

# Extractive Q&A

## Performance Comparison between Learning Methods Context Learning and Fine-Tuning

Leonardo Augusto da Silva Pacheco  
Marcus Vinícius Borela de Castro

Extractive Answering (The answer to a question given a piece of context is a direct substring of the context (fragmento do texto) - Analogy; A highlighter)

Sem considerar o retriever, que seria um "limitador" de performance

Escopo

Abstractive Answering-The answer to a question given a piece of context is a free-form phrase based on the context-Analogy - A pen

Não escopo

Descrição

Exemplo

[Inglês SQuAD 1.0](#)

[Português \(SQuAD v1\\_pt\)](#)

[Stanford Question Answering Dataset](#) (base de teste)

Dataset

A binary metric that gives  $EM = 1$  if the characters in the predicted and ground truth answers match exactly, and  $EM = 0$  otherwise. If no answer is expected, the model gets  $EM = 0$  if it predicts any text at all.

Exact Match (EM)

Measures the harmonic mean of the precision and recall. (considerar palavras coincidentes e diferentes)

F1

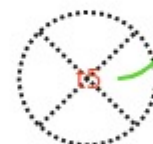
Definição

Métricas

Model (261 mb) is a fine-tune checkpoint of DistilBERT-base-cased, fine-tuned using (a second step of) knowledge distillation on SQuAD v1.1 (F1=87.1 on the dev set) (for comparison, BERT bert-base-cased reaches F1= 88.7).

distilbert-base-cased-distilled-squad

Inglês



Não encontrado em inglês; em português, desempenho inferior

Fine Tuning

Modelo Q&A (1.33 gb)  
com refinamento no  
squad\_v1\_pt. (F1:84.4 ;  
EM=72.68)

[pierreguillou/bert-large-cased-squad-v1.1-portuguese](https://huggingface.co/pierreguillou/bert-large-cased-squad-v1.1-portuguese)

Português (baseados no BERTimbau)

Modelos

Modelo Q&A (433mb) com  
refinamento no squad\_v1\_pt.  
(F1:82.82.5; EM=70.49)

~~[pierreguillou/bert-base-cased-squad-v1.1-portuguese](https://huggingface.co/pierreguillou/bert-base-cased-squad-v1.1-portuguese)~~

desempenho inferior

[GPT-neo 1.3B](https://huggingface.co/openai/gpt-neo-1.3B)

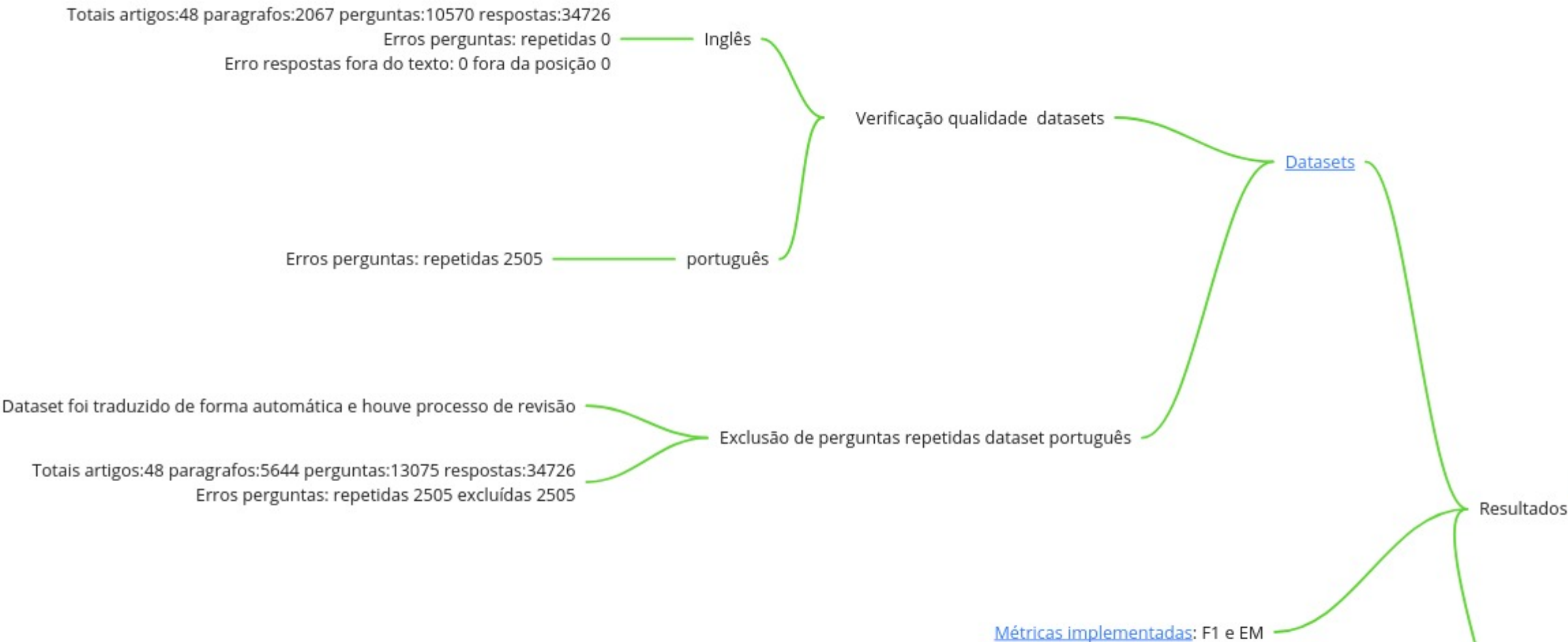
[GPT-J 6B](https://huggingface.co/openai/gpt-j-6B)

[GPT-NEOx-20b](https://huggingface.co/openai/gpt-neox-20b)

