



AZ INTERNETES KERESŐK LEKÉRDEZÉSI HATÉKONYSÁGÁNAK VIZSGÁLATA

Doktori (PhD) értekezés tézisei

TÓTH ERZSÉBET

Debreceni Egyetem
Informatikai Kar
Debrecen, 2007.

Tartalomjegyzék – Table of Contents

I. Bevezetés, irodalmi előzmények.....	1
II. Kutatási célkitűzések	3
III. Új eredmények.....	8
IV. Irodalomjegyzék	13
V. Publikációk jegyzéke.....	19
I. Introduction.....	22
II. Research objectives	24
III. Results.....	29
IV. References	34
V. Publications	40

I. Bevezetés, irodalmi előzmények

- A keresők vizsgálata során számos nehézséggel kell szembesülnünk, amelyek közül csak néhányat emelnék ki: a tanulmányok nagy része nem számol be az általa használt módszertanról, hiányzik belőlük a szisztematikus megközelítés, a keresési folyamat sokféleképpen befolyásolható az értékelések során, a keresők állandóan változnak, továbbá a világháló is egy dinamikusan változó közeg. A problémák ellenére megállapíthatjuk, hogy mégis van értelme vizsgálni a keresőket, de szabványos értékelési módszereket erre a feladatra jelenleg nem alkalmazhatunk.

- A dolgozat tárgyalja a keresők működési alapelveit és az alkotóelemeik feladatait, hangsúlyozza a keresők meghatározó tevékenységét a találatrangsorolásban. Röviden áttekinti a keresők vizsgálatával kapcsolatos problémákat és hiányosságokat. Összefoglalja az információkeresés területén használt legkorábbi statisztikai módszereket, majd pedig fókuszál a jelentősebb statisztikai vizsgálatok bemutatására, részletezve azok módszereit, eredményeit és alkalmazott mérőszámait. Bemutatja, hogy az angol és a magyar nyelvű keresők milyen eredményességgel kezelik és értelmezik a lefuttatott keresőkérdéseket morfológiai szempontból. Értékeli a magyar nyelvű keresők hatékonyságát a pontosság és a teljesség mértékek szempontjából.

- A magyar nyelvű keresők lekérdezési hatékonyságának vizsgálatát az is alátámasztja, hogy eddig nagyon kevés tanulmány született a szakirodalomban, amely a nem angol nyelvű keresésekkel behatóbban foglalkozott volna.

- Az információkeresés területén végzett kutatások rámutatnak arra, hogy a kutatóknak rendkívül nehéz megbízható mértékeket találniuk az elemzésekhez. Sokszor élénk vita folyik közöttük arról, hogy melyik mértéket miért lehetne előnyös vagy hátrányos használniuk a vizsgálatok során. Ebből adódóan a keresők értékelésével foglalkozó kutatás jelenleg nem egységes az alkalmazott módszerek tekintetében, ezért nagy szükség lenne megfelelő szabvány kidolgozására és használatára ezeknél az elemzéseknél.

- A keresők lekérdezési teljesítményének vizsgálata a kutatók érdeklődésének középpontjában áll, mert különféle megközelítéseket és módszereket használnak a mérésekben. Néhány jelentősebb kutató tevékenységére szeretném felhívni a figyelmet a témával kapcsolatban:

Sroka a helyi fejlesztésű keresőket és az angol nyelvű keresők lengyel változatát tanulmányozta. Mérésében a pontosságra helyeződött a hangsúly, amit relevancia ítéletek alapján állapított meg [Sroka2000].

Clarke és Willett a keresésekre kapott első tíz találat relevanciáját mérték. Kiszámították a pontosság, a teljesség és a téma lefedettség átlagértékeit [Clarke-Willett97].

Leighton és Srivastava öt különböző kísérletet végeztek, ahol a keresőktől kapott első 20 találat pontosságát mérték. Megfelelő mérőszámokat dolgoztak ki a találatok pontosságának a meghatározására [Leighton-Srivastava99].

Chignell, Gwizdka és Bodner első kísérletükben azt vizsgálták, hogy a napi idő és a keresési stratégia hogyan befolyásolta a keresés feldolgozási idejét mindegyik kereső esetében. Második kísérletükben azt tanulmányozták, hogy a földrajzi kiterjedés és a domain nevek hogyan hatottak a keresők teljesítményére [Chignell-Gwizdka-Bodner99].

Lucas és Topi azt elemezték, hogy a kereső operátorok és a keresőkifejezések kiválasztása milyen mértékben befolyásolta a kapott találatok relevanciáját [Lucas-Topi2002].

Radev, Libner és Fan azzal foglalkoztak, hogy a legnépszerűbb keresők mennyire hatékonyan válaszolták meg a természetes nyelvű kereséseket [Radev-Libner-Fan2002].

Bar-Ilan a keresők működését hosszabb időtartamon keresztül követte nyomon. A találatszolgáltatásuk stabilitását számos mérőszám irányából közelítette meg [Bar-Ilan2002].

Bar-Ilan és Gutman morfológiai szempontból tanulmányozták a nem angol nyelvű keresőkérdések értelmezését. Mindketten arra koncentráltak, hogy a keresők mennyire képesek figyelembe venni a vizsgált nyelvek egyedi sajátosságait [Bar-Ilan-Gutman2005].

- Leggyakrabban a pontosságot és a teljességet használják jellemzőként a mérésekben, azok azonban gyakran rivalizálnak egymással. Mindkét mérték érzékeny arra, hogy a relevanciát hogyan definiáljuk és mérjük. Az eddig megjelent tanulmányok csak a pontosságot vették figyelembe. Minimális mértékben, vagy egyáltalán nem kísérelték meg a különböző keresők teljességének a mérését.

II. Kutatási célkitűzések

- A disszertációban ismertetett vizsgálattal az volt a *célom*, hogy összehasonlítsam a magyar nyelvű keresők lekérdezési hatékonyságát. Az információkereső rendszerek hatékonyságát leginkább a pontosság és a teljesség mértékek fejezik ki, ezért ezen két mutató alapján elemeztem a magyar nyelvű keresőket. *Pontosság alatt a releváns dokumentumoknak a visszakeresett dokumentumok összességén belüli arányát értem. A teljesség pedig a megtalált releváns dokumentumoknak az összes (akár talált, akár nem) releváns dokumentumhoz viszonyított arányát jelenti ([Ungváry2001] p. 196-197).*

- A kutatási területen való előzetes tájékozódás során felfigyeltem egy másik szempontra, amely arra irányult, hogy a keresők mennyire képesek figyelembe venni a vizsgált nyelvek egyedi sajátosságait és milyen hatékonyan válaszolnak a nem angol nyelvű keresésekre. Ennek megfelelően végeztem egy másik vizsgálatot azzal a célkitűzéssel, hogy felmérjem az angol és a magyar nyelvű keresők nyelvi lehetőségeit különböző keresésekre. *Arra a kérdésre kerestem a választ, hogy ezek a keresőszolgáltatások mennyire hatékonyan kezelik és értelmezik a keresőkérdéseket morfológiai szempontból. Kutatásomnak nem volt célja, hogy a vizsgált keresőkről abszolút kijelentéseket fogalmazzak meg, hanem néhány tesztadat segítségével ráirányítsam a figyelmet arra, hogy ezek a keresők az adott kérdés kapcsán mennyire egymástól eltérően viselkednek. Ebben a kontextusban a vizsgálataim és a megállapításaim is sokkal inkább kvalitatív, mint kvantitatív jellegűek.* Habár a tesztadatok igen gondosan lettek megválasztva világos, hogy a vizsgált csekély számú példa nem alkalmas statisztikai következtetések levonására. Azonban egy későbbi kutatás során a felvázolt problémák statisztikai vizsgálata is érdekes lehet.

A morfológiai vizsgálatban az angol nyelvű keresőkön angol és magyar nyelvű keresőkérdéseket egyaránt lefuttattam, a magyar nyelvű keresőkön azonban csak magyar nyelvű kereséseket hajtottam végre, mert azok főként magyar nyelvű weboldalakat indexelnek, kivéve az Origo-Vizsla és az Altavizsla szolgáltatásokat.

- Az általam végzett morfológiai vizsgálat főbb módszertani lépései a következők voltak:
 - *A vizsgálat az angol és a magyar nyelv esetében a megfelelő szavak, kifejezések felkutatásával kezdődött. Körültekintően választottam ki a tesztelésre szánt keresőkérdések halmazát.*
 - *Ezt követően próbakereséseket hajtottam végre mindegyik keresőn, annak ellenőrzésére, hogy a kiválasztott keresőkérdések egyértelműen kifejezik-e a vizsgált nyelvi problémát és, hogy megfelelnek-e a tesztelés célkitűzéseinek. A keresőkérdések kiválasztásánál*

alapvetően arra törekedtem, hogy olyan kifejezéseket válasszak, amelyek mindkét nyelv speciális nehézségeit tükrözik. Döntésemben az előzetes megfigyeléseimre támaszkodtam.

- *A vizsgálat megkezdése előtt áttekintettem mindegyik keresőszolgáltatás tájékoztató anyagát, hogy pontosan tájékozódjam mindazokról a fontos jellemzőkről, sajátosságokról, amelyek relevánsak lehetnek a tesztelés szempontjából az adott keresőre nézve.*
- *A lefuttatott keresésekre az első 100 találatot vizsgáltam meg. Egyedüli kivételt képezett a csonkolás vizsgálata a magyar nyelvben, ahol keresőkérdésként egy szűkebb találathalmazt eredményező szakkifejezést adtam meg a “májműtét”, ami egy parazitát, illetve férget jelöl. Elsődlegesen a találatok leírását vettem alapul, de ahol indokolt volt azok tartalmába is belenéztem.*
- *Vizsgálatomat 2005 júliusa és szeptembere között végeztem el. A tesztelésbe három angol nyelvű keresőt vontam be: a Google-t, az AltaVista-t és az AlltheWeb-et. Öt jelentősebb magyar nyelvű keresőt teszteltem, amelyek a következők voltak: Heuréka, Origo-Vizsla, Kurzor, Góliát és az Altavizsla.*
- *A vizsgálatához Bar-Ilan és Gutman közös tanulmányában [Bar-Ilan-Gutman2005] lévő szempontokat vettem alapul, amelyek a következők voltak:*
 - *Morfológiai elemzés (stemming) alatt azt értem, amikor a kereső megtalálja egy keresőkérdés toldalékolt alakjait. Tehát azt vizsgáltam, hogy az adott kereső megtalálja-e egy keresőkérdés többes számú alakját vagy sem. Keresőkérdéseim a következők voltak: “dog-dogs”, “ház-házak”, “kocsi-kocsik”, “kutya-kutyák”. Az első két magyar példánál megfigyelhető, hogy a tövégi magánhangzó nem változik a többes számú alakban az egyes számúhoz képest, míg a “kutya-kutyák” esetében igen. Választottam egy fosztóképzővel ellátott főnevet is – “tisztességtelen” – a morfológiai elemzés elvégzésére. Tettem ezt azzal a céllal, hogy megfigyeljem a keresőszolgáltatás ennek az összetett kifejezésnek milyen más egyéb szóalakjait találja meg, azaz az -e valamilyen morfológiai elemzést erre a kifejezésre.*
 - *A stopszavak kezelésénél azt ellenőriztem, hogy a keresőkérdés megjelenik-e a határozott és a határozatlan névelőkkel együtt a találatleírásokban. A keresésekre kapott találatokban azt tanulmányoztam, hogy a keresők külön keresnek-e a megadott névelőkre vagy sem. A következő keresőkérdéseket vizsgáltam meg az angol nyelvben: “a dog” (=egy kutya), “an aunt” (=egy nagynéni), “the car” (=az autó). A magyar nyelv vonatkozásában pedig “a ház”, “az ember”, “egy kocsi” kereséseket használtam.*

- Az ékezetek kezelésénél arra koncentráltam, hogy a megadott keresőkérdéseknek az ékezetes változatait is lekérdezi-e a kereső a találatai között vagy sem. Ez csak a magyar nyelv esetében volt releváns vizsgálati szempont. Itt nevezetesen két keresőkérdést használtam a *“kertem”* és az *“alma”* kereséseket.
- Minden esetben a * karaktert használtam csonkolásra a keresőkifejezés után. Az angol nyelv esetében az *“Olympi*”* keresőkérdést adtam meg azzal a céllal, hogy megtaláljam az összes olimpiai játékokról szóló oldalt az alábbi kifejezésekkel, pl.: *“Olympic”* (=olimpiai), *“Olympics”* (=olimpia, olimpiai játékok), *“Olympia”* (=földrajzi név), *“Olympian”* (=olimpuszi), stb. *A magyar nyelv esetében szándékosan törekedtem egy szűkebb, behatároltabb találathalmaz vizsgálatára annak érdekében, hogy a keresőkifejezés ragozott szóalakjait könnyebben találjam meg és le tudjam ellenőrizni azok tényleges előfordulását a találathalmazon belül.* Ezért a *“májmétely”* nevű parazita ragozott szóalakjaira kerestem a *“májmétely*”* keresőkérdéssel. Továbbá megvizsgáltam a különböző keresőknél, hogy a *“májmétely*”* keresőkifejezésre kapott találathalmazok ténylegesen tartalmazzák-e a keresőkifejezés ragozott szóalakjait vagy sem.
- Végül pedig azt teszteltem, hogy a kereső felkutatja-e egy keresőkérdés szinonimáit vagy sem. Tanulmányoztam, hogy a szinonimák hogyan jelennek meg a kapott találatokban, például azok ki vannak-e emelve a találatleírásokból vagy sem, azok a megadott keresőkérdéssel együtt fordulnak-e elő a találatleírásokban, vagy azok attól függetlenül is szerepelnek bennük. Az angol nyelv esetében az alábbi keresőkérdéseket vizsgáltam meg ilyen szempontból: *“car”* (=autó), *“glasses”* (=szemüveg). A magyar nyelvben pedig a *“kutya”* és a *“vetélkedő”* keresőkérdések szinonimáira kerestem.

