

Extractive Q&A

Performance Comparison between Learning Methods Context and Transfer Learning

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

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

Exemplo

[Inglês SQuAD 1.1](#)

[Português \(SQuAD_v1_1_pt\)](#)

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

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.

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

EM@3

F1@3

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

Transfer Learning

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](#)

Português (baseados no BERTimbau)

Modelos

[GPT-neo_1.3B](#)

[GPT-J_6B](#)

[GPT-NEOx-20b](#)

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

Context Learning

Resultados preliminares - verificação e tratamento dos dados

Totais artigos:48 paragrafos:2067 perguntas:10570 respostas:34726

Erros perguntas: repetidas 0

Erro respostas fora do texto: 0 fora da posição 0

Inglês

Erros perguntas: repetidas 2505

português

Avaliação da qualidade dos datasets

Datasets

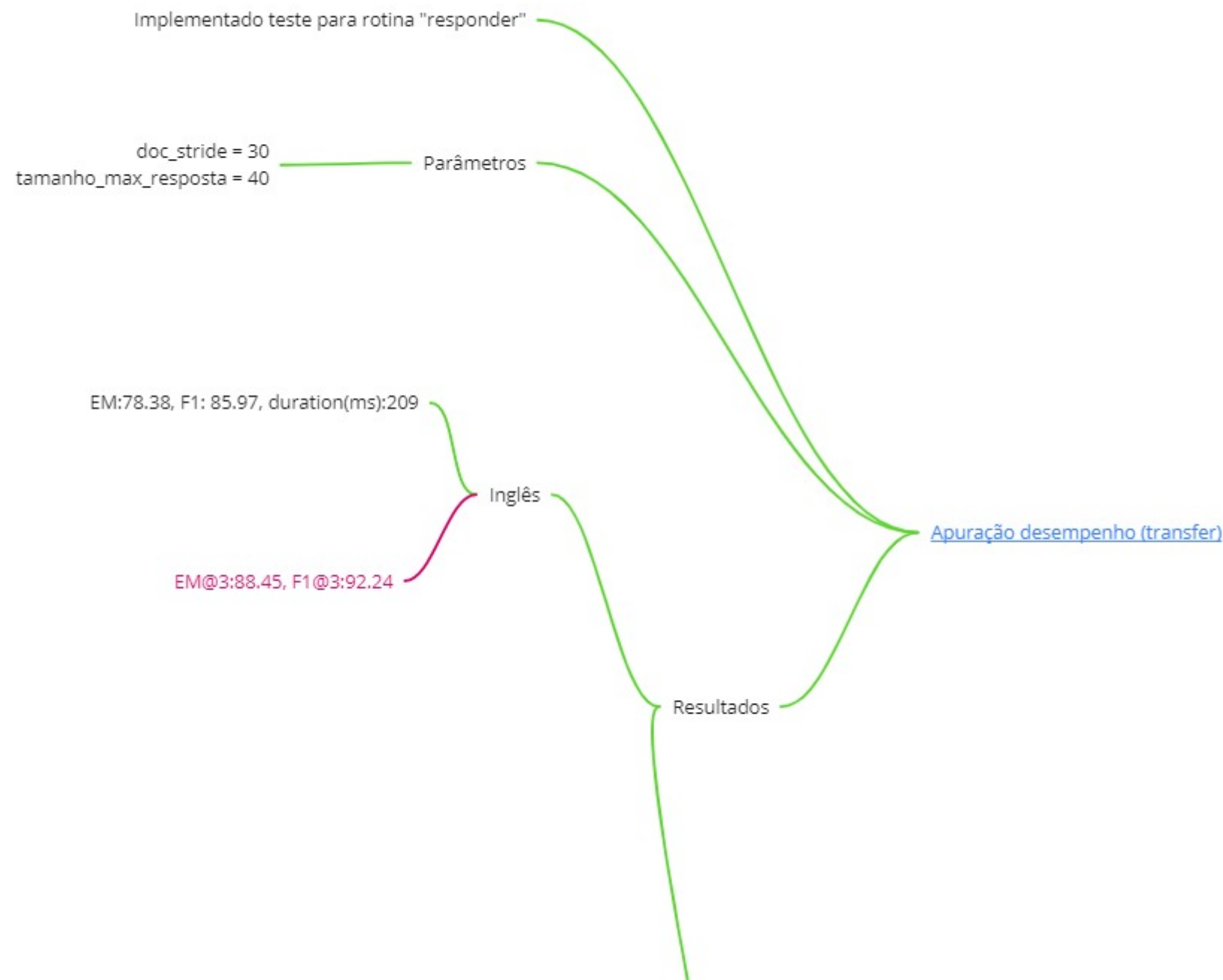
Dataset foi traduzido de forma automática e houve processo de revisão

