Extractive Q&A

Performance Comparison between Learning Methods

Context Learning and Fine-Tuning

Leonardo Augusto da Silva Pacheco
Marcus Vinícius Borela de Castro

Final Project at Discipline IA025 Introdução ao Aprendizado Profundo (2022/01)

Link para github do projeto

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.

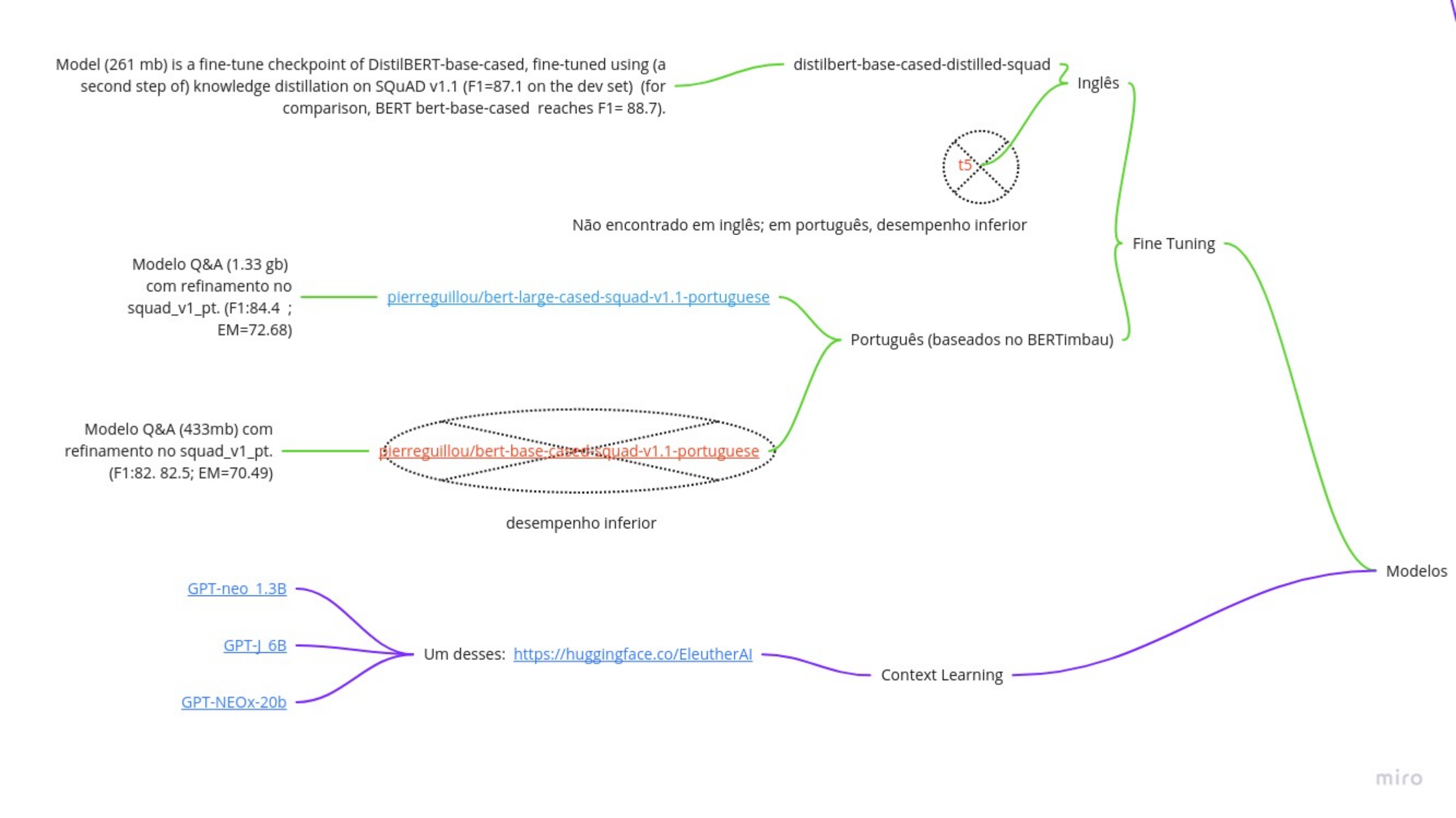
Exact Match (EM)

