

Súvislosť textov piesní pomocou NLP

- Naučila som sa o NLP (anglické „Natural Language Processing“), o modeloch a rôznych úlohách, ktoré sa riešia. Taktiež som niektoré skúsila.
- Naštudovala som si úlohu „sentence similarity“, ktorá sa zaoberá s podobnosťou textov
- Nájdenie datasetu, zdrojov údajov, ktoré v práci využívam. Stiahla som ho zo stránky kaggle. Obsahuje piesne a údaje o nich. Pre mňa sú podstatné stĺpce „title“, „artist“, ktoré opisujú meno autora a piesne, „tag“, opisujúci žáner a „lyrics“ so samotnými textami piesní.
- Na stránke Hugging Face je model, ktorý som v projekte použila.

Postup:

- Najprv som stiahla dáta. Následne som ich zredukovala a upravila tak, že som ich rozdelila podľa žánrov. Pre každý žáner som pracovala s 500 piesňami.

<https://www.kaggle.com/datasets/carlosgdj/genius-song-lyrics-with-language-information>

- Pre každý žáner som spracovala všetky texty pomocou modelu „**sentence-transformers/all-MiniLM-L6-v2**“ a získala som tak vektory, ktoré ich reprezentujú. Z nich sa pomocou nástroja *util* vypočítali hodnoty podobností. Z týchto som vypočítala priemery.

<https://huggingface.co/tasks/sentence-similarity#:~:text=Semantic%20Textual%20Similarity%20is%20the,a%20list%20of%20similarity%20scores.>

- Pre každý žáner mám priemery súvislostí textov v nich. Teraz porovnávam medzi žánrami, že ktoré obsahujú najviac podobné texty.

Výsledky:

- Pre môj obmedzený dataset vyšlo, že v žánri „rap“ je najväčšia sémantická podobnosť, hodnota 47,47
- Žáner „rock“ obsahuje najmenej podobné texty, s hodnotou 35,62
- Dataset obsahoval kategóriu „misc“, predstavujúcu viac žánrov. Pre túto kategóriu vyšlo, že podobnosť je len 26,52, čo nasvedčuje tomu, že texty v jednom žánri majú určitú sémantickú podobnosť

