



AZ INTERNETES KERESŐK LEKÉRDEZÉSI HATÉKONYSÁGÁNAK VIZSGÁLATA

Doktori (PhD) értekezés

TÓTH ERZSÉBET

**Debreceni Egyetem
Informatikai Kar
Debrecen, 2007.**

Ezen értekezést a Debreceni Egyetem TTK Matematika és Számítástudományok Doktori Iskola Informatika Programja keretében készítettem a Debreceni Egyetem TTK doktori (PhD) fokozatának elnyerése céljából.

Debrecen, 200.....

a jelölt aláírása

Tanúsítom, hogy doktorjelölt 200...-200... között a fent megnevezett Doktori Iskola Informatika Programjának keretében irányítással végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javaslom.

Debrecen, 200.....

a témavezető aláírása

Tartalomjegyzék

Bevezetés.....	1
1. Az internetes keresők működése.....	15
2. Az internetes keresők értékelése	17
3. Az internetes keresők nyelvi lehetőségeinek vizsgálata	19
4. A magyar nyelvű keresők vizsgálata.....	21
5. Összegzés és további kutatási feladatok	23
Irodalom a Bevezetéshez.....	23
1. Az internetes keresők működése	26
1.1. Keresőszolgáltatások előretörése az Interneten.....	26
1.2. Weboldalak begyűjtése és indexelése	28
1.3. A találatok sorrendezése, rangsorolása	30
1.4. A Google kereső PageRank algoritmus.....	33
1.5. Problémák az internetes kereséssel és a megoldási kísérletek	36
Irodalom az 1. fejezethez	38
2. Az internetes keresők értékelése	40
2.1. A keresők mérésére irányuló szabvány létrehozása	40
2.2. Statisztikai módszerek alkalmazása az információkeresés területén	42
2.3. Az internetes keresők elemzésével kapcsolatos problémák.....	43
2.4. A keresőszolgáltatásokon végzett jelentősebb statisztikai vizsgálatok.....	46
2.4.1. Egy tesztelés során bevezetett új mértékek.....	54
2.4.2. A SERVQUAL minőségbiztosítási modell használata a keresők értékelésére	58
Irodalom a 2. fejezethez	62
3. Az internetes keresők nyelvi lehetőségeinek vizsgálata	65
3.1. A tesztelés célja és motivációja	65
3.2. A tesztelés módszere.....	66
3.3. A tesztelésben részt vevő keresők.....	67
3.4. A kiválasztott vizsgálati szempontok ismertetése.....	67
3.5. Az angol nyelvű keresőknél talált nyelvi megoldások.....	69
3.5.1. Morfológiai elemzés	69
3.5.2. Stopszavak kezelése	69
3.5.3. Ékezetek kezelése	70
3.5.4. Csonkolás az angol nyelvben.....	70

3.5.5. Csonkolás a magyar nyelvben	70
3.5.6. Szinonimák visszakeresése az angol nyelvben	71
3.5.7. Szinonimák megtalálása a magyar nyelvben	73
3.6. A magyar nyelvű keresőknél talált nyelvi megoldások	73
3.6.1. Morfológiai elemzés	73
3.6.2. Stopszavak kezelése	74
3.6.3. Ékezetek kezelése	74
3.6.4. Csonkolás	74
3.6.5. Szinonimák megtalálása	75
3.7. A vizsgálat eredményeinek összefoglalása	76
Irodalom a 3. fejezethez	77
4. A magyar nyelvű keresők vizsgálata	78
4.1. A vizsgálat célja	78
4.2. A vizsgálat koncepciója	78
4.3. A weboldalak relevanciája	79
4.4. A vizsgálatban szereplő keresőeszközök	80
4.5. A vizsgálat során használt keresőkulcsok	80
4.6. A keresőkulcsok kiválasztása	81
4.7. A pontosság vizsgálata	83
4.7.1. „Rákóczi-szabadságharc” témakör	83
4.7.2. „Szökőár” témakör	83
4.7.3. „Evészavar” témakör	84
4.7.4. „Kegyhely” témakör	84
4.8. A teljesség vizsgálata	84
4.8.1. „Csiperkegomba” témakör	85
4.9. A teljességre kapott eredmények értékelése	86
4.9.1. A releváns találatok száma és az összes találatok száma	86
4.9.2. A relatív teljesség vizsgálata	87
4.10. A pontosságra kapott eredmények értékelése	88
4.10.1. A találati pontosság változása keresőkulcsenként	89
4.10.1.1. A diagramokból levonható konklúziók	92
4.10.2. A találatok pontosságára vonatkozó következtetés vizsgálata	93
4.10.3. A szórás és a várható érték elemzése	95
4.10.4. A vizsgálat eredményeinek összefoglalása	100
Irodalom a 4. fejezethez	103

5. Összegzés és további kutatási feladatok.....	104
6. Summary and further research tasks.....	110
7. Felhasznált irodalom.....	115
8. Publikációk jegyzéke.....	121

Függelék

1. melléklet: „Csiperkegomba” – releváns találatok.....	I
2. melléklet: „Evészavar” – releváns találatok.....	VI
3. melléklet: „Kegyhely” – releváns találatok.....	XI
4. melléklet: „Rákóczi-szabadságharc” – releváns találatok.....	XVI
5. melléklet: „Szökőár” – releváns találatok.....	XXIII
6. melléklet: Az egyes keresőkulcsokra visszakeresett releváns és összes találatszámok ...	XXIX
7. melléklet: A keresőkulcsokra kapott releváns és összes találatszámok.....	XXX

Bevezetés

Korunk számos folyamata közül csak a legjelentősebbeket emelném ki, amelyek a következők: a politikai, katonai értelemben vett egyhatalmú világ kialakulása, a demokratizálódás, az európai integráció, a fokozódó ázsiai gazdasági együttműködés, a technológia forradalma, a globalizáció, az információs és a fogyasztói társadalom kialakulása. Az első és az utolsó három a világ minden részére közvetlenül, vagy közvetve ható, globális folyamatnak tekinthető. Napjainkban a „globális” jelzőt több, tartalmában eltérő jelenségre is rá lehet húzni, azonban helyesebb, ha csak a gazdaság, a tőke, az ipar, a kereskedelem, a szolgáltatás földrajzi és módszertani terjeszkedését nevezzük globalizációnak.

Az információs és kommunikációs technológia (IKT) rohamos fejlődésének és konvergenciájának eredményeként a társadalomban egy új életforma, újszerű működés és viselkedés alakult ki. Új értékrendek jöttek létre. Ezt a széles körben elterjedt új életmódot, magatartást, információs technológiára épülő gazdaságot nevezzük információs társadalomnak ([Fodor2000] p. 95-98). Az „információs társadalom” kifejezés az 1960-as évek második felében jelent meg, az '50-es '60-as évek fordulóján jöttek létre a „tudástársadalom”, „tudásgazdaság” stb. összetételek. Az „információs társadalom” a világban fellelhető információk általános gazdagságát tükrözi, míg a „tudástársadalom” arra a gazdagságra hívja fel a figyelmet, amit a tudás teremt, és arra a szegénységre, amelyet a tudástársadalom viszonyai között a tudás hiánya okoz ([Nyíri2000] p. 49). Az információs társadalom kialakulása országoként különböző időszakban és ritmusban zajlik. A társadalom tág értelemben vett fejlettségi szintje és a kultúra nagymértékben befolyásolja azt. Lényeges, hogy erre a folyamatra sokkal „könnyebben” lehet hatni, társadalmi szinten jó irányba terelni, mint a globalizációra. A fogyasztói társadalom megjelenése a globalizációhoz és az információs társadalomhoz kapcsolódó harmadik jelentős folyamat ([Fodor2000] p. 95-97).

A könyvtárak viszonylag korán felismerték annak fontosságát, hogy szükséges eleget tenniük az információs társadalom kihívásainak. Az Európai Unió könyvtárakkal kapcsolatos tevékenysége az EU 3-4. Kutatás-Fejlesztési, azon belül Telematikai Keretprogramja keretében ment végbe. 1991-1994 között a 3. keretprogramban három pályázati felhívás jelent meg, melyek eredményeképpen 81 ún. akcióterv kezdődött el. Ebből mintegy kétszáz intézmény részvételével 51 közösen finanszírozott projectre került sor. 1995-1998 között a 4. keretprogram nyitott volt a közép-kelet-európai országok számára is. Két pályázatot írtak ki, 15 kutatási project, 7 összehangolt közös nagy munka és 20 újabb, ún. horizontális támogatási program indult. A könyvtárakkal kapcsolatos Európai Uniós programok részletes ismertetését lásd [CORDIS]-ban.

A globalizáció gyorsuló és könyörtelen versennyel jár együtt. Csak úgy lehetünk versenyképesek, hogyha az információs társadalom legfőbb értékét, magát az információt részesítjük előnyben. Egy adott szakmai kérdés megválaszolásának a leggyakoribb kiindulási pontja lehet számunkra az Internet, amely a minket körülvevő globális társadalomnak egyik fontos eszköze [Pajor2006]. Azonban az Internetről nem tételezhető fel, hogy az minden feltett kérdésünkre kielégítő választ fog nyújtani. A világhálón történő információkereséskor egyre nagyobb gondot jelent számunkra a minőségi, releváns információk felkutatása és kiválogatása a ránk zúduló információáradatból. Ebben támogatnak minket a rendelkezésre álló keresőszolgáltatások, bár azok nem minden esetben nyújtanak tökéletes megoldást számunkra.

Az Interneten megjelenő keresőeszközök köré külön iparág szerveződött, amelybe kisebb-nagyobb méretű cégek, vállalatok nagy pénzüsségeket fektetnek be. Mindezt pedig saját versenyképességük, sikerességük és hatékonyságuk javítása érdekében teszik. A számadatok érzékeltetésére a Search Engine Marketing Professionals Organization (SEMPO) egyik felmérésének lényeges megállapításaira utalnék: az Egyesült Államokban és Kanadában 5,75 milliárd dollárt költöttek 2005-ben keresőmarketingre (SEM: Search Engine Marketing). Ez az összeg 44%-al haladta meg a 2004-es költségeket. A teljes költség 83%-át fizetett helyezésekre (paid placement) fordították. Az előrejelzések szerint a keresőmarketingbe fektetett pénzüsszeg 2010-re elérheti a 11 milliárd dollárt Észak-Amerikában. *Safa Rashtchy* Internet média és marketing elemző szerint a keresőpiac fizetős része 2005-ben megközelítőleg 10 billió dollár hasznot termelt globálisan, ami 41%-kal fog növekedni 2006-ban. Előrejelzése szerint a keresőpiacnak ez a része globálisan 37%-os éves növekedésre számíthat 2010-ig, ami több mint 33 billió dollárnak felel meg. A felmérésből kiderült, hogy a keresőmarketing kampányok elsődleges célja a „branding” és az értékesítés volt. A kisebb cégek inkább a termékek eladására fókuszáltak, míg a nagyobbak (500 alkalmazott feletti) a weboldalukra érkező forgalom növelésére [Sherman2006]. A felmérés adatai rávilágítanak arra, hogy a keresőszolgáltatások, valamint a különféle cégek, vállalatok nagyon érdekeltek abban, hogy minél több bevételre tegyenek szert. Ebben a kiélezett versenyhelyzetben a keresőeszközök folyamatosan törekszenek arra, hogy megújuljanak és minél több speciális, új szolgáltatással vonzzák a használókat maguk köré. Ezért a versenyben részt vevő szereplők számára rendkívül fontos, hogy az egyes keresőeszközök minőségét hogyan értékelik a kutatók.

Az információkeresésre irányuló kutatás több mint két évtizedes múltra tekint vissza. Ezen a területen a vizsgálódás egyik lehetséges iránya a kérdést feltevő felhasználók viselkedésének tanulmányozása, azaz milyen kérdést, hogyan, milyen társadalmi rétegből, milyen tanultsággal rendelkezők tettek fel. A kutatók 1981-től folyamatosan számos modellt alkottak meg. A

modellek kialakítását nagymértékben befolyásolta a kutatók világlátása, kutatási területe és jártassága. Ennek függvényében beszélhetünk kognitív perspektivikus, szociális, szociális-kognitív vagy szervezeti modellekről ([Pajor2006] p. 40). Mindez azt tükrözi számunkra, hogy a felhasználók weben történő keresése több szinten vizsgálható, beleértve a társadalmi és a szervezeti szintet, az információkeresés szintjét, az ember és a számítógép közötti kapcsolat szintjét, valamint a megfogalmazott keresőkérdés szintjét [Spink-Jansen2004]. Csak jelzésszerűen hivatkoznék néhány jeles kutatóra, akik ezen a téren komoly eredményeket értek el: *Spink, Jansen, Saracevic, Ingwersen*. Az információkeresési viselkedésekkel, modellekkel kapcsolatos kutatási eredmények megtalálhatók [Pajor2006]-ban.

