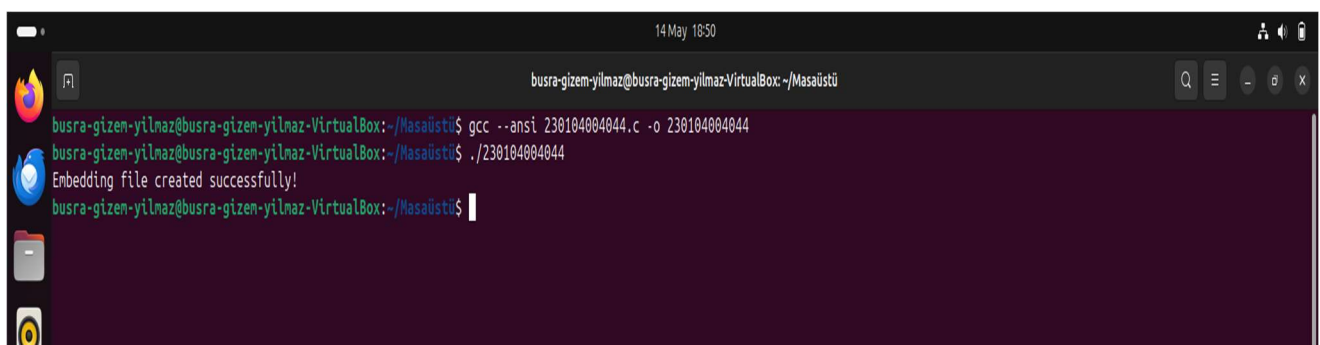


## OUTPUTS

- \*When the program starts, it reads data from the file named "database.txt".
- \*The database consists of multiple question-answer pairs separated by a line containing three hyphens (---).
- \*Each question and answer is read line by line and stored for further processing.
- \*The program then tokenizes each sentence by splitting it into words using spaces as the separator.
- \*Each word is broken down into characters to perform character-based one-hot encoding.
- \*A unique set of characters is determined from the entire dataset to define the one-hot vector space.
- \*Each character in the word is represented as a one-hot vector, and these vectors are combined to form the word representation.
- \*Each word is left-padded with zero vectors so that all word embeddings have the same length as the longest word in the dataset. Additionally, sentence embeddings for both questions and answers are left-padded to ensure that all sentence embeddings reach a uniform length based on the longest encoded sentence.
- \*Sentence embeddings for both questions and answers are generated by combining the word embeddings for each sentence.
- \*All embedding data is stored in a multi-dimensional array, with fixed dimensions for consistency across all entries.
- \*The final embeddings are written into the file named "embeddings.txt" which consists of two main sections:
  - A metadata section, including:
    - The maximum sentence embedding length.
    - The dimension of the word embeddings.
    - The total number of question-answer pairs.
  - A data section where each question and answer is listed with its corresponding embedding.
    - Each entry in the data section is separated by --- as in the input.
- \*The program is modular and organized into specific functions:
  - read\_data() reads and separates the question-answer pairs.
  - tokenize() splits sentences into words.
  - one\_hot\_encode\_word() performs character-based one-hot encoding.
  - one\_hot\_encode\_sentence() ensures equal length across all sentence embeddings.
  - write\_embeddings\_file() formats and writes the output into "embeddings.txt".



```
14 May 18:50
busra-gizem-yilmaz@busra-gizem-yilmaz-VirtualBox: ~/Masaüstü
busra-gizem-yilmaz@busra-gizem-yilmaz-VirtualBox:~/Masaüstü$ gcc --ansi 230104004044.c -o 230104004044
busra-gizem-yilmaz@busra-gizem-yilmaz-VirtualBox:~/Masaüstü$ ./230104004044
Embedding file created successfully!
busra-gizem-yilmaz@busra-gizem-yilmaz-VirtualBox:~/Masaüstü$
```

