

Posudek oponenta bakalářské práce

Název: Identifikace parafráze
Autor: Duc Vuong Tran
Oponent: Ing. Michal Konkol, Ph.D.

Cíl práce

Cílem práce bylo prozkoumat technologie pro identifikaci parafráze a na základě tohoto průzkumu vytvořit a otestovat systém pro identifikaci parafráze.

Text práce

Se strukturou práce nejsem spokojen. Práce postrádá globální pohled na problém, který by nastínil důležitost a smysl jednotlivých kapitol. Působí proto útržkovitým dojmem. Dále je nevhodně kombinována teorie a implementace (např. kapitola 2 „Použité knihovny a algoritmy“). Je obtížné se v práci orientovat, protože autor nepoužívá číslování kapitol. Důležitost kapitoly je pak dána pouze velikostí písma. V obsahu jsou sekce naopak číslovány, což je nekonzistentní. Navíc kapitola 5 má v obsahu chybné číslování.

Nyní k obsahu práce. Autor sice popisuje základní aspekty a stavební kameny svého řešení, ale často nedostatečně (málo informací, zmatenost, faktické chyby). Bodu zadání „Seznamte se s metodami pro identifikaci parafráze“ se v textu autor prakticky nevěnuje. Autor si v podstatě zvolil jednu metodu řešení, ale žádná další metoda není zmíněna (až na porovnání v závěru práce). Očekával bych popis několika nejlepších systémů, které se daný problém snaží řešit nebo alespoň výčtem alternativ. Zvolená metoda navíc není dostatečně popsána.

V abstraktu práce autor píše „Cílem této práce je seznámit se s algoritmy používané v oblasti strojového překladu.“, přičemž cílem práce je (jak napovídá název práce) identifikace parafrází.

Práce je psaná špatnou češtinou. Některé části jsou špatně srozumitelné. Přestože nejsem zastáncem počítání stránek, práce se mi zdá neopodstatněně krátká. V práci je velké množství nevyužitého místa (např. str. 20-22).

Software

Na CD se nachází maven projekt `paraphrase-parent` s třemi moduly `paraphrase`, `paraphrase-engine` a `paraphrase-fileparser`. Nevím, co autora vedlo k rozdělení aplikace do několika modulů. Vzhledem k její velikosti mi to připadá zbytečné. Modul `paraphrase-fileparser` má jednu třídu a jedno rozhraní a modul `paraphrase-engine` má 5 tříd.

Nikde jsem nenašel uživatelskou dokumentaci ani `README` soubor. Při standardním sestavení pomocí nástroje Maven dojde k chybě (chybějící knihovny). Kód je částečně komentovaný.

Závěr

Student dosáhl velmi dobrých výsledků v úloze identifikace parafráze a vytvořený software považuji za dobrý. Text práce je ovšem velmi špatný a dle mého názoru nedosahuje minimální úrovně nutné pro obhájení bakalářské práce. Navrhuji hodnocení známkou **nevyhověl** a práci nedoporučuji k obhajobě.

V Plzni dne 12. 8. 2016



Ing. Michal Konkol, Ph.D.