A keresők nyelvi megoldásait tehát ezen szempontok alapján vizsgáltam meg a különböző keresőkérdésekre. Ezzel a vizsgálati módszerrel két felállított hipotézisem tanulmányozható, amelyek a következőképpen foglalhatók össze: *a magyar és az angol nyelvű keresők teljesítménye vizsgálható és az különbözik a megvalósított nyelvi lehetőségeik tekintetében.*

- *A másik vizsgálatot a pontosság és a teljesség jellemzőkre irányulón végeztem el 2006 januárja és júniusa között. A pontosság vizsgálatakor csak az első hatvan visszakeresett találat relevanciáját vettem figyelembe. Döntésemet a Google kereső fejlesztőinek azon észrevételével támasztottam alá, mely szerint: a Google fejlesztésénél elsődlegesen arra törekedtek, hogy minden egyes keresőkérdésre nagyon releváns találatok jelenjenek meg a találati lista elején, mert a felhasználók csak az első néhány tíz találatot nézik meg kereséskor ([Brin-Page98] p. 3).*

- *A pontosságot a használat és a hatékonyság szempontjából elemeztem. A használat szempontja azt jelenti számomra, hogy a vizsgált keresők mennyi releváns találatot keresnek vissza az első, a második és a harmadik húsz találatban a különböző keresésekre. Ez lényeges kérdés a felhasználók számára, hogy melyik keresőt használják a releváns információk felkutatására. Hatékonyság alatt azt értem, hogy a keresők mennyire egyenletesen szolgáltatják a releváns dokumentumokat az első, a második és a harmadik húsz találatban, illetve az első hatvan találatban a lefuttatott keresőkérdésekre. Ezt pedig a keresők visszakeresett releváns találatainak az átlagértékei és tapasztalati szórás értékei alapján állapítottam meg. Kiszámítottam egy százaléktérteket a találatok pontosságára vonatkozóan, ami azt fejezte ki, hogy hány releváns találatot kaptam az első, a második és a harmadik húsz visszakeresett találatból. Tanulmányoztam az egyes keresőkérdések vonatkozásában a találati pontosság értékek változását az első, a második és a harmadik húsz találatban. Ezen elemzések alapján a pontosságra vonatkozóan különböző eredményeket kaptam. Ezzel az összetett vizsgálati módszerrel két hipotézisem vizsgálható, amelyek a következők: a találati lista elején lévő találatok a legnagyobb pontosságúak, majd ezután fokozatosan csökken a soron következő találatok pontossága; a vizsgálat eredményeként megkapom azokat a magyar nyelvű keresőket, amelyek a pontosságot a legnagyobb, illetve a legkisebb mértékben valósítják meg a találatszolgáltatásukban.*

- *Ezt követően megvizsgáltam, hogy lehetséges-e meggyőző következtetést meghatároznom a releváns találatok számára vonatkozóan a találati lista hátralévő részében. Ezzel az elemzéssel az alábbi hipotézisemet vetettem el: a pontosság változik az egyes vizsgált keresők esetében, azaz tudok valamilyen következtetést adni a pontosság alakulására vonatkozóan, ami egyben érvényes a találathalmaz egészére.*

- *A teljesség vizsgálatát a szakirodalomból ismert relatív teljesség alapján végeztem el, ami a következőképpen számolható ki: az egy adott kereső által visszakeresett releváns dokumentumok száma osztva az összes vizsgált keresők valamelyike által megtalált releváns dokumentumok számával ([Clarke-Willett97] p. 185-187). Ennek értelmében a vizsgált keresők összes visszakeresett találatából összeállítottam egy releváns találatokból álló halmazt „csiperkegomba” témában és azt elemeztem, hogy az egyes keresők milyen mértékben keresték vissza ennek a halmaznak a releváns dokumentumait. Kiszámoltam egy százaléktérteket a relatív teljességre vonatkozóan, ami azt tükrözte, hogy az egyes keresők ezen találathalmaz releváns dokumentumainak hány százalékát találták meg. A kapott százaléktértekek alapján pedig egy rangsort állítottam fel a keresők között. Ezzel a módszerrel a következő hipotézis tanulmányozható: a vizsgálat eredményeként értékelhetem, hogy a magyar nyelvű keresők milyen mértékben képesek*

visszakeresni a releváns találatok halmazát, azaz találathalmazaik milyen arányban tartalmazzák azokat.

- *A vizsgálat elején definiáltam, hogy milyen feltételek mellett tekintek egy weboldalt relevánsnak, azonban relevancia kategóriákat nem hoztam létre a dokumentumok relevanciájának eldöntésére.*
- *Összesen öt magyar nyelvű kereső vett részt ebben a vizsgálatban, amelyek a következők voltak: Heuréka, Origo-Vizsla, Kurzor, Góliát és az Altavizsla. Öt témában kizárólag magyar nyelvű keresőkérdéseket futtattam le ezeken a keresőkön. Választott témáim a következők voltak: “csiperkegomba”, “Rákóczi-szabadságharc”, “evészavar”, “kegyhely”, “szökőár”. Az öt téma közül egyetlenegy irányult a teljesség vizsgálatára (“csiperkegomba”), a fennmaradó négy téma pedig a pontosság tanulmányozására helyezte a hangsúlyt. A teljesség vizsgálatára egy olyan összetett keresést fogalmaztam meg, amely teljes mértékben figyelembe vette a megadott téma különböző sajátosságait. A “csiperkegomba” témához kapcsolódóan a következő keresőkérdéseket határoztam meg: “sampinyon”, “termesztett csiperke”, “kétspórás csiperke”, “agaricus bisporus”. Fontos megjegyezni, hogy a megfelelő keresőkérdések kiválasztása mindig valamilyen háttértudást követel meg. A vizsgálat megtervezésekor jelentős mértékben támaszkodtam a próbakeresések során nyert előzetes tapasztalataimra.*

III. Új eredmények

A két empirikus vizsgálat eredményeit a felállított hipotéziseim alapján foglalom össze, továbbá ismertetem az eredményekből származó téziseket.

III.1. Az angol nyelvű keresők teljesítménye vizsgálható a megvalósított nyelvi lehetőségeik szempontjából.

A vizsgálat eredményei arra engednek következtetni, hogy az angol nyelvű keresők rosszabbul kezelik és értelmezik a magyar nyelvű kereséseket, mint az angol nyelvűeket.

Ezekben a keresőeszközökben a csonkolás és a szinonimákra történő keresés eredményesen működik az angolban, de meglehetősen problémás a magyar nyelvben.

A Google kereső nem veszi figyelembe a határozott és a határozatlan névelőket az angol nyelvű kereséseknél, azonban még nem oldotta meg ezt a kérdést a magyar nyelvben. Az Altavista kereső megtalálja az angol nyelvű keresőkérdések többes számú alakját, ezért ez az egyedüli olyan szolgáltatás, ahol a morfológiai elemzés hatékonyan működik az angolban. Az angol nyelvű keresők közül egyik sem oldotta meg ezt a problémát a magyar nyelvben. Elmondható a Google és az Altavista szolgáltatásokról, hogy azok azonos teljesítményt nyújtottak a keresőkérdések értelmezésében, utánuk pedig az AlltheWeb kereső következik a rangsorban. A kapott eredmények alapján levonható az a következtetés, hogy ezeknek a keresőknek még nagyobb hangsúlyt kellene fektetniük a morfológiai elemzésre, a stopszavak és az ékezetek kezelésére a jövőbeli fejlesztésükben.

1. tétel: Az angol nyelvű keresőknél a csonkolás és a szinonimákra történő keresés hatékonyan működik az angolban, azonban az nem eredményes a magyar nyelvben.

2. tétel: Az angol nyelvű keresőknek javítaniuk kell a morfológiai elemzést, a stopszavak és az ékezetek kezelését a jövőbeli fejlesztésükben.

Módszertanilag újdonság értékét az adja ennek a megállapításnak, hogy az angol nyelvű keresőkön magyar és angol nyelvű keresőkérdéseket vizsgáltam ugyanabból a szempontból. Ezáltal pedig lehetővé vált számomra az angol nyelvű keresők teljesítményének az összehasonlítása mindkét nyelv vonatkozásában. Tudomásom szerint a szinonimákra történő keresés új megközelítés, ami még nem fordult elő a korábbi vizsgálatokban.

III.2. A magyar nyelvű keresők teljesítménye vizsgálható és különbözik a megvalósított nyelvi lehetőségeik tekintetében.

A magyar nyelvű keresők többsége helyesen kezeli az ékezeteket, ebből adódóan azok egy lényeges felhasználói követelménynek tesznek eleget.

A magyar nyelvű keresők nyelvi megoldásait értékelve megállapítható, hogy a Heuréka kereső nyújtotta a legjobb teljesítményt a morfológiai elemzés és a szinonimákra történő keresés területén. Teljesítménye elfogadható volt két másik területen, a stopszavak és az ékezetek kezelésében.

Ez után következnek a Góliát és az Altavizsla keresők, amelyek nagyon hasonló teljesítményt nyújtottak a tesztelés során. Mindketten pontosan kezelték az ékezeteket. Azonban a stopszavakat illetően fejleszteniük kell a jelenlegi működésüket.

A rangsorban az Origo-Vizsla és a Kurzor keresők az utolsó helyre kerültek ugyanazzal a teljesítménnyel. Az Origo-Vizsla szolgáltatás számos területen mutatott hiányosságokat, mint például morfológiai elemzés, csonkolás, ékezetek és stopszavak kezelése.

A Kurzor kereső hiányosságokkal rendelkezett az alábbi területeken: morfológiai elemzés, stopszavak kezelése, csonkolás, szinonimákra történő keresés. A vizsgálat eredményei alapján kijelenthető, hogy majdnem mindegyik magyar nyelvű keresőnek fejlesztenie kell a saját teljesítményét a csonkolás és a morfológiai elemzés területén.

3. tézis: *A magyar nyelvű keresők többsége pontosan kezeli az ékezeteket.*

4. tézis: *Azonban majdnem mindegyiknek fejlesztésekre kell törekednie a csonkolás és a morfológiai elemzés területén.*

A ragozott szóalakok felkutatása lehetővé tette a csonkolás mélyebb szintű elemzését a magyar nyelv esetében, erre a megközelítésre az eddigi vizsgálatok során nem találtam példát.

III.3. A pontosság változik az egyes vizsgált keresők esetében, azaz tudok valamilyen következtetést adni a pontosság alakulására vonatkozóan, ami egyben érvényes a találathalmaz egészére.

Megjegyzem, hogy a pontosság minimális mértékben változott a Góliát és az Altavizsla keresőknél, mert azok közel azonos számú releváns találatot kerestek vissza az első, a második és a harmadik húsztalálatban mind a négy keresőkérdésre. Ezenkívül mindkét kereső megegyező indexet használt a visszakeresésre, amit a vizsgálat eredményei is megerősítettek, hiszen ugyanazokat a releváns dokumentumokat keresték vissza mind az öt témára vonatkozóan egyetlenegy weboldal kivételével. A vizsgálat során nyert adatokból kiderült, hogy nem fedeztem fel evidencia szintű összefüggést az első hatvan találatban lévő releváns találatok száma és a keresőkérdésekre kapott összes találat száma között. Ebből adódóan nem határozhattam meg meggyőző következtetést a releváns találatok számára vonatkozóan a találathalmaz hátralévő részében. Tehát a harmadik hipotézist el kellett vetnem.