A vizsgált témakör fogalmai

Fontosnak tartom, hogy meghatározzam ennek a komplex tárgykörnek az alapvető fogalmait és a közöttük lévő kapcsolatrendszerét. Először a metaadat fogalmának meghatározásával kezdeném, hiszen ezeknek az információknak a megléte nélkülözhetetlen az internetes keresők működése számára. Metaadat kifejezésen a weblapok intellektuálisan vagy automatikusan létrehozott másodlagos adatait értjük, amelyek magát a dokumentumot helyettesítik. A keresés mindig a metaadatokat tekinti végig ([Ungváry2000] p. 66). Egy másik meghatározás szerint: metaadat alatt mindazokat a többletinformációkat értjük, amelyeket a keresők a weboldalakhoz kapcsolnak a keresőkérdés pontosabb megválaszolása reményében [Szeredi-Lukácsy-Benkő-Nagy2005]. Ezen adatok körébe tartoznak: a bibliográfiai leírás szabványosított adatelemei, a dokumentum tartalmát leíró kulcsszavak, tárgyszavak, deskriptorok és az osztályozási jelzetek. Elengedhetetlen követelmény volt a metaadatok egységes elektronikus kezelése, ami kiterjedt ezeknek az adatoknak az elsődleges dokumentumokból való kinyerésére és a dokumentumok számítógépes leírására ([Ungváry2000] p. 66). Számos metaadat-rendszer jött létre a hálózati információ feldolgozására, például az OCLC InterCat, a DublinCore, a WWW Semantic Header, a TEI (Text Encoding Initiative) fejléc stb. Ezek közül a metaadat-rendszerek közül a DublinCore jelentőségét hangsúlyoznám, hiszen napjainkban ez az egyik legáltalánosabban elterjedt metaadat-alkalmazás. A Dublin Core formátum 15 leíró elemet tartalmaz. Továbbá az áll a legközelebb a könyvtári katalogizáláshoz. Elterjedését elősegítette, hogy adatelemeit az európai szabványosítási szervezet, a European Committee for Standardization (CEN) is elfogadta ([Rácz2003] p. 223-225).

A keresőszolgáltatásoknak két típusát különböztethetjük meg: az indexelőszolgáltatásokat és az internetkatalógusokat. Az előbbieken belül

különleges változatként fordulnak elő a gyűjtő- és a metakeresők. A metakeresők (meta search engines, Meta-Suchmaschinen, métamoteur, métachercheur) segítségével több indexelőszolgáltatásban kereshetünk párhuzamosan anélkül, hogy az egyes szolgáltatásokkal külön foglalkoznunk kellene. A rendszer mindegyik keresőszolgáltatás adatbázisában végrehajtja a keresést, megjelenítve a találatoknál, hogy melyik szolgáltatás adatbázisában találta meg a rekordot. A metakeresők előnye, hogy rövid idő alatt valószínűsíthetően több releváns találathoz juthatunk ([Ungvárya2000] p. 15-17). Továbbá azok nem veszik figyelembe azokat az oldalakat, amelyek mindenféle trükkös megoldásokkal a javukra befolyásolják a keresők találatrangsorolását, azonban ezek az oldalak igazi, használható tartalommal nem rendelkeznek. Ezeket az oldalakat "spam"-eknek hívjuk. A metakeresők azért képesek a "spam" oldalak kiszűrésére, mert azok általában egy-egy keresőre szakosodnak és egyszerre több keresőt már nem tudnak becsapni [Szeredi-Lukácsy-Benkő-Nagy2005]. Átmeneti típusnak tekinthető a gyűjtőszolgáltatás (configurable unified search interface [CUSI], all-in-one formular, sample service, Sammeldienst), amely felajánl több keresőszolgáltatást, de mindig csak egyet választhatunk ki a lekérdezésre ([Ungvárya2000] p. 15-17).

Az indexelőszolgáltatások (search engines, Suchmaschinen, moteur de recherche) emberi munka nélkül, számítógépes programok segítségével végzik a keresést a hálózaton. Ezek a szolgáltatások két fő részből állnak: a keresőrobotból (crawler, web spider, web robot) és az indexelőből (indexer). A robotok állandóan figyelemmel követik és begyűjtik a weboldalakat a világhálóról a keresőszolgáltatás adatbázisába. Az indexelő elemzi a begyűjtött dokumentumokat, amelyekből előállítja az indexkifejezéseket. Létrehoz egy indexet, amely minden releváns kifejezéshez hozzárendeli az őt tartalmazó Uniform Resource Locator-ok (URL) listáját. A keresőszolgáltatás erre az indexre támaszkodik, amely révén elvégzi a keresést a felhasználó számára [Szeredi-Lukácsy-Benkő-Nagy2005]. A keresőrobotot és az indexelőt integráló egységet „keresőgépnek”, „keresőmotornak”, „keresőműnek” (search engine), „keresőrendszernek” (search system) nevezik. Hibásan a teljes keresőszolgáltatást „keresőgépnek”, „keresőmotornak”, „robotnak” hívják, ami a szolgáltató rendszernek csak az egyik részét jelenti. Ebbe beletartozik még a felhasználói felület és a szolgáltatott tartalom is ([Ungvárya2000] p. 10). Ezek a keresőszolgáltatások általában rendelkeznek egy egyszerű és egy összetett keresési lehetőséggel. Egyszerű kereséskor rendkívül nagy a visszakeresett, nem releváns dokumentumok száma, azaz a zaj. Annak csökkentése érdekében tanácsos használnunk

a részletes keresési lehetőséget (advanced search, powered search) ([Ungváry-Vajda2002] p. 112).

Az internetkatalógusokat (directories, annuaires Internet, répertoires Internet) ([Ungváry-Vajda2002] p. 113) „böngészőszolgáltatásnak” (browsing service, browsing Dienste) ([Ungváryb2000] p. 55), „tárgyszótárnak”, „tématárnak” (subject directory, Themenverzeichnis, annuaire thématique) ([Ungváry2001] p. 458), valamint „webes katalógusnak” (annuaire Web, répertoire Web) is nevezik ([Szeredi-Lukácsy-Benkő-Nagy2005] p. 53-54; 59). Ezek a keresőszolgáltatások hierarchikus osztályozási rendszert használnak. Adatbázisaik többnyire intellektuálisan feldolgozott weboldalak rekordjait foglalják magukba, valamint kapcsolatokat más adatbázisokhoz. Az osztályozást és a tartalmi kivonatok készítését szerkesztőségben végzik. Ezekben a katalógusokban osztályok alapján böngészhetünk, de lehetőségünk van arra is, hogy a keresőkérdés megadásával, célzott kereséssel találjuk meg a kívánt osztályt. Vannak olyan katalógusok is, melyek indexelőszolgáltatásként is működnek, ilyen például az Altavizsla. Az internetkatalógusok adatbázisai sokkal kisebbek, mint az indexelőszolgáltatásokéi, azonban a keresés lényegesen kisebb zajjal jár az intellektuális feldolgozásnak és a gondos osztályozásnak köszönhetően. A szakterületre specializálódott keresők nagy része internetkatalógusnak tekinthető. Egy részüket híres kutatóintézetek gondozzák, más részük kereskedelmi szolgáltatásnak minősül ([Ungváry-Vajda2002] p. 113-115).

Dolgozatom címében az internetes keresők megnevezés túl általánosnak tűnhet a fenti csoportosítás alapján, vizsgálataimban elsősorban indexelőszolgáltatások tanulmányozására törekedtem. Azonban az általam vizsgált angol és magyar nyelvű keresőeszközök oldalainak többségénél megfigyelhető volt, hogy azok internetkatalógusokat is felkínáltak kiegészítő szolgáltatásokként. Ebből adódóan azok többnyire mindkét minőségben egyaránt használhatók voltak. A továbbiakban áttekintem, hogy a vizsgált témakör milyen más egyéb kutatási területekhez kapcsolódik.

Kapcsolódó kutatási területek

Elsősorban hangsúlyozom a szemantikus webnek, mint perspektivikusan fejlődő területnek a jelentőségét, melynek feladata a jelentés megtalálása a webes tartalmakban. A szemantikus web kialakítására irányuló törekvések nyomán jelentek meg az ún. ontológiák. *Gruber* megfogalmazása szerint: az „ontológia megegyezésen alapuló fogalmi rendszer formális, egyértelmű leírása” [Gruber93]. Ebben a meghatározásban a „megegyezésen alapuló” kitétel lényeges, hiszen azt a szemléletet tükrözi, hogy az ontológiák szemantikai szabályrendszerek,

melyek a dolgok rendezésére használhatók [Ungváry2004]. Az ontológiák lehetővé teszik számunkra, hogy tisztázzuk az alapvető fogalmakat és a közöttük lévő relációkat. Továbbá elősegítik számunkra, hogy az erre vonatkozó tudásunkat formálisan és gépi következtetésre alkalmasan fogalmazzuk meg [Szeredi2007].

Számos fejleménynek kellett ahhoz bekövetkeznie, hogy ontológiák jöhessenek létre. Ezek közül csak a legfontosabbakat emelném ki. 2000-ben közreadtak egy „tématérképnek” (topic map) nevezett hierarchikus fogalmi struktúrát kezelő szabványt. A weben jelenleg elérhető vizualizált fogalmi struktúrák többsége ezen vagy ehhez hasonló fejlesztéseken alapszik [XML-Topic-Map]. A W3C konzorcium irányítása alatt egy másik irányban kezdődött el a fejlesztés. Ennek egyik fontos eredménye, hogy 2000-ben a web metaadatainak leírására egy szabványt hoztak létre, az XML-en alapuló webforrás leíró nyelvet (Resource Description Framework=RDF). A weben található hierarchikus fogalmi struktúrák formális leírására is ezt a nyelvet használták fel. 2002-ben a W3C konzorcium kezdeményezésére hozzákezdtek az ontológiák szabványának tekinthető webontológia-nyelv (Ontology Web Language) kidolgozásához. Az ideiglenes szabványajánlást 2003-ban adta közre a W3C konzorcium [Ungváry2004]. Jelenleg elérhető és már létező általános ontológiáknak tekinthetők, például a Dublin Core, a Magyar Egységes Ontológia. Szakterületi ontológiaként megemlíthető a Galen, amely orvostudományi szakterületen használatos [Szeredi2007]. Az ontológiáknak érdekes felhasználási területe a virtuális múzeum, amely gondolatvilágában közel áll a könyvtárakéhoz. A „Museo24” projekt leírását lásd [Szász-Saraniva-Bognár-Unzeitig-Karjalainen2006]-ban. Jelenleg egyfajta közeledés figyelhető meg hazánkban a könyvtári és az informatikai szakmai közösségek között az ontológiák terén, amit a W3C konzorcium magyar irodája szakmai előadások szervezésével egyaránt támogat [Szeredi2007]; [Ungváry2007].

A szakirodalomban az „invisible web” (láthatatlan web), „hidden web” (rejtett web), vagy „deep web” (mély web) angol kifejezéseket használják mindazon dokumentumok körének az összefoglalására, amelyek számos okból kifolyólag nem érhetők el a keresőszolgáltatások számára. A láthatatlan web csoportjába sorolhatók: a dinamikus weblapok (azaz a kereshető adatbázisokból nyert oldalak), azok az oldalak, amelyek csak regisztráció után érhetők el, a nem szöveges dokumentumok, valamint a keresőmotorok által kizárt oldalak. Fontos hangsúlyoznunk, hogy a web csak egy szolgáltatás az Interneten, tehát az nem azonos vele. Egy olyan hipertext struktúrára épül, amelyben szabadon böngészhetünk a szöveges formában megjelenített információk közötti kapcsolatok (linkek) alapján. Ha egy weblapra nem mutat egyetlen link sem, vagy abból nem vezetnek ki linkek, akkor az nem kerül bele a kereső adatbázisába. Azokat a weboldalakat, amelyeket a keresők keresőmotorjai megtalálnak

„felszíni webnek” (surface web) vagy „statikus webnek” nevezzük. Ennek nagysága a teljes web méretének a 16%-ára becsülhető. Ezzel szemben a láthatatlan web információmennyisége 550-szer nagyobb, mint a felszínié és növekedése, gyarapodása is sokkal gyorsabb ütemű [Rutkovszky, E.-Rutkovszky, Á. 2003]; [Pajor2006]. Sokféle törekvéssel igyekeztek a rejtett webet „láthatóvá tenni”, például bizonyos metakeresőkkel, intelligens keresőprogramokkal (ágensek), témakatalógusok kialakításával, egyéb speciális keresőkkel stb. Mindezeket a lehetséges megoldási kísérleteket, eszközöket bővebben lásd kifejtve [Pajor2006]-ban.

Ehhez a területhez kapcsolódóan hivatkoznék a szövegbányászat és az adatbányászat ígéretes lehetőségeire, amelyek a rejtett tudás kinyerésére törekednek a weben található, nagy mennyiségű strukturálatlan vagy félig strukturált HTML és egyéb formátumú dokumentumokból. Fiatal kutatási területnek számít a web mining, amely kiterjed az adatbányászatra, az Internet technológiákra, valamint a szemantikus webre [Scime2005].

A weben találkozhatunk speciális keresőszolgáltatásokkal is, például képek, videóanyagok visszakeresésére alkalmas keresőkkel, amelyek nagy népszerűségnek örvendenek a használók körében. Megjelenésük azt jelzi, hogy a használók rendkívül nagy mértékben igénylik a nem szöveges dokumentumok eredményes megtalálását. Ezen az új kutatási területen a megfelelő nyelvi keresőeszközök létrehozása és azok további fejlesztése elengedhetetlenül fontos feladat.

Kutatási célkitűzések és feladatok

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. Ezzel a témával – tudomásom szerint – egy korábbi, hazai disszertáció foglalkozott csupán az akronimák vonatkozásában. Szerzője bemutatja, hogy az intézményi névrövidítések mennyire alkalmasak rá, hogy a használót a megfelelő honlapra vezessék. Továbbá foglalkozik azzal a kérdéssel is, hogy az akronimák mennyire képesek azonosítani magát az intézményt, azok használhatók-e egyfajta tárgyszóként kereséskor. Vizsgálatát a szerző a magyar és a dán intézmények (államigazgatási szervek, politikai pártok, felsőoktatási intézmények) csoportján végezte el [Skrop2006].

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). Mindkét mérték valamilyen módon igényli a relevancia meghatározását. A relevanciával kapcsolatban különböző meghatározásokat találtam a szakirodalomban, amelyeket a továbbiakban ismertetek.

