Unisanta Science and Technology, **2020, 1, December**
Published Online 2020 Vol.9 N°2 http://periodicos.unisanta.br/index.php/sat)

UNISANTA
Science and Technology
ISSN 2317-1316

Robô de Conversação Baseado em Inteligência Artificial para Treinamento na Lei Geral de Proteção de Dados Pessoais

Athirson Passos Ferreira De Jesus, Douglas Augusto Nunes Da Silva,
Jorge De Souza Jacintto Junior, Marcelo Barreto Buongermino, Nathalia Carolina Murat,
Tatiane de Aparecida Oliveira Gouvêa, Luiz Antonio Ferraro Mathias

UNISANTA – Universidade Santa Cecília – Faculdade de Sistemas de Informação
Rua Oswaldo Cruz, 266 - Santos-SP, Brasil - CEP: 11045-100

E-mail*:* lmathias@unisanta.br
Received March*, 2020*

Resumo: Este trabalho refere-se ao desenvolvimento de uma ferramenta de atendimento automático, cujo principal objetivo é auxiliar usuários a obterem de forma instantânea informações relacionadas à Lei Geral de Proteção de Dados. Automatizando o atendimento e respondendo perguntas de tal forma que as pessoas tenham a impressão de estar conversando com outra pessoa e não com um programa de computador. Para o desenvolvimento da aplicação, foi utilizada *Microsoft Bot Framework*™, uma plataforma própria para criação de *bots* capazes de interagirem naturalmente com usuários, a linguagem de programação utilizada foi o node.js. Para que haja uma comunicação em linguagem natural com o usuário, foi utilizado o serviço LUIS (*Language Understanding Intelligent Service*), que permite que a comunicação entre o usuário e o *bot* ocorra de maneira orgânica. Processamento de Linguagem Natural, também conhecido como PLN, é o meio que permite a interação entre as máquinas e a linguagem natural. Essa tecnologia possibilita que os robôs entendam e compreendam o que as pessoas falam, independente da forma e conteúdo. Seu objetivo é fornecer aos computadores a capacidade de entender e compor textos. "Entender" um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos, analisar sentimentos e até aprender conceitos com os textos processados. Para armazenar as informações sobre a LGPD foi utilizado o QnA Maker como base de conhecimento. Em vista disso, a utilização do *bot* no ambiente empresarial obtém uma grande facilidade na busca de informações, além de dispor de um recurso inovador para as organizações.

Palavras-chave: Chatbot, Inteligência Artificial, Linguagem Natural, LGPD, Processamento de Linguagem Natural.

Conversation Robot Based on Artificial Intelligence for Training in the General Data Protection Regulation

Abstract: This work refers to the development of an automatic answering tool, the main objective of which is to assist users to instantly obtain information related to the General Data Protection Regulation. Automating service and answering questions in such a way that people have the impression that they are talking to someone else and not with a computer program. For the development of the application, Microsoft Bot Framework ™ was used, a proprietary platform for creating bots capable of interacting naturally with users, the programming language used was node.js. In order to have a natural language communication with the user, the LUIS (Language Understanding Intelligent Service) service was used, which allows communication between the user and the bot to occur organically. Natural Language Processing, also known as NLP, is the medium that allows the interaction between machines and natural language. This technology allows robots to understand and comprehend what people say, regardless of form and content. Its purpose is to provide computers with the ability to understand and compose texts. "Understanding" a text means recognizing the context, performing syntactic, semantic, lexical and morphological analysis, creating summaries, extracting information, interpreting the senses, analyzing feelings and even learning concepts from the processed texts. To store information about GDPR, QnA Maker was used as a knowledge base. In view of this, the use of the bot in the business environment makes it very easy to search for information, in addition to having an innovative resource for organizations.

Keywords: Artificial Intelligence, Chatbot, GDPR, Natural Language, Natural Language Processing.

1. Introdução

As empresas vêm tentando, sistematicamente, melhorar o relacionamento com os clientes e, neste sentido, vêm buscando ferramentas que suportem a automatização de processos de negócio que otimizem a qualidade do atendimento aos usuários.

Levando em conta essa necessidade, o emprego de robôs de conversação na interatividade com clientes,

Copyright © 2020 ISESC.

tende, além da melhora do atendimento, automatização deste e reduções significativas nos custos das organizações.

Chatbot, *chatterbot* ou simplesmente *bot* são robôs de conversação, ou seja, ferramentas de comunicação que simulam a fala humana e são capazes de dialogar com usuários [6,7].

Um "*Chatbot* é um sistema que utiliza linguagem natural para se comunicar com o usuário, respondendo como um humano [7].

Este trabalho de pesquisa acadêmica trata como tema principal a área de inteligência artificial, que envolve utilizar a conceituação do processamento de linguagem natural, a fim de utilizar uma ferramenta de conversação para apoiar e facilitar a maneira de busca de informações sobre a Lei Geral de Proteção de Dados (LGPD)[1], que foi inspirada na GDPR (*General Data Protection Regulation*, em português, Regulamento Geral de Proteção de Dados) lei que está em vigor na Europa.

A execução dessa pesquisa vem de uma necessidade atual por parte das organizações em se adequar a nova legislação que entrará em vigor. Essa legislação dispõe sobre o tratamento de dados pessoais, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. A LGPD oferece aos indivíduos maior controle sobre seus dados pessoais, garantindo maior transparência sobre a utilização dos dados.