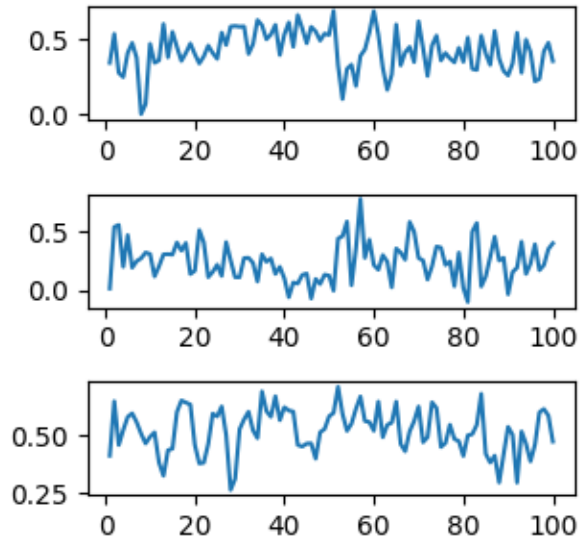
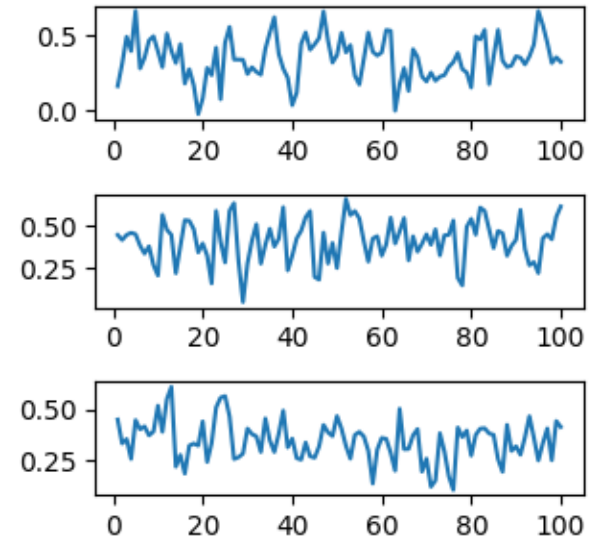
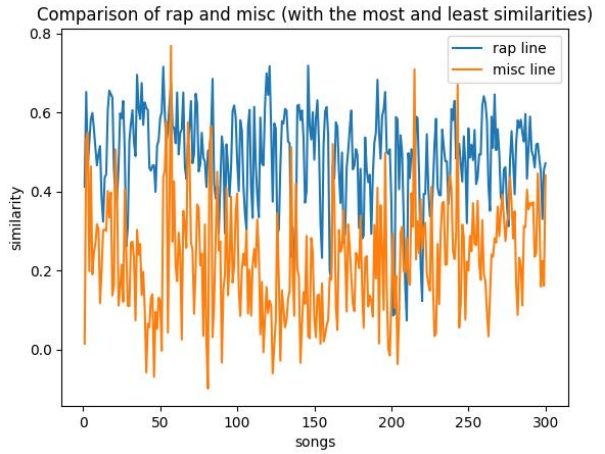
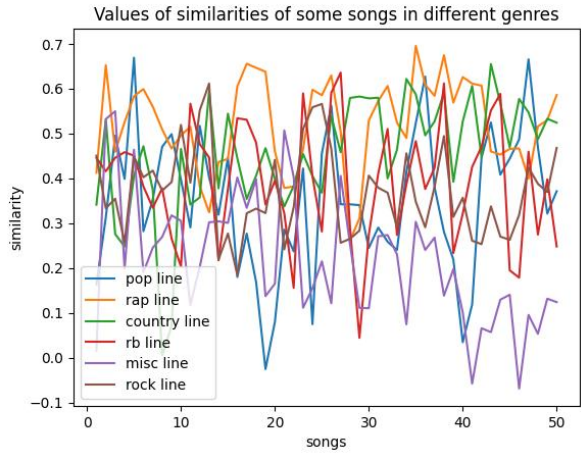
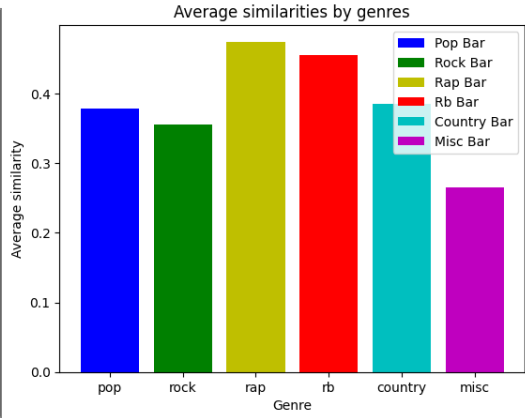
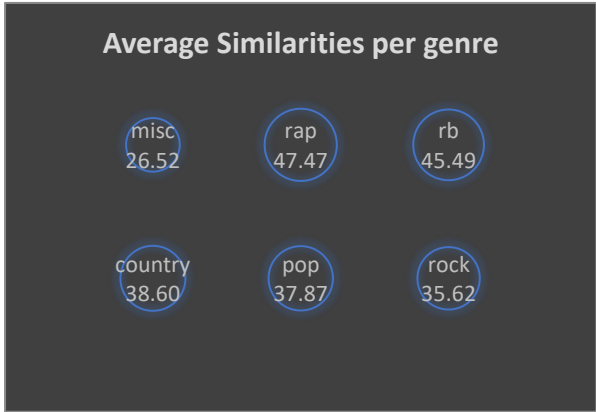
Nástroje:

- Jazyk Python
- Moduly numpy, scipy, torch, kaggle, sentence-transformers

```
pip install numpy scipy torch kaggle
pip install -U sentence-transformers
```

- Dataset zo stránky <https://www.kaggle.com/datasets/carlosgdj/genius-song-lyrics-with-language-information> alebo stiahnuť pomocou

```
!kaggle datasets download -d carlosgdj/genius-song-lyrics-with-language-information
```



LABELS: Pop
Rb
Rock

Country
Misc
Rap