



Hotels Search Engine

What you did?

After downloading the datasets, we preprocessed the dataset:

- Fixed ambiguous misspellings (e.g., Türkiye, Turkiye)
- Removed duplicate rows

After so, create a dataframe that contains:

- Hotel name
- Hotel review
- Locality
- Country
- Rating value

Next, we grouped all the reviews according to hotel name. We applied the summary model: [facebook/bart-large-cnn](#) to summarize the hotel reviews.

For embedding, [nomic-ai/nomic-embed-text-v1.5](#) model was the one utilized, feeding it all hotel information.

For the search engine, we used Inverted File Index to obtain the top results of the requested query.

To create a user-friendly interface, we implemented Gradio.

How did you collate the data on the hotel?

The provided hugging face dataset is: [traversaal-ai-hackathon/hotel_datasets](https://huggingface.co/traversaal-ai-hackathon/hotel_datasets).

Simple feedback on your search

- We demonstrated strong problem-solving abilities, particularly in overcoming technical challenges.
- Brainstorming sessions were productive, leading to innovative solutions and optimizations in the search engine.

Ruba Alhudyani, Nwair Alsahli, Razan Alzahrani

Pages:

Attachments

[ASSIGNMENT3_HOTEL_Nawir_Ruba_Razan.ipynb](#)