Para alcançar os objetivos do trabalho, esta pesquisa tem por propósito uma abordagem qualitativa, por estar sendo analisado, interpretado e descrito o desenvolvimento de um robô de conversação, com objetivo descritivo que visa identificar questões que serão utilizadas no bot. Tem como base o método experimental, que consiste em conhecer as vantagens de robôs de conversação em aplicativos de mensagens instantâneas, por meio de estudo em pesquisas e trabalhos científicos já existentes na área.

1.1 Inteligência Artificial

Com o passar do tempo, o ser humano ambicionou a criação se um ser semelhante ao seu, um ser artificial com capacidade exponencial de inteligência para poder talvez substituir o humano em suas atividades e ações cotidianas, para isso, desde os anos 40 vem sendo desenvolvida e aprimorada, a chamada "Inteligência Artificial" que, ao decorrer dos anos, vem se mostrando uma área de estudo com potencial para ajudar e complementar a Inteligência Humana.

Para John McCarthy [4] a inteligência artificial "É a ciência e engenharia de criar máquinas inteligentes, especialmente programas de computador inteligentes. Está relacionada à tarefa semelhante de utilizar computadores para entender a inteligência humana, porém a Inteligência Artificial não necessariamente deve se confinar aos métodos que são biologicamente observáveis.".

Nascida no âmbito da Ciência da Computação, a Inteligência Artificial também conhecida popularmente como IA foi criada em 1940. Pensado de forma a dar mais uma utilidade para o computador, com o intuito de emular a forma que o humano pensa, a IA possuí algumas categorias como: máquinas que pensam como seres humanos; máquinas que se comportam como seres humanos; máquinas que pensam racionalmente e máquinas que se comportam de forma racional. É difícil impor uma definição assertiva para o que é inteligência artificial, mas a ideia é tentar por meio dessa emulação do raciocínio humano, gerar soluções para problemas complexos. Segundo Coppin [2], "Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos".

1.2 Linguagem Natural

O *Homo Sapiens* é separado de outras espécies pela capacidade da linguagem [6]. A linguagem natural é utilizada como principal meio de comunicação entre seres humanos.

Diferentemente das linguagens de programação (Java, C#, JavaScript), que pode-se classificar como linguagens formais – que são projetadas para que não aconteça ambiguidade – a linguagem natural, seja ela escrita ou falada, pode conter ambiguidade, isto é, uma única sentença pode conter mais de um significado, que muitas vezes são difíceis de interpretar até por humanos, quanto mais por um computador. As linguagens naturais (e.g. português, inglês, espanhol) são difíceis de lidar, pois estão em constante evolução e mutação e, além disso, possuem gírias, dialetos e abreviações, além da já citada ambiguidade, o que se torna desafiador para o desenvolvimento de ferramentas que precisam compreendê-las. Para tanto, dentro do campo de estudo da IA, existe um conjunto de técnicas que são usadas no processamento destas linguagens.

1.2.1 Processamento de Linguagem Natural

O processamento de linguagem natural, ou PLN, é compreendido como uma área da computação que tem

1 Lei 13.709/2018 (Lei Geral de Proteção de Dados) que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado.

Copyright © 2020 ISESC

como objetivo extrair representações e significados mais completos em linguagem natural, de maneira inteligente.

Segundo Malta [3] é uma ciência que abrange um conjunto de técnicas e métodos que facilitam a análise e textual por um computador. Antes, cabe recapitular que linguagem natural se compreende pelas tecnologias utilizadas por seres humanos para comunicação, expressão ou interação com outro ser humano.

De acordo com Barbosa [1], a PLN utiliza-se conceitos linguísticos como classes de palavras, como verbo, substantivo, adjetivo, estruturas e classes gramaticais, conhecidas como *Part-Of-Speench*, bem como de situações mais complexas, como sinônimos, metáforas, anáforas, catáforas, semânticas, léxicas e sintática.

A PLN é utilizada não apenas para textos digitados via documentos digitais, mas também para textos escritos em papeis e a língua falada, utilizando-se também a tradução automática de texto entre as linguagens, considerando, em alguns casos, os sentimentos e opiniões transmitidas pela linguagem.

Todavia, nem sempre a linguagem humana, seja ela escrita ou falada, não é usada corretamente. Torna-se difícil a sua legitima interpretação, sua consideração ou seu real sentido, sendo um problema difícil para computadores.

Descobrir quem fez o que, a quem, quando, onde, como e o porquê não basta. O ser humano expressa-se por seus sentidos e impressões em sua comunicação, sendo necessária também sua interpretação pela ferramenta.

Scarpa [19] relata que alguns problemas comuns são o uso de sinônimos, que não entendem o real sentido da palavra em determinada oração. Outro problema apresentado pela autora é a polissemia, que é o múltiplo significado de uma palavra, como é o caso da palavra manga, papel etc.

1.2.2 Pré-Processamento

Primeiramente, para realização da análise léxica do documento, se faz necessário um pré-processamento dos textos recebidos, processo esse que analisa as entradas de cada palavra em uma oração, retirando todas as palavras que aparecem com determinada frequência, em um processo chamado *stopwords* (Scarpa [19]).