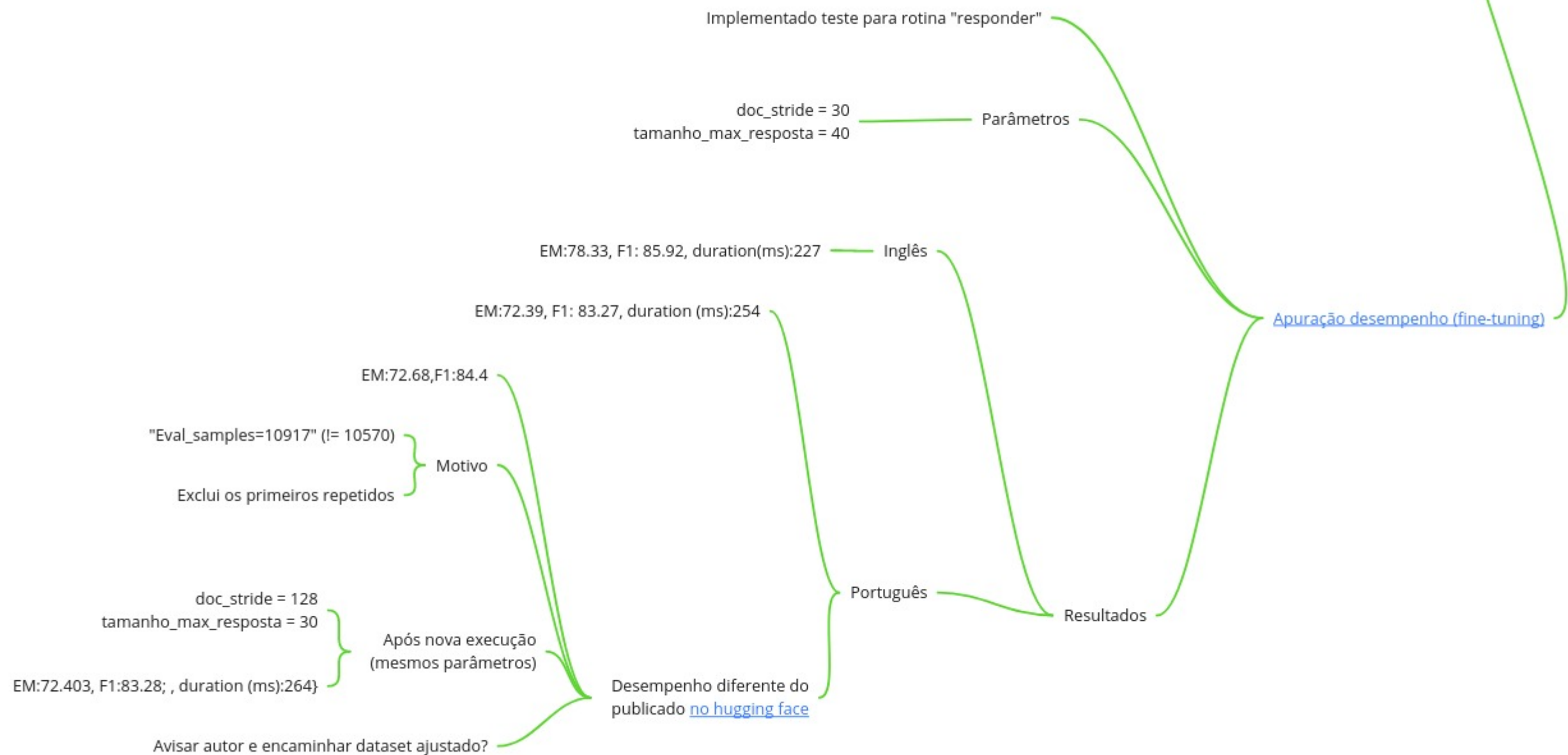
Um desses: <https://huggingface.co/EleutherAI>

