Gerardo Alemán Manzanarez

https://github.com/Manzanarez/rosaespino

Instructions to execute Indexing\_files2.py:

# 1.- Download the program Indexing\_files2.py which is in the branch indexing\_docs (<https://github.com/Manzanarez/rosaespino/tree/indexing_docs/code>).

# 2.- Create a folder to download Rosa Espino’s poems from <https://github.com/Manzanarez/rosaespino/tree/main/corpus/rosa> .

# 3.- Download all of Rosa Espino’s poems to the folder created in step 3.

# 4.- Open the program Indexing\_files2.py.

# 5.- Adjust variable path\_ix to the path where the index file will be created.

# 6.- Adjust variable path file to the path where are the Rosa Espino poems (step 3 and step 4)

# 7.- Save the program Indexing\_files2.py.

# 8.- Execute Indexing\_files2.py in your terminal (python3 Indexing\_files2.py).

# 9.- The program will display the poem title and the file in which it exists.

# 10.- Then the user will need to write a promt to look at the indexed files when it appears the following message “Prompt to search: “, the user can give another prompt as soon as the result is displayed, to exit the prompt search type “Exit”.

# 11.- The program displays the results with the following data:

## - Poem: Gloria

## - The score: 12.59482666549441

## - The rank: 0

## - The document number: 55 (this is de documen number in the index)

## -Terms matched in hit [('content', b'lo'), ('content', b'bien'), ('content', b'creo')]

Note: The program searches for the words that appear at every poem so the prompt words can appear in many poems.

Example:

Prompt to search: Bien lo creo

Here are the results

Bien lo creo

Terms matched in the results {('content', b'bien'), ('content', b'creo'), ('content', b'lo')}

Poem: Gloria

The score: 12.59482666549441

The rank: 0

The document number: 55

Terms matched in hit [('content', b'lo'), ('content', b'bien'), ('content', b'creo')]

Poem: Las Plegarias.

The score: 10.25645388374799

The rank: 1

The document number: 57

Terms matched in hit [('content', b'lo'), ('content', b'bien'), ('content', b'creo')]