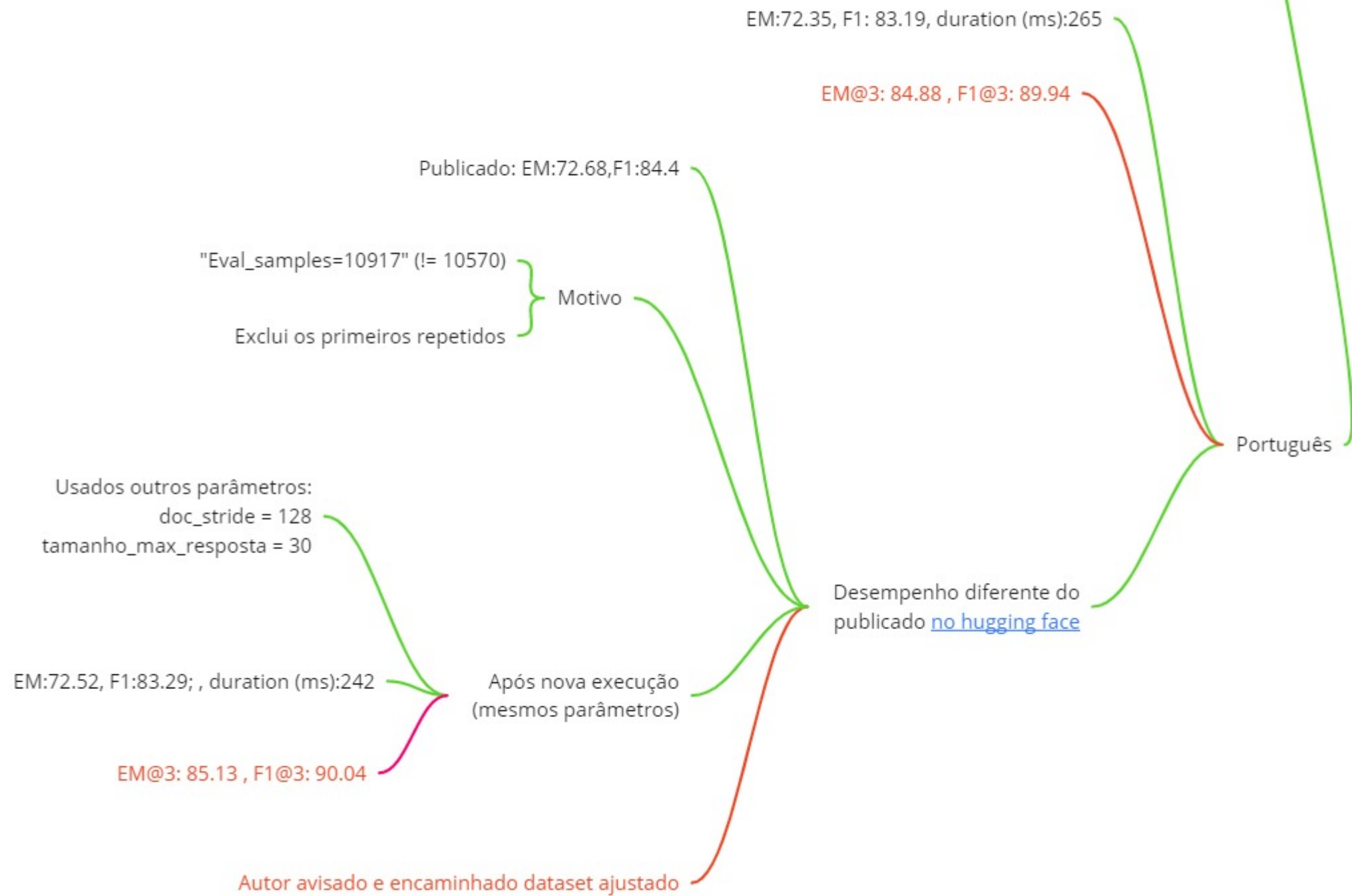
Totais artigos:48 paragrafos:5644 perguntas:13075 respostas:34726

Erros perguntas: repetidas 2505 excluídas 2505

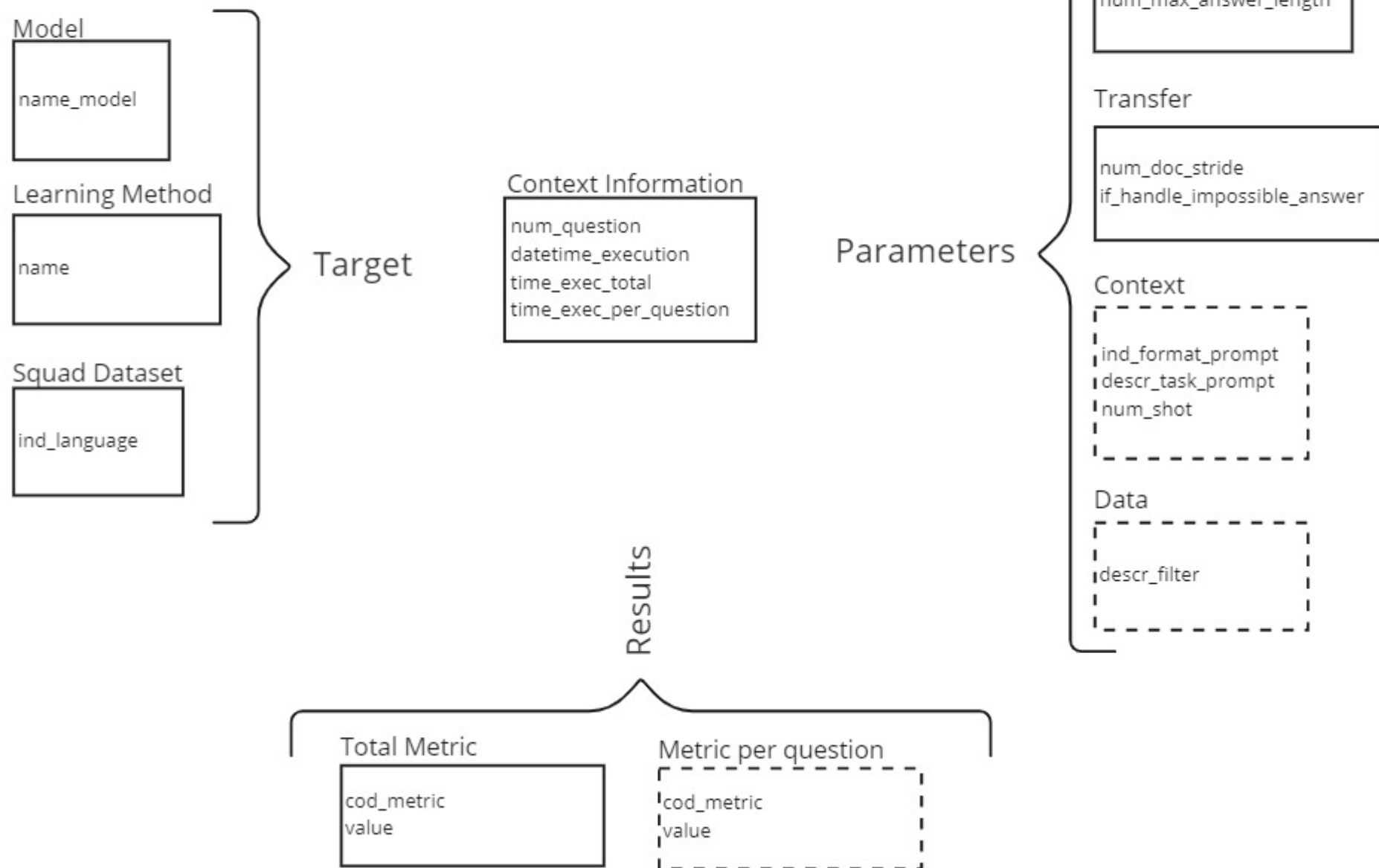
Exclusão de perguntas repetidas dataset português