Context Learning

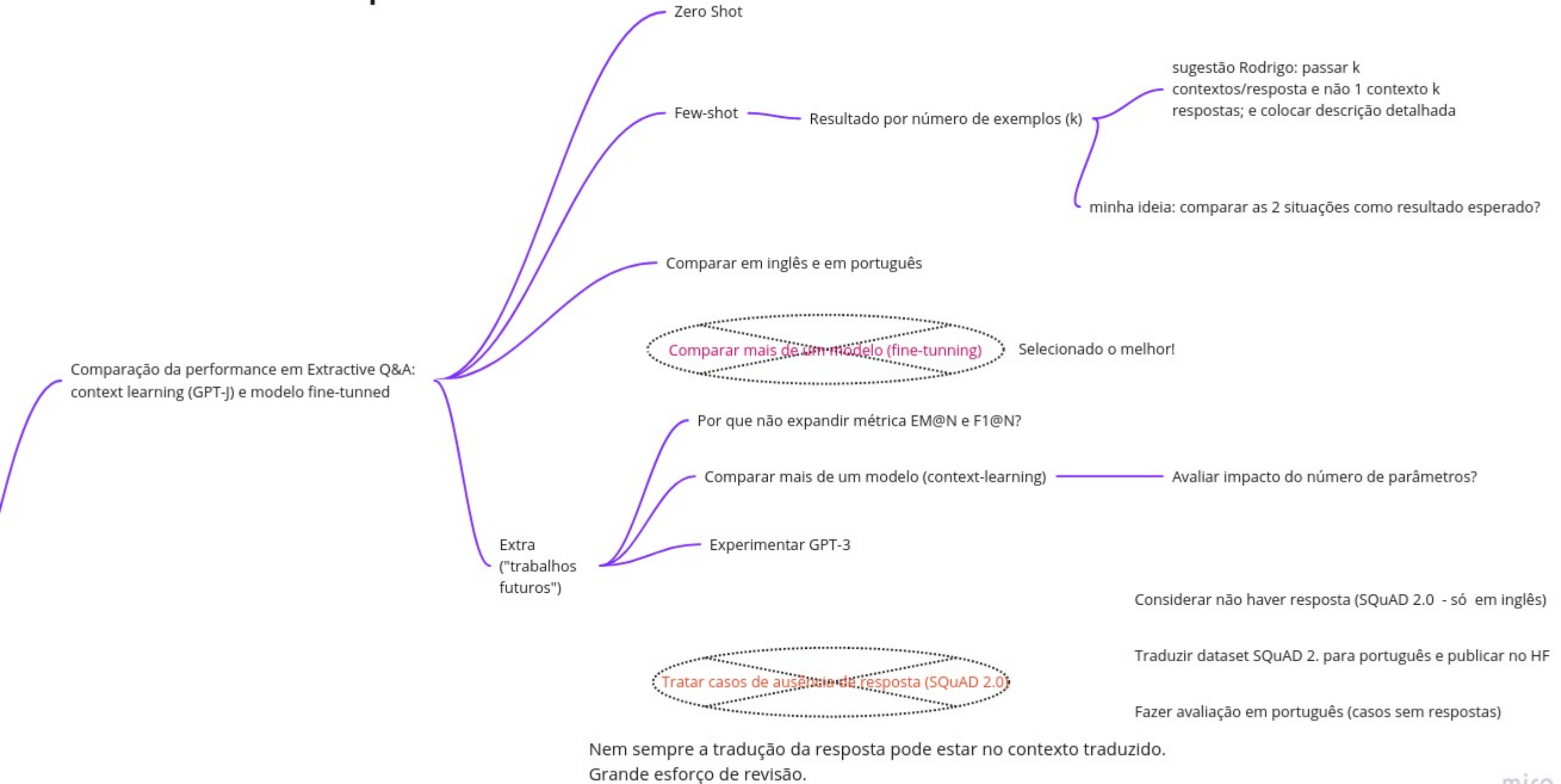
# Resultados preliminares



# Resultados preliminares

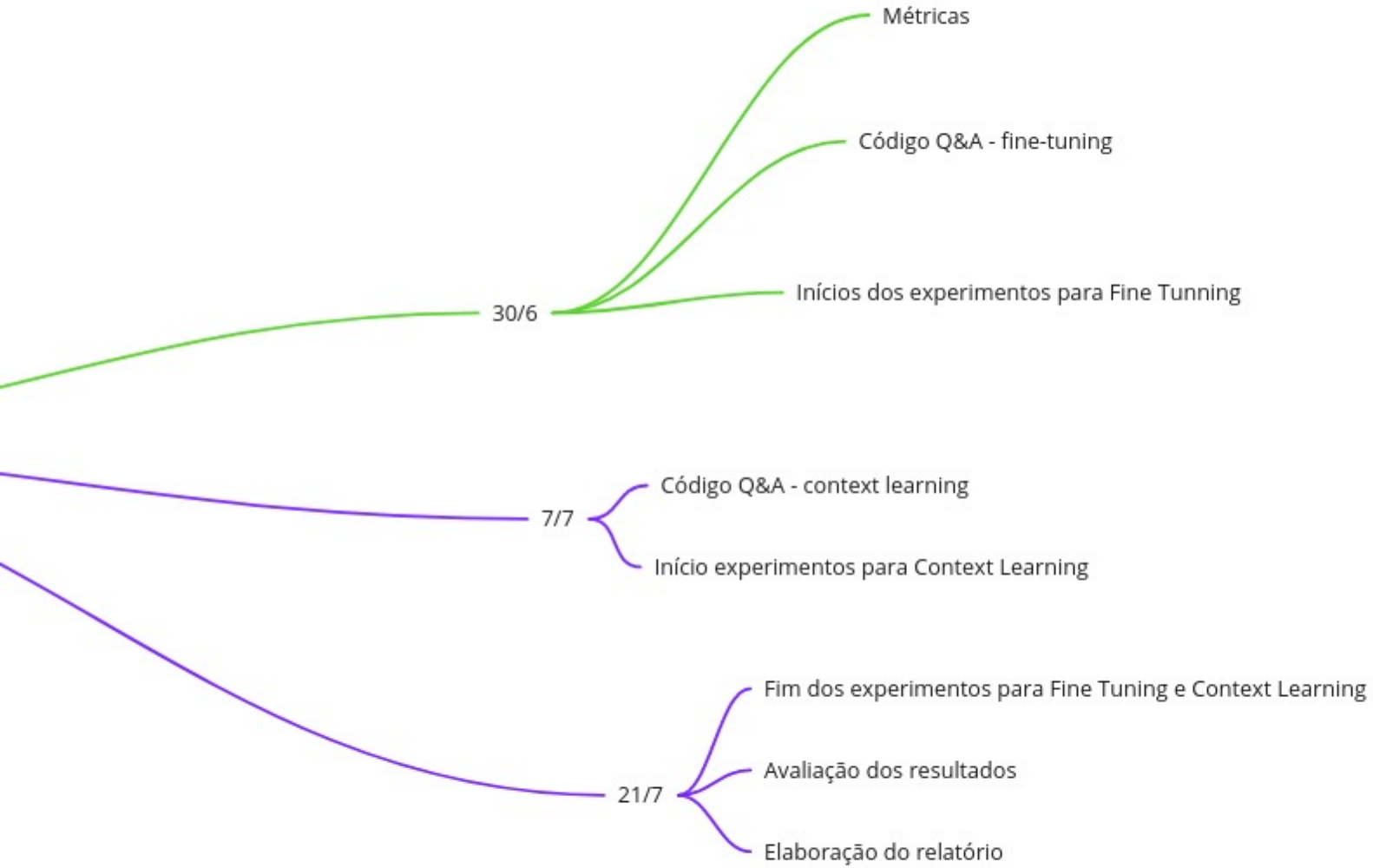


# Resultados esperados





# Cronograma



# Referências

## Fine Tunning

[Natural Language Processing with Transformers - Chapter 7 - Question Answering](#)

[Hands on Transfer Learning with BERT](#)

[Cloudera Fast Forward: CFF builds a state-of-the-art QA application with the latest NLP techniques](#)

## Context Learning

[Exemplo de execução no collab do GPT-J](#)

[Site OpenAI - sobre uso da API do gpt3](#)

[Livro Machine Translation with the Transformer - The Rise of Suprahuman Transformers with GPT-3 Engines](#)

[Livro: Exploring GPT-3](#)

cs224u\_ Few-shot OpenQA with ColBERT retrieval

[Video](#)

[Caderno](#)