Para exemplificar, essa etapa é responsável por receber as palavras, separando-as em grupos, retirar palavras idênticas e separar em grupos menores. Após isso, é retirado seus prefixos e sufixos das palavras e, o seu resultado são armazenados em um banco de dados.

Malta [3] ressalta que, para começar o pré-processamento, é necessário fazer um processo de *stemming* (stemização), em morfologia linguística é o processo de análise de uma palavra reduzindo-a a sua forma primitiva, como um radical, removendo seus prefixos (exemplo: "gato" para "gat", "meninas" para "menin". Como resultado, são gerados *tokens*[2].

Após essa análise léxica, é necessário fazer a análise sintática da oração, sendo próprio deste procedimento durante a extração de palavras chaves. Ribeiro [18] diz que é *"a tarefa de identificação automática de um conjunto de termos que descreva satisfatoriamente o assunto de um documento"*, sendo de suma importância para o estudo da PNL.

Há algoritmos disponíveis para a realização desta tarefa, separados em duas categorias: os que utilizam de métodos estatísticos ou métodos estáticos.

Para métodos estáticos, é realizado o processo *gazetter*[3]*, que*, segundo Ribeiro [18] acontece *"quando realiza a extração através do uso de uma lista pré-definida de palavras"* e armazenada em um banco de dados.

1.3 Privacidade de Dados

A privacidade de dados é, atualmente, um tema discutido amplamente, embora sua importância já exista a algum tempo, esse assunto tem ganhado mais força a cada dia devido a veiculação de notícias nos últimos anos sobre violação de dados em empresas de arquivos de crédito e companhias aéreas.

O que é impressionante é que até mesmo em organizações que deveriam gerenciar de uma forma mais responsável esses dados pessoais como, por exemplo, um fornecedor de proteção de dados, essas informações são expostas.

A preocupação em manter a privacidade dos dados pessoais surgiu quando foi observado que dados são capazes de gerar valor e podem ser usados de forma estratégica para influenciar desde consumo de produtos, por exemplo, a até situações que possam causar maior impacto, como por exemplo, campanhas políticas.

Um exemplo de como isso pode ser usado para tal fim, em 2016 a empresa Cambridge Analytica usou inadequadamente dados de 87 milhões de usuários, a partir do Facebook para influenciar na campanha eleitoral nos Estados, ocasião em que a empresa havia comprado ilegalmente os dados das pessoas sem o consentimento delas.

2 ***TOKENS*** **-** *Token* são cadeias de caracteres que possuem significado. Em linguagens de programação podem ser palavras reservadas, identificados, números, sinais, operadores etc. No caso apresentado refere-se ao resultado da extração dos radicais.

3 ***GAZETTEER*** - *Gazetteer* é Dicionário ou diretório das palavras. No exemplo apresentado, servirá como índice das palavras constantes no documento.

Copyright © 2020 ISESC

2. Materiais e Métodos

A partir de uma abordagem qualitativa, uma vez que esta pesquisa científica procura analisar, interpretar e descrever o desenvolvimento de um robô de conversação, o presente estudo tem objetivo descritivo por identificar e mostrar questões que serão utilizadas para a criação do *bot*, como o Serviço LUIS™, a ferramenta QnA Maker [7], entre outros métodos que serão abordados no do correr desse tópico. A pesquisa tem como base o método experimental, pois consiste em desenvolver um *software* (robô de conversação - *Chatbot*) para plataformas de conversação instantânea [5].

2.1 Serviço LUIS™ (*Language Understanding*)

O *chatbot* necessita de um serviço inteligente, que consiga reconhecer padrões em frases e identificar a qual intenção a frase pertence. No mercado atual, existem algumas ferramentas que suprem essa necessidade, como o Watson, sistema cognitivo da IBM que oferece uma API de PLN para ser consumido por outras aplicações, bem como o LUIS que envolve uma série de serviços cognitivos oferecidos pela Microsoft™. O serviço LUIS usa três aspectos principais para entender idiomas [8]: **Enunciados:** um enunciado é uma entrada do usuário que seu aplicativo precisa interpretar; **Intenções:** uma intenção representa uma tarefa ou ação que o usuário deseja executar. É uma finalidade ou uma meta expressa no enunciado de um usuário; **Entidades:** uma entidade representa uma palavra ou frase dentro do enunciado que você deseja extrair.

Este trabalho utilizou o serviço LUIS™, o qual permitiu a definição de um conjunto de intenções provido pelo serviço, relacionadas a frases acrescentadas neste pelos autores da pesquisa. Essa ferramenta foi empregada com o propósito de permitir que o *chatbot* pudesse compreender a intenção da pergunta do usuário relacionada com a LGPD, verificando se esta intenção, que o usuário digitou por meio de plataformas de mensagens instantâneas, corresponde com alguma intenção que foi estabelecida dentro do serviço, possibilitando responder de forma mais natural.

No *Bot* LGPD[12] ao receber um enunciado enviado pelo usuário, o serviço aciona um algoritmo baseado em técnicas de inteligência artificial que irá calcular e analisar as semelhanças entre o enunciado que foi digitado pelo usuário da solução através de uma *interface*, com as outras já existentes no serviço, identificando a qual intenção aquela frase mais se assemelha através de um sistema de pontuação.

