Lazy song

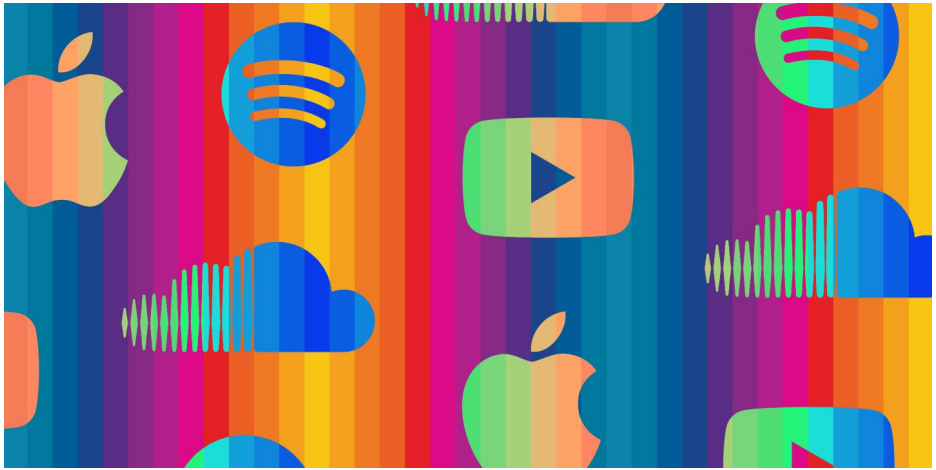
progetto per il corso di Gestione dell'Informazione
dell'università di Modena e Reggio Emilia anno 2020/2021

[Lorenzo Stigliano e Alessandra Bonaccorso]

Indice

1. idea
2. raccolta documenti
3. preprocessing e algoritmo di indexing
4. search engine e gui
5. esempi di query di ricerca
6. benchmark (grafici e valutazioni)
7. conclusioni
8. fine

Idea di base



Come abbiamo appreso dal corso di gestione dell’informazione, i dati intorno a noi si materializzano in varie forme ed in modi differenti.

Per questo progetto ci è stato chiesto di realizzare un search engine tematico. Abbiamo scelto la musica, perché è qualcosa che accomuna tutti e che ci circonda ogni giorno.

Nel nostro search engine sono presenti i testi di vari artisti ed è possibile ricercare un termine ed avere come risultato canzoni che contengono quel termine (o simili).

Raccolta del materiale

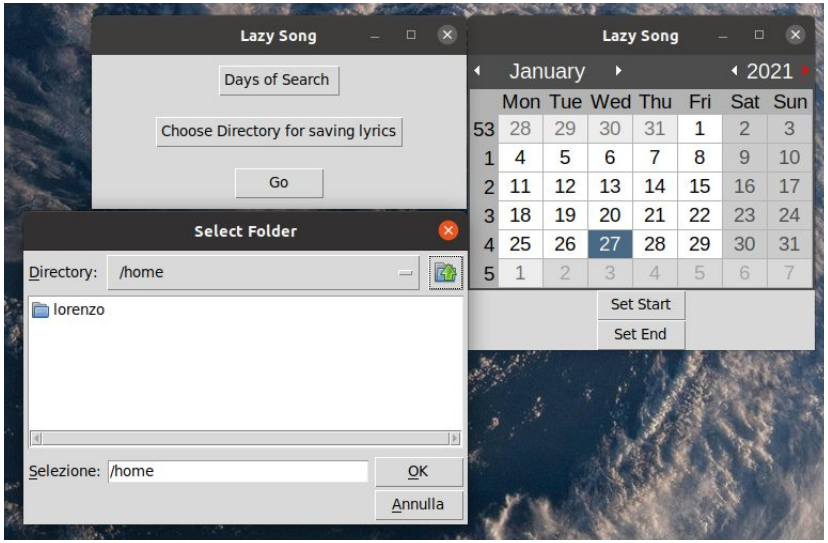
Il vero programmatore/informatico è Lazy! Dunque non ci sembrava corretto scaricare tramite copia ed incolla ogni canzone manualmente.

