

**Universidade de Évora - Escola de Ciências e Tecnologia**

**Mestrado em Engenharia Informática**

Dissertação

**Desenvolvimento de Chatbots para responder a  
perguntas frequentes**

João Quirino Machado e Silva

Orientador(es) | Irene Pimenta Rodrigues

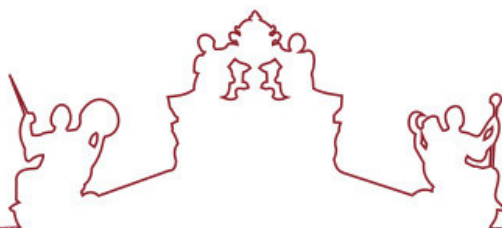
Évora 2021

---

---

---

---



**Universidade de Évora - Escola de Ciências e Tecnologia**

**Mestrado em Engenharia Informática**

Dissertação

**Desenvolvimento de Chatbots para responder a  
perguntas frequentes**

João Quirino Machado e Silva

Orientador(es) | Irene Pimenta Rodrigues

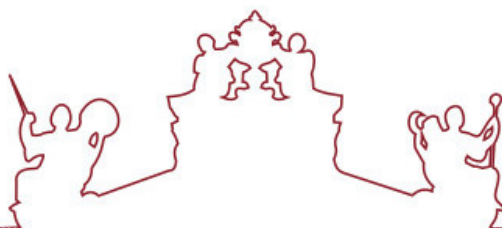
Évora 2021

---

---

---

---



A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências e Tecnologia:

Presidente | Lígia Maria Ferreira (Universidade de Évora)

Vogais | Dora Regina Oliveira Melo (Instituto Politécnico de Coimbra) (Arguente)  
Irene Pimenta Rodrigues (Universidade de Évora) (Orientador)



*A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original.*



# Agradecimentos

Gostaria de referir que a realização desta dissertação não teria sido exequível sem o contributo, direto ou indireto, de todos os que me rodeiam.

Em primeiro lugar, como não poderia deixar de ser, começo por agradecer à minha Orientadora, Irene Pimenta Rodrigues. Agradeço todo o tempo que dedicou às nossas reuniões, todo o conhecimento que acrescentou a esta dissertação e, acima de tudo, agradeço por não ter desistido de me orientar quando eu quase desisti de terminar a Tese.

Em segundo lugar, gostaria de agradecer aos meus pais, à minha irmã e à minha namorada por me rodearem com a sua presença imprescindível. Independentemente da minha situação, graças a eles, nunca me senti sozinho e sem apoio.

Por fim, mas não menos importante, queria deixar uma palavra a todos os amigos que, de formas várias, me motivaram a continuar.

A todos, muito obrigado.





# Conteúdo

<b>Conteúdo</b>	<b>ix</b>
<b>Lista de Figuras</b>	<b>xv</b>
<b>Lista de Tabelas</b>	<b>xix</b>
<b>Glossário</b>	<b>xxi</b>
<b>Sumário</b>	<b>xxiii</b>
<b>Abstract</b>	<b>xxv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Objetivos . . . . .	1
1.3 Plano da Tese . . . . .	2
<b>2 Estado da Arte</b>	<b>3</b>
2.1 História dos Chatbots . . . . .	4
2.2 Tipos de Chatbots . . . . .	5
2.2.1 Domínio da Conversa . . . . .	5
2.2.2 Geração da Resposta . . . . .	6
2.3 Técnicas Utilizadas . . . . .	6
2.3.1 Pré-processamento do Texto . . . . .	6
2.3.2 Tokenização do Texto . . . . .	7
2.3.3 Normalização do Texto . . . . .	7
2.3.4 Distâncias entre Textos . . . . .	7

2.4	Problemas dos Chatbots . . . . .	9
<b>3</b>	<b>Desenvolvimento</b>	<b>11</b>
3.1	Enquadramento do Chatbot . . . . .	11
3.2	Escolha do Software . . . . .	12
3.2.1	Python . . . . .	12
3.2.2	Python Regular Expression . . . . .	13
3.2.3	Python Natural Language Tool Kit . . . . .	13
3.2.4	Python NumPy . . . . .	15
3.2.5	Python Gensim . . . . .	15
3.3	Arquitetura do Sistema . . . . .	16
3.4	Descrição do Sistema . . . . .	17
3.4.1	Início de Conversa . . . . .	17
3.4.2	Questão . . . . .	17
3.4.3	Seleção de Resposta . . . . .	17
3.4.4	Resposta . . . . .	18
3.4.5	Fim de Conversa . . . . .	18
3.5	Cálculo de Medidas . . . . .	18
3.5.1	Arquitetura da Seleção de Resposta . . . . .	18
3.5.2	Grau de Similaridade . . . . .	19
3.5.3	Cálculo da Resposta . . . . .	20
3.6	Constituição da implementação do Sistema . . . . .	21
3.6.1	Chatbot - Variantes . . . . .	21
3.6.2	Chatbot - Funções Auxiliares . . . . .	22
3.6.3	FAQs utilizadas . . . . .	25
3.7	Chatbot - Implementação . . . . .	25
<b>4</b>	<b>Avaliação do Sistema</b>	<b>31</b>
4.1	Descrição das FAQs utilizadas . . . . .	32
4.2	Descrição das Experiências . . . . .	33
4.3	Métodos de Avaliação . . . . .	33
4.4	Resultados . . . . .	34
4.4.1	Covid . . . . .	34
4.4.2	OnePlus . . . . .	36
4.4.3	UEFA . . . . .	39
4.4.4	Évora . . . . .	41
4.4.5	Parlamento . . . . .	43

4.4.6	Todas as Questões Reformuladas (Inglês)	45
4.4.7	Todas as Questões Reformuladas (Português)	46
4.4.8	FAQs com múltiplos Domínios	47
<b>5</b>	<b>Caso de Estudo - Instituto da Vinha e do Vinho</b>	<b>49</b>
5.1	Passo 1 - Seleção da melhor resposta	49
5.2	Passo 2 - Seleção da melhor variante	50
5.3	Passo 3 - Resultados	50
5.3.1	Instituto da Vinha e do Vinho - Questões Não Relacionadas	50
5.3.2	Instituto da Vinha e do Vinho - Questões Reformuladas	51
5.3.3	Instituto da Vinha e do Vinho- Questões Originais	52
5.4	Comparação de resultados	52
<b>6</b>	<b>Conclusões e trabalho Futuro</b>	<b>55</b>
6.1	Trabalho futuro	56
<b>A</b>	<b>Funções Auxiliares - Código</b>	<b>59</b>
<b>B</b>	<b>Chatbot "Pergunta e Resposta - Remove as Stopwords" - Código</b>	<b>63</b>
<b>C</b>	<b>Covid - Inglês</b>	<b>79</b>
<b>D</b>	<b>OnePlus - Inglês</b>	<b>87</b>
<b>E</b>	<b>UEFA - Inglês</b>	<b>93</b>
<b>F</b>	<b>Évora - Português</b>	<b>99</b>
<b>G</b>	<b>Parlamento - Português</b>	<b>107</b>
<b>H</b>	<b>Instituto da Vinha e do Vinho - Português</b>	<b>113</b>
<b>I</b>	<b>Resultados "Resposta - Remove as Stopwords"</b>	<b>119</b>
I.1	Covid	120
I.1.1	Covid - Questões Não Relacionadas	120
I.1.2	Covid - Questões Reformuladas	120
I.1.3	Covid - Questões Originais	121
I.2	OnePlus	122
I.2.1	OnePlus - Questões Não Relacionadas	122
I.2.2	OnePlus - Questões Reformuladas	122

I.2.3	OnePlus - Questões Originais . . . . .	123
I.3	UEFA . . . . .	124
I.3.1	UEFA - Questões Não Relacionadas . . . . .	124
I.3.2	UEFA - Questões Reformuladas . . . . .	124
I.3.3	UEFA - Questões Originais . . . . .	125
I.4	Évora . . . . .	126
I.4.1	Évora - Questões Não Relacionadas . . . . .	126
I.4.2	Évora - Questões Reformuladas . . . . .	126
I.4.3	Évora - Questões Originais . . . . .	127
I.5	Parlamento . . . . .	128
I.5.1	Parlamento - Questões Não Relacionadas . . . . .	128
I.5.2	Parlamento - Questões Reformuladas . . . . .	128
I.5.3	Parlamento- Questões Originais . . . . .	129
I.6	Todas as Questões Reformuladas (Inglês) . . . . .	130
I.6.1	Todas as Questões Reformuladas (Inglês) . . . . .	130
I.7	Todas as Questões Reformuladas (Português) . . . . .	131
I.7.1	Todas as Questões Reformuladas (Português) . . . . .	131
<b>J</b>	<b>Resultados "Resposta - Não remove as Stopwords"</b>	<b>133</b>
J.1	Covid . . . . .	134
J.1.1	Covid - Questões Não Relacionadas . . . . .	134
J.1.2	Covid - Questões Reformuladas . . . . .	134
J.1.3	Covid - Questões Originais . . . . .	135
J.2	OnePlus . . . . .	136
J.2.1	OnePlus - Questões Não Relacionadas . . . . .	136
J.2.2	OnePlus - Questões Reformuladas . . . . .	136
J.2.3	OnePlus - Questões Originais . . . . .	137
J.3	UEFA . . . . .	138
J.3.1	UEFA - Questões Não Relacionadas . . . . .	138
J.3.2	UEFA - Questões Reformuladas . . . . .	138
J.3.3	UEFA - Questões Originais . . . . .	139
J.4	Évora . . . . .	140
J.4.1	Évora - Questões Não Relacionadas . . . . .	140
J.4.2	Évora - Questões Reformuladas . . . . .	140
J.4.3	Évora - Questões Originais . . . . .	141

J.5	Parlamento . . . . .	142
J.5.1	Parlamento - Questões Não Relacionadas . . . . .	142
J.5.2	Parlamento - Questões Reformuladas . . . . .	142
J.5.3	Parlamento- Questões Originais . . . . .	143
J.6	Todas as Questões Reformuladas (Inglês) . . . . .	144
J.6.1	Todas as Questões Reformuladas (Inglês) . . . . .	144
J.7	Todas as Questões Reformuladas (Português) . . . . .	145
J.7.1	Todas as Questões Reformuladas (Português) . . . . .	145

<b>Bibliografia</b>		<b>147</b>
---------------------	--	------------



# Lista de Figuras

3.1	Python - lista das Bibliotecas utilizadas . . . . .	12
3.2	Regular Expression . . . . .	13
3.3	NLTK Tokenize - Exemplo . . . . .	13
3.4	NLTK Corpus - Exemplo . . . . .	13
3.5	NLTK Metrics - Levenshtein Distance . . . . .	14
3.6	NLTK Metrics - Jaccard Distance . . . . .	14
3.7	NLTK Stem - Stemming . . . . .	14
3.8	NLTK Stem - Lemmatization . . . . .	15
3.9	NumPy - ArgSort . . . . .	15
3.10	Arquitetura do Sistema . . . . .	17
3.11	Arquitetura da Seleção de Resposta . . . . .	19
4.1	FAQs Covid - As stopwords são removidas . . . . .	35
4.2	FAQs Covid - As stopwords não são removidas . . . . .	36
4.3	FAQs OnePlus - As stopwords são removidas . . . . .	37
4.4	FAQs OnePlus - As stopwords não são removidas . . . . .	38
4.5	FAQs UEFA - As stopwords são removidas . . . . .	39
4.6	FAQs UEFA - As stopwords não são removidas . . . . .	40
4.7	FAQs Évora - As stopwords são removidas . . . . .	41
4.8	FAQs Évora - As stopwords não são removidas . . . . .	42
4.9	FAQs Parlamento - As stopwords são removidas . . . . .	43
4.10	FAQs Parlamento - As stopwords não são removidas . . . . .	44
4.11	Todas as Questões Reformuladas (Inglês) - As stopwords são removidas . . . . .	45

4.12	Todas as Questões Reformuladas (Inglês) - As stopwords são removidas . . . . .	45
4.13	Todas as Questões Reformuladas (Português) - As stopwords são removidas . . . . .	46
4.14	Todas as Questões Reformuladas (Português) - As stopwords são removidas . . . . .	47
4.15	FAQs com Múltiplos Domínios - As stopwords são removidas . . . . .	48
4.16	FAQs com Múltiplos Domínios - As stopwords não são removidas . . . . .	48
5.1	Instituto da Vinha e do Vinho - "Pergunta e Resposta - Remove as stopwords - Questões Não Relacionadas" . . . . .	50
5.2	Instituto da Vinha e do Vinho - "Pergunta e Resposta - Remove as stopwords - Questões Reformuladas" . . . . .	51
5.3	Instituto da Vinha e do Vinho- "Pergunta e Resposta - Remove as stopwords - Questões Originais" . . . . .	52
5.4	Questões Reformuladas - Resultados de cada Domínio . . . . .	53
I.1	Covid - "Resposta - Remove as stopwords - Questões Não Relacionadas" . . . . .	120
I.2	Covid - "Resposta - Remove as stopwords - Questões Reformuladas" . . . . .	120
I.3	Covid - "Resposta - Remove as stopwords - Questões Originais" . . . . .	121
I.4	OnePlus - "Resposta - Remove as stopwords - Questões Não Relacionadas" . . . . .	122
I.5	OnePlus - "Resposta - Remove as stopwords - Questões Reformuladas" . . . . .	122
I.6	OnePlus - "Resposta - Remove as stopwords - Questões Originais" . . . . .	123
I.7	UEFA - "Resposta - Remove as stopwords - Questões Não Relacionadas" . . . . .	124
I.8	UEFA - "Resposta - Remove as stopwords - Questões Reformuladas" . . . . .	124
I.9	UEFA - "Resposta - Remove as stopwords - Questões Originais" . . . . .	125
I.10	Évora - "Resposta - Remove as stopwords - Questões Não Relacionadas" . . . . .	126
I.11	Évora - "Resposta - Remove as stopwords - Questões Reformuladas" . . . . .	126
I.12	Évora - "Resposta - Remove as stopwords - Questões Originais" . . . . .	127
I.13	Parlamento - "Resposta - Remove as stopwords - Questões Não Relacionadas" . . . . .	128
I.14	Parlamento - "Resposta - Remove as stopwords - Questões Reformuladas" . . . . .	128
I.15	Parlamento - "Resposta - Remove as stopwords - Questões Originais" . . . . .	129
I.16	Todas as Questões Reformuladas (Inglês) - "Resposta - Remove as stopwords" . . . . .	130
I.17	Todas as Questões Reformuladas (Português) - "Resposta - Remove as stopwords" . . . . .	131
J.1	Covid - "Resposta - Não remove as stopwords - Questões Não Relacionadas" . . . . .	134
J.2	Covid - "Resposta - Não remove as stopwords - Questões Reformuladas" . . . . .	134
J.3	Covid - "Resposta - Não remove as stopwords - Questões Originais" . . . . .	135
J.4	OnePlus - "Resposta - Não remove as stopwords - Questões Não Relacionadas" . . . . .	136
J.5	OnePlus - "Resposta - Não remove as stopwords - Questões Reformuladas" . . . . .	136



J.6	OnePlus - "Resposta - Não remove as stopwords - Questões Originais" . . . . .	137
J.7	UEFA - "Resposta - Não remove as stopwords - Questões Não Relacionadas" . . . . .	138
J.8	UEFA - "Resposta - Não remove as stopwords - Questões Reformuladas" . . . . .	138
J.9	UEFA - "Resposta - Não remove as stopwords - Questões Originais" . . . . .	139
J.10	Évora - "Resposta - Não remove as stopwords - Questões Não Relacionadas" . . . . .	140
J.11	Évora - "Resposta - Não remove as stopwords - Questões Reformuladas" . . . . .	140
J.12	Évora - "Resposta - Não remove as stopwords - Questões Originais" . . . . .	141
J.13	Parlamento - "Não remove - Remove as stopwords - Questões Não Relacionadas" . . . . .	142
J.14	Parlamento - "Não remove - Remove as stopwords - Questões Reformuladas" . . . . .	142
J.15	Parlamento - "Não remove - Remove as stopwords - Questões Originais" . . . . .	143
J.16	Todas as Questões Reformuladas (Inglês) - "Resposta - Não remove as stopwords" . . . . .	144
J.17	Todas as Questões Reformuladas (Português) - "Não remove - Remove as stopwords" . . . . .	145



# Lista de Tabelas

5.1 Domínios das FAQ - Melhor Método e Pior Método . . . . .	53
--	----



# Glossário

**Array** Estrutura de Dados utilizada para armazenar informação

**Chatbot** Programa de computador que tenta interpretar o que os seres Humanos dizem, procurando responder de forma adequada

**Open Source** Modelo de desenvolvimento em que não existem licenças comerciais associadas

**Retrieval-Based** Sistemas suportados por um corpus que serve de base de dados

**Rule-Based** Sistemas que funcionam com base em regras e ações

**Stopwords** Também denominadas palavras vazias, são palavras que não acrescentam muito significado ao texto, podendo ser removidas

**Tokenização** Processo de converter uma sequência de caracteres numa sequência de tokens



# Sumário

Hoje em dia, seja qual for o sítio que visitamos na Internet, dificilmente não nos deparamos com um Chatbot: desde as pequenas caixas de diálogo que nos assistem em compras online, até aos agentes virtuais que executam os pedidos que fazemos através de comandos de voz. Nesta dissertação, o objetivo principal é a implementação de um Chatbot para responder a perguntas frequentes. Para tal, foi realizado um estudo sobre o Processamento de Linguagem Natural, formas de preparar o texto para que fique pronto a ser utilizado pelo computador, distâncias que podem ser aplicadas a texto, similaridade semântica e, por fim, qual o impacto da remoção das palavras vazias. Após a realização deste estudo e obtenção da informação necessária, foi implementado um Chatbot para responder a perguntas frequentes, independentemente do tema dessas perguntas. Esse Chatbot foi, posteriormente, testado em perguntas frequentes relacionadas com o Instituto da Vinha e do Vinho.

**Palavras chave:** Processamento de Linguagem Natural, Chatbot, Perguntas frequentes, Python, Palavras vazias, Stemização, Lematização, Similaridade de Cosine, Distância de Levenshtein, Distância de Jaccard, Similaridade Semântica





# Abstract

## Chatbots development to answer frequently asked questions

Nowadays, whatever the site we visit on the Internet, we hardly encounter a Chatbot: from the small dialog boxes that assist us in online shopping to the virtual agents that execute the orders we make through voice commands. In this dissertation, the main goal is to implement a Chatbot to answer frequently asked questions. With this in mind, a study was carried out on natural language processing, ways to prepare the text so that it is ready for use by the computer, distances that can be applied to text, semantic similarity and, finally, what is the impact of removing stopwords. After conducting this and obtaining the necessary information, a Chatbot was implemented to answer frequently asked questions, regardless of the domain of those questions. This Chatbot was subsequently tested on frequently asked questions related to the Institute of Vine and Wine.

**Keywords:** Natural Language Processing, Chatbot, Frequently asked questions, Python, Stopwords, Stemming, Lemmatization, Cosine Similarity, Levenshtein Distance, Jaccard Distance, Semantic Similarity



# 1

## Introdução

### 1.1 Motivação

Apesar de já ter algumas décadas de história, o termo Chatbot (ou Agente de Diálogo) nunca esteve tão presente no nosso quotidiano. De fato, no nosso dia a dia, seja através de uma simples navegação na Internet ou de um pedido de suporte para resolver algum problema que tenhamos tido com uma encomenda, dificilmente não nos cruzamos com Chatbots que, de alguma forma, dependendo do foco da nossa navegação na Internet e das nossas pretensões, têm como finalidade facilitar a nossa vida.

Dado o crescimento significativo deste tipo de software, a motivação para esta dissertação reside na curiosidade sobre a história dos Chatbots, que tipos de Chatbots existem, como são implementados e, atualmente, quais os maiores desafios que enfrentam.

### 1.2 Objetivos

De forma a explorar o tópico acima mencionado (na "Motivação") e os conteúdos inerentes mais significativos, o objetivo desta dissertação é a implementação de um Chatbot que consiga responder, com o maior

sucesso possível, a perguntas frequentes, também conhecidas como FAQs.

Pretende-se, com a implementação, verificar qual a fiabilidade de determinadas técnicas e o respetivo desempenho quando as colocamos em prática, através da análise dos resultados obtidos.

Após a obtenção dos resultados e com base nos mesmos, será proposta uma implementação que possa ser utilizada nas FAQs do Instituto da Vinha e do Vinho.

### 1.3 Plano da Tese

No segundo capítulo, "Estado da Arte", vai ser possível ver a história de alguns Chatbots, desde os que apareceram no início até alguns dos mais recentes. Também vai ser introduzida uma divisão dos Chatbots em tipos, de forma que o sistema desenvolvido possa ser enquadrado futuramente. Além disto, vão ser mencionadas algumas técnicas relacionadas com o Processamento de Linguagem Natural, sendo que algumas delas serão, certamente, utilizadas na implementação do sistema.

Quanto ao terceiro capítulo, "Desenvolvimento", é possível verificar o enquadramento do sistema implementado, de acordo com a divisão feita no Estado da Arte. Também podemos ver todas as escolhas que existiram ao nível do software, desde a linguagem de Programação ao editor de código. De seguida, é possível observar a constituição da Arquitetura do Sistema, de forma a que tenhamos uma imagem teórica geral do comportamento que existe na prática. Depois, à luz das partes que constam na Arquitetura do Sistema, vai ser feita uma descrição individual de cada uma dessas partes. Além disto, também vamos poder ver qual a constituição da implementação do sistema, seguida de um caso específico, mais detalhado, dessa mesma implementação.

Relativamente ao quarto capítulo, "Avaliação do Sistema", é possível observar a descrição das FAQs que foram utilizadas para testar o sistema. Também podemos verificar a descrição de todas as experiências realizadas, seguida dos métodos de avaliação que foram postos em prática nessas experiências. Além disto, vamos poder observar os resultados que foram obtidos.

Em relação ao quinto capítulo, "Caso de Estudo - Instituto da Vinha e do Vinho", consoante os resultados obtidos no capítulo da "Avaliação do Sistema", vamos poder verificar qual foi a abordagem escolhida para testar nas FAQs relacionadas com o Instituto da Vinha e do Vinho, e quais foram os resultados da utilização dessa abordagem, comparando-a com os resultados obtidos no capítulo da "Avaliação do Sistema".

Quanto ao sexto capítulo, "Conclusões e trabalho futuro", é possível verificar uma contextualização do que foi feito ao longo desta dissertação e quais as conclusões que retiradas do trabalho desenvolvido e dos resultados obtidos. Além disto, podemos observar também o trabalho futuro, onde são feitas sugestões que, teoricamente, poderão melhorar o desempenho do sistema.

Por fim, existem também alguns anexos que podem ser consultados de forma complementar, tais como: o código utilizado na implementação, as FAQs utilizadas, e alguns resultados que não foram colocados no capítulo da "Avaliação do Sistema".



# 2

## Estado da Arte

Apesar do termo Chatbot ter surgido apenas em 1994, proposto por Michael Mauldin [Ori94], através da junção das palavras "chatter" (a pessoa que conversa) e "bot" (abreviatura de robot), o conceito de ter um bot a interagir com um humano já cimentava a sua história desde os anos 50.

Foi precisamente em 1950, com a publicação do artigo "Computer Machinery and Intelligence" [TUR50], que Alan Turing trouxe ao mundo a questão "Can machines think?". Hoje em dia, o trabalho desenvolvido por Turing, que visava testar a capacidade das máquinas em ter comportamentos inteligentes equivalentes aos de um ser Humano, é conhecido como Turing Test [Har06].

Desde então, muitos foram os Chatbots [AFHC16] que deixaram a sua marca nesta história, entre eles: ELIZA, PARRY, A.L.I.C.E., Jabberwacky, Mitsuku, Dr. Sbaitso, Smarter Child, Siri, IBM Watson, Google Now, Amazon Alexa, Google Allo, Tay, Xiaoice.

É de referir que existe uma competição anual, denominada Loebner Prize [Pow98], no ramo da Inteligência Artificial, cujo propósito é premiar os programas que "pareçam mais humanos". Isto é, quanto maior for a probabilidade de um humano confundir um programa com outro humano, maior será a probabilidade desse programa ganhar o prémio Loebner Prize.

## 2.1 História dos Chatbots

ELIZA [Wei66] foi um software desenvolvido por Joseph Weizenbaum, em 1966, no laboratório de Inteligência Artificial do MIT. Este Chatbot tinha como propósito simular conversas básicas entre máquina e Humano através do reconhecimento de padrões.

PARRY [Col75] foi um software desenvolvido pelo Psiquiatra Kenneth Colby, em 1972, na Universidade de Standford. Este Chatbot tinha como propósito simular uma pessoa com Esquizofrenia através de uma estratégia de conversa, o que o tornava mais avançado que o Chatbot ELIZA.

Dr. Sbaits [Sti19] foi um software desenvolvido em 1991 pela empresa Creative Labs. O propósito deste Chatbot era comportar-se como Psicólogo nas conversas que tinha com os utilizadores. No entanto, este Chatbot não conseguia interagir com os utilizadores de forma complexa.

A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) [Wal09] foi um software desenvolvido por Richard Wallace, em 1995, na Universidade de Lehigh. Este Chatbot tinha como propósito melhorar a implementação do Chatbot ELIZA, isto é, caso fosse feita uma pergunta que não fosse reconhecida, Richard Wallace iria adicionar uma resposta para essa pergunta.

Jabberwacky [Rol97] foi um software desenvolvido pelo programador Rollo Carpenter, em 1997. Este Chatbot tinha como propósito simular conversas humanas através de entretenimento e humor, sendo que foi o primeiro Chatbot a tentar incorporar interação através de voz.

Smarter Child [Wik20c] foi um software desenvolvido em 2001 pela empresa ActiveBuddy (hoje chamada Colloquis). O propósito deste Chatbot era permitir que os utilizadores tivessem acesso a notícias e a outras informações em tempo real, tendo sido utilizado em plataformas como o MSN Messenger e o Yahoo Messenger.

Mitsuku [CA20] foi um software desenvolvido por Steve Worswick, em 2005. O propósito deste Chatbot não é fazer-se passar por (ou substituir) um ser Humano, mas sim estar disponível para conversar com quem precisar. Este Chatbot está disponível em plataformas como o Facebook Messenger, Instagram, Telegram, entre outras.

Watson [FLB<sup>+</sup>13] foi um software desenvolvido em 2006 pela IBM, na sequência do projeto DeepQA. O propósito deste Chatbot era responder a perguntas de um concurso televisivo Americano, o "Jeopardy!". Após ter sido bem sucedido nesse projeto inicial, o Chatbot acabou por ser alargado a outras indústrias.

Siri [Wik20b] foi um software desenvolvido em 2010, como uma aplicação para IOS, tendo sido adquirida pela Apple dois meses depois do seu lançamento. O seu principal propósito é responder a questões dos utilizadores e fazer recomendações. Este software é um dos mais famosos do mercado, permitindo aos utilizadores vários tipos de interações e funcionalidades.

Google Now [Tha16] foi um software desenvolvido em 2012 pela Google. O propósito deste software era ser uma funcionalidade para auxiliar os utilizadores na pesquisa, tanto em dispositivos Android como em dispositivos IOS, fazendo recomendações com base em pesquisas anteriores feitas pelos utilizadores.

Xiaoice [ZGLS18] foi um software desenvolvido em 2014, pela Microsoft. O propósito deste software era conversar com os utilizadores, tendo em conta um quociente emocional. Desde então o Chatbot já teve várias versões, tendo adquirido outras funcionalidades como reconhecimento de imagem, noção de contexto, entre outras funcionalidades.

Amazon Alexa [LRK<sup>+</sup>19] foi um software desenvolvido em 2014 pela Amazon. O propósito deste software é funcionar como um Assistente Virtual do utilizador, permitindo ao mesmo tempo desempenhar inúmeras

funções através do sistema, sendo que todos os anos são adicionadas novas funcionalidades que revolucionam o dia a dia do utilizador graças à inovação e à utilidade que essas funcionalidades acrescentam à vida de quem usufrui delas. Um exemplo disso é a facilidade com que podemos utilizar comandos de voz para diversas ações (desligar a luz, ligar para alguém, pedir a tradução de uma palavra, entre outros).

Google Allo [Wik20a] foi um software desenvolvido em 2016 pela Google. O propósito deste software era permitir a troca de mensagens entre os utilizadores, utilizando como identificador o número de telemóvel, sugerindo respostas a mensagens recebidas. A sugestão de resposta não se limitava a mensagens textuais, isto é, se o utilizador recebe como resposta uma imagem, também era gerada uma sugestão de resposta.

Tay [Dav16] foi um software desenvolvido em 2016, via Twitter, pela Microsoft. O propósito deste software era ser capaz de ter conversas casuais através da aprendizagem que ia adquirindo de conversas anteriores. Contudo, este Chatbot acabou por se tornar polémico por ter aprendido e replicado Tweets ofensivos, tendo sido desligado no mesmo ano em que foi trazido a público.

Zo [Wik20d] foi um Software desenvolvido em 2016, pela Microsoft, como sendo a versão em Inglês do Chatbot Xiaoice. O propósito deste Chatbot foi substituir o seu antecessor, Tay, após a polémica já mencionada. Em 2019, o Chatbot Zo acabou por ser descontinuado na maior parte das plataformas onde se encontrava inserido.

Como é possível verificar, desde o aparecimento dos primeiros Chatbots, foram notórias as modificações que aconteceram no mundo e, conseqüentemente, levaram a que este tipo de tecnologia fosse cada vez mais complexa. O que não mudou, desde então, foi a infinidade de possibilidades e utilidades que podemos dar a este tipo de software, desde uma simples conversa para saber qual a temperatura numa determinada zona do país até ao controlo das atividades do dia a dia numa casa inteligente.

## 2.2 Tipos de Chatbots

De forma geral, dependendo do seu propósito e do âmbito onde se encontram inseridos, os Chatbots têm de ser capazes de analisar o que o utilizador diz, interpretar o significado da análise que fizeram e, por fim, dar uma resposta, o mais adequada possível, ao utilizador.

Os Chatbots podem ser categorizados relativamente ao domínio da conversa que conseguem ter com o utilizador e relativamente à forma como fazem a geração da resposta que vão dar ao utilizador.

### 2.2.1 Domínio da Conversa

Quanto ao Domínio da Conversa, os Chatbots podem ter Domínio Fechado ou Domínio Aberto.

Quando falamos num Chatbot que está limitado ao tipo de conversa que pode ter, falamos num Chatbot de Domínio Fechado [NSP19], também conhecido como Agente de Diálogo. As respostas deste tipo de Chatbot vão ser sempre restritas a um determinado âmbito como, por exemplo, ajudar os utilizadores a fazer reservas num restaurante.

Depois temos os Chatbots de Domínio Aberto [RDG<sup>+</sup>20], também conhecidos como Agentes Conversacionais. Este tipo de Chatbot não está restrito a um assunto específico, pelo que consegue, junto do utilizador, explorar um amplo número de tópicos, por exemplo, servir de "amigo" ou terapeuta para os utilizadores, ajudando no combate à solidão.

Olhando para estes dois domínios, os Chatbots de Domínio Fechado tendem a ser mais eficientes, isto é,



como são desenvolvidos para lidar com um assunto específico, têm mais sucesso a desempenhar as suas funções. Quanto aos Chatbots de Domínio Aberto, tendem a ser mais complexos e muitos mais difíceis de avaliar, uma vez que não se focam num assunto em particular.

### 2.2.2 Geração da Resposta

Quanto à Geração de Resposta, existem três Modelos principais, sendo que os Chatbots podem ser Rule-Based, Retrieval-Based, ou Generativos.

Os Chatbots são considerados Rule-Based [EMR18] quando as respostas estão totalmente predefinidas e são retornadas ao utilizador de acordo com um conjunto de regras. Um exemplo de software que se enquadra neste âmbito é o Chatbot ELIZA.

Os Chatbots são considerados Retrieval-Based [JLL14] quando as respostas dadas ao utilizador são retiradas de um Corpus de diálogos existente. Existem várias técnicas para interpretar as intenções do utilizador e, após essa interpretação, o Chatbot devolve a melhor resposta possível, dentro do Corpus de diálogos disponível.

Os Chatbots são considerados Generativos [KD20] quando as respostas dadas ao utilizador são geradas pelo próprio Chatbot, não estando dependentes de uma Base de Dados com respostas predefinidas. Contudo, a implementação destes Chatbots é muito mais complexa e requer uma abundância relativa de dados. Além disto, se não forem bem treinados, estes Chatbots podem dar respostas sem sentido e com erros gramaticais.

Independentemente da forma como é feita a Geração de Resposta, estes Chatbots partilham a mesma base: analisar o que o utilizador diz, interpretar essa análise e, por fim, dar uma resposta.

É de salientar que é possível utilizar combinações de diferentes Modelos de forma a produzir resultados tão ótimos quanto possível.

## 2.3 Técnicas Utilizadas

A base do que podemos fazer com os Chatbots, independentemente do seu enquadramento, está no Processamento de Linguagem Natural. Isto é, o que fazemos com os inputs que temos por parte dos utilizadores e de que forma os computadores podem interpretar esses inputs.

### 2.3.1 Pré-processamento do Texto

Um dos primeiros passos para com o texto é proceder à sua limpeza. Isto é, colocar o texto na melhor forma possível para que esteja pronto a ser utilizado num contexto específico, seja ele qual for. De notar que, dado que se trata de texto, este poderá ter vários formatos, pelo que, ao fazer o pré-processamento do texto, teremos de ter isso em conta.

Alguns exemplos de conteúdos comuns que queremos ver removidos de um texto são: caracteres especiais, espaços em branco, formatações especiais, marcas de pontuação, entre outros.

### 2.3.2 Tokenização do Texto

Existem muitas tarefas relacionadas com o Processamento de Linguagem Natural que requerem a tokenização do texto e isso passa por partir o texto em bocados. Esses bocados podem ser letras, palavras completas ou até mesmo frases. Os bocados em que partimos o texto, sejam eles quais forem, são chamados de tokens.

A tokenização pode ser um processo bastante útil quando queremos encontrar quantas palavras tem um texto, ou quantas vezes uma determinada palavra aparece nesse texto e, sabendo quantas vezes uma determinada palavra aparece, pode levar-nos a saber, por exemplo, o sentimento presente no texto e de forma é que quem o escreveu deixou isso transparecer.

### 2.3.3 Normalização do Texto

A normalização do texto também é algo bastante comum no Processamento de Linguagem Natural. A ideia principal da normalização de texto é dar alguma consistência aos dados antes de serem utilizados nas fases seguintes.

Alguns exemplos de normalização de texto são: colocar todo o texto em letras maiúsculas ou minúsculas, remoção das stopwords, aplicar Stemming, aplicar Lemmatization, entre outros.

As stopwords (ou palavras vazias) são palavras filtradas antes ou depois de proceder ao Processamento de Linguagem Natural. De notar que as stopwords podem ser qualquer lista de palavras que queiramos remover de um texto; no entanto, por norma, são as palavras mais comuns do idioma em questão.

Stemming é um processo que tem como objetivo truncar as palavras de forma a chegar ao seu radical. Este processo é utilizado, entre outros, em motores de pesquisa, de forma a melhorar os resultados devolvidos.

Lemmatization, por sua vez, é um processo que tem como objetivo converter as palavras à sua forma de raiz. À semelhança do processo anterior, a Lemmatization também é utilizada de forma a melhorar os resultados obtidos numa pesquisa, tendo também outras utilizações possíveis.

### 2.3.4 Distâncias entre Textos

No Processamento de Linguagem Natural, também é bastante comum representar as palavras como sendo vectores numéricos. Isto permite que possamos encontrar o contexto onde uma determinada palavra se encontra inserida, ou se certas palavras ocorrem em contextos semelhantes. Alguns exemplos de distâncias entre vectores são: distância Euclidiana, distância de Manhattan, similaridade de Cosine, distância de Jaccard, entre outras.

#### Distância Euclidiana

Apesar da distância Euclidiana ter várias fórmulas a si associadas, a equação 2.3.4 expressa a fórmula utilizada na implementação da função "euclidean", que podemos ver na Biblioteca "scipy", do Python [comb].

$$euclidean = \left( \sum (w_i |u_i - v_i|^2) \right)^{\frac{1}{2}}$$

Na fórmula 2.3.4, temos:

- $w$  representa um array, opcional, que contém os pesos para cada valor em  $u$  e  $v$
- $u$  representa um array unidimensional
- $v$  representa um array unidimensional

### Distância de Manhattan

A fórmula que podemos ver em baixo é a fórmula utilizada na implementação da função "cityblock", que podemos ver na Biblioteca "scipy", do Python [Mül18].

$$manhattan = \sum_i |u_i.v_i|$$

Na fórmula em cima, temos:

- $u$  representa um array unidimensional
- $v$  representa um array unidimensional

### Similaridade de Cosine

Na similaridade de Cosine [SGGAtAP14], o objetivo não é saber propriamente a distância entre dois vectores, mas sim o ângulo entre eles. A fórmula da similaridade é a seguinte:

$$similaridade = \cos(\theta) = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Na fórmula em cima, temos:

- $A$  representa um vector não nulo
- $B$  representa um vector não nulo

### Distância de Jaccard

A fórmula que podemos ver de seguida, referente à distância de Jaccard [Kho20], é utilizada na implementação da função "jaccard\_distance", que podemos ver na Biblioteca "nltk", do Python. De notar que, na seguinte fórmula, assumimos a existência de dois objetos, "A" e "B", com atributos binários.

$$jaccard = \frac{A_{01} + A_{10}}{A_{01} + A_{10} + A_{11}}$$

Na fórmula em cima, temos:

- $A_{01}$  representa o número total de atributos onde o atributo de "A" é zero e o atributo de "B" é um
- $A_{10}$  representa o número total de atributos onde o atributo de "A" é um e o atributo "B" é zero
- $A_{11}$  representa o número total de atributos onde "A" e "B" ambos têm o valor um

## 2.4 Problemas dos Chatbots

Como vimos ao longo do Estado da Arte, os Chatbots têm vindo a crescer bastante nas últimas décadas. As aplicações deste tipo de software são quase infinitas, podendo ser adaptados em qualquer área ou sector, facilitando a vida dos utilizadores.

No entanto, apesar de toda a evolução registada até hoje, existem pontos fulcrais que precisamos de ter em conta como, por exemplo, privacidade, considerações de Ética, abuso do utilizador, entre outros.

No que respeita à privacidade [IAV<sup>+</sup>19], o que é que acontece aos dados que os utilizadores partilham durante a conversa? Estarão esses dados protegidos? Os utilizadores devem ser informados sobre estes aspetos antes de partilharem dados com o Chatbot.

E quanto à Ética [KTP<sup>+</sup>19], será que todos os utilizadores têm noção de que estão a interagir com um Chatbot? Antes de existir interação com o utilizador, deve ser claro que quem está do outro lado é um Chatbot e não um ser Humano.

