

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název práce:	Natural Language in Web-based Search Engines
Autor/ka práce:	Bc. Barbora POLÁKOVÁ
Studijní program:	Informační studia a knihovnictví (ÚISK FF UK)
Vedoucí diplomové práce:	Mgr. Jan Brejcha
Oponent diplomové práce:	Mgr. Věra Pilecká
Hodnocení:	Velmi dobře

Cíl práce:

Cílem práce je posouzení využití přirozeného jazyka ve webových vyhledávacích a zasazení této problematiky do širšího kontextu dalších oborů (lingvistika, neuropsychologie apod.).

Shoda se zadáním diplomové práce:

Práce se shoduje se pokyny pro vypracování uvedenými v zadání diplomové práce.

Obsahová stránka práce:

Diplomová práce se zaměřuje na problematiku zpracování přirozeného jazyka pro účely vyhledávání informací pomocí webových vyhledávačů. V úvodních kapitolách autorka zasazuje téma do širšího teoretického kontextu; popisuje obecně proces vyhledávání informací i jeho jednotlivé typy a dále nástroje používané pro vyhledávání informací v prostředí webu. Jádrem diplomové práce jsou kapitoly 4 až 6, které se podrobněji zabývají zpracováním přirozeného jazyka ve webových vyhledávacích, „lidským informačním systémem“ a kognitivními procesy.

Jako nejpřínosnější z celé práce se jeví čtvrtá kapitola. Autorka popisuje specifické rysy přirozeného i umělého jazyka a z nich vycházející problémy při zpracování pro účely vyhledávačů. Velmi zajímavé jsou popisy konkrétních postupů používaných pro převod dotazu zadaného v přirozeném jazyce do jazyka systému. Některé systémy (AskMSR, Precise), o kterých se autorka zmiňuje, bohužel nejsou citovány v seznamu použitých zdrojů ani nejsou např. uvedeny odkazy na jejich webové stránky v poznámce k příslušnému místu textu.

Autorka se dobře zorientovala v terminologii a problematice oborů souvisejících se zvoleným tématem (neuropsychologie, kognitivní psychologie), což je zřejmé v páté a šesté kapitole, které obsahují stručná, ale jasná vysvětlení jevů jako např. přenos impulzů mezi neurony v lidském mozku, paměť jako kognitivní proces, mentální modely.

V diplomové práci autorka často využívá zkratky, jejichž význam sice vyplývá z kontextu, ale bylo by lepší je při prvním uvedení celého termínu odkázat v závorce (např. HIS pro *human information system*, IE pro *information environment* – v tomto případě se zkratka může plést s běžně používanou zkratkou pro Internet Explorer, pokud nebyla vysvětlena).

Celkově je po obsahové stránce diplomová práce na velmi dobré úrovni.

Výběr informačních zdrojů:

Zvolené informační zdroje odpovídají zaměření a rozsahu práce. Autorka prostudovala velké množství informačních pramenů převážně v anglickém jazyce, bibliografie je na diplomovou práci značně rozsáhlá a obsahuje stěžejní tituly.

Stylistická a gramatická úroveň práce:

Diplomová práce je specifická tím, že byla napsána v anglickém jazyce. Autorka v předmluvě uvádí jako primární důvod zpracování textu v tomto jazyce záměr pokračovat ve studiu doktorského programu v zahraničí. Přestože diplomantka odvedla nepochybně velký kus práce při zpracování cizojazyčného textu takového rozsahu, z textu je na mnoha místech patrné, že se nejedná o její rodný jazyk. Tento fakt bohužel čtení celé práce ovlivňuje a na některých místech znesnadňuje porozumění prezentovaných faktů a závěrů.

Gramatická úroveň práce je celkově dobrá, mezi častější chyby patří např. chybějící (zejména neurčité) členy podstatných jmen, použití čárek, neshodující se jednotné/množné číslo u podstatného jména a slovesa (*this paper...interlink*), chybějící sloveso ve větě, chybný slovosled (*important cognitive processes language related; later in 90s was this cognitive approach extended*), chybné využití anglického slovního druhu, např. záměna přídavných jmen a příslovcí (*language is dynamics, s. 18, presented in different syntactically and semantically language formats, s. 24, more accurately information retrieval, s. 58*), záměna podobných anglických výrazů (ve významu kromě *beside* místo *besides*). Autorka také na některých místech využívá správně apostrofu z anglické klávesnice, v mnoha případech je ovšem použita česká čárka.

Stylisticky je často patrná česká struktura věty; je zřejmé, kde autorka vycházela z dostupné anglické literatury a kde formulovala text sama či překládala z češtiny (např. s. 19 nahoře, závěr). Zcela poslední odstavec textu v závěru je esencí uvedených chyb - český čtenář patrně pochopí, co měla autorka na mysli, rodilý mluvčí by ale podle mého názoru měl s porozuměním této pasáži značný problém.

Formální stránka práce (struktura, formální a grafická úprava):

Práce má poměrně přehlednou a logickou strukturu, jednotlivé podkapitoly na sebe vhodně navazují.

Po formální stránce splňuje diplomová práce veškeré náležitosti.

Práce má velmi dobrou grafickou úroveň, oceňuji např. zvýraznění důležitých termínů v textu pomocí tučného písma. U obrázků 1 a 2 je nešťastné, že v textu se o nich hovoří pouze na s. 13 a samotné obrázky se nacházejí až na stranách 15 a 16.


Doplňující otázky (témata) pro obhajobu:

1. V práci hovoříte o zpracování přirozeného jazyka (NLP) zejména v souvislosti s možností zodpovídání dotazů uživatelů v přirozeném jazyce. Jsou možné i další aplikace NLP?

Závěrečné hodnocení:

S ohledem na výše uvedené hodnocení doporučuji diplomovou práci k obhajobě s návrhem klasifikace „velmi dobře“.

13. 9. 2010


Mgr. Věra Pilecká