# FIM

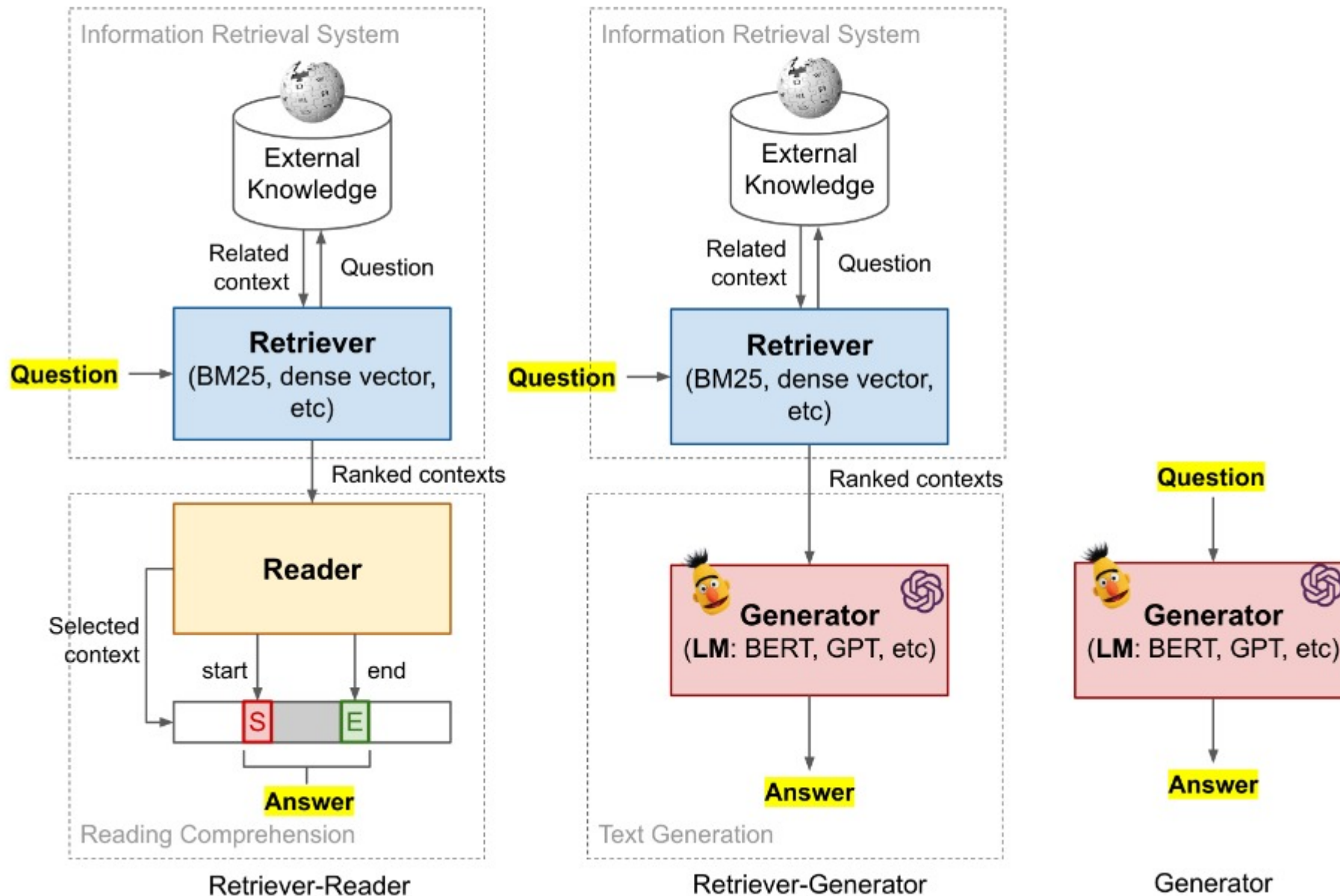
## Github do Projeto

The screenshot shows the GitHub repository page for `marcusboela/exqa-complearning`. The repository is public and has 12 commits. The file list includes:

File	Description	Time
<code>borela com classe dataset; exclusao duplicata ok</code>		23 minutes ago
<code>data/dataset/squad</code>	mod ingles; dataset pt; ajustes...	2 days ago
<code>docs</code>	com classe dataset; exclusao duplicata ok	23 minutes ago
<code>models/fine_tuning/distilbert-base-c...</code>	mod ingles; dataset pt; ajustes...	2 days ago
<code>source</code>	com classe dataset; exclusao duplicata ok	23 minutes ago
<code>.gitignore</code>	mod ingles; dataset pt; ajustes...	2 days ago
<code>README.md</code>	First presentation and paste structure	5 days ago
<code>estudo.py</code>	mod ingles; dataset pt; ajustes...	2 days ago
<code>requirements.txt</code>	calculating fine-tuning; start code organization	9 hours ago

The repository description is: **exqa-complearning**  
Extractive Q&A - Performance Comparison between Learning Methods: Context Learning and Fine-Tuning

