

Posudek vedoucího diplomové práce

Lemmatizér českého jazyka

Diplomant: Bc. Ivo Snoza

Vedoucí diplomové práce: RNDr. Miroslav Benedikovič,
Katedra softvérových technologií, FEI, Univerzita Pardubice

Cílem diplomové práce bylo vytvoření systému pro automatickou lemmatizaci a anotaci slov českého jazyka a vytvoření softvérové aplikace umožňující převod vstupního textu v českém jazyce na lemmatizovaný text. Aplikace má umožňovat definování a editaci informací ve vytvořené databázi slov tak, aby mohla být automaticky zpracovatelná vhodným softvérovým vybavením (statistické informace, kategorizace článků, apod. ...).

1 - Jaké metody diplomant ve své práci uplatnil?

Diplomant uplatnil, kromě znalostí programátorských technik, znalosti získané z předmětů oblasti datových struktur a kompilátorů. Protože úlohou vytvořené aplikace je analýza živého jazyka, diplomant byl nucen podrobněji nastudovat i pravidla českého jazyka za účelem vytvoření algoritmů popisujících gramatickou i významovou stránku analyzovaného textu.

2- Co diplomant při vypracování své DP vytvořil?

V úvodu se zaměřil na stručný popis několika příkladů z historie komunikace člověka a počítače na úrovni přirozených jazyků (ELIZA). Ve druhé kapitole vymezil základní pojmy z oblasti přirozených a umělých jazyků, které jsou využity v další části práce.

Jádrum diplomové práce jsou následující kapitoly, které se zabývají procesem počítačového zpracování přirozeného jazyka. V první etapě ve fázi lexikální analýzy je nutno provést tokenizaci textu a nalezení tzv. stemmat. Diplomant se v této části věnuje zejména problémům při extrahování stemmat v českém jazyce. Čtvrtá kapitola je zaměřena na morfologickou stránku, kdy je zjišťován slovní druh stemmatu za účelem uložení příslušné informace do databáze.

Pátá kapitola se týká popisu návrhu databáze a samotné aplikace lemmatizátoru. Diplomant vhodně zvolil databázový systém Oracle a pro vývoj aplikace jazyk C# na platformě .NET Framework 4.0. Přirozený jazyk je příliš složitý na plně automatizované zpracování, proto diplomant řešil tento problém kombinací automatického a interaktivního zpracování vstupního textu. V případě nejednoznačného řešení je nucen uživatel ovlivnit výběrem a ručním potvrzením správné možnosti před uložením relevantní informace do databáze. Databázi je možno vyexportovat ve formátu XML.

Součástí diplomové práce je medium CD se zdrojovým kódem vytvořené aplikace. Příloha obsahuje „Uživatelskou příručku“ a „Popis struktury“.

3 - Jak diplomant prokázal správnost navrženého řešení problému?

Diplomant vytvořil funkční aplikaci lematizéru vytvářející požadovanou databázi. Zdrojový kód aplikace je sestaven z malých modulů popsaných komentáři. Relativně přehledně volené moduly umožňují snadnou úpravu zdrojových textů podle požadavků dalšího vývoje aplikace. K tomu následující poznámka: Některé malé moduly tyto komentáře neobsahují, případně jsou stručnější. Doporučoval bych nešetřit komentáři i v tom případě, když se to může zdát zbytečné.

4 - Podařilo se diplomantovi splnit cíle práce, které mu byly uloženy?

Z pohledu složitosti problému analýzy přirozeného jazyka podle mého názoru cíly zadání diplomant splnil.

5 - Jaká kvalita textu diplomové práce z hlediska jeho struktury, srozumitelnosti ... ?

Práce je psaná srozumitelným způsobem, vyznačuje se v mnoha případech přílišnou stručností.

6 - Které nejasnosti vyskytující se v DP by měl diplomant objasnit při obhajobě a jaké jsou další připomínky k DP?

Ze závěru práce je naznačena úvaha o dalším vývoji této aplikace. Předkládám k diskusi následující připomínky:

- Jak si představujete další využitelnost databáze vytvořené aplikace?
- Pokuste se popsat z hlediska pohledu autora stupeň uživatelské přívětivosti vytvořené aplikace a možnosti vylepšení její komunikativnosti s uživatelem.
- V čem jsou, podle vašeho názoru, možnosti dalšího vývoje této aplikace (hromadný import dat a podobně...).

7 - Doporučení a hodnocení klasifikačním stupněm.

Předloženou práci doporučuji k obhajobě a navrhuji hodnotit ji klasifikačním stupněm

výborně mínus (1-)

Pardubice, 2. 9. 2011

RNDr. Miroslav Benedikovič,
KST - FEI, Univerzita Pardubice