



Knowledge Graph Dataset

Generate Human Understandable Text

10 questions about this sentence:



Finding the most relevant queries

Inputs

Source Sentence
Tripple
Sentences to compare to
1st Question
10th Question

Cross Encoder

Outputs

1st Question 0.83
10th Question 0.33

Sort

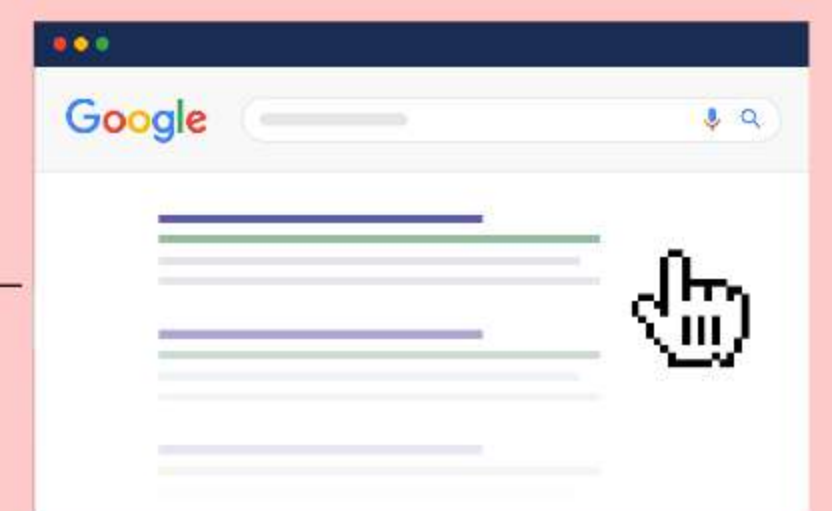
If the top question score Upper threshold

NO

YES

Google Search

Process & Extract Links



Submit N Top Question + Main Question To google and fetch 100 first result

Use

Create A data pool

Context

Chunks

Top K retrieved chunks

Embedding for retrieval tasks (Smaller Chunks)

Embeddings

Similarity Cutoff Strategy

Synthesis

LLM

LLM

LLM

LLM

Response

majority votes

Ask the final judge to resolve the conflicts in the majority vote.

Paid Models

Large Param Model

PASSED

FAILED

2 In this section, we filter the main source of data creation.

(ie: Wikipedia for FactBench Dataset)

3 We use a strategy to check for similarity, and if the similarity percentage falls below a certain threshold, we remove that data.

However, in some cases, removing the data leaves us with no input data.

When this happens, we stop using the strategy altogether.

4 When two of the language models believe the answer is correct, and two believe it is wrong, we switch to a more advanced model, such as a commercial one or a model with more parameters.

This model is then used to make the final decision.