O resultado dessa pontuação é melhorado com um processo de "treinamento", permitindo ao serviço LUIS™ uma melhor identificação das particularidades de cada intenção. Um outro recurso importante desta plataforma de PLN são as Entidades, representadas por dados essenciais que foram digitados pelo usuário e que devem ser colhidas da frase para que o *bot* possa fazer a busca da informação desejada.

A figura 1 abaixo demonstra um trecho das entidades inseridas no sistema do LUIS™.

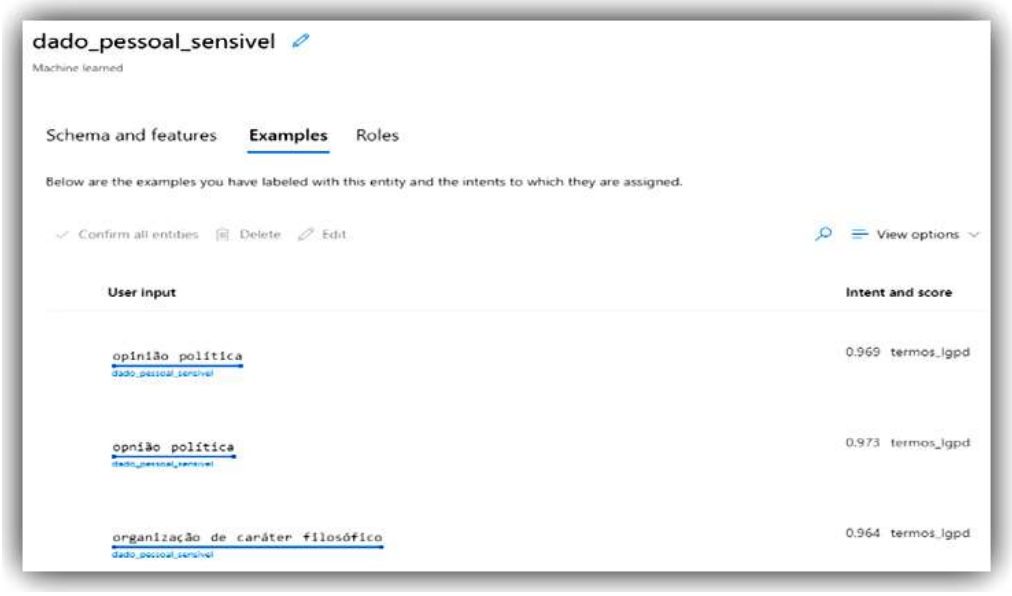

Figura 1. Trecho da Entidade Dado Pessoal Sensível no LUIS™.

Copyright © 2020 ISESC

Na figura anterior, é possível ver o trecho da entidade "Dado Pessoal Sensível" e algumas declarações que estão relacionadas a esta entidade, como o "dado biométrico" e o "dado genérico".

2.2 QNA Maker™

O QnA Maker™ é uma base de dados com serviço de Processamento de Linguagem Natural (PNL) disponibilizado pela Microsoft™ em sua plataforma Azure (ambiente de apoio tecnológico baseados em serviços em Nuvem), para criação de camada de conversação natural sobre esta base, podendo ser "utilizado para encontrar a resposta mais apropriada para qualquer entrada de idioma natural proveniente de sua *Knowledge Base* (base de dados de conhecimento) personalizada de informações." [5]. Em suma, o QnA Maker™ utiliza sua base de conhecimento para, através da intenção capturada do enunciado do usuário pelo LUIS, responder o questionamento utilizando linguagem natural.

Esta ferramenta está sendo utilizada como uma base de conhecimento que fornece informações sobre a LGPD aos usuários.

Para o *Bot* LGPD a base de dados, contendo as informações sobre a Lei Geral de Proteção de Dados, foi construída a partir de um arquivo em formato de planilha eletrônica.

As vantagens deste tipo de serviço são a agilidade na entrega da resposta, a simplicidade na inserção de novos dados sem a utilização de linhas de código de programação, dimensionamento dinâmico através da plataforma e facilidade de integração em outros serviços Microsoft seja pelo portal da Azure ou usando Aplicações externas.

Para obter a resposta mais assertiva, o serviço usa a base de dados de conhecimento treinada para devolver o resultado correto e/ou qualquer aviso correlacionado que possa ser utilizado para refinar a pesquisa. Os formatos da fonte de dados importados são estruturados e agrupados em metadados "chave/valor" ou, caso prefira, em níveis, retornando uma resposta formatada em JSON, conforme figura 2 abaixo. Segundo sua documentação, o serviço "utiliza a resposta JSON para tomar decisões sobre como continuar a conversa. Essas decisões podem incluir mostrar a resposta principal e apresentar mais opções para refinar a pesquisa a fim de obter a melhor resposta."

Com seu ciclo iterativo de vida da base de dados de conhecimento no QnA Maker, há melhor aprendizado e publicação e coleta de dados de consulta nos *Bots* Publicados, análise e alterações dos dados em sua base. Após a publicação do *bot*, o serviço está disponível para uso.