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

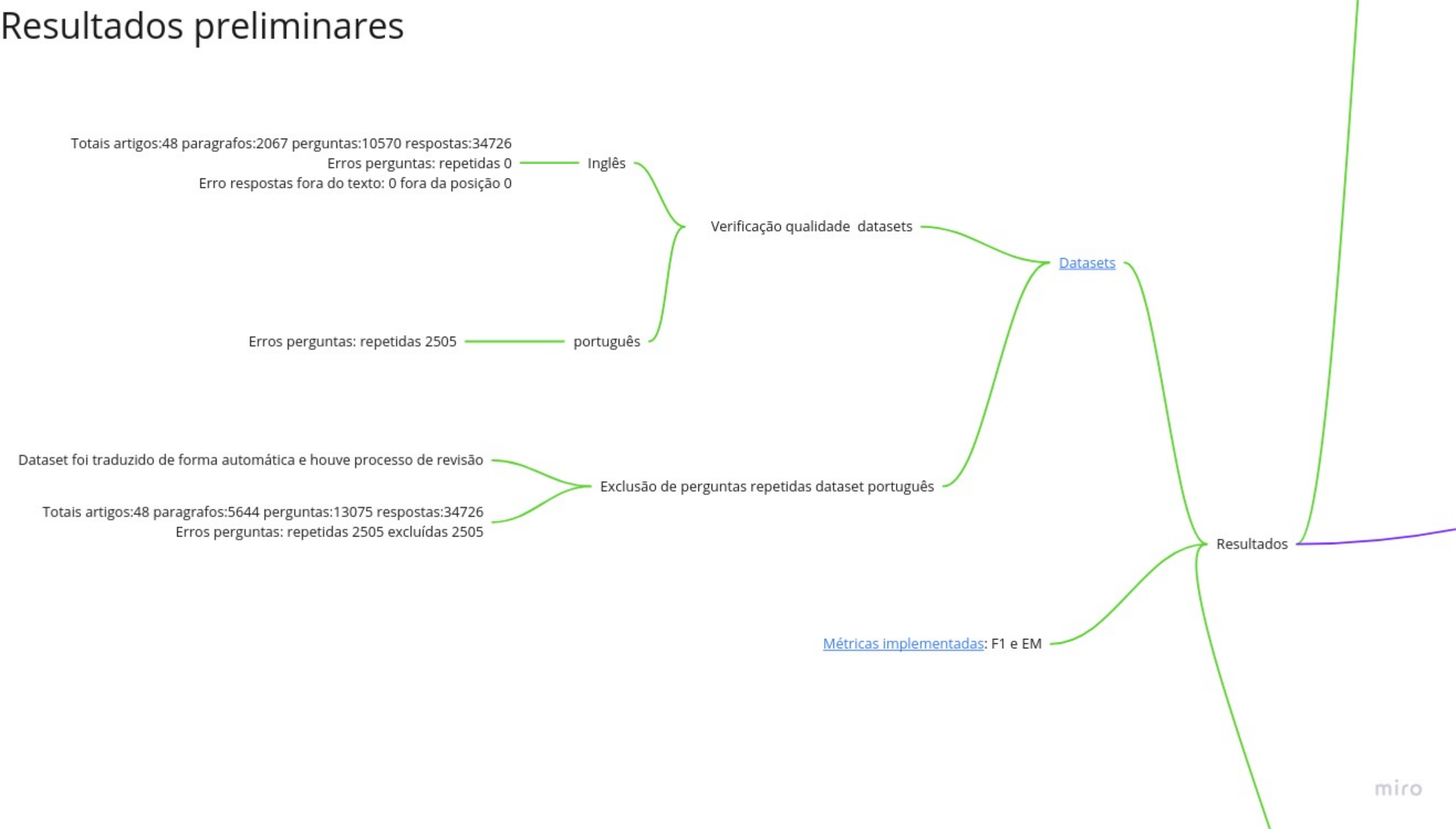
F1

Definição

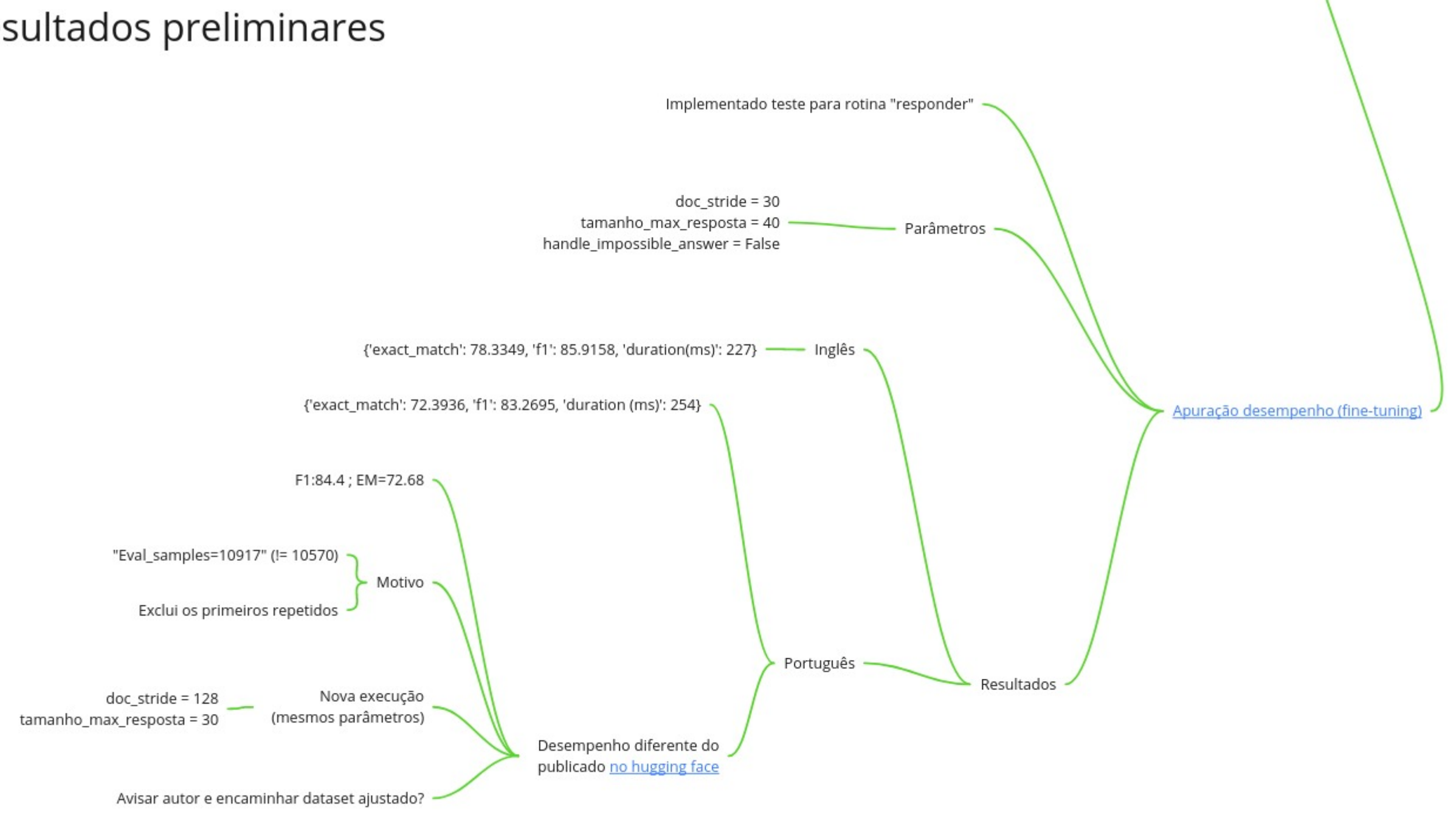
Métricas


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
t5
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
Português (baseados no BERTimbau)
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
desempenho inferior
Modelos
GPT-neo 1.3B
GPT-J 6B
GPT-NEOx-20b
Um desses: https://huggingface.co/EleutherAI
Context Learning

