

Posudek oponenta

Název DP: Automatické zodpovídání dotazů založené na sumarizaci textů

Diplomant: Tárík Saleh Salem

Troufalým cílem práce autora, vzhledem k náročnosti, je usnadnění vyhledávání informací na internetu. Autorovi se, dle mého názoru, správně nelíbí odpovědi vyhledávačů k pokládaným dotazům ve formě seznamu potenciálně relevantních webových stránek, a proto vytváří vyhledávač vlastní, který navíc z potenciálně relevantních webových stránek vytváří extrakt pomocí techniky sumarizace.

K realizaci zadané úlohy autor nastudoval velké množství technik, postupů a principů popsaných v českém i anglickém jazyce. K těmto patří především principy vyhledávacích systémů, nástroje pro extrakci textů z HTML stránek, nástroje pro zpracování textů, různé techniky sumarizace a konkrétní sumarizaci založenou na latentní sémantické analýze využívající singulární rozklad.

Text práce je pěkně čitelný, hezky strukturovaný, plný názorných příkladů, tabulek a obrázků. Přesto autorovi vytýkám několik následujících nedostatků. Především je to velké množství překlepů, kdy již ve stručném úvodu lze nalézt slova jako „prace“ či „impolementovaného“. Také se mi nelíbí absence českého abstraktu k jinak v češtině psané práci za současné přítomnosti abstraktu anglického. Struktura práce a především to, jak spolu jednotlivé části souvisí, nejsou naznačené v úvodu. Na straně 26 autor zaměňuje pojmy „latentní sémantická analýza“ a „latentní sémantické indexování“ bez vysvětlení. Také bych vytknul nejednotný styl referencí na literaturu, kde absentují např. roky vydání.

Přes tyto drobnější nedostatky navrhuji práci k obhajobě se známkou „výborně“. Důvodem je především velká náročnost úlohy (autor mj. použil 40 zdrojů literatury) a vysoká technická úroveň diplomové práce, kdy autor prozkoumal, pěkně popsal, porovnal a vybral vhodné nástroje k řešení úlohy, které následně upravil a zkomponoval do výsledné aplikace pojmenované ASI. Práce je také, dle mého názoru, jednoznačným přínosem do oboru vyhledávání a autor na ní ukázal, že inženýrský titul si více než zaslouží.

Otázka k obhajobě: Navrhněte způsob úpravy ASI tak, aby si poradil s webovými stránkami bez diakritiky.

----- výborně -----

V Plzni, dne 4. června 2012

Lubomír Krčmář

