

Hľadanie podobností v textoch ľudových piesní

Cieľom práce bolo implementovať porovnávanie ľudových piesní tak, aby sme vedeli povedať, že takéto texty sú **tá istá pieseň**:

Dzifče počarovne,
ňezaľup še do mňe
ja chlopec
vandrovni, co ci
budze zo mňe

Dzifče počarovne,
nezaľub se do mne
ja chlapec
vandrovny, co ci
budze zo mňe

Dzivče počarovne,
nepopatraj na mne
ja chlapec
vandrovny, co ci
budze zo mne

Dôvody sú napríklad jednotný správny zápis v databáze,
výskum regionálnych rozdielov a iné

Michal Mižák

Vedúci práce: doc. RNDr. Stanislav Krajči PhD



Femplin

- ❖ Texty sme postupne
 - ✓ zozbierali a upravili
 - ✓ porovnali podľa štandardného *vector space* algoritmu
- ❖ Tento algoritmus sme vylepšili napr. o
 - ✓ kontextuálnu podobnosť slov
 - ✓ zgrupovanie slov do *tried ekvivalencie*
 - ✓ rozdelenie do fáz ako napr. fáza váženía pojmov alebo fáza porovnávania slov
 - ✓ možnosť jednotlivo konfigurovať všetky fázy algoritmu
- ❖ Vytvorili sme priestor pre
 - ✓ skúmanie efektivity jednotlivých konfigurácií algoritmu
 - ✓ odbornú analýzu a zjednocovanie databázy textov

❖ <https://github.com/MichalMizak>



Michal Mižák

Vedúci práce: doc. RNDr. Stanislav Krajči, PhD.