Resultados preliminares



Resultados preliminares



Resultados esperados

Comparação da performance em Extractive Q&A: context learning (GPT-J) e modelo fine-tunned

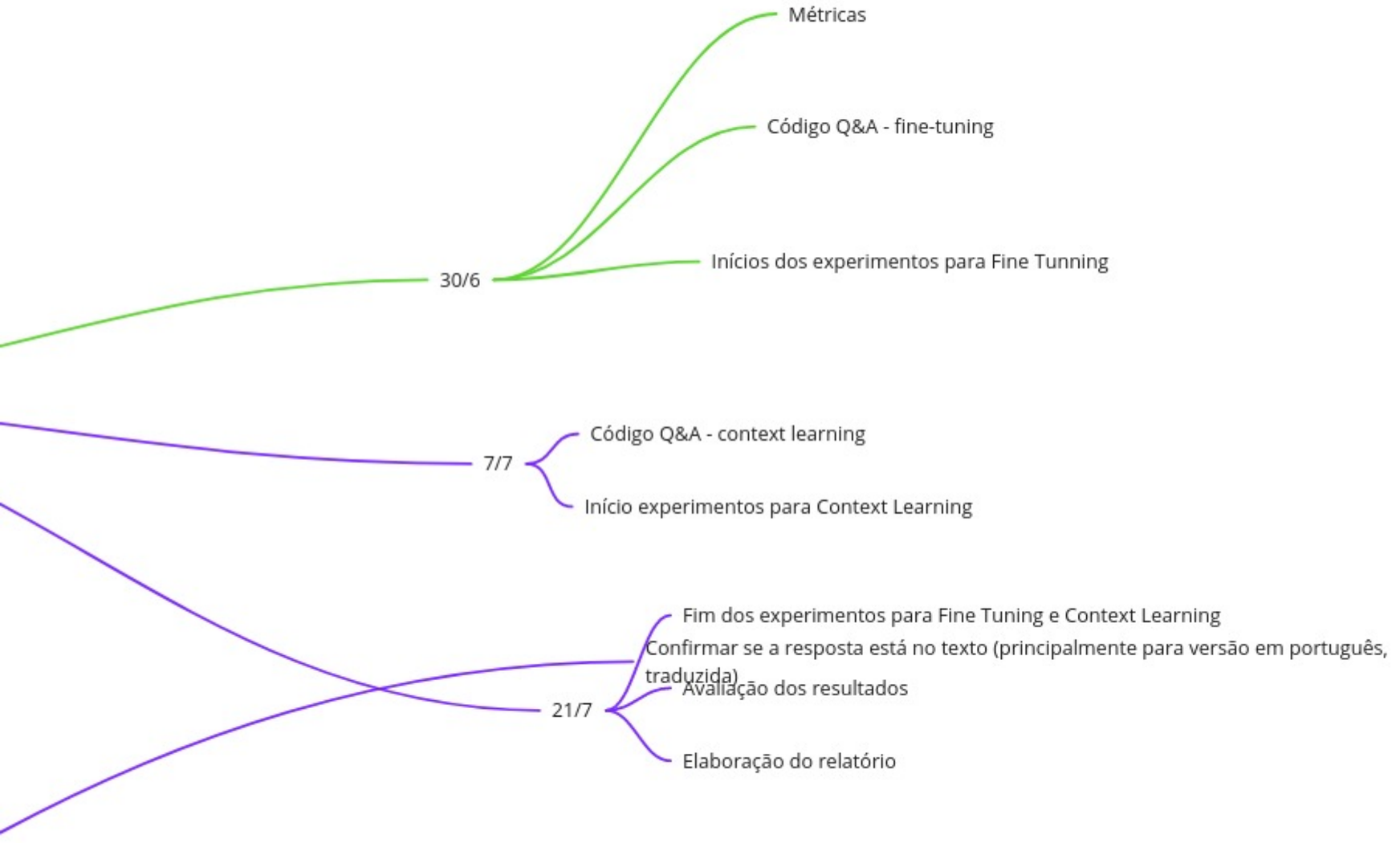
- Zero Shot
- Few-shot
  - Resultado por número de exemplos (k)
    - sugestão Rodrigo: passar k contextos/resposta e não 1 contexto k respostas; e colocar descrição detalhada
    - minha ideia: comparar as 2 situações como resultado esperado?
- Comparar em inglês e em português
- Comparar mais de um modelo (fine-tunning) — Selecionado o melhor!
- Extra ("trabalhos futuros")
  - Por que não expandir métrica EM@N e F1@N?
  - Comparar mais de um modelo (context-learning) — Avaliar impacto do número de parâmetros?
  - Experimentar GPT-3

Tratar casos de ausência de resposta (SQuAD 2.0)