„Mennyiségileg a relevancia adott keresőkérdés vonatkozásában a visszakeresett tételek azon hányada, amely objektív értelemben megfelel a feltett kérdésnek” ([Ungváry-Vajda2002] p. 153). Egy másik meghatározás szerint: „a relevancia azt fejezi ki, hogy mekkora a közelség a felhasználói kérdés és a talált dokumentumok tartalma között (azaz azok a dokumentumok, melyek a kérdésnek megfelelnek, relevánsak)” ([Ungváry-Vajda2002] p. 155). A relevanciát úgyis definiálhatjuk, hogy az hasonlósági kapcsolatot jelent a tárolt információ reprezentációja (indextétel) és a kérdés reprezentációja (keresőprofil) között. ([Horváth-Sütő 2003] p. 175). A relevanciában sok szubjektív vonás található, például különböző használók eltérő álláspontot képviselhetnek adott dokumentumnak egy kérdésre vonatkozó relevanciáját vagy irrelevanciáját illetően ([Ungváry2001] p. 197). *Mortimer Taube* 1965-ben kiállt a szubjektív relevanciafogalom használata mellett, és tiltakozott mindenfajta „matematizált” relevancia bevezetése ellen. Az utóbbin alapuló képleteket pszeudo-matematikai konstrukcióknak tekintette, melyekkel az információkereső rendszerek hatékonyságát számszerűen értékelték [Taube65]. *Van Rijsbergen* szerint elvileg van olyan relevanciafogalom, amelyet objektívnek tarthatunk, s amelyet „logikai relevanciának” nevezhetünk ([Ungváry2001] p. 198). A fenti értelmezések közül vizsgálatom során a relevanciának az első meghatározásából indultam ki és abban mélyedtem el.

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 „májmetelyt”, 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 tővégi 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 végez-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 csokolásra 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*” (=olimposzi), 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 „*májmétey*” nevű parazita ragozott szóalakjaira kerestem a „*májmétey**” keresőkérdéssel. Továbbá megvizsgáltam a különböző keresőknél, hogy a „*májmétey**” 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 különbözik a megvalósított nyelvi lehetőségeik tekintetében.*

A másik vizsgálatot a pontosság és a teljesség jellemzőkre irányulóan 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 százaléértéket 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 vettem 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 százalékértéket a relatív teljességre vonatkozóan, ami azt tükrözte, hogy az egyes keresők ezen találatthalmaz releváns dokumentumainak hány százalékát találták meg. A kapott százalékértékek 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 találatthalmazaik 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.

A disszertáció egyes fejezeteinek bemutatása

Disszertációm négy, egymással szorosan összekapcsolódó fejezetből áll, amelyeket a továbbiakban szeretnék ismertetni.

- A disszertáció első fejezetében áttekintem a keresők lényeges összetevőit és az általuk végzett feladatokat. Részletezem a weboldalak begyűjtését ellátó keresőrobotok tevékenységét, ehhez kapcsolódóan tárgyalom az adattár jelentőségét, ahová a begyűjtött

weboldalak kerülnek. Ezenkívül áttekintem az indexelő feladatait, amely elemzi és egyben kigyűjti a releváns kifejezéseket a meglátogatott weboldalakból. A keresők elsődleges feladata, hogy megtalálják a feltett keresőkérdésre pontosan illeszkedő indexelt dokumentumokat. A keresők működésének legfontosabb jellemzője a találatok rangsorolása, ami különböző rangsorolási szempontok alapján történhet. Sorra veszem a találatrangsorolás fő elveit, amelyeket a keresők általában alkalmaznak. Jelenleg a Google kereső PageRank algoritmusát kitüntetett figyelemben részesül, ismertetem annak a szakmában pontatlanul elterjedt és a korrigált változatát. Kitérek ennek az algoritmusnak az alapfelfogására, valamint arra, hogy az hogyan modellezi egy véletlen szörfölő viselkedését. Végezetül, részletezem az internetes kereséssel kapcsolatos különböző problémákat és az azokra nyújtott megoldási kísérleteket.

- A dolgozat második fejezetében felhívom a figyelmet arra, hogy különféle módszerekkel és megközelítésekkel értékelték eddig a keresőket és hiányzik a mértékeknek egy általánosan alkalmazható szabványa az elemzésekhez. Ezért egyre sürgetőbb igény jelentkezik egy ilyen szabvány kidolgozása iránt, amely segítségével az értékeléseket megfelelő szintű következetességgel végezhetnék el a kutatók. Továbbá ismertetek egy javaslatot arra vonatkozóan, hogy egy lehetséges szabványnak milyen mutatókat kellene magába foglalnia a mérésekhez. Összefoglalom az információkeresés területén alkalmazott különböző statisztikai módszereket. Ezután pedig rátérek a keresők mérésével kapcsolatos problémák bemutatására. A szakirodalmi áttekintésből kiderül, hogy élénk vita folyik a kutatók között, hogy mit és hogyan mérjenek. Kezdetben a cranfieldi vizsgálatokban értékelték az osztályozási rendszerek hatékonyságát, amelyek során a pontosságot és a teljességet alapvető mértékeként alkalmazták ([Horváth-Sütő2003] p. 176-179). Felvázolom e két mérték vizsgálatával kapcsolatos nehézségeket. Ezután a teljesség igénye nélkül hét jelentősebb statisztikai vizsgálatot mutatok be, tárgyalva az azok során használt statisztikai módszereket, mértékeket és a kapott eredményeket. Befejezésül egy olyan vizsgálat eredményeiről számolok be, amely sikeresen adaptálta a SERVQUAL minőségbiztosítási modellt az internetes keresők minőségének a vizsgálatára.
- A dolgozat harmadik fejezete szorosan kapcsolódik Bar-Ilan és Gutman kutatók vizsgálatához, akik morfológiai szempontból tanulmányozták a nem angol nyelvű keresőkérdések értelmezését az angol nyelvű és a helyi fejlesztésű keresőkönél [Bar-Ilan-Gutman2005]. Ehhez kapcsolódóan végeztem egy vizsgálatot, amely arra irányult, hogy az angol és a magyar nyelvű keresők milyen hatékonysággal képesek nyelviileg értelmezni a lefuttatott kereséseket. Az angol nyelvű keresőkön magyar és angol nyelvű keresőkérdéseket

egyaránt tanulmányoztam. Azonban a magyar nyelvű keresők esetében csak magyar nyelvű kereséseket vizsgáltam, mivel azok többnyire magyar nyelvű weboldalakat indexelnek, kivéve az Origo-Vizslát és az Altavizslát. Kifejtem, hogy milyen módszerrel választottam ki a teszteléshez szükséges keresőkérdéseket, milyen keresők vettek részt a vizsgálatban. Áttekintem a vizsgálat szempontjait: a morfológiai elemzést, az ékezetek, valamint a stopszavak kezelését, a csonkolást, és a szinonimákra történő keresést. Részletezem, hogy ezen szempontok alapján milyen kérdésekre kerestem a választ a vizsgálat során. Végül pedig összegzem az angol és a magyar nyelvű keresőknél talált nyelvi megoldásokat. A vizsgálat elején felállított hipotéziseim alapján összefoglalom a keresők nyelvi lehetőségeire kapott eredményeket.

- A disszertáció negyedik fejezetében egy másik vizsgálatot ismertetek, amely a magyar nyelvű keresők lekérdezési hatékonyságának az elemzésére irányul. Az információkereső rendszerek hatékonyságát két alapvető mérték fejezi ki a pontosság és a teljesség, ezért azok alapján vizsgáltam a magyar nyelvű keresőket. Összesen öt magyar nyelvű keresőn csak magyar nyelvű kereséseket elemeztem. A pontosság vizsgálatakor a lefuttatott keresőkérdésekre csak az első hatvan találatot vettem figyelembe, amit 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 teljességet pedig a szakirodalomból ismert relatív teljesség alapján elemeztem ([Clarke-Willett97]] p. 185-187). A vizsgálat elvégzéséhez pontosan meghatároztam, hogy milyen feltételek mellett tekintetem egy weboldalt relevánsnak. Ezt követően bemutattam a vizsgálatban szereplő keresőket, valamint a lefuttatott keresőkérdéseket. Rávilágítottam a vizsgálatához szükséges megfelelő keresőkérdések kiválasztásának problémájára, ami minden esetben valamilyen háttértudást követelt meg. Tárgyaltam, hogy milyen jellegű dokumentumokat tekintetem relevánsnak, illetve irrelevánsnak az egyes keresőkérdések vonatkozásában. Kifejtettem a teljességre és a pontosságra kapott eredményeket a felállított hipotéziseim alapján.

A dolgozat négy fejezete szoros kapcsolatban áll egymással, egy egységes egészet alkot. Az első fejezet összefoglalja az internetes keresők értékelésével kapcsolatos nehézségeket, a teljesség igénye nélkül az hét jelentősebb statisztikai vizsgálat módszereit és eredményeit mutatja be. A második fejezet áttekinti a keresők működési alapelveit és azok lényeges alkotóelemeit. A harmadik fejezet egy olyan vizsgálatot ismertet, amely az angol és a magyar nyelvű keresők eredményességét elemzi a keresőkérdések nyelvi értelmezése szempontjából. A negyedik fejezet

egy másik vizsgálatról számol be, amely a magyar nyelvű keresők hatékonyságának az értékelésére fókuszál.

A Bevezetés további részében ismertetem a disszertáció négy fejezetének legfontosabb megállapításait. A dolgozat megfelelő fejezeteinél felsorolom azokat a publikációimat, amelyekben a dolgozat eredményeinek közzlése megtörtént.

1. Az internetes keresők működése

Ebben a fejezetben ismertetem egy keresőrendszer működési elveit, valamint az összetevői feladatait. A találatok rangsorolása a keresők egyik legmeghatározóbb tevékenysége, ehhez kapcsolódóan bemutatom a Google kereső PageRank módszerét. Tárgyalom a PageRank algoritmus alap gondolatát és rávilágítok arra, hogy az hogyan modellezi a felhasználói viselkedést egy keresés során.

1.1. Keresőszolgáltatások előretörése az Interneten

Áttekintem mindazokat a fő folyamatokat, amelyek az utóbbi tíz évben vagy esetleg még korábban kezdődtek el, és az információs társadalom kialakulásához vezettek. Röviden kitérek mindazokra az alkalmazási területekre, amelyeknek a fejlődését gyökeresen befolyásolta az információs társadalom megjelenése.

A keresők többsége kutatási programként indult a '90-es évek közepétől. Megnövekedett a keresésre specializálódó szoftvertermékek száma, például webes katalógusok, metakeresők, szakterületi szolgáltatások, kereső agensek és "push" szolgáltatások jelentek meg.

1.2. Weboldalak begyűjtése és indexelése

Egy kereső alkotóelemei a következő feladatokat látják el:

- *Keresőrobot*: a keresők ennek a speciális szoftvernek a segítségével végzik a weboldalak meglátogatását és begyűjtését. A robotnak le kell töltenie az általa meglátogatott oldalt, és azt át kell adnia az indexelőnek. Nem végez semmilyen elemzést a meglátogatott dokumentumon, hanem csak nyomon követi a hivatkozásokat és letölti a felkutatott oldalakat.
- *Adattár*: a keresőrobotok által begyűjtött oldalak ide kerülnek. Az adattár elsődleges feladata az oldalak egyenkénti tömörítése és szekvenciális tárolása. Annak további feladata a meglévő, begyűjtött dokumentumok frissítése is.
- *Indexelő*: fő feladata az adatbázisban lévő meglátogatott oldalak elemzése és az indexelendő kifejezések belőlük történő kigyűjtése. Az indexelő tulajdonképpen az adattárra

támaszkodik. Az indexelő létrehoz egy indexet, amely minden releváns kifejezéshez hozzákapcsolja az őt tartalmazó URL-ek listáját.

1.3. A találatok sorrendezése, rangsorolása

A keresőknek jelentős alkotóeleme a *Rangsoroló modul*, amely egy adott keresőkérdésre automatikusan sorrendezi a találatokat fontosság szerint.

A keresők leglényegesebb tevékenysége a találatok megfelelő fontossági sorrendben történő megjelenítése a felhasználók számára. Mivel a felhasználók csak az első 10-20 találatot képesek áttekinteni egy adott keresésre, ezért rendkívüli fontossággal bír, hogy a kereső milyen találatokat jelenít meg a találati lista elején.

Az egyik leggyakoribb sorrendezési szempont a keresőkifejezés helyének és előfordulási gyakoriságának vizsgálata a dokumentumban. Megfigyelhető az a tendencia, hogy a könnyedén manipulálható rangsorolási szempontok egyre inkább háttérbe kerülnek és csökken a súlyuk a végső sorrend kialakításában. Helyettük pedig olyan kritériumokra helyeződik a hangsúly, amelyeket nehezebb befolyásolni. Itt megemlíthetők, például olyan módszerek, amelyek az oldalak közti linkstruktúrát veszik figyelembe.

1.4. A Google kereső PageRank algoritmus

A PageRank algoritmus alapgondolata az, hogy amikor egy oldal hivatkozik egy másik weblapra, akkor a forrásweboldal tulajdonképpen ajánlja a hivatkozott lapot. Tehát az oldal létrehozója azért tüntette fel a linket az oldalán, mert a másik lapot valamilyen szempontból fontosnak tekintette. Emellett azt is figyelembe kell vennünk, hogy a hivatkozó oldal mennyire fontos, mert egy jelentős oldalnak többet ér a hivatkozása. Eredményül egy rekurzív algoritmust kapunk, ami azt fejezi ki, hogy egy oldal fontos, ha mérvadó oldalak hivatkoznak rá.

A PageRank algoritmus egy olyan modellnek is tekinthető, amely a véletlen szörfölő viselkedését tükrözi. Egy ilyen felhasználó véletlenszerűen elindul egy weboldalról és folyamatosan előrehaladva különböző hivatkozásokra kattint. Nem is figyeli meg, hogy hova kattint, hanem egyenletes eloszlás szerint választ a meglévő hivatkozások közül. Ezzel magyarázható az, hogy a PageRank algoritmus a kimenő linkek számával elosztja egy bizonyos oldal fontosságát. Mindez addig tart, amíg szörfölünk meg nem unja a kattintgatást és egy másik véletlenszerűen kiválasztott weboldalon nem indul el. Ez az egyenlet egy valószínűségi eloszlást határoz meg, ahol egy-egy weboldal PageRank értéke egy valószínűségnek (0 és 1 közötti valószínűségnek) felel meg. Ebben a modellben az összes weboldal PageRank értékeinek összege maximum 1 lehet. Ez a megállapítás csak abban az esetben lehet igaz, hogyha a felhasználónk

egy adott oldalon mindig talál legalább egy hivatkozást, amelyen továbbhaladhat [Szeredi-Lukácsy-Benkő-Nagy2005].