III.4. A találati lista elején lévő találatok a legnagyobb pontosságúak, majd ezután fokozatosan csökken a soron következő találatok pontossága.

A vizsgált keresők többsége arra törekedett, hogy a legnagyobb mennyiségű releváns találatot szolgáltatassa az első húsz találatban mind a négy keresőkérdésre.

Az első hatvan találatban azok háromféle tendenciában keresték vissza a releváns találatokat, amelyek a következők voltak: szigorúan monoton csökkenő, monoton csökkenő és nem monoton. Ebben az esetben a keresőket a használat szempontjából tanulmányoztam, ahol kizárólag az első, a második és a harmadik húsz találatban visszakeresett releváns találatok mennyiségét vettem figyelembe.

A hatékonyság szempontjából elemezve a keresőket azt mondhatom, hogy mind a négy keresőkérdés esetében az első húsz találatban bizonyultak a legjobbnak, mivel a releváns találataik átlagértéke ezen a helyen volt a legmagasabb.

Azonban a keresők hatékonysága erőteljesen lecsökkent a második és a harmadik húsz találatban mind a négy keresőkérdésre, mert a releváns találataik átlagértéke ugyanezekben a helyeken nagyon alacsonynak minősült. Továbbá a keresők hatékonyságában minimális eltérés volt tapasztalható a második és a harmadik húsz találatban mind a négy keresőkérdésre, hiszen csekély különbség jelentkezett a releváns találataik átlagértékei között ezeken a vizsgált helyeken.

5. tézis: A magyar nyelvű keresők többsége az első húsz találatban mind a négy keresőkérdésre szolgáltatta a legnagyobb mennyiségű releváns találatot.

6. tézis: Továbbá azok a leghatékonyabbnak bizonyultak az első húsz találatban mind a négy keresőkérdésre, hiszen a releváns találataik átlagértéke ezen a helyen volt a legmagasabb.

A magyar nyelvű keresőket a használat szempontjából közelítettem meg, ami a felhasználók számára lényeges kérdés. A használat szempontjából kapott eredményt a releváns találatok átlagértékei és tapasztalati szórás értékei egyaránt alátámasztották a hatékonyság szempontjából.

III.5. A vizsgálat eredményeként megkapom azokat a magyar nyelvű keresőket, amelyek a pontosságot a legnagyobb, illetve a legkisebb mértékben valósítják meg a találatatszolgáltatásukban.

Használat szempontjából a Kurzor és az Origo-Vizsla keresők minősültek a legjobbnak a felhasználók számára, mert mindkét kereső arra törekedett, hogy a lehető legtöbb releváns találatot kutassa fel az első húsz találatban a lefuttatott keresésekre. Amennyiben a keresők találatatszolgáltatásának az egyenletes minőségére helyezem a hangsúlyt az első hatvan találatban, és nem csupán a visszakeresett releváns találataik mennyiségére fókuszálok ugyanezen a helyen, akkor ez utóbbitól eltérő eredményeket kapok. Megjegyzem, hogy egy kereső találatatszolgáltatásának az

egyenletes minőségét a hatékonyság szempontja fejezi ki számomra, amit a releváns találatok átlagértékei és tapasztalati szórás értékei figyelembevételével kapok meg.

Az Origo-Vizsla kereső bizonyult a leghatékonyabbnak az első hatvan találatban mind a négy keresőkérdésre, mert a releváns találatainak az átlagértéke a legnagyobb volt a többi vizsgált keresőéhez képest.

A hatékonyság szempontjából a Heuréka kereső került a második helyre, bár a releváns találatainak az átlagértéke sokkal kisebb volt az Origo-Vizslához képest az első hatvan találatban mind a négy keresőkérdésre.

Ugyanebből a szempontból a Kurzor kereső szerepelt a harmadik helyen a rangsorban, hiszen a releváns találatainak az átlagértéke alacsonynak minősült – a tapasztalati szórás figyelembevételével – az első hatvan találatban mind a négy keresőkérdésre.

Végül pedig az utolsó helyen osztozkodtak a Góliát és az Altavizsla keresők, azonban azok hatékonysága nem különbözött egymástól, mert a releváns találataik átlagértékei és a tapasztalati szórás értékeik közel azonosak voltak az első hatvan találatban mind a négy keresőkérdésre.

7. tézis: A magyar nyelvű keresőket rangsorolhatom a hatékonyságuk alapján, figyelembe véve azok releváns találatainak átlagértékeit és tapasztalati szórás értékeit az első hatvan találatban mind a négy keresőkérdésre.

A hatékonyság itt azt tükrözi számomra, hogy a vizsgált keresők mennyire egyenletesen képesek szolgáltatni a releváns dokumentumokat az első hatvan találatban. Ennek értelmében a pontosságot egy új aspektusból vizsgáltam, nem csupán a használat szempontjából tanulmányoztam azt.

III.6. A vizsgálat eredményeként értékelhetem, hogy a magyar nyelvű keresők milyen mértékben képesek visszakeresni a releváns találatok halmazát, azaz találathalmazaik milyen arányban tartalmazzák azokat.

A relatív teljességre kapott százaléktételek alapján a következő rangsort állítottam fel a magyar nyelvű keresők között. Az Origo-Vizsla kereső találta meg a vizsgált releváns találathalmaz legtöbb dokumentumát, mert azok 42,85%-át fedezte fel a weben. A rangsorban a második helyre kerültek a Kurzor, a Góliát és az Altavizsla keresők 23,80%-kal, hiszen azok azonos mennyiségű releváns dokumentumot kerestek vissza ebből a találathalmazból. Végül pedig a Heuréka kereső kissé lemaradt az előző három kereső visszakeresési teljesítményétől, mert az a találathalmazban lévő releváns dokumentumok 19,04%-át kutatta fel a weben.

Figyelemreméltó mutató számomra, hogy a magyar nyelvű keresők átlagosan 5,6 releváns dokumentumot találtak meg a weben „csiperkegomba” témával kapcsolatban, ami a vizsgált halmazban lévő huszonegy találat 26,66%-ának felelt meg. Tehát a magyar nyelvű keresők a

releváns találatok halmazának több mint az egynegyedét keresték vissza átlagosan „csiperkegomba” témában, ami meglehetősen szerény teljesítménynek minősül.

8. tézis: *A magyar nyelvű keresők visszakeresési teljesítménye rangsorolható a relatív teljesség alapján. Azok a releváns találatok halmazának több mint az egynegyedét keresték vissza átlagosan „csiperkegomba” témában.*

Korábban mások is vizsgálták a keresőket a relatív teljesség szempontjából. A meglévő kutatások számára érdekes adalékul szolgálhat ez az eredmény a magyar nyelvű keresők vonatkozásában.

III.7. További kutatási feladatok

- A megfogalmazott tézisek alapját képezhetik egy további, részletesebb statisztikai elemzésnek. Vizsgálataimban egy tesztkulcssor összeállítására törekedtem, amit óvatosan lehet bővíteni egy további statisztikai elemzés számára. Érdekes lenne összevetni a vizsgálataimból származó eredményeket egy másik, újabb elemzés eredményeivel.
- Véleményem szerint a szemantikus web területén található kutatások eredményei jól hasznosíthatóak lennének egy újabb vizsgálat számára, amelyek a keresőkérdések és a felkutatható dokumentumok jelentésével kapcsolatosak.
- Tervezek és egyben kezdeményezek projekt szintű együttműködést finn, észt kutatókkal, akik keresőszolgáltatásaikban minőségi, releváns információ-visszakeresés megvalósítására törekednek. Ezenkívül lehetőség szerint törekszem projekt kialakítására olyan céggel, amely épít az általam feltárt eredményekre.
- Szükségesnek tartom, hogy hallgatóimat bevonjam az internetes keresők hatékonyságával kapcsolatos vizsgálatokba, kutatásokba. Továbbá fontos, hogy megismertessem velük ennek a területnek a fejlesztéseit, tudományos eredményeit. (Eddig szakdolgozatok, szemináriumi dolgozatok és feladatok keretében volt erre lehetőségem.)

IV. Irodalomjegyzék

- [1] Albert, L.–Locsmándi, Cs.–Vasas, G.: *Ismerjük fel a gombákat!* Gabó, 1995. 191 p.
- [2] *AlltheWeb tájékoztató segédlete*
<http://www.alltheweb.com/help/index> 2007.03.10.
- [3] *Altavista tájékoztató segédlete*
<http://www.altavista.com/help/search/default> 2007.03.10.
- [4] Bar-Ilan, J.: Methods for measuring search engine performance over time = *Journal of the American Society for Information Science*, vol. 53. no. 4. (2002), p. 308-319.
- [5] Bar-Ilan, J.: Criteria for evaluating information retrieval systems in highly dynamic environments. In.: *Proceedings of the 2nd International Workshop on Web Dynamics*, Honolulu, Hawaii, 7th May 2002.
<http://www.dcs.bbk.ac.uk/webDyn2/onlineProceedings.html> 2007.03.02.
- [6] Bar-Ilan, J.–Gutman, T.: How do search engines respond to some non-English queries? = *Journal of Information Science* vol. 31. no. 1. (2005), p. 13-28.
- [7] Benczúr, A.–Bíró, I.–Csalogány, K.–Rácz B.–Sarlós T.–Uher, M.: PageRank és azon túl: Hiperhivatkozások szerepe a keresésben = *Magyar Tudomány*, 167. évf. 11. sz. (2006), p. 1325-1331.
<http://www.matud.iif.hu/06nov/07.html> 2007.07.17.
- [8] Berners-Lee, T.–Fielding, R.–Frystyk, H.: *Hypertext Transfer Protocol – HTTP/1.0. (RFC1945)*. May 1996.
<http://www.rfc-editor.org> 2007.08.10.
- [9] Bilodeau, M.–Brenner, D.: *Theory of multivariate statistics*. Springer, 1999.
- [10] Bocsán Róbert [Személyes közlés] 2007.09.04.
- [11] Brin, S.–Page, L.: The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, vol. 30. no. 1-7. (1998), p. 107-117.
<http://infolab.stanford.edu/pub/papers/google.pdf> 2007.03.02.
- [12] Chignell, M. H.–Gwizdka, J.–Bodner, R. C.: Discriminating meta-search: A framework for evaluation = *Information Processing and Management*, vol. 35. (1999), p. 337-362.
- [13] Clarke, S. J.–Willett, P.: Estimating the recall performance of web search engines = *Aslib Proceedings*, vol. 49. no. 7. (1997), p. 184-189.
- [14] *Creating a European library space: Telematics for libraries programmes 1990-1998*.
<http://cordis.europa.eu/libraries/en/intro.html> 2008.01.22.
- [15] Cronbach, L. J.: Coefficient alpha and the internal structure of tests = *Psychometrika*, vol. 16. (1951), p. 257-334.

- [16] Ding, W.–Marchionini, G.: A comparative study of web search performance in global complexity: Information, chaos and control. In.: *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, Baltimore, Maryland, 21-24 October 1996. p. 136-142.
- [17] Fodor, I.: Merre megy a világ gazdasága, merre mehetünk mi? In.: *Az információs társadalom*. Összeáll. Demetrovics J., Keviczky L. Bp.: MTA, 2000. p. 95-113.
- [18] Friedman, E.–Uher, M.–Windhager, E.: Keresés a világhálón = *Híradástechnika*, 58. évf. 3. sz. (2003), p. 20-24.
<http://www.ilab.sztaki.hu/websearch-data/Publications/kereses.pdf> 2007.07.17.
- [19] Froehlich, T. J.: Case study 5.1: Developing search engine evaluation criteria. In.: *Library evaluation*. Libraries Unlimited, 2001. p. 185-200.
- [20] *Google tájékoztató segédlete*
<http://www.google.com/support> 2007.03.10.
- [21] Gordon, M.–Pathak, P.: Finding information of the World Wide Web: The retrieval effectiveness of search engines = *Information Processing and Management*, vol. 35. (1999), p. 141-180.
- [22] Gorsuch, R. L.: *Factor Analysis*. Hillsdale: Lawrence Erlbaum Associates, 1983.
- [23] *Góliát tájékoztató segédlete*
<http://info.goliat.hu/haszn.htm> 2007.03.10.
- [24] Greenacre, M. J.: *Theory and applications of correspondence analysis*. London: Academic Press, 1984.
- [25] Gruber, T.: A translation approach to portable ontology specifications = *Knowledge Acquisition*, vol. 5. no. 2. (1993), p. 199-220.
- [26] Hair, J. F.–Anderson, R. E.–Tatham, R. L.–Black, W. C.: *Multivariate Data Analysis*. Prentice-Hall, 1998.
- [27] Harter, S. P.: Variations in relevance assessments and the measurement of retrieval effectiveness = *Journal of the American Society for Information Science*, vol. 47. (1996), p. 37-49.
- [28] Harter, S. P.–Hert, C. A.: Evaluation of information retrieval systems: Approaches, issues and methods = *Annual Review of Information Science and Technology*, vol. 32. (1997), p. 3-79.
- [29] Hawking, D.–Craswell, N.: Very large scale retrieval and web search. In.: *TREC: Experiment and evaluation in information retrieval* /Ellen Voorhees, Donna Harman editors. MIT Press, 2005.
http://es.csiro.au/pubs/trecbook_for_website.pdf 2007.07.10.
- [30] *Heuréka tájékoztató segédlete*
<http://www.heureka.hu/heureka/tippekek.html> 2007.03.09.
- [31] Hódi Imre [*Személyes közlés*] 2007.08.13.
- [32] Horváth T.–Sütheő P.: A tartalmi feltárás. In.: *Könyvtárosok kézikönyve*. 2. köt. *Feltárás és visszakeresés*. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2001.