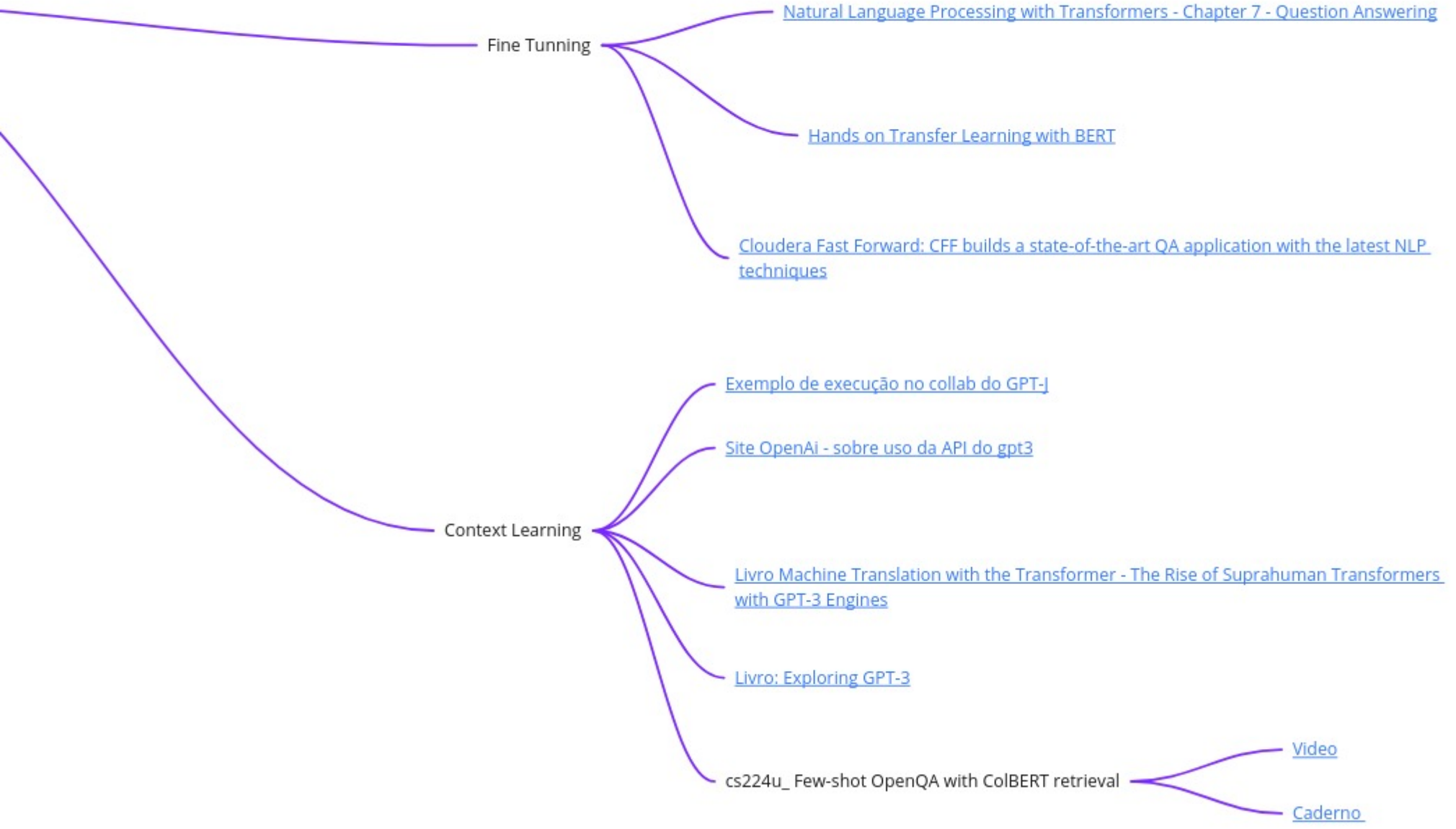
- Considerar não haver resposta (SQuAD 2.0 - só em inglês)
- Traduzir dataset SQuAD 2. para português e publicar no HF
- Fazer avaliação em português (casos sem respostas)

Nem sempre a tradução da resposta pode estar no contexto traduzido.
Grande esforço de revisão.