Por fim, o que devemos fazer em situações em que o utilizador é abusivo para com o Chatbot? Será que devemos responder com humor, ignorando esses abusos, ou simplesmente acabar a conversa? Um exemplo disto, já referido, foi o Chatbot Tay, que aprendeu conteúdos ofensivos através dos utilizadores com quem interagiu e replicou esse conteúdo.



# 3

## Desenvolvimento

Através da pesquisa relacionada com o tema, foi possível observar a forma como os Chatbots evoluíram ao longo do tempo, que Chatbots existiram, que Chatbots ainda existem e o seu contributo para, de certa forma, facilitar a vida de quem interage com eles. Também foi feita uma observação das ferramentas e técnicas utilizadas no desenvolvimento de Chatbots.

Pretende-se, ao longo deste capítulo, fazer um enquadramento do Chatbot desenvolvido, mencionar as escolhas feitas, ao nível do software, para o desenvolvimento do Chatbot, explorar a representação que é feita a partir do texto que o utilizador escreve, descrever o funcionamento do sistema, desenhar a arquitetura do sistema implementado e, por fim, dizer que técnicas foram utilizadas para desenvolver o sistema.

### 3.1 Enquadramento do Chatbot

Tal como foi mencionado no Estado da Arte, podemos classificar o Chatbot relativamente ao Domínio da Conversa e à Geração de Resposta.

Quanto ao Domínio da Conversa, podemos dizer que o Chatbot implementado é um sistema de Domínio Fechado. Esta classificação deve-se ao sistema estar restrito a um tema de conversa específico, neste caso

ao tema das FAQs em questão, seja qual for o tema dessas FAQs. Posto isto, o sistema implementado não conseguirá manter uma conversa de âmbito livre, podendo dar respostas sem sentido ou descontextualizadas, nomeadamente se a conversa fugir ao tema das FAQs.

Quanto à Geração de Resposta, podemos dizer que o Chatbot implementado é um sistema Retrieval-Based. Esta classificação deve-se ao sistema estar ligado a um conjunto de FAQs, um Corpus, do qual será retirada a resposta que o sistema considerar mais adequada, tendo em conta a questão do utilizador e a técnica utilizada para seleccionar a resposta mais adequada.

## 3.2 Escolha do Software

Em primeiro lugar, é de salientar que o sistema foi desenvolvido em ambiente Windows [Cor], mais especificamente na versão 10.

Após escolher o ambiente no qual viria a ser desenvolvido o sistema, foi feita uma pesquisa do Software disponível para auxiliar o desenvolvimento. O primeiro passo foi a escolha da Linguagem de Programação, sendo o Python a linguagem escolhida por ser uma das mais adequadas devido a todo o material open source que existe à disposição dos programadores, nomeadamente todo o conjunto de recursos ligados ao Processamento de Linguagem Natural. O segundo passo foi escolher o editor de texto para código, tendo a escolha acabado por recair sobre o Sublime Text.

Relativamente ao Sublime Text [HQ], além de algumas funcionalidades que já se encontram presentes por defeito, existem muitos outros packages que podem ser instalados consoante a necessidade dos utilizadores, o que torna o Sublime Text bastante personalizável. Entre as funcionalidades mais úteis temos a syntax highlighting (de acordo com a Linguagem em questão) e a possibilidade de dividir uma janela em várias colunas, o que permite estar a trabalhar com diferentes ficheiros ao mesmo tempo.

Devido às características mencionadas, apesar de existirem outras opções igualmente válidas, como o Visual Studio Code, entre outros, a escolha do Sublime Text não desiludiu e veio a mostrar-se bastante útil no desenvolvimento do sistema.

### 3.2.1 Python

Python [Foub] é uma linguagem de Programação orientada por objetos de alto nível, conhecida pela sua vasta flexibilidade devido às diferentes áreas em que as suas utilidades podem ser aplicadas, além de que é uma linguagem relativamente fácil de aprender e de utilizar.

Relativamente ao sistema implementado, as Bibliotecas (do Python) mais utilizadas foram as seguintes:

- Regular Expression Operations (também conhecida como "re")
- Natural Language Toolkit (também conhecida como "NLTK")
- NumPy
- Gensim

Figura 3.1: Python - lista das Bibliotecas utilizadas

### 3.2.2 Python Regular Expression

Python re [CS16] é um Módulo que está ligado às Expressões Regulares. Quanto ao sistema implementado, este Módulo foi utilizado para remover tudo o que não fossem palavras e espaços em branco do input do utilizador. Isto é, tem como função remover as marcas de pontuação. Por exemplo, como podemos observar na figura 3.2, as marcas de pontuação foram removidas.

```
Input: Hey! Let me know more about Phones, please.
Output: Hey Let me know more about Phones please
```

Figura 3.2: Regular Expression

### 3.2.3 Python Natural Language Tool Kit

Python nltk [LB02] é um conjunto de Bibliotecas que está ligado ao Processamento de Linguagem Natural, tendo como conteúdo uma grande variedade de programas que facilitaram o desenvolvimento do sistema.

Dentro da documentação do NLTK, podemos verificar que existe uma vasta variedade de packages que nos podem ser úteis quando procuramos ferramentas para Processamento de Linguagem Natural. No desenvolvimento deste sistema foram utilizados quatro desses packages: tokenize, corpus, metrics e stem.

#### Python NLTK - Tokenize

Do Package nltk.tokenize [VJ<sup>+</sup>16] foi importado o Método "word\_tokenize". Este Método tem por objetivo receber numa frase e criar uma lista com todas as palavras nessa frase. Isto pode ser útil para analisar a importância individual de cada palavra ou para efeitos de comparação, por exemplo, após ter uma lista com todas as palavras da frase, verificar se essas palavras existem em determinadas frases. Por exemplo:

```
Input: Hey Let me know more about Phones please
Output: ['Hey', 'Let', 'me', 'know', 'more', 'about', 'Phones', 'please']
```

Figura 3.3: NLTK Tokenize - Exemplo

#### Python NLTK - Corpus

Do Package nltk.corpus [SL20] foi importada uma lista de palavras chamada "stopwords". Esta lista contém as palavras mais comuns de uma determinada língua, neste caso do Inglês e do Português, e é particularmente útil quando queremos fazer uma análise comparativa, por exemplo, entre a questão do utilizador e cada uma das possíveis respostas, em termos de palavras, e queremos remover as palavras mais comuns da língua em questão de forma a que tenham influência mínima no resultado. Por exemplo:

```
Input: ['Hey', 'Let', 'me', 'know', 'more', 'about', 'Phones', 'please']
Output: ['Hey', 'Let', 'know', 'Phones', 'please']
```

Figura 3.4: NLTK Corpus - Exemplo

#### Python NLTK - Metrics

Do Package nltk.metrics [Proa] foram importados os Métodos "edit\_distance" e "jaccard\_distance".



O Método "edit\_distance" tem como objetivo calcular a Distância de Levenshtein [Roo19] entre dois bocados de texto. Esta distância é calculada através do número total de remoções, inserções, ou substituições que sejam precisas para um determinado bocado de texto se tornar em outro. Por exemplo, na imagem que se segue, podemos dizer que a diferença entre "Desert" e "Dessert" é apenas um "s", pelo que a distância será um.

```
Text A: Desert
Text B: Dessert

Distance between Text A and Text B: 1
```

Figura 3.5: NLTK Metrics - Levenshtein Distance

Em relação ao Método "jaccard\_distance", este tem por objetivo calcular a distância de Jaccard entre dois conjuntos (sendo que esses conjuntos podem ser frases, por exemplo). Esta distância é calculada da seguinte forma: obtemos o número de elementos não repetidos de ambos os conjuntos (união), obtemos o número de elementos comuns a ambos os conjuntos (interseção), depois subtraímos os dois valores obtidos e, por fim, dividimos o valor obtido pelo valor da união. Por exemplo, na imagem que se segue, teríamos união de dez elementos e interseção de sete elementos, pelo que seguindo os cálculos a distância será 0.3.

```
Text A: This is a exam
Text B: This is a test

Distance between Text A and Text B: 0.3
```

Figura 3.6: NLTK Metrics - Jaccard Distance

### Python NLTK - Stem

Do Package nltk.stem [BVS<sup>+</sup>17] foram importadas as Classes PorterStemmer e WordNetLemmatizer. Ambos os processos estão relacionados com a normalização de texto.

A Classe PorterStemmer [Prob] é utilizada para aplicar um processo conhecido como Stemming. Podemos dizer que, ao aplicar Stemming a uma palavra, estamos à procura da forma dessa palavra, podendo muitas vezes ficar com algo que não faz muito sentido (no exemplo "this" passa a "thi"). Por exemplo, na imagem que se segue, podemos observar algumas mudanças antes e depois do Stemming ser aplicado.

```
Text before Stemming: ['Lets', 'check', 'Stemming', 'on', 'this', 'phrase']
Text after Stemming: ['let', 'check', 'stem', 'on', 'thi', 'phrase']
```

Figura 3.7: NLTK Stem - Stemming

A Classe WordNetLemmatizer [Proc] é utilizada para aplicar um processo conhecido como Lemmatization. Pode-se dizer que, ao aplicar Lemmatization a uma palavra, estamos à procura do significado da palavra, obtendo sempre algo que faz sentido (ao contrário de quando fazemos Stemming). Por exemplo, na imagem que se segue, podemos observar algumas mudanças antes e depois da Lemmatization ser aplicada.

```
Text before Lemmatization: ['Do', 'you', 'see', 'dogs', 'next', 'to', 'churches']  
Text after Lemmatization: ['Do', 'you', 'see', 'dog', 'next', 'to', 'church']
```

Figura 3.8: NLTK Stem - Lemmatization

### 3.2.4 Python NumPy

Python nymPy [HMvdW<sup>+</sup>20] é uma Biblioteca que contém uma grande quantidade de funções Matemáticas. Para o caso específico do Sistema implementado, a função utilizada foi a "argsort" [coma]. A ideia desta função é devolver uma lista de índices ordenados, seja de forma ascendente ou descendente. Por exemplo, na imagem que se segue, podemos ver a lista de valores que é introduzida, sendo que vai ser retornada uma outra lista com os índices dos valores da lista introduzida, ordenada de forma descendente.

```
Input list: [0.25, 0.75, 0.5, 1]  
  
List of indices (descending order): [3 1 2 0]
```

Figura 3.9: NumPy - ArgSort

### 3.2.5 Python Gensim

Python gensim [SD18] é um conjunto de Bibliotecas que está ligado, entre outros, à deteção de similaridade e modelação de tópicos. Esta Biblioteca é bastante útil no que toca ao Processamento de Linguagem Natural.

Dentro da documentação do Gensim, podemos verificar que existem vários Packages que nos podem ser úteis quando procuramos ferramentas para Processamento de Linguagem Natural. No desenvolvimento deste sistema foram utilizados quatro desses packages: download, corpora, models, e similarities.

#### Python Gensim - Downloader

O Módulo "Downloader" [Rehb] é uma API (Application Programming Interface) que é utilizada para download, tal como o nome indica, mas também para carregamento de conjuntos de dados e de modelos pré treinados. No caso específico deste projeto, o Módulo em questão foi utilizado para carregar o conjunto de dados "glove-wiki-gigaword-50" [PSM14].

#### Python Gensim - Corpora

Do Package corpora foi importado o Módulo "Dictionary" [Sax18]. Este Módulo é utilizado para fazer um mapeamento entre palavras e os seus identificadores inteiros. Se tivermos uma lista com as palavras da questão do utilizador e uma lista de listas com as palavras das possíveis respostas, vai ser construído um Dicionário com todas as palavras, sem repetições, que fazem parte tanto da questão como das respostas. No Dicionário, cada uma dessas palavras vai ser uma chave, sendo que o valor de cada uma dessas chaves é a posição da palavra no Dicionário.

#### Python Gensim - Models

Do Package models foram utilizados os Módulos "TfidfModel" e "WordEmbeddingSimilarityIndex".

O Módulo "TfidfModel" [Rehd] permite criar uma medida estatística sobre a relevância de cada palavra num determinado Corpus, sendo que o nosso Corpus é a questão do utilizador e todas as possíveis respostas. Para este caso específico, o Módulo "TfidfModel", recebe o Dicionário mencionado na secção "Python Gensim - Corpora" para verificar, então, a importância de cada palavra no nosso Corpus.

Já o Módulo "WordEmbeddingSimilarityIndex" [Reha] permite criar um índice de similaridade através de um conjunto de dados que é carregado para esse propósito. Para o caso específico deste projeto, o conjunto de dados utilizado foi o "glove-wiki-gigaword-50", que já foi mencionado na secção "Python Gensim - Downloader".

### Python Gensim - Similarities

Do Package similarities foram utilizados os Módulos "SparseTermSimilarityMatrix" e o "SoftCosineSimilarity".

O Módulo "SparseTermSimilarityMatrix" [Rehc] coloca à disposição de quem o utiliza um conjunto de Classes que permitem lidar com similaridades de termos. No caso específico deste projeto, este Módulo foi utilizado para construir uma matriz de similaridade. Essa matriz foi construída com base nas informações obtidas no conjunto de dados "glove-wiki-gigaword-50" (secções "Python Gensim - Downloader" e "Python Gensim - Models"), no Dicionário (secção "Python Gensim - Corpora"), e na medida estatística sobre cada palavra (secção "Python Gensim - Models").

Em relação ao Módulo "SoftCosineSimilarity" [Reha] permite calcular a Soft Cosine Similarity em relação a um conjunto de documentos (que serão as FAQs utilizadas). Para o caso específico deste projeto, a Soft Cosine Similarity é aplicada utilizando o Dicionário (secção "Python Gensim - Corpora") e a matriz de similaridade (secção "Python Gensim - Similarities").

## 3.3 Arquitetura do Sistema

O sistema desenvolvido está dividido em cinco partes principais: início de conversa, questão, seleção de resposta, resposta, e fim de conversa.

Relativamente ao início de conversa, o objetivo é dar conhecimento ao utilizador de que está a falar com um Chatbot, obter o nome do utilizador e perguntar ao utilizador qual a sua questão.

Quanto à seleção de resposta, o objetivo é selecionar a resposta mais adequada para a questão que o utilizador colocou. Isto é feito através dos Métodos implementados.

Em relação à resposta, o objetivo é dar a resposta selecionada ao utilizador e perguntar se tem mais alguma questão. Caso o utilizador tenha mais questões, voltamos para a seleção de resposta. Caso não tenha mais questões, passamos para o fim de conversa.

Relativamente ao fim de conversa, esta parte do sistema é alcançada quando o utilizador pretende terminar a conversa com o Chatbot.

Na imagem que se segue é possível observar um desenho da arquitetura, de forma a ilustrar melhor o que foi dito até agora.

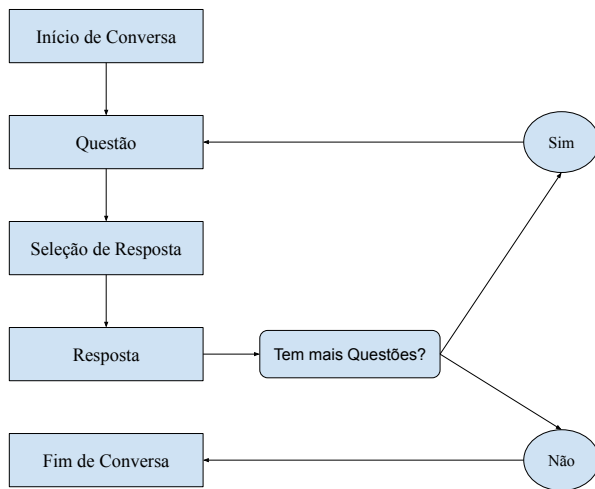


Figura 3.10: Arquitetura do Sistema

## 3.4 Descrição do Sistema

Tal como foi mencionado, o sistema desenvolvido encontra-se dividido em cinco partes: início de conversa, questão, seleção de resposta, resposta, e fim de conversa.

### 3.4.1 Início de Conversa

Quando um utilizador tem uma questão e procura ajuda para a esclarecer, entramos no âmbito do início de conversa.

O início de conversa tem dois objetivos: dar conhecimento ao utilizador de que está a falar com um Chatbot e obter o nome do utilizador.

De notar que, com a obtenção do nome do utilizador, pretende-se tratar o utilizador pelo seu nome de forma a tornar a conversa um pouco mais informal e criar alguma proximidade.

### 3.4.2 Questão

A partir do momento em que o Chatbot obtém o nome do utilizador, entramos no âmbito da questão.

A questão tem dois objetivos: perguntar ao utilizador, tratando-o pelo nome inserido, qual a questão que pretende ver esclarecida e obter a questão que o utilizador coloca.

### 3.4.3 Seleção de Resposta

Assim que o utilizador coloca a questão que pretende ver esclarecida, entramos no âmbito da seleção de resposta.

A seleção de resposta tem dois objetivos: interpretar a questão do utilizador e selecionar uma resposta adequada.

Após interpretar a questão do utilizador, recorrendo a um conjunto de regras descritas com mais pormenores na secção 3.5, é selecionada a resposta que é considerada a mais adequada para a questão que o utilizador colocou.

#### **3.4.4 Resposta**

Após selecionar uma das respostas existente no Corpus, o sistema deve dar essa resposta ao utilizador. Aqui entramos no âmbito da resposta, onde o utilizador pode não ficar esclarecido e querer colocar mais questões, ou pode ficar completamente esclarecido.

Quando o utilizador não fica esclarecido e pretende colocar outra questão, o processo repete-se. Isto é, voltamos para o âmbito da questão, onde o sistema irá obter a nova questão colocada pelo utilizador e, de seguida, passamos para o âmbito da seleção de resposta, onde o sistema irá selecionar a resposta do Corpus que achar mais adequada e, por fim, passar essa resposta para o âmbito da resposta.

Caso o utilizador fique completamente esclarecido, pode informar o Chatbot de que a sua intenção é terminar a conversa. Quando isto acontece, passamos para o âmbito do fim de conversa.

#### **3.4.5 Fim de Conversa**

Relativamente ao fim de conversa, esta parte do sistema é alcançada quando o utilizador ficou esclarecido e pretende terminar a conversa com o Chatbot.

De notar que também se pode dar o caso do utilizador não ter ficado esclarecido, mas é preciso que todas as respostas dadas ao utilizador não tenham cumprido o seu propósito ou tenham sido descontextualizadas.

### **3.5 Cálculo de Medidas**

Como pudemos ver na secção anterior, o sistema está dividido em cinco partes fundamentais. E, apesar de serem todas importantes, nomeadamente pelo papel que cada uma desempenha, a Seleção de Resposta é fulcral para que o utilizador possa ser auxiliado da melhor forma possível.

#### **3.5.1 Arquitetura da Seleção de Resposta**

Como pudemos observar na secção "Arquitetura do Sistema", a Seleção de Resposta é uma das cinco partes em que se encontra dividido o sistema.

Na figura em baixo, podemos verificar o interior da Seleção de Resposta, assumindo que estamos a trabalhar com um grau de similaridade que será a soma de quatro valores. Este valor, quatro, é meramente indicativo, pelo que poderíamos ter mais ou menos, consoante a decisão tomada relativamente à implementação.

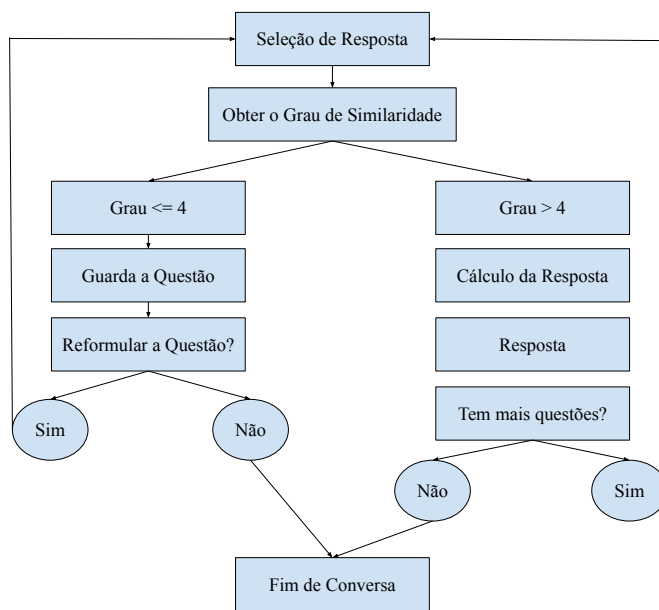


Figura 3.11: Arquitetura da Seleção de Resposta

### 3.5.2 Grau de Similaridade

A Seleção de Resposta procura, numa primeira fase, obter um grau de similaridade entre o input do utilizador e o conteúdo da FAQ utilizada. Esse grau é obtido da seguinte forma:

1. são removidas as stopwords
2. através da comparação literal das palavras, contam-se quantas palavras comuns existem entre o input do utilizador e cada um dos pares "questão-resposta" presente nas FAQ e guarda-se o valor do par "questão-resposta" com mais palavras em comum
3. o processo do ponto anterior é repetido, mas aplica-se a técnica Stemming antes de fazer a comparação e guardar o valor obtido
4. volta-se a repetir o mesmo processo, mas desta vez aplica-se a técnica Lemmatization antes de fazer a comparação e guardar o valor
5. depois aplica-se a técnica Similaridade de Cosine, que não faz a contagem do número literal das palavras como as anteriores, mas de onde é retirada a distância do par "questão-resposta" que mais se aproxima do input do utilizador
6. quando estes quatro valores forem calculados, podemos somá-los de forma a obter o grau de similaridade

Lista 3.1: Algoritmo para obtenção do Grau de Similaridade - Lista de passos

De seguida, caso o valor do grau de similaridade seja superior ao número de somas que temos de fazer para o obter, neste caso quatro, então o input do utilizador é passado a todos os diferentes Métodos (que podem ser vistos na lista "Métodos utilizados para responder ao utilizador", na próxima página) para que cada um deles possa obter a resposta que, de acordo com os seus cálculos específicos, faz mais sentido. Caso o valor do grau de similaridade seja inferior ou igual a quatro, dizemos ao utilizador para reformular a questão.

Antes de avançar para a importância do grau de similaridade, porquê "caso o valor do grau de similaridade seja superior ao número de somas que temos de fazer para o obter"? Porque, para existir um mínimo de similaridade entre o input do utilizador e algum (ou alguns) dos pares "questão-resposta", é expectável que pelo menos uma palavra seja igual. Posto isto, a ideia é verificar se é possível confirmar a existência dessa similaridade antes de devolver uma resposta ao utilizador. Neste caso, a existência dessa similaridade é verificada de quatro formas, daí a soma dos quatro valores, mas isto é facultativo e depende da implementação.

O grau de similaridade é bastante útil para que as respostas apresentadas ao utilizador não sejam descontextualizadas da questão colocada e, apesar de não ser cem por cento eficaz, reduz bastante o número de respostas sem sentido, como vai ser possível verificar no capítulo 4 ("Avaliação do Sistema").

De notar que, nas variantes do Chatbot em que não são removidas as stopwords, não faz sentido existir este grau de similaridade, pelo que o mesmo é removido da implementação (quando as stopwords não são removidas).

### 3.5.3 Cálculo da Resposta

Assumindo que o utilizador coloca uma questão relacionada com o tema das FAQs e que o grau de similaridade é superior ao número de somas que temos de fazer para o obter, a próxima fase é seleccionar a resposta mais adequada à questão do utilizador.

De forma a fazer a seleção da resposta, o input do utilizador é passado a oito Métodos que o vão tentar obter de formas diferentes. Esses Métodos são:

- Básico
- Stemming
- Lemmatization
- Cosine
- Levenshtein - Frase
- Levenshtein - Palavra
- Jaccard
- Semantic Similarity

Lista 3.2: Métodos utilizados para responder ao utilizador

O Básico vai verificar qual dos pares "questão-resposta" tem mais palavras em comum com o input do utilizador, devolvendo a resposta associada a esse par.

O Stemming vai aplicar a técnica Stemming ao input do utilizador e a todos os pares "questão-resposta" da FAQ e verificar qual dos pares tem mais palavras em comum com o input do utilizador, devolvendo a resposta associada a esse par.

A Lemmatization vai aplicar a técnica Lemmatization ao input do utilizador e a todos os pares "questão-resposta" da FAQ e verificar qual dos pares tem mais palavras em comum com o input do utilizador, devolvendo a resposta associada a esse par.

A Similaridade de Cosine vai aplicar a fórmula associada a esta técnica ao input do utilizador e aos pares "questão-resposta". No final, teremos uma lista com todas as distâncias (valores entre zero e um) do input

do utilizador a cada um dos pares "questão-resposta", sendo devolvida a resposta associada ao par que tiver mais perto (mais perto do valor um) do input do utilizador.

A "Levenshtein - Frase" vai verificar a distância de Levenshtein entre o input do utilizador e cada um dos pares "questão-resposta". De notar que esta distância é calculada sobre as frases inteiras, tanto do input como dos pares "questão-resposta". No final, teremos uma lista com todas as distâncias, sendo que a resposta devolvida ao utilizador será a resposta cuja distância estiver mais perto do zero.

A "Levenshtein - Palavra" vai fazer praticamente o mesmo que a distância do ponto anterior. No entanto, em vez de verificar a distância da frase total, faz a verificação palavra a palavra, guardando as distâncias mais curtas entre as palavras do input do utilizador e as palavras dos pares "questão-resposta". No final, teremos uma lista com todas as distâncias, sendo que a resposta devolvida ao utilizador será a resposta com distância mais perto de zero.

O "Jaccard" vai verificar a distância de Jaccard entre o input do utilizador e cada um dos pares "questão-resposta". No final destes cálculos, teremos uma lista com todas as distâncias, com valores entre zero e um, sendo que a resposta devolvida ao utilizador vai a correspondente ao par "questão-resposta" que estiver mais perto de zero.

A "Semantic Similarity" vai comparar input do utilizador e os pares "questão-resposta" com base no quão similares são as palavras em ambos, criando uma matriz de similaridade que vai conter a similaridade entre pares de palavras, ponderada utilizando o modelo TF-IDF. Depois é calculada a Soft Cosine Similarity entre o input do utilizador e cada par, sendo que cada um destes valores, compreendidos entre zero e um, será colocado numa lista. A resposta devolvida será aquela cujo valor estiver mais perto de um.

## 3.6 Constituição da implementação do Sistema

Até agora, como podemos observar em secções anteriores, vimos quais as escolhas feitas relativamente ao software utilizado e, depois disso, qual a arquitetura do sistema. Posto isto, falta ver o que é que faz parte da implementação, na prática.

Existem três partes que são consideradas fundamentais para que o sistema funcione: o Chatbot em si, algumas funções auxiliares que servem de apoio ao Chatbot e, por fim, as FAQs que vão servir para testar o sistema.

### 3.6.1 Chatbot - Variantes

Relativamente ao Chatbot, existem quatro implementações do mesmo. Apesar de maior parte do código ser igual, foi decidido que se iriam criar estas quatro variantes de forma a conseguir obter uma melhor organização quando fosse o momento de testar cada uma delas. De notar que todas as variantes são diferentes na Seleção de Resposta, ou seja, após o utilizador colocar a questão. Podemos distingui-las da seguinte forma:

- Resposta - Remove as Stopwords
- Resposta - Não Remove as Stopwords
- Pergunta e Resposta - Remove as Stopwords
- Pergunta e Resposta - Não Remove as Stopwords

Lista 3.3: Chatbot - Variantes



Para mais informações sobre as stopwords, é possível consultar a subsecção "Python NLTK - Corpus" (no capítulo do Desenvolvimento, secção "Python Natural Language Tool Kit").

Na primeira implementação, "Resposta - Remove as Stopwords", após obter a questão que o utilizador coloca, tenta responder ao utilizador apenas com base nas respostas presentes na FAQ que estiver a ser utilizada, ignorando as questões da FAQ. De notar que, nesta implementação, são removidas as stopwords.

Quanto à segunda implementação, "Resposta - Não Remove as Stopwords", após obter a questão que o utilizador coloca, também tenta responder ao utilizador apenas com base nas respostas das FAQ. No entanto, nesta implementação não são removidas as stopwords.

Relativamente à terceira implementação, "Pergunta e Resposta - Remove as Stopwords", após obter a questão que o utilizador coloca, tenta responder ao utilizador com base tanto nas perguntas como nas respostas presentes na FAQ que estiver a ser utilizada. Nesta implementação são removidas as stopwords.

Em relação à última implementação, "Pergunta e Resposta - Não Remove as Stopwords", após obter a questão que o utilizador coloca, tenta responder com base tanto nas perguntas como nas respostas presentes na FAQ. No entanto, nesta implementação, não são removidas as stopwords.

### 3.6.2 Chatbot - Funções Auxiliares

Tal como foi referido, as funções auxiliares são uma das partes fundamentais para que o sistema funcione. Existem dois ficheiros com funções auxiliares: um para os Chatbots em que as stopwords são removidas, outro para os Chatbots em que as stopwords não são removidas. A única diferença entre estes dois ficheiros está relacionada com as stopwords serem ou não removidas, sendo o restante conteúdo dos ficheiros igual.

Apesar das funções auxiliares servirem de apoio ao Chatbot, é de notar que, inicialmente, a existência destas funções auxiliares não estava prevista, a criação destes ficheiros surgiu com a necessidade de diminuir a complexidade do código escrito no ficheiro do Chatbot, fazendo cálculos auxiliares. De forma a utilizar as funções auxiliares, basta importá-las dentro do ficheiro do Chatbot.

O ficheiro das funções auxiliares é composto por seis funções. Todo o código dessas funções pode ser visto em anexo (no anexo A).

#### Função "preprocessing"

A primeira função, "preprocessing", tem como parâmetro o input do utilizador e, tal como o nome indica, tem como propósito fazer o pré-processamento desse input. Esta função coloca todo o input do utilizador em minúsculas, remove as marcas de pontuação existentes, remove (ou não, dependendo da variante do Chatbot) as stopwords e, por fim, retorna o input com o pré-processamento feito. Esta função vai auxiliar todos os Métodos do Chatbot que necessitarem de pré-processar o input do utilizador. O código desta função pode ser visto em baixo.

---

```
#User input preprocessing
def preprocessing(user_input):
    user_input = user_input.lower()
    user_input = re.sub(r'[\W\s]', "", user_input)

    tokens = word_tokenize(user_input)

    without_stop_words = {i for i in tokens if i not in stop_words}
```

```
return without_stop_words
```

---

Listing 3.1: Função "preprocessing"

### Função "similar\_words"

Já a segunda função, "similar\_words", tem como parâmetros o input do utilizador e uma possível resposta. Esta função vai procurar palavras comuns entre o input do utilizador e a possível resposta e, cada vez que encontra uma palavra comum entre ambas, incrementa o valor de uma variável numa unidade. No final, retorna o valor de palavras comuns. Esta função vai auxiliar um Método específico do Chatbot que serve para selecionar a resposta que vai ser devolvida ao utilizador. O código desta função pode ser visto em baixo.

```
#Check similar words
def similar_words(user_input, possible_answer):
    similar_words = 0
    for word in user_input:
        if(word in possible_answer):
            similar_words += 1

    return similar_words
```

---

Listing 3.2: Função "similar\_words"

### Função "product\_between\_lists"

No que toca à terceira função, "product\_between\_lists", tem como parâmetros duas listas. Esta função tem por objetivo fazer o produto do valor existente em cada posição da primeira lista pelo valor que está exatamente na mesma posição da segunda lista. Todos produtos são somados e guardados numa variável, sendo que a função retorna a soma final desses produtos. Esta função vai auxiliar um Método específico do Chatbot que serve para selecionar uma resposta utilizando a Distância de Cosine. O código desta função pode ser visto em baixo.

```
#Product between lists
def product_between_lists(list_one, list_two):
    product_sum = 0
    for i in range(len(list_one)):
        product_sum += list_one[i] * list_two[i]

    return product_sum
```

---

Listing 3.3: Função "product\_between\_lists"

### Função "sum\_lists\_in\_list"

Relativamente à quarta função, "sum\_lists\_in\_list", tem como parâmetro uma lista (lista exterior) de listas (listas interiores). Por cada uma das listas interiores, os valores em cada posição vão ser somados,

sendo a soma final dos valores nessa lista acrescentada a uma lista auxiliar. Ou seja, se no início tivermos uma lista exterior com três listas interiores, a função irá retornar uma lista com três posições, estando cada uma dessas posições ocupadas com os valores da soma de cada uma das listas interiores. Esta função vai auxiliar um Método específico do Chatbot que serve para selecionar uma resposta utilizando a Distância de Cosine. O código desta função pode ser visto em baixo.

---

```
#Sum lists (inner) inside list (outer)
def sum_lists_in_list(list_one):
    final_sum = []

    for inner_list in list_one:
        aux = 0
        for number in inner_list:
            aux += number
        final_sum.append(aux)

    return final_sum
```

---

Listing 3.4: Função "sum\_lists\_in\_list"

### Função "words\_edit\_distance"

Em relação à quinta função, "words\_edit\_distance", tem como parâmetros duas listas. Esta função tem por objetivo verificar a Distância de Levenshtein (secção "Python NLTK - Metrics") entre todas as palavras da primeira e da segunda lista. Por cada palavra que estiver na primeira lista, vai ser verificada a distância entre essa palavra e cada uma das palavras da segunda lista, sendo essa distância colocada numa lista auxiliar. No final da primeira iteração do ciclo, teremos todas as distâncias entre a primeira palavra da lista um e todas as palavras da lista dois inseridas na lista auxiliar. Dessas distâncias, vai ser selecionada a menor e colocada numa segunda lista auxiliar. O ciclo repete-se até todas as distâncias mínimas estarem nesta segunda lista auxiliar. No final, todas estas distâncias mínimas são somadas e é esse o valor que a função retorna. Esta função vai auxiliar um Método específico do Chatbot que serve para selecionar uma resposta utilizando a Distância de Levenshtein. O código desta função pode ser visto em baixo.

---

```
#Levenshtein Distance - auxiliar
def words_edit_distance(list_one, list_two):
    distances_list = []

    for word in list_one:
        aux_list = []
        for words in list_two:
            aux_list.append(edit_distance(word, words))

        distances_list.append(min(aux_list))

    sum_distances = sum(distances_list)

    return sum_distances
```

---

Listing 3.5: Função "words\_edit\_distance"

**Função "words\_jaccard\_distance"**

Quanto à sexta função, "words\_jaccard\_distance", tem como parâmetros duas listas. Esta função faz exatamente o mesmo que a função anterior. No entanto, como a função que estamos a importar (ver secção "Python NLTK - Metrics") para fazer o cálculo desta distância utiliza união e interseção de conjuntos, precisamos de utilizar uma Classe específica do Python para obter o output desejado, daí esta função e a função do parágrafo anterior terem implementações separadas. Esta função vai auxiliar um Método específico do Chatbot que serve para selecionar uma resposta utilizando a Distância de Jaccard. O código desta função pode ser visto em baixo.

---

```
#Jaccard Distance - auxiliar
def words_jaccard_distance(list_one, list_two):
    distances_list = []
    for word in list_one:
        aux_list = []
        for words in list_two:
            aux_list.append(jaccard_distance(set(word), set(words)))

        distances_list.append(min(aux_list))

    sum_distances = sum(distances_list)

    return sum_distances
```

---

Listing 3.6: Função "words\_jaccard\_distance"

**3.6.3 FAQs utilizadas**

O sistema implementado pode ser utilizado sobre qualquer FAQ. No entanto, para este caso específico, foram selecionadas seis FAQs de Domínios diferentes, cada uma com vinte questões e respetivas respostas.

Cada FAQ está guardada num único ficheiro, sendo que a Estrutura de Dados escolhida para armazenar as questões e as respostas das FAQs foi o Dicionário, sendo este o nome Python atribuído à representação de uma coleção cujos elementos são do tipo "chave-valor" [Foua]. Para o caso específico deste projeto, tendo em conta que os Dicionários guardam pares "chave-valor", as questões serão as chaves do Dicionário e as respostas a essas questões serão os valores.

De forma a utilizar o sistema sobre uma FAQ específica, basta importar o ficheiro da FAQ em questão dentro do ficheiro das funções auxiliares.

**3.7 Chatbot - Implementação**

Tal como foi mencionado na secção "Chatbot - Variantes", foram implementados quatro Chatbots, sendo que maior parte do código foi reutilizado, modificando só o necessário para responder ao que se queria testar em cada uma das variantes.

Posto isto, nesta secção, de forma a não tornar o exemplo de implementação de um desses Chatbots algo exaustivo, foi decidido que apenas se colocaria uma das implementações. Essa implementação será a do Chatbot que devolve uma resposta ao utilizador com base nas questões e nas respostas da FAQ, removendo as stopwords ("Pergunta e Resposta - Remove as Stopwords").

A Classe Chatbot é composta por quinze Métodos. Todo o código desses Métodos pode ser visto em anexo (no anexo B).

### **Método "start\_chat"**

Este é o Método que inicializa o Chatbot. O propósito deste Método é informar o utilizador das palavras que pode escrever para sair da conversa, obter o nome do utilizador para criar alguma proximidade, e tratar o utilizador pelo nome, perguntando qual a questão que deseja colocar.

### **Métodos "questions\_Preprocessing", "answers\_Preprocessing", e "user\_Input\_Preprocessing"**

O primeiro dos três Métodos, "questions\_Preprocessing", recebe a lista de questões existentes na FAQ utilizada e faz o pré-processamento dessas questões. O segundo Método, "answers\_Preprocessing", recebe a lista de respostas existente na FAQ e faz o pré-processamento dessas respostas. Por fim, o Método "user\_Input\_Preprocessing", recebe a pergunta do utilizador e faz o pré-processamento dessa pergunta. Todos estes Métodos utilizam a função auxiliar "preprocessing".

### **Método "select\_answer"**

Este Método, que tem como parâmetros três listas (input do utilizador, lista das questões existentes na FAQ, e lista de respostas existentes na FAQ), tem como principal objetivo selecionar a resposta mais adequada, tendo em conta o input do utilizador. O Método vai comparar o input do utilizador com todas as questões e todas as respostas na FAQ, criando um grau de similaridade (palavras comuns entre o input do utilizador e as questões da FAQ, e palavras comuns entre o input do utilizador e as respostas da FAQ). Se o input do utilizador for mais parecido com uma das questões na FAQ, então é devolvida a resposta da FAQ associada a essa questão. Caso o input do utilizador seja mais parecido com uma das respostas da FAQ, é devolvida essa resposta. Este Método utiliza a função auxiliar "similar\_words".