- [33] Koster, M.: *A method for web robots control. Technical report*, Internet Engineering Task Force (IETF), 1996.
<http://www.robotstxt.org/wc/norobots-rfc.html> 2007.07.10.
- [34] Kurzor tájékoztató segédlete
<http://www.kurzor.hu/help/help.htm> 2007.03.10.
- [35] Landoni, M.–Bell, S.: Information retrieval techniques for evaluating search engines: A critical overview = *Aslib Proceedings*, vol. 52. no. 3. (2000), p. 124-129.
- [36] Leighton, H. V.: *Performance of four World Wide Web (WWW) index services: Infoseek, Lycos, WebCrawler and WWWorm*. 1995.
<https://www.winona.edu/library/staff/vl/webind.htm> 2007.03.10.
- [37] Leighton, H. V.–Srivastava, J.: *Precision among World Wide Web search services (search engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos*. 1997.
<https://www.winona.edu/library/staff/vl/webind2/webind2.htm> 2007.03.10.
- [38] Leighton, H. V.–Srivastava, J.: First 20 precision among World Wide Web search services (search engines) = *Journal of the American Society for Information Science*, vol. 50. no. 10. (1999), p. 870-881.
- [39] Lucas, W.–Topi, H.: Form and function: The impact of query term and operator usage on web search results = *Journal of the American Society for Information Science*, vol. 53. no. 2. (2002), p. 95-108.
- [40] *Magyar információs társadalom jelentés 1998-2008.: Jelentés az elmúlt évtizedről*. Készít. a BME-UNESCO Információs Társadalom- és Trendkutató Központjának (ITTK) kutatócsoportja, GKINET és az MTA Infokommunikációs Jogi Centrum kutatói. Bp. 2007. december
http://www.ittk.hu/web/docs/ITTK_MITJ_1998-2008.pdf 2008.01.28.
- [41] Nyíri, K.: Globális társadalom, helyi kultúra. In.: *Az információs társadalom*. Összeáll. Demetrovics J., Keviczky L. Bp.: MTA, 2000. p. 43-64.
- [42] Oppenheim, C.–Morris, A.–McKnight, C.–Lowley, S.: The evaluation of WWW search engines = *Journal of Documentation*, vol. 56. no. 2. (2000), p. 190-211.
- [43] Origo-Vizsla tájékoztató segédlete
http://www.vizsla24.hu/pages/sugo_tartalomjegyzek.html 2005.08.11.
- [44] Page, L.–Brin, S.–Motwani, R.–Winograd, T.: *The pagerank citation ranking: Bringing order to the web. Technical report*, Stanford Digital Library Technologies Project, 1998.
<http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf> 2007.07.17.
- [45] Pajor, E.: *A láthatatlan/mély web felhasználása a könyvtári tájékoztatásban. [Doktori (PhD) értekezés]*. Bp.: ELTE BTK, 2006. 214 p.
<http://www.szilleri.tvn.hu/> 2008.01.15.
- [46] Parasuraman, A.–Zeithaml, V. A.–Berry, L. L.: *Delivering quality service: Balancing customer perceptions and expectations*. New York: Free Press, 1990.

- [47] Parasuraman, A.–Zeithaml, V. A.–Berry, L. L.: Refinement and reassessment of the SERVQUAL scale = *Journal of Retailing*, vol.67. no. 4. (1991), p. 420-450.
- [48] Prószyky, G.–Novák, A.: Computational morphologies for small uralic languages. In.: *Inquiries into words, constraints and contexts*. Antti Arppe et al. (Eds.) p. 116-125.
<http://csli-publications.stanford.edu/koskenniemi-festschrift/12-proszeky-novak.pdf> 2008.01. 15.
- [49] Rácz, Á.: A kiadványok bibliográfiai számbavétele; leíró katalogizálás. In.: *Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés*. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2003.
- [50] Radev, D. R.–Libner, K.–Fan, W.: Getting answers to natural language questions on the Web = *Journal of the American Society for Information Science*, vol. 53. no. 5. (2002), p. 359-364.
- [51] Raghavarao, D.: *Constructions and combinatorial problems in design of experiments*. Dover Publications, 1988.
- [52] Rummel, R. J.: *Applied Factor Analysis*. Evanston: Northwestern University Press, 1970.
- [53] Rutkovszky, E.–Rutkovszky, Á.: *A láthatatlan web keresése*. (2003). [Előadásanyag]
<https://nws.niif.hu/ncd2003/docs/ehu/EHU-61.htm> 2008.01.15.
- [54] Schwartz, C.: Web search engines = *Journal of the American Society for Information Science*, vol. 49. no. 11. (1998), p. 973-982.
- [55] Scime, A.: *Web mining: Applications and techniques*. Idea Group Inc. (IGI), 2005. 427 p.
- [56] Sherman, C.: *The state of search engine marketing*. (2006).
<http://searchenginewatch.com/showPage.html?page=3575926> 2008.01.20.
- [57] Siegal, S.–Castellan, N. J.: *Nonparametric statistics for the Behavioral Sciences*. Singapore: McGraw-Hill, 1988.
- [58] Silverstein, C.–Henzinger, M.–Marais, J.–Moricz, M.: Analysis of a very large web search engine query log = *SIGIR Forum*, vol. 33. no. 1. (1999), p. 6-12.
- [59] Skrop, A.: *New methods in web information retrieval effectiveness = Új módszerek a webes információ-visszakeresés hatékonyságának területén. [PhD dissertation]*. Veszprém: Pannon Egyetem, Műszaki Informatikai Kar, 2006. 138 p.
- [60] Spink, A.–Greisdorf, H.: Regions and levels: Measuring and mapping users' relevance judgments = *Journal of the American Society for Information Science*, vol. 52. no. 2. (2001), p. 161-173.
- [61] Spink, A.–Jansen, B. J.: A study of web search trends = *Webology*, vol. 1. no. 2. (2004)
<http://www.webology.ir/2004/v1n2/a4.html> 2008.01.27.
- [62] Sroka, M.: Web search engines for Polish information retrieval: Questions of search capabilities and retrieval performance = *International Information & Library Review*, vol. 32. (2000), p. 87-98.

- [63] Su, L. T.: Developing a comprehensive and systematic model of user evaluation of Web-based search engines. In.: *National Online Meeting: Proceedings*, (1997), p. 335-345. Medford, NJ: Information Today
- [64] Sullivan, D.: *Checking your listing in search engines*, October 2001.
<http://searchenginewatch.com/webmasters/article.php/2167861> 2007.07.10.
- [65] Szász, B.–Saraniva, A.–Bognár, K.–Unzeitig, M.–Karjalainen, M.: *Cultural heritage on the semantic web – the Museum24 project*. (2006). [Előadásanyag]. 10 p.
<http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/articles/Szasz-museum24Paris.pdf> 2008.01.14.
<http://www.museo24.fi> (“Museo24” portál honlapja) 2008.01.14.
- [66] Szeredi, P.: *Ontológiák – egy matematikus-informatikus szemével* = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. [Előadásanyag]. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> 2008.01.10.
- [67] Szeredi, P.–Lukácsy, G.–Benkő, T.–Nagy, Zs.: A szemantikus világháló. In.: *A szemantikus világháló elmélete és gyakorlata*. Szerz. Szeredi P., Lukácsy G., Benkő T., Nagy Zs. Bp.: Typotex, 2005. p. 17-59.
- [68] Taube, M.: A note on the pseudo-mathematic of relevance = *American Documentation*, vol. 16. (1965), p. 69-72.
- [69] *Taxaurusz: Növény- és állatrendszertani fogalmak makrotezaurusza: 3.0 változat*. [Összeáll.] Ungváry R.; [közr. az] Országos Széchényi Könyvtár. Bp.: OSZK, 2000. XVII, 4, 16, 98 p.
<http://mek.oszk.hu/00000/00077> 2007.04.30.
- [70] Tinsley, H. E. A.: *Handbook of applied multivariate statistics and mathematical modeling*. Academic Press, 2000.
- [71] Tóth, E.: Statistical methods in measuring search engine performance = *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, vol. 20. no. 1. (2004), p. 105-113.
http://www.emis.de/journals/AMAPN/vol20_1/12.html 2008.01.27.
- [72] Tóth, E.: Exploring the capabilities of English and Hungarian search engines for various queries = *Libri*, vol. 56. no. 1. (2006), p. 38-47.
- [73] Tóth, E.: Kísérlet az internetes keresők nyelvi lehetőségeinek felmérésére = *Könyvtári Figyelő*, 53. évf. 2. sz. (2006), p. 303-311.
- [74] Ungváry, R.: A tartalom szerinti információkeresés az interneten: I. indexelőszolgáltatások = *Tudományos és Műszaki Tájékoztatás*, 47. évf. 1. sz. (2000), p. 3-17.
http://tmt.omikk.bme.hu/show_news.html?id=1624&issue_id=15 2008.01.27.
- [75] Ungváry, R.: A tartalom szerinti információkeresés az interneten: II. internetkatalógusok = *Tudományos és Műszaki Tájékoztatás*, 47. évf. 2. sz. (2000), p. 55-67.
http://tmt.omikk.bme.hu/show_news.html?id=1625&issue_id=16 2008.01.27.

- [76] Ungváry R.: Az információkeresés értékelése. In.: *Oszályozás és információkeresés: kommentált szöveggyűjtemény. 2. köt. Az információkeresés és elmélete.* Szerk. Ungváry R., Orbán É. Bp.: OSZK, 2001.
<http://mek.oszk.hu/01600/01683/pdf/01683-2.pdf> 2007.11.17.
- [77] Ungváry, R.: Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása = *Tudományos és Műszaki Tájékoztatás*, 51. évf. 5. sz. (2004), p. 175-191.
http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450 2008.01.27.
- [78] Ungváry, R.: Az ontológia fogalma, avagy az eltűnt tezaurusz = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. [Előadásanyag]. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> 2008.01.10.
- [79] Ungváry, R.–Vajda, E.: *Könyvtári információkeresés. 2. jav. kiad.* Bp.: Typotex, 2002. 170 p.
- [80] *A világ előrehaladása az információs társadalom terén 1998-2008.: World Progress Report 2008.* Készít. a BME-UNESCO Információs Társadalom- és Trendkutató Központjának (ITTK) kutatócsoportja. Bp. 2007. március
http://www.ittk.hu/web/docs/ITTK_WPR1998-2008.pdf 2008.01.28.
- [81] Wang, H.–Xie, M.–Goh, T. N.: Service quality of Internet search engines = *Journal of Information Science*, vol. 25. no. 6. (1999), p. 499-507.
- [82] Xie, M.–Wang, H.–Goh, T. N.: Quality dimensions of Internet search engines = *Journal of Information Science*, vol. 24. no. 5. (1998), p. 87-94.
- [83] *XML-Topic-Map (XTM) Standard*, ISO/IEC 13250: 2000. XTM TopicMaps Org.
<http://www.topicmaps.org/xtm> 2008.01.14.
- [84] Zádori Janka [Személyes közlés] 2007.08.10.

V. Publikációk jegyzéke

Hivatkozások:

1. Helyzetkép az internetes keresők értékeléséről = Könyvtári Figyelő, 47. évf. 2. sz. (2001.), p. 301-307.
Hiv.: Poprády G.: Könyvtári trendek. In: Könyvtárosok kézikönyve. 5. Segédletek. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2003. p.13-62.
2. Az automatikus osztályozással kapcsolatos kísérleti kutatások eredményei = Tudományos és Műszaki Tájékoztatás, 49. évf. 5. sz. (2002.), p. 191-198.
Hiv.: Poprády G.: Könyvtári trendek. In: Könyvtárosok kézikönyve. 5. Segédletek. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2003. p.13-62.