Cronograma

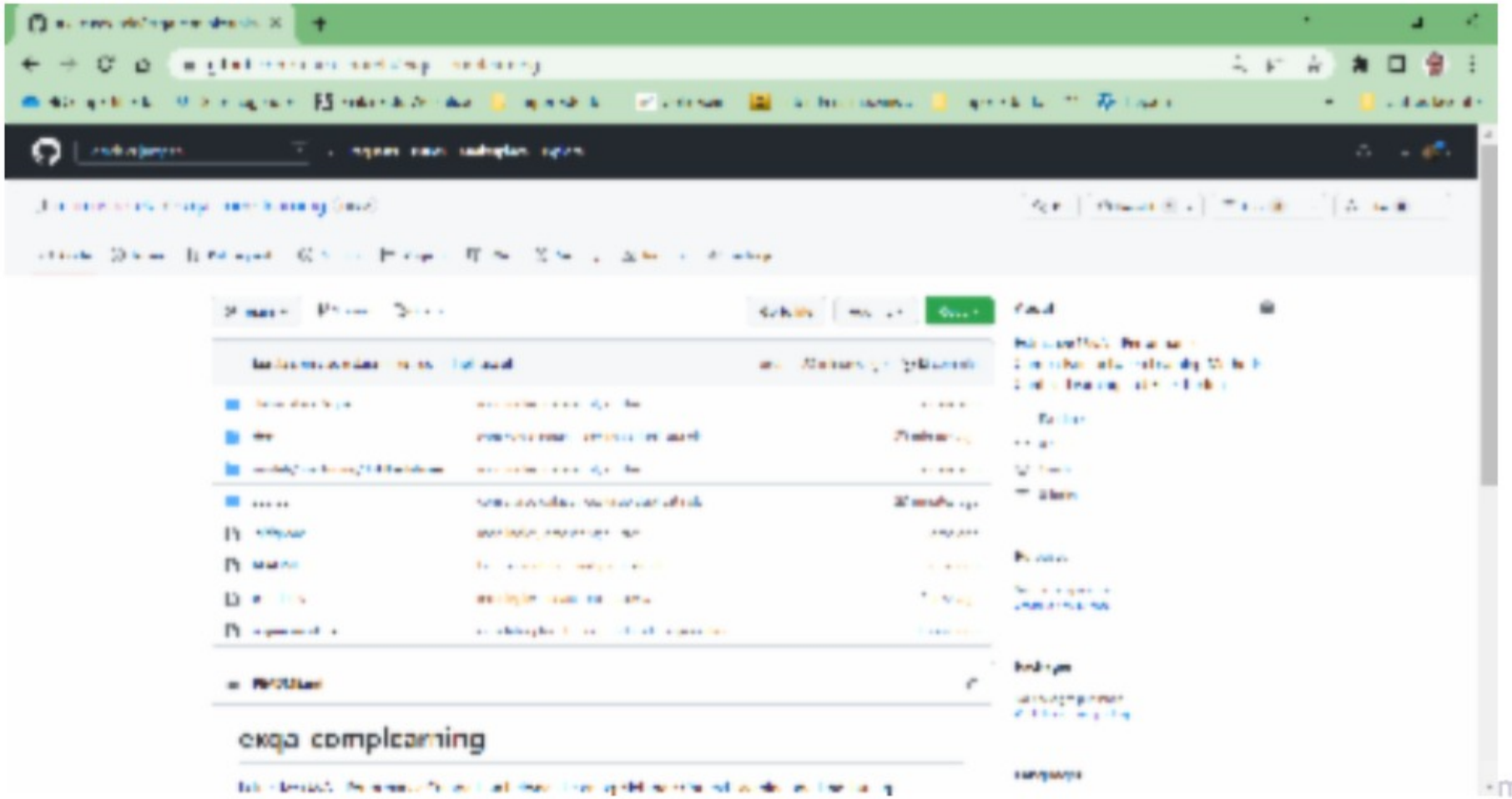


Referências

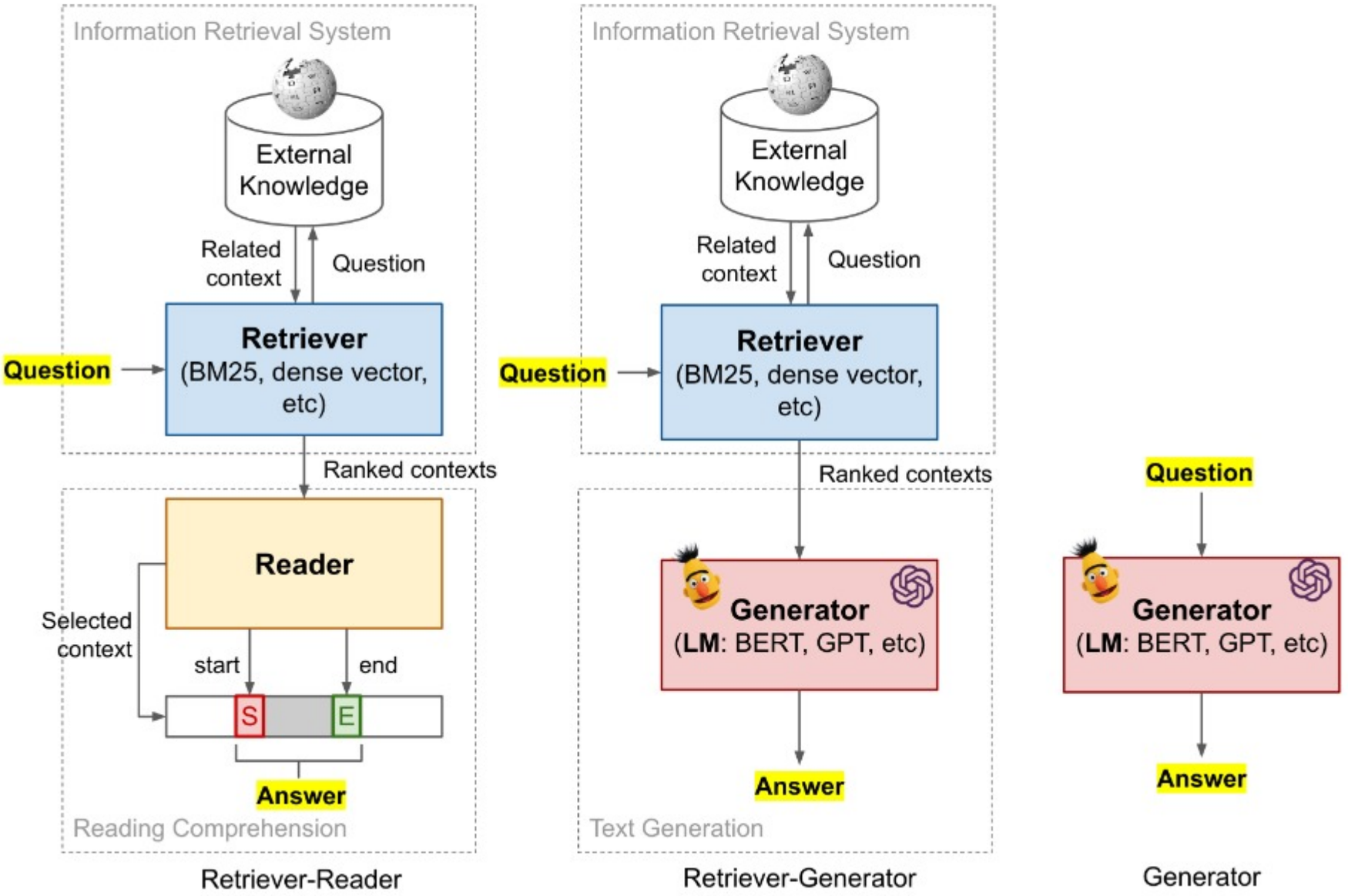


FIM

[Github do Projeto]



