

Code Documentation

1. app.py

Module Contents

This is the main file to run includes the GUI interface to interact with the the index.

| | |
|---------------|--|
| class | UI_MainWindow(object) |
| method | setupUi(self, MainWindow) Sets up the UI interface of the main window. |
| method | retranslateUi(self, MainWindow) Renames the Object Names |
| class | MainWindow(QtWidgets.QMainWindow) |
| var | scene Graphics Scene where the graphic items are staged in the graphics view. We can add QGraphicsItems to it. |
| var | data The variable stores a pandas dataframe of the original song data to display results. |
| var | ind The variable stores the constructed index |
| method | __init__(self) Initializes the MainWindow of GUI |
| method | extractInfo(self, Results) Returns result with required song_info as a pandas dataframe |
| method | viewResults(self, res, cluster, Results) Prints Relevant results to the screen |

method **search(self)**

The function implements the custom search function with appropriate ranking and returns ordered song_ids and clusters where the query was matched

method **home(self)**

Sets up the homepage of the GUI

Dependencies :- PyQt5 (GUI), Pandas (Data Manipulation)

2. spellcheck — correction of typos

1.11 Module Contents:

The spellcheck module defines the following functions:

spellcheck.all_words(text)

Returns a string in lower case consisting of only words and deleting all other characters from the given input string

spellcheck.corrected(w)

Returns the most likely correct word for the given typo word 'w'

spellcheck.e_dist1(w)

Returns the set of all possible strings which have an edit distance of 1 from the given string 'w'

spellcheck.e_dist2(w)

Returns the set of all possible strings which have an edit distance of 2 from the given string 'w'

spellcheck.possibilities(w)

Returns the list of all possible words that were closely related to the typo word 'w'.

spellcheck.present(w)

Returns the set of words in the given list of strings 'w' and also in the DICT dictionary

spellcheck.Prob(w, N=sum(DICT.values()))

Returns the probability of the word w occurring in the given data file 'DICT'

1.12 Dependencies:

This module requires the modules 're' and 'collections'

1.2 soundex — indexing sounds by phonetics

2.11 Module Contents:

The soundex module defines the following functions:

soundex.index(str_input)

Returns a string denoting the soundex value of the given string_inp

soundex.similar_words(index)

Returns a list of all strings having the soundex value equal to 'index'

soundex.soundex_list(str_input)

Returns a list of all strings having the soundex value equal to the given string

2.12 Dependencies:

This module requires the modules 're', 'json' and 'collections'

Module : indexer.py

Builds the index for the retrieval engine. Employs soundex.py . Dependencies include nltk, re, and json packages.

Module : search.py

Takes a phrase query and returns the appropriate DocIDs and positions of query tokens.

soundex_check(tok):

Utility function to check if user has input an escape sequence to run all soundex versions for the token.

add_trial(l, pos1, pos2):

Function to group multiple pair-wise positional index searches. Is called by pos_intersect.

pos_intersect(p1, p2, k):

Positional index intersection using the algorithm described in class.