Referált publikációk:

3. Helyzetkép az internetes keresők értékeléséről = Könyvtári Figyelő, 47. évf. 2. sz. (2001.), p. 301-307.
DIALOG H. W. Wilson Record Number: BLIB02008916
4. Innovative solutions in automatic classification: a brief summary = Libri, vol. 52. no. 1., (2002.), p. 48-53.
DIALOG H. W. Wilson Record Number: BLIB02004470
5. Adatbányászatra irányuló törekvések a könyvtári területen = Könyvtári Figyelő, 48. évf. 3. sz. (2002.), p. 527-532.
DIALOG H. W. Wilson Record Number: BLIB03102651
6. Az automatikus osztályozással kapcsolatos kísérleti kutatások eredményei = Tudományos és Műszaki Tájékoztatás, 49. évf. 5. sz. (2002.), p. 191-198.
LISA, MAKSZAB
7. Statistical methods in measuring search engine performance = Acta Mathematica Academiae Paedagogicae Nyíregyháziensis, vol. 20. no. 1. (2004.), p. 105-113.
ZENTRALBLATT ZBL 1065.68018
8. Exploring the capabilities of English and Hungarian search engines for various queries = Libri, vol. 56. no. 1. (2006.), p. 38-47.
DIALOG 04376483 Genuine Article: 048VZ
9. Hogyan mérhető az internetes keresők szolgáltatása? = Híradástechnika, 61. évf. 1. sz. (2006.), p. 35-40.
10. Kísérlet az internetes keresők nyelvi lehetőségeinek felmérésére = Könyvtári Figyelő, 53. évf. 2. sz. (2006.), p. 303-311.
DIALOG H. W. Wilson Record Number: BLIB6110598

Egyéb publikációk:

11. Eszenyiné B. M., Tóth E.: Quality Management Techniques and Informatics = Problemü ekonomicsnovo ta szosalnava rozvitku regionu i praktika naukova eksperimentu, Kijev-Uzsgorod-Nyíregyháza: Karpati Kiadó, 2000. p. 276-280.
12. Eszenyiné B. M., Tóth E.: Quality Management Techniques and Informatics = Informatizacija dialnoszti pidpriemctv malovo ta szerednovo biznesu: mehanizm, problemi, rozbitok, "Naukovij visznyik Uzsgorodszkava Gyerzsavnovo Universzityetu" Szerija "Ekonomika", No. 5., Uzsgorod, 2000. p. 64-69.
13. Az internetes keresők működésének minőségi vizsgálata = Természettudományi Közlemények, Nyíregyháza: Nyíregyházi Főiskola, (2005.), p. 427-438.

14. Quality and search engines' retrieval = Természettudományi Közlemények, Nyíregyháza: Nyíregyházi Főiskola, (2007.), p. 125-134.

Konferenciák kiadványaiban megjelent publikációk:

15. An overview of the experimental methods used in automatic classification = MicroCAD 2002, International Scientific Conference Miskolc, Hungary, 7-8. March 2002., Volume of Section H: Applied Information Engineering p. 205-209. Full paper submitted
16. Szabó Sz., Kormos J., Tóth E.: Társadalmi informatika és a rendszerfejlesztés interdiszciplináris megközelítései = Informatika a felsőoktatásban 2002, Debrecen: Debreceni Egyetem, 2002. aug. 28-30., 6 oldal, cd kiadvány
17. Adatbányászattal kapcsolatos kérdések a könyvtári területen = Informatika a felsőoktatásban 2002, Debrecen: Debreceni Egyetem, 2002. aug. 28-30., 5 oldal, cd kiadvány
18. The application of intelligent search agents = MicroCAD 2003, International Scientific Conference Miskolc, Hungary, 6-7. March 2003., Volume of Section N: Applied Information Engineering p. 151-157. Full paper submitted
19. Tóth E., Fazekas G.: Statistical methods used in search engine evaluations, Proceedings of the 6th International Conference on Applied Informatics (ICAI04), Eger, Hungary, 27-31. January 2004. Volume I. p. 155-162.
20. Az internetes keresők szolgáltatásának minőségi vizsgálata = Informatika a felsőoktatásban 2005, Debrecen: Debreceni Egyetem, 2005. augusztus 24-26., 6 oldal, cd kiadvány

Konferencia előadások:

21. Internetes keresők használata a tartalomszolgáltatásban, Magyar Tudomány Napja 2000. alkalmából rendezett Szabolcs-Szatmár-Bereg Megyei Tudományos Konferencia, Nyíregyháza, 2000. november 7.
22. Információkeresés az interneten, MKE Műszaki Könyvtáros Szekció Orvosi Csoportja és a Magyar Orvosi Könyvtárak Szövetsége ülése, Nyíregyháza, 2001. augusztus 9.
23. Az internetes keresők értékelésével kapcsolatos kérdések, Magyar Tudomány Napja 2001. alkalmából rendezett Szabolcs-Szatmár-Bereg Megyei Tudományos Konferencia, Nyíregyháza, 2001. október 29.
24. An overview of the experimental methods used in automatic classification, MicroCAD 2002, International Scientific Conference, Miskolc, 2002. március 7.
25. Társadalmi informatika és a rendszerfejlesztés interdiszciplináris megközelítései, Informatika a felsőoktatásban 2002, Debrecen, 2002. augusztus 29.
26. Adatbányászattal kapcsolatos kérdések a könyvtári területen, Informatika a felsőoktatásban 2002, Debrecen, 2002. augusztus 30.
27. The application of intelligent search agents, MicroCAD 2003, International Scientific Conference, Miskolc, 2003. március 6.
28. Az internetes keresők mérésének egy lehetséges módszertana, MTA Szabolcs-Szatmár-Bereg Megyei Tudományos Testületének ülése, Nyíregyháza, 2003. szeptember 27.
29. Statistical methods used in search engine evaluations, 6th International Conference on Applied Informatics, Eger, 2004. január 30.

30. Az internetes keresők szolgáltatásának minőségi vizsgálata, Informatika a felsőoktatásban 2005, Debrecen, 2005. augusztus 25.
31. Az internetes keresők nyelvi lehetőségei, 30 éves a nyíregyházi könyvtárosképzés alkalmából rendezett tudományos emlékülés, Nyíregyháza, 2006. október 16.

I. Introduction

- During the examination of search engines we have to face with several difficulties, I would emphasize only some of them: the majority of the studies do not report its own applied methodology, the evaluations do not have a systematic approach, the search process can be biased in the evaluations, search engines are changing all the time and the web is a dynamic, constantly changing environment as well. Although several problems associated with search engine evaluations, we can say that we have to make an effort to measure search engines currently in use. However, we can not use standardized evaluation approaches for this task.
- The dissertation discusses the principles of search engines' operation and the tasks of their components. It emphasizes search engines' determining role in hit ranking. It briefly reviews problems and deficiencies relating to the examination of search engines. It summarizes the earliest statistical methods used in the field of information retrieval. Then it focuses on the introduction of the most significant statistical analyses, detailing their methods, findings and their applied measures. It shows how successful English and Hungarian search engines are in handling and interpreting the executed queries from a morphological aspect. It assesses the effectiveness of the Hungarian search engines from the aspect of precision and recall measures.
- The analysis of the Hungarian search engines' retrieval effectiveness is supported by the fact that only a few research studies were written so far in the literature which dealt with queries in non English language more thoroughly.
- The information retrieval studies point out that it is extremely difficult for researchers to find reliable measures for the analyses. Several times there is a great debate on which measure would be advantageous or disadvantageous for them to use during the examinations. Following from this, research on search engine evaluation is inconsistent in applied methods, so there is a real need for working out and using a standard set of measures in these examinations.
- The analysis of search engine performance is in the center of researchers' interest, because they apply various approaches and methods in the assessments. I would like to call your attention to some prominent researchers' activities in relation with this topic:

Sroka studied local search engines and the Polish versions of the English search engines. In his measurement precision was emphasized that he determined on the basis of human relevance judgements [Sroka2000].

Clarke and *Willett* measured the relevance of the first ten results for the queries. They calculated mean values for precision, recall and coverage [Clarke-Willett97].

Leighton and *Srivastava* conducted five different experiments and they evaluated the precision of the first twenty results retrieved by search engines. They worked out appropriate measures for determining the precision of the hits [Leighton-Srivastava99].

Chignell, *Gwizdka* and *Bodner* analysed the effect of time of day and the effect of query strategy on query processing time for each search engine in their first experiment. They studied the influence of geographical coverage and Internet domains on search engine performance in their second experiment [Chignell-Gwizdka-Bodner99].

Lucas and *Topi* examined the effect of query operator and term selection on the relevancy of search results [Lucas-Topi2002].

Radev, *Libner* and *Fan* dealt with the following issue: how successful the most popular search engines were in finding accurate answers to natural language questions [Radev-Libner-Fan2002].

Bar-Ilan tracked the operation of the search engines over a longer period of time. She approached the stability of their hit retrieval from the aspect of several measures [Bar-Ilan2002].

Bar-Ilan and *Gutman* analysed from a morphological aspect the interpretation of non-English queries. Both of them concentrated on how search engines took into account the specific characteristics of the analysed languages [Bar-Ilan-Gutman2005].

- Most often precision and recall are used as characteristics in evaluations, but they compete with each other. Both of them are sensitive to how we define and measure relevance. The research studies written so far took into account only precision. They did not make any attempts to measure the recall of the different search engines, only to a minimum extent.

II. Research objectives

- With the analysis discussed in the dissertation *my aim was to compare the retrieval effectiveness of the Hungarian search engines*. Precision and recall measures express mostly the effectiveness of the information retrieval systems, therefore I analysed the Hungarian search engines on the basis of these two attributes. *By precision I mean the proportion of the relevant documents within all retrieved documents. Recall means the proportion of the retrieved relevant documents compared to all (retrieved or not retrieved) relevant documents ([Ungváry2001] p. 196-197).*

- During my previous enquiry in the research field I paid attention to another aspect which focused on how search engines took into consideration the specific characteristics of the analysed languages and how effective they were in responding to non-English queries. According to this, I conducted another examination with a goal to explore the language capabilities of the English and the Hungarian search engines for various queries. *I looked for an answer to the following question: how effective search engines were in handling and interpreting queries from a morphological aspect. It was not the aim of my research to make absolute statements about the analysed search engines, but using a few test data to draw attention to the fact that these search tools behave differently concerning a given issue. In this context my examinations and statements are much rather qualitative than quantitative in their nature.* Although test data have been thoroughly chosen, it is clear that a small number of analysed examples is not suitable for drawing statistical conclusions. However, the statistical analysis of the outlined issues can be interesting during a following research.

In the morphological analysis English search engines were examined on queries in both English and Hungarian. However, Hungarian search engines were analysed only on the Hungarian queries because, except for Origo-Vizsla and Altavizsla services, they mainly index only Hungarian Web pages.

- The most important methodological steps of the conducted morphological analysis were the following:
 - *In the case of English and Hungarian languages the examination began with finding appropriate search terms.* A set of search terms was carefully constructed for testing.
 - *Then trial searches were run on each search engine to check if the results from the selected terms would reflect clearly the issue analysed and if they would correspond to the research goals. I selected those search terms that emphasized the linguistic difficulties of these languages* and I also relied on my previous observations.