### **Método "basic\_answer"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador da forma mais básica possível. O Método faz o seguinte:

1. pré-processamento do input do utilizador
2. pré-processamento das questões existentes na FAQ
3. pré-processamento das respostas existentes na FAQ
4. seleciona a resposta que parece ser a mais adequada
5. devolve a resposta selecionada

Lista 3.4: Método "basic\_answer" - Funcionamento

### **Método "intermediate\_Answer\_Stemming"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar o processo Stemming. O Método faz o seguinte:

1. cria um novo Stemmer utilizando a Classe "PorterStemmer"
2. pré-processamento do input do utilizador
3. pré-processamento das questões existentes na FAQ
4. pré-processamento das respostas existentes na FAQ
5. aplica Stemming ao input do utilizador pré-processado
6. aplica Stemming às questões pré-processadas
7. aplica Stemming às respostas pré-processadas
8. seleciona a resposta que parece ser a mais adequada
9. devolve a resposta selecionada

Lista 3.5: Método "intermediate\_Answer\_Stemming" - Funcionamento

#### **Método "intermediate\_Answer\_Lemmatization"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar o processo Lemmatization. O Método faz o seguinte:

1. cria um novo Lemmatizer utilizando a Classe "WordNetLemmatizer"
2. pré-processamento do input do utilizador
3. pré-processamento das questões existentes na FAQ
4. pré-processamento das respostas existentes na FAQ
5. aplica Lemmatization ao input do utilizador pré-processado
6. aplica Lemmatization às questões pré-processadas
7. aplica Lemmatization às respostas pré-processadas
8. seleciona a resposta que parece ser a mais adequada
9. devolve a resposta selecionada

Lista 3.6: Método "intermediate\_Answer\_Lemmatization" - Funcionamento

#### **Método "advanced\_Answer\_Cosine\_Similarity"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar a técnica Cosine Distance. O Método faz o seguinte:

1. pré-processamento do input do utilizador
2. pré-processamento das questões existentes na FAQ
3. pré-processamento das respostas existentes na FAQ
4. cria uma lista de vectores em que cada vector tem a união entre o input do utilizador pré-processado e cada uma das questões pré-processadas, individualmente
5. cria duas listas (exteriores) de listas (interiores), verifica se uma determinada palavra existe num determinado vector e, em que cada lista interior só coloca zeros e uns, em que zero significa a não existência dessa palavra no vector e um significa a existência dessa palavra no vector
6. cria uma lista em que cada posição contém a soma do produto de todos os elementos (no mesmo índice em ambas as listas) das listas criadas no ponto anterior
7. cria duas listas em que cada uma das posições nessas listas vai conter a soma dos valores contidos nas listas interiores do ponto cinco
8. cria uma lista para guardar os valores obtidos através da aplicação da fórmula apresentada no capítulo Estado da Arte (Similaridade de Cosine)
9. seleciona a resposta que parece ser a mais adequada
10. devolve a resposta selecionada

Lista 3.7: Método "advanced\_Answer\_Cosine\_Similarity" - Funcionamento

### **Método "advanced\_Levenshtein\_Distance\_Phrase"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar a técnica Levenshtein Distance. Este Método, como podemos ver, tem no nome indicado "Phrase", isto deve-se ao fato de a distância em questão ser aplicada entre frases completas. O Método faz o seguinte:

1. pré-processamento do input do utilizador
2. pré-processamento das questões existentes na FAQ
3. pré-processamento das respostas existentes na FAQ
4. coloca o input do utilizador numa variável
5. cria uma lista de com as questões pré-processadas
6. cria uma lista com as respostas pré-processadas
7. cria uma lista de distâncias em que cada posição é a distância de Levenshtein entre o input do utilizador e cada uma das questões
8. cria uma lista de distâncias em que cada posição é a distância de Levenshtein entre o input do utilizador e cada uma das respostas
9. seleciona o valor mínimo de ambas as listas de distâncias
10. vai buscar o índice desse valor
11. devolve a resposta que estiver nesse índice ao utilizador

Lista 3.8: Método "advanced\_Levenshtein\_Distance\_Phrase" - Funcionamento

**Método "advanced\_Levenshtein\_Distance\_Word"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar a técnica Levenshtein Distance. Este Método, como podemos ver, tem no nome indicado "Word", isto deve-se ao fato de a distância em questão ser aplicada palavra a palavra. Este Método utiliza a função auxiliar "words\_edit\_distance". O Método faz o seguinte:

1. pré-processamento do input do utilizador
2. pré-processamento das questões existentes na FAQ
3. pré-processamento das respostas existentes na FAQ
4. cria uma lista de distâncias em que cada posição é a distância de Levenshtein entre o input do utilizador e cada uma das questões com a ajuda da função auxiliar "words\_edit\_distance"
5. cria uma lista de distâncias em que cada posição é a distância de Levenshtein entre o input do utilizador e cada uma das respostas com a ajuda da função auxiliar "words\_edit\_distance"
6. seleciona o valor mínimo de ambas as listas de distâncias
7. vai buscar o índice desse valor
8. devolve a resposta que estiver nesse índice ao utilizador

Lista 3.9: Método "advanced\_Levenshtein\_Distance\_Word" - Funcionamento

**Método "advanced\_Jaccard\_Distance"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar a técnica Jaccard Distance. Este Método utiliza a função auxiliar "words\_jaccard\_distance". O Método faz o seguinte:

1. pré-processamento do input do utilizador
2. pré-processamento das questões existentes na FAQ
3. pré-processamento das respostas existentes na FAQ
4. cria uma lista de distâncias em que cada posição é a distância de Jaccard entre o input do utilizador e cada uma das questões com a ajuda da função auxiliar "words\_jaccard\_distance"
5. cria uma lista de distâncias em que cada posição é a distância de Jaccard entre o input do utilizador e cada uma das respostas com a ajuda da função auxiliar "words\_jaccard\_distance"
6. seleciona o valor mínimo de ambas as listas de distâncias
7. vai buscar o índice desse valor
8. devolve a resposta que estiver nesse índice ao utilizador

Lista 3.10: Método "advanced\_Jaccard\_Distance" - Funcionamento

**Método "advanced\_Semantic\_Similarity"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo devolver uma resposta ao utilizador após aplicar a técnica de Similaridade Semântica. O Método faz o seguinte:



1. pré-processamento do input do utilizador
2. pré-processamento das questões existentes na FAQ
3. pré-processamento das respostas existentes na FAQ
4. carrega um modelo pré treinado para uma variável
5. constrói um Dicionário que faz um mapeamento entre palavras e os identificadores dessas palavras
6. constrói um Modelo "Term Frequency - Inverse Document Frequency" a partir do Dicionário
7. constrói uma matriz de Similaridade de Termos a partir do dados obtidos nos três pontos anteriores
8. calcula a Soft Cosine Similarity entre a questão do utilizador e cada um dos pares "questão-resposta", guardando os valores obtidos numa lista
9. seleciona o índice que contém o maior valor calculado
10. devolve a resposta que estiver nesse índice ao utilizador

Lista 3.11: Método "Advanced\_Semantic\_Similarity" - Funcionamento

### **Método "body\_Chat"**

Este Método, que tem como parâmetro o input do utilizador, tem como objetivo invocar todos os Métodos que podem devolver uma resposta ao utilizador. O Método faz o seguinte:

1. verifica o valor do grau de similaridade
2. mostra ao utilizador, de forma individual, as respostas devolvidas por cada um dos Métodos que foram mencionados anteriormente
3. pergunta ao utilizador se ficou esclarecido ou quer colocar mais alguma questão

Lista 3.12: Método "body\_Chat" - Funcionamento

### **Método "exit\_Chat"**

Este Método, que tem como parâmetro o input do utilizador, serve terminar a conversa entre o sistema e o utilizador. O Método faz o seguinte:

1. verifica se o input do utilizador consta numa lista de palavras que permitem ao utilizador terminar a conversa
2. caso o input do utilizador seja no sentido de terminar a conversa, a conversa com o Chatbot é terminada

Lista 3.13: Método "exit\_Chat" - Funcionamento

# 4

## Avaliação do Sistema

Após a implementação do Sistema e da descrição do seu funcionamento, que pode ser vista no Capítulo do Desenvolvimento, resta avaliar o comportamento do Sistema na prática, utilizando FAQs reais. As FAQs em questão foram retiradas da Internet e pertencem a Domínios diferentes, podendo ser vistas em anexo (anexo C a anexo H).

É de salientar que a Avaliação será feita por mim, que também implementei o Sistema. O ideal seria que esta fosse feita de forma independente, por vários utilizadores, contudo, foi decidido que esta também seria uma forma válida de fazer a Avaliação.

Tal como já foi referido, as stopwords (ou palavras vazias) são palavras que não adicionam muito significado a uma frase e que podem ser removidas quando na aplicação de algumas técnicas para interpretação das frases e, neste caso específico, os idiomas dessas frases serão o Inglês e o Português. O objetivo deste Capítulo é identificar o impacto das stopwords no desempenho dos Métodos utilizados, que podem ser lembrados na lista que se segue:

- Básico
- Stemming
- Lemmatization
- Cosine
- Levenshtein - Frase
- Levenshtein - Palavra
- Jaccard
- Semantic Similarity

Lista 4.1: Lista dos Métodos utilizados na Avaliação do Sistema

## 4.1 Descrição das FAQs utilizadas

As FAQs utilizadas para testar o Sistema implementado, tal como foi referido anteriormente, foram retiradas da Internet. É de salientar que, à exceção das FAQs relacionadas com o Instituto da Vinha e do Vinho, não houve uma preferência específica por um Domínio em particular, apenas foi pretendido que todos os Domínios utilizados fossem diferentes, de forma a poder verificar o comportamento do Sistema em cada um deles.

Posto isto, foram selecionadas seis FAQs, cada uma com vinte questões e respetivas respostas, das quais três estão em Inglês (Covid, OnePlus, e UEFA) e três estão em Português (Évora, Parlamento, e Instituto da Vinha e do Vinho). Aqui, mais uma vez, ao selecionar FAQs de línguas diferentes, pretende-se analisar o comportamento do Sistema em ambas as situações.

O primeiro Domínio está relacionado com o novo Coronavírus, estando as FAQs em Inglês. As FAQs foram retiradas da página "Centers for Disease Control and Prevention" [oHHS]. Essas FAQs estão relacionadas com prevenção, sintomas e sinais de emergência, pessoas com risco de doença severa, e rastreamento de contactos.

O segundo Domínio está relacionado com a marca de telemóveis OnePlus, estando as FAQs em Inglês. As FAQs foram retiradas da página "Never Settle - OnePlus (Portugal)" [One]. Essas FAQs estão relacionadas com pós-venda, apoio técnico, e envios de encomendas.

O terceiro Domínio está relacionado com a UEFA, estando as FAQs em Inglês. As FAQs foram retiradas da página "FAQ | UEFA Euro 2020 Hospitality" [UEF]. Essas FAQ estão relacionadas com informações gerais e compras.

O quarto Domínio está relacionado com a Universidade de Évora, estando as FAQs em Português. As FAQs foram retiradas da página "Serviços Académicos da UÉVORA - Universidade de Évora" [d]. Essas FAQs estão relacionadas com informações sobre como contactar os Serviços Académicos, sobre matrículas e propinas.

O quinto Domínio está relacionado com o Parlamento da República Portuguesa, estando as FAQs em Português. As FAQs foram retiradas da página "Perguntas frequentes - Parlamento" [dR]. Essas FAQs estão relacionadas com a atividade parlamentar, os deputados, e a cidadania e participação.

O sexto e último Domínio está relacionado com o Instituto da Vinha e do Vinho, estando as FAQs em

Português. As FAQs foram retiradas da página " IVV // FAQ's" [dVedV]. Essas FAQs estão relacionadas com autorizações de plantação de vinha, declarações obrigatórias, e promoção em países terceiros.

## 4.2 Descrição das Experiências

Como já foi mencionado em Capítulos anteriores, o Sistema utiliza vários Métodos para devolver uma resposta ao Utilizador. A ideia principal da Avaliação é verificar o desempenho de cada Método em determinadas condições de forma a saber quais os Métodos com maior acerto e, possivelmente, como é que podem ser melhorados.

A cada um dos Domínios das FAQs, individualmente, vão ser colocadas sessenta questões: vinte que fogem completamente ao tema do Domínio (Questões Não Relacionadas) e que são comuns a todos os Domínios, vinte que estão relacionadas com o tema do Domínio mas que foram reformuladas a partir das questões originais presentes nas FAQs (Questões Reformuladas), e as vinte questões originais presentes nas FAQs daquele Domínio (Questão Original). Estes são os três principais cenários que visam três categorias de utilizadores: os utilizadores que estão equivocados na sua pesquisa, os que têm uma ideia do que querem perguntar, e os que são claros na questão que colocam.

Cada questão colocada terá uma nota de zero, dois, ou cinco, sendo que zero indica que a resposta devolvida a essa questão está errada, dois indica que o Sistema devolveu uma resposta a dizer que não percebeu a questão do utilizador (resposta por defeito), e cinco indica que a resposta devolvida a essa questão foi acertada.

Para responder de forma acertada a uma questão não relacionada, o Método tem de dizer ao utilizador que não percebeu a questão colocada, pedindo ao utilizador para a reformular. De forma a responder acertadamente a uma questão reformulada e a uma questão original, o Método tem de devolver a resposta associada a essa questão nas FAQs originais, tendo em conta que, apesar das questões reformuladas e a questões originais serem diferentes, a resposta é a mesma para ambas.

É de salientar que, além da avaliação destes cinco Domínios, vai ser feita uma avaliação em que todas as questões reformuladas do mesmo idioma (Inglês ou Português) são colocadas num ficheiro. Ou seja, para o Inglês vamos ter sessenta questões reformuladas e respetivas respostas (vinte relacionadas com Covid, vinte relacionadas com OnePlus, e vinte relacionadas com a UEFA), e para o Português vamos ter quarenta questões reformuladas e respetivas respostas (vinte relacionadas com Évora e vinte relacionadas com o Parlamento).

A ideia fundamental desta segunda avaliação é verificar se o fato dos Domínios estarem ou não separados influencia o desempenho dos Métodos.

## 4.3 Métodos de Avaliação

A seleção da resposta que o Sistema devolve ao utilizador foi testada de quatro formas diferentes: seleção com base apenas nas respostas das FAQs (removendo as stopwords, e sem remover as stopwords) e seleção de resposta com base nas questões e nas respostas das FAQs (removendo as stopwords, e sem remover as stopwords).

Antes de mais, tendo em conta que o objetivo é criar um Sistema que devolva respostas com base em FAQs, faz todo o sentido que, quando o utilizador coloca uma questão, a pesquisa por algo relacionado com essa questão seja feita nas questões e nas respostas das FAQs e não apenas nas respostas, como sugerem os

dois primeiros testes. No entanto, tendo em conta que todas as questões (tirando as originais das FAQs) foram criadas por mim, foi decidido que, apesar destes testes não serem expostos nos resultados, seria feita menção aos mesmos e seriam colocados em anexo. Além disto, acaba por ser mais uma perspetiva relativamente à avaliação.

Em todos os quatro testes, para cada Domínio, foram colocadas sessenta questões ao Sistema. Primeiro as vinte questões não relacionadas, depois as vinte questões reformuladas e, por fim, as vinte questões originais. Após cada questão colocada, os resultados de cada Método foram anotados e guardados na respetiva tabela do seu Domínio específico.

## 4.4 Resultados

Os resultados que se seguem são relativos aos testes em que o Sistema devolve uma resposta ao utilizador tendo em conta o conteúdo das questões e das respostas presentes nas FAQs. É, também, feita a distinção de quando são removidas as stopwords e de quando não são removidas as stopwords.

Relativamente às questões não relacionadas, considera-se que um Método deu uma resposta certa quando ele diz que não percebe a pergunta que lhe está a ser colocada, pedindo ao utilizador para a reformular.

Quanto às restantes, tanto as questões reformuladas como as questões originais, considera-se que um Método deu uma resposta certa quando ele devolve a resposta associada àquela questão específica.

### 4.4.1 Covid

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para as FAQs relacionadas com Covid (em Inglês), tendo em conta o tipo de questão (não relacionada, reformulada, e original) e tendo em conta se as stopwords são ou não removidas.

## Covid - Resultados removendo as stopwords

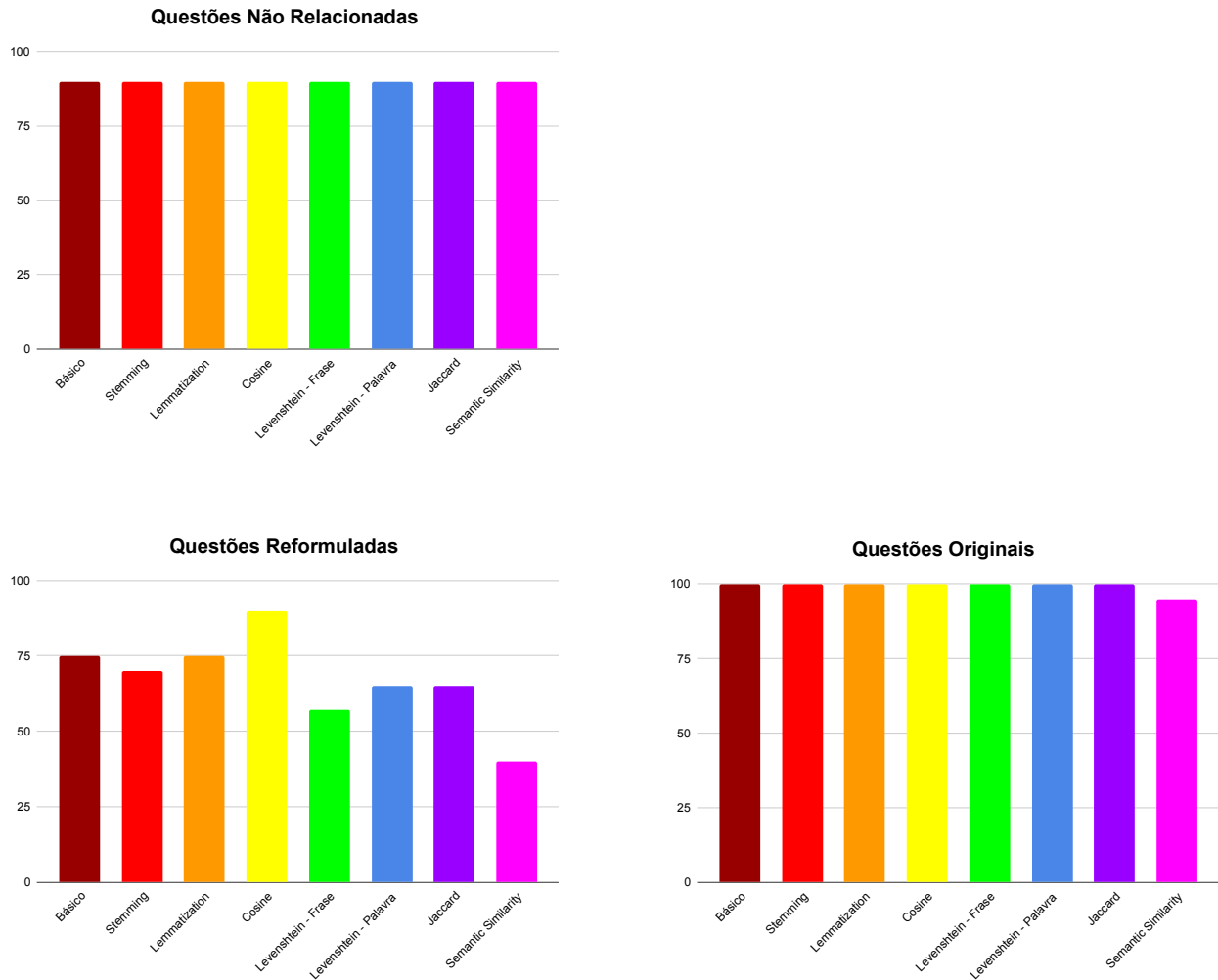


Figura 4.1: FAQs Covid - As stopwords são removidas

Como é possível verificar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram o mesmo desempenho, respondendo que não percebiam a questão do utilizador em dezoito das vinte questões colocadas.

Em relação às questões reformuladas, é possível observar que o desempenho dos Métodos variou bastante, sendo que o Método que teve pior desempenho foi o "Semantic Similarity", tendo acertado apenas em oito das vinte questões que lhe foram colocadas. O Método que teve melhor desempenho foi o "Cosine", acertando em dezoito das vinte questões.

Já o último gráfico mostra que, quando foram colocadas as questões originais que constam nas FAQs, todos os Métodos acertaram as vinte respostas, exceto o "Semantic Similarity", que falhou uma.

## Covid - Resultados não removendo as stopwords

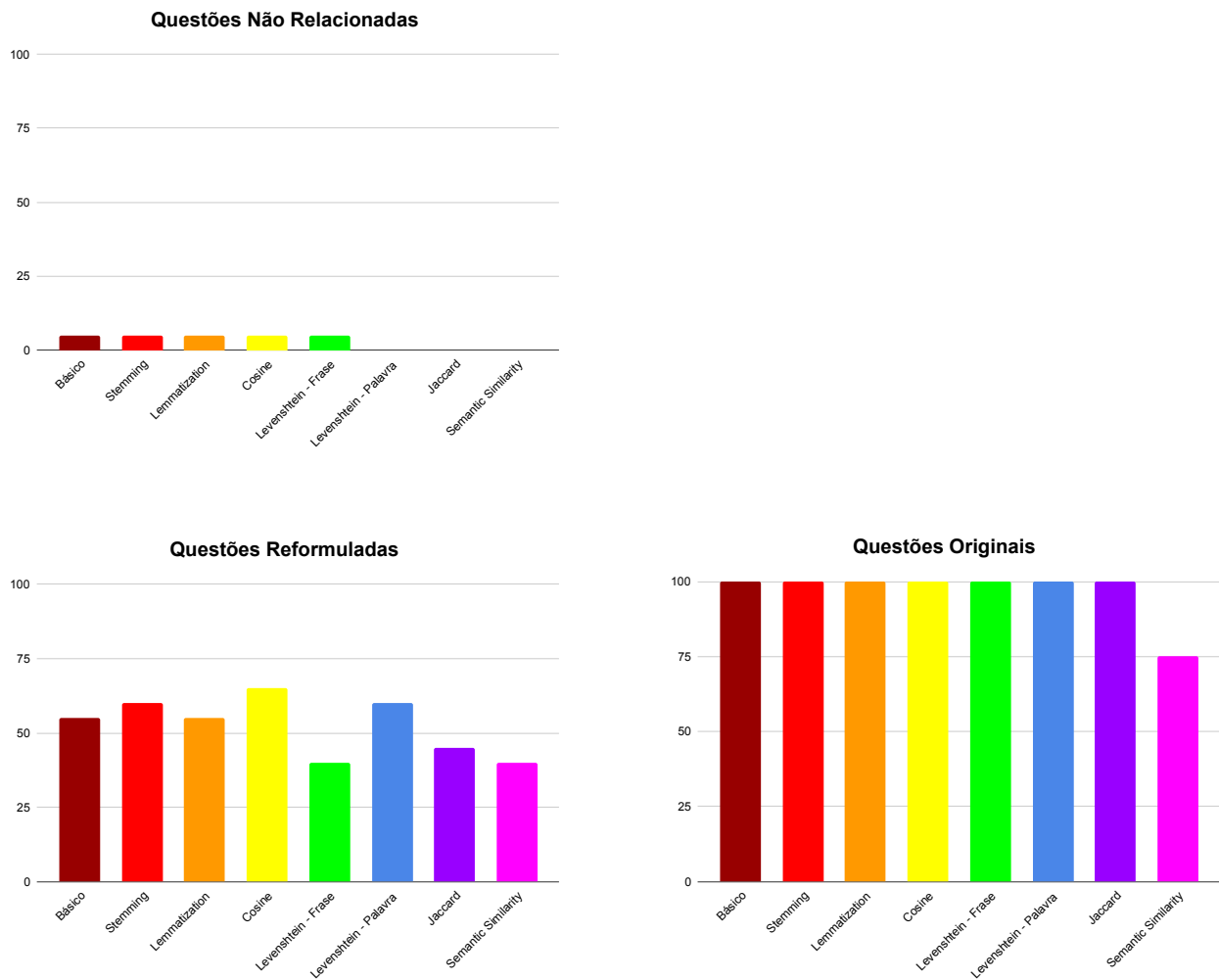


Figura 4.2: FAQs Covid - As stopwords não são removidas

Como podemos observar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram um mau desempenho, devolvendo respostas presentes nas FAQs que nada tinham a ver com a questão colocada, quando apenas deviam ter dito ao utilizador que não tinham entendido a questão.

Relativamente às questões reformuladas, podemos considerar que os Métodos tiveram um desempenho mediano, sendo que os três Métodos que tiveram pior desempenho apenas acertaram em oito das vinte questões colocadas. Já o Método com melhor desempenho, o "Cosine", acertou em treze das vinte questões.

No que toca às questões originais, é possível observar que a pior percentagem de acerto pertenceu ao Método "Semantic Similarity", tendo respondido de forma acertada a quinze das vinte questões. Todos os outros Métodos tiveram a pontuação máxima.

### 4.4.2 OnePlus

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para as FAQs relacionadas com a OnePlus (em Inglês), tendo em conta o tipo de questão (não relacionada, reformulada, e original) e tendo em conta se as stopwords são ou não removidas.

### OnePlus - Resultados removendo as stopwords

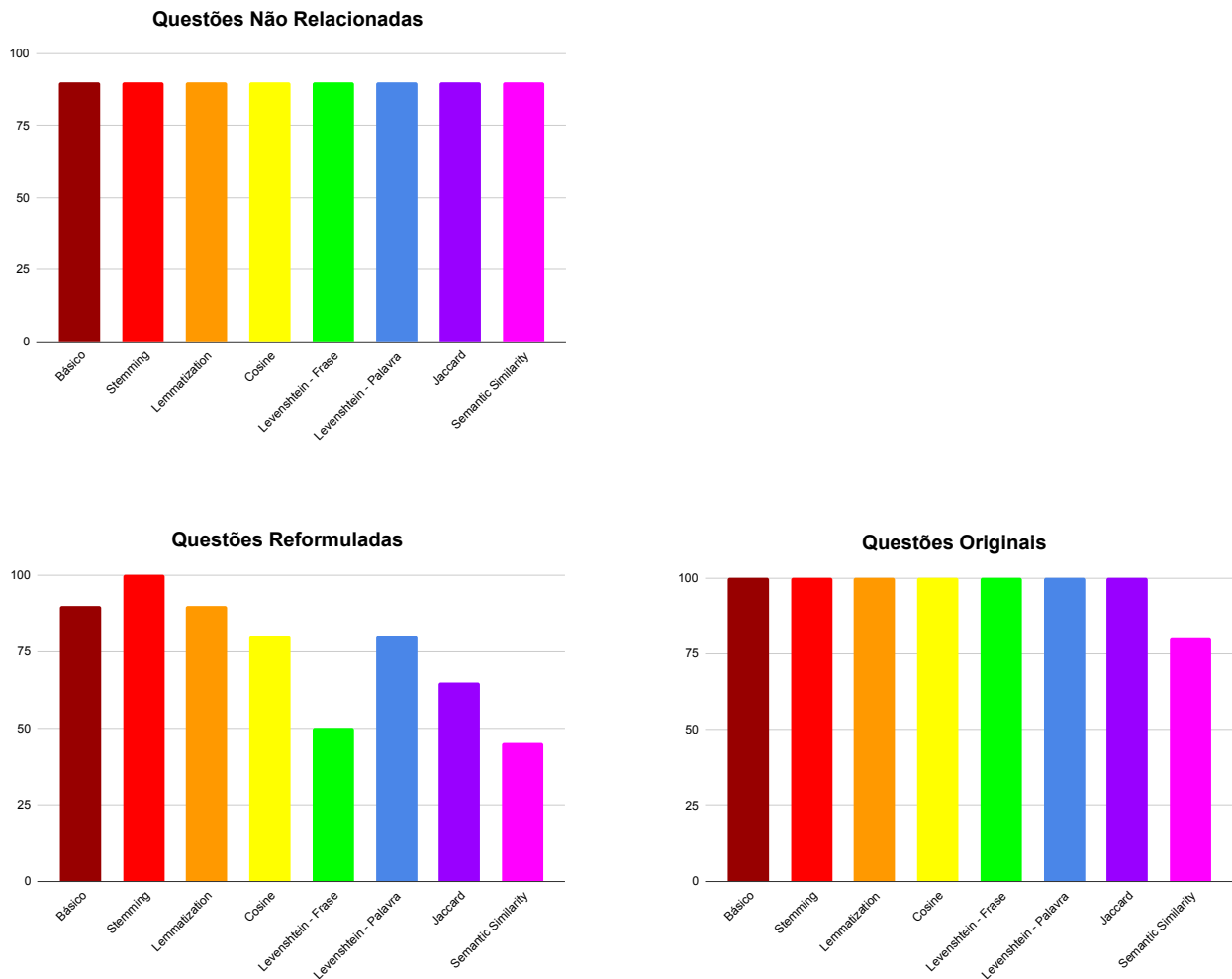


Figura 4.3: FAQs OnePlus - As stopwords são removidas

Como é possível verificar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram o mesmo desempenho, respondendo que não percebiam a questão do utilizador em dezoito das vinte questões colocadas.

Quanto às questões reformuladas, é possível observar que o desempenho dos Métodos variou bastante, sendo que o Método que teve pior desempenho foi o "Semantic Similarity", tendo acertado apenas em nove das vinte questões que lhe foram colocadas. O Método que teve melhor desempenho foi o "Stemming", acertando em todas as questões.

No gráfico das questões originais, podemos observar que todos os Métodos acertaram as vinte respostas, exceto o "Semantic Similarity", que falhou quatro.



### OnePlus - Resultados não removendo as stopwords

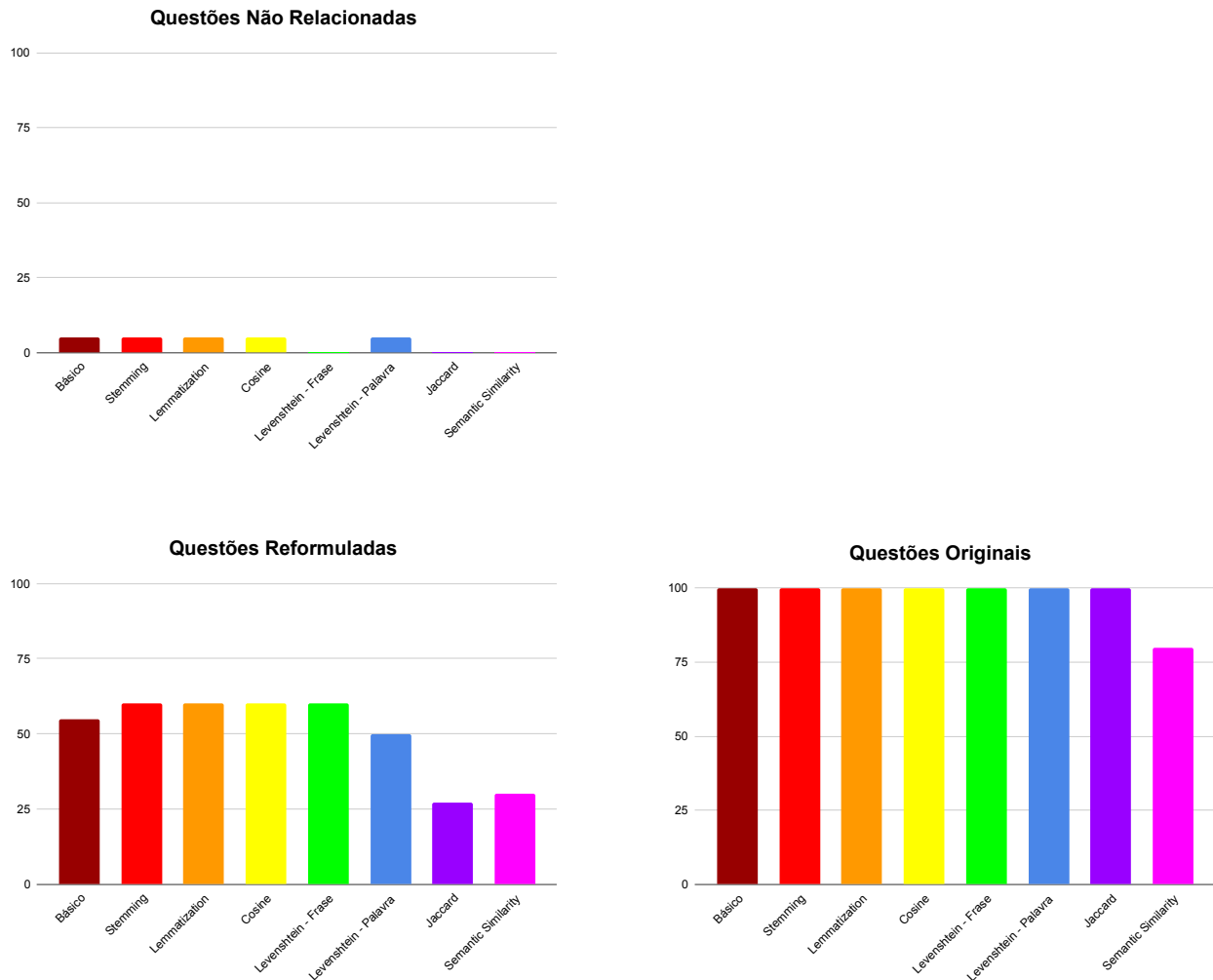


Figura 4.4: FAQs OnePlus - As stopwords não são removidas

Como é possível observar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram um mau desempenho, devolvendo respostas presentes nas FAQs que nada tinham a ver com a questão colocada, quando apenas deviam ter dito ao utilizador que não tinham entendido a questão, de forma a que o utilizador a pudesse reformular.

Relativamente às questões reformuladas, podemos considerar que os Métodos tiveram um desempenho razoável, sendo que o Método que teve pior desempenho apenas acertou em cinco das vinte questões colocadas, tendo também pedido ao utilizador para reformular umas das questões, daí a percentagem de acerto ser de vinte e sete por cento. A maior percentagem de acerto pertenceu a quatro Métodos, tendo todos acertado em doze das vinte questões.

No que toca às questões originais, é possível observar que a pior percentagem de acerto pertenceu ao Método "Semantic Similarity", tendo respondido de forma acertada a dezasseis das vinte questões. Todos os outros Métodos tiveram a pontuação máxima.

### 4.4.3 UEFA

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para as FAQs relacionadas com a UEFA (em Inglês), tendo em conta o tipo de questão (não relacionada, reformulada, e original) e tendo em conta se as stopwords são ou não removidas.

#### UEFA - Resultados removendo as stopwords

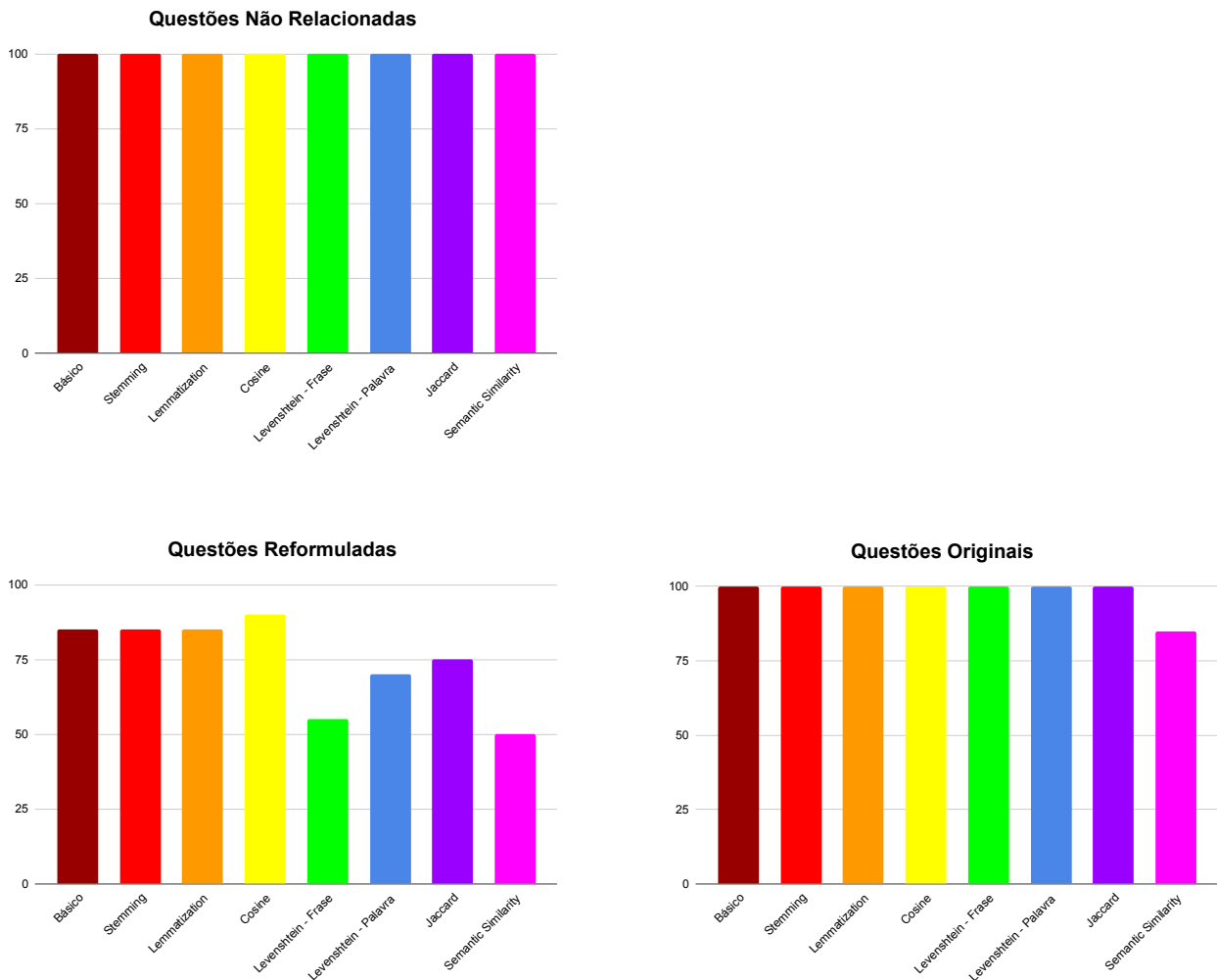


Figura 4.5: FAQs UEFA - As stopwords são removidas

Como é possível verificar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram o mesmo desempenho, respondendo que não percebiam a questão do utilizador em todas as questões colocadas.

Quanto às questões reformuladas, é possível observar que o desempenho da maior parte Métodos foi bom, sendo que o Método que teve pior desempenho foi o "Semantic Similarity", tendo acertado apenas em dez das vinte questões que lhe foram colocadas. O Método que teve melhor desempenho foi o "Cosine", acertando em dezoito das vinte questões.

No gráfico das questões originais, é possível observar que todos os Métodos acertaram as vinte respostas, exceto o "Semantic Similarity", que falhou três.