# Não é contexto: abstractive Q&A



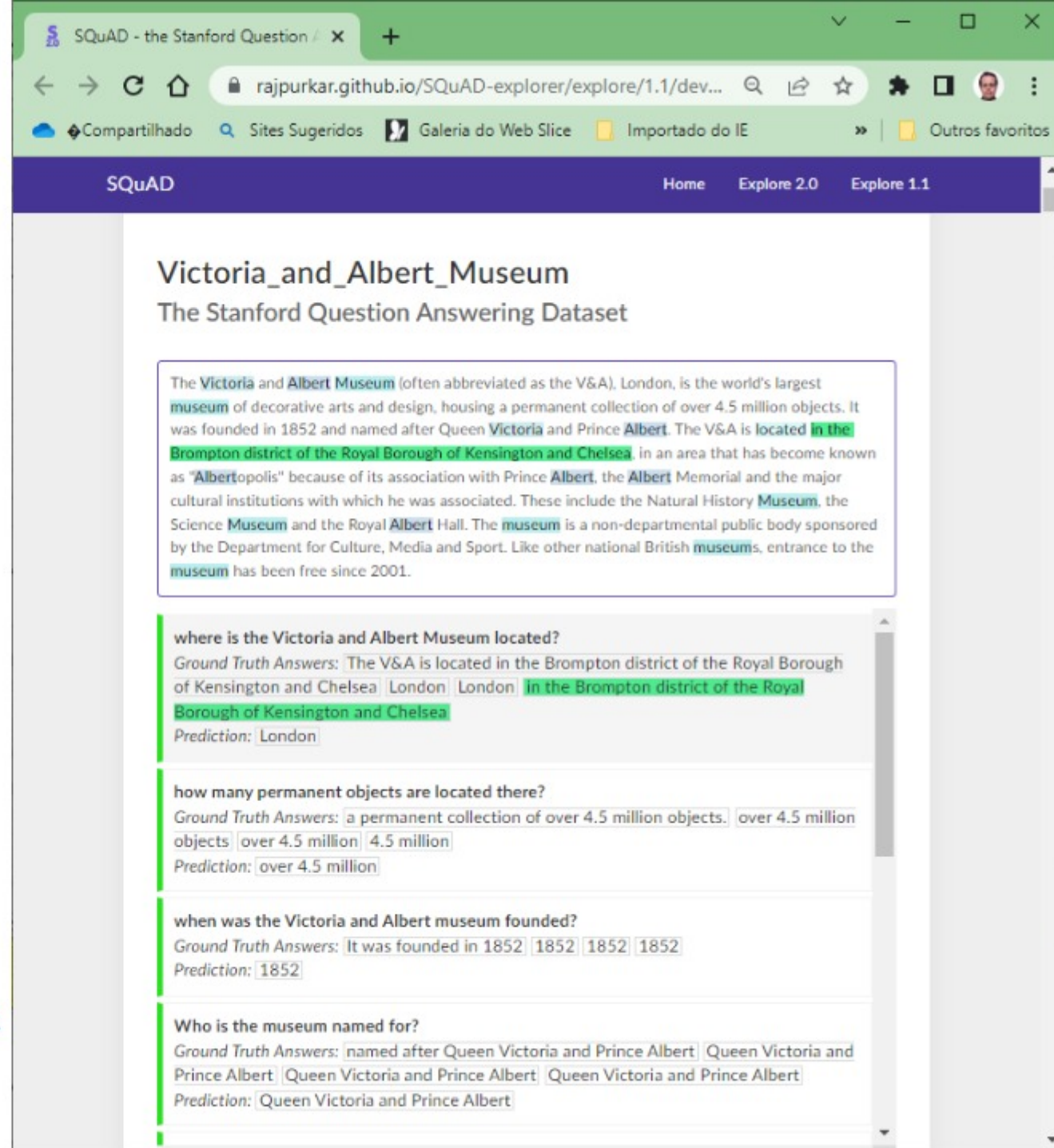
Extractive Q&A - Performance Comparison between  
Learning Methods:  
Context Learning and Fine-Tuning

Modelos

Cronograma

# Descrição

## O que é Extractive Q&A?



The screenshot displays the SQuAD (Stanford Question Answering Dataset) web interface. The browser address bar shows the URL: [rajpurkar.github.io/SQuAD-explorer/explore/1.1/dev...](http://rajpurkar.github.io/SQuAD-explorer/explore/1.1/dev...). The page title is "SQuAD" and the navigation bar includes links for "Home", "Explore 2.0", and "Explore 1.1".

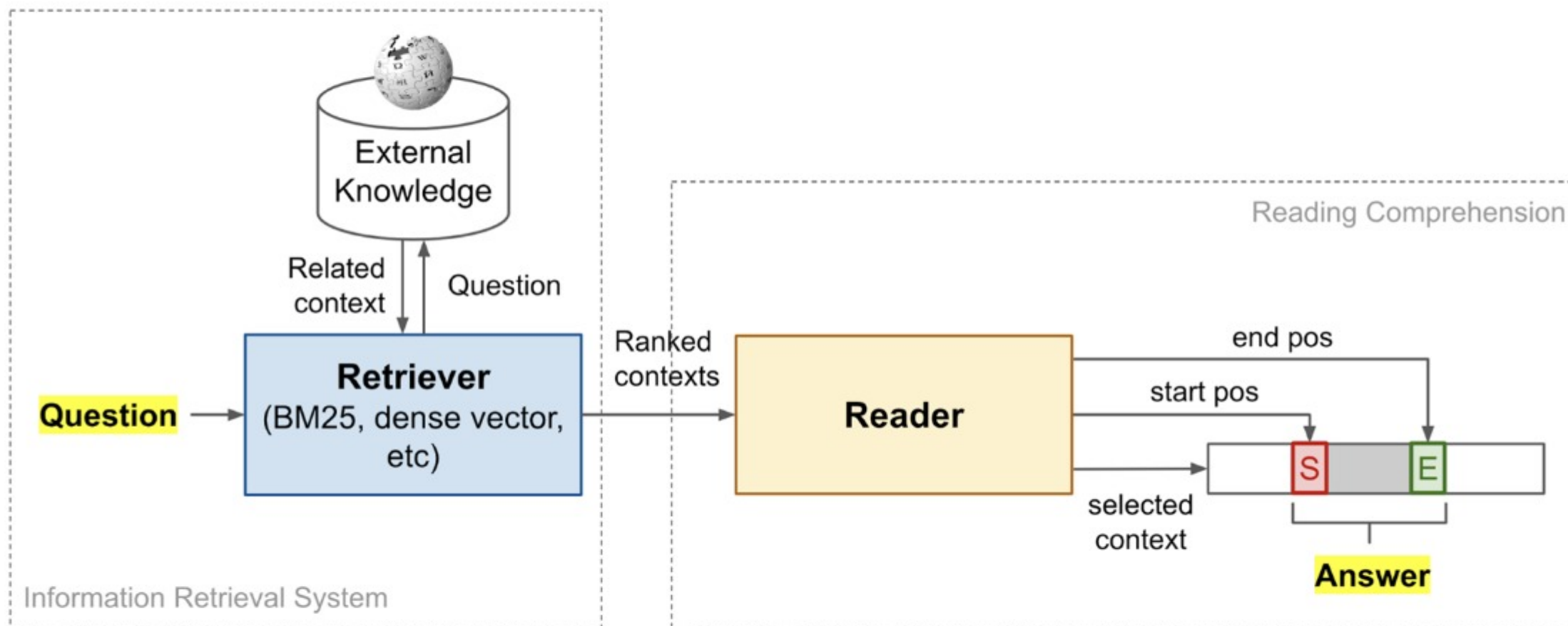
The main content area features a passage titled "Victoria\_and\_Albert\_Museum" from "The Stanford Question Answering Dataset". The passage text is: "The **Victoria and Albert Museum** (often abbreviated as the V&A), London, is the world's largest **museum** of decorative arts and design, housing a permanent collection of over 4.5 million objects. It was founded in 1852 and named after Queen **Victoria** and Prince **Albert**. The V&A is **located in the Brompton district of the Royal Borough of Kensington and Chelsea**, in an area that has become known as "**Albertopolis**" because of its association with Prince **Albert**, the **Albert** Memorial and the major cultural institutions with which he was associated. These include the Natural History **Museum**, the Science **Museum** and the Royal **Albert** Hall. The **museum** is a non-departmental public body sponsored by the Department for Culture, Media and Sport. Like other national British **museums**, entrance to the **museum** has been free since 2001."