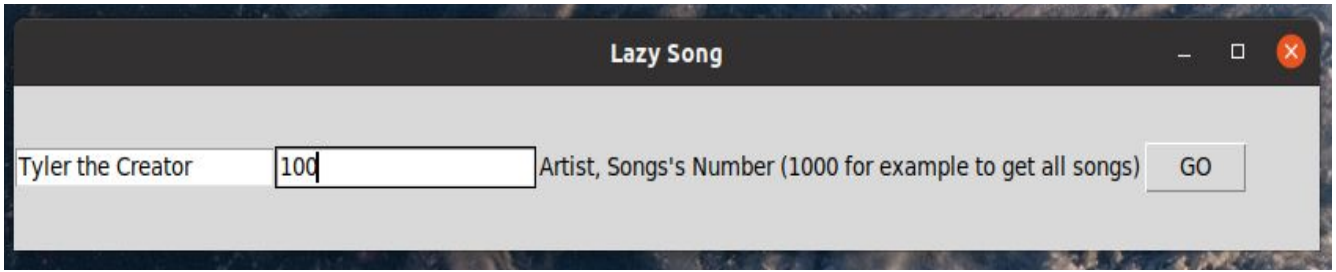
Una rapida ricerca su internet ci ha permesso di scoprire il pacchetto per python GeniusLyrics, che dato in pasto il nome di un artista scarica i testi delle sue canzoni dal sito più famoso di lyrics, Genius.com.

GENIUS

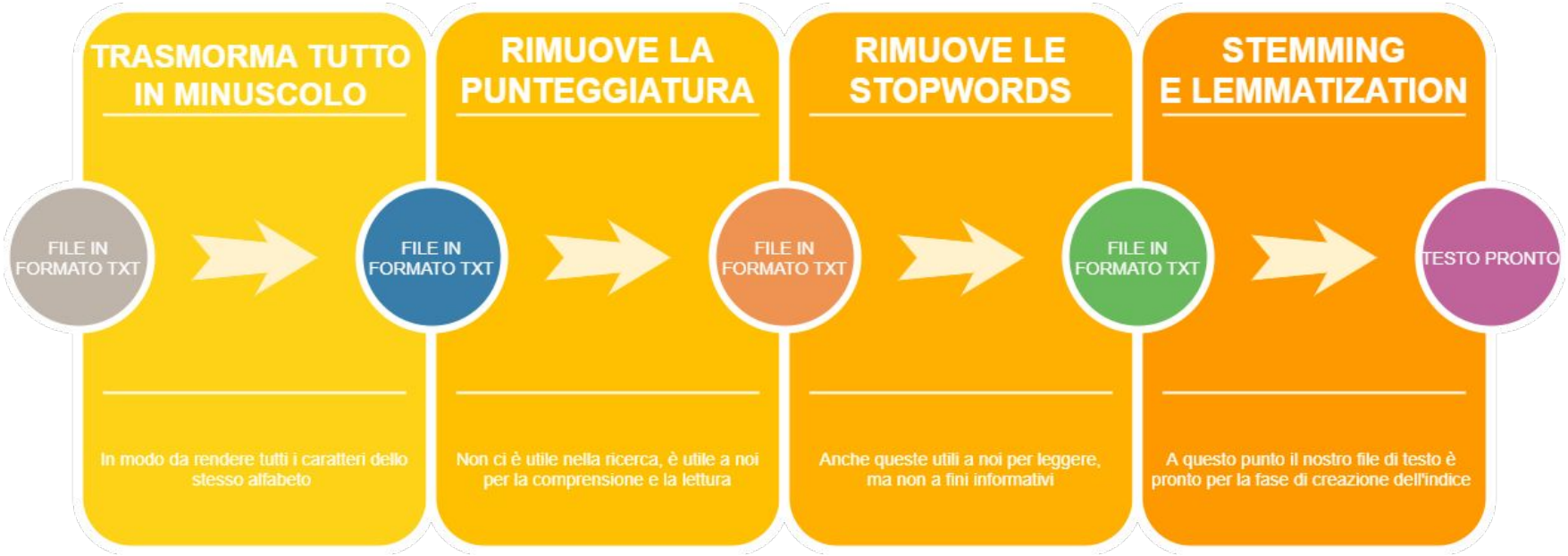
Per scegliere gli artisti abbiamo selezionato quelli che hanno raggiunto le classifiche negli ultimi anni. Di ognuno abbiamo selezionato la top ten delle sue canzoni, in modo da avere una selezione più omogenea.