### UEFA - Resultados não removendo as stopwords

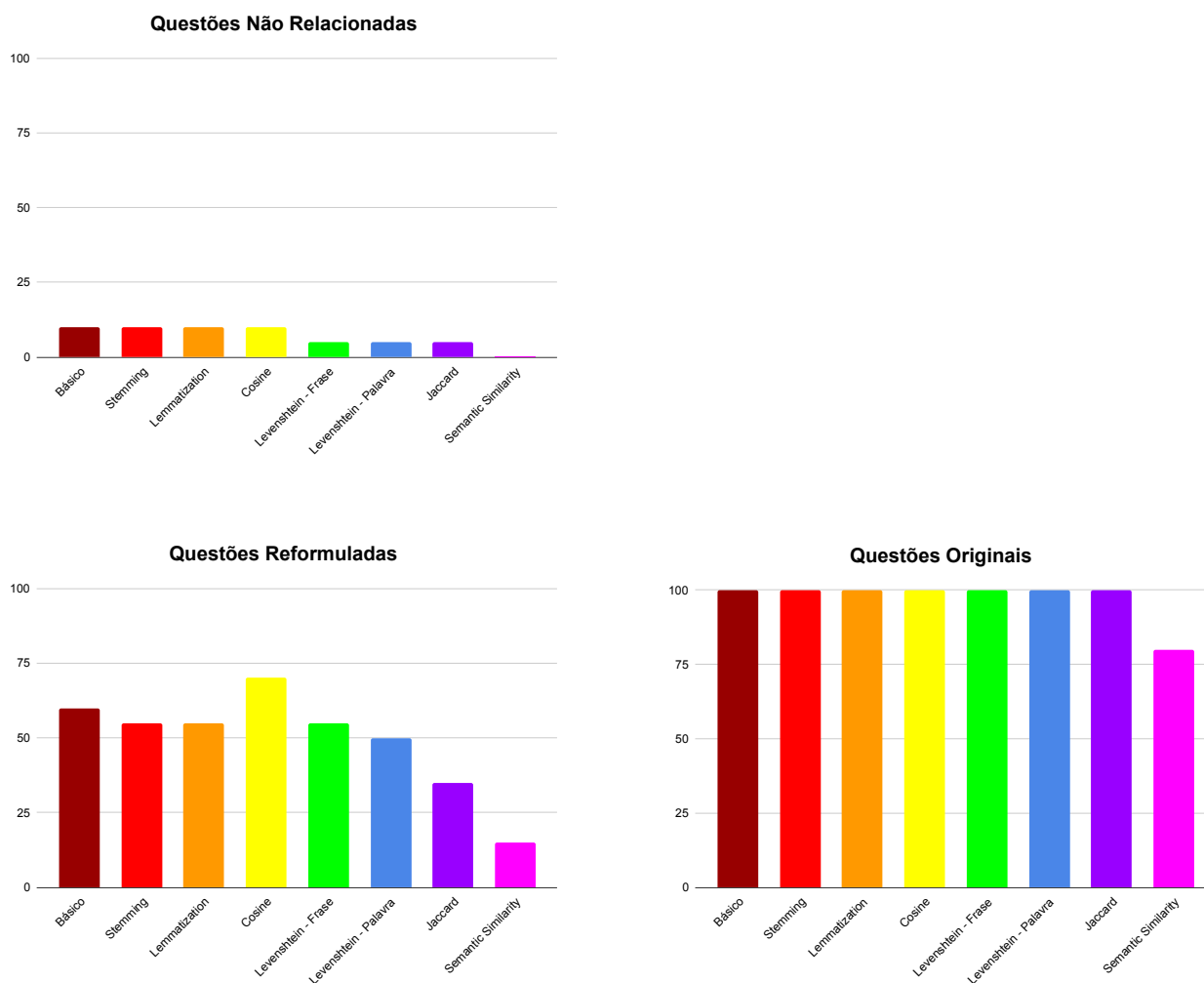


Figura 4.6: FAQs UEFA - As stopwords não são removidas

Como é possível observar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram um mau desempenho, devolvendo respostas presentes nas FAQs que nada tinham a ver com a questão colocada, quando apenas deviam ter dito ao utilizador que não tinham entendido a questão, de forma a que o utilizador a pudesse reformular.

Relativamente às questões reformuladas, podemos considerar que os Métodos tiveram um desempenho aceitável, sendo que o Método que teve pior desempenho foi o "Semantic Similarity", acertando apenas em cinco das vinte questões colocadas. A maior percentagem de acerto pertenceu ao Método "Cosine", tendo acertado em catorze das vinte questões.

Relativamente às questões originais, é possível observar que a pior percentagem de acerto pertenceu ao Método "Semantic Similarity", tendo respondido de forma acertada a dezasseis das vinte questões. Todos os outros Métodos tiveram a pontuação máxima.

#### 4.4.4 Évora

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para as FAQs relacionadas com a Universidade de Évora (em Português), tendo em conta o tipo de questão (não relacionada, reformulada, e original) e tendo em conta se as stopwords são ou não removidas.

##### Évora - Resultados removendo as stopwords

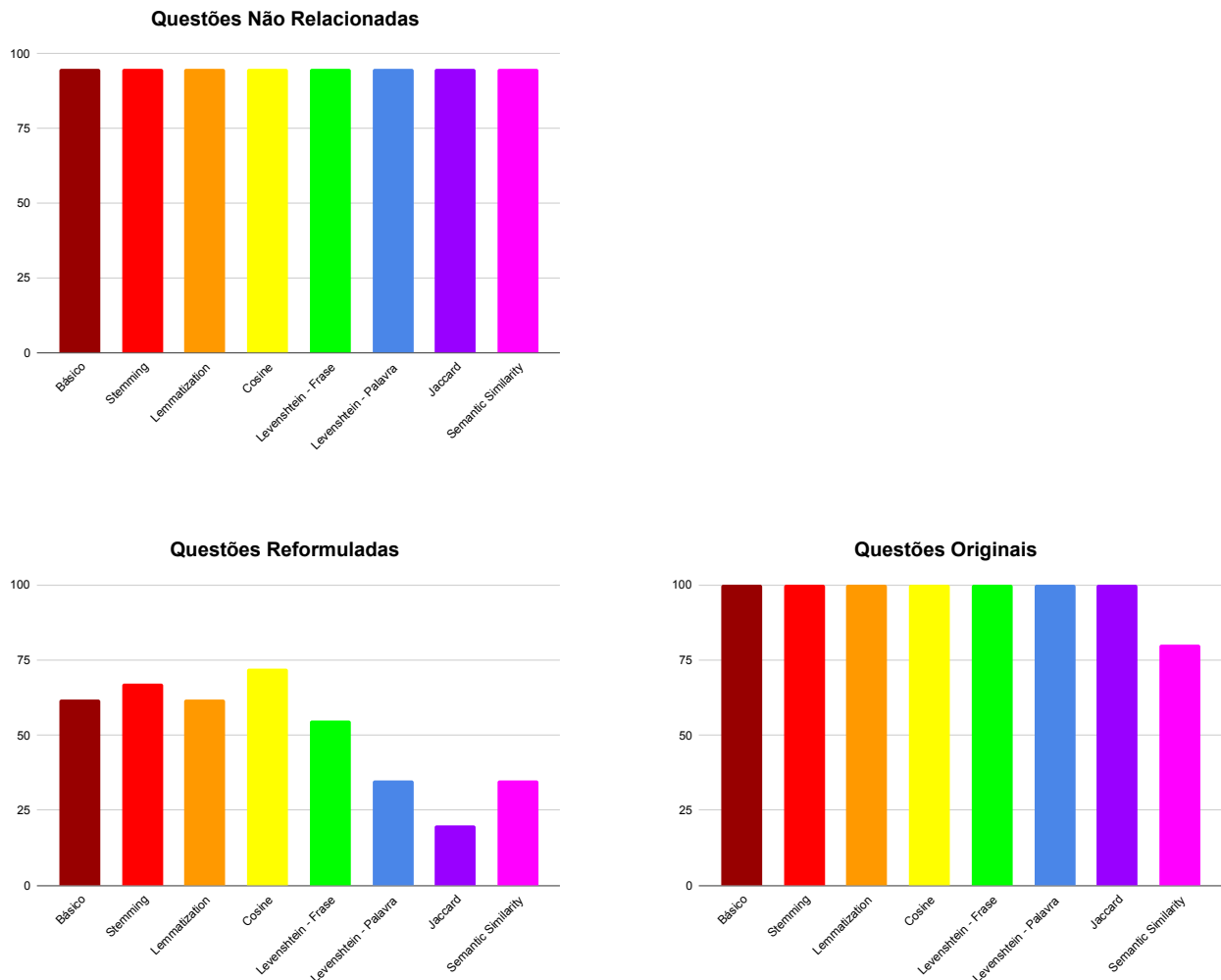


Figura 4.7: FAQs Évora - As stopwords são removidas

Como podemos observar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram o mesmo desempenho, respondendo que não percebiam a questão do utilizador em dezanove das vinte questões colocadas.

Relativamente às questões reformuladas, é possível verificar que o desempenho da maior parte Métodos foi mediano, sendo que o Método que teve pior desempenho foi o "Jaccard", tendo acertado apenas em sete das vinte questões que lhe foram colocadas. Já o Método que obteve melhor desempenho foi o "Cosine", acertando em catorze das vinte questões, sendo que a percentagem de setenta e dois por cento deve-se ao fato de ter respondido que não percebeu uma das questões.

Quanto às questões originais, é possível observar que todos os Métodos acertaram as vinte respostas, exceto

o "Semantic Similarity", que falhou quatro.

### Évora - Resultados não removendo as stopwords

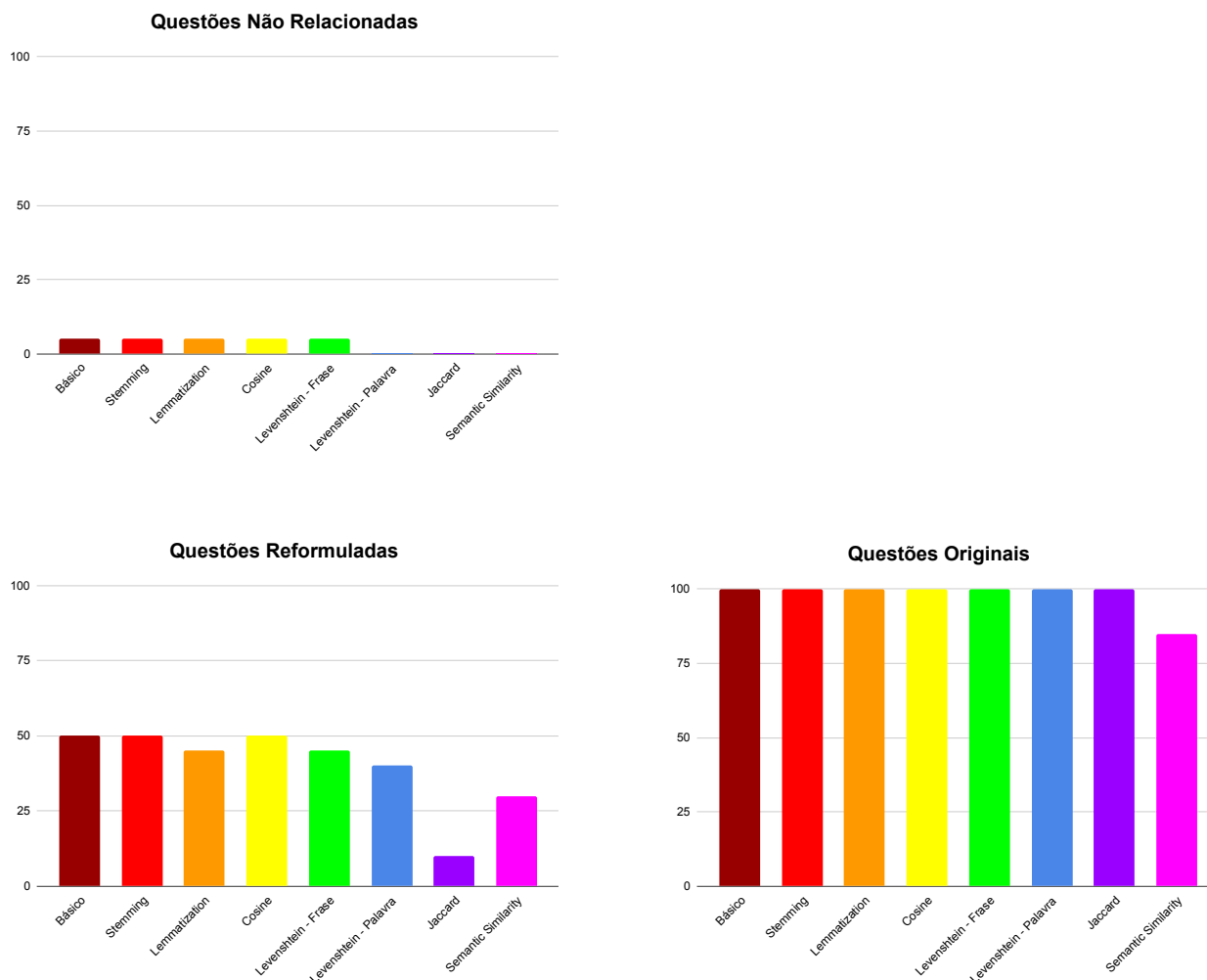


Figura 4.8: FAQs Évora - As stopwords não são removidas

Como podemos observar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram um mau desempenho, devolvendo respostas presentes nas FAQs que nada tinham a ver com a questão colocada, quando apenas deviam ter dito ao utilizador que não tinham entendido a questão, de forma a que o utilizador a pudesse reformular.

Relativamente às questões reformuladas, podemos considerar que os Métodos tiveram um desempenho abaixo de mediado, sendo que o Método que teve pior desempenho foi o "Jaccard", acertando apenas em duas das vinte questões colocadas. A maior percentagem de acerto pertenceu a três Métodos, acertando todos em dez das vinte questões.

Já nas questões originais, é possível observar que a pior percentagem de acerto pertenceu ao Método "Semantic Similarity", tendo respondido de forma acertada a dezassete das vinte questões. Todos os outros Métodos tiveram a pontuação máxima.

### 4.4.5 Parlamento

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para as FAQs relacionadas com o Parlamento (em Português), tendo em conta o tipo de questão (não relacionada, reformulada, e original) e tendo em conta se as stopwords são ou não removidas.

#### Parlamento - Resultados removendo as stopwords

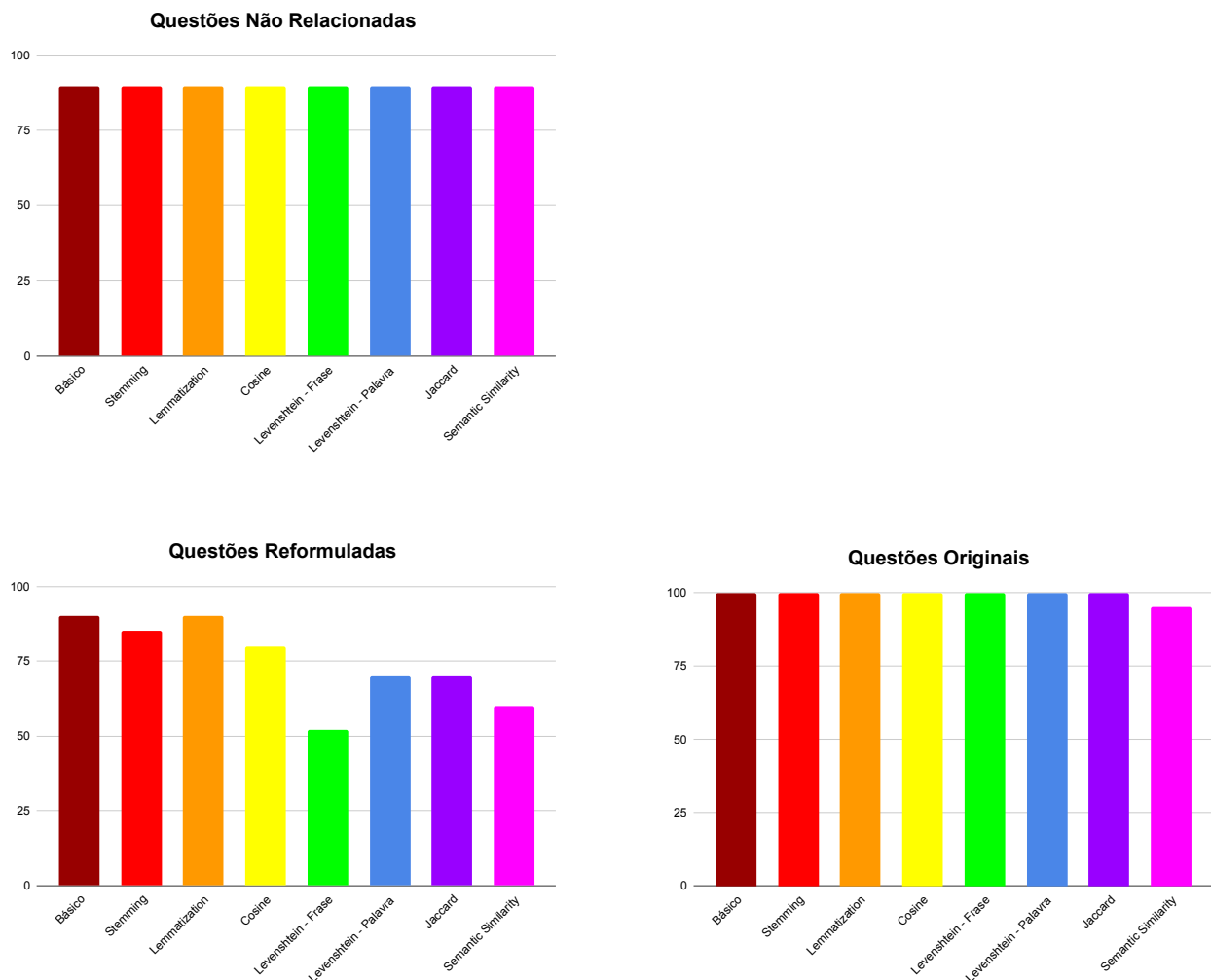


Figura 4.9: FAQs Parlamento - As stopwords são removidas

Como é possível verificar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram o mesmo desempenho, respondendo que não percebiam a questão do utilizador em dezoito das vinte questões colocadas.

Em relação às questões reformuladas, é possível verificar que o desempenho da maior parte Métodos foi bom, sendo que o Método que teve pior desempenho foi o "Levenshtein - Frase", tendo acertado apenas em dez das vinte questões que lhe foram colocadas, sendo a percentagem de cinquenta e dois por cento se deve ao fato de ter dito que não percebeu uma das questões que lhe foi colocada. Existiram dois Métodos com o melhor desempenho, tendo ambos acertado em dezoito das vinte questões.

Relativamente às questões originais, é possível observar que todos os Métodos acertaram as vinte respostas, exceto o "Semantic Similarity", que falhou uma.

### Parlamento - Resultados não removendo as stopwords

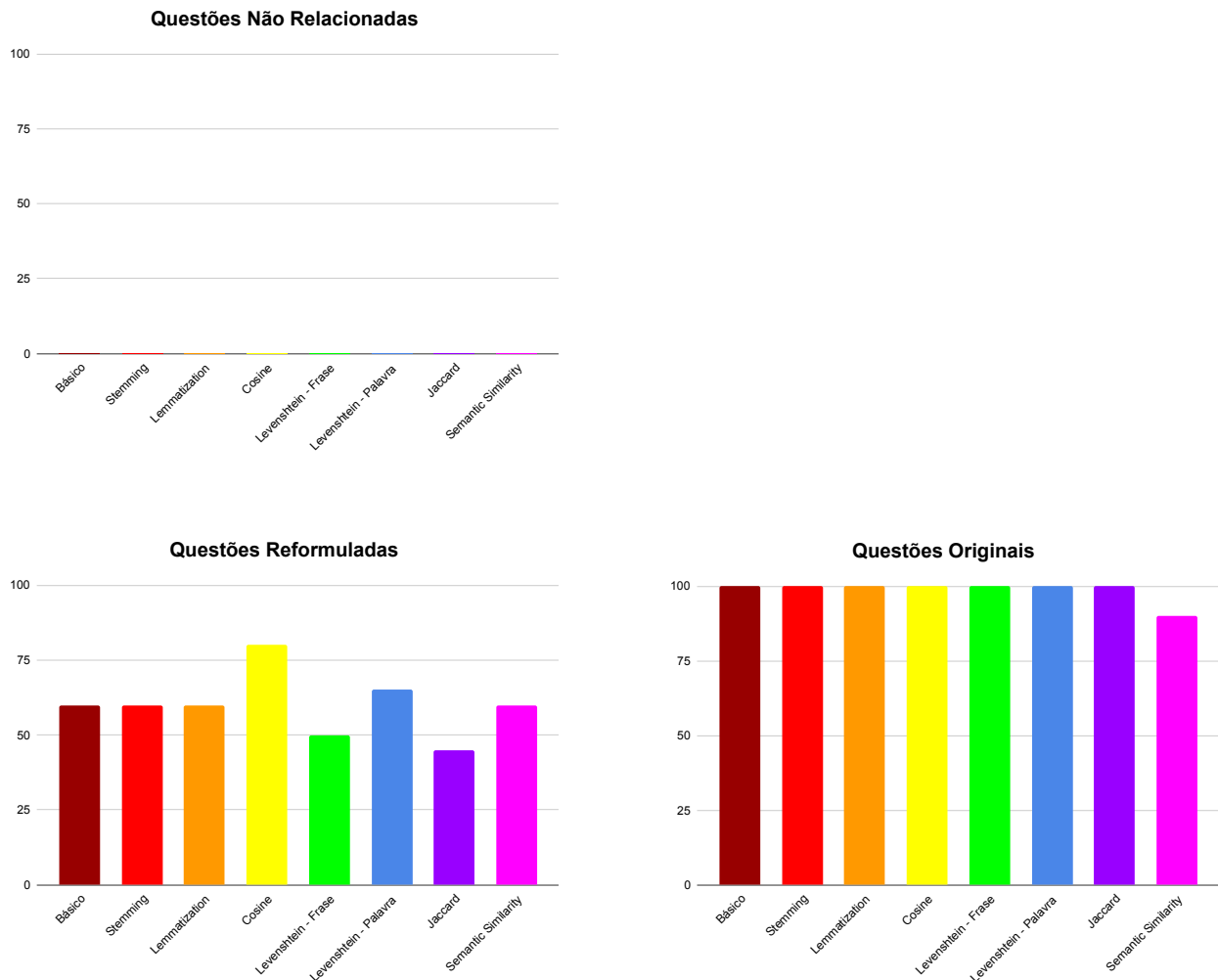


Figura 4.10: FAQs Parlamento - As stopwords não são removidas

Como podemos verificar, quando foram colocadas questões não relacionadas, todos os Métodos tiveram um mau desempenho, devolvendo respostas presentes nas FAQs que nada tinham a ver com a questão colocada, quando apenas deviam ter dito ao utilizador que não tinham entendido a questão, de forma a que o utilizador a pudesse reformular.

Relativamente às questões reformuladas, podemos considerar que os Métodos tiveram um desempenho razoável, sendo que o Método que teve pior desempenho foi o "Jaccard", acertando apenas em nove das vinte questões colocadas. O Método com melhor desempenho foi o "Cosine", tendo acertado em dezasseis das vinte questões.

Em relação às questões originais, é possível verificar que a pior percentagem de acerto pertenceu ao Método "Semantic Similarity", tendo respondido de forma acertada a dezoito das vinte questões. Todos os outros Métodos tiveram a pontuação máxima.

#### 4.4.6 Todas as Questões Reformuladas (Inglês)

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para todas as questões reformuladas dos diferentes Domínios que estão em Inglês (Covid, OnePlus, UEFA), tendo em conta se as stopwords são ou não removidas.

##### Todas as Questões Reformuladas (Inglês) - Resultados removendo as stopwords

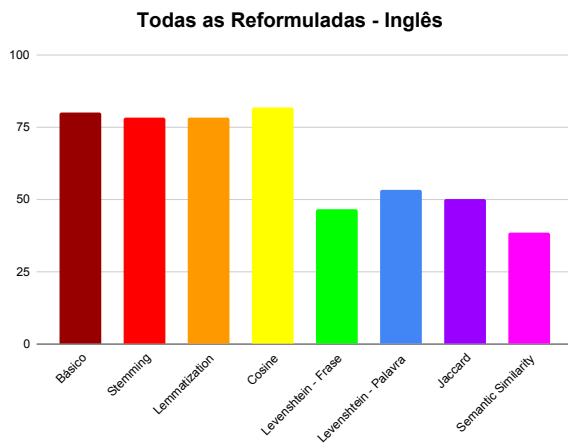


Figura 4.11: Todas as Questões Reformuladas (Inglês) - As stopwords são removidas

Como é possível verificar, o desempenho de metade dos Métodos foi razoável e o desempenho da outra metade foi bom, sendo que o Método que teve pior desempenho foi o "Semantic Similarity", tendo acertado em vinte e três das sessenta questões que lhe foram colocadas. O Método que obteve o melhor desempenho foi o Cosine, tendo acertado quarenta e nove das sessenta questões.

##### Todas as Questões Reformuladas (Inglês) - Resultados não removendo as stopwords

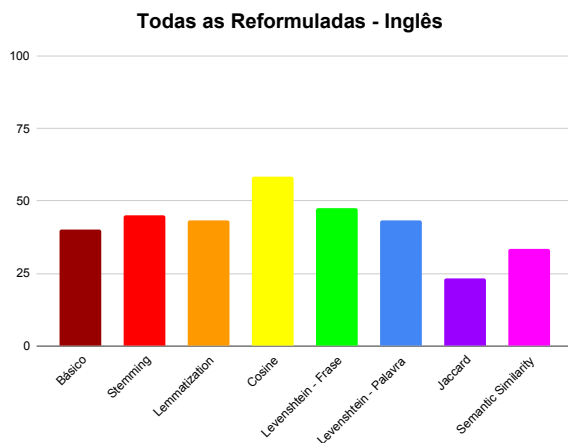


Figura 4.12: Todas as Questões Reformuladas (Inglês) - As stopwords são removidas



Como podemos constatar através dos gráficos, o desempenho de sete dos oito Métodos ficou abaixo dos cinquenta por cento de acerto, sendo que o Método que teve pior desempenho foi "Jaccard", tendo apenas acertado catorze das sessenta questões colocadas. O Método que conseguiu maior percentagem de acerto foi o "Cosine", tendo acertado em trinta e cinco das sessenta questões.

#### 4.4.7 Todas as Questões Reformuladas (Português)

Nos gráficos em baixo pode ser vista a percentagem de acerto de cada Método para todas as questões reformuladas dos diferentes Domínios que estão em Português (Évora e Parlamento), tendo em conta se as stopwords são ou não removidas.

#### Todas as Questões Reformuladas (Português) - Resultados removendo as stopwords

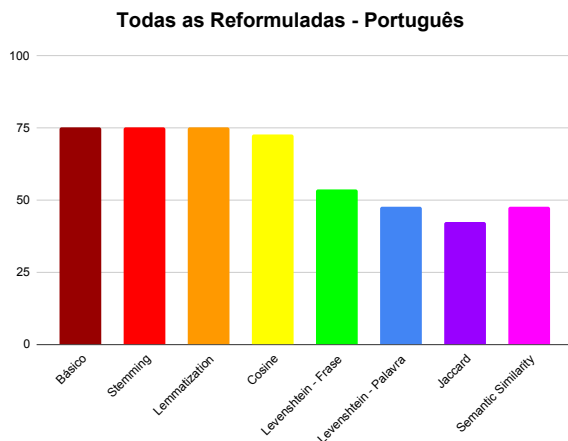


Figura 4.13: Todas as Questões Reformuladas (Português) - As stopwords são removidas

Como podemos observar, o desempenho de metade dos Métodos foi razoável e o desempenho da outra metade foi bom, sendo que o Método que teve pior desempenho foi o "Jaccard", tendo acertado em dezasseis das quarenta questões que lhe foram colocadas. Existiram três Métodos com o melhor desempenho, tendo todos acertado trinta das quarenta questões.

### Todas as Questões Reformuladas (Português) - Resultados não removendo as stopwords

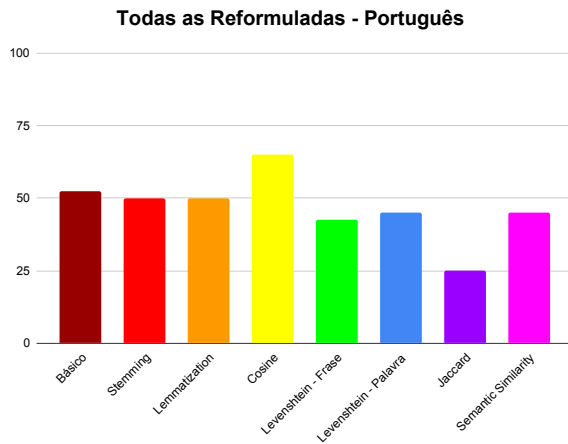


Figura 4.14: Todas as Questões Reformuladas (Português) - As stopwords são removidas

Como é possível verificar, o desempenho geral dos Métodos foi mediano, sendo que o Método que teve pior desempenho foi "Jaccard", tendo apenas acertado dez das quarenta questões colocadas. O Método que conseguiu maior percentagem de acerto foi o "Cosine", tendo acertado em vinte e seis das sessenta questões.

#### 4.4.8 FAQs com múltiplos Domínios

Tal como foi referido no início deste Capítulo, o propósito de juntar todas as questões reformuladas do mesmo idioma num só ficheiro é averiguar até que ponto o desempenho dos Métodos é afetado. Para tal, vai ser comparado o número de respostas certas a questões reformuladas, para todos os Domínios, antes e depois das FAQs estarem juntas no mesmo ficheiro.

Para poder estudar esta abordagem, foram criados dois ficheiros: um com todas as questões reformuladas (e respetivas respostas) em Inglês, e outro com todas as questões reformuladas em Português.

Relativamente às tabelas que se seguem, um sinal de igual significa que, para um determinado Domínio e utilizando um determinado Método, o número de respostas certas manteve-se. Caso tenhamos um valor antecedido de um sinal de menos, significa que o Método em questão acertou menos esse número de respostas. Já um sinal de mais, indica que esse Método acertou mais esse número de respostas nestas condições de avaliação.

### FAQs com múltiplos Domínios - Resultados removendo as stopwords

FAQs com Múltiplos Domínios

	Básico	Stemming	Lemmatization	Cosine	Levenshtein - Frase	Levenshtein - Palavra	Jaccard	Semantic Similarity
Covid	=	=	=	=	=	=	=	-1
OnePlus	=	=	=	=	-3	-2	-2	-3
UEFA	=	-3	-1	-1	-3	-8	-8	+1
Évora	-1	-1	-1	-2	-1	=	=	-3
Parlamento	=	=	=	=	+1	-2	-1	+3

Figura 4.15: FAQs com Múltiplos Domínios - As stopwords são removidas

Como é possível verificar na tabela, regra geral, todos os Métodos erram mais questões quando misturamos os Domínios das FAQs no mesmo ficheiro. Os casos mais flagrantes da tabela pertencem aos Métodos "Levenshtein - Palavra" e "Jaccard", que para as questões reformuladas da UEFA erraram mais oito respostas do que tinham errado quando os Domínios das FAQs não estavam misturados. É de salientar, também, o fato do Método "Semantic Similarity", para as questões reformuladas do Parlamento, ter acertado mais três respostas do que quando os Domínios não estavam misturados.

### FAQs com múltiplos Domínios - Resultados não removendo as stopwords

FAQs com Múltiplos Domínios

	Básico	Stemming	Lemmatization	Cosine	Levenshtein - Frase	Levenshtein - Palavra	Jaccard	Semantic Similarity
Covid	-1	-1	-1	-1	=	=	=	-2
OnePlus	-1	=	=	-3	-3	-1	-3	-2
UEFA	-8	-7	-7	=	=	-4	-5	+7
Évora	=	=	=	=	=	=	=	=
Parlamento	-1	-2	-1	=	-2	-3	-1	=

Figura 4.16: FAQs com Múltiplos Domínios - As stopwords não são removidas

Como é possível verificar na tabela, à semelhança do que aconteceu na anterior, todos os Métodos erram mais questões quando misturamos os Domínios das FAQs no mesmo ficheiro. O Método "Básico", por exemplo, para as questões reformuladas da UEFA, errou mais oito respostas do que tinha errado quando os Domínios das FAQs não estavam misturados. É de referir que o Método "Semantic Similarity", para as questões reformuladas da UEFA, acertou mais sete respostas do que quando os Domínios não estavam misturados.



# 5

## Caso de Estudo - Instituto da Vinha e do Vinho

Tal como já foi mencionado, o objetivo desta dissertação passa pelo desenvolvimento de um Chatbot que possa, eventualmente, ser utilizado pelo Instituto da Vinha e do Vinho.

Nas secções que se seguem, é possível verificar quais os passos que foram dados até ser escolhida a melhor variante do Chatbot, de forma a que pudesse ser essa a utilizada na FAQ relacionada com o Instituto da Vinha e do Vinho.

Após escolher a variante, a mesma é testada e os respetivos resultados são apresentados para todo o tipos de questões: Questões Não Relacionadas, Questões Reformuladas, e Questões Originais.

### **5.1 Passo 1 - Seleção da melhor resposta**

Como pudemos verificar em capítulos anteriores, nomeadamente no "Desenvolvimento" e na "Avaliação do Sistema", o Chatbot implementado tem quatro variantes: "Resposta - Remove as Stopwords", "Resposta

- Não Remove as Stopwords", "Pergunta e Resposta - Remove as Stopwords", e "Pergunta e Resposta - Não Remove as Stopwords".

Dessas quatro variantes, ficou claro que, se a resposta devolvida ao utilizador for obtida com base nos pares "pergunta-resposta" que constam na FAQ, o número de respostas certas é superior em comparação com as variantes nas quais a resposta devolvida apenas se baseia nas respostas existentes na FAQ. Por essa razão, os resultados das variantes "Resposta - Remove as Stopwords" e "Resposta - Não Remove as Stopwords" foram remetidos para anexo (anexo I e anexo J).

## 5.2 Passo 2 - Seleção da melhor variante

Posto isto, como é possível verificar no capítulo anterior, "Avaliação do Sistema", foi feito um estudo completo sobre as duas variantes do Chatbot que devolvem uma resposta ao utilizador com base nos pares "pergunta-resposta" presentes na FAQ, são elas: "Pergunta e Resposta - Remove as Stopwords" e "Pergunta e Resposta - Não Remove as Stopwords".

Após obter os resultados que as duas variantes acima mencionadas tiveram em todos os Domínios, foi decidido que a variante "Pergunta e Resposta - Remove as Stopwords" seria a escolhida para ser utilizada na FAQ relacionada com o Instituto da Vinha e do Vinho.

## 5.3 Passo 3 - Resultados

Nesta secção é possível verificar os resultados para o Domínio relacionado com o Instituto da Vinha e do Vinho. Tal como foi referido em secções anteriores, a variante do Chatbot utilizada foi a "Pergunta e Resposta - Remove as Stopwords", uma vez que foi, durante os testes, a que obteve melhores resultados.

### 5.3.1 Instituto da Vinha e do Vinho - Questões Não Relacionadas



Figura 5.1: Instituto da Vinha e do Vinho - "Pergunta e Resposta - Remove as stopwords - Questões Não Relacionadas"

Como é possível verificar no gráfico em cima, quando foram colocadas questões não relacionadas com o conteúdo apresentado nas FAQs do Instituto da Vinha e do Vinho, todos os Métodos tiveram o mesmo desempenho, respondendo que não percebiam a questão do utilizador em dezanove das vinte questões colocadas.

### 5.3.2 Instituto da Vinha e do Vinho - Questões Reformuladas

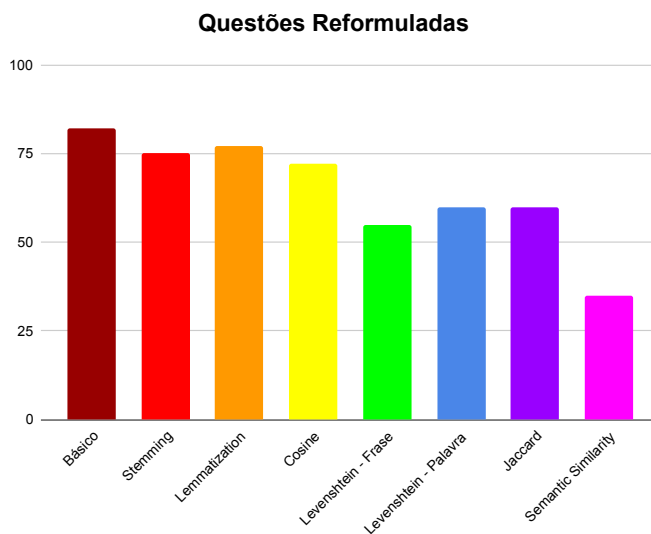


Figura 5.2: Instituto da Vinha e do Vinho - "Pergunta e Resposta - Remove as stopwords - Questões Reformuladas"

Em relação ao gráfico das questões reformuladas, como podemos observar, o desempenho dos Métodos foi, no geral, acima da média. O Método que teve o pior desempenho foi o "Semantic Similarity", tendo acertado apenas sete das vinte questões que lhe foram colocadas. Já o Método com melhor desempenho foi o "Básico", acertando dezasseis das vinte questões que lhe foram colocadas, tendo ainda dito que não percebia uma das questões, daí a a percentagem de oitenta e dois por cento.

### 5.3.3 Instituto da Vinha e do Vinho- Questões Originais

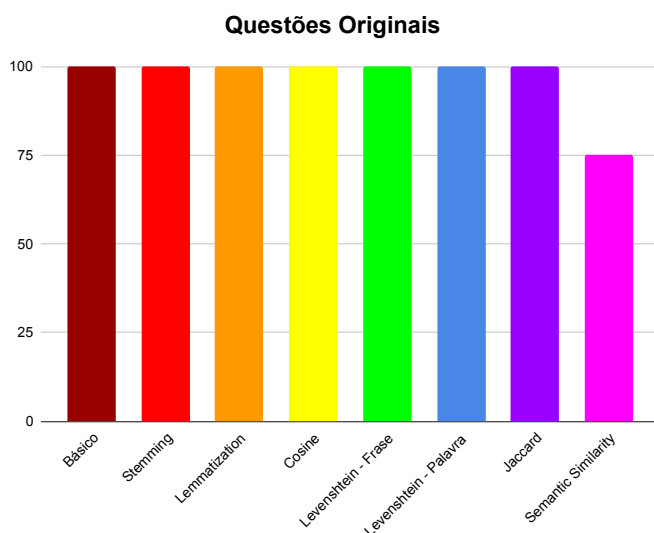


Figura 5.3: Instituto da Vinha e do Vinho- "Pergunta e Resposta - Remove as stopwords - Questões Originais"

Em relação ao último gráfico, este mostra que, quando foram colocadas as questões originais que constam nas FAQs, todos os Métodos acertaram as vinte respostas, exceto o "Semantic Similarity", que falhou cinco.

