#ComprovanteBemFamiliarTempoEntrega	Comprovante do grupo familiar- tempo médio de entrega dos documentos	6
12-#ComprovanteBolsaEstudo	Comprovante do grupo familiar – informações da bolsa de estudo	3
13-#ComprovanteDoencaCronica	Documentos informando doenças crônicas	1
14-#ComprovanteFinanciamento	Documentos de financiamento	5
15-#ComprovanteResidencia2Anos	Comprovante de residência com mais de 2 anos	2
16-#ComprovanteResidenciaData	Data mínima do comprovante de residência	3
17-#ComprovanteResidenciaNenhuma	Não possui comprovante de residência	1
18-#ComprovanteTipoMoradiaBemNosso	Documentos do tipo de moradia	3
19-#ComprovanteTipoMoradiaCarro	Documentos com informações veiculares	3
20-#Confirmação	Compreende confirmações	5
21-#Despedida	Compreende despedidas	5
22-#DocumentoBemFamiliarAgricultor	Documentos de agricultor	5
23-#DocumentoBemFamiliaCarroVendido	Documentos informando venda de veículo	4
24-#DocumentoBemFamiliarCertidaoNaoEncontrada	Não encontrou as certidões do grupo familiar	4
25-#DocumentosEnsinoMedio	Documentos do ensino médio	4

26-#General_About_You	Solicite atributos pessoais genéricos.	14
27-#General_Agent_Capabilities	Solicite recursos de robô.	22
28-#General_Connect_to_Agent	Solicite um agente humano.	22
29-#General_ending	Termine a conversa.	26
30-#General_human_or_bot	Pergunte se está falando com um humano ou um robô.	19
31-#General_Greetings	Cumprimente o robô.	10
32-#General_Positive_Feedback	Exprima sentimento positivo ou gratidão.	23
33-#General_Negative_Feedback	Expresse feedback desfavorável.	21
34-#General_Security_Assurance	Exprima preocupações sobre a segurança do robô.	23
35-#GrupoFamiliarMoroSozinho	Documentos informando que o usuário mora sozinho	2
36-#HistoricoEscolarEGraduacao	Documentos de histórico escolar e graduação	3
37-#MudancaCadastro	Pergunta se houve alteração no cadastro	7
38-#Negacao	Compreende negações	5
39-#OqueEUNIEDU	Pergunta sobre o que é o UNIEDU	5