

Posudek oponenta bakalářské práce

Autor práce: **Stanislav Škovran**

Název práce: **Výběr informací z lékařských zpráv**

Student v rámci své bakalářské práce vytvořil dva programy. Jeden anonymizuje jména v lékařských zprávách a druhý v nich vyhledává zadaná klíčová slova.

Práce byla výrazně vylepšena a zadání splňuje. Po formální stránce je v pořádku, literatura je dostatečná a relevantní. Bohužel se nepodařilo dostáhnout požadované kvality při použití dostupných NLP nástrojů a vytvořilo se vlastní řešení. Ačkoli metoda je z pohledu analýzy textů jednoduchá, řešení i rozsah textu již odpovídá bakalářské práci.

Autor vytvořil vlastní rozpoznávání pojmenovaných entit, bohužel bez jeho vyhodnocení nejde posoudit jeho kvalitu. Neuvažování kontextu a víceznačnost mohou způsobit chyby při vyhledávání lékařských pojmů, což by si také zasloužilo vyhodnocení.

Databáze je v pořádku, jednoduchá prezentační vrstva splňuje základní potřeby vyhledávání.

K práci mám následující dotazy:

1. Anonymizátor jsem zkoušel na následujícím textu:

ČSSD se musí omluvit Johnovi za výroky o tunelování dopisem i na webu. Bývalý ministr vnitra Radek John částečně uspěl s žalobou na sociální demokraty.

výstup:

ČSSD se musí omluvit Adam001 za výroky o tunelování dopisem i na webu. Bývalý ministr vnitra Adam002 Adam001 částečně uspěl s žalobou na sociální demokraty.

Jak náročné by bylo při dané implementaci zjistit, že se jedná o stejnou entitu a správně určit křestní jméno/příjmení?

2. Souvisí s příkladem: Jak náročné by bylo dodržovat koncovky i v případě anonymizovaného textu? (Adam001 -> Adamovi001)
3. Výkonnost anonymizátoru: 36k zpráv za 2h – zdá se mi trochu pomalé. Co je zde úzké hrdlo?

Navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni 20. 5. 2016

Doc. Ing. Josef Steinberger, Ph.D.

