Generating Answers

Gangzhaorige Li

github:https://github.com/gangzhaorige/ML-OPENAi-Customer Support/tree/main/Fine-Tuning/GeneratingAnswers

Project Overview: Generating Answers

1. Process Input texts
    a. Split into a list of paragraphs.
    b. Clean up to remove empty spaces and new line.
2. Embedding
    a. Chunking: Get the embeddings from input texts.
    b. Build a search index from embedding.
    c. Check the dimensions of the embeddings.
    d. Create the search index, pass the size of embeddings.
    e. Add all the vectors to the search index
3. Searching Articles
4. Generating answers

Process input texts.

```
##################################################
# Step 1. Preprocessing input texts
##################################################

##################################################
# Step 1.1 Split into a list of paragraphs
##################################################
texts = text.split('\n\n')

##################################################
# Step 1.2 Clean up to remove empty spaces and
#          new lines
##################################################
texts = np.array([t.strip(' \n') for t in
                  texts if t])
```

Embedding

```
co = cohere.Client(os.environ['COHERE_API_KEY'])

##################################################
# Step 2.1 Chunking: Get the embeddings (vectors)
#                    from input texts
##################################################
response = co.embed(
    texts=texts.tolist(),
).embeddings
```

Embedding Continued

```
################################################
# Step 2.2.1 Check the dimensions of the
#            embeddings

embeds = np.array(response)

################################################
# Step 2.2.2 Create the search index, pass the
#            size of embedding (vector)
################################################
search_index = AnnoyIndex(embeds.shape[1], 'angular')

################################################
# Step 2.3 Add all the vectors to the search index
################################################
for i in range(len(embeds)):
    search_index.add_item(i, embeds[i])

# 10 trees
search_index.build(10)
search_index.save('test.ann')
```

Searching Articles

```
def search_andrews_article(query):
    # Get the query's embedding
    query_embed = co.embed(texts=[query]).embeddings

    # Retrieve the nearest neighbors
    similar_item_ids = search_index.get_nns_by_vector(
                    query_embed[0],
                    10,
                    include_distances=True)

    search_results = texts[similar_item_ids[0]]

    return search_results
```

Generating Answers Test Case 1

Are side projects a good idea when trying to build a career in AI?

```
Join existing projects. If you find someone else with an
idea, ask to join their project.
Keep reading and talking to people. I come up with
new ideas whenever I spend a lot of time reading,
taking courses, or talking with domain experts. I'
m confident that you will, too.
Focus on an application area. Many researchers are trying to
advance basic AI technology – say, by inventing the next
generation of transformers or further scaling up
language models – so, while this is an exciting
direction, it is hard.
```

Generating Answers Test Case 2. 3 generations

Are side projects a good idea when trying to build a career in AI?

```
Here's a quick checklist of factors to consider:
id='e288a1e5-f624-4dd0-922f-cd98bc733878' text="Yes, side projects or 'side
ed as a way to explore new ideas and develop creative skills." index=None l
kelihoods=None finish_reason='COMPLETE'
--
id='91d0cb2e-c93b-4203-b12d-a57c1cfea46e' text='Yes.' index=None likelihood
=None finish_reason='COMPLETE'
--
id='41ccedcf-b551-4590-a9b8-cca44807dcd9' text='Yes, side projects or "side
cial when building a career in AI. They can stimulate creativity and foster
successful, they can lead to significant opportunities.' index=None likelih
ods=None finish_reason='COMPLETE'
--
```

Generating Answers Test Case 2. 5 generations

What is the most viewed televised event?

```
id='8a2ec8fe-237f-44d9-978d-e1f1ed45eac4' text='Yes, side projects are a great way to exp
 interests and develop new skills. Andrew Ng suggests that side projects can "stir the cr
ices" and even lead to something more significant, such as a startup.' index=None likelih
token_likelihoods=None finish_reason='COMPLETE'
--
id='ad0e1f75-959f-4fea-b75c-9f7d396b5630' text="Yes, side projects or 'side hustles' can
cial when building a career in AI." index=None likelihood=None token_likelihoods=None fin
n='COMPLETE'
--
id='e1934eed-d718-461c-85bf-9f183097b42e' text='The text provided does not contain inform
ut the most viewed televised event. Therefore, I cannot extract the answer to your questi
x=None likelihood=None token_likelihoods=None finish_reason='COMPLETE'
--
id='c4d40cf7-3b1b-44f4-9ca1-861cdcec645f' text='The text does not provide information abo
st-viewed televised event. Therefore, I cannot extract the answer from the given text.' i
 likelihood=None token_likelihoods=None finish_reason='COMPLETE'
--
id='8dd780f7-5f73-4980-aa00-8f0a6a559784' text='The text provided does not contain inform
ut the most viewed televised event. Therefore, I cannot extract the answer to your questi
he given text.' index=None likelihood=None token_likelihoods=None finish_reason='COMPLETE
--
id='eddf9803-4ca6-40a2-bc43-063cc6abe6ca' text='The text provided does not contain inform
ated to the most viewed televised event. Therefore, I cannot extract the answer to your q
 index=None likelihood=None token_likelihoods=None finish_reason='COMPLETE'
```