1.5. Problémák az internetes kereséssel és a megoldási kísérletek

Röviden összefoglalom a kereséssel kapcsolatos nehézségeket, amelyek a következők:

- Az Internet hatalmas mérete, ami az oldalak keresését és begyűjtését nagymértékben befolyásolja. A weblapok meglátogatása és feltérképezése időigényes feladatot jelent még a legjobb keresők számára is.
- Az utolsó begyűjtés óta eltelt idő alatt az Internet tartalma és szerkezete gyorsan megváltozik.
- A keresőrendszerek számára általában elérhetetlenek azok az Interneten meglévő tartalmak, amelyek a mély web körébe sorolhatók, például a weben keresztül lekérdezhető adatbázisok, a nem szöveges formában található dokumentumok.
- A keresőrobotok nem követik a dinamikus weblapokra mutató hivatkozásokat, ezáltal azok sok információhoz nem férnek hozzá.
- A keresők nem a fellelhető dokumentumok és a keresőkérdés jelentésével foglalkoznak, hanem csupán a szöveges alakkal.

Ebben a részben az egyes felvetett problémákra körvonalazom a megoldási lehetőségeket is.

2. Az internetes keresők értékelése

Ebben a fejezetben az a célom, hogy összefoglaljam az internetes keresők mérésével kapcsolatos problémákat, továbbá, hogy áttekintsem mindazokat a kutatási eredményeket, amelyeket a különböző statisztikai vizsgálatok során elértek.

2.1. A keresők mérésére irányuló szabvány létrehozása

Jelenleg komoly vita folyik a kutatók között, hogy milyen megbízható mutatókat használjanak a mérések során. Ehhez hozzájárul még az, hogy rendkívül nehéz ezeknek a megbízható mértékeknek a megtalálása.

A mutatók általános szabványának hiánya nagy problémát jelent az értékelésekben. Ezen hiányosság miatt a keresőszolgáltatások értékelésével foglalkozó kutatás jelenleg nem egységes az alkalmazott módszerek tekintetében, ezért nagy szükség lenne megfelelő szabvány kidolgozására és használatára a méréseknél. Ebben a részben ismertetek egy javaslatot arra vonatkozóan, hogy egy ilyen szabvány milyen mértékeket tartalmazzon a mérésekhez.

2.2. Statisztikai módszerek alkalmazása az információkeresés területén

Ebben a részben felvázolom, hogy a kutatók milyen statisztikai módszereket használtak a különböző problémák vizsgálatára az információkeresés területén, például a keresési találatok elemzésére, a keresőrendszerek lekérdezési teljesítményének az összehasonlítására, a releváns dokumentumok becslésére egy adott gyűjteményen belül stb.

2.3. Az internetes keresők elemzésével kapcsolatos problémák

Számos probléma merül fel a keresők vizsgálatakor, amelyek a következők:

- a keresők értékelésével kapcsolatos vizsgálatok nem egységes módszerekkel valósultak meg.
- A tanulmányok nagy része nem számol be az általa használt módszertanról.
- Az elemzések kevés tesztelési eredménnyel rendelkeznek, ezért azok nem alkalmasak a mélyebb szintű statisztikai vizsgálatok végzésére.
- Hiányzik a szisztematikus megközelítés ezekből a tanulmányokból.
- A kutatóknak nincs egységes kialakult véleménye arról, hogy mit mérjenek, és hogyan mérjenek egy szolgáltatást.
- A keresési folyamat sokféleképpen befolyásolható az értékelések során.
- Az internetes keresők állandóan változnak, továbbá a világháló is egy dinamikusan változó közeg.

Ezen nehézségek ellenére megállapíthatjuk, hogy mégis van értelme elemezni a használatban lévő keresőket, de szabványos értékelési módszereket erre a feladatra nem alkalmazhatunk.

2.4. A keresőszolgáltatásokon végzett jelentősebb statisztikai vizsgálatok

Ebben a részben hét jelentősebb statisztikai vizsgálat módszereit és eredményeit foglalom össze. Ahol indokoltnak találtam azok alkalmazható mértékeit is részleteztem. Itt csak a vizsgálatok elért eredményeit tekintem át lényegre törően, és nem térek ki azok módszereire.

- *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 gyűjtemény lefedettség átlagértékeit [Clarke-Willett97].

- *Leighton és Srivastava* öt különböző kísérletet végeztek, ahol a keresőktől kapott első 20 találat pontosságát mérték. Megfelelő mértékeket 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* azt vizsgálták, hogy a legnépszerűbb keresők mennyire hatékonyan válaszolták meg a természetes nyelvű kereséseket [Radev-Libner-Fan2002].
- *Bar-Ilan* a keresők működését követte nyomon hosszabb időtartamon keresztül. Találatszolgáltatásuk stabilitását számos mérték irányából közelítette meg [Bar-Ilan2002a].

Ez a fejezet tehát áttekintést nyújt a korábbi kutatásokról, valamint azokról a statisztikai módszerekről, amelyeket sikeresen használtak a mérések során.

Az alábbi publikációim tartalmazzák az első fejezet eredményeit, megállapításait.

1. Tóth E.: Helyzetkép az internetes keresők értékeléséről = Könyvtári Figyelő 47. évf. 2. sz. (2001), p. 301-307.
2. Tóth, E.: Statistical methods in measuring search engine performance = Acta Mathematica Academiae Paedagogicae Nyíregyháziensis, vol. 20. no. 1. (2004), p. 105-113.
3. Tóth E.: Hogyan mérhető az internetes keresők szolgáltatása? = Híradástechnika, 61. évf. 1. sz. (2006), p. 35-40.

3. Az internetes keresők nyelvi lehetőségeinek vizsgálata

Az ebben a fejezetben ismertetett morfológiai vizsgálat a keresőkérdések nyelvi értelmezésére irányul az angol és a magyar nyelvű keresőknél.

3.1. A tesztelés célja és motivációja

Vizsgálatommal az volt a céлом, hogy felmérjem az angol és a magyar nyelvű keresők nyelvi lehetőségeit a megadott keresésekre. Az angol nyelvű keresőkön angol és magyar nyelvű keresőkérdéseket egyaránt lefuttattam, a magyar nyelvű keresőkön azonban csak magyar nyelvű kereséseket végeztem, mert azok főként magyar nyelvű weboldalakat indexelnek, kivéve az Origo-Vizsla és az Altavizsla szolgáltatásokat. Arra a kérdésre kerestem a választ, hogy ezek a

keresőszolgáltatások mennyire hatékonyan kezelték a keresőkérdéseket morfológiai szempontból.

3.2. A tesztelés módszere

A vizsgálat során végzett módszertani lépéseket mutatom be.

3.3. A tesztelésben részt vevő keresők

A tesztelésbe három angol nyelvű keresőt vontam be: a Google-t, az AltaVista-t és az AlltheWeb-et. A választásom azért esett erre a három angol nyelvű keresőre, mert azok lehetővé teszik a magyar nyelvű weboldalakra történő keresést. Öt jelentősebb magyar nyelvű keresőt teszteltem, amelyek a következők voltak: *Heuréka*, *Origo-Vizsla*, *Kurzor*, *Góliát* és az *Altavizsla*. A keresők kiválasztásánál fontos kritérium volt számomra, hogy azok rövid válaszidő alatt szolgáltatassanak megfelelő mennyiségű találatot a keresésekre.

3.4. A kiválasztott vizsgálati szempontok ismertetése

A teszteléshez *Bar-Ilan* és *Gutman* közös tanulmányában [Bar-Ilan-Gutman2005] található szempontokat vettem alapul, amelyek az alábbiak voltak: morfológiai elemzés (stemming), stopszavak és ékezetek kezelése, csonkolás, szinonimákra történő keresés. A keresők nyelvi lehetőségeit tehát ezen szempontok alapján vizsgáltam. Különböző keresőkérdéseket futtattam le a keresőkön, amelyek hűen tükrözték ezeket a szempontokat és megfeleltek a tesztelés célkitűzéseinek.

3.5.-3.6. Az angol és a magyar nyelvű keresőknél talált nyelvi megoldások

Ezekben az alfejezetekben áttekintem az angol és a magyar nyelvű keresőknél kapott eredményeket a vizsgálati szempontok alapján.

3.7. A vizsgálat eredményeinek összefoglalása

Vizsgálatom eredményeit a felállított hipotéziseim alapján ismertetem.

Az alábbi publikációim tartalmazzák a harmadik fejezet eredményeit, megállapításait.

Tóth E.: Exploring the capabilities of English and Hungarian search engines for various queries = *Libri*, vol. 56. no. 1. (2006), p. 38-47.

Tóth E.: Kísérlet az internetes keresők nyelvi lehetőségeinek felmérésére = *Könyvtári Figyelő*, 53. évf. 2. sz. (2006), p. 303-311.

4. A magyar nyelvű keresők vizsgálata

Az ebben a fejezetben bemutatott vizsgálat a magyar nyelvű keresők hatékonyságát értékeli.

4.1. A vizsgálat célja

Vizsgálatom elsődleges célkitűzése, hogy összehasonlítsam a magyar nyelvű keresők lekérdezési hatékonyságát.

4.2. A vizsgálat koncepciója

Összefoglalom, hogy milyen elméleti, módszertani megfontolások alapján hajtottam végre a vizsgálatomat.

4.3. A weboldalak relevanciája

A vizsgálat során egy weboldal akkor minősült relevánsnak, hogyha:

- tartalmazta a megadott keresőkulcsot és/vagy annak más egyéb szinonimáit;
- technikailag bármikor elérhető volt;
- magyar nyelven jelent meg;
- tartalmában és teljes terjedelmében szorosan kapcsolódott a megjelölt témához. Ennek az utóbbi feltételnek a teljesülését az biztosította, hogyha a weboldal tartalmát minden egyes esetben alaposan ellenőriztem és meggyőződtem annak megbízhatóságáról és hitelességéről formai szempontok alapján. Amennyiben ezek a definiált feltételek együttesen nem álltak fenn, akkor egy weboldal nem minősülhetett relevánsnak a vizsgálat számára.

4.4. A vizsgálatban szereplő keresőeszközök

Összesen öt magyar nyelvű keresőt vontam be a vizsgálatba, amelyek a következők voltak: Heuréka, Origo-Vizsla, Kurzor, Góliát és az Altavizsla. Azért választottam ezeket a keresőket, mert a vizsgálat ideje alatt azok bizonyultak a legjelentősebb keresőeszközöknek a magyar felhasználók számára. Találatszolgáltatásukat megbízhatónak találtam, mert azok rövid válaszdőn belül megfelelő mennyiségű találatot nyújtottak a lefuttatott keresésekre.

Keresőkulcs, illetve *keresőkulcs érték* alatt azokat a felhasználók által megadott releváns kulcsszavakat értem, amelyekre a kereső a weboldalak visszakeresését végzi. Mindig egy speciális keresőkulcs érték jelölhet egy *témakört*, például a „szökőár” keresőkulcs érték fejezi ki a „szökőár” témakört.

4.5. A vizsgálat során használt keresőkulcsok

Áttekintem mindazokat a keresőkulcsokat, amelyek alapján a vizsgálatomat elvégeztem.

4.6. A keresőkulcsok kiválasztása

Ebben a részben arra szeretnék rávilágítani, hogy a megfelelő keresőkulcsok megtalálása mindig valamilyen háttértudást követel meg, továbbá az adott téma sajátosságainak a figyelembevételét. A vizsgálat megtervezésekor nagymértékben támaszkodtam a próbakeresések során nyert előzetes tapasztalataimra, amelyeket ebben az alfejezetben részletezek.

4.7. A pontosság vizsgálata

Ebben a részben különböző eseteket mutatok be, amikor a weboldalak relevanciájával kapcsolatban döntéseket hoztam az egyes témakörök esetében.

4.8. A teljesség vizsgálata

Ebben a részben áttekintem azokat az eseteket, amikor a visszakeresett dokumentumok relevanciáját határoztam meg „*csiperkegomba*” témakörben.

4.9. A teljességre kapott eredmények értékelése

A keresők által visszakeresett összes találatból létrehoztam a releváns találatoknak egy halmazát „*csiperkegomba*” témakörben. Ezután pedig azt tanulmányoztam, hogy az egyes keresők milyen arányban keresték vissza ennek a halmaznak a releváns dokumentumait. Kiszámoltam egy százaléértéket a relatív teljességre vonatkozóan, ami azt tükrözte, hogy az egyes keresők ezen találathalmaz releváns dokumentumainak hány százalékát keresték vissza. A kapott százaléértékek alapján pedig felállítottam egy rangsort a keresők között.

4.10. A pontosságra kapott eredmények értékelése

Vizsgálatom során csak az első hatvan találat pontosságát vettem figyelembe. A találatok pontosságára vonatkozóan kiszámítottam egy százaléértéket, ami azt fejezte ki, hogy hány releváns találatot kaptam az első, a második és a harmadik húsz lekérdezett találatból. Tanulmányoztam az egyes keresőkulcsok vonatkozásában a találati pontosság értékek változását az első, a második és a harmadik húsz találatban. Ezt követően megvizsgáltam, hogy lehetséges-e meggyőző következtetést meghatároznom a releváns találatok számára vonatkozóan a találati

lista hátralévő részében. Végül pedig a releváns találatok tapasztalati szórását, várható értékét, valamint a mintaátlag szórását különböző megközelítésekben elemeztem.