Below the passage, four questions are listed, each with its ground truth answers and a predicted answer:

- Question:** where is the Victoria and Albert Museum located?  
**Ground Truth Answers:** The V&A is located in the Brompton district of the Royal Borough of Kensington and Chelsea | London | London | in the Brompton district of the Royal Borough of Kensington and Chelsea  
**Prediction:** London
- Question:** how many permanent objects are located there?  
**Ground Truth Answers:** a permanent collection of over 4.5 million objects. | over 4.5 million objects | over 4.5 million | 4.5 million  
**Prediction:** over 4.5 million
- Question:** when was the Victoria and Albert museum founded?  
**Ground Truth Answers:** It was founded in 1852 | 1852 | 1852 | 1852  
**Prediction:** 1852
- Question:** Who is the museum named for?  
**Ground Truth Answers:** named after Queen Victoria and Prince Albert | Queen Victoria and Prince Albert | Queen Victoria and Prince Albert | Queen Victoria and Prince Albert  
**Prediction:** Queen Victoria and Prince Albert

Exemplo: [Stanford Question Answering Dataset SQuAD](http://rajpurkar.github.io/SQuAD-explorer/explore/1.1/dev...)

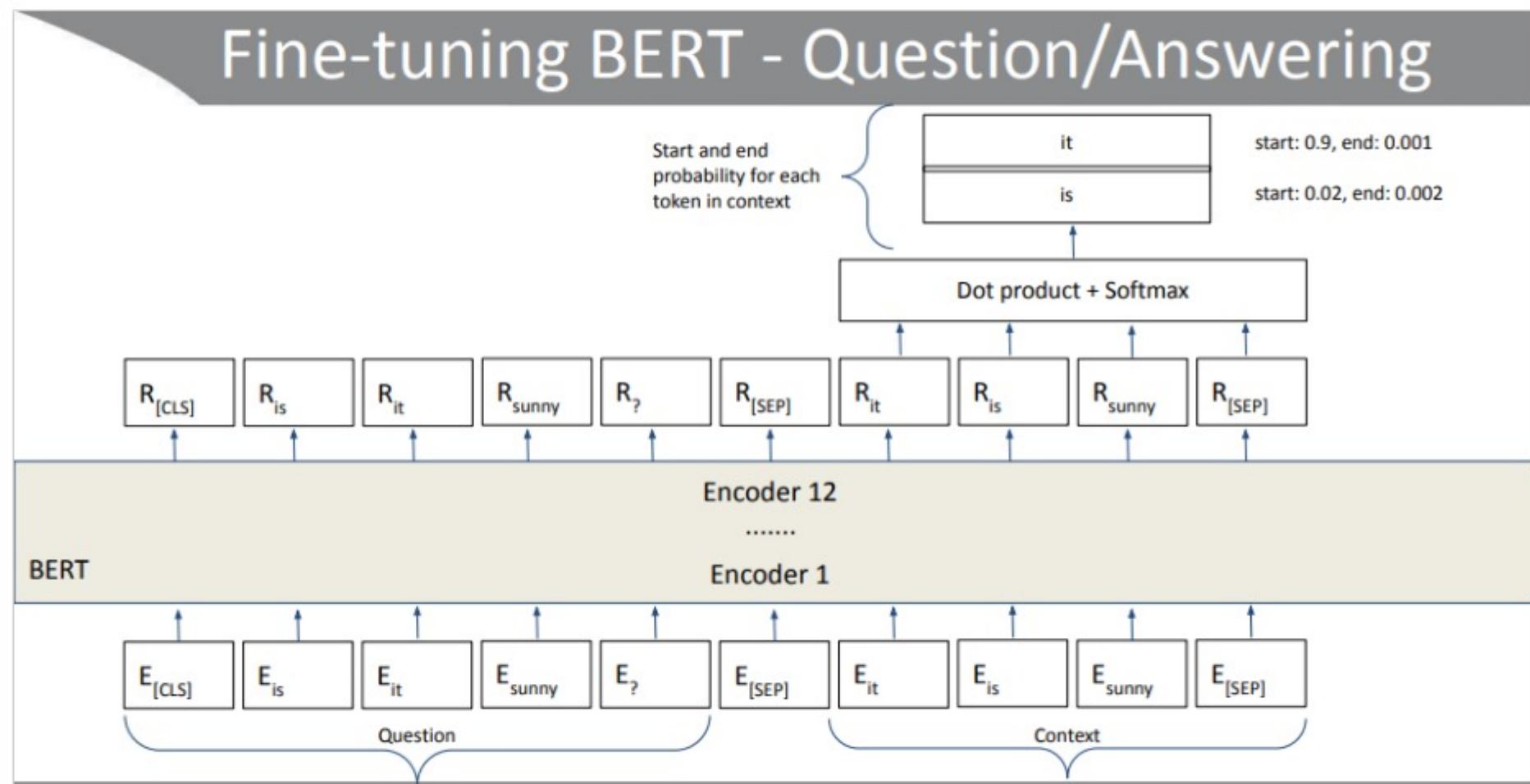
# IR Extractive Q&A (over Information Retrieval)



Fonte: [How to Build an Open-Domain Question Answering System?](#)



# Extractive Q&A in Fine Tuning





# Q&A in Context Learning

Contexto

Few Shot  
 $k=3$

Exemplo: [GPT-3](#) x SQuAD

The screenshot shows the OpenAI Playground interface. The browser tab is "Playground - OpenAI API". The URL is "beta.openai.com/playground/p/default-qa?mode=e...". The page has a navigation bar with "Overview", "Documentation", "Examples", and "Playground". The "Playground" section is active, showing a "Q&A" mode. The "Input" section contains a context paragraph about the Victoria and Albert Museum, followed by three question-answer pairs (few-shot examples). The "Instructions" section at the bottom says "Answer the question using context fragment". The "Model" dropdown is set to "text-davinci-edit-001". The "Temperature" slider is at 0, and the "Top P" slider is at 1. A "Submit" button is at the bottom left. A notification at the bottom right says "Editing is free while in beta. We'd love your feedback."