Screenshoots Dump

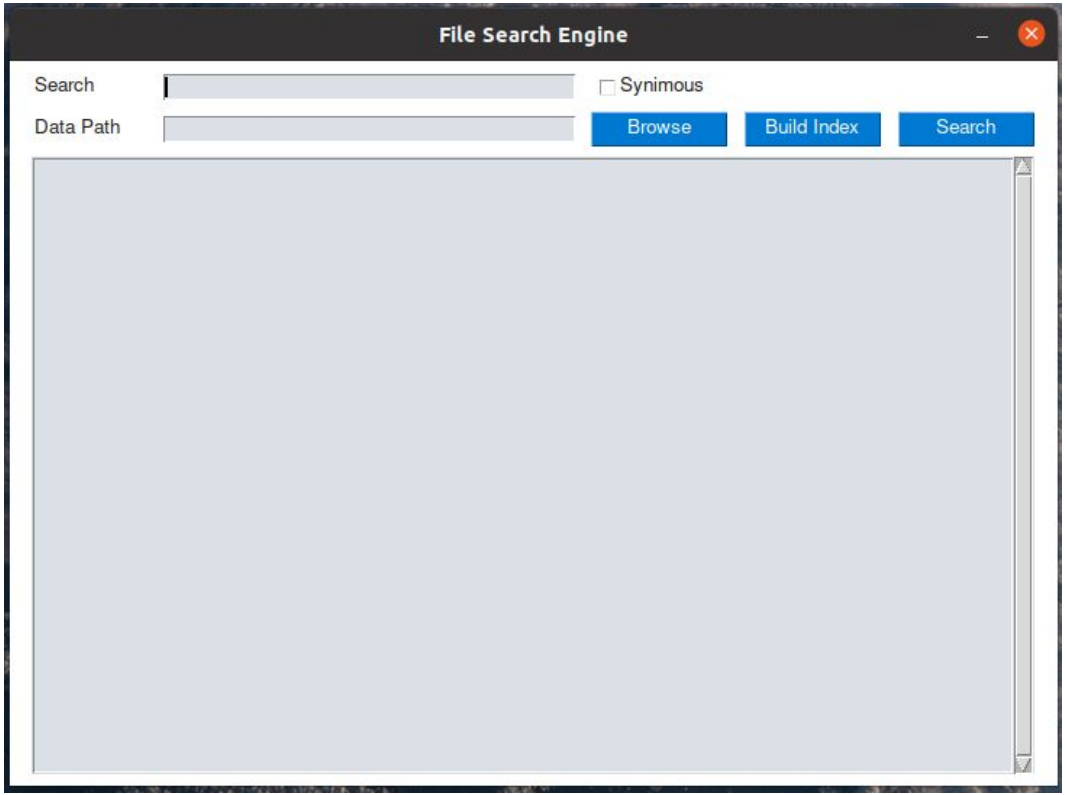




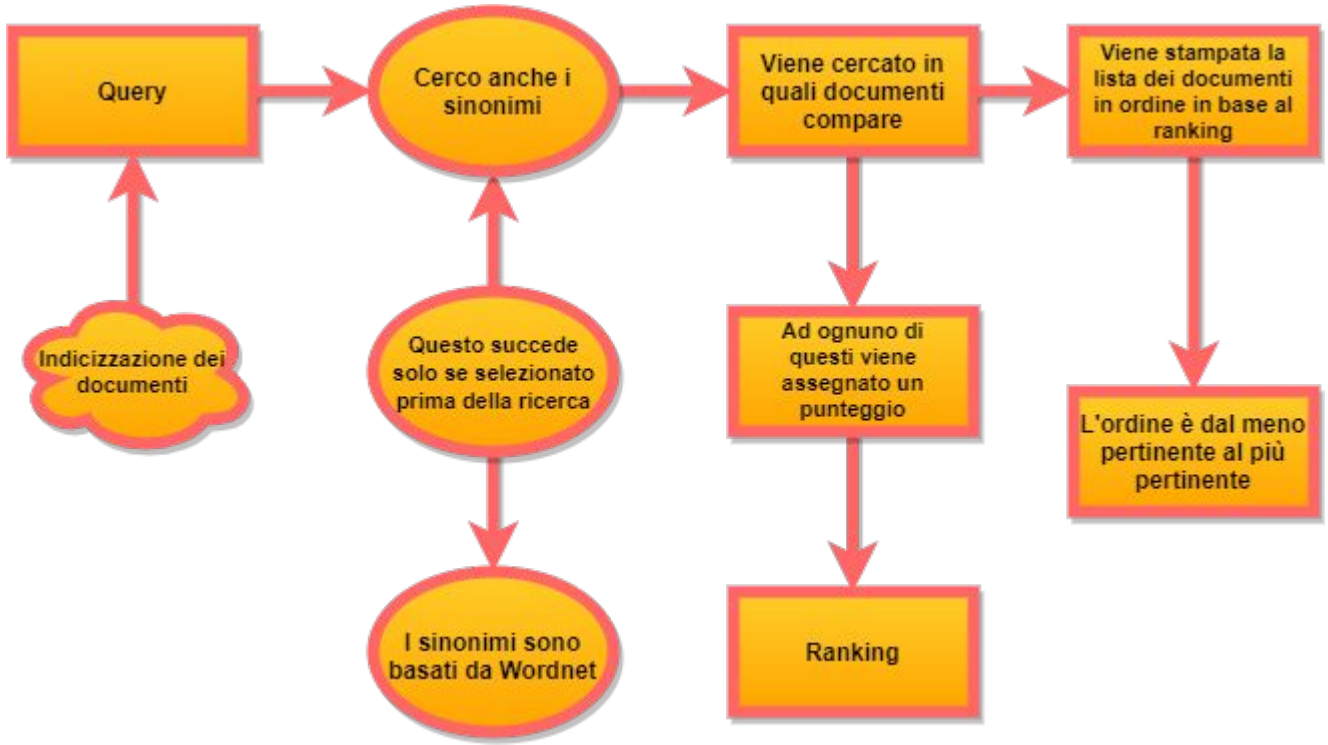