5. Összegzés és további kutatási feladatok

A két empirikus vizsgálatom eredményeit a felállított hipotéziseim alapján foglalom össze, valamint ismertetem az eredményekből származó téziseket. Ezenkívül felvázolom a további kutatási feladatokat.

Irodalom:

[Bar-Ilan2002a] 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.

[Bar-Ilan-Gutman2005] 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.

[Brin-Page98] 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.

[Chignell-Gwizdka-Bodner99] 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.

[Clarke-Willett97] Clarke, S. J.–Willett, P.: Estimating the recall performance of web search engines = Aslib Proceedings, vol. 49. no. 7. (1997), p. 184-189.

[CORDIS] Creating a European library space: Telematics for libraries programmes 1990-1998.
<http://cordis.europa.eu/libraries/en/intro.html> 2008.01.22.

[Fodor2000] 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.

[Gruber93] Gruber, T.: A translation approach to portable ontology specifications = Knowledge Acquisition, vol. 5. no. 2. (1993), p. 199-220.

[Horváth-Sütheő 2003] Horváth T.–Sütheő P.: A tartalmi feltárás. In.: Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2003.

[Leighton-Srivastava99] 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.

[Lucas-Topi2002] 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.

[Nyíri2000] Nyíri, K.: Globális társadalom, helyi kultúra. In.: Az információs társadalom. Összeáll. Demetrovics J., Keviczky L. Bp.: MTA, 2000. p. 43-64.

- [Pajor2006] 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.
- [Rácz2003] Rácz, Á.: A kiadványok bibliográfiai számbavétele; leíró katalogizálás. In.: Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2003.
- [Radev-Libner-Fan2002] 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.
- [Rutkovszky, E.-Rutkovszky, Á. 2003] Rutkovszky, E.–Rutkovszky, Á.: A láthatatlan web keresése. (2003). [Előadásanyag]
<https://nws.niif.hu/ncd2003/docs/ehu/EHU-61.htm> 2008.01.15.
- [Scime2005] Scime, A.: Web mining: Applications and techniques. Idea Group Inc. (IGI), 2005. 427 p.
- [Sherman2006] Sherman, C.: The state of search engine marketing. (2006).
<http://searchenginewatch.com/showPage.html?page=3575926> 2008.01.20.
- [Skrop2006] Skrop, A.: New methods in web information retrieval effectiveness = Új módszerek a web-es információ-visszakeresés hatékonyságának területén. [PhD dissertation]. Veszprém: Pannon Egyetem, Műszaki Informatikai Kar, 2006. 138 p.
- [Spink-Jansen2004] 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.
- [Sroka2000] 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.
- [Szász-Saraniva-Bognár-Unzeitig-Karjalainen2006] 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.
- [Szeredi2007] Szeredi, P.: Ontológiák–egy matematikus-informatikus szemével = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> 2008.01.10.
- [Szeredi-Lukácsy-Benkő-Nagy2005] 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.

- [Taube65] Taube, M.: A note on the pseudo-mathematic of relevance = American Documentation, vol. 16. (1965), p. 69-72.
- [Ungvárya2000] 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.
- [Ungváryb2000] 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.
- [Ungváry2001] 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.
- [Ungváry2004] 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.
- [Ungváry2007] Ungváry, R.: Az ontológia fogalma, avagy az eltűnt tezaurusz = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> 2008.01.10.
- [Ungváry-Vajda2002] Ungváry, R.–Vajda, E.: Könyvtári információkeresés. 2. jav. kiad. Bp.: Typotex, 2002. p. 170.
- [XML-Topic-Map] XML-Topic-Map (XTM) Standard, ISO/IEC 13250:2000. XTM TopicMaps Org.
<http://www.topicmaps.org/xtm> 2008.01.14.

1. Az internetes keresők működése

A világháló új helyzetet teremtett a hagyományos információkeresés területén, hiszen a rendszerezettség, homogenitás és a rend helyett azt láthatjuk, hogy bárki létrehozhat rajta dokumentumot, amelynek minőségét és megbízhatóságát senki nem vizsgálja. Tehát a világháló heterogén szintaktikájú és szemantikájú, továbbá többnyire nem ellenőrzött tartalmú dokumentumok halmazát reprezentálja. Ebből adódóan az internetes keresés alapvetően eltér egy meglévő, lassan változó, kontrollált dokumentumgyűjteményben való kereséstől. Ez a különbség többek között abban is megnyilvánul, hogy a keresőknek meg kell találniuk a releváns webes tartalmaknak azt a halmazát, amelyek jól hasznosíthatók a felhasználók számára, nem pedig egy meglévő dokumentumgyűjteményből kell kiválogatniuk a keresőkérdésre pontosan illeszkedő dokumentumokat. Kereséskor a legjobb találatoknak egyéb jellemzőik is vannak (frissítési gyakoriság, minőség, hivatkozási szám, népszerűség stb.), amit a keresőknek szintén figyelembe kell venniük és nem elegendő csupán a keresésnek pontosan megfelelő dokumentumokat szolgáltatniuk. Egy-egy keresésre különböző válaszokat adhatunk, ezért nagyon lényeges, hogy mely találatokat jelenítjük meg elsőként a felhasználóknak [Szeredi-Lukácsy-Benkő-Nagy2005]. Még mielőtt rátérnék a téma részletes tárgyalására, szeretném pontosan meghatározni az internetes keresők fogalmát. Internetes keresők alatt a programoknak egy olyan általános csoportját értem, amely lehetővé teszi a weben történő dokumentumkeresést a felhasználók számára. Ezek a programok dokumentumokat indexelnek és arra törekednek, hogy megtalálják a feltett keresőkérdésre a releváns találatokat.

1.1. Keresőszolgáltatások előretörése az Interneten

Lényeges változásnak tekinthetjük, hogy az ezredfordulóra az országok döntő többsége információs társadalomként jelent meg a világtérképen. Végbement egy technológiai forradalom, a digitális eszközök mindennapjaink részévé váltak, azokat nem csak eszközöknek tekintjük, hanem emocionálisan is kötődünk hozzájuk. Mindez kihatott a médiafogyasztásra, megváltoztak a kulturális objektumok átvételi csatornái. Átrajzolódott a gazdaság, a kormányzás, a tartalomipar, átalakult a fogyasztó és az előállító viszonya. Legfontosabb jelenségként tapasztalhattuk az informatika, az Internet és a számítógép héthöznapivá válását, bár ezek a folyamatok jóval korábban (10-20-30 éve) kezdődtek el. Az Internet használata egy új, szűk kör által használt újdonságból a világ lakosságának

nagyjából hatoda által használt eszközzé vált. 1998-2008 között az internetezők száma több mint tízszeresére növekedett világszerte ([ITTKa2007] p. 7-8, 13, 23).

Az elmúlt tíz év viszonylatában Magyarország nem tekinthető sikeresnek. Hazánk rosszul teljesít a különböző információs társadalmi mutatókban, nemzetközi összehasonlításban az továbbra sem tartozik, sem az élenjárók, sem a közvetlen követők csoportjába. Megfigyelhető a korábban megszerzett pozícióinak az elvesztése. Magyarország bekerül egy olyan nemzetközi munkamegosztásba, ahol gyakran alárendelt szerepben találhatók a hazai szereplők. Az elmúlt tíz év azonban gyökeresen megváltoztatta a hazai viszonyokat, hiszen az Európai Unióhoz való csatlakozás révén hatalmas történelmi esély nyílt a felzárkózásunkra ([ITTKb2007] p. 109).

A 2000-es évek elején az Internet terjedésében egy második, minőségi hullám kezdődött el, a gyors, állandó online jelenlétet biztosító kapcsolatoké. A szélessávú Internet elérések vezető technológiája az xDSL. A jövőt nézve a jelszó a következő milliárd behálózása – nem csak fizikai, hanem kulturális, attitűd értelemben is. A jövőben megjelenik a drótnélküli és a fényvezetős technika is, ezek az utóbbi két évben indultak rohamos fejlődésnek. A mobilkészüléket is az IKT eszközök meghatározó tényezőjének tekintjük. A mobiltechnológia terjedésében megfigyelhető egy éles minőségi váltás: a kezdeti beszédforradalmat az adatátvitel forradalma váltja fel. A mobilkészülékek elterjedése kulturális változást is eredményezett: a mobilkommunikáció életünk természetes és szerves részévé vált. Az IKT eszközök egyenlőtlen elterjedése következtében társadalmi megosztottság alakult ki, amelyet paradox módon ugyanezen eszközök segítségével próbálnak megszüntetni. Számos megoldás született a digitális egyenlőtlenségek felszámolására (pl. hiper olcsó laptop és mobilkészülék terjesztése), ennek ellenére azok jelentős változást nem eredményeztek.

Az informatika és az oktatás magas szintű kapcsolatban áll egyre inkább egymással. Az elmúlt évtizedben az iskolákban terjednek az IKT eszközök, a digitális tananyagok, a tanulást segítő szoftverek. Sőt az évtized vége felé már virtuális oktatási terekről, és a szórakoztatva tanításban rejlő lehetőségekről esik szó.

Az utóbbi évtizedben nagy változáson ment keresztül az e-kormányzat területe. Kezdetben a közigazgatás elektronizálása még csak a számítógépek, és az Internet használatának elterjesztését jelentette. Napjainkban azonban az elektronikus közigazgatás a teljes közigazgatási munka és a szolgáltató oldali folyamatok modernizálását, egyszerűsítését, az ügyfél oldali szolgáltatások

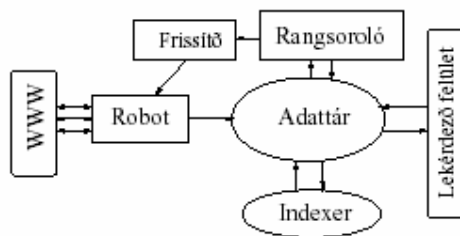
kényelmesebbé, olcsóbbá válását foglalja magába. A teljes körű e-kormányzás még csak álom, annak megvalósulása nehézkes.

Az információs társadalom egyik rendkívül fontos területe a kutatás-fejlesztés, illetve annak hasznosítása a gazdasági versenyképesség. A kutatás-fejlesztésre és az innovációra fordított pénzüsszegek reálértékben és GDP-arányosan is folytonosan növekednek.

Az Internet nagy hatással van a politikai közéletre is. A világháló terjedésével egyre jelentősebb szerepet játszik a politika minden megjelenési formájában, a kampányokban, a tájékoztatásokban stb. Az Interneten mind több a kísérlet a szabályozás és az ellenőrzés növelésére. A cenzúra egyre kifinomultabb megjelenési formái mellett az elektronikus demokrácia térhódítása is egyre inkább jellemzőbb ([ITTKa2007] p. 8-12).

1994-ben jelentek meg az első internetes keresők a weboldalak gyors mennyiségi növekedésének köszönhetően. A keresőknek nagy része kutatási programként indult. A kezdeti időszakban néhány kezdeményezés kudarcba fulladt, mert a vállalt feladat meghaladta a korlátozott emberi és technikai erőforrásokat. A fennmaradt keresőszolgáltatásokat főként vállalati tőkéből, reklámokból, tőkebefektetésekből, illetve kutatási kezdeményezésekből finanszírozták. 1996-ban a különböző kereskedelmi folyóiratok, üzleti és napilapok komoly figyelmet szenteltek a keresőknek. Megnövekedett a keresésre specializálódó szoftvertermékek száma, például webes katalógusok, metakeresők, szakterületi szolgáltatások, kereső ágensek és “push” szolgáltatások jelentek meg [Schwartz98].

A továbbiakban részletesen ismertetem egy keresőrendszer alkotóelemeit és a rájuk bízott elvégzendő feladatokat. A következő ábra vázlatos áttekintést nyújt számunkra egy kereső szerkezeti felépítéséről.



1. ábra. Egy keresőrendszer felépítése

Forrás: ([Friedman-Uher-Windhager2003] p. 20)

1.2. Weboldalak begyűjtése és indexelése

A keresők első, lényeges feladata az oldalak meglátogatása és begyűjtése, amit speciális szoftverek, ún. keresőrobotok (angolul: crawlers, web robots, web spiders) segítségével valósítanak meg. Ezek a programok folyamatosan és bizonyos időközönként

átfésülik a webet. Egy keresőrobot választhat egy népszerű, de megbízható oldalt kiindulópontjául, illetve dolgozhat egy korábbi, meglévő adatbázis alapján is. A robotnak le kell töltenie az általa meglátogatott oldalt, és át kell adnia azt az indexelőnek. Ezután az oldalon lévő hiperhivatkozásokat nyomon követve ugyanígy kell annak eljárnia a hivatkozott oldalakkal is.

Számos esetben bizonyos időkülönbség jelentkezik a begyűjtés és az indexelés, valamint az eredmény keresőbe történő beépülése között. Ezért az oldalak begyűjtését és indexelését két, párhuzamosan zajló feladatnak kell tekintenünk. A keresőrobotok tehát nem végeznek semmilyen elemzést a meglátogatott dokumentumon, hanem csak nyomon követik a hivatkozásokat és letöltik a felfedezett oldalakat. Látszólag a robotok nagyon hasonló módon működnek, azonban jelentős különbségek figyelhetők meg a viselkedésükben.

Egy robot számára fontos szempont az, hogy mely hivatkozásokat kövesse nyomon, és mely oldalakat keresse fel, valamint lényeges kérdés, hogy milyen gyakran végezze el az oldalak begyűjtését. Egy keresőrendszer általában több robotot alkalmaz a weblapok begyűjtésére. Emiatt a hálózati forgalom megnövekszik. A robotok igyekeznek nem folyamatosan leterhelni egy szerveret különböző kérésekkel, hanem időben elosztva küldik neki a kéréseket [Szeredi-Lukácsy-Benkő-Nagy2005].