Não é contexto: abstractive Q&Ä



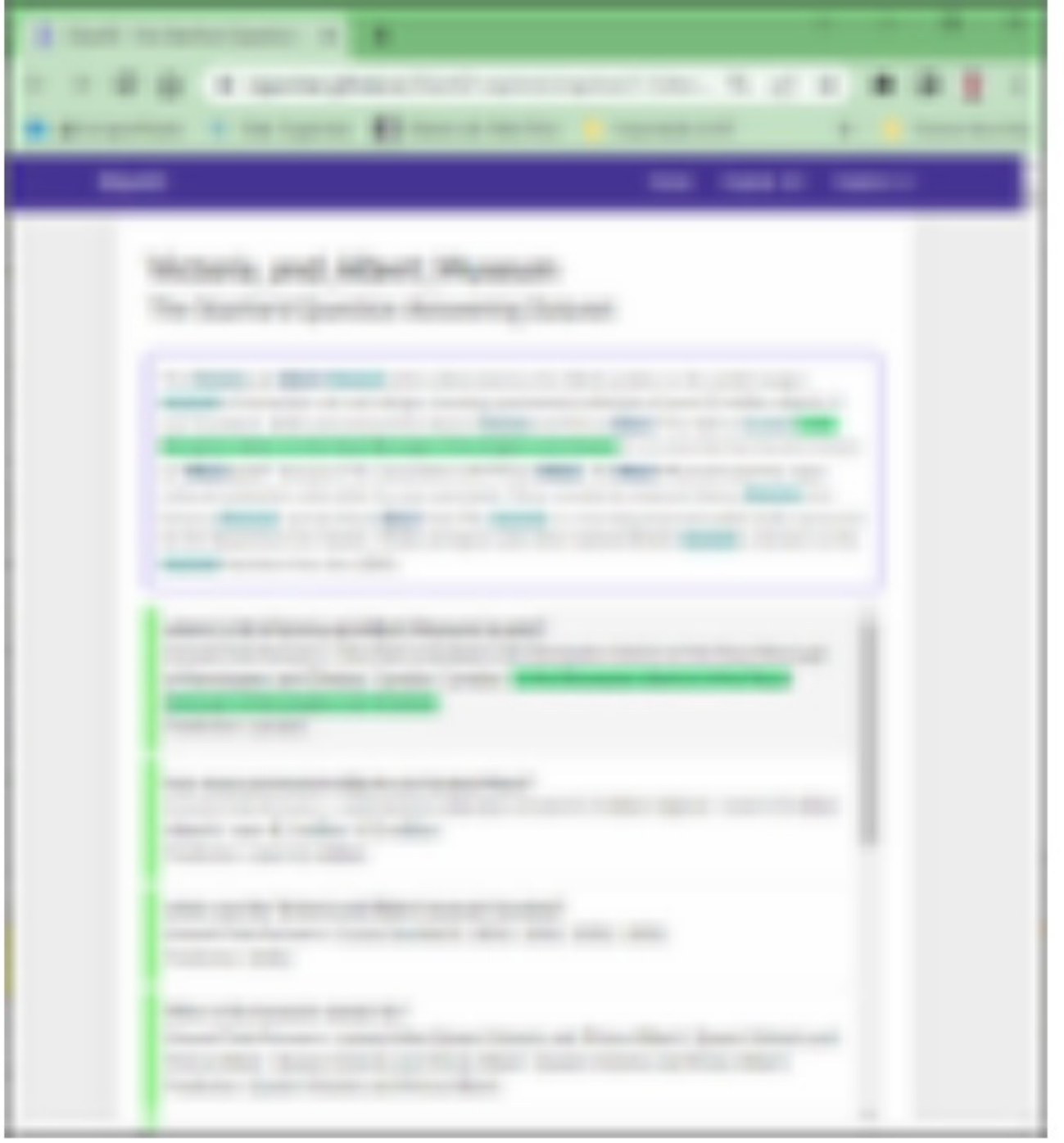
Extractive Q&A - Performance Comparison between Learning Methods:
Context Learning and Fine-Tuning

Modelos

Cronograma

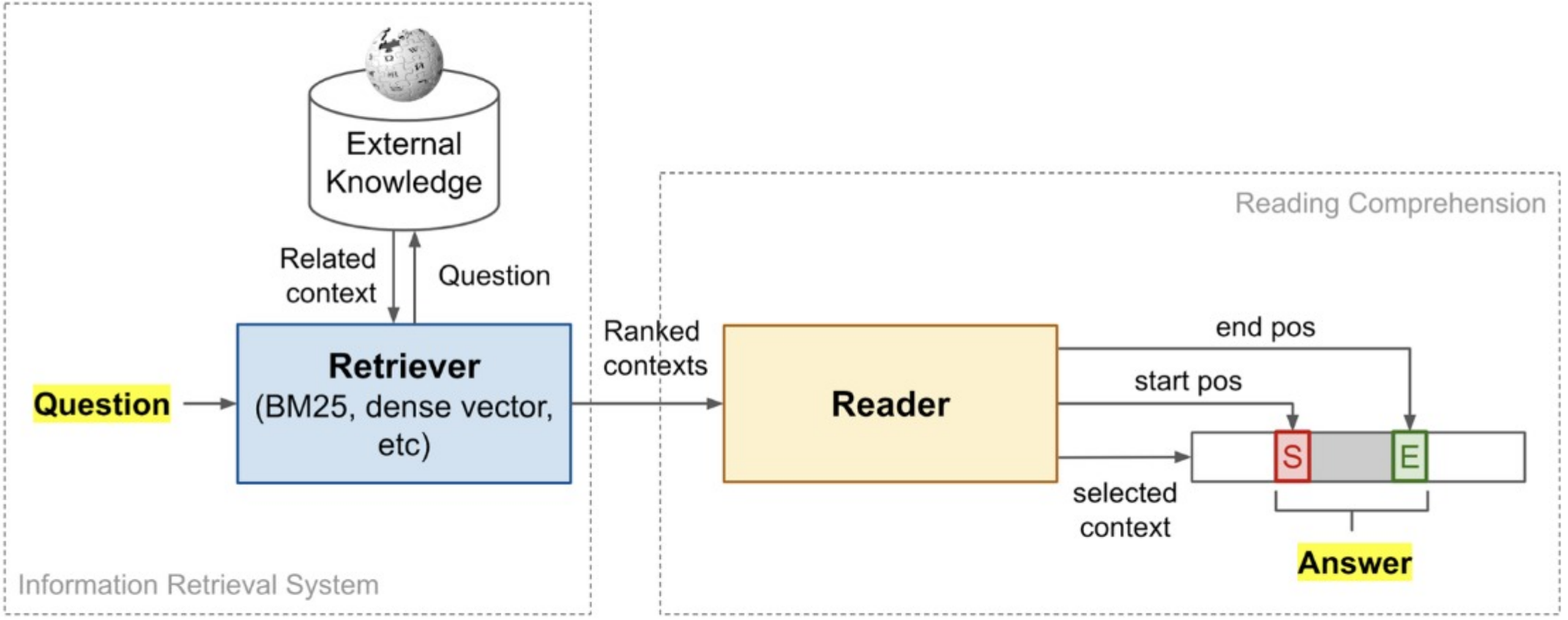
Descrição

O que é Extractive Q&A?