## 5.4 Comparação de resultados

Como podemos observar ao longo deste Capítulo, os resultados obtidos para a FAQ relacionada com o Instituto da Vinha e do Vinho foram bastante semelhantes aos resultados obtidos para os restantes Domínios, nas mesmas condições de avaliação.

Tendo em conta que os resultados para as questões não relacionadas e para as questões originais quase nunca variam, seja qual for o Domínio, foi decidido que esses resultados não iriam ser comparados.

Precisamente no sentido oposto, temos as questões reformuladas. Neste tipo de questão, como foi possível verificar ao longo do capítulo "Avaliação do Sistema", os resultados dos Métodos variam bastante. Nesse sentido, o gráfico que se segue contém, para cada Domínio das FAQs, o desempenho do melhor Método (à esquerda, a azul), e o desempenho do pior Método (à direita, a vermelho).

De forma a tornar mais clara a leitura do gráfico que vamos ver de seguida (após a tabela), é de salientar que deve ser tida em consideração a informação contida na seguinte tabela:



Domínio da FAQ	Melhor Método (azul)	Pior Método (vermelho)
Covid	Cosine	Semantic Similarity
OnePlus	Stemming	Semantic Similarity
UEFA	Cosine	Semantic Similarity
Évora	Cosine	Jaccard
Parlamento	Básico/Lemmatization	Levenshtein - Frase
Instituto da Vinha e do Vinho	Básico	Semantic Similarity

Tabela 5.1: Domínios das FAQ - Melhor Método e Pior Método

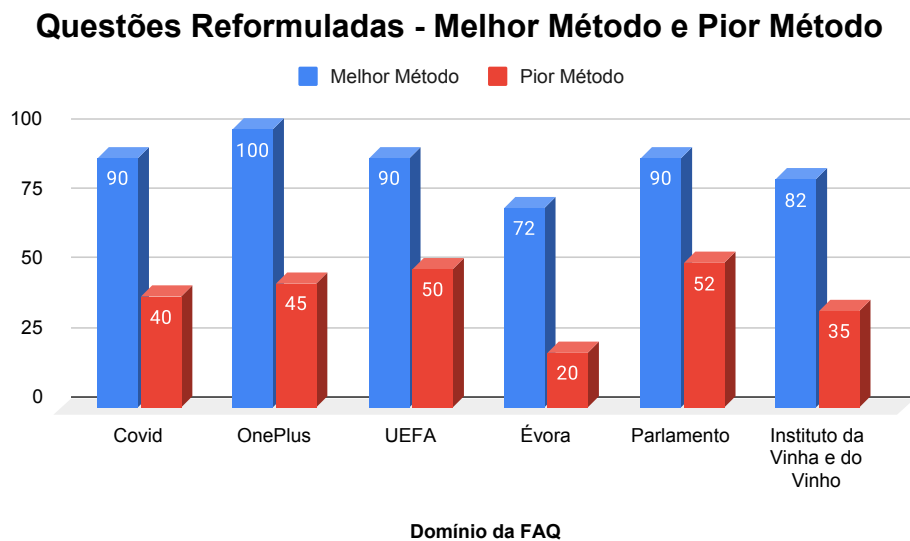


Figura 5.4: Questões Reformuladas - Resultados de cada Domínio

Como podemos verificar pela leitura conjunta da tabela e do gráfico, o Método "Semantic Similarity" é, em quatro dos seis Domínios, aquele que acerta menos respostas, quando colocadas questões reformuladas.

Quanto ao Métodos que acertam mais respostas, podemos ver que o Cosine é mais regular, tendo a melhor percentagem de acerto em três dos seis Domínios.

Posto isto, relativamente às questões reformuladas do Instituto da Vinha e do Vinho, conseguimos verificar, após os resultados apresentados, que o desempenho dos Métodos neste Domínio se manteve na linha dos resultados apresentados para os outros Domínios, no capítulo "Avaliação do Sistema".



# 6

## Conclusões e trabalho Futuro

Nesta dissertação foram mencionados vários Chatbots, sublinhando o papel que todos eles tiveram na história e o seu contributo para que o progresso deste tipo de software fosse cada vez maior e, no fundo, cada vez mais cimentado no nosso dia a dia, não obstante do longo caminho que ainda falta percorrer, não só a nível das funcionalidades, mas também a nível Ético. Também pudemos observar quais os tipos de Chatbot que existem, classificando-os quanto ao tipo de conversas que podem ter e de que forma podem gerar ou obter a resposta que, posteriormente, vão devolver ao utilizador. De forma a conseguir uma implementação sólida no desenvolvimento do sistema, foram estudadas algumas técnicas utilizadas no desenvolvimento de Chatbots, técnicas essas relacionadas com o Processamento de Linguagem Natural, isto é, com a forma como o texto é tratado a partir do momento em que o utilizador coloca alguma questão. Durante o estudo dessas técnicas foi reforçada a importância de fases como o pré-processamento de texto, a tokenização do texto, e a normalização do texto. Como o propósito desta dissertação está relacionado com o desenvolvimento de um Chatbot para responder a FAQs, foi feito um estudo para perceber de que forma é que podemos verificar qual das respostas presentes na FAQ é a mais apropriada para esclarecer a questão que o utilizador coloca. Nesse sentido, foi fundamental perceber que tipos de distâncias podem ser aplicadas entre textos e qual a forma correta de as utilizar.

Após a elaboração do Estado da Arte, e com base nele, teve início o Desenvolvimento do Chatbot. O primeiro passo foi definir todo o ambiente em que seria desenvolvido o sistema, desde a linguagem de Programação ao editor de Código. De notar que, na escolha da linguagem de Programação, foi fulcral escolher uma linguagem com ferramentas prontas a utilizar no que toca ao Processamento de Linguagem Natural. Relativamente aos problemas que surgiram durante a implementação, é de referir que um dos passos mais importantes foi a criação de um grau de similaridade que tem de ser obtido antes de responder ao utilizador. Este grau de similaridade é fundamental para que o sistema não responda a questões que não estejam relacionadas com o tema das FAQs utilizadas, sendo que, quando o grau de similaridade não cumpre os requisitos mínimos, é pedido ao utilizador que reformule a questão colocada. Também é importante referir que, da perspectiva que podemos ter da imprevisibilidade do utilizador, foi muito importante, à luz das questões originais presentes nas FAQs, criar um conjunto de questões reformuladas de forma a permitir que o número de questões acertadas por cada Método fosse o mais próximo possível da realidade.

Com o avançar da implementação, surgiu a necessidade de testar o comportamento de vários sistemas e não de apenas um, como tinha sido planeado inicialmente. Nesse sentido, foram criadas quatro variantes do sistema que apenas diferem em dois aspectos: na forma como procuram obter a resposta devolvida ao utilizador, e no fato de procederem ou não à remoção das stopwords. A ideia, a partir deste ponto, foi testar cada uma das variantes de forma sólida e coerente para que, à luz dos resultados de todas elas, fosse escolhida uma, neste caso a que obtivesse melhores resultados, para ser utilizada nas FAQs relacionadas com o Instituto da Vinha e do Vinho.

Em relação às variantes do sistema, é de referir que todas elas têm Métodos iguais, não existindo nenhuma diferença nesse aspecto. Quanto aos Métodos utilizados, não houve uma preferência específica, simplesmente foram surgindo através do estudo que decorreu durante a elaboração do Estado da Arte, tendo em conta todos os materiais que existem disponíveis. De todo o modo, é de salientar que, obviamente, existiu a preocupação de escolher Métodos que se pudessem complementar e que, de certa modo, conseguissem trazer algo diferente à forma como a resposta era selecionada e, à posteriori, devolvida ao utilizador.

No que toca aos resultados propriamente ditos, foi possível observar que o número de questões acertadas por parte de cada Método variou quando se mudava o Domínio das FAQs, sendo que isto também se deve à imprevisibilidade das questões reformuladas a partir das originais, tendo em conta todas as possibilidades que existem de fazer a mesma pergunta de formas diferentes. Também pudemos verificar que o sistema tende a devolver mais respostas certas quando as FAQs estão em Inglês. Contudo, ficou claro através dos resultados obtidos que, a melhor combinação possível que se pode fazer ao nível das variantes é a seguinte: responder ao utilizador tendo em conta os pares "pergunta-resposta" existentes nas FAQs, e removendo as stopwords.

Nesse sentido, para o Domínio das FAQs relacionadas com o Instituto da Vinha e do Vinho, que se encontram disponíveis em Português, a variante utilizada foi a "Pergunta e Resposta - Remove as Stopwords". Os resultados, como pudemos observar no capítulo anterior ("Avaliação do Sistema"), foram bastante semelhantes aos resultados obtidos nos outros Domínios, o que prova não só a consistência dos Métodos, como a consistência desta variante do Chatbot.

## 6.1 Trabalho futuro

Relativamente a possíveis modificações que melhorariam o sistema, seria interessante que, cada vez que fosse criada uma instância do Chatbot, essa instância pudesse receber como argumentos um identificador da língua que estamos a utilizar nas FAQs e o ficheiro das FAQs. Atualmente, como é possível ver no código em anexo, tanto o identificador como o ficheiro das FAQs são modificados no local do código correspondente, de forma não automatizada.

Em relação ao pré-processamento das questões e das respostas existentes nas FAQs, seria vantajoso que isto fosse feito assim que o sistema tem acesso ao Dicionário onde constam estes pares "questão-resposta". Desta forma, em vez deste pré-processamento ser realizado em cada Método, seria feito uma só vez para os Métodos onde fosse necessário.

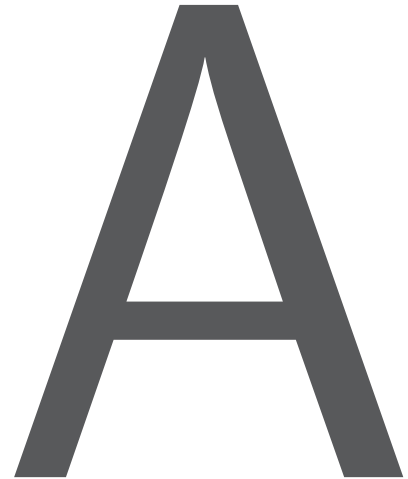
Como é possível verificar no capítulo do "Desenvolvimento", as questões do utilizador para as quais o Chatbot não tem resposta são guardadas num Dicionário. Neste sentido, seria benéfico para o sistema, antes de assumir que estas questões não têm resposta, fazer uma verificação através de sinónimos de forma a tentar identificar uma possível resposta para aquela questão. Se continuasse sem selecionar uma resposta, então adicionava-a ao Dicionário para que fosse revista mais tarde.

Outra modificação que poderia acrescentar algo ao sistema implementa seria a utilização de medidas diferentes daquelas que são utilizadas, com o propósito de verificar se o número de respostas certas aumenta através da utilização de outras medidas.

Quanto aos oito Métodos existentes para calcular respostas, não é de todo plausível que o utilizador coloque uma questão e lhe sejam devolvidas oito respostas. De forma a melhorar este aspeto do sistema, seria fundamental a existência de um classificador automático que conseguisse obter a melhor combinação possível dos diferentes Métodos. Situações como a de um Método acertar muitas vezes quando outro falha, por exemplo, seriam identificadas através de um classificador, sendo que isso levaria, quase de certeza, a um aumento do número de respostas certas.

Outra mudança pertinente seria a adição de medidas que dessem mais peso à parte Semântica, uma vez que a implementação atual se preocupa maioritariamente com a parte Lexical.





## Funções Auxiliares - Código

Neste anexo é possível ver o código de todas as funções auxiliares do Chatbot. A descrição destas funções pode ser vista no capítulo do Desenvolvimento, na secção "Chatbot - Funções Auxiliares".

### Função "preprocessing"

---

```
#User input preprocessing
def preprocessing(user_input):
    user_input = user_input.lower()
    user_input = re.sub(r'[\w\s]', "", user_input)

    tokens = word_tokenize(user_input)

    without_stop_words = {i for i in tokens if i not in stop_words}

    return without_stop_words
```

---

Listing A.1: Função "preprocessing"

### Função "similar\_words"

---

```
#Check similar words
def similar_words(user_input, possible_answer):
    similar_words = 0
    for word in user_input:
        if(word in possible_answer):
            similar_words += 1

    return similar_words
```

---

Listing A.2: Função "similar\_words"

### Função "product\_between\_lists"

---

```
#Product between lists
def product_between_lists(list_one, list_two):
    product_sum = 0
    for i in range(len(list_one)):
        product_sum += list_one[i] * list_two[i]

    return product_sum
```

---

Listing A.3: Função "product\_between\_lists"

### Função "sum\_lists\_in\_list"

---

```
#Sum lists (inner) inside list (outer)
def sum_lists_in_list(list_one):
    final_sum = []

    for inner_list in list_one:
        aux = 0
        for number in inner_list:
            aux += number
        final_sum.append(aux)

    return final_sum
```

---

Listing A.4: Função "sum\_lists\_in\_list"



### Função "words\_edit\_distance"

---

```
#Levenshtein Distance - auxiliar
def words_edit_distance(list_one, list_two):
    distances_list = []

    for word in list_one:
        aux_list = []
        for words in list_two:
            aux_list.append(edit_distance(word, words))

        distances_list.append(min(aux_list))

    sum_distances = sum(distances_list)

    return sum_distances
```

---

Listing A.5: Função "words\_edit\_distance"

### Função "words\_jaccard\_distance"

---

```
#Jaccard Distance - auxiliar
def words_jaccard_distance(list_one, list_two):
    distances_list = []
    for word in list_one:
        aux_list = []
        for words in list_two:
            aux_list.append(jaccard_distance(set(word), set(words)))

        distances_list.append(min(aux_list))

    sum_distances = sum(distances_list)

    return sum_distances
```

---

Listing A.6: Função "words\_jaccard\_distance"



# B

## Chatbot "Pergunta e Resposta - Remove as Stopwords" - Código

Neste anexo é possível ver o código da implementação do Chatbot "Pergunta e Resposta - Remove as Stopwords". A descrição destes Métodos pode ser vista no capítulo do Desenvolvimento, na secção "Chatbot - Exemplo de Implementação".

**Imports necessários do Chatbot**


---

```
#####
### SEMANTIC SIMILARITY ###
import numpy as np
np.seterr(divide = 'ignore', invalid = 'ignore')

import gensim.downloader as api
from gensim.corpora import Dictionary
from gensim.models import TfIdfModel
from gensim.models import WordEmbeddingSimilarityIndex
from gensim.similarities import SparseTermSimilarityMatrix
from gensim.similarities import SoftCosineSimilarity

glove = api.load("glove-wiki-gigaword-50")
#####

#####
### PASTAS INDIVIDUAIS ###
import sys

# insert at 1, 0 is the script path (or '' in REPL)
sys.path.insert(1, 'C:/Users/spart/Desktop/New_Tese/Projecto_Implementação/')

sys.path.insert(1,
    'C:/Users/spart/Desktop/New_Tese/Projecto_Implementação/auxiliar_functions/')

sys.path.insert(1, 'C:/Users/spart/Desktop/New_Tese/Projecto_Implementação/faqs/')
#####

### IMPORTAR FUNÇÕES AUXILIARES ###
from removeStopWords_auxiliar_functions import preprocessing, product_between_lists,
    similar_words, sum_lists_in_list, words_edit_distance, words_jaccard_distance #importa
    das Funções no ficheiro "test.py"

### IMPORTAR AS FAQS ###
from faqs_IVV import answers_dict

### GUARDAR PERGUNTAS SEM RESPOSTA ###
from questions_Need_Answer import questions_dict

from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
from nltk.metrics import edit_distance, jaccard_distance

#LISTA com as possíveis Respostas
answers_list = list(answers_dict.values())

#LISTA com as Perguntas
questions_list = list(answers_dict.keys())
```

---

Listing B.1: Imports necessários do Chatbot

## Classe "ChatBot"

---

```

###Método para começar a conversa###
class ChatBot:
    #LISTA com palavras que permitem ao utilizador sair do chat
    exit_words = ("exit", "goodbye", "quit", "stop")

```

---

Listing B.2: Classe "ChatBot"

## Método "start\_Chat"

---

```

###Método para começar a conversa###
def start_Chat(self):
    print("\n#####")
    print("### You can type 'exit', 'goodbye', 'quit', or 'stop' to leave the chat! ###")
    print("#####")

    get_user_name = input("\nFAQS_Bot: Hello, I'm your FAQS Chatbot! Can you tell me your
        name, please?\n> ")
    while(get_user_name == ""):
        get_user_name = input("\nFAQS_Bot: Hello, I'm your FAQS Chatbot! Can you tell me your
            name, please?\n> ")

    user_input = input("\nFAQS_Bot: Hi, {user_name}, what can I help you with?\n>
        ".format(user_name = get_user_name)).lower()
    while(user_input == ""):
        user_input = input("\nFAQS_Bot: Hi, {user_name}, what can I help you with?\n>
            ".format(user_name = get_user_name)).lower()

    while(not self.exit_Chat(user_input)):
        user_input = self.body_Chat(user_input)

```

---

Listing B.3: Método "start\_Chat"

## Método "questions\_Preprocessing"

---

```

###Método para fazer o préprocessamento das possíveis Respostas###
def questions_Preprocessing(self, questions_list):
    questions_preprocessed = []

    for question in questions_list:
        question_preprocess = preprocessing(question.lower())
        questions_preprocessed.append(question_preprocess)

    return questions_preprocessed

```

---

Listing B.4: Método "questions\_Preprocessing"

### Método "answers\_Preprocessing"

---

```
###Método para fazer o processamento das possíveis Respostas###
def answers_Preprocessing(self, answers_list):
    answers_preprocessed = []

    for answer in answers_list:
        answer_preprocess = preprocessing(answer.lower())
        answers_preprocessed.append(answer_preprocess)

    return answers_preprocessed
```

---

Listing B.5: Método "answers\_Preprocessing"

### Método "user\_Input\_Preprocessing"

---

```
###Método para fazer o processamento do Input do Utilizador###
def user_Input_Preprocessing(self, user_input):
    user_input_preprocessed = preprocessing(user_input)

    return user_input_preprocessed
```

---

Listing B.6: Método "user\_Input\_Preprocessing"

## Método "select\_Answer"

---

```
###Método para seleccionar a Resposta mais parecida com o Input do Utilizador###
def select_Answer(self, listUser, list_Question, list_Answer):
    similar_answer = []
    for answer in list_Answer:
        overlap = similar_words(listUser, answer)
        similar_answer.append(overlap)

    similar_question = []
    for question in list_Question:
        overlap = similar_words(listUser, question)
        similar_question.append(overlap)

    similar_question_max = max(similar_question)
    similar_answer_max = max(similar_answer)

    if(similar_question_max < similar_answer_max):
        select_answer_index = similar_answer.index(similar_answer_max)
    else:
        select_answer_index = similar_question.index(similar_question_max)

    answer_selected = answers_list[select_answer_index]

    similar_max = max(similar_question_max, similar_answer_max)

    return answer_selected, similar_max
```

---

Listing B.7: Método "select\_Answer"

**Método "basic\_Answer"**


---

```

###Método para retornar a Resposta Básica###
def basic_Answer(self, user_input):
    user_input_preprocessed = self.user_Input_Preprocessing(user_input)

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = self.answers_Preprocessing(answers_list)

    answer_selection = self.select_Answer(user_input_preprocessed, questions_preprocessed,
        answers_preprocessed)

    return answer_selection

```

---

Listing B.8: Método "basic\_Answer"

**Método "intermediate\_Answer\_Stemming"**


---

```

###Método para retornar a Resposta Intermédia - Stemming###
def intermediate_Answer_Stemming(self, user_input):
    stemmer = PorterStemmer()

    user_input_preprocessed = self.user_Input_Preprocessing(user_input)

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = self.answers_Preprocessing(answers_list)

    user_input_stemmed = []
    for word in user_input_preprocessed:
        user_input_stemmed.append(stemmer.stem(word))

    questions_list_stemmed = []
    for question in questions_preprocessed:
        aux_list = []
        for word in question:
            aux_list.append(stemmer.stem(word))
        questions_list_stemmed.append(aux_list)

    answers_list_stemmed = []
    for answer in answers_preprocessed:
        aux_list = []
        for word in answer:
            aux_list.append(stemmer.stem(word))
        answers_list_stemmed.append(aux_list)

    answer_selection = self.select_Answer(user_input_stemmed, questions_list_stemmed,
        answers_list_stemmed)

    return answer_selection

```

---

Listing B.9: Método "intermediate\_Answer\_Stemming"



## Método "intermediate\_Answer\_Lemmatization"

---

```

###Método para retornar a Resposta Intermédia - Lemmatization###
def intermediate_Answer_Lemmatization(self, user_input):
    lemmatizer = WordNetLemmatizer()

    user_input_preprocessed = self.user_Input_Preprocessing(user_input)

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = self.answers_Preprocessing(answers_list)

    user_input_lemmatized = []
    for word in user_input_preprocessed:
        user_input_lemmatized.append(lemmatizer.lemmatize(word))

    questions_list_lemmatized = []
    for question in questions_preprocessed:
        aux_list = []
        for word in question:
            aux_list.append(lemmatizer.lemmatize(word))
        questions_list_lemmatized.append(aux_list)

    answers_list_lemmatized = []
    for answer in answers_preprocessed:
        aux_list = []
        for word in answer:
            aux_list.append(lemmatizer.lemmatize(word))
        answers_list_lemmatized.append(aux_list)

    answer_selection = self.select_Answer(user_input_lemmatized, questions_list_lemmatized,
        answers_list_lemmatized)

    return answer_selection

```

---

Listing B.10: Método "intermediate\_Answer\_Lemmatization"

**Método "advanced\_Answer\_Cosine\_Similarity"**


---

```

###Método para retornar a Resposta Avançada - Cosine###
def advanced_Answer_Cosine_Similarity(self, user_input):
    user_input_preprocessed = self.user_Input_Preprocessing(user_input)

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = self.answers_Preprocessing(answers_list)

    #VERSÃO 2
    vector_user_questions = []
    for question in questions_preprocessed:
        vector_user_questions.append(user_input_preprocessed.union(question))

    #VERSÃO 1
    vector_user_answers = []
    for answer in answers_preprocessed:
        vector_user_answers.append(user_input_preprocessed.union(answer))

    #VERSÃO 2
    user_vector_v2 = []
    questions_vector = []
    for i in range(len(vector_user_questions)):
        aux_list_user_v2 = []
        aux_list_question = []
        for word in vector_user_questions[i]:
            if(word in user_input_preprocessed):
                aux_list_user_v2.append(1)
            else:
                aux_list_user_v2.append(0)

            if(word in questions_preprocessed[i]):
                aux_list_question.append(1)
            else:
                aux_list_question.append(0)
        user_vector_v2.append(aux_list_user_v2)
        questions_vector.append(aux_list_question)

    #VERSÃO 1
    user_vector = []
    answers_vector = []
    for i in range(len(vector_user_answers)):
        aux_list_user = []
        aux_list_answer = []
        for word in vector_user_answers[i]:
            if(word in user_input_preprocessed):
                aux_list_user.append(1)
            else:
                aux_list_user.append(0)

            if(word in answers_preprocessed[i]):
                aux_list_answer.append(1)
            else:
                aux_list_answer.append(0)

```

```

user_vector.append(aux_list_user)
answers_vector.append(aux_list_answer)

#VERSÃO 2
product_sum_user_questions = []
for i in range(len(user_vector_v2)):
    product_sum_user_questions.append(product_between_lists(user_vector_v2[i],
        questions_vector[i]))

#VERSÃO 1
product_sum_user_answers = []
for i in range(len(user_vector)):
    product_sum_user_answers.append(product_between_lists(user_vector[i],
        answers_vector[i]))

#VERSÃO 2
list_of_user_sums_v2 = sum_lists_in_list(user_vector_v2)
list_of_question_sums = sum_lists_in_list(questions_vector)

#VERSÃO 1
list_of_user_sums = sum_lists_in_list(user_vector)
list_of_answer_sums = sum_lists_in_list(answers_vector)

#VERSÃO 2
cosine_distance_user_questions = []
for i in range(len(product_sum_user_questions)):
    try:
        formula = product_sum_user_questions[i] / float((list_of_user_sums_v2[i] *
            list_of_question_sums[i]) ** 0.5)
    except ZeroDivisionError:
        cosine_distance_user_questions.append(0)
    cosine_distance_user_questions.append(formula)

#VERSÃO 1
cosine_distance_user_answers = []
for i in range(len(product_sum_user_answers)):
    try:
        formula = product_sum_user_answers[i] / float((list_of_user_sums[i] *
            list_of_answer_sums[i]) ** 0.5)
    except ZeroDivisionError:
        cosine_distance_user_answers.append(0)
    cosine_distance_user_answers.append(formula)

similar_question_max = max(cosine_distance_user_questions)
similar_answer_max = max(cosine_distance_user_answers)

if(similar_question_max < similar_answer_max):
    select_answer_index = cosine_distance_user_answers.index(similar_answer_max)
else:
    select_answer_index = cosine_distance_user_questions.index(similar_question_max)

answer_selected = answers_list[select_answer_index]

similar_max = max(similar_question_max, similar_answer_max)

return answer_selected, similar_max

```

---

Listing B.11: Método "advanced\_Answer\_Cosine\_Similarity"

### Método "advanced\_Answer\_Levenshtein\_Distance\_Phrase"

---

```
###Método para retornar a Resposta Avançada - Distância Levenshtein Frase a Frase###
def advanced_Answer_Levenshtein_Distance_Phrase(self, user_input):
    user_input_preprocessed = self.user_Input_Preprocessing(user_input)

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = self.answers_Preprocessing(answers_list)

    user_input_join = " ".join(user_input_preprocessed)

    questions_join = []
    for question in questions_preprocessed:
        questions_join.append(" ".join(question))

    answers_join = []
    for answer in answers_preprocessed:
        answers_join.append(" ".join(answer))

    levenshtein_distance_user_questions = []
    for question in questions_join:
        levenshtein_distance_user_questions.append(edit_distance(user_input_join, question))

    levenshtein_distance_user_answers = []
    for answer in answers_join:
        levenshtein_distance_user_answers.append(edit_distance(user_input_join, answer))

    similar_question_min = min(levenshtein_distance_user_questions)
    similar_answer_min = min(levenshtein_distance_user_answers)

    if(similar_question_min > similar_answer_min):
        select_answer_index = levenshtein_distance_user_answers.index(similar_answer_min)
    else:
        select_answer_index = levenshtein_distance_user_questions.index(similar_question_min)

    answer_selected = answers_list[select_answer_index]

    similar_min = min(similar_question_min, similar_answer_min)

    return answer_selected, similar_min
```

---

Listing B.12: Método "advanced\_Answer\_Levenshtein\_Distance\_Phrase"

## Método "advanced\_Answer\_Levenshtein\_Distance\_Word"

---

```

###Método para retornar a Resposta Avançada - Distância Levenshtein Palavra a Palavra###
def advanced_Answer_Levenshtein_Distance_Word(self, user_input):
    user_input_preprocessed = list(self.user_Input_Preprocessing(user_input))

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = list(self.answers_Preprocessing(answers_list))

    distance_list_user_questions = []
    for question in questions_preprocessed:
        distance_list_user_questions.append(words_edit_distance(user_input_preprocessed,
            question))

    distance_list_user_answers = []
    for answer in answers_preprocessed:
        distance_list_user_answers.append(words_edit_distance(user_input_preprocessed,
            answer))

    similar_question_min = min(distance_list_user_questions)
    similar_answer_min = min(distance_list_user_answers)

    if(similar_question_min > similar_answer_min):
        select_answer_index = distance_list_user_answers.index(similar_answer_min)
    else:
        select_answer_index = distance_list_user_questions.index(similar_question_min)

    answer_selected = answers_list[select_answer_index]

    similar_min = min(similar_question_min, similar_answer_min)

    return answer_selected, similar_min

```

---

Listing B.13: Método "advanced\_Answer\_Levenshtein\_Distance\_Word"

**Método "advanced\_Answer\_Jaccard\_Distance"**


---

```

###Método para retornar a Resposta Avançada - Distância Palavra a Palavra###
def advanced_Answer_Jaccard_Distance(self, user_input):
    user_input_preprocessed = list(self.user_Input_Preprocessing(user_input))

    questions_preprocessed = self.questions_Preprocessing(questions_list)

    answers_preprocessed = list(self.answers_Preprocessing(answers_list))

    distance_list_user_questions = []
    for question in questions_preprocessed:
        distance_list_user_questions.append(words_jaccard_distance(user_input_preprocessed,
            question))

    distance_list_user_answers = []
    for answer in answers_preprocessed:
        distance_list_user_answers.append(words_jaccard_distance(user_input_preprocessed,
            answer))

    similar_question_min = min(distance_list_user_questions)
    similar_answer_min = min(distance_list_user_answers)

    if(similar_question_min > similar_answer_min):
        select_answer_index = distance_list_user_answers.index(similar_answer_min)
    else:
        select_answer_index = distance_list_user_questions.index(similar_question_min)

    answer_selected = answers_list[select_answer_index]

    similar_min = min(similar_question_min, similar_answer_min)

    return answer_selected, similar_min

```

---

Listing B.14: Método "advanced\_Answer\_Jaccard\_Distance"

## Método "advanced\_Answer\_Semantic\_Similarity"

---

```

###Método para retornar a Resposta Avançada - Distância Palavra a Palavra###
def advanced_Answer_Semantic_Similarity(self, user_input):
    query = self.user_Input_Preprocessing(user_input)

    corpus_questions = self.answers_Preprocessing(questions_list)

    corpus_answers = self.answers_Preprocessing(answers_list)

    similarity_index = WordEmbeddingSimilarityIndex(glove)

    #VERSÃO 2
    # Build the term dictionary, TF-idf model - USER, QUESTIONS
    dictionary_user_questions = Dictionary(corpus_answers + [query])
    tfidf_user_questions = TfidfModel(dictionary = dictionary_user_questions)

    #VERSÃO 1
    # Build the term dictionary, TF-idf model - USER, ANSWERS
    dictionary_user_answers = Dictionary(corpus_answers + [query])
    tfidf_user_answers = TfidfModel(dictionary = dictionary_user_answers)

    #VERSÃO 2
    # Create the term similarity matrix.
    similarity_matrix_user_questions = SparseTermSimilarityMatrix(similarity_index,
        dictionary_user_questions, tfidf_user_questions)

    #VERSÃO 1
    # Create the term similarity matrix.
    similarity_matrix_user_answers = SparseTermSimilarityMatrix(similarity_index,
        dictionary_user_answers, tfidf_user_answers)

    #VERSÃO 2
    query_tf_user_questions = tfidf_user_questions[dictionary_user_questions.doc2bow(query)]

    #VERSÃO 1
    query_tf_user_answers = tfidf_user_answers[dictionary_user_answers.doc2bow(query)]

    #VERSÃO 2
    index_user_questions =
        SoftCosineSimilarity(tfidf_user_questions[[dictionary_user_questions.doc2bow(document)
            for document in corpus_questions]], similarity_matrix_user_questions)

    #VERSÃO 1
    index_user_answers =
        SoftCosineSimilarity(tfidf_user_answers[[dictionary_user_answers.doc2bow(document)
            for document in corpus_answers]], similarity_matrix_user_answers)

    #VERSÃO 2
    doc_similarity_scores_user_questions = index_user_questions[query_tf_user_questions]

    #VERSÃO 1
    doc_similarity_scores_user_answers = index_user_answers[query_tf_user_answers]

    #VERSÃO 2

```

```
sorted_indexes_user_questions = np.argsort(doc_similarity_scores_user_questions[::-1])

#VERSÃO 1
sorted_indexes_user_answers = np.argsort(doc_similarity_scores_user_answers[::-1])

if(sorted_indexes_user_answers[0] > sorted_indexes_user_questions[0]):
    pick_index = sorted_indexes_user_answers[0]
else:
    pick_index = sorted_indexes_user_questions[0]

return answers_list[pick_index]
```

---

Listing B.15: Método "advanced\_Answer\_Semantic\_Similarity"



## Método "body\_Chat"

---

```

###Método para responder ao utilizador###
def body_Chat(self, user_input):

    degree = self.basic_Answer(user_input)[1] +
             self.intermediate_Answer_Stemming(user_input)[1] +
             self.intermediate_Answer_Lemmatization(user_input)[1] +
             self.advanced_Answer_Cosine_Similarity(user_input)[1]

    if(degree <= 4):
        if(user_input not in questions_dict.values()):
            questions_dict_length = len(questions_dict)
            questions_dict[questions_dict_length] = user_input

            with open("../questions_Need_Answer.py", "w") as addNewQuestions:
                addNewQuestions.write("questions_dict = " + str(questions_dict))

            while(user_input == ""):
                user_input = input("\nFAQS_Bot: I can't find any answer to your question. Can you
                ask in a different way, please?\n> ").lower()

            return user_input

    best_basic_response = self.basic_Answer(user_input)[0]
    print("\nFAQS_Bot [Basic_Answer]: " + best_basic_response)

    best_intermediate_response_stemmed = self.intermediate_Answer_Stemming(user_input)[0]
    print("\nFAQS_Bot [Intermediate_Answer_Stemming]: " + best_intermediate_response_stemmed)

    best_intermediate_response_lemmatized =
        self.intermediate_Answer_Lemmatization(user_input)[0]
    print("\nFAQS_Bot [Intermediate_Answer_Lemmatization]: " +
        best_intermediate_response_lemmatized)

    best_advanced_response_cosine = self.advanced_Answer_Cosine_Similarity(user_input)[0]
    print("\nFAQS_Bot [Advanced_Answer_Cosine_Similarity]: " + best_advanced_response_cosine)

    best_advanced_response levenshtein_phrase =
        self.advanced_Answer_Levenshtein_Distance_Phase(user_input)[0]
    print("\nFAQS_Bot [Advanced_Answer_Levenshtein_Distance_Phase]: " +
        best_advanced_response levenshtein_phrase)

    best_advanced_response levenshtein_word =
        self.advanced_Answer_Levenshtein_Distance_Word(user_input)[0]
    print("\nFAQS_Bot [Advanced_Answer_Levenshtein_Distance_Word]: " +
        best_advanced_response levenshtein_word)

    best_advanced_response_jaccard = self.advanced_Answer_Jaccard_Distance(user_input)[0]
    print("\nFAQS_Bot [Advanced_Answer_Jaccard_Distance]: " + best_advanced_response_jaccard)

    best_advanced_response_semantic_similarity =
        self.advanced_Answer_Semantic_Similarity(user_input)

```

```

print("\nFAQS_Bot [Advanced_Answer_Semantic_Similarity]: " +
      best_advanced_response_semantic_similarity)

user_input = input("\nFAQS_Bot: Do you need help with anything else?\n> ").lower()
while(user_input == ""):
    user_input = input("\nFAQS_Bot: Do you need help with anything else?\n> ").lower()

return user_input

```

---

Listing B.16: Método "body\_Chat"

### Método "exit\_Chat"

---

```

###Método para sair da conversa###
def exit_Chat(self, user_input):
    for exit_word in self.exit_words:
        if(exit_word in user_input):
            print("\nFAQS_Bot: I hope you got your questions answered! See you later...\n")
            return True
    return False

```

---

Listing B.17: Método "exit\_Chat"

### Criar Instâncias da Classe

---

```

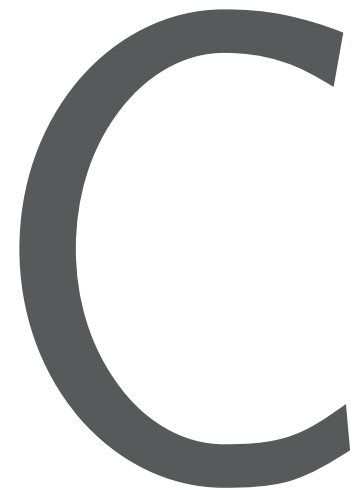
#Instâncias da CLASSE
firstInstance = ChatBot()
firstInstance.start_Chat()

```

---

Listing B.18: Criar Instâncias da Classe





## Covid - Inglês

Neste anexo encontram-se as FAQs referentes ao Domínio relacionado com o Covid, em Inglês.

As FAQs estão organizadas da seguinte forma para cada Domínio: FAQs Originais, Questões Reformuladas, e Questões Não Relacionadas.

Nas FAQs Originais é possível encontrar as FAQs que foram retiradas do site em questão, tal como se encontram nesse site.

Quanto às Questões Reformuladas, é possível verificar a existência de vinte questões que foram reformuladas a partir das FAQs Originais. Esta reformulação é apenas um exemplo, dado que poderiam ter sido reformuladas de qualquer outra forma, desde que fizesse sentido relativamente à questão original.

Por último, nas Questões Não Relacionadas podemos observar um conjunto de vinte questões que não estão, de todo, relacionadas com o tema das FAQs.