Pre-processing



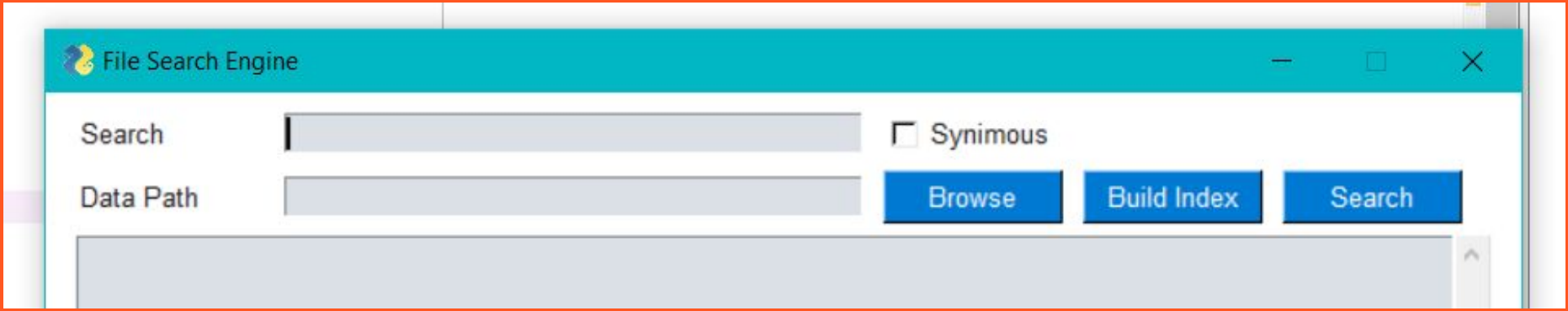
Search Engine