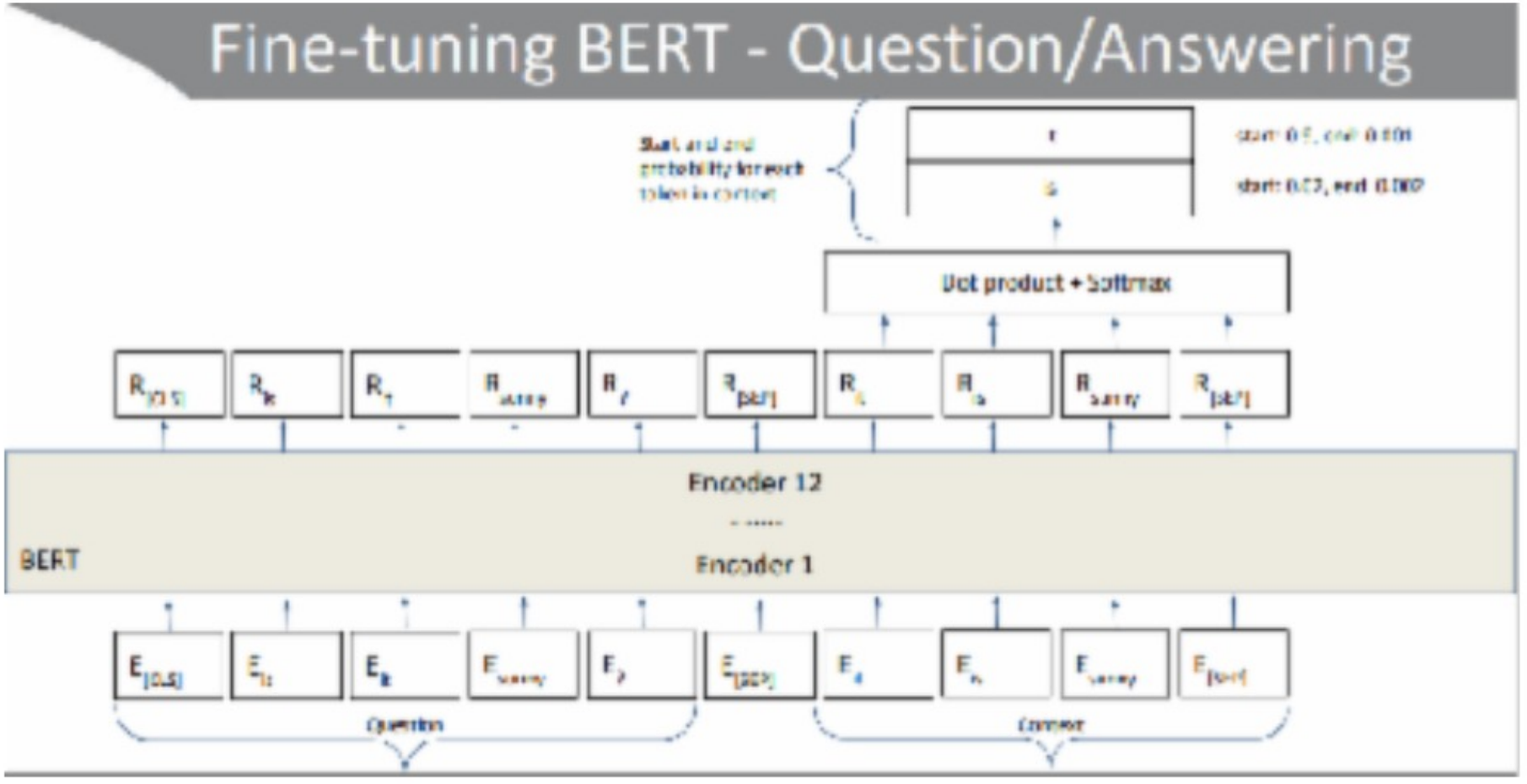
Exemplo: [Stanford Question Answering Dataset SQuAD]

IR Extractive Q&A (over Information Retrieval)



