



**FAKULTA
INFORMAČNÍCH
TECHNOLOGIÍ
ČVUT V PRAZE**

Zadání bakalářské práce

Název: Dvojjazyčné vyhledávání v dokumentech
Student: Lukáš Rynt
Vedoucí: Ing. David Bernhauer
Studijní program: Informatika
Obor / specializace: Znalostní inženýrství
Katedra: Katedra aplikované matematiky
Platnost zadání: do konce letního semestru 2022/2023

Pokyny pro vypracování

Fulltextové vyhledávání v dokumentech naráží na problémy s použitím synonym nebo s použitím různých jazyků.

Popište základní algoritmy pro indexaci a vyhledávání v textových dokumentech (minimálně booleovský a vektorový model).

Popište princip word embedding algoritmů obecně a dále se zaměřte na Word2Vec.

Proveďte rešerši některých přístupů pro vícejazyčné použití Word2Vec.

Využijte vícejazyčného znění zákonů (např. nový občanský zákoník, právní akty EU, ...) pro natrénování dvojjazyčného Word2Vec (čeština a angličtina).

Navrhněte a implementujte prototyp jednoduchého webového vyhledávače v dokumentech na základě natrénovaného dvojjazyčného Word2Vec modelu.

Elektronicky schválil/a Ing. Karel Klouda, Ph.D. dne 19. listopadu 2021 v Praze.