A robotok az oldalak begyűjtése közben egy prioritási sort használnak, amelyben a még meg nem látogatott oldalak címei szerepelnek fontossági sorrendben. A sor elejéről kiveszik a címeket és a hozzájuk tartozó oldalakat, letöltik és kigyűjtik belőlük a hivatkozásokat. A felderített linkekről a robotok eldöntik, hogy melyiket kell követniük, ezeket beteszik a prioritási sorba, a többit pedig elhagyják. A begyűjtés addig tart, amíg a helyi erőforrások, mint például a tárolókapacitás, el nem fogynak ([Friedman-Uher-Windhager2003] p. 20).

A webszervereknek módjukban áll a robotok számára megtiltani egyes oldalak begyűjtését, az oldalon lévő hivatkozások követését és az oldal archiválását. Ezt a Robot Kizárási Szabványban (angolul: Robot Exclusion Standard) megszabott módon tehetik meg [Koster96]. Ha egy weblapra nem hivatkozik egy másik oldal, akkor a keresőrobot nem fogja megtalálni azt. Ezért az új honlapokat tanácsos manuálisan regisztrálnunk az egyes keresőknél, amelyek így indexelni tudják azokat. A keresőket lekérdezhethetjük arról, hogy egy adott oldal indexelve van-e náluk. Azonban a különböző keresőknél eltérő kereséseket kell alkalmaznunk erre a célra [Sullivan2001].

A keresőrobotok által begyűjtött oldalak az adattárba kerülnek (angolul: repository). Az adattár elsődleges feladata az oldalak egyenkénti tömörítése és szekvenciális tárolása. Ezenkívül a rendszer nyilvántartja egy állományban a dokumentumok pontos elhelyezkedését. Az adattár további feladata a meglévő, begyűjtött dokumentumok frissítése is. Ha például módosul egy

oldal, akkor annak az új metaadatait el kell helyezni az adatbázisban, a régit pedig törölni kell ([Friedman-Uher-Windhager2003] p. 20).

A keresők másik lényeges összetevője az `indexelő` (angolul: `indexer`), amelynek fő feladata az adatbázisban lévő meglátogatott oldalak elemzése és az indexelendő kifejezések belőlük történő kigyűjtése [Friedman-Uher-Windhager2003]; [Szeredi-Lukácsy-Benkő-Nagy2005]. Az indexelő tulajdonképpen az adattárra támaszkodik. A feldolgozás elején két problémával találkozhat az indexelő. A weben előforduló oldalak elemzése összetett feladat. Ezt nem csupán a dokumentumok heterogén kialakítása okozza, hanem az egy-egy adott formátum esetén előforduló hibák is, például gépelési, szintaktikai hibák a HTML dokumentumokban. A másik probléma az, hogy az indexelőnek szét kell tudnia választani a fontos és a kevésbé fontos kifejezéseket egy dokumentumban.

Erre egy lehetséges megoldás az, hogy figyelembe vesszük a szavak gyakoriságát és eldobjuk a legkisebb, valamint a legnagyobb gyakoriságú szavakat. Az előbbieket azért, mert azok nem lehetnek fontosak, hogyha csak néhány alkalommal fordulnak elő, az utóbbiakról nagy valószínűséggel állítható, hogy azok felesleges szavak a dokumentumban. Ezenkívül feltételezzük azt, hogy a töltelék- és egyéb szavak eloszlása eltérő egy dokumentumban. Tehát a szavak eloszlásának elemzésével bizonyos szavak kategóriákba sorolhatók.

Azonban a gyakorlatban elterjedt egy másik megközelítés is. Ebben nyelvenként létrehozunk egy ún. `tiltott szó` (angolul: `stopwords`) listát, amely magába foglalja a tartalmi szempontból feleslegesnek tekintett szavakat [Szeredi-Lukácsy-Benkő-Nagy2005]. Az ilyen lista meggátolja a névelők, a kötőszavak és más, szinte minden dokumentumban előforduló szavak indexelését ([Friedman-Uher-Windhager2003] p. 21). Tehát ez a módszer rendkívül gyors, egyszerű és könnyen használható.

A megmaradt releváns kifejezéseket bizonyos jellemzőivel együtt gyűjti ki a dokumentumból az indexelő. Fontos jellemzőnek minősül a szó előfordulásának helye, mint például az oldal címe, a metaelemek, az oldalon belüli pozíció [Szeredi-Lukácsy-Benkő-Nagy2005]. Továbbá az indexelő létrehoz egy indexet, amely minden releváns kifejezéshez hozzákapcsolja az őt tartalmazó URL-ek listáját ([Friedman-Uher-Windhager2003] p. 21). A kigyűjtött indexelendő kifejezéseket és jellemzőiket a tényleges keresés és sorrendezés során veszik alapul a keresők.

1.3. A találatok sorrendezése, rangsorolása

A keresők működésének a leglényegesebb vonása a találatok megfelelő fontossági sorrendben történő megjelenítése a felhasználók számára. Ezért a keresőknek jelentős alkotóeleme a `Rangsoroló` modul, amely egy adott keresésre automatikusan sorrendezi a találatokat

fontosság szerint ([Friedman-Uher-Windhager2003] p. 21). Az indexelt adatmennyiség megnövekedésével vált egyre fontosabb feladattá a találatok pontos sorrendezése. Mivel a felhasználók csak az első 10-20 találatot képesek áttekinteni egy adott keresőkérdésre, ezért rendkívüli fontossággal bír, hogy a kereső milyen találatokat jelenít meg a találati lista elején ([Szeredi-Lukácsy-Benkő-Nagy2005] p. 36-37). Ezért kereséskor célunk a témában íródott legszínvonalasabb weblapok felkutatása, melyhez az oldalakat rangsorolni kell [Friedman-Uher-Windhager2003]. Az egyes keresők által használt rangsorolási szempontokról általában keveset tudunk, de a fő elvek ismertek.

Az egyik legalapvetőbb sorrendezési szempontnak minősül a keresőkifejezés helyének és előfordulási gyakoriságának vizsgálata a dokumentumban. A keresők nagyon gyakran előnyben részesítik azokat az oldalakat, amelyeknek a címében is megtalálható a keresendő kifejezés. A találatok sorrendezésénél azt is figyelembe vehetik, hogy a dokumentum mely részében jelenik meg először a keresőkifejezés. Itt az alapelv az, hogy a weblap szempontjából releváns kifejezések nagy valószínűséggel fordulnak elő már a bevezetésben is, vagy legalábbis a dokumentum elején. Egyes keresők az oldal fontosságának meghatározásához szemügyre veszik a fontméretet is, következtetésekre jutnak a szavak közti távolságokból is, valamint elemzik a HTML-metaelemeket. A metaadatok segítségével közölhetjük honlapunk tartalmának összefoglalóját, valamint az oldalunkra vonatkozó kulcsszavakat. Ezeket a háttér-információkat is hasznosíthatják a keresők a rangsorolás, valamint a keresés közben is.

Másik jelentős vizsgálati szempont a keresőkifejezések előfordulási gyakorisága. Itt azzal a feltételezéssel élhetünk, hogyha egy dokumentumban egy bizonyos kifejezés gyakran fordul elő, akkor az fontos a téma szempontjából. Ebben az esetben természetesen kivételt képeznek a tiltott szavak listáján lévő kifejezések. Továbbá lényeges, hogy ne csak az egyes szavak előfordulási gyakoriságát kövessük nyomon, hanem az adott szóösszetételekét is. A keresők sokszor tanulmányozzák felhasználóik reakcióit is. Amennyiben például a felhasználók többsége nem az első találatra kattint a szolgáltatott találatlistában, akkor nagy a valószínűsége annak, hogy rossz a találatok rangsorolása és nem az első helyen szereplő oldal a legrelevánsabb.

Ezek a felsorolt sorrendezési szempontok sajnos lehetővé teszik, hogy könnyedén befolyásoljuk a találatok rangsorolását. Megfigyelhető az a tendencia, hogy a könnyedén manipulálható rangsorolási szempontok egyre inkább háttérbe kerülnek és csökken a súlyuk a végső sorrend kialakításában. Helyettük pedig olyan kritériumokra helyeződik a hangsúly, amelyeket nehezebb

befolyásolni. Itt megemlíthetők például olyan módszerek, amelyek az oldalak közti linkstruktúrát veszik figyelembe ([Szeredi-Lukácsy-Benkő-Nagy2005] p. 37-38).

A találatok rangsorolásánál kényes etikai kérdésként merülhet fel az, hogy a kereső jó pénzért nem árul-e kulcsszavakat a cégek számára. A megvásárolt kulcsszóért cserébe az adott cég webhelye az első 10 találat között szerepelhet. Ez nem jellemző a nagyobb keresőkre, azonban a felhasználói kulcsszavakhoz kapcsolódó reklámok eladása széles körben elterjedt gyakorlat. Ezekben az esetekben a szoftverfejlesztők úgy változtatják meg a keresők relevancia rangsorolási algoritmusát, hogy az eladott kulcsszó a felhasználót rögtön vezesse arra a webhelyre, amely korábban megvásárolta azt [Froehlich2001].

Egyes keresők a linkhez tartozó szöveget nem a linket tartalmazó, hanem a link által hivatkozott oldalhoz tartozónak veszik. Az ilyen típusú linket *horgony*-nak hívjuk, amit bizonyos keresők a találatok rangsorolásakor használnak fel ([Szeredi-Lukácsy-Benkő-Nagy2005] p. 38). A Google együttkezei a linkek szövegét azokkal a weboldallal, amelyekre azok ténylegesen hivatkoznak. Ennek a módszernek számos előnye van: a linkek sok esetben pontosabb leírást nyújtanak a hivatkozott oldalakról, mint maguk az oldalak. Ezenkívül olyan nem szöveges dokumentumokról is megtudhatunk bizonyos dolgokat, amelyeket egy kereső amúgy nem indexelhet. Ezzel a módszerrel olyan oldalakat is megkaphatunk, amelyeket a keresőrobot nem gyűjtött be a webről. Előfordulhat az is, hogy olyan oldal kerül lekérdezésre, amely korábban nem létezett, azonban az mégis rendelkezik a rámutató hivatkozással. Ilyenkor azonban meggondolandó, hogy érdemes-e a nem ellenőrzött tartalmat visszakereshetővé tenni. A linkek szövegének hatékony felhasználása technikailag nehezen oldható meg, mert az nagy mennyiségű adat feldolgozását igényli [Brin-Page98]; [Szeredi-Lukácsy-Benkő-Nagy2005].

Az Internet megjelenése előtt az egyik legismertebb és sokat használt információ-visszakeresési technika a Vektortér Modell volt. Azonban az közvetlenül nem használható az internetes keresés esetén. Ennek oka egyrészt a világháló mérete és annak állandóan változó tartalma. Másrészt pedig az a mindennapos gyakorlat, hogy az internetes keresőket használók többsége nem definiálja pontosan a keresőkérdést, ez pedig sokszor rossz találatokat eredményez a Vektortér Modell esetében [Szeredi-Lukácsy-Benkő-Nagy2005]. Megjegyzem, hogy a Google kereső (a rendszerről publikusan elérhető információk és tapasztalatok alapján) nem a Vektortér Modellt alkalmazza, hanem az ún. Boole modellt és egy lexikális keresőt épít [Page-Brin-Motwani-Winograd]. A Google döntése ellenére számos próbálkozás irányul arra, hogy a Vektortér Modellt webes környezetben is használható változattá fejlesszék.

Sokan vélekednek úgy, hogy a Google kereső népszerűségét annak köszönheti, hogy a találatokat minőségileg jobban rangsorolja, mint a többi kereső. A Google alkalmazza a fentebb ismertetett

általános módszereket, továbbá kialakít egy olyan speciális algoritmust is, amely kizárólag a linkstruktúrát alapul véve határozza meg az egyes dokumentumok fontosságát. Ezt a fontosságot a kereső megfelelően súlyozza és a többi faktort egyaránt figyelembe véve, dönt a végső sorrend kialakításáról [Szeredi-Lukácsy-Benkő-Nagy2005].

1.4. A Google kereső PageRank algoritmus

A PageRank (PR) egy valós szám, ami egy adott oldal fontosságát tükrözi. A Google kereső a PageRank algoritmust alkalmazza az általa indexelt oldalak fontosságának meghatározásához, amit figyelembe vesz a rangsorolás során. A Google más egyéb szempontokat is felhasznál a sorrend kialakításakor, amelyek közül csak egy a PageRank érték, azonban ez az egyik legfontosabb. A PageRank-kel kapcsolatos eredmények megtalálhatók [Szeredi-Lukácsy-Benkő-Nagy2005]-ben. Fontosságuk miatt néhány alapvető megállapítást részletezek a továbbiakban.

A PageRank algoritmus alap gondolata az, hogy amikor egy oldal hivatkozik egy másik weblapra, akkor a forrásweboldal tulajdonképpen ajánlja a hivatkozott weblapot. Tehát az oldal létrehozója azért tüntette fel a linket az oldalán, mert a másik lapot valamilyen szempontból fontosnak tekintette. Emellett azt is figyelembe kell vennünk, hogy a hivatkozó oldal mennyire fontos, mert egy fontos oldalnak többet ér a hivatkozása. Eredményül egy rekurzív algoritmust kapunk, ami azt fejezi ki, hogy egy oldal fontos, ha mérvadó oldalak hivatkoznak rá. Ez a modell természetesen vitatható, hiszen lehetséges, hogy csak rossz példaként hozunk fel egyes weboldalakot és nem arra szeretnénk velük célozni, hogy azok értékes oldalak. A gyakorlat azonban az eredeti alapötlet sikerességét igazolja, hiszen kevésbé meghatározóak ez utóbbi linkek az Interneten ([Szeredi-Lukácsy-Benkő-Nagy2005] p. 38).