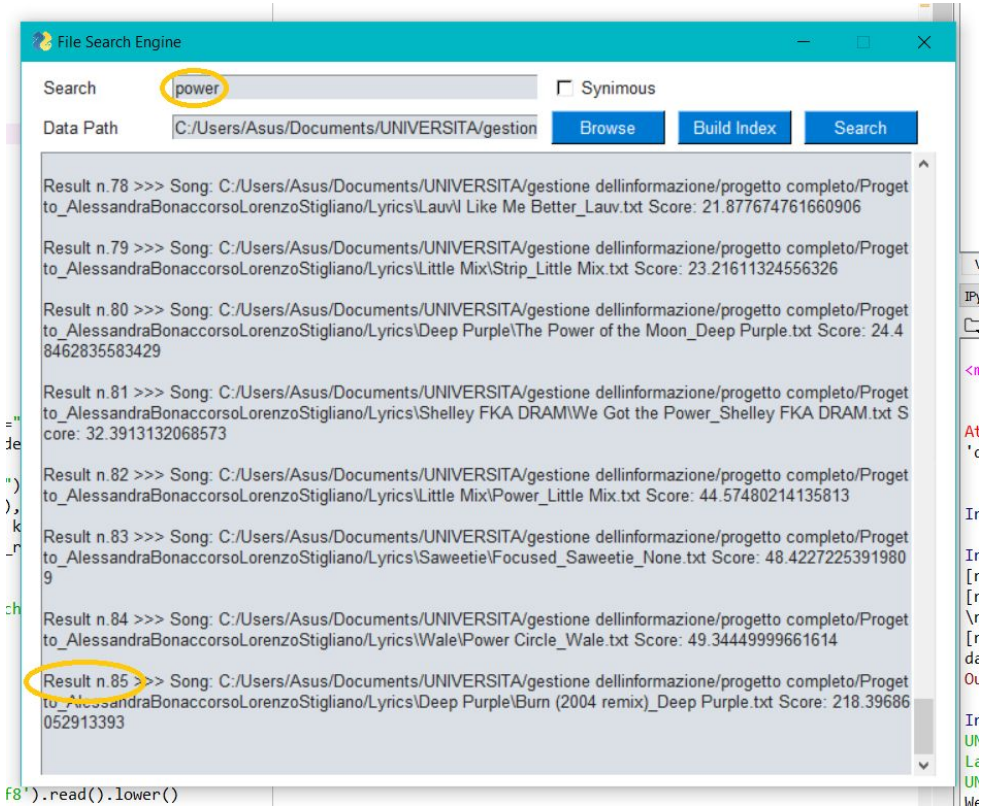
Search engine



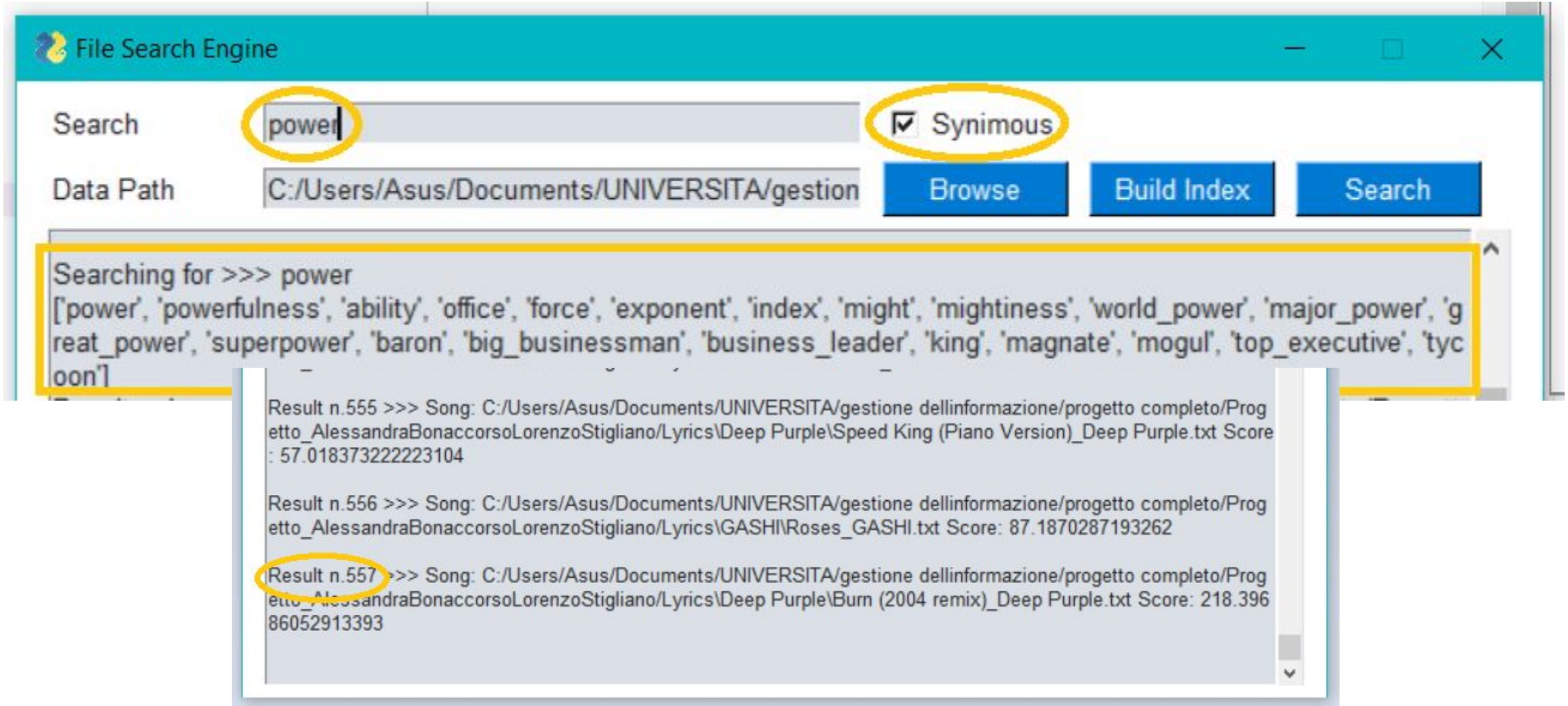