2.3 Microsoft *Bot* Framework

O *Microsoft Bot Framework* é uma ferramenta para a criação, testes, implantação e gerenciamento de *chatbots*. É um recurso do Serviço de *Bot* da plataforma em nuvem Azure, da Microsoft™, e nesta plataforma os serviços que normalmente seriam armazenados em servidores locais, agora poderão ser abrigados na nuvem[4].

Conjuntos de ferramentas como o *Microsoft Bot Framework* são chamados de SDK's (*Software Development Kits* – Kits para desenvolvimento de Software). Através da estrutura provenientes do SDK do *Bot Framework* e em conjunto com serviços de IA, neste contexto, torna-se possível a criação de robôs de conversação capazes de compreender a linguagem natural e interagir com pessoas através de perguntas e repostas.

2.4 Integração dos Serviços LUIS™ E QNA Maker™

Apesar de ambos os serviços utilizarem Processamento de Linguagem Natural, o sistema do QnA Maker utiliza modelo estático, não sendo possível o melhor tratamento das perguntas, além de não realizar o tratamento de perguntas via reconhecimento de voz, bem como não possuir grandes tratamentos nas intenções reais dos usuários, sendo necessário o uso de outro sistema cognitivo ou aplicação para melhor retorno destas respostas. Por outro lado, o serviço LUIS™, apesar de conseguir tratar respostas duplas dentro da aplicação, não retorna as respostas sem o tratamento por uma aplicação externa para criação de modelos estatísticos.


```
[
  {"displayOrder":0,"qnaId":2,"displayText":"Level 2 Question A"},
  {"displayOrder":0,"qnaId":3,"displayText":"Level 2 - Question B"}
]
```


Figura 2. Exemplo de Código JSON Gerado.

[4] Computação em **nuvem** (do inglês *cloud computing*) é um conceito que faz referência a uma tecnologia que permite o acesso a programas, arquivos e serviços por meio da internet, sem a necessidade de instalação de programas ou armazenamento de dados

Copyright © 2020 ISESC

Segundo a Microsoft [10,11], em conjunto com o sistema cognitivo LUIS, é possível analisar a "intenção do texto de um usuário (conhecido como uma expressão), enquanto QnA Maker™ determina a resposta para o texto de um usuário (conhecido como consulta).". Como apresentado na figura 3, onde é demonstrado o fluxograma do funcionamento da solução, representando a arquitetura geral da aplicação.

O LUIS™ é recomendado quando se precisa saber a verdadeira intenção do enunciado como processo no *chatbot*, podendo extrair melhor a entidade para outro serviço, para obtenção da resposta. Para o QnA Maker este deve ser usado quando se tem a base de dados de conhecimento estática de respostas.

Conforme figura 4 abaixo, é apresentado o fluxograma da aplicação que fará a conexão entre o LUIS™ e o QnA Maker, esta deverá fazer o monitoramento da resposta apresentada, utilizando pontuações de confiança e, dependendo deste grau, entregará à melhor resposta [13,14,15].

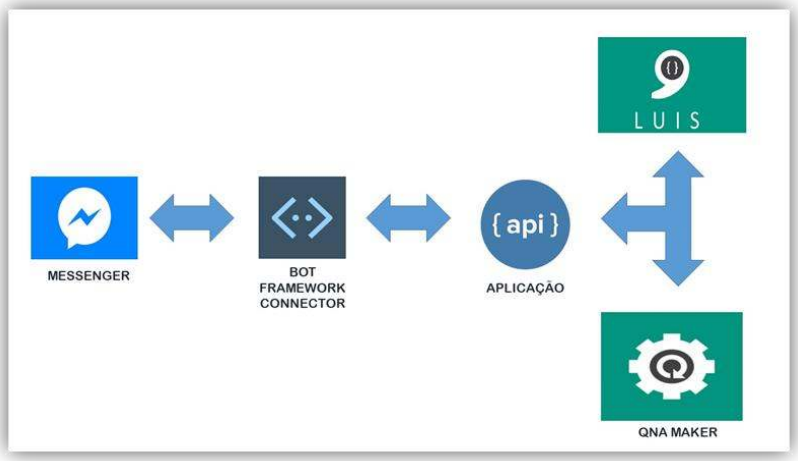

Figura 3. Fluxograma da arquitetura geral do trabalho.