- *I reviewed the help files of each search engine* to have more information about their relevant features and capabilities.
 - *The first 100 hits were examined on the queries. The only exception was the analysis of truncation in Hungarian where a parasite name called 'májmetely' (=liver fluke) was entered as a query that provided a limited set of hits.* Primarily the abstracts of the results were used for analysis, but the contents of the Web pages were also checked where necessary.
- *My analysis was conducted between July and September 2005.* Altogether three English search engines were examined: Google, AltaVista and AlltheWeb. As a test case, five significant Hungarian search engines were considered: Heuréka, Origo-Vizsla, Kurzor, Góliát and Altavizsla.
 - For testing, the aspects in *Bar-Ilan* and *Gutman's* [Bar-Ilan-Gutman2005] common study served as a basis which were the following:
 - I mean by *stemming* if the search engine retrieved any other suffixed word forms of a search phrase or not. So it was examined to determine if the search engine retrieved the plural form of a search phrase or not. The queries entered were as follows: '*dog-dogs*' (in English), '*ház-házak*' (=house-houses), '*kocsi-kocsik*' (=car-cars), '*kutya-kutyák*' (=dog-dogs). In the first two Hungarian examples the final vowel of the stem does not change in the plural form compared to the singular form. However, in the third Hungarian example it does change in the plural form compared to the singular form. A noun phrase with a privative suffix was deliberately chosen for studying stemming in Hungarian, which was the '*tisztességtelen*' (=dishonest) phrase. Using this phrase I could monitor if the search engine retrieved any other word forms of this complex phrase and applied stemming to it.
 - In the case of *stopwords*, whether the search phrase appeared with definite and indefinite articles in the results was checked to determine whether the search engine searched separately for the articles entered or not. Thus the following queries were analysed in English: '*a dog*', '*an aunt*' and '*the car*'. Concerning Hungarian the following terms were entered: '*a ház*' (=the house), '*az ember*' (=the man), '*egy kocsi*' (=a car).
 - In connection with *diacritics*, whether the search engine found the diacritical versions of the entered queries among its hits or not was concentrated on. It was a relevant aspect to analyse only in Hungarian. Here two search phrases were used, namely '*kertem*' (=my garden) and '*alma*' (=apple).
 - For *truncation* an asterisk (*) was utilized after the search term in each case. In English the query '*Olympi**' was entered to find all those Web pages that are about the Olympic

Games containing the following phrases: ‘Olympic’, ‘Olympics’, ‘Olympia’, ‘Olympian’, etc. *In Hungarian a limited set of hits was chosen to find more easily the suffixed stem forms of the query and to check their real appearance within the set of hits.* By entering the query ‘*májmétely**’ the suffixed stem forms of ‘*májmétely*’ (=liver fluke) parasite were retrieved. In addition to this, the sets of hits retrieved by the ‘*májmétely**’ query through all search engines were examined to check if they really included the suffixed stem forms of the query.

- Finally, whether the search engines retrieved the *synonyms* of a query or not was tested. How the synonyms were visible in the hits was also studied, for example if they were highlighted from the abstracts, if they appeared together with the query in the abstracts, or if they were available alone. The following English queries were examined from this aspect: ‘*car*’, ‘*glasses*’. In Hungarian the synonyms of ‘*kutya*’ (=dog) and ‘*vetélkedő*’ (=contest) were searched for.

So the language solutions of the search engines were examined on the basis of these aspects for various queries. With this method for the analysis two of my constructed hypotheses can be studied which can be summarized in the following way: *the performance of the Hungarian and the English search engines can be analysed and differs in the respect of their implemented language capabilities.*

- *I carried out another analysis focusing on precision and recall between January and June 2006. In the examination of precision I considered only the first sixty hits for the executed queries.* My decision was supported by the software developers of the Google who claimed that in the development of Google the main principle was that the most relevant documents should appear at the beginning of the hitlist on each query because users mostly were willing to look at only the first few tens of results in the hitlist during a search ([Brin-Page98] p. 3).

- *I analysed precision from the aspect of usage and effectiveness. Usage means for me that how many relevant Web pages are retrieved by the examined search engines in the first, second and third twenty hits for various queries.* It is an essential question for users which search engines they should use for finding relevant information. *I mean by effectiveness how equally search engines provide relevant documents in the first, second and third twenty hits or in the first sixty hits for the executed queries.* I determined effectiveness on the basis of the mean and standard deviation values of the relevant hits retrieved by the search engines. *I calculated a percentage value for the precision of the hits that reflected how many relevant hits I received from the retrieved first, second and third twenty results.* Concerning each query the change of the hit precision values was studied in the first,

second and third twenty results. On the basis of these examinations I received different findings for precision. With this complex method for the analysis two of my hypotheses can be examined which are the following: *the results at the beginning of the hitlist are the most precise, after that the precision of the following hits decreases gradually; as a result of the analysis I obtain those Hungarian search engines that accomplish precision in their hit retrieval to the greatest or to the smallest extent.*

- Then I examined whether it is possible to determine a convincing conclusion for the number of relevant hits in the remaining part of the hitlist. With this examination I rejected the following hypothesis: *precision changes in the case of each analysed search engine and some kind of conclusion can be given for the variations in precision which is at the same time valid for the whole set of hits.*

- The analysis of recall was done on the basis of relative recall known from the literature which can be calculated in the following way: the number of relevant documents retrieved by a given search engine is divided by the total number of relevant documents found by at least one of all the analysed search engines ([Clarke-Willett97] p. 185-187). According to this, I created a set of relevant results from all the retrieved hits of the search engines in 'csiperkegomba' (=mushroom) subject and I examined a given search engine's performance in retrieving this set of relevant documents. A percentage value was calculated for relative recall which expressed that how many percentages of the relevant pages from this set were found by a given search engine. On the basis of these percentage values I made a ranking list among the analysed search engines. With this method the following hypothesis can be studied: *as a result of the analysis I can evaluate the Hungarian search engines' performance in retrieving the set of relevant hits, i.e. to what extent their set of hits contain these relevant results.*

- At the beginning of the analysis I defined on what conditions I considered a Web page relevant. However, I did not create any relevance categories for determining the relevance of the documents.

- Altogether five Hungarian search engines were analysed which were the following: Heuréka, Origo-Vizsla, Kurzor, Góliát and Altavizsla. In five subjects queries only in Hungarian were executed on these search tools. My chosen topics were as follows: 'csiperkegomba' (=mushroom), 'Rákóczi-szabadságharc' (=Rákóczi's fight for freedom), 'evészavar' (=eating disorder), 'kegyhely' (=shrine), 'szökőár' (=tidal wave). Only one of them focused on analysing recall ('csiperkegomba' =mushroom) and the other four subjects put an emphasis on studying relevance. I constructed a

complicated query for analysing recall that completely took into consideration the special characteristics of the given topic. Relating to 'csiperkegomba' (=mushroom) subject the following queries were determined: 'sampinyon' (=champignon), 'termesztett csiperke' (=cultivated mushroom), 'kétspórás csiperke' (common mushroom) and 'Agaricus Bisporus' (the Latin name of common mushroom). It is important to note that *the selection of the appropriate queries always requires some background information*. In the design of the examination I relied on my previous observations.

III. Results

The findings of two empirical analyses are summarized on the basis of my constructed hypotheses and theses arising from the results are also presented.

III.1. The performance of the English search engines can be assessed from the aspect of their implemented language capabilities.

The results of this study let me draw a conclusion that English search engines handle queries for Hungarian terms worse than they handle English terms.

In these search tools truncation and searching for synonyms work properly in English but it is problematic in Hungarian.

Google omits definite and indefinite articles in the English queries, but it has not solved this issue yet for Hungarian queries. AltaVista is good at finding the plural form of the English queries, so this is the only service where stemming works properly in English. However, none of the English services coped with this problem in the Hungarian language. It can be said that Google and AltaVista are equal in their performance of interpreting queries, followed by AlltheWeb. On the basis of these findings, I can conclude that more emphasis should be placed on stemming, handling of stopwords and diacritics in the future development of these search engines.

Thesis 1.: In English search tools truncation and searching for synonyms work properly in English but it is not efficient in Hungarian.

Thesis 2.: English search engines have to improve their functionality in stemming, handling of stopwords and diacritics in their future development.

Methodologically a recent value in this statement is that I examined queries in English and Hungarian on the English search engines from the same aspect. Thus it was possible for me to compare the performance of the English search engines concerning both languages. According to my knowledge, searching for synonyms is a new approach, which has not occurred yet in the earlier analyses.

III.2. The performance of the Hungarian search engines can be analysed and differs in the respect of their implemented language capabilities.

The majority of the Hungarian search engines handle diacritics efficiently, and in this way they meet an essential user requirement.

If I evaluate the linguistic capabilities of Hungarian search engines I can say that Heuréka provided the best performance in fields such as stemming and searching for synonyms. There were two other fields where its performance was acceptable, namely handling of stopwords and diacritics.

The next best performance was achieved by Góliát and Altavizsla, which showed very similar performance in the tests. They handled diacritics precisely, but they need to improve their functionality in dealing with stopwords.

In this ranking Origo-Vizsla and Kurzor are last with a similar level of performance. Origo-Vizsla showed inadequacies in several fields such as stemming, truncation, handling of diacritics and stopwords.

Kurzor had deficiencies in the following fields: stemming, handling of stopwords, truncation and searching for synonyms. The results of these tests have established that almost every Hungarian search tool has to improve its performance in the field of truncation and stemming.

***Thesis 3.:** The majority of the Hungarian search engines handle diacritics efficiently.*

***Thesis 4.:** But almost every one of them has to develop its performance in the field of truncation and stemming.*

Searching for the suffixed stem forms of the query made possible the thorough analysis of truncation in Hungarian. For this approach I have not found an example yet during the analyses conducted so far.

III.3. Precision changes in the case of each analysed search engine and some kind of conclusion can be given for the variations in precision, which is at the same time valid for the whole set of hits.

I note that precision changed to a minimal extent in the case of Góliát and Altavizsla services, because they retrieved almost the same number of relevant results in the first, second and third twenty hits for all the four queries. Besides this both search tools used the same index for retrieval confirmed by the results of the analysis as well as, because they retrieved the same relevant documents for all the five topics except for one Web page.

It turned out from the data received that I have not found any relationship with evidence level between the number of relevant results in the first sixty hits and the total number of results retrieved for the queries. For this reason I could not determine a convincing conclusion for the number of relevant results in the remaining part of the hitlist. So I had to reject this hypothesis.

III.4. Results at the beginning of the hitlist are the most precise, after that the precision of the following hits decreases gradually.

The majority of the analysed search engines made an effort to retrieve the greatest quantity of the relevant results in the first twenty hits for all the four queries.

In the first sixty hits they retrieved relevant results with three kinds of tendencies which were the following: strictly monotonous reducing, monotonous reducing and not monotonous. In this case I

studied search engines from the aspect of usage, where I took into consideration only the quantity of the relevant results retrieved in the first, second and third twenty hits.

Analysing search engines from the aspect of efficiency I can say that in the case of all the four queries in the first twenty hits they proved to be the best, because the mean of their relevant results was the highest at this place.

However, the efficiency of search engines reduced to a great extent in the second and third twenty hits for all the four queries, because the average of their relevant results was very low at these places. In addition to this, a minimal difference was experienced in the efficiency of search engines in the second and third twenty hits for all the four queries, since a slight difference appeared between the means of their relevant results at these places.

***Thesis 5.:** The majority of the Hungarian search engines have retrieved the greatest amount of the relevant results in the first twenty hits for all the four queries.*

***Thesis 6.:** Besides this, they proved to be the most efficient in the first twenty hits for all the four queries, since the mean of their relevant results was the highest at this place.*

I have approached the Hungarian search engines from the aspect of usage that was an essential issue for the users. The result received from the aspect of usage was also supported by the means and the standard deviations of the relevant hits from the aspect of efficiency.

III.5. As a result of the analysis I obtain those Hungarian search engines which accomplish precision in their hit retrieval to the greatest or to the smallest degree.

From the aspect of usage Kurzor and Origo-Vizsla turned out to be the best for the users, because both of them made an effort to retrieve the greatest number of relevant results in the first twenty hits for the executed queries. If I lay an emphasis on the equal quality of the search engines' hit retrieval in the first sixty results and I do not focus only on the quantity of their retrieved relevant hits at this place, then I receive different results from this latter one. I note that the aspect of efficiency expresses for me the equal quality of the search engine's hit retrieval that I obtain by considering the means and the standard deviations of the relevant results.

Origo-Vizsla proved to be the most efficient in the first sixty hits for all the four queries, because the mean of its relevant results was the highest compared to that of the other search engines.

From the aspect of efficiency Heuréka was the following one in the ranking, but the average of its relevant results was much smaller than that of Origo-Vizsla in the first sixty hits for all the four queries.

From the same aspect Kurzor occupied the third place in the ranking, since the mean of its relevant results was qualified as low – considering standard deviation – in the first sixty hits for all the four queries.

Finally, Góliát and Altavizsla services were the last in the ranking, but their efficiency was not different from each other, because the means and the standard deviations of their relevant results were almost the same in the first sixty hits for all the four queries.

Thesis 7.: I can rank the Hungarian search engines on the basis of their efficiency, taking into account the means and the standard deviations of their relevant results in the first sixty hits for all the four queries.

Efficiency here reflects for me how equally the analysed search engines can retrieve the relevant documents in the first sixty hits. According to this, I have examined precision from a new aspect and I have not studied it only from the aspect of usage.

III.6. As a result of the analysis I can evaluate the Hungarian search engines' performance in retrieving the set of relevant hits, i.e. to what extent their set of hits contain the relevant results.