Esempi di query



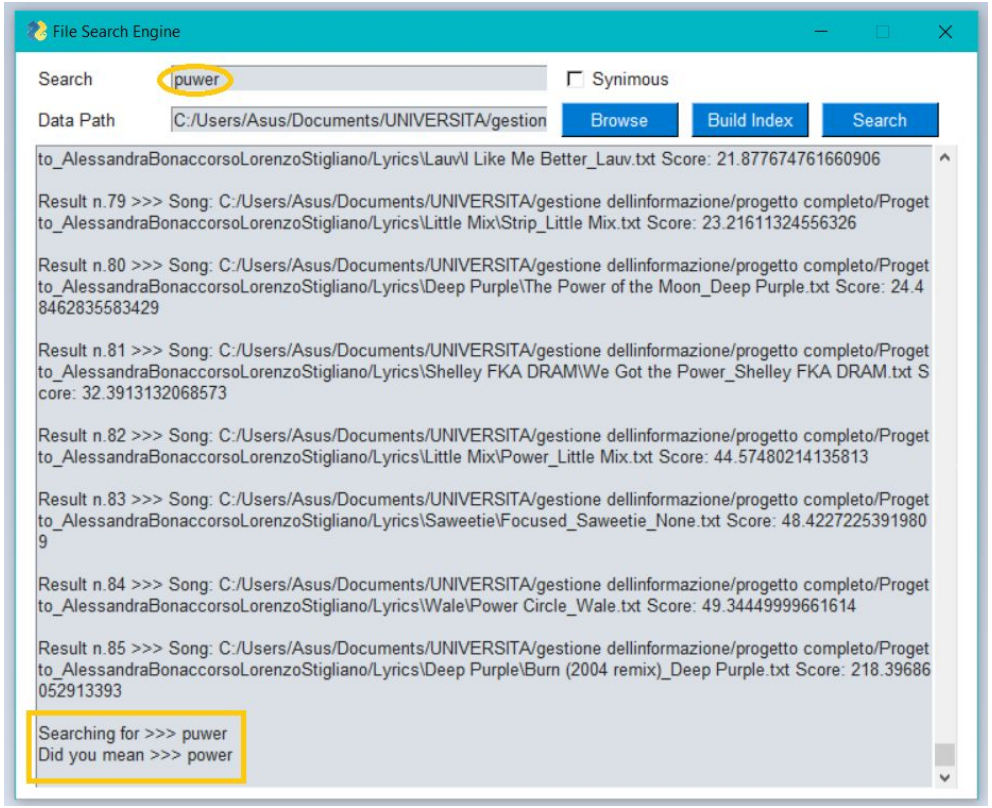
Ricerca della parola “power”



