**UNIVERSIDADE MUNICIPAL DE SÃO CAETANO DO SUL USCS PRÓ-REITORIA DE GRADUAÇÃO**

**ESCOLA POLITÉCNICA - USCS**

**André Picaro Michelli**

**Fernando Navarro de Souza**

**Mauricio Carvalho Pinheiro**

**Marcelo Rodrigues Pascoal**

**Expl.AI.n ChatBot: Manual Cognitivo para Estruturar Dados e Gerar Informação**

**UNIVERSIDADE MUNICIPAL DE SÃO CAETANO DO SUL USCS PRÓ-REITORIA DE GRADUAÇÃO**

**ESCOLA POLITÉCNICA - USCS**

**André Picaro Michelli**

**Fernando Navarro de Souza**

**Mauricio Carvalho Pinheiro**

**Marcelo Rodrigues Pascoal**

**Expl.AI.n ChatBot: Manual Cognitivo para Estruturar Dados e Gerar Informação**

**Trabalho de Conclusão de Curso apresentado à Escola Politécnica da Universidade Municipal de São Caetano do Sul como requisito obrigatório para obtenção do título de bacharel em Ciência da Computação**

ORIENTADOR : Prof° Raphael Lopes de Souza

**Resumo**

O seguinte projeto para obtenção do título de bacharel em ciência da computação, curso vinculado a Escola Politécnica da Universidade Municipal de São Caetano do Sul tem como objetivo criar um software que processa arquivos textuais, para facilitar o acesso à determinada informação através de um algoritmo de busca. Onde por meio de um *chatbot* o usuário insere uma busca textual, e o mesmo faz a tratativa dos dados de entrada, e assim a aplicação realiza o processamento dos documentos previamente salvos, como por exemplo um arquivo *PDF ou DOCX*. Futuramente, pretendemos criar outras funcionalidades na aplicação, e assim poderemos também inserir dados por voz.

Atualmente, a sociedade produz dados de forma exponencial. De forma

paralela, as organizações também seguem esse mesmo padrão. Através de dados textuais em arquivos PDF ou DOCX extensos, contendo por exemplo as políticas, regras de negócio ou processos que devem ser seguidos dentro da corporação, e estão em constante alteração para se adequarem à metamorfose das relações organizacionais contemporâneas. Com isso surge a necessidade dos atores estarem consultando esses materiais constantemente, a fim de manterem-se atualizados em relação aos protocolos e ou informações existentes nesses arquivos.

A demanda que visamos atingir são os indivíduos que de alguma forma vão ter algum contato com determinado arquivo textual PDF ou DOCX contendo informações que precisam ser acessadas de forma rápida.

O processo de comunicação com o chatbot consiste em quatro componentes: Emissor, receptor, mensagem e canal, onde o emissor cria uma busca para ser transmitida por uma interface de chat para um receptor, O IBM Watson Assistant, sendo este responsável por encaminhar a mensagem para a API do chatbot, e caso a busca seja processada, permite à API emitir uma resposta ao Watson Assistant, que por sua vez devolve ao emissor a resposta. Sendo assim, utilizaremos o processo de *PLN (Processamento de Linguagem Natural).*

**Palavras-Chave**: Inteligência artificial, big data, manual cognitivo, processamento da linguagem natural, machine learning.

Abstract

The following project to obtain a bachelor's degree in computer science, a course linked to the Polytechnic School of the Municipal University of São Caetano do Sul, aims to create a software that processes text files, to facilitate understanding through a search algorithm. Where, through a chatbot, the user enters a textual search, and it handles the input data, and thus the application processes previously saved documents, such as a PDF or DOCX file. In the future, we intend to create other functionalities in the application, so that we will also be able to enter data by voice.

Currently, society produces data exponentially. In parallel, organizations also follow this same pattern. Through textual data in extensive PDF or DOCX files, containing for example the policies, business rules or processes that must be followed within the corporation, and are constantly changing to adapt to the metamorphosis of contemporary organizational relationships. With this comes the need for actors to be constantly consulting these materials, in order to keep up-to-date in relation to the protocols and/or information existing in these files.