On the basis of the percentage values calculated for relative recall I made the following ranking list among the Hungarian search engines: Origo-Vizsla retrieved the greatest amount of relevant hits from the analysed set, because it discovered 42.85% of them on the web. The next best performance was achieved by Kurzor, Góliát and Altavizsla with 23.80%, because they retrieved the same amount of relevant hits from this set. At last Heuréka lagged behind slightly from the retrieval performance of the previous three search engines, because it found 19.04% of the relevant documents in the set on the web.

It is a remarkable index for me that Hungarian search engines have found 5.6 relevant documents on average on the web in 'mushroom' topic that corresponded to 26.66% of the twenty-one hits in the set. So Hungarian search engines have retrieved more than a quarter of the analysed set on average in 'mushroom' topic, which is qualified as rather moderate performance.

Thesis 8.: The retrieval performance of the Hungarian search engines can be ranked on the basis of relative recall. They have retrieved more than a quarter of the set of relevant hits on average in 'mushroom' topic.

Previously others also analysed search engines from the aspect of relative recall. This result can be an interesting contribution to the research studies in this field concerning Hungarian search engines.

III.7. Further research tasks

- The formulated theses can serve as basis for a further, more detailed statistical analysis. In my examinations I made an effort to construct a set of test keys that can be carefully enlarged for a further statistical analysis. It would be worth comparing the results arising from my examinations with that of another, further analysis.

- In my opinion the findings of research projects in the field of semantic web would be well exploited for a new analysis that are related to the meaning of queries and documents to be found.
- I plan and also initiate a project with Finnish and Estonian researchers who would like to accomplish a relevant information retrieval of quality in their search services. In addition to this, as far as possible I make an effort to develop a project with a firm that builds on my explored findings.
- I think it is necessary to initiate my students in the analyses and research activities related to search engine effectiveness. It is also important to make them acquainted with the developments and scientific results of this field. (So far I had an opportunity for this in the framework of theses, essays and tasks).

IV. References

- [1] Albert, L.–Locsmándi, Cs.–Vasas, G.: *Ismerjük fel a gombákat!* Gabó, 1995. 191 p.
- [2] *AlltheWeb tájékoztató segédlete*
<http://www.alltheweb.com/help/index> 2007.03.10.
- [3] *Altavista tájékoztató segédlete*
<http://www.altavista.com/help/search/default> 2007.03.10.
- [4] Bar-Ilan, J.: Methods for measuring search engine performance over time = *Journal of the American Society for Information Science*, vol. 53. no. 4. (2002), p. 308-319.
- [5] Bar-Ilan, J.: Criteria for evaluating information retrieval systems in highly dynamic environments. In.: *Proceedings of the 2nd International Workshop on Web Dynamics*, Honolulu, Hawaii, 7th May 2002.
<http://www.dcs.bbk.ac.uk/webDyn2/onlineProceedings.html> 2007.03.02.
- [6] Bar-Ilan, J.–Gutman, T.: How do search engines respond to some non-English queries? = *Journal of Information Science*, vol. 31. no. 1. (2005), p. 13-28.
- [7] Benczúr, A.–Bíró, I.–Csalogány, K.–Rácz B.–Sarlós T.–Uher, M.: PageRank és azon túl: Hiperhivatkozások szerepe a keresésben = *Magyar Tudomány*, 167. évf. 11. sz. (2006), p. 1325-1331.
<http://www.matud.iif.hu/06nov/07.html> 2007.07.17.
- [8] Berners-Lee, T.–Fielding, R.–Frystyk, H.: *Hypertext Transfer Protocol – HTTP/1.0. (RFC1945)*. May 1996.
<http://www.rfc-editor.org> 2007.08.10.
- [9] Bilodeau, M.–Brenner, D.: *Theory of multivariate statistics*. Springer, 1999.
- [10] Bocsán Róbert [*Személyes közlés*] 2007.09.04.
- [11] Brin, S.–Page, L.: The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, vol. 30. no. 1-7. (1998), p. 107-117.
<http://infolab.stanford.edu/pub/papers/google.pdf> 2007.03.02.
- [12] Chignell, M. H.–Gwizdka, J.–Bodner, R. C.: Discriminating meta-search: A framework for evaluation = *Information Processing and Management*, vol. 35. (1999), p. 337-362.
- [13] Clarke, S. J.–Willett, P.: Estimating the recall performance of web search engines = *Aslib Proceedings*, vol. 49. no. 7. (1997), p. 184-189.
- [14] *Creating a European library space: Telematics for libraries programmes 1990-1998*.
<http://cordis.europa.eu/libraries/en/intro.html> 2008.01.22.
- [15] Cronbach, L. J.: Coefficient alpha and the internal structure of tests = *Psychometrika*, vol. 16. (1951), p. 257-334.

- [16] Ding, W.–Marchionini, G.: A comparative study of web search performance in global complexity: information, chaos and control. In.: *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, Baltimore, Maryland, 21-24 October 1996. p. 136-142.
- [17] Fodor, I.: Merre megy a világ gazdasága, merre mehetünk mi? In.: *Az információs társadalom*. Összeáll. Demetrovics J., Keviczky L. Bp.: MTA, 2000. p. 95-113.
- [18] Friedman, E.–Uher, M.–Windhager, E.: Keresés a világhálón = *Híradástechnika*, 58. évf. 3. sz. (2003), p. 20-24.
<http://www.ilab.sztaki.hu/websearch-data/Publications/kereses.pdf> 2007.07.17.
- [19] Froehlich, T. J.: Case study 5.1: Developing search engine evaluation criteria. In.: *Library evaluation*. Libraries Unlimited, 2001. p. 185-200.
- [20] *Google tájékoztató segédlete*
<http://www.google.com/support> 2007.03.10.
- [21] Gordon, M.–Pathak, P.: Finding information of the World Wide Web: The retrieval effectiveness of search engines = *Information Processing and Management*, vol. 35. (1999), p. 141-180.
- [22] Gorsuch, R. L.: *Factor Analysis*. Hillsdale: Lawrence Erlbaum Associates, 1983.
- [23] *Góliát tájékoztató segédlete*
<http://info.goliat.hu/haszn.htm> 2007.03.10.
- [24] Greenacre, M. J.: *Theory and applications of correspondence analysis*. London: Academic Press, 1984.
- [25] Gruber, T.: A translation approach to portable ontology specifications = *Knowledge Acquisition*, vol. 5. no. 2. (1993), p. 199-220.
- [26] Hair, J. F.–Anderson, R. E.–Tatham, R. L.–Black, W. C.: *Multivariate Data Analysis*. Prentice-Hall, 1998.
- [27] Harter, S. P.: Variations in relevance assessments and the measurement of retrieval effectiveness = *Journal of the American Society for Information Science*, vol. 47. (1996), p. 37-49.
- [28] Harter, S. P.–Hert, C. A.: Evaluation of information retrieval systems: Approaches, issues and methods = *Annual Review of Information Science and Technology*, vol. 32. (1997), p. 3-79.
- [29] Hawking, D.–Craswell, N.: Very large scale retrieval and web search. In.: *TREC: Experiment and evaluation in information retrieval* /Ellen Voorhees, Donna Harman editors. MIT Press, 2005.
http://es.csiro.au/pubs/trecbook_for_website.pdf 2007.07.10.
- [30] *Heuréka tájékoztató segédlete*
<http://www.heureka.hu/heureka/tippekek.html> 2007.03.09.
- [31] Hódi Imre [*Személyes közlés*] 2007.08.13.
- [32] Horváth T.–Sütheő P.: A tartalmi feltárás. In.: *Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés*. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2001.

- [33] Koster, M.: *A method for web robots control. Technical report*, Internet Engineering Task Force (IETF), 1996.
<http://www.robotstxt.org/wc/norobots-rfc.html> 2007.07.10.
- [34] Kurzor tájékoztató segédlete
<http://www.kurzor.hu/help/help.htm> 2007.03.10.
- [35] Landoni, M.–Bell, S.: Information retrieval techniques for evaluating search engines: A critical overview = *Aslib Proceedings*, vol. 52. no. 3. (2000), p. 124-129.
- [36] Leighton, H. V.: *Performance of four World Wide Web (WWW) index services: Infoseek, Lycos, WebCrawler and WWWorm*. 1995.
<https://www.winona.edu/library/staff/vl/webind.htm> 2007.03.10.
- [37] Leighton, H. V.–Srivastava, J.: *Precision among World Wide Web search services (search engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos*. 1997.
<https://www.winona.edu/library/staff/vl/webind2/webind2.htm> 2007.03.10.
- [38] Leighton, H. V.–Srivastava, J.: First 20 precision among World Wide Web search services (search engines) = *Journal of the American Society for Information Science*, vol. 50. no. 10. (1999), p. 870-881.
- [39] Lucas, W.–Topi, H.: Form and function: the impact of query term and operator usage on web search results = *Journal of the American Society for Information Science*, vol. 53. no. 2. (2002), p. 95-108.
- [40] *Magyar információs társadalom jelentés 1998-2008.: Jelentés az elmúlt évtizedről*. Készít. a BME-UNESCO Információs Társadalom- és Trendkutató Központjának (ITTK) kutatócsoportja, GKINET és az MTA Infokommunikációs Jogi Centrum kutatói. Bp. 2007. december
http://www.ittk.hu/web/docs/ITTK_MITJ_1998-2008.pdf 2008.01.28.
- [41] Nyíri, K.: Globális társadalom, helyi kultúra. In.: *Az információs társadalom*. Összeáll. Demetrovics J., Keviczky L. Bp.: MTA, 2000. p. 43-64.
- [42] Oppenheim, C.–Morris, A.–McKnight, C.–Lowley, S.: The evaluation of WWW search engines = *Journal of Documentation*, vol. 56. no. 2. (2000), p. 190-211.
- [43] Origo-Vizsla tájékoztató segédlete
http://www.vizsla24.hu/pages/sugo_tartalomjegyzek.html 2005.08.11.
- [44] Page, L.–Brin, S.–Motwani, R.–Winograd, T.: *The pagerank citation ranking: Bringing order to the web. Technical report*, Stanford Digital Library Technologies Project, 1998.
<http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf> 2007.07.17.
- [45] Pajor, E.: *A láthatatlan/mély web felhasználása a könyvtári tájékoztatásban. [Doktori (PhD) értekezés]*. Bp.: ELTE BTK, 2006. 214 p.
<http://www.szilleri.tvn.hu/> 2008.01.15.
- [46] Parasuraman, A.–Zeithaml, V. A.–Berry, L. L.: *Delivering quality service: Balancing customer perceptions and expectations*. New York: Free Press, 1990.

- [47] Parasuraman, A.–Zeithaml, V. A.–Berry, L. L.: Refinement and reassessment of the SERVQUAL scale = *Journal of Retailing*, vol.67. no. 4. (1991), p. 420-450.
- [48] Prószéky, G.–Novák, A.: Computational morphologies for small uralic languages. In.: *Inquiries into words, constraints and contexts*. Antti Arppe et al. (Eds.) p. 116-125.
<http://csli-publications.stanford.edu/koskenniemi-festschrift/12-proszeky-novak.pdf> 2008.01. 15.
- [49] Rácz, Á.: A kiadványok bibliográfiai számbavétele; leíró katalogizálás. In.: *Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés*. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2003.
- [50] Radev, D. R.–Libner, K.–Fan, W.: Getting answers to natural language questions on the Web = *Journal of the American Society for Information Science*, vol. 53. no. 5. (2002), p. 359-364.
- [51] Raghavarao, D.: *Constructions and combinatorial problems in design of experiments*. Dover Publications, 1988.
- [52] Rummel, R. J.: *Applied Factor Analysis*. Evanston: Northwestern University Press, 1970.
- [53] Rutkovszky, E.–Rutkovszky, Á.: *A láthatatlan web keresése*. (2003). [Előadásanyag]
<https://nws.niif.hu/ncd2003/docs/ehu/EHU-61.htm> 2008.01.15.
- [54] Schwartz, C.: Web search engines = *Journal of the American Society for Information Science*, vol. 49. no. 11. (1998), p. 973-982.
- [55] Scime, A.: *Web mining: Applications and techniques*. Idea Group Inc. (IGI), 2005. 427 p.
- [56] Sherman, C.: *The state of search engine marketing*. (2006).
<http://searchenginewatch.com/showPage.html?page=3575926> 2008.01.20.
- [57] Siegal, S.–Castellan, N. J.: *Nonparametric statistics for the Behavioral Sciences*. Singapore: McGraw-Hill, 1988.
- [58] Silverstein, C.–Henzinger, M.–Marais, J.–Moricz, M.: Analysis of a very large web search engine query log = *SIGIR Forum*, vol. 33. no. 1. (1999), p. 6-12.
- [59] Skrop, A.: *New methods in web information retrieval effectiveness = Új módszerek a webes információ-visszakeresés hatékonyságának területén. [PhD dissertation]*. Veszprém: Pannon Egyetem, Műszaki Informatikai Kar, 2006. 138 p.
- [60] Spink, A.–Greisdorf, H.: Regions and levels: Measuring and mapping users' relevance judgments = *Journal of the American Society for Information Science*, vol. 52. no. 2. (2001), p. 161-173.
- [61] Spink, A.–Jansen, B. J.: A study of web search trends = *Webology*, vol. 1. no. 2. (2004)
<http://www.webology.ir/2004/v1n2/a4.html> 2008.01.27.
- [62] Sroka, M.: Web search engines for Polish information retrieval: Questions of search capabilities and retrieval performance = *International Information & Library Review*, vol. 32. (2000), p. 87-98.

