WEEK 10 HOMEWORK 4: GENERATING ANSWERS: INPUT TEXT ==> CHUNKING ==> EMBEDDING ==> SEARCH INDEX ==> QUERY ==> SEARCH ==> QUESTION ==> ANSWER

STEP 1: Preprocessing Input text

The given text is split into list of paragraphs, then newline and empty spaces are removed from it.

```python
texts = text.split('\n\n')
```
✓ 0.0s

```python
texts = np.array([t.strip(' \n') for t in texts if t])
```
✓ 0.0s

STEP 2: Embedding

Get the embeddings and build a search index from embeddings. Add all vectors to the search index.

```python
embeds = np.array(response)

search_index = AnnoyIndex(embeds.shape[1], 'angular')
```
✓ 0.0s

```python
for i in range(len(embeds)):
    search_index.add_item(i, embeds[i])

# 10 trees
search_index.build(10)
search_index.save('test.ann')
```
✓ 0.6s

True

STEP 3: Searching

```python
embeds = np.array(response)

search_index = AnnoyIndex(embeds.shape[1], 'angular')
```
✓ 0.0s

```python
for i in range(len(embeds)):
    search_index.add_item(i, embeds[i])

# 10 trees
search_index.build(10) |
search_index.save('test.ann')
```
✓ 0.6s

True

STEP 4: Generating answers.

```python
results = ask_andrews_article(
    "Are side projects a good idea when trying to build a career in AI?",
    num_generations=3
)

for gen in results:
    print(gen.text)
    print('--')
```
[43] ✓ 4.4s

... Yes, side projects or "side hustles" can be beneficial when building a career in AI, as they can stimulate creativity and strengthen connections with colleagues.
--
Yes, side projects or "side hustles" are recommended as they can stimulate creativity and foster collaboration.
--
Yes, side projects or "side hustles" can be beneficial and stimulate creativity, as well as strengthen relationships with collaborators.
--

```python
results = ask_andrews_article(
    "What is the most viewed televised event?",
    num_generations=5
)

for gen in results:
    print(gen.text)
    print('--')
```
[44] ✓ 0.9s

... The text provided does not contain information related to the most viewed televised event. Therefore, I cannot extract the answer to your question.
--
The text provided does not contain information about the most-viewed televised event. Therefore, I cannot extract the answer to your question.
--
The text provided does not contain information related to the most viewed televised event. Therefore, I cannot extract the answer to your question.
--
The text provided does not contain information related to the most viewed televised event. Therefore, I cannot extract the answer to your question.
--
The text provided does not contain information related to the most viewed televised event. Therefore, I cannot extract the answer to your question.
--

GITHUB LINK: https://github.com/savithashreem07/Fine-Tuning/tree/feature/generating-answers/Generating%20Answers

GOOGLE SLIDE LINK:
https://docs.google.com/presentation/d/1wxEczgnn4tvaQPU5N5b3KTIgAnU8hAVtCXb-IZUWA38/edit?usp=sharing