Az alap algoritmust [Page-Brin-Motwani-Winograd98]-ban közölték először. (Nagy valószínűséggel feltételezhetjük azt, hogy a Google most már egy másik változatát használja az itt tárgyaltaknak, amiről azonban nem tájékoztatják a nyilvánosságot) [Szeredi-Lukácsy-Benkő-Nagy2005]. Ez a rekurzív egyenlet a weboldal fontosságára egy megközelítőleges becslést nyújt. A Page Rank normalizálja az egy adott weboldalon található hivatkozások számát (ezek lehetnek az oldalra mutató linkek, valamint ugyanarról az oldalról más egyéb dokumentumokra történő hivatkozások) [Brin-Page98]. Érdekes, hogy a szerzők egyik cikkükben pontatlanul adták meg az egyenlet első tagját és az így terjedt el a szakmában széles körben. Ez a változat megtekinthető az alábbiakban:

$$PR(A) = (1-d) + d \cdot \left(\frac{PR(t_1)}{C(t_1)} + \dots + \frac{PR(t_n)}{C(t_n)} \right) \quad (2.4.1.)$$

Az egyenlet az A oldal PageRank értékét határozza meg. Az egyenletben $t_1 \dots t_n$ jelöli azokat az oldalakat, amelyek A oldalra mutatnak. $PR(t_i)$ fejezi ki az i . ilyen oldal PageRank értékét, azaz annak a fontosságát. A d paramétert egy skálázó faktornak tekintjük, aminek értéke 0 és 1 közé eshet. A d értékét a szerzők $0,85$ -nek határozták meg [Brin-Page98]; [Szeredi-Lukácsy-Benkő-Nagy2005]. C -vel jelöljük az egy oldalon lévő összes kimenő hivatkozás darabszámát. Például, hogyha $C(t_i)$ értékét 24 -nek vesszük, az azt jelenti, hogy az i . oldal összesen 24 darab kimenő hivatkozást tartalmaz, amelyek közül egy biztosan az A oldalra hivatkozik. Az eredeti algoritmus nem számol azzal az esettel, hogy mi történik akkor, hogyha egy oldalról több link is hivatkozik egy másik oldalra.

Az egyenlet tehát a következőt jelenti: az A oldal az első olyan oldaltól, amely hivatkozik rá, $PR(t_1)/C(t_1)$ -nyi szavazatot kap, azaz a t_1 -es oldal egyenletesen elosztja a saját fontosságát a kimenő hivatkozásai között. Ha t_1 oldalon egyetlen kimenő link található, akkor A megkapja a teljes $PR(t_1)$ értéket, ha három, akkor csak t_1 fontosságának a harmadát stb. Ugyanezt az elvet követjük az összes többi olyan oldal esetén, ahonnan találunk hivatkozást A -ra. Ezután ezeket a fontosságokat összeadjuk és megkapjuk A oldal fontosságát. Ebből tehát az következik számunkra, hogy kedvezőbb PR értéket kapunk, ha egy alacsonyabb PR értékű lap mutat ránk, mintha egy magasabb, amennyiben az alacsonyabb fontosságú lapon nem sok kimenő link található. Egy dolgot azonban biztosan kijelenthetünk, hogyha oldalunkra több oldal hivatkozik, nem számít, hogy milyen rangos oldalak, valamilyen mértékben nőni fog a fontosságunk.

A d faktornak köszönhetően egy bizonyos oldal nem a teljes fontosságát osztja szét a kimenő linkjei között, hanem annak csak a 85% -át. Ahhoz, hogy megértsük ezt az összefüggést, szükségünk van egyrészt a javított PageRank egyenletre és a PageRank algoritmus egy újabb jelentésének bemutatására. A javított PageRank egyenlet tehát a következőképpen adható meg, ahol N az összes indexelt weblap számát jelenti.

$$PR(A) = \frac{(1-d)}{N} + d \cdot \left(\frac{PR(t_1)}{C(t_1)} + \dots + \frac{PR(t_n)}{C(t_n)} \right) \quad (2.4.2.)$$

A PageRank algoritmus egy olyan modellnek is tekinthető, amely a „véletlen szörfölő” viselkedését tükrözi. Egy ilyen felhasználó véletlenszerűen elindul egy weboldaltól és a hivatkozásokra véletlenszerűen kattintva folyamatosan előrehalad. Nem is figyeli meg, hogy hova kattint, hanem egyenletes eloszlás szerint választ a meglévő hivatkozások közül. Ezzel magyarázható az, hogy a PageRank algoritmus a kimenő linkek számával elosztja egy bizonyos oldal fontosságát. Mindez addig tart, amíg szörfölőnk meg nem unja a kattintgatást és egy másik

véletlenszerűen kiválasztott weboldalon nem indul el. Ez az egyenlet egy valószínűségi eloszlást határoz meg, ahol egy-egy weboldal PageRank értéke egy valószínűségnek (0 és 1 közötti valószínűség) felel meg. Ebben a modellben az összes weboldal PageRank értékeinek összege maximum 1 lehet. Ez a megállapítás csak abban az esetben igaz, hogyha a felhasználónk egy adott oldalon mindig talál legalább egy hivatkozást, amelyen továbbhaladhat [Brin-Page98]; [Szeredi-Lukácsy-Benkő-Nagy2005].

Hogyha webszájtunk olyan oldalt tartalmaz, amelyre ugyan mutat link, de belőle nem indul kimenő hivatkozás, akkor a szájt nem veszi fel a maximális PageRank értéket. Lógó (angolul: dangling) oldalnak hívjuk az ilyen oldalakat. A Google figyelmen kívül hagyja a lógó oldalakat, mert azok ellentmondanak a PageRank algoritmus által használt „véletlen szörfölő” modellnek. Tehát a Google kereső szűri a lógó oldalakat (az elhagyások miatt esetlegesen újonnan keletkezett lógó oldalakat rekurzív módon szintén figyelmen kívül hagyja). A megmaradt linkstruktúrában kiszámolja a pontos PR értékeket. Ezután fokozatosan visszahelyezi a lógó oldalakat és kiszámolja azok fontosságát is a már kiszámított PR értékek alapján.

A Google nem csupán a linkstruktúrát elemzi, hanem egyéb tényezőket is figyelembe vesz az oldalak rangsorolásakor. Például sokszor negatívan értékeli azt, ha bizonyos, megjelölt oldalakra mutató hivatkozásokat tüntetünk fel az oldalunkon. Nyomon követi azt is, hogy az oldalra történő hivatkozások ugyanabból a domainből, földrajzi területről származnak-e. Tehát a rangsorolás szempontjából többet ér az, ha valaki „független” hivatkozik ránk, mint ha egy „ismerős” szavaz nekünk bizalmat ([Szeredi-Lukácsy-Benkő-Nagy2005] p. 43-44).

A PageRank módszer manipulálása sokkal nehezebb feladat, mint a szöveges dokumentumok sorrendjének befolyásolása. Ennek oka az, hogy a web nagyobb részét kell módosítanunk, valamint hivatkozások sűrű szövevényével kell ellátnunk azt. A Google által alkalmazott rangsorolási módszer ismert a nagy nyilvánosság számára, ezért a világban számos cég specializálódott különféle manipulatív megoldások használatára, amelyekkel a saját forgalmukat tudják indokolatlanul befolyásolni. A cégeknek ezt a törekvését finomabb változatban „kereső optimalizálásnak” hívjuk, erősebb változatban pedig „hivatkozás spam-nek”. A PageRank támadásának egyik közkezdvelt módszere a linkfarmok létrehozása. Ilyenkor nagyszámú és sok szerverre kiterjedő, részben értékes oldalak másolatát, részben számítógéppel előállított oldalakat tartalmazó oldalcsoportot állítanak elő. Itt az oldalak mindegyike a céloldalra hivatkozik, ezáltal magas fontosságot tulajdonítanak annak. Ha elegendő számban áll rendelkezésünkre ismert, illetve spam oldal, akkor a PageRank eljárás megváltoztatásával könnyen kaphatunk olyan büntető eljárásokat, amelyek a spam oldalakat elkerülik ([Benczur-Bíró-Csalogány-Rácz-Sarlós-Uher2006] p. 1329-1330).

1.5. Problémák az internetes kereséssel és a megoldási kísérletek

A kereséssel kapcsolatos problémákat öt fő csoportba soroljuk, amelyek a következők:

1. Általános problémának tekinthető az Internet hatalmas mérete, ami nemcsak a keresést, hanem az oldalak begyűjtését is nagymértékben befolyásolja. A weblapok meglátogatása és feltérképezése időigényes feladatot jelent még a legjobb keresők számára is.
2. Az utolsó begyűjtés óta eltelt idő alatt az Internet tartalma és szerkezete gyorsan megváltozik, ami további nehézségeket eredményez [Hawking-Craswell2005].
3. A keresőrendszerek számára általában elérhetetlenek azok az Interneten meglévő tartalmak, amelyek a mély web (angolul: deep web) körébe sorolhatók.
4. A keresőrobotok nem gyűjtik be a dinamikus weblapokra mutató hivatkozásokat.
5. Az internetes keresők nem a felkutatható dokumentumok és a keresőkérdés jelentésével foglalkoznak, hanem csupán a szöveges alakkal.

A továbbiakban részletezem, hogy milyen lehetőségekkel oldják meg eredményesen ezeket a felmerülő problémákat.

Az óriási adattömeg visszakeresését oldják meg a metakeresők, amelyek párhuzamosan más keresőkkel kerestetnek. Így azok az Internet nagyobb részét képesek átfésülni [Szeredi-Lukácsy-Benkő-Nagy2005]. Növelik a találati esélyünket az ismeretlen témák esetében, valamint átfogóbb képet nyújtanak számunkra a weben fellelhető információkról egy adott témában.

A gyorsan változó tartalom kezelésére használható az oldalak begyűjtésének fókuszált módja (angolul: focused crawling). Ennek a módszernek lényege az, hogy nem követünk minden hivatkozást, hanem valamilyen szempontrendszer szerint egy bizonyos területhez kapcsolódó oldalakra szűkítjük a keresési teret, például nevezetes hírportálok meglátogatására. A fókuszált begyűjtést végző robotokkal kialakíthatunk egy-egy adott területre specializálódott keresőt is. Létrehozhatunk például egy olyan keresőt, amely orvosi tartalmak indexelésére és orvosi szakterületen feltett kérdések megválaszolására alkalmas.

A mély web csoportjába tartoznak a weben keresztül lekérdezhető adatbázisok, valamint a nem szöveges formában található dokumentumok. Ez az adatmennyiség azért nem elhanyagolható számunkra, mert becslések szerint a mély weben nagyságrendekkel több információt tárolunk,

mint a hagyományos weboldalakon. Ezzel szemben *sekély webnek* (angolul: *surface web*) hívjuk az Internet statikusan elérhető részét, azaz azokat a kész, meglévő weboldalakat, amelyeket a szerver külön kérésre eljuttat a klienshez az Interneten keresztül. A mély- és a sekély web közötti határ élesen nem különül el egymástól. A kettő között húzódik az ún. szürke zóna (angolul: *gray zone*), amelyben nem egyértelmű, hogy milyen jellegű tartalmak lelhetők fel. A mély web kezelését úgy támogathatjuk, hogyha a keresők számára is elérhető metainformációkat közlünk az adatbázisok tartalmáról, valamint különböző csatoló programokat hozunk létre a nem szöveges állományokhoz (PDF, Excel, JPG stb.).

Újabb nehézség az is, hogy a keresőrobotok nem követik a dinamikus weblapokra mutató hivatkozásokat, ezáltal azok sok információhoz nem férnek hozzá. Ennek oka az, hogy a dinamikus linkek gyakran hoznak létre hatalmas vagy esetleg végtelen keresési tereket. Ezeket *keresőcsapdának* (angolul: *spider trap*) hívjuk, amelyeket a keresőrobotok megpróbálnak elkerülni. Gyakran előfordul az, amikor bizonyos szerverek megkísérlik álcázni magukat és egy keresőrobotnak eltérő tartalmat nyújtanak, mint például egy böngészőnek. Napjainkban számos technika terjedt el a dinamikus oldalak indexelésének támogatására, amelyek lényege az, hogy elhitetjük a keresőrobotokkal, hogy statikus hivatkozást követnek.

Az internetes keresők számára a legnagyobb probléma az, hogy nem a fellelhető dokumentumok és a keresőkérdés jelentésével foglalkoznak, hanem csupán a szöveges alakokkal. A nyelvi problémákat tulajdonképpen az okozza, hogy a mai eszközökkel történő információ-visszakeresés túlságosan a letárolt szöveges információ tényleges alakjára épül. Ennek egyik következménye az, hogy a nem szöveges dokumentumok által tárolt információk nem kereshetők vissza automatikusan. További hiányosságként kiemelhetjük azt is, hogy a keresőrendszerek nem ismerik a fogalmak jelentését és a fogalmak közötti kapcsolatokat, ezért azok nem képesek különféle következtetések levonására [Szeredi-Lukácsy-Benkő-Nagy2005]. Ezt a problémát a szemantikus keresők orvosolják hatékonyan.

Az internetes keresőknek létezik egy másik fajtája, a *webes katalógusok*, amelyek emberek által összegyűjtött oldalakat tesznek visszakereshetővé. Ezek a katalógusok eredményesen oldják meg a jelentés, azaz a szemantika megragadását, ami az oldalak begyűjtését és indexelését végző emberek feladata. Előnyük az, hogy oldalaik megbízhatóak és minőségük egyaránt garantált, hiszen emberek válogatják ki azokat. Itt nemcsak szöveges keresésre van lehetőségünk, hanem témakategóriák között is böngészhetünk. Ezekben a gyűjteményekben nagy segítséget jelent számunkra az, hogy az oldalak kategorizálásakor emberek olvasták végig azokat. Legnagyobb hátrányuk viszont az, hogy lassan frissülnek és meglehetősen szűk körűek. Ezenkívül meg kell

említenünk a kérdésátalakító keresőket is, amelyek szintén a jelentés megragadására törekednek. Feladatuk, hogy megpróbálják jobban értelmezni a feltett keresőkérdést és azt úgy átalakítani, hogy az új keresőkérdés már jobb találatokat eredményezzen. Egy ilyen átalakításhoz a keresőknek rendelkezniük kell bizonyos háttértudással, amely valamilyen matematikai formalizmussal írható le.