### Covid - FAQs Originais

Number	Questions	Answers
1	How can I protect myself?	Visit the How to Protect Yourself & Others page to learn about how to protect yourself from respiratory illnesses, like COVID-19.
2	Does CDC recommend the use of masks to prevent COVID-19?	Wear masks in public settings when around people not living in your household and particularly where other social distancing measures are difficult to maintain, such as grocery stores, pharmacies, and gas stations. Masks may slow the spread of the virus and help people who may have the virus and do not know it from transmitting it to others. COVID-19 can be spread by people who do not have symptoms and do not know that they are infected. That's why it's important for everyone to practice social distancing (staying at least 6 feet away from other people) and wear masks in public settings. Masks provide an extra layer to help prevent the respiratory droplets from traveling in the air and onto other people. The masks recommended are not surgical masks or N-95 respirators. Those are critical supplies that must continue to be reserved for healthcare workers and other medical first responders, as recommended by current CDC guidance. More information about masks can be found on our masks site. <a href="#">How to Wear</a> , <a href="#">How to Wash</a> , <a href="#">How to Make</a>
3	Is it safe to get care for my other medical conditions during this time?	It is important to continue taking care of your health and wellness. Continue your medications, and do not change your treatment plan without talking to your healthcare provider. Continue to manage your disease the way your healthcare provider has told you. Have at least a 2-week supply of all prescription and non-prescription medications. Talk to your healthcare provider about whether your vaccinations are up-to-date. Call your healthcare provider: if you have any concerns about your medical conditions, or if you get sick OR to find out about different ways you can connect with your healthcare provider for chronic disease management or other conditions. Do not delay getting emergency care for your health problems or any health condition that requires immediate attention: If you need emergency help, call 911 OR Emergency departments have infection prevention plans to protect you from getting COVID-19 if you need care for your medical condition. Continue to practice everyday prevention. Wash your hands often, avoid close contact, wear a mask, cover coughs and sneezes, and clean and disinfect frequently touched surfaces often. For more information, see <a href="#">Groups at Higher Risk for Severe Illness</a> .
4	Am I at risk for COVID-19 from mail, packages, or products?	There is still a lot that is unknown about COVID-19 and how it spreads. Coronaviruses are thought to be spread most often by respiratory droplets. Although the virus can survive for a short period on some surfaces, it is unlikely to be spread from domestic or international mail, products or packaging. However, it may be possible that people can get COVID-19 by touching a surface or object that has the virus on it and then touching their own mouth, nose, or possibly their eyes, but this is not thought to be the main way the virus spreads. Learn more about <a href="#">safe handling of deliveries and mail</a> .
5	Is it okay for me to donate blood?	In healthcare settings across the United States, donated blood is a lifesaving, essential part of caring for patients. The need for donated blood is constant, and blood centers are open and in urgent need of donations. CDC encourages people who are well to continue to donate

		blood if they are able, even if they are practicing social distancing because of COVID-19. CDC is supporting blood centers by providing recommendations that will keep donors and staff safe. Examples of these recommendations include spacing donor chairs 6 feet apart, thoroughly adhering to environmental cleaning practices, and encouraging donors to make donation appointments ahead of time.
6	Should contact lens wearers take special precautions to prevent COVID-19?	Currently there is no evidence to suggest contact lens wearers are more at risk for acquiring COVID-19 than eyeglass wearers. Contact lens wearers should continue to practice safe contact lens wear and care hygiene habits to help prevent against transmission of any contact lens-related infections, such as always washing hands with soap and water before handling lenses. People who are healthy can continue to wear and care for their contact lenses as prescribed by their eye care professional. Find more information about how coronavirus spreads and how to protect yourself. Visit CDC's contact lens website for more information on healthy contact lens wear and care.
7	Is contact lens disinfecting solution effective against COVID-19?	Hydrogen peroxide-based systems for cleaning, disinfecting, and storing contact lenses should be effective against the virus that causes COVID-19. For other disinfection methods, such as multipurpose solution and ultrasonic cleaners, there is currently not enough scientific evidence to determine efficacy against the virus. Always use solution to disinfect your contact lenses and case to kill germs that may be present. Handle your lenses over a surface that has been cleaned and disinfected. Find more information about how coronavirus spreads and how to protect yourself. Visit CDC's contact lens website for more information on healthy contact lens wear and care.
8	Should I use soap and water or hand sanitizer to protect against COVID-19?	Handwashing is one of the best ways to protect yourself and your family from getting sick. Wash your hands often with soap and water for at least 20 seconds, especially after blowing your nose, coughing, or sneezing; going to the bathroom; and before eating or preparing food. If soap and water are not readily available, use an alcohol-based hand sanitizer with at least 60% alcohol.
9	What cleaning products should I use to protect against COVID-19?	Clean and disinfect frequently touched surfaces such as tables, doorknobs, light switches, countertops, handles, desks, phones, keyboards, toilets, faucets, and sinks. If surfaces are dirty, clean them using detergent or soap and water prior to disinfection. To disinfect, most common EPA-registered household disinfectants will work. See CDC's recommendations for household cleaning and disinfection.
10	What are the symptoms and complications that COVID-19 can cause?	People with COVID-19 have reported a wide range of symptoms – from mild symptoms to severe illness. Symptoms may appear 2-14 days after exposure to the virus. If you have fever, cough, or other symptoms, you might have COVID-19.
11	When should I seek emergency care if I have COVID-19?	Look for emergency warning signs* for COVID-19. If someone is showing any of these signs, seek emergency medical care immediately: Trouble breathing, Persistent pain or pressure in the chest, New confusion, Inability to wake or stay awake, Bluish lips or face. *This list is not all possible symptoms. Please call your medical provider for any other symptoms that are severe or concerning to you. Call 911 or call ahead to your local emergency facility: Notify the operator that you are seeking care for someone who has or may have COVID-19.
12	Is it possible to have the flu and COVID-19 at the same time?	Yes. It is possible to test positive for flu (as well as other respiratory infections) and COVID-19 at the same time.

13	Who is at increased risk for developing severe illness from COVID-19?	People at increased risk include: Older adults, People of all ages with certain underlying medical conditions. Pregnant people might also be at increased risk of severe illness from COVID-19. Long-standing systemic health and social inequities have put many people from racial and ethnic minority groups at increased risk of getting sick and dying from COVID-19. In addition to those at increased risk, there are certain groups of people who require extra precautions during the pandemic.
14	Are there any medications I should avoid taking if I have COVID-19?	Currently, there is no evidence to suggest that taking any specific medications, like blood pressure medication or ibuprofen, leads to more severe illness from COVID-19. Continue to take your medications and to follow your treatment plan as prescribed by your healthcare provider. Any changes to your medications should only be made after talking with your healthcare provider. Contact your healthcare provider if you have questions or concerns. For more information, see People with Underlying Medical Conditions.
15	Are people with disabilities at higher risk?	Adults with disabilities are more likely to have an underlying medical condition that may put them at increased risk of severe illness from COVID-19 including, but not limited to, heart disease, stroke, diabetes, chronic kidney disease, cancer, high blood pressure, and obesity. In addition, having a disability may make it harder to practice social distancing, wear a mask, and practice hand hygiene. For more information, see People with Disabilities and People who May Need Extra Precautions.
16	What is contact tracing?	Contact tracing has been used for decades by state and local health departments to slow or stop the spread of infectious diseases. Contact tracing slows the spread of COVID-19 by: Letting people know they may have been exposed to COVID-19 and should monitor their health for signs and symptoms of COVID-19, Helping people who may have been exposed to COVID-19 get tested, Asking people to self-isolate if they have COVID-19 or self-quarantine if they are a close contact of someone with COVID-19. During contact tracing, the health department staff will not ask you for: Money, Social Security number, Bank account information, Salary information, Credit card numbers
17	What will happen with my personal information during contact tracing?	Discussions with health department staff are confidential. This means that your personal and medical information will be kept private and only shared with those who may need to know, like your health care provider. If you have been diagnosed with COVID-19, your name will not be shared with those you came in contact with. The health department will only notify people you were in close contact with that they might have been exposed to COVID-19. Each state and jurisdiction use their own method for collecting and protecting health information. To learn more, contact your state or local health department. You may also be interested in: If I participate contact tracing for COVID-19 using a digital tool, is my personal health information secure?
18	Who is considered a close contact to someone with COVID-19?	For COVID-19, a close contact is anyone who was within 6 feet of an infected person for a total of 15 minutes or more. An infected person can spread COVID-19 starting 48 hours (or 2 days) before the person has any symptoms or tests positive for COVID-19.
19	Am I considered a close contact if I was wearing a mask?	Yes, you are still considered a close contact even if you were wearing a mask while you were around someone with COVID-19. Masks are meant to protect other people in case you are infected, and not to protect you from becoming infected.

20	If I am a close contact, will I be tested for COVID-19?	<p>If you have been in close contact with someone who has COVID-19, you should be tested, even if you do not have symptoms of COVID-19. The health department may be able to provide resources for testing in your area. While you are waiting for your COVID-19 test result, stay home away from others (self-quarantine) and monitor your health for symptoms of COVID-19 to protect your friends, family, and others from possibly getting COVID-19. If your test is positive, you should continue to stay home and self-isolate away from others and monitor your health. If you have symptoms of COVID-19 and they worsen or become severe, you should seek emergency medical care. Severe symptoms include trouble breathing, persistent pain or pressure in the chest, confusion, inability to wake or stay awake, or bluish lips or face. Someone from the health department may call you to Check on your health, Discuss who you have been around, and Ask where you have spent time while you may have been able to spread COVID-19 to others. If your test is negative and you don't have symptoms, you should continue to stay home and self-quarantine away from others for 14 days after your last exposure to COVID-19 and follow all recommendations from the health department. This is important because symptoms can appear up to 14 days after you've been exposed and are infected. A negative result before the end of your quarantine period does not rule out possible infection. Additionally, you do not need a repeat test unless you develop symptoms, or if you require a test to return to work. If your test is negative and you have symptoms, you should continue to self-quarantine away from others for 14 days after your last exposure to COVID-19 and follow all recommendations from the health department. Additional medical consultation and a second test may be needed if your symptoms do not improve.</p>
----	---	--



### Covid - Questões Reformuladas

Number	Questions
Question 1	How can I protect my own health?
Question 2	Where should I use masks for prevention?
Question 3	What should I do if I have other health problems?
Question 4	Does COVID-19 spreads by mail or packages?
Question 5	Can I donate blood, or it is safer not doing it?
Question 6	Should I have extra caution knowing that I use contact lens?
Question 7	Should I use any disinfection methods for my contact lens?
Question 8	What are the differences between using soap and water or hand sanitizer?
Question 9	What products can I use to protect against COVID-19?
Question 10	Which are the symptoms caused by COVID-19?
Question 11	What is the correct time to ask for emergency help when infected with COVID-19?
Question 12	Can flu and COVID-19 coexist?
Question 13	What types of people can suffer more from COVID-19?
Question 14	What medications should I avoid if I am infected with COVID-19?
Question 15	Will people with disabilities suffer more from COVID-19?
Question 16	What happens if, somehow, my parcel is lost?
Question 17	Does contact tracing bring any problems to my personal information?
Question 18	How can I be considered a close contact?
Question 19	Let us suppose I am using a mask. Can I be considered a close contact?
Question 20	If I am considered a close contact, does it mean that I have to be tested?

### Questões Não Relacionadas (Inglês)

Number	Questions
Question 1	I need a doctor, is this the correct chat?
Question 2	What is the weather for tomorrow in Lisbon, Portugal?
Question 3	I want a Pixel to take good photos
Question 4	I can get you some lemon with ice, do you want?
Question 5	I want the price of 5 apples
Question 6	Are you bored, Chatbot? You can come to the U.S.
Question 7	Where can I find a Sushi restaurant?
Question 8	If you don't mind, I would like to talk about politics
Question 9	My neighbours make too much noise, how can I avoid this?
Question 10	Computers give humans a great power
Question 11	This summer was very hot
Question 12	Should I call someone if I want extra information?
Question 13	I was told that Bots run very quickly
Question 14	I can't power my computer, any solution for this?
Question 15	Books can tell us as many stories as we want
Question 16	Do you have a football club? If so, do you get angry when they lose?
Question 17	Can you send me anything by boat?
Question 18	What can I do if I want to track my dog?
Question 19	Storms always have some damage to do
Question 20	When I play FPS games, I always want that other players cover me





## OnePlus - Inglês

Neste anexo encontram-se as FAQs referentes ao Domínio relacionado com a OnePlus, em Inglês.

As FAQs estão organizadas da seguinte forma para cada Domínio: FAQs Originais, Questões Reformuladas, e Questões Não Relacionadas.

Nas FAQs Originais é possível encontrar as FAQs que foram retiradas do site em questão, tal como se encontram nesse site.

Quanto às Questões Reformuladas, é possível verificar a existência de vinte questões que foram reformuladas a partir das FAQs Originais. Esta reformulação é apenas um exemplo, dado que poderiam ter sido reformuladas de qualquer outra forma, desde que fizesse sentido relativamente à questão original.

Por último, nas Questões Não Relacionadas podemos observar um conjunto de vinte questões que não estão, de todo, relacionadas com o tema das FAQs.

### OnePlus - FAQs Originais

Number	Questions	Answers
1	Can I purchase specific parts for my device?	Unfortunately, we do not currently sell specific parts.
2	Can I purchase add-ons during the repair process?	Our repair services are strictly limited to repairing or replacing your device. We will not be able to handle add-on accessories you purchased during this process.
3	What is the dead pixel policy?	If your device has more than 3 dead pixels or if the dead pixel is greater than 0.15mm in size, we can offer a replacement.
4	Can I get a refund for price drop?	If you have ordered a OnePlus device on oneplus.com at the original price fewer than 14 days before a price drop announcement, we will issue you a refund for the price difference.
5	How do I contact the insurance provider?	You may contact Simpleurance via email at support.pt@simpleurance.de or call them at 0049 30 688316450
6	Can I change my charger to a U.S. / Europe / U.K. one in my order?	We stock region-specific chargers in our regional warehouses. Alternative versions are not available. However, all of our power adapters are compatible with 100-240V (50-60Hz) power outlets. You can use yours internationally with a proper converter.
7	Where do I find a manual?	You can find the manual at <a href="https://www.oneplus.com/support/manuals">https://www.oneplus.com/support/manuals</a>
8	I can't get a signal	Most signal errors have something to do with your SIM card or network coverage. Make sure the SIM Card is inserted correctly, In OnePlus One if you're not getting a signal at all, make sure your card facing the bottom of the OnePlus One. While NanoSIM to MicroSIM adaptors usually work, some adapters may get stuck in the SIM slot or cause damage to the motherboard. We recommend that you get a new MicroSIM card for your OnePlus one – they are usually free or very inexpensive from your mobile operator. If you're connected, while 2G and 3G networks are usually available, LTE connectivity can often be weaker depending on your area. Make sure that your SIM card's contacts are clean, and if you're using an adapter, it must not move around once placed on the SIM tray. Confirm APN settings with your network provider to avoid no network/low network issues.
9	How do I locate my IMEI number?	The following steps will help you locate your IMEI number: Power up your OnePlus device and open up the dial pad. Enter the code *#06#. Your IMEI number should appear on your screen. We include your IMEI number on the outside of your OnePlus handset's box.
10	What should I do if my phone is not powering on?	Power cycle the device by holding the Power button for 20 seconds, 5 times. Attempt to start the phone by holding down the Power button along with the Volume down button to start recovery mode and performing a factory reset. If these do not solve your problem, Please follow these steps: 1) Let your phone rest for 12 hours (do not charge); 2) Charge with the official charger and usb for 6hrs; 3) After 6hrs disconnect then reconnect your phone to the charger, a low battery screen should appear; 4) Do not attempt to boot. Allow it charge for 1 more hour; 5) After 1 hour, your device should be ready to boot and power-on.

11	Is it normal for my device to overheat while watching movies or playing games?	Device may produce varying levels of heat depending on usage. We have seen temperatures soar while running games and third party applications running in the background. However, we assure you that our devices and hardware have been tested for quality; so rest assured friends, there is nothing to worry about. Please contact customer support if you have any questions or need help.
12	People cannot hear me during calls	Please make sure that your main microphone is not blocked by dust particles. Confirm whether your device's software is up-to-date. If you have the same issue with third-party applications, try using another SIM card.
13	My device's battery drains very quickly	Clear any unused applications that are running the background. Check if your device software is up to date. Please note that latest Android system version with high performance and advanced features, it may not have with same battery backup compared to old versions but it will be optimized with future updates (battery backup may differ based on usage)
14	My device experiences random reboots	Confirm whether the issue is related with a specific application. If you are experiencing problems with a single app, uninstall and reinstall that app. Make sure that your device is up-to-date. If the issue persists, please back up your device and perform a hard reset.
15	My charger doesn't work, and my device won't charge properly	Please make sure the charger/cable is properly connected to the outlet and the device. Try a different charger/cable or power source if possible. If the issue persists, please contact our support team. Please note that long term use of a third-party charger may damage the device and could void your warranty.
16	What should I do if my parcel is lost?	If you feel that your parcel is lost or missing, please contact OnePlus Support ASAP. We will carry out a full investigation to determine the next steps. Once we obtain all of the relevant information it will take 5 to 7 business days to process. We will do our best to handle your case ASAP.
17	What shipping methods do you offer?	We have two shipping methods for most regions - Standard and Priority. You can select one of these during checkout. The cost will be automatically calculated for your order.
18	How do I track my order?	You may track your order by visiting links below depending on the courier used? DHL: DHL Express, USPS: USPS, UPS: UPS, FEDEX: Fedex Express, PLPOST: PLPOST, DEDHL: DEDHL, SFHK: SF Express, Or you can track your order by visiting: OnePlus tracking
19	What should I do if my parcel was damaged upon delivery?	If you noticed that your parcel is damaged when you sign for it, please request the courier to make an abnormal remark on the shipping label. Please keep the original packaging and contact support ASAP. When you contact us, be sure to include pictures and evidence along with your claim. Once we obtain all of the relevant information it will usually take 5 to 7 business days to process.
20	What is covered by the insurance plan?	On-Guard 12 and On-Guard 24 cover breakage (eg. display breakage), damage due to incorrect use, damage by drops and falls, damage from liquids, excessive voltage, induction and short circuit, sabotage and vandalism. More detailed information is available under Term & Conditions upon checkout.

### OnePlus - Questões Reformuladas

Number	Questions
Question 1	The glass of my mobile is broken. Can I buy that specific part alone?
Question 2	Hi guys. I sent my mobile for repairing. Can you send some add-ons together with my mobile?
Question 3	I just noticed that my phone has some dead pixels. What can I do about it?
Question 4	I bought a OnePlus device last week and today the same device had a price drop. Can i get a refund?
Question 5	If I have any problem with my device how can I contact insurance provider?
Question 6	Is it possible to change my charger adapter?
Question 7	How can I find instructions manuals?
Question 8	I'm having some problems with my signal and my network is not working
Question 9	I need to know my IMEI, where can I find the number?
Question 10	My device is not powering on even though my power button is not stuck, what can I do?
Question 11	Should I be worried if my device overheats?
Question 12	I think that my device's microphone is not working
Question 13	My device is losing battery very quickly... Any solution for this?
Question 14	My device started rebooted alone after a certain app was installed, what can I do?
Question 15	My charger is not working, it simply does not power my device
Question 16	What happens if, somehow, my parcel is lost?
Question 17	What kind of shipping methods do you have?
Question 18	How can I track my order?
Question 19	My parcel arrived and the package is damaged
Question 20	What kinds of damage is covered by insurance plan?

### Questões Não Relacionadas (Inglês)

Number	Questions
Question 1	I need a doctor, is this the correct chat?
Question 2	What is the weather for tomorrow in Lisbon, Portugal?
Question 3	I want a Pixel to take good photos
Question 4	I can get you some lemon with ice, do you want?
Question 5	I want the price of 5 apples
Question 6	Are you bored, Chatbot? You can come to the U.S.
Question 7	Where can I find a Sushi restaurant?
Question 8	If you don't mind, I would like to talk about politics
Question 9	My neighbours make too much noise, how can I avoid this?
Question 10	Computers give humans a great power
Question 11	This summer was very hot
Question 12	Should I call someone if I want extra information?
Question 13	I was told that Bots run very quickly
Question 14	I can't power my computer, any solution for this?
Question 15	Books can tell us as many stories as we want
Question 16	Do you have a football club? If so, do you get angry when they lose?
Question 17	Can you send me anything by boat?
Question 18	What can I do if I want to track my dog?
Question 19	Storms always have some damage to do
Question 20	When I play FPS games, I always want that other players cover me





# E

## UEFA - Inglês

Neste anexo encontram-se as FAQs referentes ao Domínio relacionado com a UEFA, em Inglês.

As FAQs estão organizadas da seguinte forma para cada Domínio: FAQs Originais, Questões Não Relacionadas, e Questões Reformuladas.

Nas FAQs Originais é possível encontrar as FAQs que foram retiradas do site em questão, tal como se encontram nesse site.

Quanto às Questões Reformuladas, é possível verificar a existência de vinte questões que foram reformuladas a partir das FAQs Originais. Esta reformulação é apenas um exemplo, dado que poderiam ter sido reformuladas de qualquer outra forma, desde que fizesse sentido relativamente à questão original.

Por último, nas Questões Não Relacionadas podemos observar um conjunto de vinte questões que não estão, de todo, relacionadas com o tema das FAQs.

### UEFA - FAQs Originais

Number	Questions	Answers
1	What is official hospitality?	Official Hospitality is the combination of a match ticket and a catering service. Hospitality packages are available on a first-come-first-served basis and can be purchased by companies or individuals.
2	What am I allowed to do with a hospitality package?	The packages can be used to invite guests (e.g. relative, close friend, employee, colleague, business client, etc.) to see the match with you. You and your guests are not permitted to resell the packages, to combine them with other products or services, to use them for marketing purposes or to use them as part of a lottery or prize for a competition.
3	Do I need to provide the details of my guests?	Yes, details of your guests are required by UEFA, at the latest 7 days before the match. Failure to provide your guests' details may result in the cancellation of your order. Please be aware that UEFA may require a provisional list of guests at any time. Once your order is confirmed, you will be able to enter the names of your guests in your order details on our client portal.
4	What is a private suite product?	It is a private/exclusive area/room with a pitch view and access granted exclusively to the client and their guests. It is always located in the stadium and the seats are located directly in front of the room or inside the room depending on the venue.
5	I am accompanying the group of an official hospitality client, can I get an accreditation or working pass?	No, we do not provide accreditations/working-pass for this purpose. You will need to use one of the purchased packages for such purpose.
6	What is a venue series?	A Venue Series product is composed of Official Hospitality Packages which are purchased for all matches taking place at one venue (i.e 7 matches in London and 4 matches in any other venue).
7	What is an ultimate series?	An Ultimate Series product is a package of Official Hospitality Packages which are purchased for one or both Semi-finals and the Final.
8	What is the main difference between a private suite and a prestige or a 2020 club package?	A Private Suite is private to the client and their guests. A Prestige or 2020 Club package gives access to a shared hospitality facility to which other clients also have access.
9	Is it possible for a company that purchases a suite to add branding elements within the suite?	No, no commercial branding can be added to any Suite.
10	Can I visit the stadium/ my suite before the match?	No, as UEFA is not the owner of the stadium, we cannot organise visits to the stadium or Suites for clients who have purchased hospitality packages.
11	Is it possible to have an extra person in the suite?	No, the number of people per suite cannot be increased.
12	Where can I find your terms and conditions?	The UEFA EURO 2020 Official Hospitality Terms and Conditions are available at the following location: Terms and Conditions
13	Where can I find your refund policy?	The UEFA EURO 2020 Official Hospitality Refund Policy is available at the following location: Refund Policy
14	Where can I find the stadium rules?	The UEFA EURO 2020 Stadium Rules are available at the following

		location: Stadium Rules
15	What is a prestige product?	It is a package that combines a category 1 seat in the stand with access to a shared lounge. In the lounge, dinner and beverages are served to the client and their guests. Depending on the venue, the lounge does not necessarily have a pitch view and can be located either inside the stadium or in a facility outside of the stadium. Facilities outside of the stadium can be within or outside of the stadium's perimeter.
16	What is a 2020 club product?	It is a package that combines a seat in the stand with access to a shared hospitality facility. Such hospitality facility can be a suite or a lounge, with dinner and beverages served to the client and their guests. Depending on the venue, the hospitality facility does not necessarily have a pitch view and can be located either inside or outside of the stadium. Facilities outside of the stadium can be within or outside of the stadium's perimeter.
17	How can I get a discount?	No discounts are offered for the purchase of Official Hospitality packages.
18	When will the kick-off times be known for all the matches?	All kick-off times will be known after the UEFA EURO 2020 final tournament draw on 30 Novembre 2019.
19	Where can I find the correspondence between the match number and the match date?	The UEFA EURO 2020 schedule is available at the following location:
20	Is there wifi in the hospitality areas?	Yes, guests will be provided with connection details upon arrival in the relevant hospitality area.

### UEFA - Questões Reformuladas

Number	Questions
Question 1	Can you provide more information about Official Hospitality?
Question 2	Are there any limitations for what you can do with a Hospitality Package?
Question 3	Do you need any information about my guests?
Question 4	Can you provide more information about Private Suite?
Question 5	Is it possible to get an accreditation if I'm with a Hospitality Client?
Question 6	Can you provide more information about Venue Series?
Question 7	Can you provide more information about Ultimate Series?
Question 8	What the main difference between Private Suite and 2020 Club Package?
Question 9	Can I put some branding elements in my Suite?
Question 10	Is it possible to visit the Stadium before the match? And my Suite?
Question 11	I have an extra friend, is it possible to take him to the Suite?
Question 12	Can you provide more information about Terms and Conditions?
Question 13	Is it possible to cancel any order and get a refund?
Question 14	Is there any information about the rules to accomplish on the Stadium?
Question 15	Can you provide more information about Prestige Product?
Question 16	Can you provide more information about 2020 Club Product?
Question 17	Is it possible to have any type of discount?
Question 18	When will you release official dates for all tournament matches?
Question 19	Where can I find Euro 2020 schedule?
Question 20	Is there any free internet? Where can I find connection details?

### Questões Não Relacionadas (Inglês)

Number	Questions
Question 1	I need a doctor, is this the correct chat?
Question 2	What is the weather for tomorrow in Lisbon, Portugal?
Question 3	I want a Pixel to take good photos
Question 4	I can get you some lemon with ice, do you want?
Question 5	I want the price of 5 apples
Question 6	Are you bored, Chatbot? You can come to the U.S.
Question 7	Where can I find a Sushi restaurant?
Question 8	If you don't mind, I would like to talk about politics
Question 9	My neighbours make too much noise, how can I avoid this?
Question 10	Computers give humans a great power
Question 11	This summer was very hot
Question 12	Should I call someone if I want extra information?
Question 13	I was told that Bots run very quickly
Question 14	I can't power my computer, any solution for this?
Question 15	Books can tell us as many stories as we want
Question 16	Do you have a football club? If so, do you get angry when they lose?
Question 17	Can you send me anything by boat?
Question 18	What can I do if I want to track my dog?
Question 19	Storms always have some damage to do
Question 20	When I play FPS games, I always want that other players cover me





## Évora - Português

Neste anexo encontram-se as FAQs referentes ao Domínio relacionado com a Universidade de Évora, em Português.

As FAQs estão organizadas da seguinte forma para cada Domínio: FAQs Originais, Questões Reformuladas, e Questões Não Relacionadas.

Nas FAQs Originais é possível encontrar as FAQs que foram retiradas do site em questão, tal como se encontram nesse site.

Quanto às Questões Reformuladas, é possível verificar a existência de vinte questões que foram reformuladas a partir das FAQs Originais. Esta reformulação é apenas um exemplo, dado que poderiam ter sido reformuladas de qualquer outra forma, desde que fizesse sentido relativamente à questão original.

Por último, nas Questões Não Relacionadas podemos observar um conjunto de vinte questões que não estão, de todo, relacionadas com o tema das FAQs.



### Évora - FAQs Originais

Número	Questões	Respostas
1	Como submeter dúvidas aos SAC?	Através deste Balcão SAC ONLINE, pode retirar as suas dúvidas, podendo através de “Pesquisa” colocar uma ou mais palavras sobre o assunto que tem dúvidas ou selecionando uma das seguintes categorias relativas ao assunto em que tem dúvidas.
2	Como contactar telefonicamente os SAC?	Os estudantes podem contactar os Serviços Académicos telefonicamente através do 266 760 220 no seguinte horário: segunda, quarta e sexta das 9:30 às 16:00; terça e quinta das 9:30 às 13:00. Chamadas telefónicas de extensões de telefones da Universidade de Évora, é possível através das seguintes extensões contactar os SAC
3	Como obter senha para atendimento presencial?	No intuito de melhorar a qualidade do atendimento presencial nos SAC e reduzir os tempos de espera e a aglomeração de pessoas face ao contexto atual, vimos por este meio informar que a partir de dia 19 de outubro, de 2020, os estudantes da UÉ podem agendar o atendimento presencial para “Procedimentos Académicos para Estudantes da UÉ”. O atendimento presencial nos SAC realiza-se no Balcão do Estudante na Sala 1, do Edifício de Santo Agostinho (Rua Duques de Cadaval) funciona no seguinte horário: segunda, quarta e sexta das 9:30 às 16:00; terça e quinta das 9:30 às 13:00
4	Onde obter informação sobre Serviços e Procedimentos Académicos?	Para além da informação que é disponibilizada neste Balcão SAC.ONLINE, os Serviços Académicos disponibilizam informação no Portal da Universidade de Évora - Estudar: Candidaturas, Mobilidade, Apoios: Prémio e Bolsas de Mérito, Apoios: Programas e Apoios a Estudantes, Estudantes UÉ: Serviços e Procedimentos Académicos, Reconhecimento de Graus Estrangeiros. No perfil de cada estudante no Sistema de Informação Integrado da Universidade de Évora (SIUE) é ainda disponibilizado o Guia de Procedimentos Académicos / Guidelines - Academic Procedures assim como a principal Regulamentação que deve ser do conhecimento dos estudantes da UÉ.
5	Como efetuar a matrícula?	Deverá efetuar a sua matrícula on-line sem ter que se deslocar à Universidade de Évora (Consulte o Guia de matrícula on-line) . Para tal deverá aceder ao Sistema de Informação Integrado da Universidade de Évora (SIUE) em <a href="http://siue.uevora.pt">http://siue.uevora.pt</a> . Procedimentos para efetuar a matrícula: 1º - Aceda ao SIUE através do endereço: <a href="http://siue.uevora.pt">http://siue.uevora.pt</a> , 2º - Utilize as credenciais de acesso – Nome de utilizador (e-mail) e Palavra-passe) – a mesma utilizada na candidatura. (ATENÇÃO: os colocados pelo Concurso Nacional de Acesso ou no Concurso Especial para titulares de cursos de vias profissionais do ensino secundário, ainda não possuem credenciais de acesso, como tal, têm de começar por se registar no SIUE de modo a poderem realizar a sua matrícula (ver Guia de Registo de novo Utilizador)), 3º - Após a autenticação, e para iniciar o processo de Matrícula, aceda à opção Matrículas (No ato da matrícula deve inserir a sua fotografia, para que seja emitido através do Banco Santander o Cartão de Estudante). No caso de ser colocado na Universidade de Évora ao abrigo do Concurso Nacional de Acesso, num curso com pré-requisitos, deverá efetuar matrícula presencialmente, devendo trazer consigo o impresso do pré-requisito do respetivo curso. Após efetuar a matrícula, receberá um e-mail com informação no qual constará o seu nome de utilizador e palavra-passe de estudante. Com estes

		dados, deve aceder ao SIIUE onde deverá efetuar inscrições, aceder aos valores em pagamento e respetivas referências multibanco para pagamento, pedidos de creditação e ainda a consulta da principal regulamentação.
6	O curso tem pré-requisitos, como proceder à matrícula?	Poderá efetuar matrícula on-line*, mas é imprescindível validar a sua matrícula devendo deslocar-se aos Serviços Académicos no mês em que efetua a matrícula ou no subsequente, nas datas previstas para o seu curso que poderá consultar aqui , sendo necessário: - documento de identificação, - visto no caso de estudantes internacionais, - documentos originais da obtenção do grau e/ou certificados de habilitações submetidos na candidatura (caso não os tenha submetido autenticados). Os titulares de habilitações obtidas no estrangeiro, os documentos a apresentar devem estar autenticados pelos Serviços Consulares portugueses no país em que obteve as habilitações estrangeiras ou com a aposição da Apostila de Haia ., - pré-requisitos, se aplicável ao curso em que se matriculou. O impresso (pré-requisito) a entregar para cada um dos cursos com pré-requisitos tem de ser os seguintes a obter no site da DGES: Enfermagem, Pré-requisitos do Grupo A - Comunicação Interpessoal, Medicina Veterinária, Pré-requisitos do Grupo B - Comunicação Interpessoal, Ciências do Desporto, Pré-requisitos do Grupo E - Aptidão Funcional e Física, Arquitetura, Pré-requisitos do Grupo F - Capacidade Visual e Motora, Artes Plásticas e Multimédia, Pré-requisitos do Grupo F - Capacidade Visual e Motora, Design, Pré-requisitos do Grupo F - Capacidade Visual e Motora, *No caso de ser colocado na Universidade de Évora ao abrigo do Concurso Nacional de Acesso, deverá efetuar matrícula presencialmente, devendo trazer consigo o impresso do pré-requisito do respetivo curso.
7	É necessário alguma validação após matrícula?	No caso de não ter ingressado pelo CNA, deverá validar/autenticar os seus documentos, pelo que deverá deslocar-se aos Serviços Académicos, sendo necessário: - documento de identificação, - visto, - documentos originais da obtenção do grau e/ou certificados de habilitações submetidos na candidatura (caso não os tenha submetido autenticados). Os titulares de habilitações obtidas no estrangeiro, os documentos a apresentar devem estar autenticados pelos Serviços Consulares portugueses no país em que obteve as habilitações estrangeiras ou com a aposição da Apostila de Haia ., - pré-requisitos, se aplicável ao curso em que se matriculou. O impresso (pré-requisito) a entregar para cada um dos cursos com pré-requisitos tem de ser os seguintes a obter no site da DGES: Enfermagem, Pré-requisitos do Grupo A - Comunicação Interpessoal, Medicina Veterinária, Pré-requisitos do Grupo B - Comunicação Interpessoal, Ciências do Desporto, Pré-requisitos do Grupo E - Aptidão Funcional e Física, Arquitetura, Pré-requisitos do Grupo F - Capacidade Visual e Motora, Artes Plásticas e Multimédia, Pré-requisitos do Grupo F - Capacidade Visual e Motora
8	Como obter o comprovativo de matrícula / inscrições?	O Comprovativo de Matrícula e Inscrições deve ser obtido online , pelo próprio aluno, através do seu perfil de ALUNO no SIIUE. Para tal, aceda ao SIIUE com os dados de autenticação habituais, e através do menu principal escolha a opção Documentos » Comprovativos, seguido da opção Matrícula e Inscrição. O Comprovativo de Matrícula e Inscrições é assinado digitalmente nos termos previstos no regime jurídico dos documentos eletrónicos e da assinatura digital, Decreto-Lei n.º 290-D/99, de 2 de Agosto, com as posteriores alterações. O documento possui ainda um código de validação que permite a qualquer entidade externa, validar através do SIIUE, a validade do mesmo. Deste modo, este Comprovativo não será assinado/carimbado pelos Serviços Académicos uma vez que o

		mesmo já possui os elementos necessários à verificação da sua validade e autenticidade. O Comprovativo de Matrícula e Inscrição é gratuito e deve ser emitido anualmente, fazendo esta prova da: Condição de estudante no presente ano letivo, Frequência do curso em determinado ano letivo, Lista de Unidades Curriculares a que o estudante está inscrito em determinado ano letivo. Este comprovativo pode ser utilizado para fins de: Abono de família, ADSE, Vistos, Pedidos de empréstimo bancário, Pedido de estatuto de Trabalhador-Estudante (junto da entidade empregadora), etc.
9	Quais os prazos de matrículas?	Os prazos de matrícula decorrem nos prazos estabelecidos no Calendário de Procedimentos Académicos. Para mais informações consulte o Portal da Universidade de Évora.
10	Como obter o cartão de estudante?	Após efetuada a sua matrícula, na qual obrigatoriamente deverá introduzir a fotografia, o cartão de estudante será emitido e será posteriormente notificado para levantar o cartão nos Serviços Académicos (no Edifício Santo Agostinho), assim que o mesmo esteja disponível, sendo necessário no SIUE proceder ao agendamento de atendimento nos SAC, para o seu levantamento. O Cartão de Estudante é obrigatório (sendo necessário para acesso aos edifícios da Universidade de Évora) e o pedido de emissão é automático quando efetua a matrícula. Se perder o cartão, poderá no se perfil de aluno no SIUE, requisitar 2ª via do cartão, sendo depois notificado quando o mesmo estiver disponível para ser levantado no Banco Santander.
11	Como proceder se pretender mudar de especialização/ramo/variante, caso o seu curso as contemple, posteriormente à escolha efetuada no ato de matrícula/inscrição?	Poderá requerer através do GESDOC (Sistema de Gestão Documental) a mudança de especialização/ramo/variante que escolheu no ato da matrícula/inscrição ou no período de alteração de inscrições, devendo ter em consideração que poderá ter obtido aprovação à UC, que não serão consideradas na nova especialização/ramo/variante para a qual pediu mudança. O pedido fica sujeita a aprovação da Reitoria.
12	Como proceder se alterar o meu endereço de email, morada, contacto telefónico, ou outros dados na matrícula?	É imprescindível manter os seus dados de matrícula atualizados. Poderá, a qualquer momento atualizar os dados de contactos, ou fotografia, diretamente através do seu perfil de aluno no SIUE. Caso pretenda a alteração de outros dados, deverá requerer através de requerimento em GESDOC, a alteração dos dados.
13	Como adicionar uma Foto no perfil de aluno do SIUE	Através do perfil ALUNO do SIUE poderá adicionar/substituir a sua fotografia de perfil. A fotografia introduzida deve ser de tipo passe e estar no formato .jpg. Para proceder a essa alteração deverá entrar no SIUE ( <a href="https://siiue.uevora.pt">https://siiue.uevora.pt</a> ) com as suas credenciais de acesso e começar por seleccionar opção de menu "Aluno", seguida da opção lateral "Dados Pessoais" e "Editar Dados Pessoais". Insira a nova fotografia e escolha "Guardar". Caso tenha dificuldade na visualização imediata da fotografia sugerimos que faça um refresh/reload à página do SIUE no seu navegador. Mais informamos que a fotografia introduzida será também utilizada com vista à eventual emissão do seu cartão de estudante. Para tal é essencial que a fotografia seja do tipo passe, caso contrário o cartão de estudante não será emitido.
14	Como proceder se desistir do curso em que está matriculado, querendo anular matrícula?	No caso de pretender anular matrícula (interromper/suspender os estudos), deve fazê-lo através de requerimento a ser submetido através do GESDOC. Se o requerer até 31 de outubro não será considerado devedor de propinas desse ano letivo. Caso o pedido de anulação seja efetuado entre 31 de outubro e 31 de dezembro fica devedor apenas das prestações do semestre ímpar (ficando com inscrição válida apenas neste semestre ímpar) e após esta data (pedidos após 31 de dezembro) fica devedor de todas as prestações do ano letivo. Excepcionalmente em 2020/2021,

		considerando a situação da pandemia e o atraso no início das aulas, o prazo limite para requerer a anulação de matrícula sem ser considerado devedor das propinas do ano letivo 2020/2021 será até 15 de novembro de 2020, sendo que se requerer após 15 de novembro e 31 de dezembro será considerado devedor das prestações do semestre ímpar
15	Se requerer a anulação da matrícula, posteriormente é possível retomar os estudos no mesmo curso na Universidade de Évora?	Ao requerer a anulação de matrícula (através do GesDOC), pode posteriormente candidatar-se a reingresso no mesmo curso ou em curso que o suceda, nos prazos estabelecidos anualmente por despacho reitoral, não podendo, contudo, fazê-lo nas seguintes condições: No caso de requerer anulação da matrícula até 31 de outubro no ano letivo em que ingressa, sendo neste caso necessário candidatar-se de novo a ingresso. No caso de requerer anulação após 31 de outubro, não pode no ano letivo subsequente reingressar, pois a lei impõe que haja um ano letivo de interrupção dos estudos para poder ser efetuado o reingresso, podendo, contudo, reingressar posteriormente. No caso do curso em que pretende reingressar não estar acreditado à data do pedido de reingresso.
16	O que é a prescrição da matrícula?	Entende-se por prescrição a perda do direito de matrícula e inscrição anual quando o estudante não cumpra critérios mínimos de aproveitamento. Neste caso o estudante ficará impedido de se candidatar e inscrever no ciclo de estudos pelo período de dois semestres consecutivos (alínea b) do n.º 3 do artigo 5.º da Lei n.º 37/2003).
17	Valor das propinas	Nos termos do Regulamento de Propinas em vigor (disponível no seu perfil do SIIUE), o valor da propina é independente do número de unidades curriculares em que o estudante se encontre inscrito e do número de ECTS obtido em creditação. Em 2020/21, o valor de propina devido por ano letivo no 1º ciclo e Mestrado Integrado é de 697,00€. No caso de estudantes internacionais a propina anual é de 2500€, a não ser que o estudante com estatuto de estudante internacional tenha bolsa de mérito ou bolsa de cooperação e desenvolvimento. Nos cursos de 2º e 3º ciclo os valores de propinas devidos constam no Despacho de Fixação de Propinas.
18	Quando e como pagar as Propinas?	Pode efetuar o pagamento das propinas da seguinte forma: Na totalidade no período da 1.ª prestação. Em 10 prestações: 1ª prestação até 9 de outubro; 2ª prestação até 31 de outubro; 3ª prestação até 30 de novembro; 4ª prestação até 31 de dezembro; 5ª prestação até 31 de janeiro; 6ª prestação até 28 de fevereiro; 7ª prestação até 31 de março; 8ª prestação até 30 de abril; 9ª prestação até 31 de maio; 10ª prestação até 30 de junho. O pagamento pode ser efetuado utilizando: a rede de Caixas Automáticas Multibanco ou processo equivalente através do Homebanking/Internet; presencialmente na tesouraria sita nos SAC; via Paypal. Esta forma de pagamento é bastante simples e segura, tendo a vantagem da transação ficar validada no SIIUE imediatamente após o pagamento. Para usar este serviço deverá possuir uma conta Paypal. A criação desta conta é gratuita. No caso de estudantes estrangeiros, em que não seja possível o pagamento através dos meios acima referidos, poderão solicitar à Tesouraria, através de e-mail (tesouraria@sac.uevora.pt), os dados para efetuarem a transferência, sendo contudo o pagamento considerado efetivado apenas quando for enviado para o e-mail da Tesouraria o comprovativo da transferência. A informação necessária para proceder ao respetivo pagamento (refª multibanco, entidade, etc.) está disponível no perfil no SIIUE, podendo consultar o Guia de Valores em Pagamento.
19	Onde consultar as referências para pagamento de propinas?	A Universidade de Évora disponibiliza o pagamento de Propinas e outros Emolumentos via rede de Caixas Automáticas Multibanco, ou processo equivalente através do Homebanking, ou ainda através de PayPal. A

		<p>informação necessária para proceder a esses pagamentos está disponível no SIIUE. Para consultar as referências de pagamento deve aceder ao SIIUE - Perfil Aluno, e através do menu Aluno poderá aceder a todos os dados Académicos do aluno, incluindo a consulta dos valores em pagamento e respectivas referências para seu pagamento via Multibanco. Na secção Disponibilização clique em Valores em Pagamento para consultar detalhadamente os valores a pagar (quando existam) e respetivas referências multibanco. Consulte o Guia Rápido SIIUE para obter ajuda sobre este procedimento. Notas importantes: :: Cumpra o prazo de pagamento definido. A falta de pagamento ou o pagamento efectuado para além do prazo limite implicará quer a aplicação de juros de mora, quer a aplicação das penalizações previstas. :: O talão emitido pela Caixa Automática Multibanco ou documento impresso do pagamento por Internet fazem prova de pagamento, devendo por isso ser conservados. :: No prazo médio de 24/48h úteis após o pagamento ser efectuado, o mesmo pode ser consultado através do Menu Ver » Pagamentos (se referente a propinas o mesmo reflectir-se-á também no quadro de propinas). :: Quando o registo do pagamento surgir no SIIUE, poderá obter o respectivo recibo.</p>
20	Dificuldades de Pagamento através de PayPal	<p>A nível do SIIUE, apenas podemos assegurar que os dados de pagamento estejam corretos e garantir a correta ligação ao serviço Paypal. Qualquer outro problema que os alunos tenham com o PayPal podem entrar com o Customer Support deste serviço, através do número 213 665 621</p>

### Évora - Questões Reformuladas

<b>Número</b>	<b>Questões</b>
Questão 1	Tenho algumas dúvidas. Como as posso colocar aos SAC?
Questão 2	Para onde posso ligar para contactar os SAC?
Questão 3	Onde posso obter senha para ser atendido pessoalmente?
Questão 4	Existe alguma informação disponível sobre os Procedimentos Académicos?
Questão 5	De que forma posso fazer a matrícula?
Questão 6	Caso o curso tenha pré-requisitos, como posso fazer a matrícula?
Questão 7	Já fiz a minha matrícula, preciso de a validar?
Questão 8	Onde posso adquirir um comprovativo em como estou matriculado?
Questão 9	Qual é o prazo para me matricular?
Questão 10	Existe algum cartão para os estudantes? Se sim, como o posso obter?
Questão 11	Já escolhi a minha especialização. No entanto, pretendo mudar. É possível?
Questão 12	Já fiz a minha matrícula, mas pretendo alterar alguns dados. É possível?
Questão 13	Pretendo adicionar uma fotografia no SIIUE. Como é que posso fazer isso?
Questão 14	Quero desistir do curso, como posso cancelar a minha matrícula?
Questão 15	Na eventualidade de cancelar a minha matrícula, é possível retomar os estudos mais tarde?
Questão 16	Como é que posso prescrever?
Questão 17	Onde posso consultar o valor que tenho de pagar em termos de propinas?
Questão 18	Como é que posso pagar as minhas propinas?
Questão 19	Existe forma de pagar as propinas através de referência bancária?
Questão 20	Queria pagar as propinas por PayPal, mas não estou a conseguir. Onde posso obter ajuda?