- [63] Su, L. T.: Developing a comprehensive and systematic model of user evaluation of Web-based search engines. In.: *National Online Meeting: Proceedings*, (1997), p. 335-345. Medford, NJ: Information Today
- [64] Sullivan, D.: *Checking your listing in search engines*, October 2001.
<http://searchenginewatch.com/webmasters/article.php/2167861> 2007.07.10.
- [65] Szász, B.–Saraniva, A.–Bognár, K.–Unzeitig, M.–Karjalainen, M.: *Cultural heritage on the semantic web – the Museum24 project*. (2006). [Előadásanyag]. 10 p.
<http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/articles/Szasz-museum24Paris.pdf> 2008.01.14.
<http://www.museo24.fi> (“Museo24” portal homepage) 2008.01.14.
- [66] Szeredi, P.: *Ontológiák – egy matematikus-informatikus szemével* = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. [Előadásanyag]. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> 2008.01.10.
- [67] Szeredi, P.–Lukácsy, G.–Benkő, T.–Nagy, Zs.: A szemantikus világháló. In.: *A szemantikus világháló elmélete és gyakorlata*. Szerz. Szeredi P., Lukácsy G., Benkő T., Nagy Zs. Bp.: Typotex, 2005. p. 17-59.
- [68] Taube, M.: A note on the pseudo-mathematic of relevance = *American Documentation*, vol. 16. (1965), p. 69-72.
- [69] *Taxaurusz: Növény- és állatrendszertani fogalmak makrotezaurusza: 3.0 változat*. [Összeáll.] Ungváry R.; [közr. az] Országos Széchényi Könyvtár. Bp.: OSZK, 2000. XVII, 4, 16, 98 p.
<http://mek.oszk.hu/00000/00077> 2007.04.30.
- [70] Tinsley, H. E. A.: *Handbook of applied multivariate statistics and mathematical modeling*. Academic Press, 2000.
- [71] Tóth, E.: Statistical methods in measuring search engine performance = *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, vol. 20. no. 1. (2004), p. 105-113.
http://www.emis.de/journals/AMAPN/vol20_1/12.html 2008.01.27.
- [72] Tóth, E.: Exploring the capabilities of English and Hungarian search engines for various queries = *Libri*, vol. 56. no. 1. (2006), p. 38-47.
- [73] Tóth, E.: Kísérlet az internetes keresők nyelvi lehetőségeinek felmérésére = *Könyvtári Figyelő*, 53. évf. 2. sz. (2006), p. 303-311.
- [74] Ungváry, R.: A tartalom szerinti információkeresés az interneten: I. indexelőszolgáltatások = *Tudományos és Műszaki Tájékoztatás*, 47. évf. 1. sz. (2000), p. 3-17.
http://tmt.omikk.bme.hu/show_news.html?id=1624&issue_id=15 2008.01.27.
- [75] Ungváry, R.: A tartalom szerinti információkeresés az interneten: II. internetkatalógusok = *Tudományos és Műszaki Tájékoztatás*, 47. évf. 2. sz. (2000), p. 55-67.
http://tmt.omikk.bme.hu/show_news.html?id=1625&issue_id=16 2008.01.27.

- [76] Ungváry R.: Az információkeresés értékelése. In.: *Osztályozás és információkeresés: kommentált szöveggyűjtemény. 2. köt. Az információkeresés és elmélete*. Szerk. Ungváry R., Orbán É. Bp.: OSZK, 2001.
<http://mek.oszk.hu/01600/01683/pdf/01683-2.pdf> 2007.11.17.
- [77] Ungváry, R.: Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása = *Tudományos és Műszaki Tájékoztatás*, 51. évf. 5. sz. (2004), p. 175-191.
http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450 2008.01.27.
- [78] Ungváry, R.: Az ontológia fogalma, avagy az eltűnt tezaurusz = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. [Előadásanyag]. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> 2008.01.10.
- [79] Ungváry, R.–Vajda, E.: *Könyvtári információkeresés*. 2. jav. kiad. Bp.: Typotex, 2002. 170 p.
- [80] *A világ előrehaladása az információs társadalom terén 1998-2008.: World Progress Report 2008*. Készít. a BME-UNESCO Információs Társadalom- és Trendkutató Központjának (ITTK) kutatócsoportja. Bp. 2007. március
http://www.ittk.hu/web/docs/ITTK_WPR1998-2008.pdf 2008.01.28.
- [81] Wang, H.–Xie, M.–Goh, T. N.: Service quality of Internet search engines = *Journal of Information Science*, vol. 25. no. 6. (1999), p. 499-507.
- [82] Xie, M.–Wang, H.–Goh, T. N.: Quality dimensions of Internet search engines = *Journal of Information Science*, vol. 24. no. 5. (1998), p. 87-94.
- [83] *XML-Topic-Map (XTM) Standard*, ISO/IEC 13250: 2000. XTM TopicMaps Org.
<http://www.topicmaps.org/xtm> 2008.01.14.
- [84] Zádori Janka [Személyes közlés] 2007.08.10.

V. Publications

Citations:

1. Helyzetkép az internetes keresők értékeléséről = Könyvtári Figyelő, 47. évf. 2. sz. (2001.), p. 301-307.
Cited by: Poprády G.: Könyvtári trendek. In: Könyvtárosok kézikönyve. 5. Segédletek. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2003. p.13-62.
2. Az automatikus osztályozással kapcsolatos kísérleti kutatások eredményei = Tudományos és Műszaki Tájékoztatás, 49. évf. 5. sz. (2002.), p. 191-198.
Cited by: Poprády G.: Könyvtári trendek. In: Könyvtárosok kézikönyve. 5. Segédletek. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2003. p.13-62.

Abstracted publications:

3. Helyzetkép az internetes keresők értékeléséről = Könyvtári Figyelő, 47. évf. 2. sz. (2001.), p. 301-307.
DIALOG H. W. Wilson Record Number: BLIB02008916
4. Innovative solutions in automatic classification: a brief summary = Libri, vol. 52. no. 1., (2002.), p. 48-53.
DIALOG H. W. Wilson Record Number: BLIB02004470
5. Adatbányászatra irányuló törekvések a könyvtári területen = Könyvtári Figyelő, 48. évf. 3. sz. (2002.), p. 527-532.
DIALOG H. W. Wilson Record Number: BLIB03102651
6. Az automatikus osztályozással kapcsolatos kísérleti kutatások eredményei = Tudományos és Műszaki Tájékoztatás, 49. évf. 5. sz. (2002.), p. 191-198.
LISA, MAKSZAB
7. Statistical methods in measuring search engine performance = Acta Mathematica Academiae Paedagogicae Nyíregyháziensis, vol. 20. no. 1. (2004.), p. 105-113.
ZENTRALBLATT ZBL 1065.68018
8. Exploring the capabilities of English and Hungarian search engines for various queries = Libri, vol. 56. no. 1. (2006.), p. 38-47.
DIALOG 04376483 Genuine Article: 048VZ
9. Hogyan mérhető az internetes keresők szolgáltatása? = Híradástechnika, 61. évf. 1. sz. (2006.), p. 35-40.
10. Kísérlet az internetes keresők nyelvi lehetőségeinek felmérésére = Könyvtári Figyelő, 53. évf. 2. sz. (2006.), p. 303-311.
DIALOG H. W. Wilson Record Number: BLIB6110598

Other publications:

11. Eszenyiné B. M., Tóth E.: Quality Management Techniques and Informatics = Problemü ekonomicsnovo ta szosalnava rozvitku regionu i praktika naukova ekszperimentu, Kijev-Uzsgorod-Nyíregyháza: Karpati Kiadó, 2000. p. 276-280.
12. Eszenyiné B. M., Tóth E.: Quality Management Techniques and Informatics = Informatizacija dialnoszti pidpriemctv malovo ta szerednovo bizneszu: mehanizm, problemi, rozbitok, "Naukovij visznyik Uzsgorodszkava Gyerzsavnovo Universzityetu" Szerija "Ekonomika", No. 5., Uzsgorod, 2000. p. 64-69.
13. Az internetes keresők működésének minőségi vizsgálata = Természettudományi Közlemények, Nyíregyháza: Nyíregyházi Főiskola, (2005.), p. 427-438.

14. Quality and search engines' retrieval = Természettudományi Közlemények, Nyíregyháza: Nyíregyházi Főiskola, (2007), p. 125-134.

Essays published in conference proceedings:

15. An overview of the experimental methods used in automatic classification = MicroCAD 2002, International Scientific Conference Miskolc, Hungary, 7-8. March 2002., Volume of Section H: Applied Information Engineering p. 205-209. Full paper submitted
16. Szabó Sz., Kormos J., Tóth E.: Társadalmi informatika és a rendszerfejlesztés interdiszciplináris megközelítései = Informatika a felsőoktatásban 2002, Debrecen: Debreceni Egyetem, 2002. aug. 28-30., 6 oldal, cd kiadvány
17. Adatbányászattal kapcsolatos kérdések a könyvtári területen = Informatika a felsőoktatásban 2002, Debrecen: Debreceni Egyetem, 2002. aug. 28-30., 5 oldal, cd kiadvány
18. The application of intelligent search agents = MicroCAD 2003, International Scientific Conference Miskolc, Hungary, 6-7. March 2003., Volume of Section N: Applied Information Engineering p. 151-157. Full paper submitted
19. Tóth E., Fazekas G.: Statistical methods used in search engine evaluations, Proceedings of the 6th International Conference on Applied Informatics (ICAI04), Eger, Hungary, 27-31. January 2004. Volume I. p. 155-162.
20. Az internetes keresők szolgáltatásának minőségi vizsgálata = Informatika a felsőoktatásban 2005, Debrecen: Debreceni Egyetem, 2005. augusztus 24-26., 6 oldal, cd kiadvány

Presentations in conferences:

21. Internetes keresők használata a tartalomszolgáltatásban, Magyar Tudomány Napja 2000. alkalmából rendezett Szabolcs-Szatmár-Bereg Megyei Tudományos Konferencia, Nyíregyháza, 2000. november 7.
22. Információkeresés az interneten, MKE Műszaki Könyvtáros Szekció Orvosi Csoportja és a Magyar Orvosi Könyvtárak Szövetsége ülése, Nyíregyháza, 2001. augusztus 9.
23. Az internetes keresők értékelésével kapcsolatos kérdések, Magyar Tudomány Napja 2001. alkalmából rendezett Szabolcs-Szatmár-Bereg Megyei Tudományos Konferencia, Nyíregyháza, 2001. október 29.
24. An overview of the experimental methods used in automatic classification, MicroCAD 2002, International Scientific Conference, Miskolc, 2002. március 7.
25. Társadalmi informatika és a rendszerfejlesztés interdiszciplináris megközelítései, Informatika a felsőoktatásban 2002, Debrecen, 2002. augusztus 29.
26. Adatbányászattal kapcsolatos kérdések a könyvtári területen, Informatika a felsőoktatásban 2002, Debrecen, 2002. augusztus 30.
27. The application of intelligent search agents, MicroCAD 2003, International Scientific Conference, Miskolc, 2003. március 6.
28. Az internetes keresők mérésének egy lehetséges módszertana, MTA Szabolcs-Szatmár-Bereg Megyei Tudományos Testületének ülése, Nyíregyháza, 2003. szeptember 27.
29. Statistical methods used in search engine evaluations, 6th International Conference on Applied Informatics, Eger, 2004. január 30.

30. Az internetes keresők szolgáltatásának minőségi vizsgálata, Informatika a felsőoktatásban 2005, Debrecen, 2005. augusztus 25.
31. Az internetes keresők nyelvi lehetőségei, 30 éves a nyíregyházi könyvtárosképzés alkalmából rendezett tudományos emlékülés, Nyíregyháza, 2006. október 16.