Resultados preliminares - Transfer Learning





Evaluations' rastró*



* Pandas Dataframe

GPT-3

para definir formato prompt e parâmetros

The screenshot shows the OpenAI Playground interface. On the left, there's a 'Get started' sidebar with instructions and a 'KEEP IN MIND' section. The main area is titled 'Playground' and contains a prompt and its completion. A probability distribution overlay is shown for the completion 'Flamengo'. On the right, there's a settings panel with various parameters like Model, Temperature, and Top P.

Get started

Enter an instruction or select a preset, and watch the API respond with a completion that attempts to match the context or pattern you provided.

You can control which model completes your request by changing the model.

KEEP IN MIND

- Use good judgment when sharing outputs, and attribute them to your name or company. [Learn more.](#)
- Requests submitted to our models may be used to train and improve future models. [Learn more.](#)
- Our default models' training data cuts off in 2021, so they may not have knowledge of current events.

Playground

Load a preset...

Save View code Share

Model: text-davinci-002

Temperature: 0

Maximum length: 256

Stop sequences: Pergunta, \n\n, Contexto

Top P: 1

Frequency penalty: 0

Presence penalty: 0

Best of: 3

Inject start text: [checked]

Inject restart text: [checked]

Show probabilities: Full spectrum

Instrução: Dado o contexto, dê a resposta à pergunta:

Contexto: Por causa da chuva, o Flamengo, melhor time do Brasil, ficou sem jogar a final da Libertadores. Há muitos torcedores pelo Flamengo, o melhor time, na minha Família. Além da bicicleta, em 1990, eu tinha a honra de ser flamenguista. Ontem, quando reli os documentos do tribunal, descobri que em 1990 quando tomei posse, minha declaração de bens só continha uma bicicleta.

Pergunta: De onde os documentos foram relidos?

Resposta: Do tribunal

Contexto: Por causa da chuva, o Flamengo, melhor time do Brasil, ficou sem jogar a final da Libertadores. Há muitos torcedores pelo Flamengo, o melhor time, na minha Família. Além da bicicleta, em 1990, eu tinha a honra de ser flamenguista. Ontem, quando reli os documentos do tribunal, descobri que em 1990 quando tomei posse, minha declaração de bens só continha uma bicicleta.

Pergunta: Quando tomei posse?

Resposta: 1990

Contexto: Por causa da chuva, o Flamengo, melhor time do Brasil, ficou sem jogar a final da Libertadores. Há muitos torcedores pelo Flamengo, o melhor time, na minha Família. Além da bicicleta, em 1990, eu tinha a honra de ser flamenguista. Ontem, quando reli os documentos do tribunal, descobri que em 1990 quando tomei posse, minha declaração de bens só continha uma bicicleta.

Pergunta: Qual o melhor time do Brasil?

Resposta: Flamengo

Flam = 72.87%

Q = 22.93%

\n = 2.27%

a = 0.69%

Q = 0.27%

Total: -0.32 logprob on 1 tokens (99.04% probability covered in top 5 logits)

Submit

500

Primeiras experimentações (Context)

miro

GPT-NEOx-20b

erro memória RAM



Teste básico GPT-NEOx ☆

Ficheiro Editar Ver Inserir Tempo de execução Ferramentas Ajuda [Todas as alterações foram guardadas.](#)

Comentário Partilhar Editar

+ Código + Texto

RAM Disco Editor

```
from transformers import GPTNeoXForCausalLM, GPTNeoXTokenizerFast

model = GPTNeoXForCausalLM.from_pretrained("EleutherAI/gpt-neox-20b")

tokenizer = GPTNeoXTokenizerFast.from_pretrained("EleutherAI/gpt-neox-20b")

device = torch.device("cuda")
model.to(device)
```