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

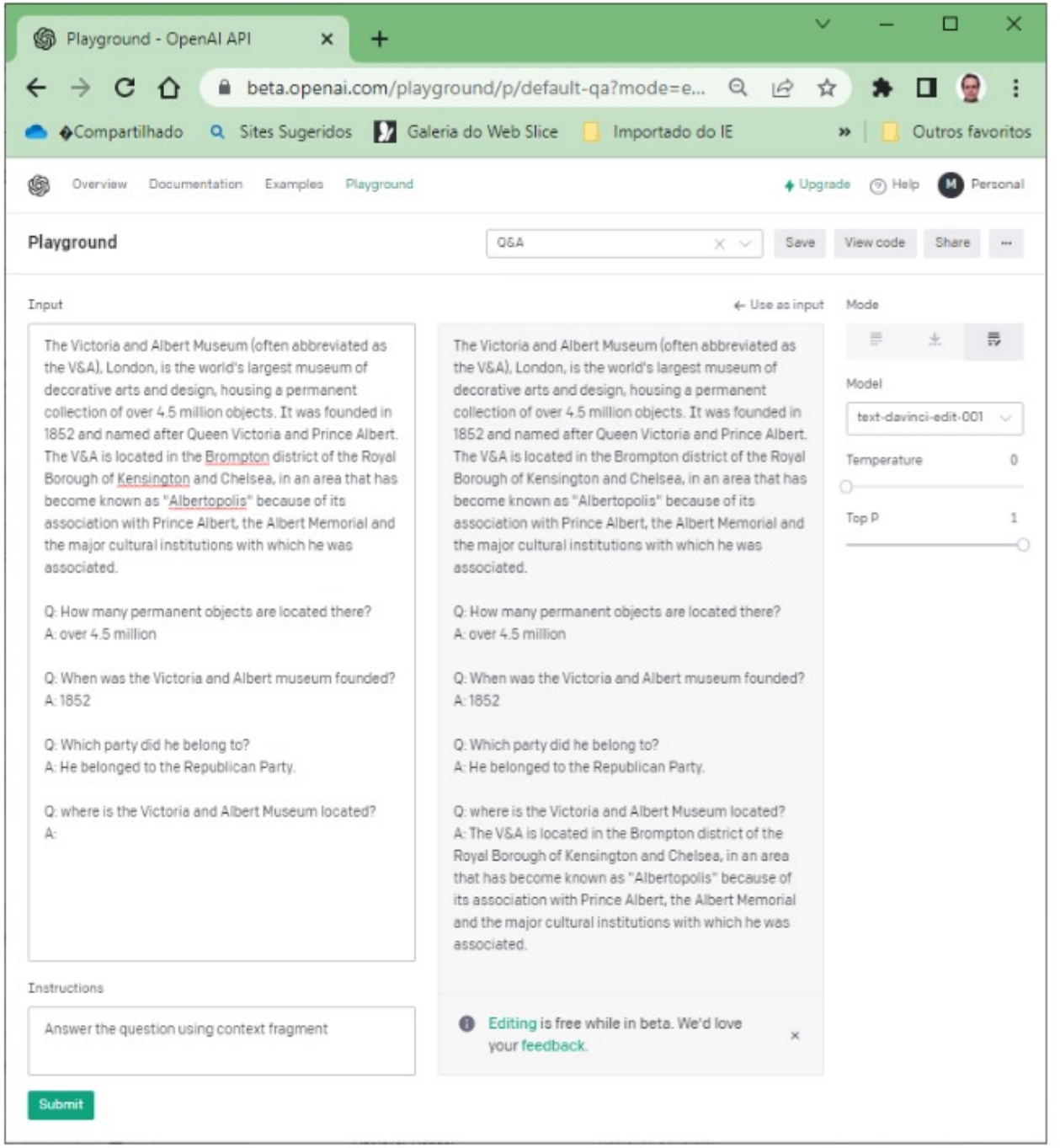
Extractive Q&A in Fine Tuning



Q&A in Context Learning

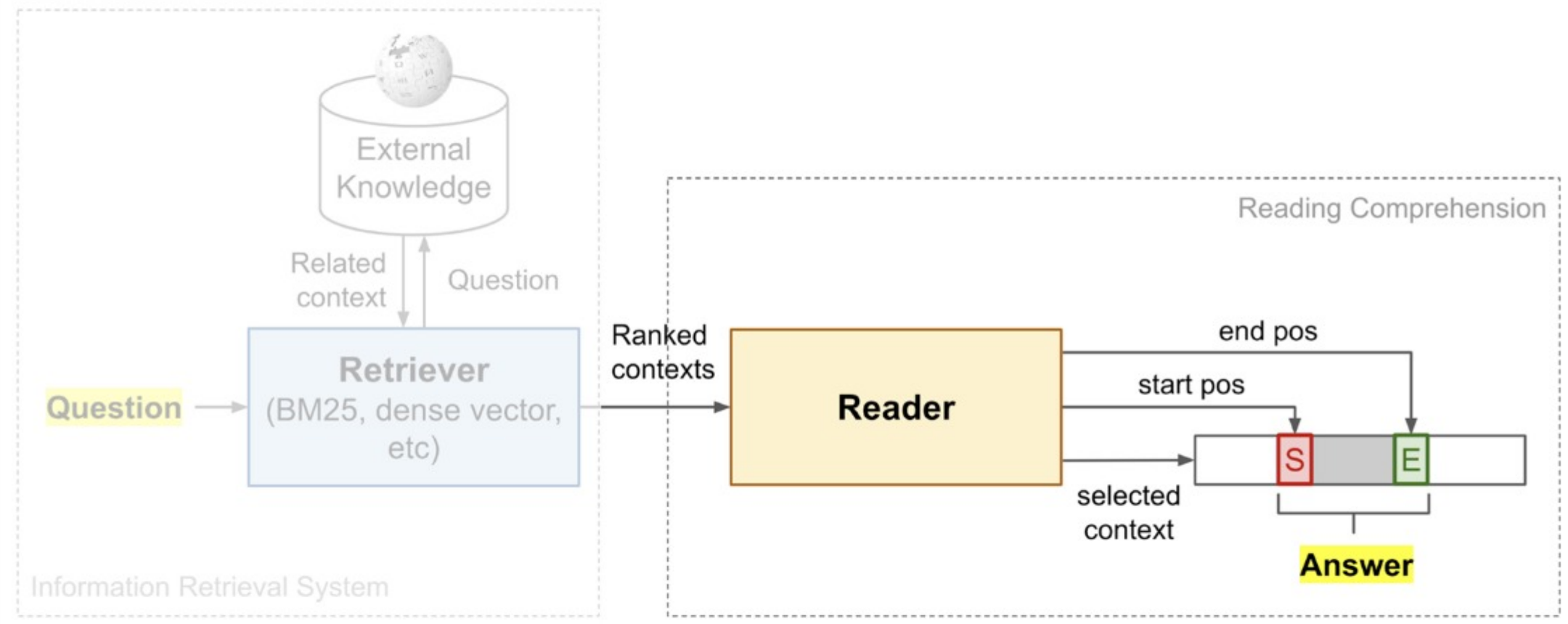
Contexto

Few Shot
k=3



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

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)