**Questões Não Relacionadas (Português)**

<b>Número</b>	<b>Questões</b>
Pergunta 1	Preciso de um médico, este é o chat correto?
Pergunta 2	Como está o tempo amanhã em Lisboa, Portugal?
Pergunta 3	Preciso de um Pixel para tirar boas fotos
Pergunta 4	Posso trazer limão com gelo, queres?
Pergunta 5	Eu quero o preço de 5 maçãs
Pergunta 6	Estás aborrecido, Chatbot? Podes vir para os EUA
Pergunta 7	Onde é que posso encontrar um restaurante de Sushi?
Pergunta 8	Se não te importares, gostaria de falar sobre política
Pergunta 9	Os meus vizinhos fazem demasiado barulho, como é que posso evitar isto?
Pergunta 10	Os computadores dão aos humanos um grande poder
Pergunta 11	Este verão foi muito quente
Pergunta 12	Devo ligar para alguém se quiser informações adicionais?
Pergunta 13	Disseram-me que os Bots correm muito depressa
Pergunta 14	Não consigo ligar o meu computador, alguma solução para isto?
Pergunta 15	Os livros podem contar-nos tantas histórias quanto quisermos
Pergunta 16	Tens um clube de futebol? Se tiveres, ficas chateado quando eles perdem?
Pergunta 17	Podes enviar-me qualquer coisa por barco?
Pergunta 18	O que posso fazer se quiser rastrear o meu cão?
Pergunta 19	As tempestades causam sempre algum dano
Pergunta 20	Quando jogo jogos FPS, quero sempre que os outros jogadores me cubram







## Parlamento - Português

Neste anexo encontram-se as FAQs referentes ao Domínio relacionado com o Parlamento Português, em Português.

As FAQs estão organizadas da seguinte forma para cada Domínio: FAQs Originais, Questões Reformuladas, e Questões Não Relacionadas.

Nas FAQs Originais é possível encontrar as FAQs que foram retiradas do site em questão, tal como se encontram nesse site.

Quanto às Questões Reformuladas, é possível verificar a existência de vinte questões que foram reformuladas a partir das FAQs Originais. Esta reformulação é apenas um exemplo, dado que poderiam ter sido reformuladas de qualquer outra forma, desde que fizesse sentido relativamente à questão original.

Por último, nas Questões Não Relacionadas podemos observar um conjunto de vinte questões que não estão, de todo, relacionadas com o tema das FAQs.

### Parlamento - FAQs Originais

Número	Questões	Respostas
1	Como são elaboradas as leis?	A iniciativa legislativa cabe aos Deputados ou aos Grupos Parlamentares - neste caso, chamam-se projetos de lei e também ao Governo ou às Assembleias Legislativas das regiões autónomas - neste caso chamam-se propostas de lei. Também grupos de cidadãos eleitores podem exercer o direito de iniciativa legislativa junto da Assembleia da República, bem como participar no procedimento legislativo a que derem origem. Depois de ser admitida pelo Presidente da Assembleia, a iniciativa é objeto de um parecer da comissão especializada a quem foi distribuída, seguindo-se o seu debate na generalidade, sempre feito em reunião plenária, que termina com a votação na generalidade (sobre as linhas gerais da iniciativa). Segue-se um debate e a votação na especialidade (artigo a artigo), que pode ser feito em Plenário ou em Comissão. Finalmente, o texto final é submetido a uma votação final global, sempre feita em Plenário. A iniciativa aprovada chama-se Decreto da Assembleia da República. Após promulgação do Presidente da República e referenda do Primeiro-Ministro, o texto é publicado como Lei no Diário da República.
2	Qual a diferença entre um projeto de lei, uma proposta de lei, uma lei e um decreto-lei?	Nos projetos de lei, a iniciativa legislativa cabe aos Deputados, aos Grupos Parlamentares ou a grupos de cidadãos. Nas propostas de lei, a iniciativa legislativa é do Governo ou das Assembleias Legislativas das Regiões Autónomas. Sendo aprovadas em Plenário da Assembleia da República, estas iniciativas legislativas irão dar origem a uma Lei. O Decreto-Lei é o resultado de um processo legislativo do Governo, cuja tramitação não decorre na Assembleia da República. No entanto, nos trinta dias subsequentes à publicação de um Decreto-Lei que não tenha sido aprovado no exercício da competência legislativa exclusiva do Governo, a requerimento de dez Deputados, pode a Assembleia da República realizar a Apreciação Parlamentar do mesmo. O Plenário poderá então confirmar o diploma, determinar a cessação de vigência ou alterar o diploma. O resultado dessa alteração toma a forma de uma Lei.
3	Como saber o estado em que se encontra uma determinada iniciativa legislativa (projeto de lei ou proposta de lei, por exemplo)?	A tramitação das iniciativas legislativas pode ser consultada aqui, permitindo saber em que fase do processo de apreciação se encontram.
4	Onde posso consultar a atividade de fiscalização da Assembleia da República?	Uma das principais competências constitucionais da Assembleia da República é a de fiscalizar o Governo e da Administração. Para o efeito, a Assembleia tem ao seu dispor vários instrumentos políticos de fiscalização que podem ser consultados (como apreciação de decretos-lei, interpelações ao Governo, perguntas e requerimentos ou inquéritos parlamentares).
5	Como consultar os últimos diplomas aprovados?	Os últimos diplomas aprovados no Parlamento podem ser consultados aqui. Os projetos de lei (da iniciativa dos Deputados, dos grupos parlamentares ou de grupos de cidadãos) e as propostas de lei (da iniciativa do Governo ou das Assembleias Legislativas das regiões autónomas, após aprovação, são publicados como Decretos da Assembleia da República na II Série do Diário da Assembleia da República. Após promulgação pelo Presidente da República e referenda pelo Primeiro-Ministro, são publicados como leis no Diário da República.

6	Como consultar os trabalhos preparatórios de uma Lei ou Resolução?	Pesquisando o diploma pretendido aqui. Se desconhecer qual a Legislatura em que foi aprovado, deverá selecionar o espaço em branco no marcador Legislatura. Depois deve selecionar o tipo de diploma, o ano e o número, ou pesquisar por assunto. No resultado da pesquisa, a partir do título do diploma, navega-se para a página onde aparecem a(s) iniciativa(s) que deram origem ao diploma, bastando carregar na identificação da(s) mesma(s) para aceder à tramitação da iniciativa, também designada por trabalhos preparatórios.
7	Como pesquisar os debates realizados no Plenário da Assembleia da República?	Os debates parlamentares podem ser consultados aqui, permitindo a ligação às gravações vídeo e aos textos publicados em Diário da Assembleia da República. Os Diários podem também ser consultados em Debates Parlamentares, que contém os jornais oficiais do Parlamento desde 1821.
8	Quando têm lugar as reuniões plenárias?	As reuniões plenárias têm lugar, em regra, nas tardes de quarta-feira e quinta-feira e na manhã de sexta-feira. Os agendamentos são disponibilizados aqui.
9	Como assistir a uma reunião plenária?	As reuniões plenárias são públicas, pelo que todos os cidadãos que, individualmente ou em grupos organizados, pretendam assistir podem aceder às galerias da Sala das Sessões, por ordem de chegada e enquanto a capacidade das mesmas o permitir. O acesso às galerias processa-se pela porta da Praça de S. Bento, depois da apresentação, no controlo de segurança, de um documento de identificação. As reuniões plenárias são transmitidas em direto pela ARTV-Canal Parlamento, seguindo-se a sua disponibilização em arquivo no mesmo site.
10	Quando têm lugar as votações?	As votações têm lugar, em regra, à sexta-feira, a partir das 12h00. As votações agendadas e o resultado das votações realizadas podem ser consultados aqui.
11	É possível assistir a reuniões das Comissões Parlamentares?	As reuniões das Comissões Parlamentares são, em regra, públicas. Para assistir, basta dirigir-se a uma das portas do Palácio de S. Bento e indicar qual a reunião a que pretende assistir. Deve ter-se em atenção, contudo, que as salas de reunião têm um número muito limitado de lugares. Porém, a maioria é gravada e difundida pelo Canal Parlamento. Através da WEB-TV (plataforma do Canal Parlamento/AR-TV disponível na Internet), é possível assistir em direto a todas as transmissões televisivas das reuniões que tenham lugar num determinado dia, escolhendo aquela que se quer ver, ao contrário do que sucede no cabo e na TDT.
12	Onde encontrar informação sobre as audições realizadas pelas comissões parlamentares?	A maior parte das audições parlamentares são transmitidas pela ARTV-Canal Parlamento, nomeadamente através do seu site da Internet, e estão também disponíveis em arquivo (organizado por comissões parlamentares). A programação da ARTV-Canal Parlamento pode ser consultada no site da Internet ou na respetiva página de Facebook. Nos sites de cada uma das comissões parlamentares podem ser pesquisadas informações adicionais..
13	Onde posso consultar legislação relevante na área da atividade parlamentar?	O portal da Assembleia da República organiza e disponibiliza a legislação relevante tocante à atividade parlamentar, começando na Constituição da República Portuguesa, passando pelo Regimento da Assembleia da República, Estatuto dos Deputados, etc. Estes diplomas podem ser consultados e descarregados aqui.

14	Como obter os estudos e documentos de apoio à atividade parlamentar produzidos pela Assembleia da República?	Os serviços parlamentares organizam e difundem dossiês de informação e direito comparado, folhas informativas e outros instrumentos de estudo que apoiam os órgãos e serviços do Parlamento. Estes documentos encontram-se disponíveis aqui
15	Qual a composição da Assembleia da República?	A Assembleia da República é composta por 230 Deputados – PS (108), PPD/PSD (79), BE (19), PCP (10), CDS-PP (5), PAN (4), PEV (2), CH (1), IL (1), e L (1), de acordo com os resultados eleitorais, de 6 de outubro de 2019.
16	Como conhecer as atividades (iniciativas apresentadas, intervenções em Plenário, participação em Comissões, atividades na área das Relações Internacionais) de um Deputado?	Através da seleção de um determinado Deputado, pesquisável aqui, deve ser escolhido o menu “Atividade Parlamentar”, dando acesso a informação relativa à atividade do Deputado por Legislatura.
17	Como saber quais os Deputados eleitos por um determinado círculo eleitoral?	A pesquisa de Deputados por círculo eleitoral, mas também por partido e Legislatura, pode ser feita aqui.
18	Como contactar um Deputado?	Acedendo à biografia de um determinado Deputado, pesquisável aqui, pode ser enviada uma mensagem, a partir da ligação “Enviar e-mail” no topo da página do lado direito. Podem também ser contactados os Grupos Parlamentares.
19	Como apresentar uma petição?	Relativamente à Assembleia da República, o direito de petição exerce-se através de uma exposição escrita, dirigida ao Presidente da Assembleia da República, que pode ser assinada por um único cidadão. A Assembleia da República disponibiliza ainda uma plataforma eletrónica para receção de petições e recolha de assinaturas pela Internet.
20	Como apresentar uma iniciativa legislativa de cidadãos?	O direito de iniciativa legislativa de cidadãos é exercido através da apresentação à Assembleia da República de projetos de lei subscritos por um mínimo de 20 000 cidadãos eleitores (definitivamente inscritos no recenseamento eleitoral, no território nacional ou no estrangeiro). Os projetos de lei referidos no número anterior são apresentados por escrito, em papel ou por via eletrónica, estando disponível uma plataforma para a submissão das iniciativas.

### Parlamento - Questões Reformuladas

Número	Questões
Questão 1	Qual o procedimento para criar uma lei?
Questão 2	Como se distingue decreto de lei, lei, projeto de lei, e proposta de lei?
Questão 3	É possível obter informações sobre os avanços de uma determinada iniciativa legislativa?
Questão 4	Como posso obter informações sobre a atividade de fiscalização da Assembleia?
Questão 5	Onde posso verificar os últimos diplomas com aprovação confirmada?
Questão 6	É possível verificar como é preparada uma Lei ou Resolução?
Questão 7	É possível ter acesso aos debates que já foram realizados?
Questão 8	Existe algum calendário relativo às reuniões plenárias?
Questão 9	Onde é que podemos assistir às reuniões plenárias?
Questão 10	Quando é que são as votações?
Questão 11	Existe possibilidade de assistir a reuniões das Comissões Parlamentares?
Questão 12	Existem informações sobre as audições das Comissões Parlamentares?
Questão 13	Qual a legislação da atividade Parlamentar?
Questão 14	Existem documentos de apoio à atividade Parlamentar? Se sim, como os posso obter?
Questão 15	Como é composta a Assembleia da República?
Questão 16	Quais as atividades de um Deputado?
Questão 17	Onde existem informações sobre os Deputados eleitos para um círculo eleitoral?
Questão 18	Queria contactar um Deputado, como é que o posso fazer?
Questão 19	Quais os passos necessários para criar uma petição?
Questão 20	De que forma posso mostrar uma iniciativa legislativa aos cidadãos?

**Questões Não Relacionadas (Português)**

<b>Número</b>	<b>Questões</b>
Pergunta 1	Preciso de um médico, este é o chat correto?
Pergunta 2	Como está o tempo amanhã em Lisboa, Portugal?
Pergunta 3	Preciso de um Pixel para tirar boas fotos
Pergunta 4	Posso trazer limão com gelo, queres?
Pergunta 5	Eu quero o preço de 5 maçãs
Pergunta 6	Estás aborrecido, Chatbot? Podes vir para os EUA
Pergunta 7	Onde é que posso encontrar um restaurante de Sushi?
Pergunta 8	Se não te importares, gostaria de falar sobre política
Pergunta 9	Os meus vizinhos fazem demasiado barulho, como é que posso evitar isto?
Pergunta 10	Os computadores dão aos humanos um grande poder
Pergunta 11	Este verão foi muito quente
Pergunta 12	Devo ligar para alguém se quiser informações adicionais?
Pergunta 13	Disseram-me que os Bots correm muito depressa
Pergunta 14	Não consigo ligar o meu computador, alguma solução para isto?
Pergunta 15	Os livros podem contar-nos tantas histórias quanto quisermos
Pergunta 16	Tens um clube de futebol? Se tiveres, ficas chateado quando eles perdem?
Pergunta 17	Podes enviar-me qualquer coisa por barco?
Pergunta 18	O que posso fazer se quiser rastrear o meu cão?
Pergunta 19	As tempestades causam sempre algum dano
Pergunta 20	Quando jogo jogos FPS, quero sempre que os outros jogadores me cubram





# Instituto da Vinha e do Vinho - Português

Neste anexo encontram-se as FAQs referentes ao Domínio relacionado com o Instituto da Vinha e do Vinho, em Português.

As FAQs estão organizadas da seguinte forma para cada Domínio: FAQs Originais, Questões Reformuladas, e Questões Não Relacionadas.

Nas FAQs Originais é possível encontrar as FAQs que foram retiradas do site em questão, tal como se encontram nesse site.

Quanto às Questões Reformuladas, é possível verificar a existência de vinte questões que foram reformuladas a partir das FAQs Originais. Esta reformulação é apenas um exemplo, dado que poderiam ter sido reformuladas de qualquer outra forma, desde que fizesse sentido relativamente à questão original.

Por último, nas Questões Não Relacionadas podemos observar um conjunto de vinte questões que não estão, de todo, relacionadas com o tema das FAQs.



### Instituto da Vinha e do Vinho - FAQs Originais

Número	Questões	Respostas
1	O que devo fazer para plantar uma vinha?	A plantação de uma vinha só pode ser feita com uma autorização de plantação que suporta legalmente a vinha. Estas autorizações representam uma licença oficial para plantar. As autorizações são válidas por 3 anos civis, a contar da data da sua concessão, não são prorrogáveis e não podem ser transferidas a outras entidades. Quem quiser iniciar a atividade de viticultor terá de obter uma autorização de nova plantação, com a apresentação de uma candidatura. Quem tem vinha com enquadramento legal válido, pode solicitar a emissão de uma autorização de replantação. A formalização dos pedidos de emissão de autorizações de replantação pode ser efetuada nas Direções Regionais de Agricultura e Pescas (DRAP) da área da parcela a arrancar, nos balcões das Associações de Agricultores (AA), nas Comissões Vitivinícolas Regionais (CVR) ou no SIVV em <a href="https://sivv.ivv.min-agricultura.pt">https://sivv.ivv.min-agricultura.pt</a> para o que terá de estar registado nesta plataforma. Estes pedidos são comunicados no SIVV através de Declaração de Arranque no prazo de 30 dias seguidos após o arranque.
2	Após a conclusão da plantação, devo comunica-la? Onde?	Depois de efetuada a plantação da vinha, esta deve ser comunicada no SIVV através de uma Declaração de Plantação (DPLAN) no prazo de 30 dias seguidos após a plantação. Posteriormente a DRAP, após análise da DPLAN, introduz a nova parcela no SIVV que passará a constar do Registo Vitícola (RV). Caso esteja registado no SIVV, pode comunicar a plantação da vinha submetendo um processo de Declaração de Plantação.
3	Existem vários tipos de autorizações?	"Sim, existem vários tipos de autorizações: Autorização de replantação com arranque (ARCA) - Após o arranque da vinha deve ser comunicado no SIVV a Declaração de Arranque, no prazo de 30 dias seguidos a esse arranque. Autorização de replantação sem arranque (ARSA) - o arranque da vinha de compensação é realizado até quatro anos sobre a data da plantação da nova vinha, mediante a constituição de garantia, a favor do IVV, no valor de € 1 500/ha, a qual é devolvida após a confirmação do arranque da vinha de compensação. O pedido de autorização inicia-se no SIVV, da mesma forma que nos processos "ARCA", ou seja através do pedido "Declaração de Arranque".
4	A quem pertencem as autorizações de replantação?	As autorizações de replantação são atribuídas ao titular da vinha arrancada ou a arrancar.
5	Como posso obter Novas Autorizações de Plantação (NAP)?	As NAP são concedidas anualmente.
6	Há algum período definido para pedir Novas Autorizações de Plantação?	A submissão das candidaturas para as NAP decorrerá anualmente entre 1 de abril e 15 de maio e a decisão sobre cada candidatura será conhecida até 1 de agosto do mesmo ano.
7	Existem regras para obter uma NAP?	As NAP estarão sujeitas à aplicação de critérios de elegibilidade e a sua distribuição será feita numa base pro rata ou por critérios de prioridade, caso a área dos pedidos exceda a área disponível.
8	Posso-me candidatar ao Regime de Apoio à Reestruturação e Reconversão de Vinhas (VITIS) com	Não, não se pode candidatar.

	uma Nova Autorização de Plantação?	
9	Qual a área a distribuir anualmente nas novas autorizações?	O número de hectares disponível, a nível nacional, é definido anualmente e pode atingir o valor máximo de 1% da área de vinha plantada.
10	Sou penalizado se não utilizar a minha autorização de nova plantação?	Os viticultores que não utilizem as novas autorizações dentro do seu prazo de validade ficam sujeitos a penalizações administrativas.
11	Quais são as Declarações Obrigatórias?	Declaração de Existências (DE): Constitui uma obrigação de todos os detentores de produtos v\u00ednicos, reportando-se aos volumes detidos a 31 de julho. Declaração de Colheita e Produção (DCP): Constitui uma obrigação de todos os operadores económicos que tenham colhido uvas e/ou tenham produzido mosto/vinho.
12	Quais as entidades que estão obrigadas a apresentar as Declarações Obrigatórias?	Declaração de Existências: As pessoas singulares ou coletivas ou os agrupamentos dessas pessoas, com exceção dos consumidores privados e dos retalhistas, apresentam anualmente uma declaração das existências de mostos de uvas, de mostos de uvas concentrados, de mostos de uvas concentrados retificados e de vinhos na sua posse em 31 de Julho. Declaração de Colheita e Produção: A apresentação da declaração de colheita e produção (DCP) constitui uma obrigação de todos os operadores económicos que tenham colhido uvas e/ou tenham produzido mosto/vinho.
13	Qual o nível de apoio máximo que uma candidatura pode ter?	O nível do financiamento comunitário é de 50% do valor das despesas elegíveis para todos os projetos aprovados. As Associações e Organizações Profissionais e Interprofissionais do sector do vinho, e os organismos públicos diretamente relacionados com o sector do vinho podem beneficiar, em acréscimo, de financiamento nacional até ao limite de 30%. A atribuição de financiamento nacional depende da pontuação final atribuída ao projeto.
14	Este apoio destina-se apenas a PME?	Não. Empresas que não sejam PME também podem concorrer e beneficiar do apoio. Porém, se as candidaturas excederem o valor da dotação orçamental do concurso, só são aprovados os projetos com pontuação mais elevada, dando-se prioridade às PME, através da adição de pontos adicionais.
15	Posso incluir ações de promoção em países não prioritários?	Sim, devendo a candidatura incluir obrigatoriamente a sua caracterização. No entanto, A inclusão de mercados não prioritários será menos valorizada na pontuação do projeto. Não são aceites candidaturas que incluam apenas mercados não prioritários. A aceitação de mercados não prioritários está condicionada à dotação orçamental disponível
16	É possível incluir ações de participação em feiras internacionais que decorram em território da UE (ex.º Prowein na Alemanha)?	Não. Os eventos de caráter internacional que decorram em países da União Europeia não são elegíveis.
17	Tenho projeto QREN aprovado para mercados que quero incluir na candidatura ao apoio à promoção. É possível?	Sim. É possível desde que sejam ações distintas, na forma e/ou temporalmente. Não podem suscitar dúvidas quanto à possibilidade de existência de duplo financiamento.
18	Numa candidatura conjunta, o que acontece se uma das entidades associadas não cumprir alguma das condições de acesso?	A candidatura não é aceite por ser considerada não elegível. Todas as entidades de uma candidatura conjunta são consideradas "beneficiários", pelo que todas têm que cumprir as condições de acesso na data de submissão da candidatura.

19	Existem limites de investimento definidos para apresentação de um projeto?	Sim. Os projetos apresentados a concurso têm de ter um valor mínimo de investimento anual de 10.000 Euros. Os projetos apresentados por empresas, grupos de empresas ou associações destas, de qualquer natureza e forma jurídica, desde que relacionadas com o sector do vinho e organizações de produtores (por ex. cooperativas), são limitados a um valor máximo de investimento anual, de 25% do valor das vendas e serviços prestados, sendo considerada a média do valor que consta nas demonstrações de resultados dos anos 2012, 2013 e 2014. Para os candidatos com atividade inferior a três anos, é considerada 25% da média dos anos existentes e, no mínimo, da demonstração de resultados do ano de 2014.
20	Há um número máximo de mercados que podem constar de um só projeto?	Não. O investimento deve estar focalizado num número de mercados adequado às capacidades técnica, económica, financeira e aos recursos humanos do candidato. Valores de investimento desadequados pioram a qualidade da candidatura, podendo mesmo levar à sua reprovação.

### Instituto da Vinha e do Vinho - Questões Reformuladas

<b>Número</b>	<b>Questões</b>
Questão 1	Quais os passos para proceder à plantação de uma vinha?
Questão 2	Caso tenha terminado a plantação, preciso de a comunicar?
Questão 3	Quais os diferentes tipos de autorização?
Questão 4	Quem são os detentores das autorizações, em casa do replantação?
Questão 5	Gostaria de obter novas autorizações para plantar. É possível?
Questão 6	Caso seja possível obter novas autorizações, qual o período em que o devo fazer?
Questão 7	Que regras estão associadas à obtenção de novas autorizações?
Questão 8	Será que me posso candidatar ao VITIS se obtiver uma nova autorização para plantar?
Questão 9	Qual o número de hectares para novas plantações?
Questão 10	Caso não possa utilizar a minha autorização de plantação, serei penalizado?
Questão 11	Existem algumas declarações que sejam obrigatórias? Se sim, quais?
Questão 12	Relativamente às declarações que são obrigatórias, quem é obrigado a apresentá-las?
Questão 13	Em termos financeiros, qual o apoio máximo que podemos ter se a candidatura for aceite?
Questão 14	Os apoios em questão destinam-se só às Pequenas e Médias Empresas?
Questão 15	Posso fazer marketing em países que não sejam prioritários?
Questão 16	Posso promover o produto em feiras internacionais?
Questão 17	Se já tiver um projeto QREN com aprovação total, posso incluir esse projeto numa possível candidatura. Posso?
Questão 18	Caso decidir avançar para uma candidatura conjunta, o que acontece se uma das partes não cumprir as condições exigidas?
Questão 19	Na eventualidade de apresentar um projeto, quais os limites em termos de investimento?
Questão 20	Existe um número limite de mercados onde pode estar o mesmo projeto?

**Questões Não Relacionadas (Português)**

<b>Número</b>	<b>Questões</b>
Pergunta 1	Preciso de um médico, este é o chat correto?
Pergunta 2	Como está o tempo amanhã em Lisboa, Portugal?
Pergunta 3	Preciso de um Pixel para tirar boas fotos
Pergunta 4	Posso trazer limão com gelo, queres?
Pergunta 5	Eu quero o preço de 5 maçãs
Pergunta 6	Estás aborrecido, Chatbot? Podes vir para os EUA
Pergunta 7	Onde é que posso encontrar um restaurante de Sushi?
Pergunta 8	Se não te importares, gostaria de falar sobre política
Pergunta 9	Os meus vizinhos fazem demasiado barulho, como é que posso evitar isto?
Pergunta 10	Os computadores dão aos humanos um grande poder
Pergunta 11	Este verão foi muito quente
Pergunta 12	Devo ligar para alguém se quiser informações adicionais?
Pergunta 13	Disseram-me que os Bots correm muito depressa
Pergunta 14	Não consigo ligar o meu computador, alguma solução para isto?
Pergunta 15	Os livros podem contar-nos tantas histórias quanto quisermos
Pergunta 16	Tens um clube de futebol? Se tiveres, ficas chateado quando eles perdem?
Pergunta 17	Podes enviar-me qualquer coisa por barco?
Pergunta 18	O que posso fazer se quiser rastrear o meu cão?
Pergunta 19	As tempestades causam sempre algum dano
Pergunta 20	Quando jogo jogos FPS, quero sempre que os outros jogadores me cubram





## Resultados "Resposta - Remove as Stopwords"

Neste anexo é possível verificar os resultados referentes à variante do Chatbot "Resposta - Remove as Stopwords". Tal como o nome da variante indica, a resposta devolvida ao utilizador apenas tem em conta as respostas presentes nas FAQs, independentemente do Domínio. Esta variante remove as stopwords.

## I.1 Covid

Nesta secção é possível verificar os resultados para o Domínio relacionado com o Covid.

### I.1.1 Covid - Questões Não Relacionadas



Figura I.1: Covid - "Resposta - Remove as stopwords - Questões Não Relacionadas"

### I.1.2 Covid - Questões Reformuladas

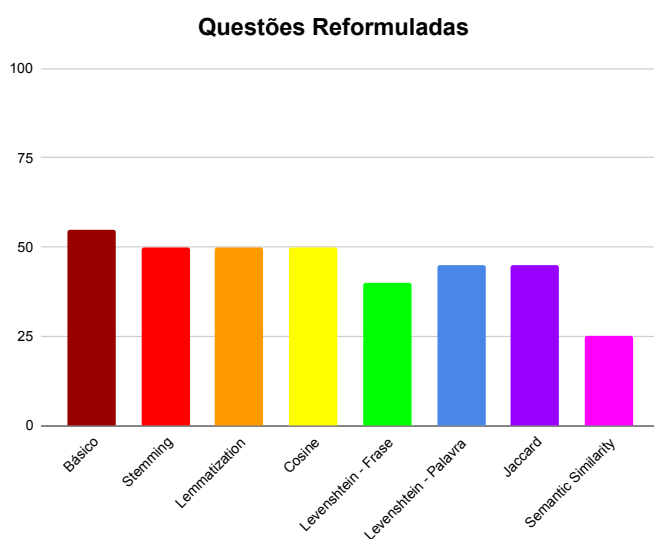


Figura I.2: Covid - "Resposta - Remove as stopwords - Questões Reformuladas"



### I.1.3 Covid - Questões Originais

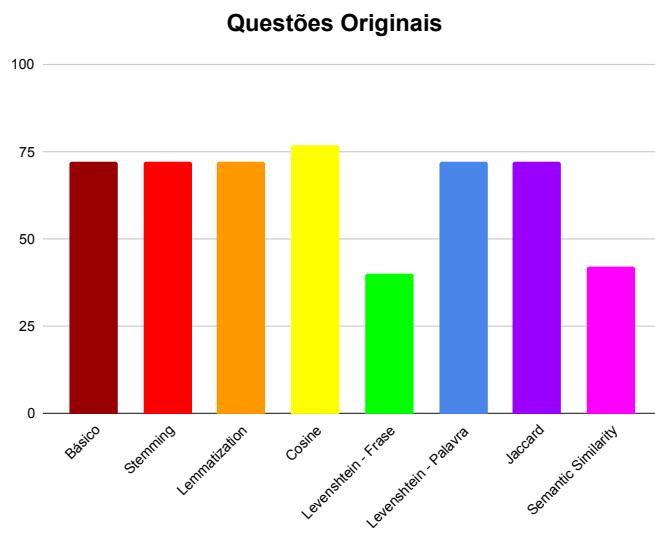


Figura I.3: Covid - "Resposta - Remove as stopwords - Questões Originais"

## 1.2 OnePlus

Nesta secção é possível verificar os resultados para o Domínio relacionado com a OnePlus.

### 1.2.1 OnePlus - Questões Não Relacionadas



Figura I.4: OnePlus - "Resposta - Remove as stopwords - Questões Não Relacionadas"

### 1.2.2 OnePlus - Questões Reformuladas

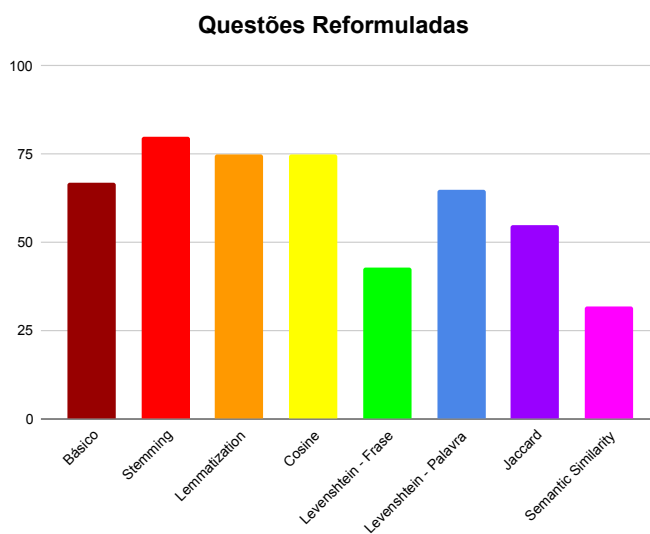


Figura I.5: OnePlus - "Resposta - Remove as stopwords - Questões Reformuladas"

### I.2.3 OnePlus - Questões Originais

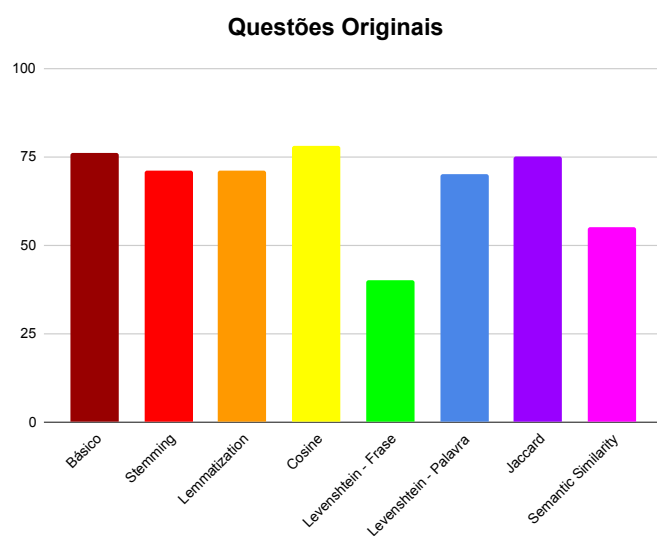


Figura I.6: OnePlus - "Resposta - Remove as stopwords - Questões Originais"

### I.3 UEFA

Nesta secção é possível verificar os resultados para o Domínio relacionado com a UEFA.