Downloading: 0% | 0.00/868M [00:00<?, ?B/s]
Downloading: 100% 576M/576M [00:08<00:00, 67.6MB/s]
Downloading: 100% 591M/591M [02:32<00:00, 2.00MB/s]

```
texto_pergunta= "Qual o melhor time do Brasil?"
texto_contexto= "Por causa da chuva, o Flamengo, melhor \
time do Brasil, ficou sem jogar a final da Libertadores. Há muitos torcedores pelo \
Flamengo, o melhor time, na minha Família. Além da bicicleta, em 1990, eu tinha a honra \
de ser flamenguista. Ontem, quando reli os documentos do tribunal, \n\
descobri que em 1990 quando tomei posse, minha declaração de bens só continha uma bicicleta."

prompt = ('Responda à pergunta extraíndo a resposta do contexto:\nContexto: ' + texto_contexto + '\nPergunta: ' + texto_pergunta + '\nResposta:')

tokenized = tokenizer(prompt, return_tensors="pt")
input_ids = tokenized.input_ids.to(device)
attention_mask = tokenized.attention_mask.to(device)

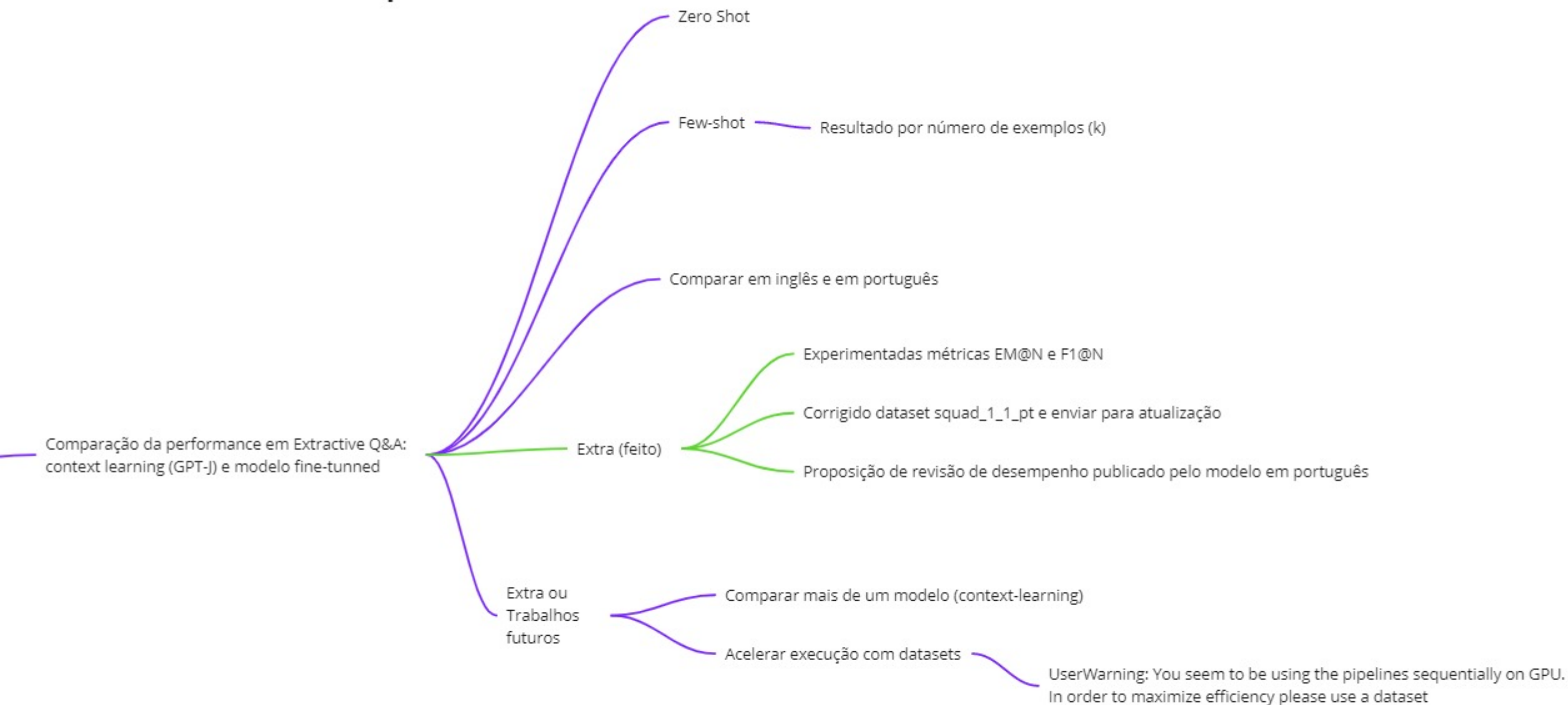
with torch.no_grad():
```

A sessão falhou depois de utilizar toda a RAM disponível. Como utilizador do Colab Pro, tem uma memória superior em comparação com os utilizadores não subscritores, mas se tiver interesse em tempos de execução RAM mais elevados, consulte o [Colab Pro+](#).

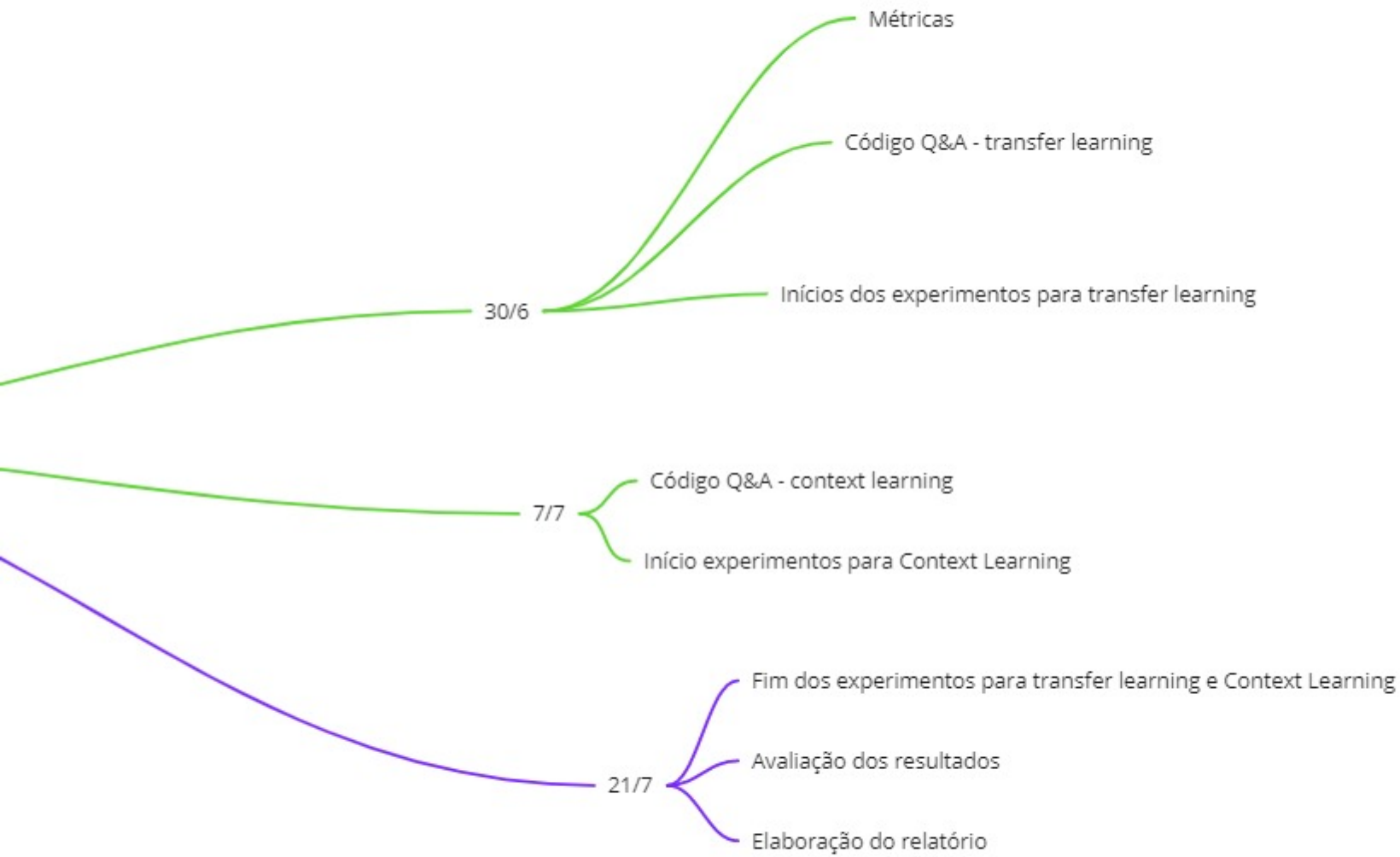
[Veja os registos do tempo de execução](#)