The demand that we aim to reach is the individuals that somehow will have some contact with a certain textual PDF or DOCX file containing information that needs to be accessed quickly. The communication process with the chatbot consists of four components: Sender, receiver, message and channel, where the sender creates a search to be transmitted through a chat interface to a receiver, The IBM Watson Assistant, which is responsible for forwarding the message to the chatbot API, and if the query is processed, it allows the API to issue a response to the Watson Assistant, which in turn returns the response to the sender. Therefore, we will use the PLN (Natural Language Processing) process.

**Keywords**: Artificial intelligence, big data, cognitive manual, natural language processing, machine learning.

Lista de Siglas

* PLN: Processamento da Linguagem Natural;
* SQL: Structured Query Language;
* PHP: HyperText PreProcessor;
* PYTHON: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics.
* Rank-BM25 Search algorithms
* JS: JavaScript;
* PDF: Portable Document Format;
* ML: Machine Learning;
* HTML: HyperText Markup Language;
* CSS: Cascading Style Sheets.

**Sumário**

* Resumo............................................................................................................................3
* Abstract............................................................................................................................4
* Lista de siglas................................................................ ..................................................5
* Introdução.........................................................................................................................7
* Motivação...................................................................................................................7
* Objetivos Gerais.........................................................................................................7
* Objetivos Específicos ................................................................................................7
* Tecnologias Utilizadas.....................................................................................................8
* Referências
* Apêndice

**Introdução**

Com o avanço da quantidade de dados na sociedade, o ser humano sente a necessidade de, cada vez mais, adquirir informações de forma mais rápida. Segundo Ángel F. Villarejo-Ramos (2020, p. 1) e Juan-Pedro Cabrera-Sánchez (2020, p. 1):

1) “Atualmente, a sociedade gera dados sobre nossas atividades a uma taxa

exponencial de crescimento”. Ou seja, a quantidade total de dados duplica a cada dois anos.

Sendo assim, surgiu a necessidade do usuário ter acesso a estes dados que são produzidos de forma mais prática e que pudesse proporcionar ao mesmo uma experiência de acesso á informação um pouco mais rápida e fácil.

**Motivação**

A motivação deste trabalho é a necessidade da estruturação de dados, desenvolvendo um software que estruture esses dados de arquivos textuais dentro das organizações e auxilie as mesmas a agilizar o processo de elucidação de novas informações que precisam ser acessadas de forma rápida e prática. A demanda que visamos atingir nas organizações são os indivíduos que, de alguma forma, vão ter algum contato com determinado arquivo textual (pdf, docx, etc).

**Objetivos Gerais**

* Criação de um software que processa arquivos textuais.
* O software auxiliará o usuário a fazer tais buscas em arquivos.

**Objetivos Específicos**

* O usuário deverá ser capaz de inserir uma query a ser processada intuitivamente seguindo as instruções do próprio software dentro d eum chatbot.
* O software deverá capturar os dados inseridos e fazer o processamento realizando o processo de PLN para fornecer uma resposta.
* O software será treinado com as respostas previamente aprendidas e fornecerá a resposta que melhor se adequar à busca feita pelo usuário.
* O programa fornecerá respostas já aprendidas caso o usuário faça uma busca feita anteriormente.

**Tecnologias Utilizadas**

Durante o desenvolvimento da aplicação, foram utilizadas algumas tecnologias de front-end, back-end e para tratamento dos dados, dentre elas temos:

* HTML e CSS para a estilização.
* Next JS: Next.js é uma estrutura da web de desenvolvimento front-end React de código aberto criada por Vercel que permite funcionalidades como renderização do lado do servidor e geração de sites estáticos para aplicativos da web baseados em React para a estrutura da aplicação.
* Python para fazer o processamento de lingauem natural.
* IBM Watson Assistant e IBM functions para o chatbot que faz a conexão entre a API que faz o servidor que faz o processamento da linguagem natural PLN.
* MySQL para armazenamento das respostas obtidas pela inteligência artificial e de documentos de entrada inseridos na base de dados pelo usuário.