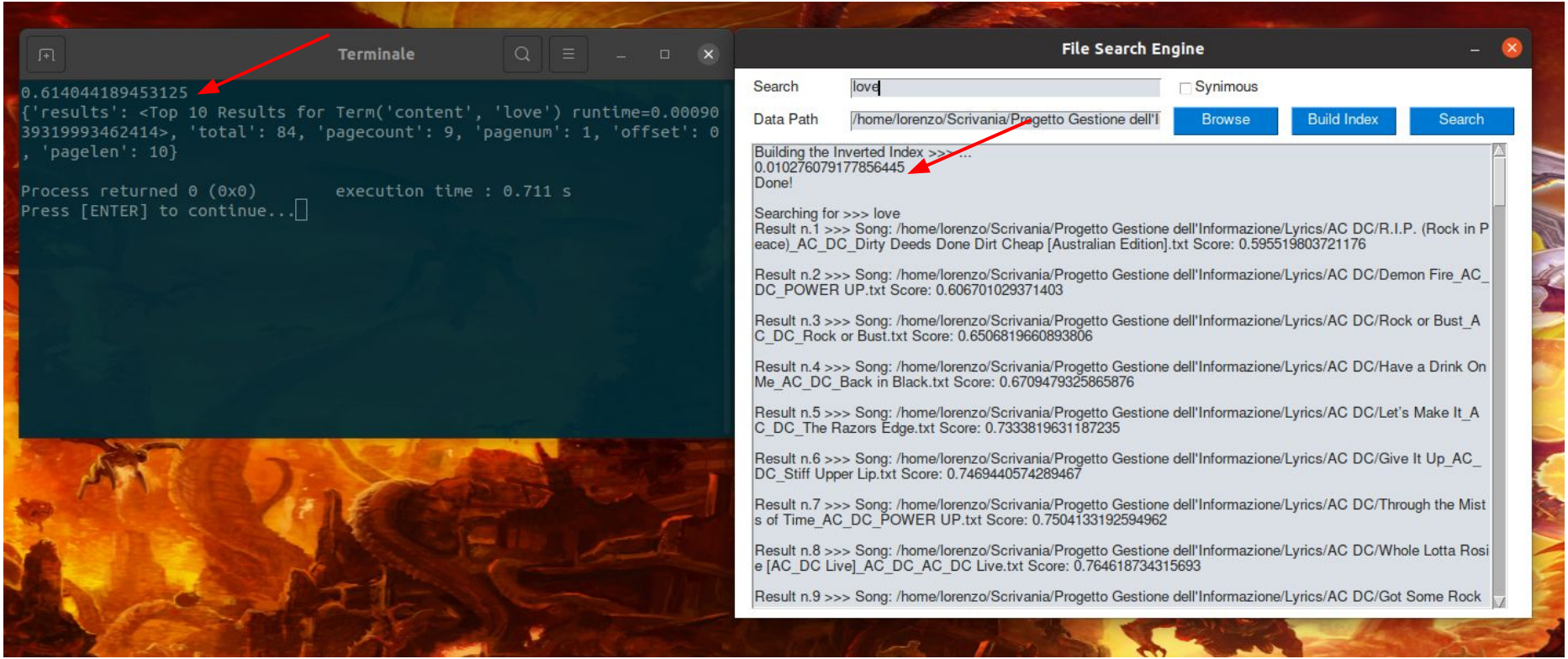
Ricerca con sinonimi di “power”