Playground - OpenAI API

beta.openai.com/playground/p/default-qa?mode=e...

Compartilhado Sites Sugeridos Galeria do Web Slice Importado do IE

Overview Documentation Examples Playground Upgrade Help Personal

Playground Q&A Save View code Share

Input Use as input Mode

The Victoria and Albert Museum (often abbreviated as the V&A), London, is the world's largest museum of decorative arts and design, housing a permanent collection of over 4.5 million objects. It was founded in 1852 and named after Queen Victoria and Prince Albert. The V&A is located in the Brompton district of the Royal Borough of Kensington and Chelsea, in an area that has become known as "Albertopolis" because of its association with Prince Albert, the Albert Memorial and the major cultural institutions with which he was associated.

Q: How many permanent objects are located there?  
A: over 4.5 million

Q: When was the Victoria and Albert museum founded?  
A: 1852

Q: Which party did he belong to?  
A: He belonged to the Republican Party.

Q: where is the Victoria and Albert Museum located?  
A:

The Victoria and Albert Museum (often abbreviated as the V&A), London, is the world's largest museum of decorative arts and design, housing a permanent collection of over 4.5 million objects. It was founded in 1852 and named after Queen Victoria and Prince Albert. The V&A is located in the Brompton district of the Royal Borough of Kensington and Chelsea, in an area that has become known as "Albertopolis" because of its association with Prince Albert, the Albert Memorial and the major cultural institutions with which he was associated.

Q: How many permanent objects are located there?  
A: over 4.5 million

Q: When was the Victoria and Albert museum founded?  
A: 1852

Q: Which party did he belong to?  
A: He belonged to the Republican Party.

Q: where is the Victoria and Albert Museum located?  
A: The V&A is located in the Brompton district of the Royal Borough of Kensington and Chelsea, in an area that has become known as "Albertopolis" because of its association with Prince Albert, the Albert Memorial and the major cultural institutions with which he was associated.

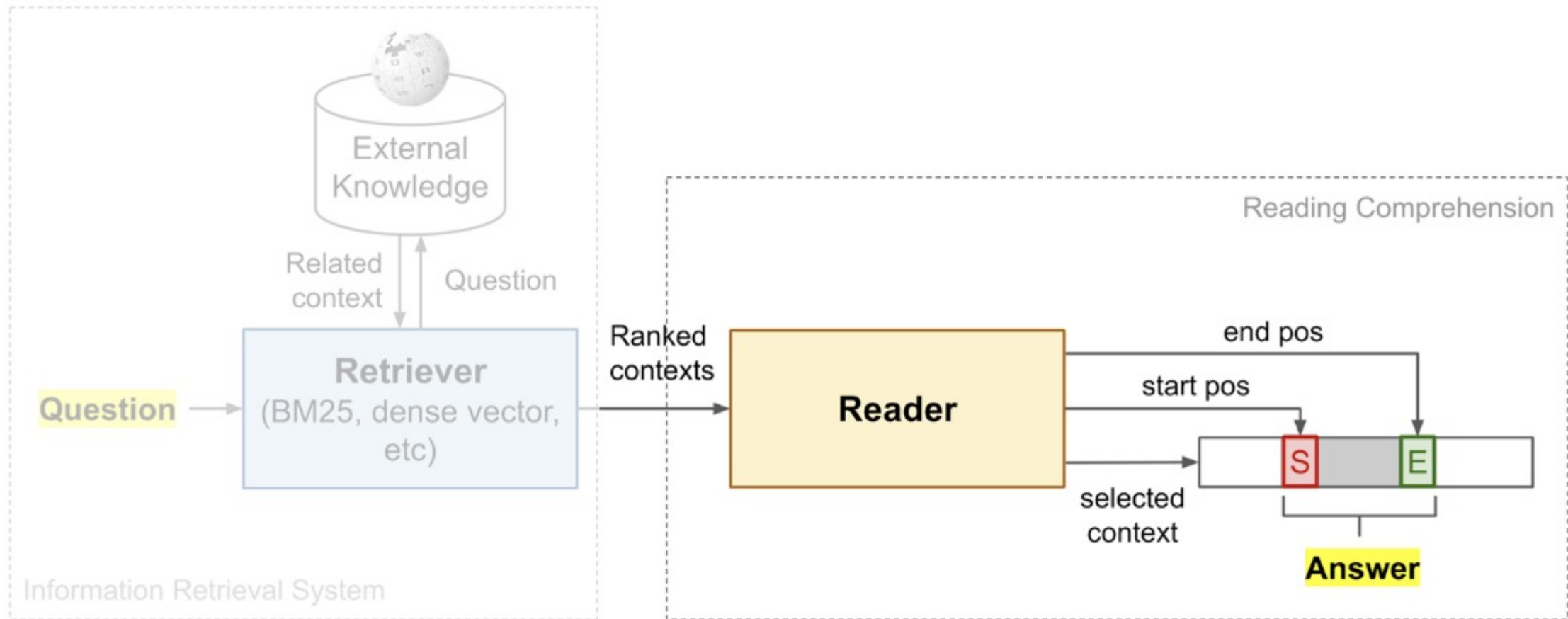
Instructions

Answer the question using context fragment

Submit

Editing is free while in beta. We'd love your feedback.