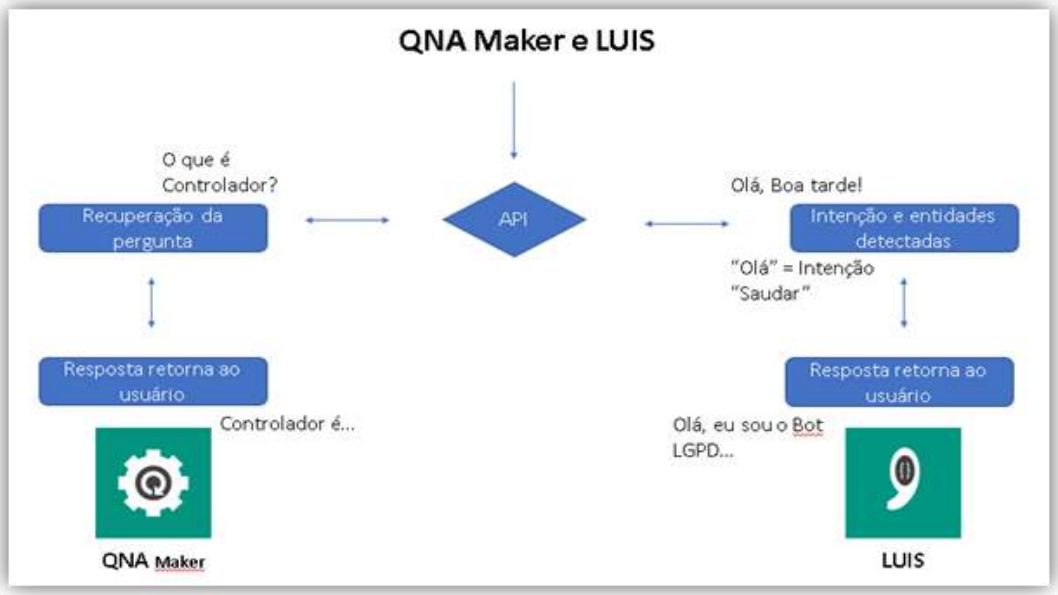

Figura 4. Fluxograma do Sistema utilizando QNA Maker / LUIS™.

Copyright © 2020 ISESC

3. Resultados e Discussões

Com As informações coletadas que tencionavam o desenvolvimento do *Bot*, foram de extrema importância para o bom funcionamento do *chatbot*, assim como foram reunidas informações de fontes integras e confiáveis sobre a LGPD (Lei Geral de Proteção de Dados) que é o foco de respostas do robô de conversação. Com a junção das informações foi possível o total desenvolvimento do *chatbot* e posteriormente, a sua publicação em canais de mensagens instantâneas como o *Microsoft Teams*, o *Telegram* e o *Web Chat*.

A interatividade entre as ferramentas QnA Maker™ e LUIS™, que utilizam Processamento de Linguagem Natural, contribuíram para com a qualidade do desempenho e refinamento na entrega de resultados das pesquisas e tempo de resposta aos usuários do *chatbot*.

Durante a construção da base de conhecimento, foi verificado que alguns termos possuem semelhanças, como o termo de "Dado Pessoal" que durante o treinamento do *Bot* foi notado uma imprecisão com o termo "Dado Pessoal Sensível", o que causou conflitos para o entendimento do *bot*, resultando em uma resposta ao usuário que não condizia com o esperado.

A fim de evitar o problema, foram criadas respostas que devolvem todas as informações de termos idênticos em forma de *cards* interativos, para melhor refinamento na pesquisa da resposta. Visando analisar o seu funcionamento após o desenvolvimento, o *chatbot Bot* LGPD foi submetido a testes de interação com perguntas realizadas pelo usuário nas plataformas *Web Chat*, *Telegram*, *Microsoft Teams*, Skype e Facebook Messenger, com a intenção de avaliar a acuracidade das respostas dadas pelo *Bot*, e examinar se os resultados desejados foram concretizados. A seguir serão detalhadas avaliações dos testes realizados, envolvendo resultados e observações que puderam ser notados.

3.1 Testes no *Telegram*

Em testes realizados na plataforma de mensagens instantâneas *Telegram*, foi simulado um cenário de início de conversa do usuário com o *Bot*. Primeiramente é solicitado ao usuário que inicie o *chatbot*, conforme figura 5.

Ao ser perguntado sobre Dados, o *Bot* questionou o usuário sobre o que ele gostaria de saber sobre a palavra dados, ofertando 3 (três) *cards* com opções existentes para essa palavra, que seriam "Dado Pessoal", "Dado Anonimizado" e "Dado Sensível" como mostrado na figura 5, possibilitando que este escolhesse qual informação seria mais apropriada para o seu questionamento.

O resultado dos testes realizados nos aplicativos *mobile* e *desktop* foram muito satisfatórios por haver uma resposta rápida ao usuário.



Figura 5. Conversa com o *Bot* pelo aplicativo *mobile Telegram*.

As *interfaces* dos aplicativos são bastante intuitivas, facilitando a interação com qualquer pessoa que os utilizará.

3.2 Testes no *Microsoft Teams*

Durante a realização dos testes no aplicativo *desktop* do *Microsoft Teams*, foram simuladas perguntas do usuário para verificar a acuracidade e assertividade das respostas.

Na figura 6 abaixo é demonstrada parte da conversa em que as perguntas são realizadas de formas distintas e o robô de conversação compreende, através do processamento de linguagem natural, que a intenção do usuário é a mesma para ambas as perguntas.

As conclusões foram bastante adequadas, visto que o robô compreendeu que a intenção das perguntas é a mesma. Além disso, os usuários têm a possibilidade de selecionar quais informações são mais relevantes para ele através de *cards* interativos. Neste contexto, pode-se observar também que o *Bot* fica disponível 24 horas por dia, podendo, assim, sanar as dúvidas das pessoas a qualquer momento.