15 min 57 s concluído à(s) 22:14

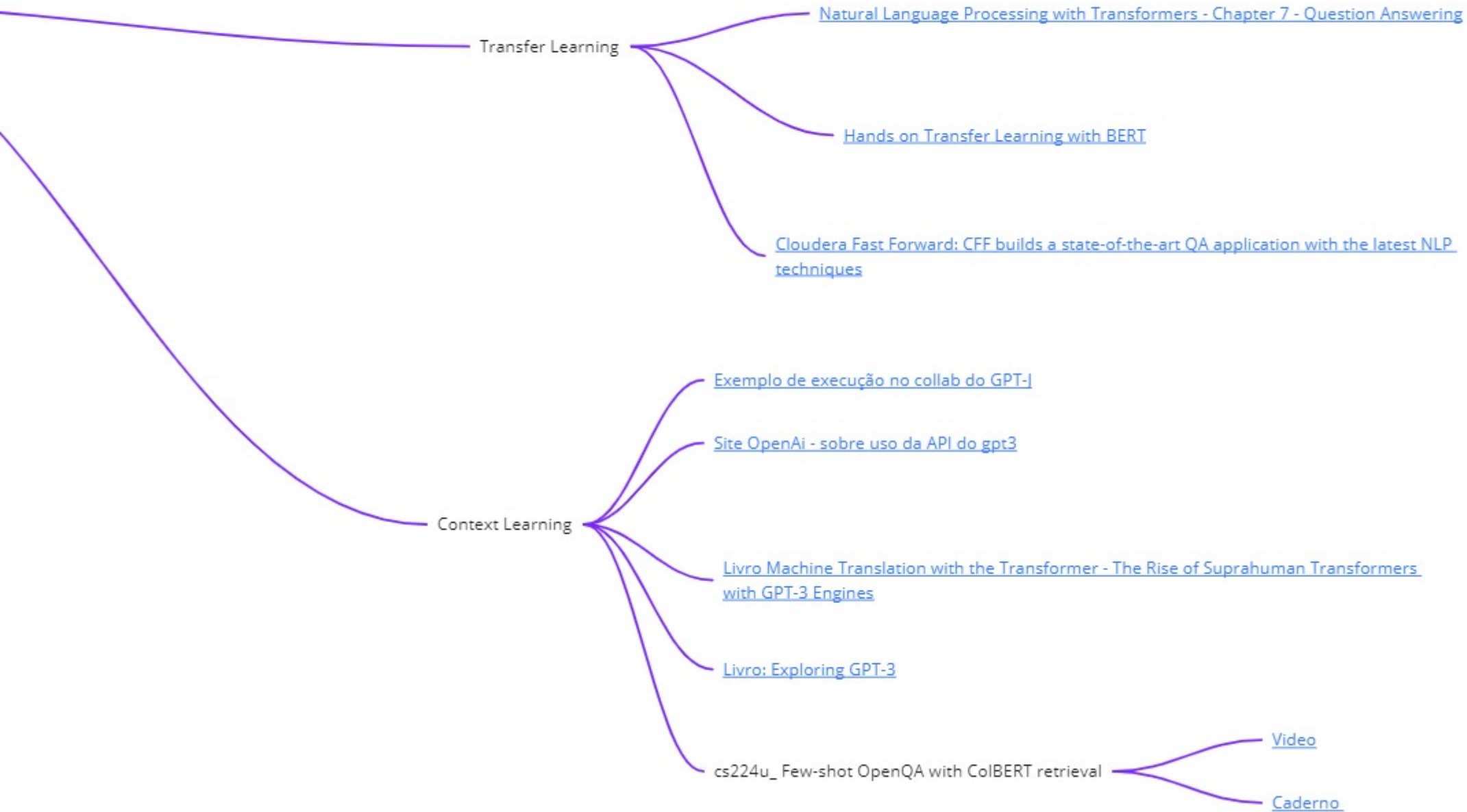
Resultados esperados



Cronograma



Referências



FIM

Github do Projeto

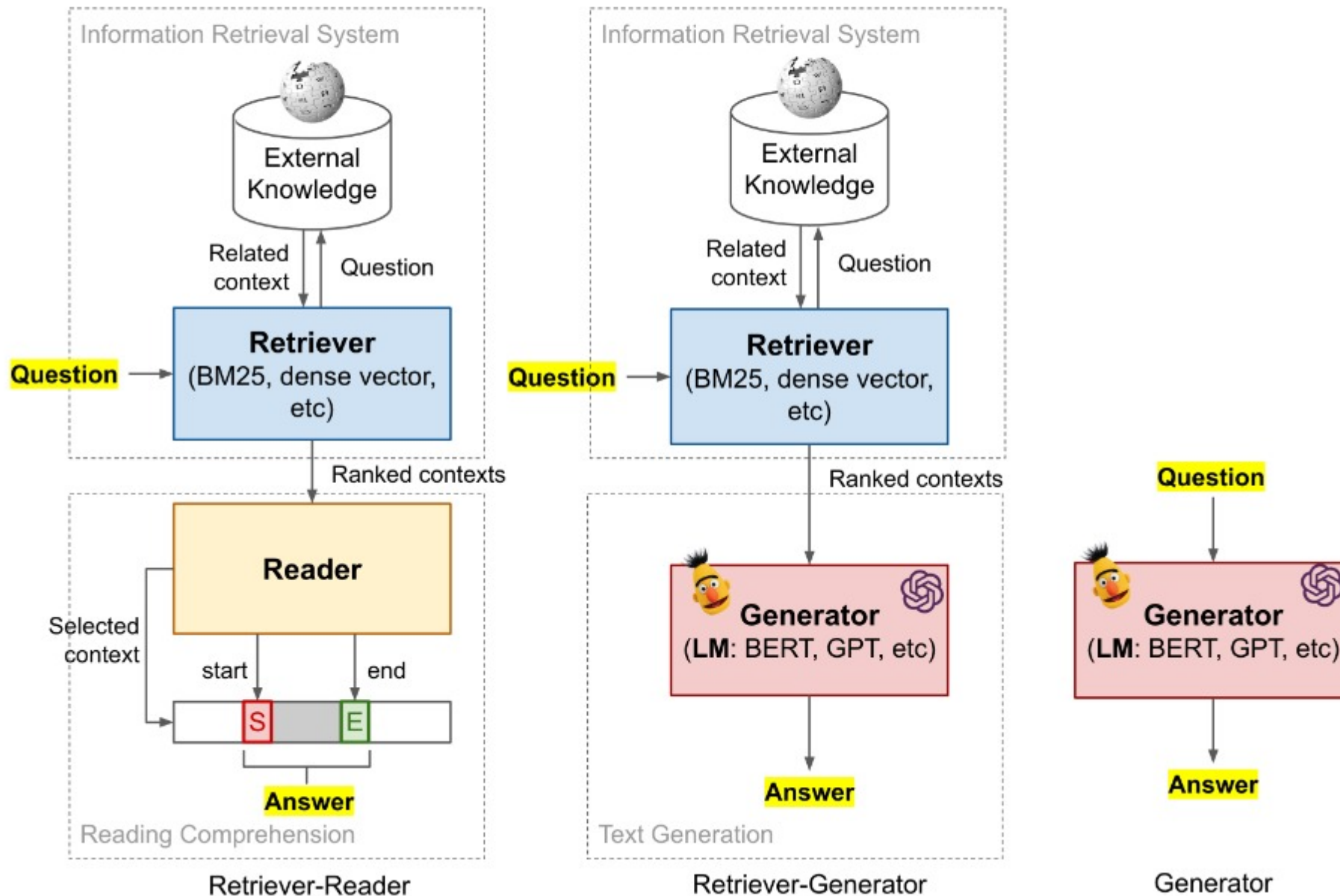
The screenshot shows the GitHub repository page for `marcusborela/exqa-complearning`. The repository is public and has 1 star, 0 forks, and 1 watch. The repository description is "Extractive Q&A - Performance Comparison between Learning Methods: Context Learning and Fine-Tuning".

The repository structure is as follows:

File/Folder	Commit Message	Commit Time
<code>data</code>	Calculo metrica@3 salvando...	4 hours ago
<code>docs</code>	F1@3 e EM@3; reestruturação	4 days ago
<code>models/transfer_lear...</code>	rename fine_tuning para transfer_learning	6 days ago
<code>source</code>	Calculo metrica@3 salvando...	4 hours ago
<code>.gitignore</code>	F1@3 e EM@3; reestruturação	4 days ago
<code>README.md</code>	readme - 1a apresentacao sem link miro	7 days ago
<code>estudo.py</code>	Calculo metrica@3 salvando...	4 hours ago
<code>requirements.txt</code>	calculating fine-tuning; start code organization	7 days ago

The repository also includes a `README.md` file, which is currently open in the editor. The repository has no releases published and no packages published.