# Igualdade de condições: sem retrieval (sem limitador de performance)



contexto: Fontes de oxigênio altamente concentradas promovem combustão rápida. Riscos de incêndio e explosão existem quando oxidantes e combustíveis concentrados são trazidos para perto; um evento de ignição, como calor ou faísca, é necessário para acionar a combustão. O oxigênio é o oxidante, não o combustível, mas, no entanto, a fonte da maior parte da energia química liberada na combustão. Os riscos de combustão também se aplicam a compostos de oxigênio com alto potencial oxidativo, como peróxidos, cloratos, nitratos, percloratos e dicromatos, porque eles podem doar oxigênio ao fogo.

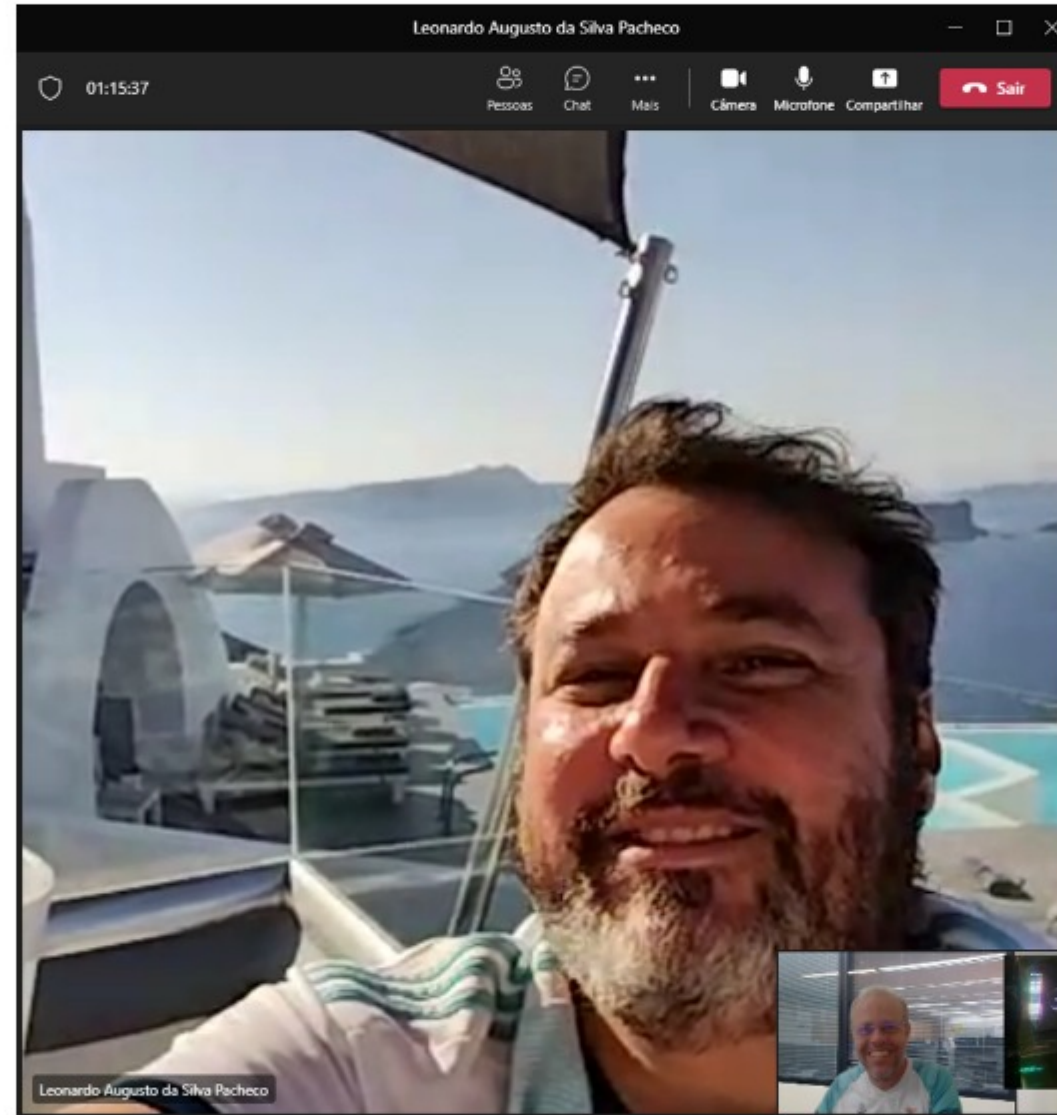
1a ocorrência: {'answers': [{'answer\_start': 10, 'text': 'oxigênio', 'answer\_start\_original': 31, 'text\_original': 'oxygen'}, {'answer\_start': 177, 'text': 'um evento de ignição', 'answer\_start\_original': 168, 'text\_original': 'an ignition event'}, {'answer\_start': 177, 'text': 'um evento de ignição, como calor ou faísca', 'answer\_start\_original': 168, 'text\_original': 'an ignition event, such as heat or a spark'}, {'answer\_start': 177, 'text': 'um evento de ignição', 'answer\_start\_original': 168, 'text\_original': 'an ignition event'}], 'question': 'O que é necessário para fazer a combustão acontecer?', 'id': '571a50df4faf5e1900b8a960', 'question\_original': 'What is needed to make combustion happen?'}

2a ocorrência: {'answers': [{'answer\_start': 204, 'answer\_start\_original': 195, 'text': 'calor ou uma faísca', 'text\_original': 'heat or a spark'}], 'id': '571a50df4faf5e1900b8a960', 'question': 'O que é necessário para fazer a combustão acontecer?', 'question\_original': 'What is needed to make combustion happen?'}

em inglês:

{'answers': [{'answer\_start': 195, 'text': 'heat or a spark'}, {'answer\_start': 31, 'text': 'oxygen'}, {'answer\_start': 168, 'text': 'an ignition event'}, {'answer\_start': 168, 'text': 'an ignition event, such as heat or a spark'}, {'answer\_start': 168, 'text': 'an ignition event'}], 'question': 'What is needed to make combustion happen?', 'id': '571a50df4faf5e1900b8a960'}

# Leonardo&Guaíra: 2.0



(Arquipélago de Santorini – Grécia)