Copyright © 2020 ISESC

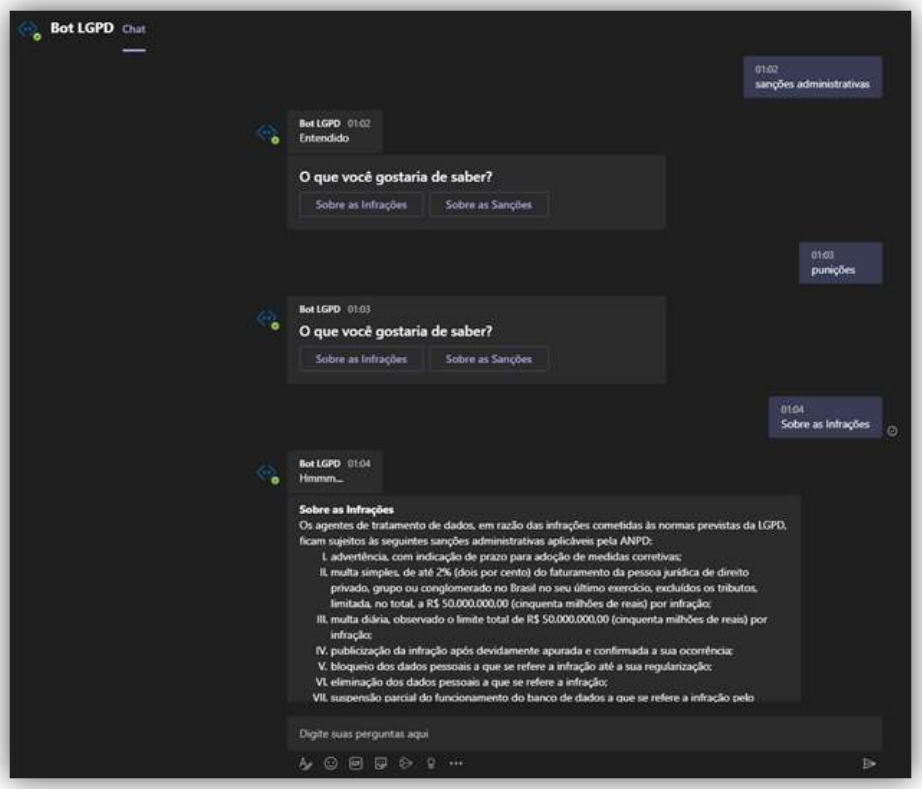

Figura 6. Conversa com o *Bot* pelo aplicativo *desktop Microsoft Teams*.

4. Considerações Finais

Atualmente é possível verificar que, com os avanços da tecnologia, as empresas estão cada dia mais próximas de atender as necessidades de seus clientes e fornecendo soluções que facilitem no cotidiano deles. Com esse intuito muitas empresas estão priorizando e agilizando o atendimento aos clientes, e, para isto, uma das soluções encontradas é a utilização de *chatbots* para que haja uma disponibilidade 24 horas por dia. Com a utilização dos *Bots*, os clientes podem sanar suas dúvidas a qualquer momento e de qualquer lugar.

Chatbots são ferramentas que promovem interações com usuários por meio de comunicação via mensagens de texto ou de voz, e simulam uma conversa humana utilizando linguagem natural.

A utilização desta ferramenta torna-se viável no processo de busca de informações em um ambiente contendo uma grande quantidade de dados.

Tendo em vista a necessidade das empresas em aderir às normas da Lei Geral de Proteção de Dados, o trabalho visou o desenvolvimento de um *chatbot* com o propósito de auxiliar os titulares de dados pessoais, ou seja, pessoas naturais (físicas) nos seus questionamentos quanto a LGPD por intermédio de aplicativos de mensagens instantâneas.

As publicações do *chatbot* para a interação com o usuário foram disponibilizadas em 3 (três) canais de comunicação: *Microsoft Teams*, por ser uma moderna plataforma de comunicação e colaboração que cada vez ganha mais espaço nas organizações; no *Telegram*, por ser uma ferramenta de mensagens instantâneas bastante prática e performática; e em uma página web, por ser de fácil acesso a todos os usuários.

Publicações em outros canais, como o *Messenger* e o *WhatsApp*, serão disponibilizadas futuramente.

Durante a pesquisa de ferramentas para o desenvolvimento da solução técnica, foram analisadas e testadas algumas opções, e as que melhor atenderam ao propósito do projeto foram o SDK[5] do *Microsoft Bot Framework*,

[5] **SDK** é a abreviação para Software Development Kit (do português, Kit de Desenvolvimento de Software.), e se trata de um conjunto de ferramentas e programas utilizados por desenvolvedores como base para construir algo de acordo com a sua necessidade.

Copyright © 2020 ISESC

a utilização da linguagem de programação *JavaScript* em ambiente *Node*.JS para a construção da API[6], os serviços cognitivos QnA Maker, utilizado como banco de conhecimento para realizar a consulta de informações sobre a LGPD e LUIS para identificar a intenção do usuário e proporcionar uma melhor naturalidade nas conversas.

As perguntas submetidas na *interface* de *chat* são compreendidas adequadamente com o auxílio do processamento de linguagem natural, processadas e gerenciadas por uma aplicação que submete a resposta mais adequada diante da intenção percebida em formato texto ou através de cartões interativos personalizados.

Foi possível notar que, com os experimentos feitos nos canais de conversação do *Telegram*, *Microsoft Teams* e via *Web Chat*, o *chatbot* demonstrou uma grande performance não só em compreender o contexto de uma pergunta, como também em responder de forma imediata às perguntas que foram submetidas, incluindo respostas que retornam ao usuário os cartões interativos personalizados, provando ser uma solução que serve de grande apoio em termos de obtenção de uma informação rápida e necessária com benefícios positivos e satisfatórios para a organização, especialmente para os que farão uso da ferramenta no dia a dia empresarial.