Não é contexto: abstractive Q&A

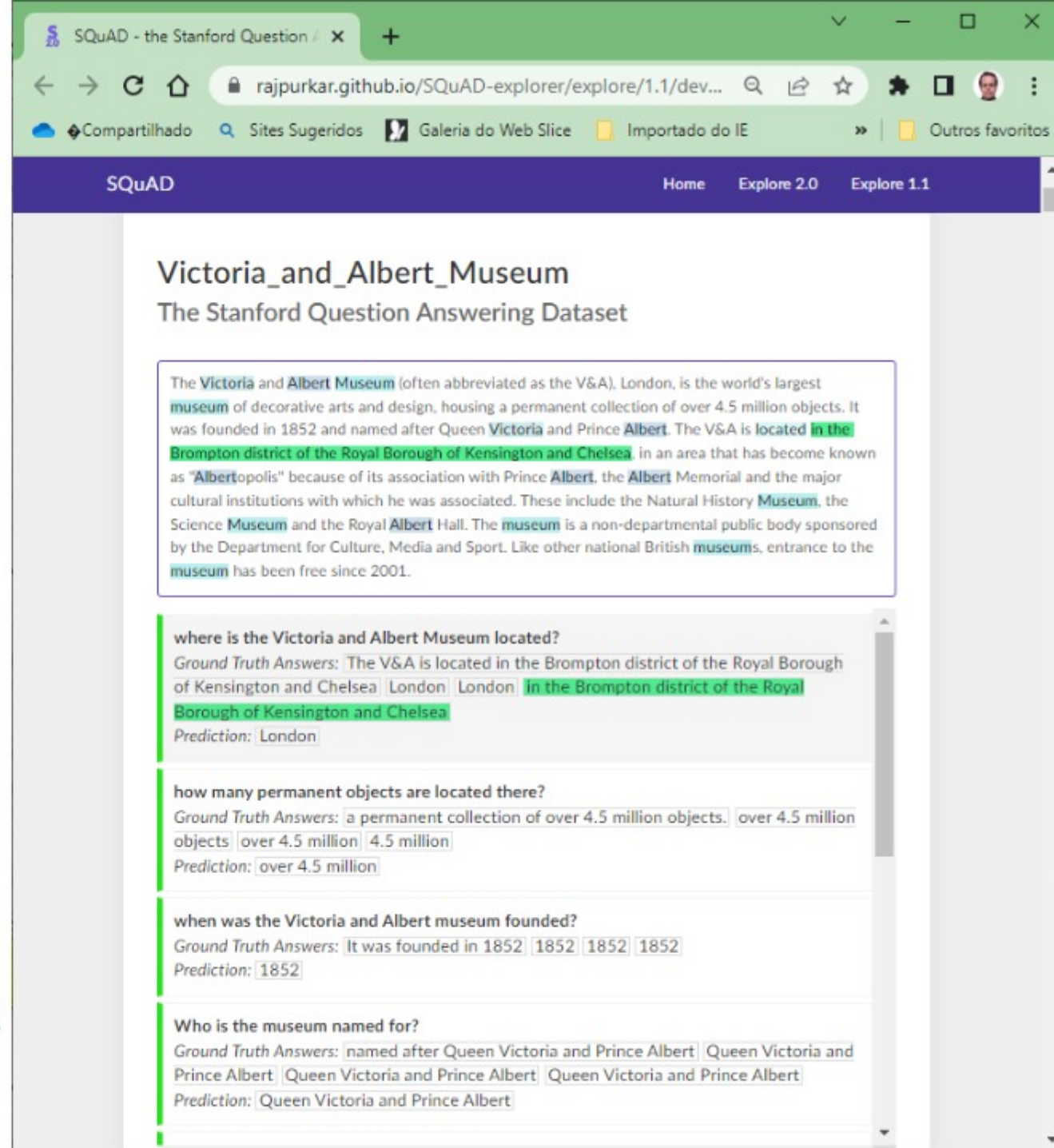


Extractive Q&A - Performance Comparison between
Learning Methods:
Context Learning and Transfer Learning

Resultados Desejados

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...`. The page title is "SQuAD" and the navigation bar includes links for "Home", "Explore 2.0", and "Explore 1.1".

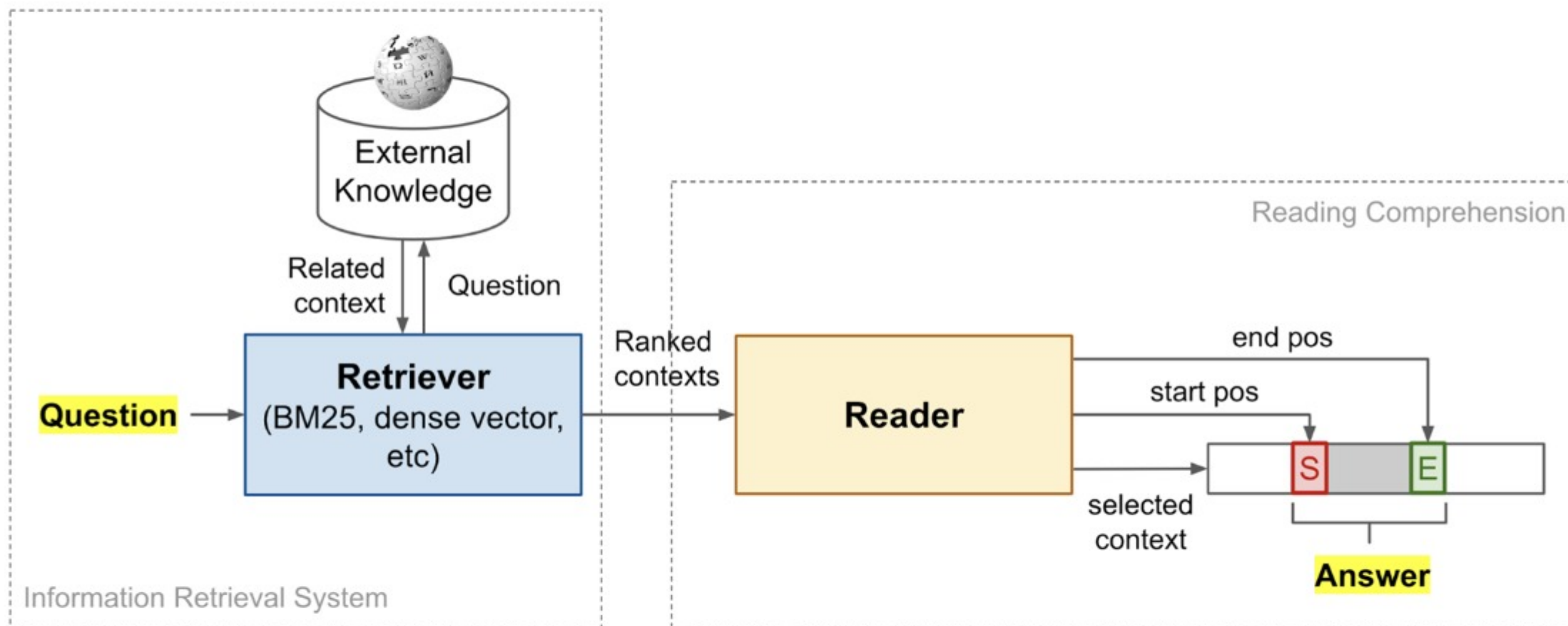
The main content area features a passage titled "Victoria_and_Albert_Museum" with the subtitle "The Stanford Question Answering Dataset". The passage text is: "The **Victoria and Albert Museum** (often abbreviated as the V&A), London, is the world's largest **museum** of decorative arts and design, housing 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, there are four questions with their corresponding ground truth and predicted answers:

