

‘dram-and-base’ documentation

Patrick Hetco

Urs Zaberer

Joshua Rivera

1 Querying functions

- `query(query: str)` Queries the database using SQL syntax.
Output: tuple of `number_of_results`, `results`
- `similarity(drama1: str, drama2: str)`
Calculates the similarity between two dramas as normalized vectors.
The similarity is based on euclidean distance.
- `top_k_most_similar(drama: str, k=5)`
Takes in drama by title (if unique) and calculates similarities to all other dramas.
Takes top k (choosable, default 5) and outputs their titles + IDs and similarity scores as list of tuples.

2 Tables of the database

2.1 dramas

Basic information

- `title`: text
- `subtitle`: text
- `author`: text
- `year`: text
- `genre`: text
- `setting`: text

Statistics

- `num_scenes`: integer
- `num_lines`: integer
- `num_stage_dirs`: integer
- `num_characters`: integer
- `longest_dialogue`: integer
- `shortest_dialogue`: integer

Emotion values Freude, Leid, Ärger, Verehrung, Liebe, Angst, Abscheu

2.2 characters

Basic information

- `id`: text
- `name`: text
- `drama`: text
- `drama_id`: text
- `gender`: text
- `relations`: text
- `dead`: text

Statistics

- `num_lines`: integer
- `num_scenes`: integer
- `percentage_lines`: real
- `main_char`: text
- `len_longest_line`: integer
- `len_shortest_line`: integer