Conclui-se que em um mercado de trabalho cada vez mais competitivo, é necessário utilizar-se de maneiras inteligentes para a absorção de informações e transformá-las em conhecimento profícuo para toda a empresa na resolução de problemas.

O maior potencial da ferramenta em sanar dúvidas dos usuários é um grande aliado na aprendizagem do indivíduo sobre a LGPD.

Referências

[1] BARBOSA, J., Santos, R.; Magalhaes J. G.; Muniz, M.; Moura, R; **Introdução ao processamento de linguagem natural usando Python.** III Escola Regional de Informática do Piauí, v. 1, 2017, p. 336-360.

[2] COPPIN, B. **Inteligência Artificial**. 1ª Edição. São Paulo: LTC, 2010.

[3] MALTA, L. H. A.; KUROIVA, M. A; R. L.; **Aprendizado de máquina e processamento de linguagem natural aplicados à identificação de discurso de ódio**. Trabalho de Conclusão de Curso. Universidade de Brasília, Brasília, 2019.

[4] **MCCARTHY,** John. **What is AI?.** Stanford University, Stanford, Estados Unidos, 12 nov. 2007.

[5] **MICROSOFT. Documentação do LUIS (Reconhecimento vocal):** O que é Reconhecimento Vocal (LUIS)?, 2020. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/luis/what-is-luis>. Acessado em 28 de jul. de 2020.

[6] **SALESFORCE**. **State of Service:** Report: The Age of AI, and Agent and Customer Empowerment. 3rd Edition. Disponível em <https://www.salesforce.com/products/service-cloud/resources/state-of-service/>. Acesso em 12 de mai. De 2020.

[7] **BOT DOC. Documentação do Serviço de Bot do Azure:** Sobre o Serviço de Bot do Azure, 2019.Disponível em: <https://docs.microsoft.com/pt-br/azure/bot-service/bot-service-overview-introduction?view=azure-bot-service-4.0>. Acessado em 06 de ago. de 2020.

[8] AZURE. **Azure resources for QnA Maker**. Disponível em: <https://docs.microsoft.com/en-us/azure/cognitive-services/QnAMaker/Concepts/azure-resources>. Acessado em 24 de ago. de 2020.

[9] **DATA SOURCE QNA MAKER**. **Ciclo de vida da base de dados de conhecimento no QnA Maker**. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/concepts/development-lifecycle-knowledge-base>. Acessado em 24 de ago. de 2020.

[10] **MICROSOFT-LUIS. Documentação do LUIS**. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/luis/>. Acessado em 26 de ago. de 2020.

[11] **MICROSOFT- DATA SOURCES**.. **Importing from data sources**. Disponível em: <https://docs.microsoft.com/en-us/azure/cognitive-services/QnAMaker/Concepts/knowledge-base>. Acessado em 25 de ago. de 2020.

[12] **QNA MAKER LS**. **Language support for a QnA Maker resource and knowledge bases**. Disponível em: <https://docs.microsoft.com/en-us/azure/cognitive-services/qnamaker/overview/language-support>. Acessado em 24 de ago. de 2020.

[6] **API** é um conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web.

Copyright © 2020 ISESC

[13] **QNA MAKER SERVICE**. **O que é o serviço QnA Maker?**. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/overview/overview>. Acessado em 24 de ago. de 2020.

[14] BD **QNA MAKER**. **Teste sua base de dados de conhecimento no QnA Maker**. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/how-to/test-knowledge-base>. Acessado em 24 de ago. de 2020.

[15] **POWER VIRTUAL AGENTS**. **Tutorial:** Adicionar sua base de dados de conhecimento ao Power Virtual Agents. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/tutorials/integrate-with-power-virtual-assistant-fallback-topic>. Acessado em 27 de ago. de 2020.

[16] **QNA MAKER**. **Usar o QnA Maker para responder a pergunta**s. Disponível em: <https://docs.microsoft.com/pt-br/azure/bot-service/bot-builder-howto-qna?view=azure-bot-service-4.0&tabs=cs>. Acessado em 24 de ago. de 2020.

[17] **NLP**. **Use serviços cognitivas com NLP (processamento de idioma natural) para enriquecer conversas de *Bot* de *chat***. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/luis/choose-natural-language-processing-service#use-both-services-when-your-knowledge-base-is-incomplete>. Acessado em 25 de ago. de 2020.

[18] Ribeiro, C. M. F. A., Bastos, D.; Santos, J.; Guerra, J.; Alves, R.; **Melhoria de Processos de Gestão em Saúde Pública: Extração automática de conhecimento e busca semântica de documentos não estruturados**. EmpíricaBR-Revista Brasileira de Gestão, Negócio e Tecnologia da Informação, v. 2, n. 1, 2017, p. 72-87.

[19] CARPA, A. D. **Técnicas de processamento de Linguagem Natural aplicadas às Ciências Sociais**. Dissertação (Mestrado) Programação de Pós-Graduação em Matemática Aplicada, Fundação Getúlio Vargas, Rio de Janeiro, 2017.

Copyright © 2020 ISESC