

Posudek oponenta  
diplomové práce Bc. Lukáše Kobíka  
„Zpřístupnění multimediálních dat pomocí nepřímých metod ve vyhledávání“

Diplomová práce pana Lukáše Kobíka se zabývá aktuální problematikou vyhledávání relevantních multimediálních informací. V současném období extenzivního nárůstu množství dat dostupných uživatelům prostřednictvím Internetu se problém efektivního vyhledání vhodných informací stává klíčovým pro jeho úspěšné využití v mnohých oborech lidské činnosti. Přetížení množstvím informací („information overload“), které jen málo anebo vůbec neposkytují hledaný informační obsah a naopak „maskují“ před uživatelem skutečně relevantní obsah může vést paradoxně k diskreditaci služeb poskytování informací jako takových. Problém je možné řešit pomocí přímých a nepřímých metod, ze kterých se diplomant zaměřil na metody nepřímé, pomocí analýzy chování uživatelů ve vytvořené sociální síti. Na základě informací poskytnutých uživateli při registraci do této sítě a zjišťování jejich zájmů jsou na tomto základě klasifikována multimediální data – soubory s obrázky.

Diplomant se ve své práci zaměřil na návrh, vytvoření, popis a analýzu možného modelu, umožňujícího zpřístupnění multimediálních dat pomocí nepřímých metod ve vyhledávání. Pro zjednodušení se zaměřil pouze na datové soubory obsahující obrazové informace, v principu by ale bylo možné zřejmě uplatnit tento postup pro jakékoliv jiné multimediální data.

Po formální stránce je práce rozčleněna do deseti částí, seznamu použité literatury a šesti příloh. Elektronická verze nebyla při posuzování k dispozici.

Po stručném úvodu jsou v první kapitole popsány specifika multimediálních dat. V druhé kapitole diplomant stručně popisuje problematiku vyhledávání multimediálních dat. Jednotlivé metody pro přímé vyhledávání jsou obsahem kapitoly 3 a nepřímými metodami se zabývá kapitola 4. Vlastním přístupem, vyhledávání pomocí sociální sítě, se věnuje diplomant v kapitole 5. Zde podrobněji uvádí příklady současných sociálních sítí, možnosti jejich analýzy a v části 5.3 popisuje navrženou metodu estimace obsahu multimediálních dat na základě aktivit účastníků sociální sítě. Stručný popis návrhu modelu pro experimenty je obsahem kapitoly 6, popis použitého nástroje pro statistické analýzy SPSS Clementine obsahem kapitoly 7.

V kapitole 8 diplomant popsal jím provedenou praktickou část diplomní práce – realizaci aplikace SFINTO - WWW stránek a multimediálního obsahu, na kterém modeloval aktivity sociální sítě. Analýze a vyhodnocení dosažených výsledků je věnována kapitola 9, dosažené výsledky jsou zhodnoceny v závěru práce. Seznam použité literatury obsahuje 34 odkazů na literární a Internetové zdroje, použité v práci.

Po formální a jazykové stránce je práce napsána čistě bez zbytečných chyb a překlepů, které se vyskytují v minimální míře ( např. str. 12 využitelný a i pro ostatní ).

Po obsahové stránce lze autorovi vytknout používání nepřesných a někdy matoucích konstatování a definic v obecné části práce, například:

Str. 12: „Multimediální data jsou oblíbena ...díky jejich velké interpretační schopnosti“ – data samy osobě snad žádné interpretační schopnosti nemají?

Str. 12: Popis cíle práce: „Na základě informací, poskytnutých uživateli při registraci bude následně získáván popis obrázků“ – pokud by to bylo jen na základě popisu, asi by se nejednalo o nepřímou metodu, avizovanou v názvu práce.

Str. 13: „MM data jsou nestrukturovaná data ...“ – přílišné zobecnění – jisté typy MM dat jsou vysoce strukturované, např. soubory obrázků ve formátu JPEG, soubory ve formátech videa MPEG a podobných.

Str. 13: „...rozdíl mezi textovými a audiovizuálními daty..“ – podle předchozího textu nic jako textová data neexistuje – pouze typ vizuální a audio..

Str. 14: Úplně nesprávné a v rozporu s předchozími tvrzeními se mi jeví konstatování o textových datech jako zvláštním druhu vektorové grafiky. Hledání textových řetězců pomocí hledání prvků vektorové grafiky by bylo asi podstatně náročnější a méně efektivní, než jsou postupy používané současnými vyhledávači Google, Yahoo, Bing a pod.

Za stěžejní část práce lze považovat část od kapitoly 4. Zde se autor zaměřil na popis nepřímých metod a konkrétně v kapitole 5 vyhledávání pomocí použití sociálních sítí. Popisuje podrobně proces spojení údajů o uživateli na soubor s multimediálními daty, jednak schematicky, jednak uvádí příslušný matematický aparát, použitý pro vyhodnocování.

V kapitole 6 je popsán návrh modelu, použitého pro vyhodnocování. Na ověřování se podle uvedených údajů zúčastnilo 112 respondentů, kteří vytvořili modelovou sociální síť.

Výsledky získané pomocí navržené WWW aplikace jsou obsahem kapitoly 9. Na třech příkladech diplomant demonstruje použití jím navržené a odzkoušené metody. Podle mého názoru uvedené příklady úplně přesvědčivé, navržená metoda by si jistě zasloužila podrobnější analýzu a více typových příkladů, co by ale velmi pravděpodobně přesáhlo rámec diplomové práce.

Při obhajobě práce by diplomant mohl podrobněji vysvětlit svou volbu použitých metod a možnost jejich kombinace s jinými metodami, naznačenou v závěru práce. V závěru pouze obecně popisuje, že „...v některých případech bylo dosaženo i 75% úspěšnosti...byly i případy, kdy byla úspěšnost třeba jen 25%“ – bylo by vhodné, kdyby tyto případy konkrétně při obhajobě prezentoval a provedl diskusi, proč k tomuto jevu dochází.

Diplomant ve své práci prokázal dobrou odbornou způsobilost a přehled v odborné problematice a vytvořil přínosnou aplikaci, která se může stát základem dalšího zkoumání v uvedené oblasti. Uvedené výtky na formální či obsahovou stránku práce nejsou zásadního charakteru. Z těchto důvodů navrhuji přijmout práci diplomanta Bc. Lukáše Kobíka k obhajobě a hodnotím ji stupněm 2 - „velmi dobře“.

V Žilině dne 7.9.2009

doc. Ing. Peter Fabián, CSc.

v.r.