

17megabyteCHATBOT

This repository contains the implementation details and code for the "17megabyteCHATBOT", a lightweight, efficient chatbot powered by the 'ms-marco-TinyBERT-L-2' model, as described in Fabio Matricardi's article on medium.com. The chatbot utilizes semantic search and a Cross-Encoder architecture to understand user queries and deliver accurate responses, making it an excellent solution for small websites or businesses looking to incorporate AI without the overhead of large models.



Features

- **Semantic Understanding:** Uses the 'ms-marco-TinyBERT-L-2' model to understand the context and semantics of user queries.
- **Lightweight:** Only requires 17 MB of memory, making it highly efficient and easy to deploy even on limited hardware.
- **No Generative AI:** Relies on a Retriever-based approach, matching user queries with the most relevant answers from a curated dataset without generating responses on the fly.
- **Easy Integration:** Designed for straightforward integration into existing systems or websites.
- **Open Source:** Fully open-source with detailed documentation, allowing for customization and improvements.

Installation

To set up the 17megabyteCHATBOT on your local machine, follow these steps:

1. Clone this repository:

```
git clone https://github.com/fabiomatricardi/17megabyteCHATBOT.git
```

2. Navigate to the cloned repository directory:

```
cd 17megabyteCHATBOT
```

3. Install the required Python packages:

```
pip install sentence_transformers rich
```

4. For the Gradio application you need to install the following package:

```
pip install gradio jsonlines
```

Note: This will also automatically install `torch` and `transformers`, which are dependencies of `sentence_transformers`.

Usage

To use the chatbot, run the provided Python script. The script will initiate a chat interface where you can input your queries, and the chatbot will respond with the most relevant answers from the curated dataset.

```
python main.py
```

Customizing Your Chatbot

To customize the chatbot for your specific needs, you can modify the `listato` variable in the script to include your own question and answer pairs. This will allow the chatbot to respond with answers tailored to your domain or business.

Contributing

Contributions to the 17megabyteCHATBOT are welcome. Please feel free to fork the repository, make your changes, and submit a pull request with your improvements.

License

This project is open-source and available under the MIT License. See the LICENSE file for more details.

Acknowledgments

- Thanks to Fabio Matricardi for the original article and concept.
- Powered by 'sentence_transformers' and 'ms-marco-TinyBERT-L-2' for semantic understanding capabilities.