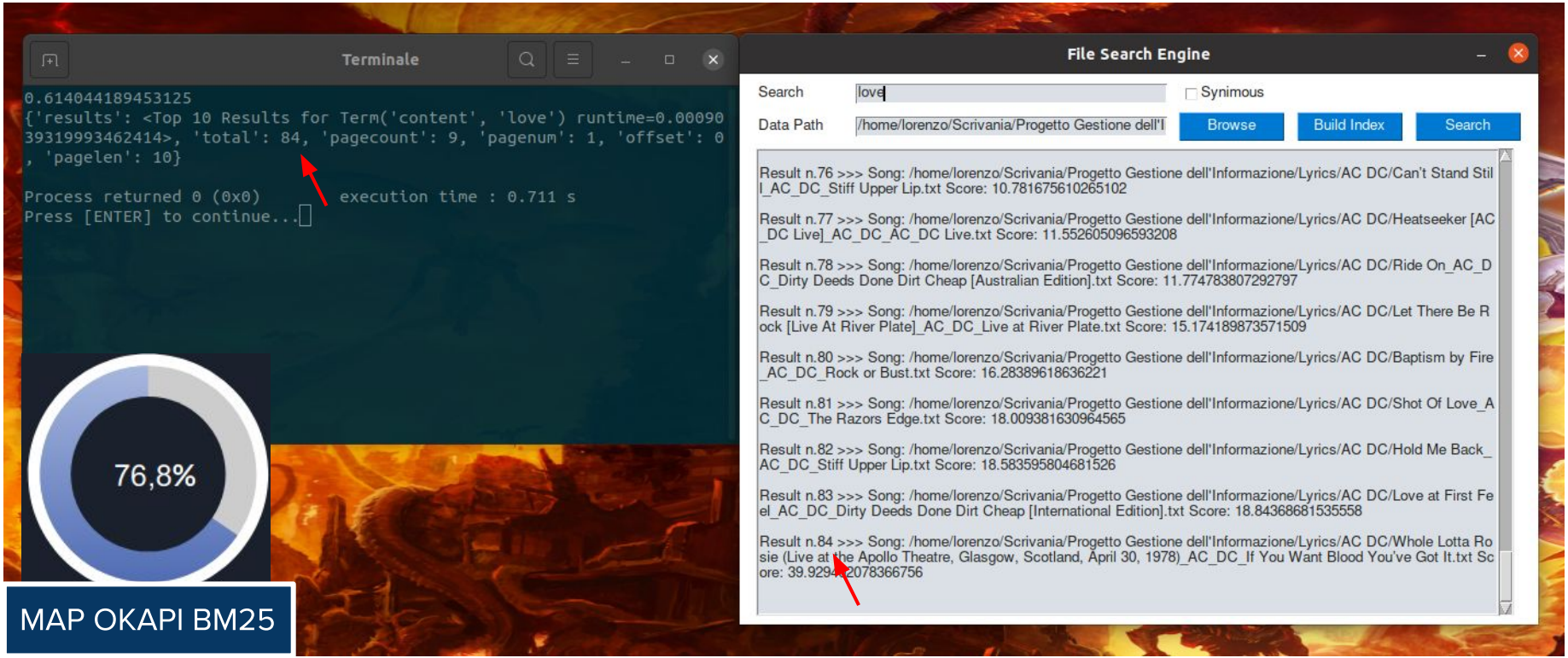
In caso di errore di spelling



Screenshoot benchmark



Screenshoot Benchmark



Conclusioni

Soluzioni alternative implementate:

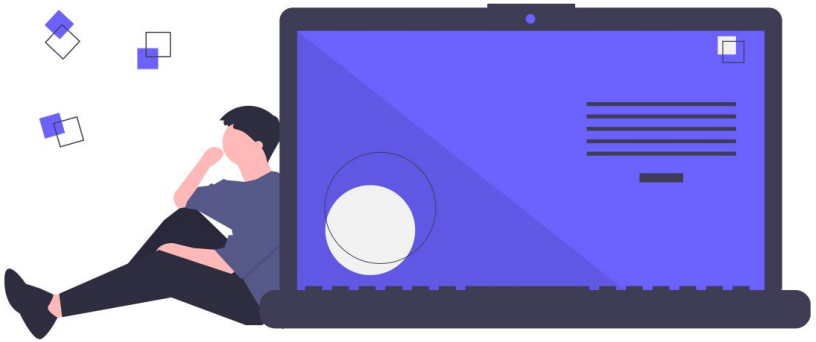
1. Modello booleano
2. Indexing (non utilizzando le librerie di whoosh e lucene)

Cosa avremmo voluto implementare:

- ricerca per emozioni/mood
- ricerca per generi

Maggiori difficoltà:

- download dump
- problemi di encoding a seconda del sistema operativo
- didattica a distanza
- lavoro di gruppo a distanza



Fine