#### I.3.1 UEFA - Questões Não Relacionadas



Figura I.7: UEFA - "Resposta - Remove as stopwords - Questões Não Relacionadas"

#### I.3.2 UEFA - Questões Reformuladas

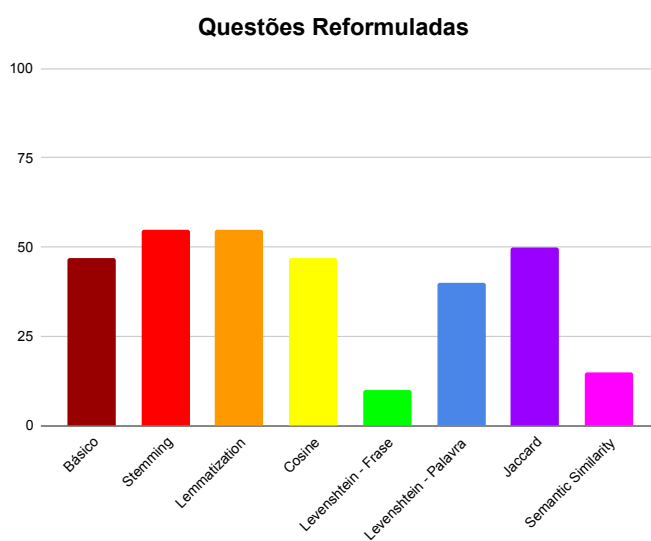


Figura I.8: UEFA - "Resposta - Remove as stopwords - Questões Reformuladas"

### I.3.3 UEFA - Questões Originais

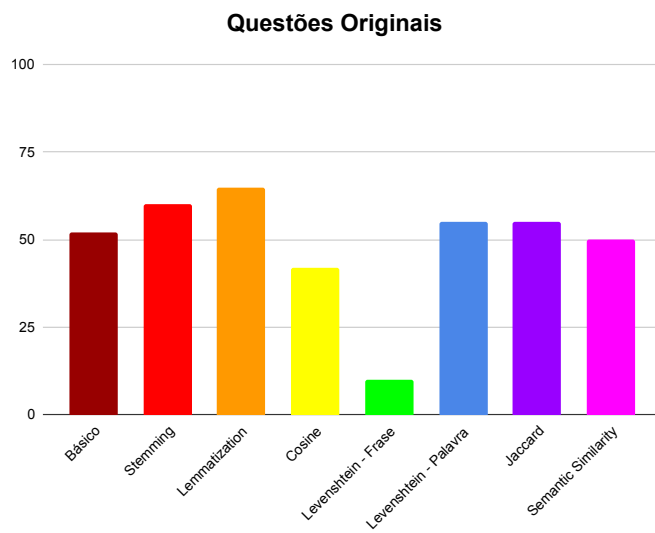


Figura I.9: UEFA - "Resposta - Remove as stopwords - Questões Originais"

## I.4 Évora

Nesta secção é possível verificar os resultados para o Domínio relacionado com a Universidade de Évora.

### I.4.1 Évora - Questões Não Relacionadas



Figura I.10: Évora - "Resposta - Remove as stopwords - Questões Não Relacionadas"

### I.4.2 Évora - Questões Reformuladas

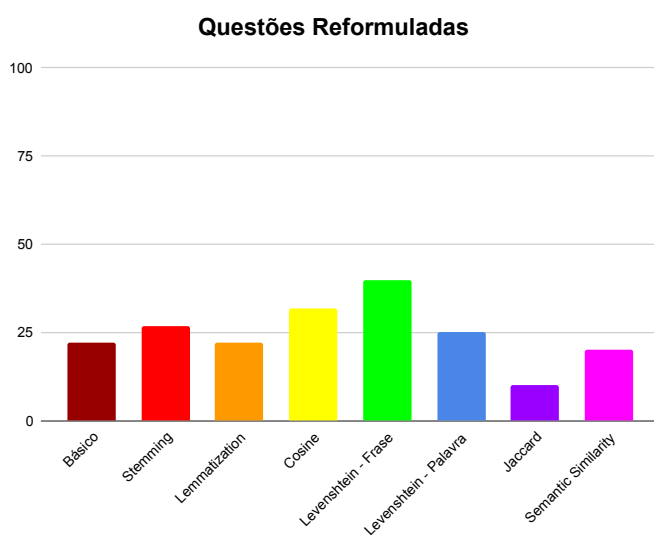


Figura I.11: Évora - "Resposta - Remove as stopwords - Questões Reformuladas"

### I.4.3 Évora - Questões Originais

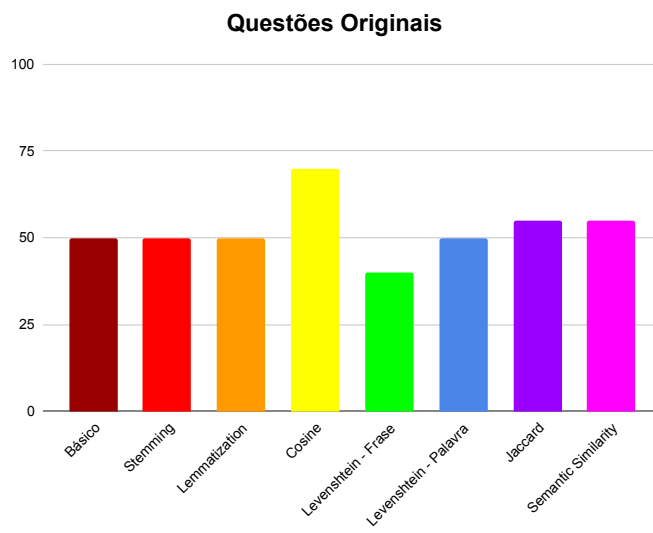


Figura I.12: Évora - "Resposta - Remove as stopwords - Questões Originais"

## 1.5 Parlamento

Nesta secção é possível verificar os resultados para o Domínio relacionado com o Parlamento Português.

### 1.5.1 Parlamento - Questões Não Relacionadas



Figura I.13: Parlamento - "Resposta - Remove as stopwords - Questões Não Relacionadas"

### 1.5.2 Parlamento - Questões Reformuladas

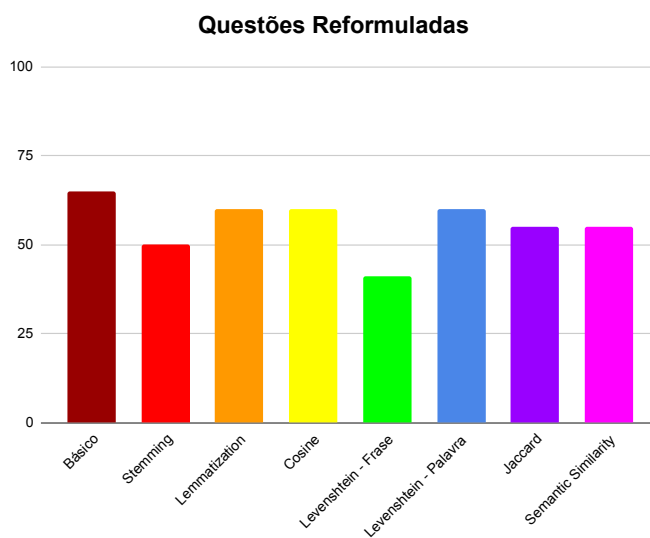


Figura I.14: Parlamento - "Resposta - Remove as stopwords - Questões Reformuladas"



### I.5.3 Parlamento- Questões Originais

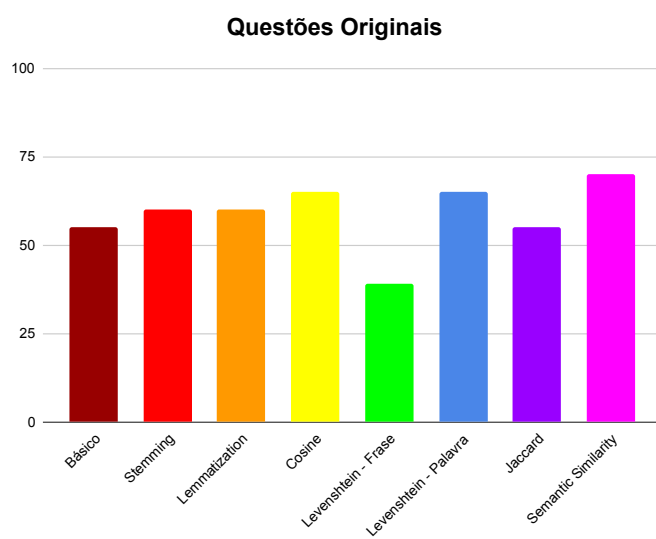


Figura I.15: Parlamento - "Resposta - Remove as stopwords - Questões Originais"

## I.6 Todas as Questões Reformuladas (Inglês)

Nesta secção é possível verificar os resultados para todas as questões reformuladas dos diferentes Domínios que estão em Inglês (Covid, OnePlus, UEFA).

### I.6.1 Todas as Questões Reformuladas (Inglês)

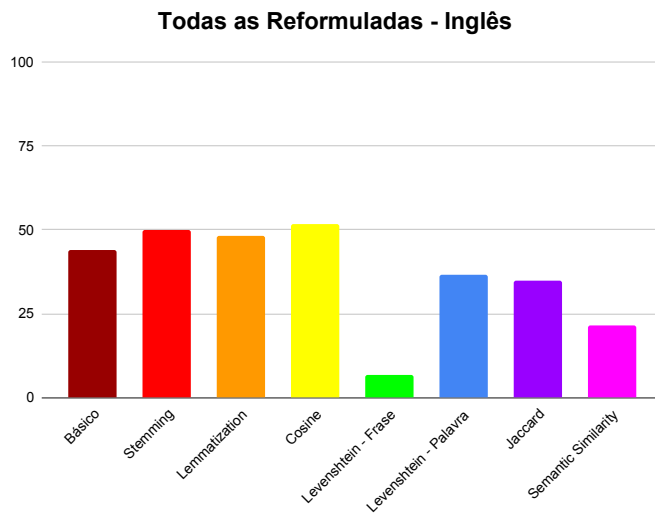


Figura I.16: Todas as Questões Reformuladas (Inglês) - "Resposta - Remove as stopwords"

## I.7 Todas as Questões Reformuladas (Português)

Nesta secção é possível verificar os resultados para todas as questões reformuladas dos diferentes Domínios que estão em Português (Évora e Parlamento).

### I.7.1 Todas as Questões Reformuladas (Português)

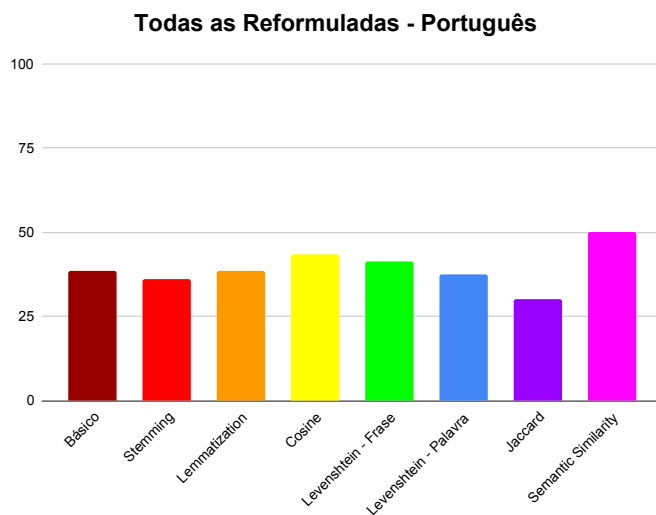


Figura I.17: Todas as Questões Reformuladas (Português) - "Resposta - Remove as stopwords"





## Resultados "Resposta - Não remove as Stopwords"

Neste anexo é possível verificar os resultados referentes à variante do Chatbot "Resposta - Não remove as Stopwords". Tal como o nome da variante indica, a resposta devolvida ao utilizador apenas tem em conta as respostas presentes nas FAQs, independentemente do Domínio. Esta variante não remove as stopwords.

## J.1 Covid

Nesta secção é possível verificar os resultados para o Domínio relacionado com o Covid.

### J.1.1 Covid - Questões Não Relacionadas

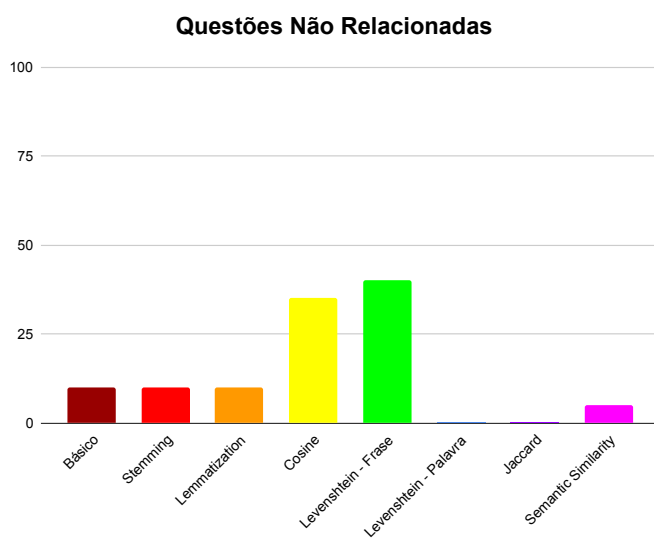


Figura J.1: Covid - "Resposta - Não remove as stopwords - Questões Não Relacionadas"

### J.1.2 Covid - Questões Reformuladas

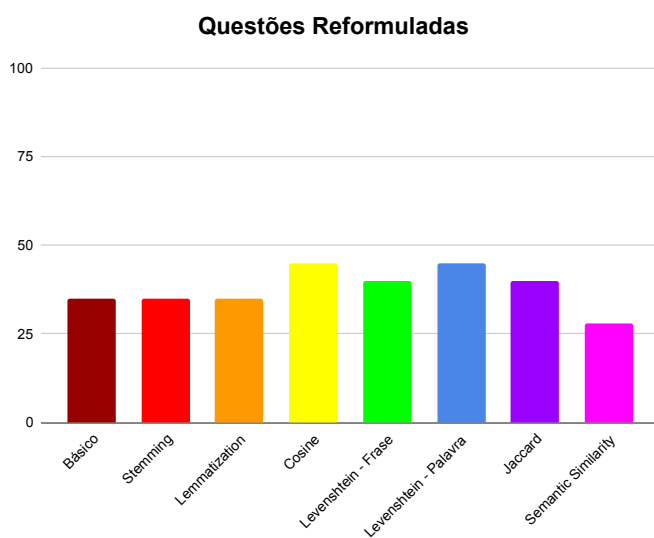


Figura J.2: Covid - "Resposta - Não remove as stopwords - Questões Reformuladas"

### J.1.3 Covid - Questões Originais

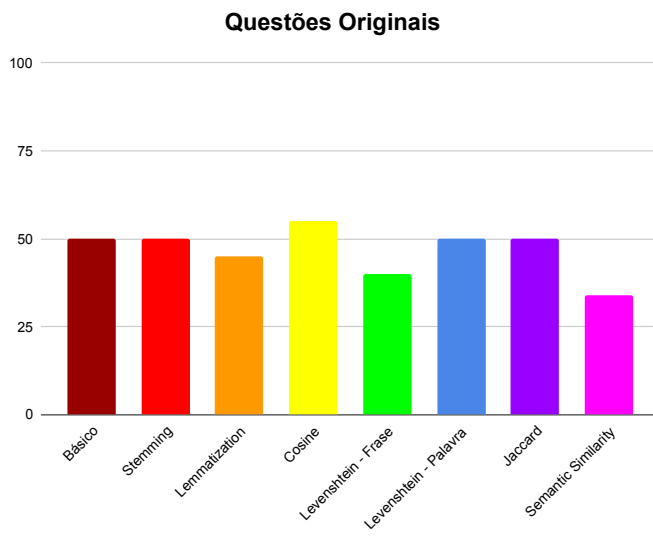


Figura J.3: Covid - "Resposta - Não remove as stopwords - Questões Originais"

## J.2 OnePlus

Nesta secção é possível verificar os resultados para o Domínio relacionado com a OnePlus.

### J.2.1 OnePlus - Questões Não Relacionadas



Figura J.4: OnePlus - "Resposta - Não remove as stopwords - Questões Não Relacionadas"

### J.2.2 OnePlus - Questões Reformuladas

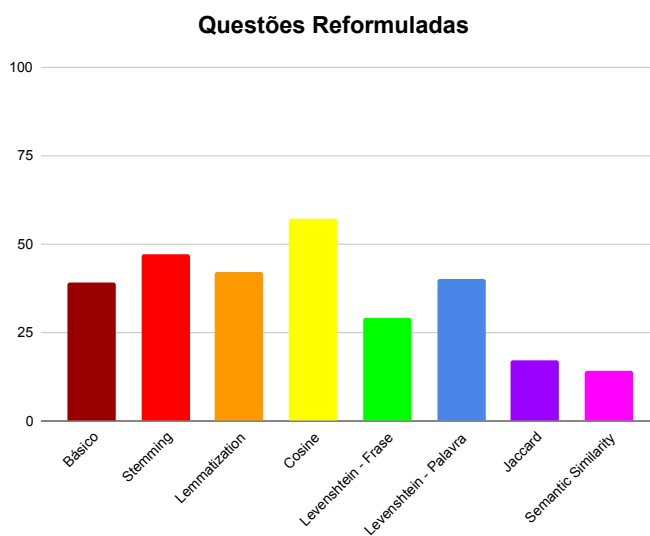


Figura J.5: OnePlus - "Resposta - Não remove as stopwords - Questões Reformuladas"



### J.2.3 OnePlus - Questões Originais

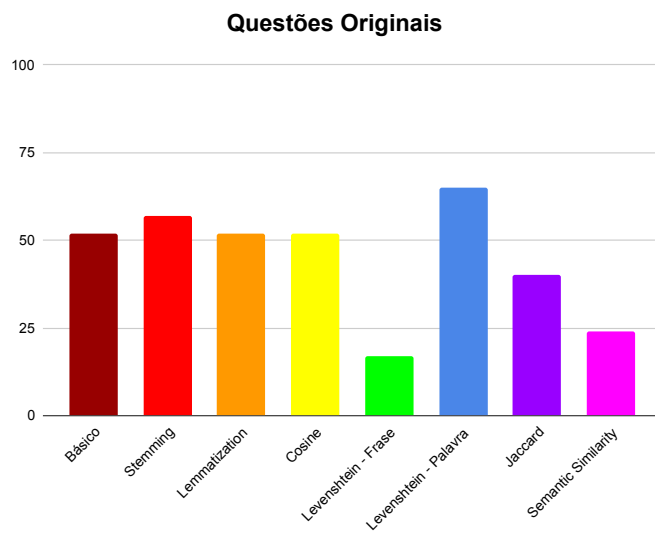


Figura J.6: OnePlus - "Resposta - Não remove as stopwords - Questões Originais"

### J.3 UEFA

Nesta secção é possível verificar os resultados para o Domínio relacionado com a UEFA.

#### J.3.1 UEFA - Questões Não Relacionadas



Figura J.7: UEFA - "Resposta - Não remove as stopwords - Questões Não Relacionadas"

#### J.3.2 UEFA - Questões Reformuladas

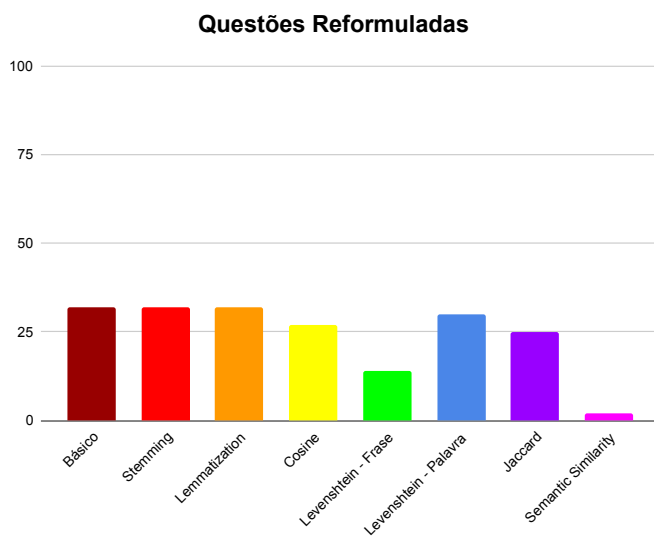


Figura J.8: UEFA - "Resposta - Não remove as stopwords - Questões Reformuladas"

### J.3.3 UEFA - Questões Originais

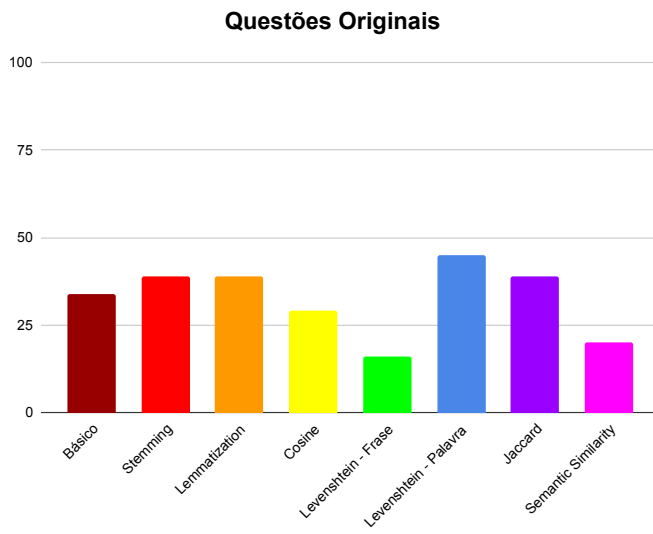


Figura J.9: UEFA - "Resposta - Não remove as stopwords - Questões Originais"

## J.4 Évora

Nesta secção é possível verificar os resultados para o Domínio relacionado com a Universidade de Évora.

### J.4.1 Évora - Questões Não Relacionadas



Figura J.10: Évora - "Resposta - Não remove as stopwords - Questões Não Relacionadas"

### J.4.2 Évora - Questões Reformuladas

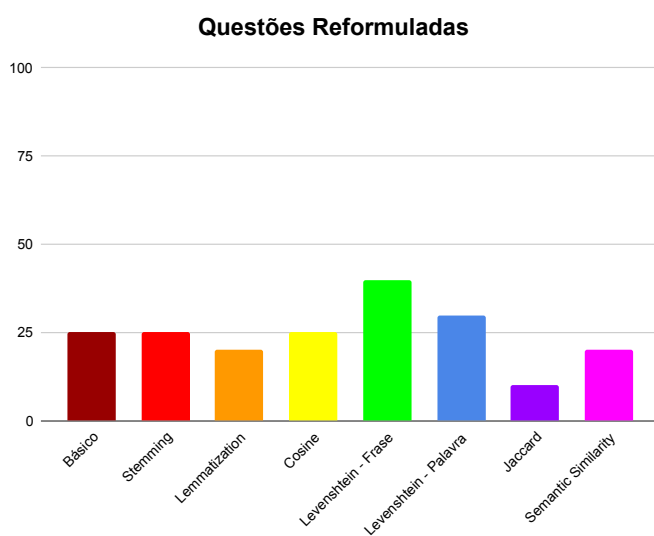


Figura J.11: Évora - "Resposta - Não remove as stopwords - Questões Reformuladas"

### J.4.3 Évora - Questões Originais

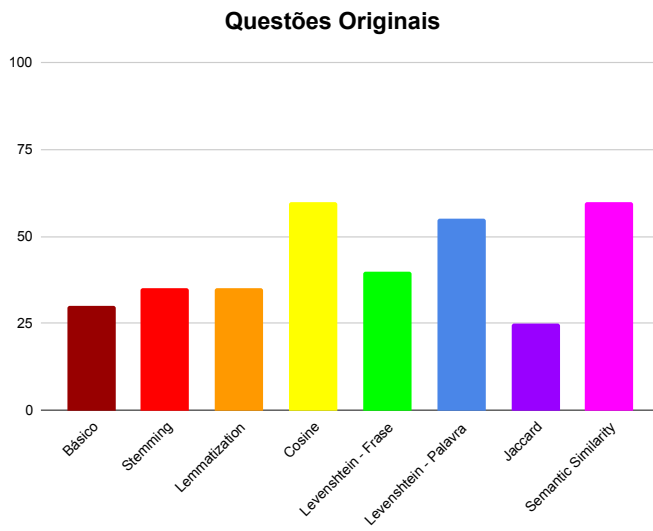


Figura J.12: Évora - "Resposta - Não remove as stopwords - Questões Originais"

## J.5 Parlamento

Nesta secção é possível verificar os resultados para o Domínio relacionado com o Parlamento Português.

### J.5.1 Parlamento - Questões Não Relacionadas



Figura J.13: Parlamento - "Não remove - Remove as stopwords - Questões Não Relacionadas"

### J.5.2 Parlamento - Questões Reformuladas

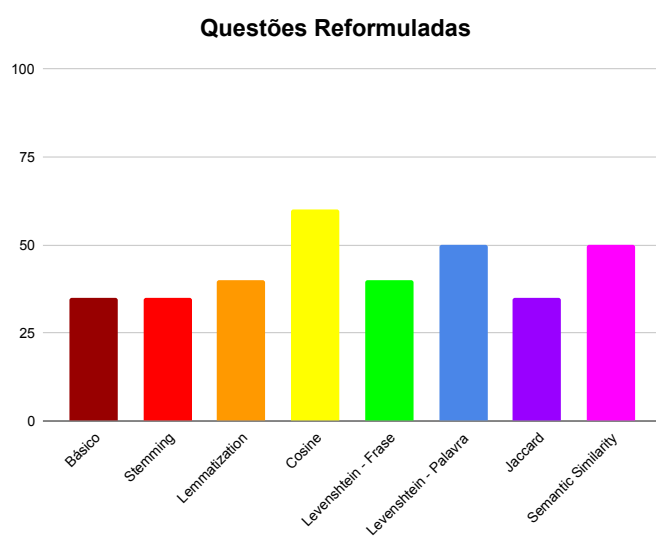


Figura J.14: Parlamento - "Não remove - Remove as stopwords - Questões Reformuladas"

### J.5.3 Parlamento- Questões Originais

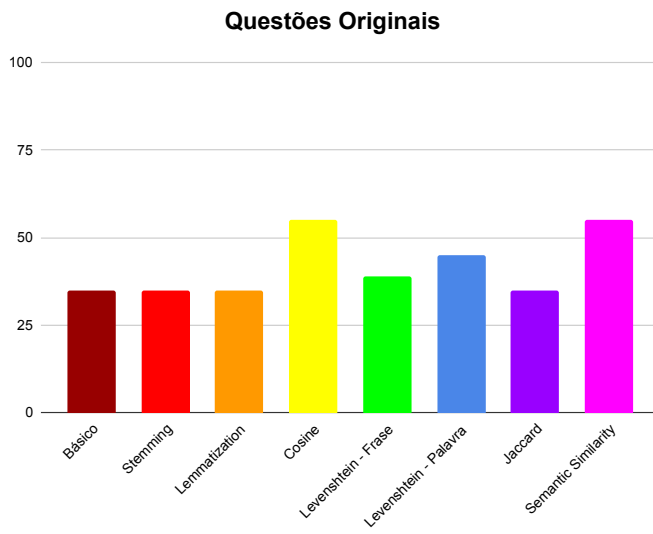


Figura J.15: Parlamento - "Não remove - Remove as stopwords - Questões Originais"

## J.6 Todas as Questões Reformuladas (Inglês)

Nesta secção é possível verificar os resultados para todas as questões reformuladas dos diferentes Domínios que estão em Inglês (Covid, OnePlus, UEFA).

### J.6.1 Todas as Questões Reformuladas (Inglês)

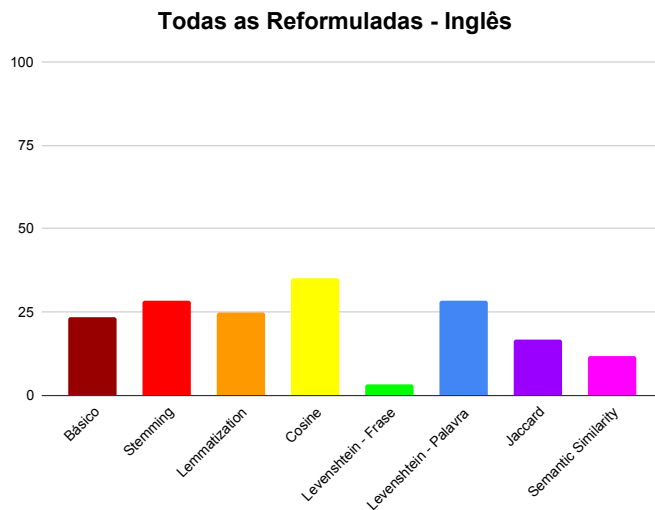


Figura J.16: Todas as Questões Reformuladas (Inglês) - "Resposta - Não remove as stopwords"



## J.7 Todas as Questões Reformuladas (Português)

Nesta secção é possível verificar os resultados para todas as questões reformuladas dos diferentes Domínios que estão em Português (Évora e Parlamento).

### J.7.1 Todas as Questões Reformuladas (Português)

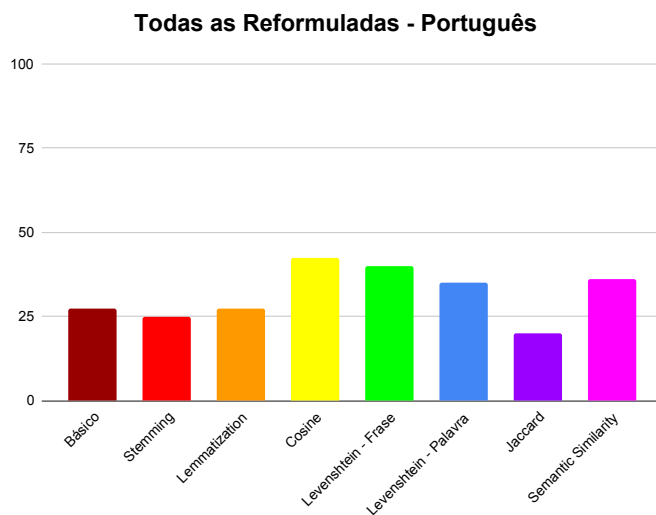


Figura J.17: Todas as Questões Reformuladas (Português) - "Não remove - Remove as stopwords"



# Bibliografia

- [AFHC16] Julie A Ask, Michael Facemire, Andrew Hogan, and HB Conversations. The state of chatbots. *Forrester. com report*, 20, 2016.
- [BVS<sup>+</sup>17] Jardeson Leandro Nascimento Barbosa, João Paulo Albuquerque Vieira, R Santos, Gilvan Veras Magalhães Junior, Mariana dos Santos MUNIZ, and Raimundo Santos Moura. Introdução ao processamento de linguagem natural usando python. *III Escola Regional de Informatica do Piauí*, 1:336–360, 2017.
- [CA20] Emmelyn AJ Croes and Marjolijn L Antheunis. Can we be friends with mitsuku? a longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, page 0265407520959463, 2020.
- [Col75] Kenneth Mark Colby. *Artificial Paranoia: A Computer Simulation of Paranoid Processes*. Elsevier Science Inc., USA, 1975.
- [coma] The SciPy community. numpy.argsort. <https://numpy.org/doc/stable/reference/generated/numpy.argsort.html>.
- [comb] The SciPy community. scipy.spatial.distance.euclidean. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.euclidean.html> scipy.spatial.distance.euclidean.
- [Cor] Microsoft Corporation. Como localizar as especificações e os requisitos de sistema para computadores windows 10. <https://www.microsoft.com/pt-pt/windows/windows-10-specifications>.
- [CS16] Carl Chapman and Kathryn T Stolee. Exploring regular expression usage and context in python. In *Proceedings of the 25th International Symposium on Software Testing and Analysis*, pages 282–293, 2016.
- [Dav16] Ernest Davis. Ai amusements: the tragic tale of tay the chatbot. *AI Matters*, 2(4):20–24, 2016.
- [dR] Assembleia da República. Perguntas frequentes. <https://www.parlamento.pt/Paginas/Perguntas-frequentes.aspx> um.

- [dVedV] O Instituto da Vinha e do Vinho. Faq's. <https://www.ivv.gov.pt/np4/faq/>.
- [d ] Universidade de Évora. Sac.online. <https://atendimento.sac.uevora.pt/>.
- [EMR18] Octavia Efraim, Vladislav Maraev, and João Rodrigues. Boosting a rule-based chatbot using statistics and user satisfaction ratings. In Andrey Filchenkov, Lidia Pivovarova, and Jan Žižka, editors, *Artificial Intelligence and Natural Language*, pages 27–41, Cham, 2018. Springer International Publishing.
- [FLB<sup>+</sup>13] David Ferrucci, Anthony Levas, Sugato Bagchi, David Gondek, and Erik T. Mueller. Watson: Beyond jeopardy! *Artificial Intelligence*, 199-200:93 – 105, 2013.
- [Foua] Python Software Foundation. 5.5. dictionaries. <https://docs.python.org/3/tutorial/datas-structures.html>.
- [Foub] Python Software Foundation. Python 3.7.6. <https://www.python.org/downloads/release/python-376/>.
- [Har06] Stevan Harnad. The annotation game: On turing (1950) on computing, machinery, and intelligence, July 2006.
- [HMvdW<sup>+</sup>20] Charles R. Harris, K. Jarrod Millman, St'efan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fern'andez del R'io, Mark Wiebe, Pearu Peterson, Pierre G'erard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [HQ] Sublime HQ. Sublime text 3.0. <https://www.sublimetext.com/blog/articles/sublime-text-3-point-0>.
- [IAV<sup>+</sup>19] Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. Privacy concerns in chatbot interactions. In *International Workshop on Chatbot Research and Design*, pages 34–48. Springer, 2019.
- [JLL14] Zongcheng Ji, Zhengdong Lu, and Hang Li. An information retrieval approach to short text conversation, 2014.
- [KD20] Jurgita Kapočiūtė-Dzikiene. A domain-specific generative chatbot trained from little data. *Applied Sciences*, 10(7):2221, 2020.
- [Kho20] Wei Hao Khoong. Deboost: A python library for weighted distance ensembling in machine learning. 2020.
- [KTP<sup>+</sup>19] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, 11:1178222619829083, 2019.
- [LB02] Edward Loper and Steven Bird. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*, 2002.
- [LRK<sup>+</sup>19] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. Talk to me: Exploring user interactions with the amazon alexa. *Journal of Librarianship and Information Science*, 51(4):984–997, 2019.

- [Mül18] Daniel Müllner. The fastcluster package: User's manual. 2018.
- [NSP19] A. Nigam, P. Sahare, and K. Pandya. Intent detection and slots prompt in a closed-domain chatbot. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 340–343, 2019.
- [oHHS] U.S. Department of Health & Human Services. Frequently asked questions. <https://www.cdc.gov/coronavirus/2019-ncov/faq.html>.
- [One] OnePlus. Shopping faqs. <https://www.oneplus.com/pt/support/shopping-help>.
- [Ori94] *AAAI '94: Proceedings of the Twelfth National Conference on Artificial Intelligence (Vol. 1)*, USA, 1994. American Association for Artificial Intelligence.
- [Pow98] David M. W. Powers. The total Turing test and the loebner prize. In *New Methods in Language Processing and Computational Natural Language Learning*, 1998.
- [Proa] NLTK Project. nltk.metrics package. <https://www.nltk.org/api/nltk.metrics.html>.
- [Prob] NLTK Project. Source code for nltk.stem.porter. [https://www.nltk.org/\\_modules/nltk/stem/porter.html](https://www.nltk.org/_modules/nltk/stem/porter.html) #PorterStemmer.
- [Proc] NLTK Project. Source code for nltk.stem.wordnet. [http://www.nltk.org/\\_modules/nltk/stem/wordnet.html](http://www.nltk.org/_modules/nltk/stem/wordnet.html) WordNetLemmatizer.
- [PSM14] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.
- [RDG<sup>+</sup>20] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot, 2020.
- [Reha] Radim Rehurek. Document similarity queries. <https://radimrehurek.com/gensim/similarities/docsim.html>.
- [Rehb] Radim Rehurek. Downloader api for gensim. <https://radimrehurek.com/gensim/downloader.html>.
- [Rehc] Radim Rehurek. Term similarity queries. <https://radimrehurek.com/gensim/similarities/termsim.html>.
- [Rehd] Radim Rehurek. Tf-idf model. <https://radimrehurek.com/gensim/models/tfidfmodel.html>.
- [RoI97] Rollo Carpenter. About the jabberwacky ai. <http://www.jabberwacky.com/j2about>, 1997. [Online; accessed 25-November-2020].
- [Roo19] Chat Room. Levenshtein distance. *algorithms*, 12(14):32, 2019.
- [Sax18] Micah D Saxton. A gentle introduction to topic modeling using python. *Theological Librarianship*, 11(1):18–27, 2018.
- [SD18] Bhargav Srinivasa-Desikan. *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018.

- [SGGAtAP14] Grigori Sidorov, Alexander Gelbukh, Helena G Añtextthreesuperiormez Adorno, and David Pinto. Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model. *Computaci A textthreesuperiorn y Sistemas*, 18:491 – 504, 09 2014.
- [SL20] Serhad Sarica and Jianxi Luo. Stopwords in technical language processing. *arXiv preprint arXiv:2006.02633*, 2020.
- [Sti19] Scott Stiefel. The chatbot will see you now. *Science and Technology Law Review*, 20(2), 2019.
- [Tha16] Shashi Thakur. Personalization for google now: User understanding and application to information recommendation and exploration. In *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, page 3, New York, NY, USA, 2016. Association for Computing Machinery.
- [TUR50] A. M. TURING. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236):433–460, 10 1950.
- [UEF] UEFA. Faq. <https://euro2020.hospitality.uefa.com/faq>.
- [VJ<sup>+</sup>16] S Vijayarani, R Janani, et al. Text mining: open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACIJ)*, 3(1):37–47, 2016.
- [Wal09] Richard Wallace. *The anatomy of A.L.I.C.E*, pages 181–210. 01 2009.
- [Wei66] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, January 1966.
- [Wik20a] Wikipedia contributors. Google allo — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Google\\_Allo&oldid=963596470](https://en.wikipedia.org/w/index.php?title=Google_Allo&oldid=963596470), 2020. [Online; accessed 11-July-2020].
- [Wik20b] Wikipedia contributors. Siri — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Siri&oldid=965027165>, 2020. [Online; accessed 10-July-2020].
- [Wik20c] Wikipedia contributors. Smarterchild — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=SmarterChild&oldid=959607750>, 2020. [Online; accessed 10-July-2020].
- [Wik20d] Wikipedia contributors. Zo (bot) — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Zo\\_\(bot\)&oldid=947107434](https://en.wikipedia.org/w/index.php?title=Zo_(bot)&oldid=947107434), 2020. [Online; accessed 12-July-2020].
- [WYF18] Chen Wei, Zhichen Yu, and Simon Fong. How to build a chatbot: Chatbot framework and its capabilities. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing, ICMLC 2018*, page 369–373, New York, NY, USA, 2018. Association for Computing Machinery.
- [ZGLS18] L. Zhou, Jianfeng Gao, Di Li, and H. Shum. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, Just Accepted:1–62, 2018.



UNIVERSIDADE DE ÉVORA  
INSTITUTO DE INVESTIGAÇÃO  
E FORMAÇÃO AVANÇADA

**Contactos:**

Universidade de Évora  
**Instituto de Investigação e Formação Avançada — IIFA**  
Palácio do Vimioso | Largo Marquês de Marialva, Apart. 94  
7002 - 554 Évora | Portugal  
Tel: (+351) 266 706 581  
Fax: (+351) 266 744 677  
email: [iifa@uevora.pt](mailto:iifa@uevora.pt)