- 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

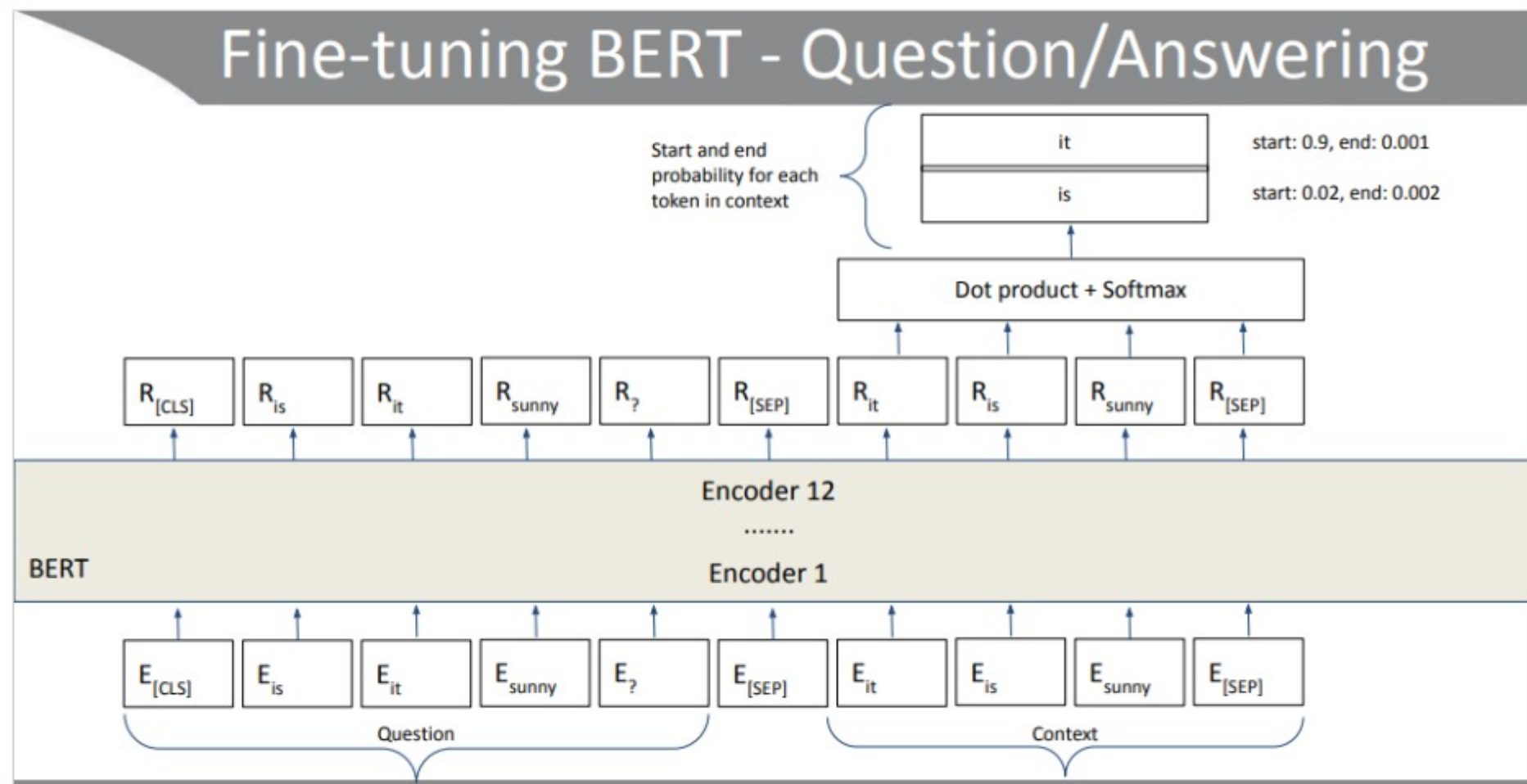
Exemplo: [Stanford Question Answering Dataset SQuAD](#)

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 address bar displays `beta.openai.com/playground/p/default-qa?mode=e...`. The page title is "Playground - OpenAI API". The navigation bar includes links for Overview, Documentation, Examples, and Playground, along with buttons for Upgrade, Help, and Personal. The main section is titled "Playground" and features a dropdown menu set to "Q&A", with buttons for Save, View code, and Share. The interface is divided into two main columns. The left column, labeled "Input", contains a large text area with a paragraph about the Victoria and Albert Museum (V&A) and three Q&A pairs. The right column, labeled "Output", shows the model's responses to these questions. On the far right, there is a "Mode" section with a dropdown menu set to "text-davinci-edit-001", and sliders for Temperature (set to 0) and Top P (set to 1). At the bottom, there is an "Instructions" section with a text input field containing "Answer the question using context fragment" and a "Submit" button. A notification at the bottom right states: "Editing is free while in beta. We'd love your feedback."

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

