

ZADÁNÍ DP/BP PRACÍ

od společnosti CCA Group a.s.

Bakalářské práce

Fulltextové nástroje pro vyhledávání ve velkých objemech dat

(BIK)

Počítačové systémy jsou dnes schopné pojmout velké objemy dat. Databáze zákazníků běžně obsahují stovky tisíc dokumentů, souhrnné velikosti dat se pohybují ve stovkách gigabytů. Zákazník přitom standardně požaduje vyhledávání v metadatech dokumentu i v jeho textu. Očekává přitom odezvy v řádu jednotek vteřin.

Cílem práce bude vybrat vhodný nástroj pro fulltextové vyhledávání, otestovat jej na velkém množství dokumentů a ukázat integraci vyhledávače v existující aplikaci zadavatele.

Zadání:

1. Seznamte se s problematikou fulltextu
2. Zvolte parametry, podle kterých lze porovnat fulltextové nástroje podle výkonu
3. Proveďte porovnání vybraných nástrojů
4. Udělejte jednoduchou implementaci napojení nástroje na existující aplikace pro správu dokumentů