

Personalizovaná sumarizácia textu

Róbert Móra

Vedúca práce: prof. Mária Bieliková

Motivácia

- problém preťaženia informáciami
- konvenčné sumarizačné metódy pracujú len s textom a nezohľadňujú ďalšie užitočné zdroje informácií:
 - charakteristiky a potreby používateľov
 - metadáta (pridané používateľmi alebo doménovými expertmi)

Predspracovanie dokumentu

Konštrukcia personalizovanej matice termov a viet

$$w(t_{ij}) = \sum_k \alpha_k R_k(t_{ij})$$

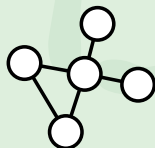
Hodnotič frekvencie a polohy termov



Hodnotič anotácií



Hodnotič relevantných doménových pojmov

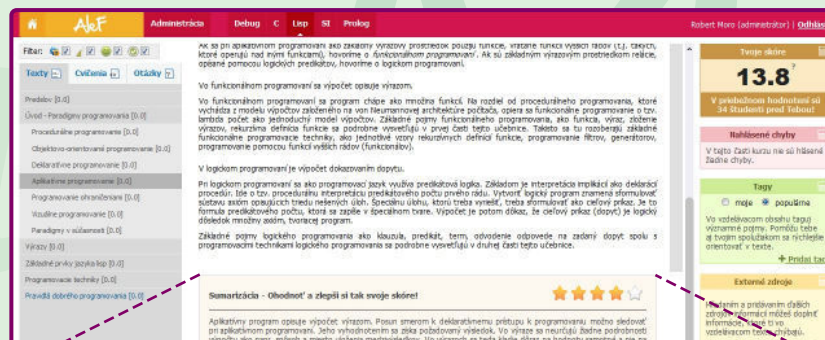


Hodnotič znalostí

SVD (rozklad matice na singulárne hodnoty) a výber viet do súhrnu

Experimentálne overenie

- v doméne výučby (scenár opakovania)
 - 2242 ohodnotení, 385 porovnaní, 479 odpovedí na otázky
- experimenty na dvoch predmetoch - programovanie a tvorba softvéru
 - 75 študentov, 5 expertov, 303 textov



Sumarizácia - Ohodnoť a zlepší si tak svoje skóre!

Aplikatívny program opisuje výpočet výrazom. Posun smerom k deklarativnému prístupu k programovaniu možno sledovať pri aplikatívnom programovaní. Jeho vyhodnotením sa získa požadovaný výsledok. Vo výraze sa neurčujú žiadne podrobnosti výpočtu ako napr. spôsob a miesto uloženia medzivýsledkov. Vo výrazech sa teda kladie dôraz na hodnoty samotné a nie na organizáciu ich uloženia. Vo funkcionálnom programovaní sa program chápe ako množina funkcií. V logickom programovaní je výpočet dokazovaním dopytu. Pri logickom programovaní sa ako programovací jazyk využíva predikátová logika. Základom je interpretácia implikácií ako deklarácií procedúr. Špeciálnu úlohu, ktorú treba vyriešiť, treba sformulovať ako cieľový príkaz. Je to formula predikátového počtu, ktorá sa zapisuje v špeciálnom tvare. Základné pojmy logického programovania ako klauzula, predikát, term, odvodenie odpovede na zadaný dopyt spolu s programovacími technikami logického programovania sa podrobne vysvetľujú v druhej časti tejto učebnice.

Slovné hodnotenie sumarizácie...

Výsledky

- návrh spôsobu kombinácie hodnotičov
 - umožňuje uvažovať rôzne parametre a kontext sumarizácie
- návrh konkrétnych hodnotičov
- doménovo a jazykovo nezávislý prístup
- prezentované na TIR@DEXA 2012 (IEEE)