A szemantikus web irányzat hatékonyan oldja meg a jelentéssel kapcsolatos problémakört, amelynek fő célja, hogy jelentést vigyen a webre. Ezt úgy teszi lehetővé, hogy a webes tartalmakhoz szabványos formában metainformációt rendel és biztosítja számunkra azt, hogy ezen metainformációk alapján következtetéseket vonjunk le. Jelenleg a metainformációk ugyanolyan heterogén formában fordulnak elő, mint maguk a webes dokumentumok. Ezért a szemantikus webnek elsődlegesen a metainformációk és a következtetéshez szükséges háttértudás egységes és feldolgozható alakban történő leírására kell törekednie [Szeredi-Lukácsy-Benkő-Nagy2005].

Irodalom:

[Brin-Page98] 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.

[Benczur-Bíró-Csalogány-Rácz-Sarlós-Uher2006] 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.

[Friedman-Uher-Windhager2003] 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.

[Froehlich2001] Froehlich, T. J.: Case study 5.1: Developing search engine evaluation criteria. In.: Library evaluation. Libraries Unlimited, 2001. p. 185-200.

[Hawking-Craswell2005] 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.

[ITTKa2007] 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.

- [ITTKb2007] Magyar információs társadalom jelentés 1998-2008: Jelentés az elmúlt évtizedről. Készít. a BME-UNESCO Információs Társadalom- és Trendkutató Központjának (ITTK) kutatócsoportja, GKIE NET és az MTA Infokommunikációs Jogi Centrum kutatói. Bp. 2007. december
http://www.ittk.hu/web/docs/ITTK_MITJ_1998-2008.pdf 2008.01.28.
- [Koster96] 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.
- [Page-Brin-Motwani-Winograd98] 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.
- [Sullivan2001] Sullivan, D.: Checking your listing in search engines, October 2001.
<http://searchenginewatch.com/webmasters/article.php/2167861> 2007.07.10.
- [Szeredi-Lukácsy-Benkő-Nagy2005] 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.

2. Az internetes keresők értékelése

Az internetes keresők sokszor megküzdnek a dinamikus változásokkal, rosszindulatú felhasználókkal, kommunikációs problémákkal, szerverhibákkal és még számos egyéb akadállyal a hálózaton. Adatbázisaiktól elvárható lenne, hogy gyorsan és hűen tükrözzék a weben bekövetkező változásokat, de azok instabilitást mutatnak. A keresőszolgáltatások sokszor szokatlan és váratlan helyzeteket kezelnek a hagyományos információkereső rendszerekhez képest. Miután egy rekord bekerül a hagyományos információkereső rendszer adatbázisába, az ritkán módosul, vagy kerül ki annak az adatbázisából. A weben azonban eltérő helyzet áll fenn, hiszen új dokumentumok jelennek meg folyamatosan, a régi dokumentumok máshová kerülnek át, vagy azok teljesen törlődnek, a meglévő dokumentumok tartalma gyakran változásokon megy keresztül. Ezekből a változásokból adódóan tehát szükséges mérni a keresők működését egy bizonyos időtartam alatt ([Bar-Ilan2002a] p. 308).

2.1. A keresők mérésére irányuló szabvány létrehozása

Kezdetben a W3C konzorcium Web Characterization Activity nevű munkacsoportja a mérési módszereknek egy sorát definiálta, azonban ezek közül egyik sem kapcsolódott a keresőeszközökhöz. 1999 végére a WCA munkacsoport megszüntette a tevékenységét, ami nem bizonyult sikeresnek ezen a területen ([Bar-Ilan2002a] p. 318). Az eddig megjelent értékelésekben a kutatók az internetes keresők minőségét számos mérték alapján mérik. Ezeknek a megbízható mutatóknak a megtalálása rendkívül nehéz feladat, számos vita folyik erről a kutatók körében. *Oppenheim* és *Froehlich* egy olyan szabvány kidolgozását javasolták, amely az alábbi mértékeket tartalmazná:

1. pontosság, azaz a visszakeresett dokumentumok összességén belül a releváns dokumentumok aránya ([Ungváry2001] p. 196) (46/I, 47/II, 47/III.);
2. teljesség, a megtalált releváns dokumentumoknak az összes (akár talált, akár nem) releváns dokumentumhoz viszonyított arányát jelenti ([Ungváry2001] p. 196-197) (47/II);
3. relatív teljesség, azaz az egy adott kereső által visszakeresett releváns dokumentumok száma osztva az összes vizsgált kereső valamelyike által megtalált releváns dokumentumok számával ([Clarke-Willett97] p. 185-187);
4. válaszidő, vagyis az, hogy mekkora az átlagos időeltérés a keresőkérdés feltétele és a válasz megadása között ([Ungváry2001] p. 196);
5. tesztelési idő, azaz a tesztelésre fordított teljes időtartam hossza ([Oppenheim-Morris-McKnight-Lowley 2000] p. 207);

6. egy adott időtartamon belül a találatok megbízhatóságának mérése, melynek során a találatokat vizsgáljuk, hogy az adott keresőkérésre mindig ugyanazokat a találatokat kapjuk ([Oppenheim-Morris-McKnight-Lowley 2000] p. 207) (46/I);
7. zsákutcás, halott hivatkozások aránya;
8. ismétlődő találatok aránya;
9. találatok minősítése felhasználókkal;
10. grafikus felhasználói felület értékelése (használatra mennyire felhasználóbarát?);
11. a súgó és a keresőprogram mennyire hasznos a kezdő és a tapasztalt felhasználók számára;
12. találatmegjelenítés, azaz az output megjelenítési módja ([Ungváry2001] p. 196);
13. reklámok jelenléte;
14. a gyűjtemény vélhető érdeklődést lefedő volta (angolul: coverage=terjedelem), vagyis, hogy milyen mértékben tartalmaz a rendszer releváns dokumentumokat. Ezt a mutatót egy adott kereső adatbázisában lévő összes releváns dokumentum és az összes vizsgált kereső adatbázisában található releváns dokumentumok hányadosa adja ([Clarke-Willett97 p. 187]), ([Landoni-Bell2000 p. 127]), ([Ungváry2001] p. 196) (47/II, 55/VII);
15. elvárt keresési hossz, azaz átlagosan hány nem releváns dokumentumot kell áttekintenie a felhasználónak ahhoz, hogy bizonyos számú releváns dokumentumot megtaláljon ([Chignell-Gwizdka-Bodner99] p. 350-351);
16. találatleírások terjedelme és olvashatósága.

Egyéb értékelési szempontok:

17. Testreszabhatóság: a keresőszolgáltatás testreszabható-e?
18. Vizuális egyértelműség: a keresőkérés és a találatmegjelenítés világos, érthető és következetes-e?
19. Navigáció: egyértelmű-e a navigációs eszközök használata? Könnyen tudunk-e a találatmegjelenítéstől a forrásdokumentumig eljutni és fordítva?
20. A keresőszolgáltatás felkínál-e valamilyen webes katalógust, amiben böngészhetünk ([Oppenheim-Morris-McKnight-Lowley 2000] p. 207), ([Froehlich2001] p. 192)?

Ez a lista természetesen bővíthet további új mérési jellemzőkkel. A mértékek általános szabványának hiánya nagy problémát jelent az értékelésekben. Ezen hiányosság miatt a keresőszolgáltatások értékelésével foglalkozó kutatás jelenleg nem egységes az alkalmazott módszerek tekintetében, ezért nagy szükség lenne megfelelő szabvány kidolgozására és

használatára a méréseknél. Statisztikai szempontból érdekes lenne azt is megvizsgálni, hogy létezik-e valamilyen összefüggés ezen mutatók között ([Tóth2004] p. 106). A rendelkezésemre álló szakirodalom azt tükrözte számomra, hogy mindenki maga választja ki a vizsgálatához szükséges mutatókat és nincs közmegegyezés arról, hogy milyen mértékek használata lenne elengedhetetlenül fontos egy mérés elvégzéséhez.

2.2. Statisztikai módszerek alkalmazása az információkeresés területén

A korábbi statisztikai módszerek összefoglalása megtalálható [Harter97]-ben. Kezdetben csekély mértékben foglalkoztak a szignifikancia tesztelésével, illetve más egyéb statisztikai jellegű kérdéssel a szakirodalomban. Ez részben annak volt köszönhető, hogy a cranfieldi eszközök csak ritkán elégítették ki a statisztikai műveletek alkalmazásának feltételeit [Harter97]. Cranfieldben számos vizsgálatot végeztek az 1960-as években, melyek során 33 osztályozási rendszer hatékonyságát hasonlították össze. Mértékeket dolgoztak ki különböző szempontok megválaszolására, mint például: az információkereső nyelv kiválasztja-e a releváns tételeket vagy sem? Visszamaradnak-e releváns tételek vagy sem? A létrehozott mutatórendszer ezeknek a szempontoknak a kölcsönös összehasonlításán nyugszik. Miután egy keresőkérdés alapján elvégeztek egy irodalomkutatást, az osztályozási rendszer tételei négy csoportba voltak sorolhatók:

a: releváns és visszakeresett dokumentumok;

b: releváns és nem visszakeresett dokumentumok;

c: nem releváns és visszakeresett dokumentumok;

d: nem releváns és nem visszakeresett dokumentumok.

Ahol $a+b+c+d$ megfelelt az *N*-nek, azaz a teljes gyűjteménynek. A releváns dokumentumokat $a+b$ jelentette, a visszakeresett dokumentumokat $a+c$, az összes irreleváns dokumentumot pedig $c+d$ jelentette.

Ez alapján a következő mérőszámokat hozták létre:

1. teljesség:
$$R (recall) = \frac{a}{a+b}$$

2. pontosság:
$$P (precision) = \frac{a}{a+c}$$

3. fölösleg:
$$F (fallout) = \frac{c}{c+d}$$

A teljesség azt jelenti, hogy a releváns dokumentumok hányada került elő. A pontosság azt mutatja, hogy a visszakeresett dokumentumok hányad része releváns. Néha az is vizsgálendő, hogy vajon a teljes gyűjteményben található irreleváns dokumentumok hányada fordul elő

találatként. A fölösleg mutatót a visszakeresett irreleváns dokumentumok és az összes irreleváns dokumentum hányadosa határozza meg ([Horváth-Sütheő 2003] p. 176-179).

Tague és *Farradane* minta keresések segítségével kiszámította a teljesség és a pontosság standard hibáit negatív binomiális eloszlás alapján. *D. Miller* és *Dattola* több módszert ismertetett a releváns dokumentumok összességének becslésére egy gyűjteményen belül. Erre a becsült számértékre a teljesség kiszámításakor volt szükség. Módszereiket empirikusan tesztelték egy Cranfield II-es 1400 dokumentumból álló gyűjteményen. *Wilbur* kifejlesztett egy olyan eljárást, amely két statisztikailag független lekérdezési módszer keresési taláatait hasonlította össze. Ezt a két módszert ugyanarra a MEDLINE adatbázisra alkalmazta.

Keen olyan statisztikai és informatikai jellegű problémákkal foglalkozott, hogy hogyan lehetne szabványosítani, összegyűjteni, megjeleníteni és összehasonlítani a teljesítményre vonatkozó mutatókat, beleértve a pontosság, a teljesség és a találatmegjelenítés felhasználó irányú megközelítéseit. *Hull* számos olyan megközelítést részletezett, amely a keresők lekérdezési teljesítményében jelentkező különbségek értékelésére irányult két vagy több statisztikai módszer esetében. Kifejtette, hogy hogyan lehet könnyedén tesztelni a normalitás és a folytonosság szükséges feltételeit. Véleménye szerint több nem paraméteres eljárás is alkalmas volt erre a célra. *Wilbur* „bootstrap módszereknek” nevezte a nem paraméteres statisztikai próbáknak egy csoportját. Úgy vélte, hogy ezek a módszerek jól alkalmazhatóak egy vagy több keresőrendszer lekérdezési teljesítményének az összehasonlítására ([Harter97] p. 12-13).

Fontos előrelépések történtek a találati eredmények statisztikai elemzésének a területén. *Shaw* és társai a hipergeometriai eloszlást használták a véletlenszerűen várható lekérdezési teljesítmény kiszámítására. Rájöttek arra, hogy az aktuális lekérdezési teljesítmény csupán néhány releváns dokumentumban tért el a véletlenszerűen várható lekérdezési teljesítménytől. Ez a megállapítás érvényes volt az összes általuk vizsgált hagyományos tesztgyűjteményre is. *Burgin* továbbítte ezt a gondolatmenetet és a Monte Carlo módszert használta a hipergeometriai eloszlás helyett, ami már gépi erőforrásokra támaszkodott. A Monte Carlo módszer számos előnye mellett lehetővé teszi azt is, hogy meghatározzuk statisztikailag szignifikáns-e a különbség egy visszakeresett találatpár között. *Burgin* műve elsőként épült már meglévő statisztikai elemzési módszerekre, amelyben az egyéni keresőkérdésekre helyezte a hangsúlyt az összegyűjtött találati eredmények és az átlagok helyett ([Harter97] p. 12-13).

2.3. Az internetes keresők elemzésével kapcsolatos problémák

Az internetes keresők elemzésére irányuló kísérletek nem egységes módszerekkel valósultak meg, ezért a legtöbb kutatási beszámoló tájékoztató jellegűnek tekinthető. Eddig főként

indexelőszolgáltatásokon végeztek ilyen jellegű méréseket, de elvileg bármilyen típusú kereső értékelhető. *Leighton és Srivastava* szerint sok összehasonlító értékelés ellentmondásos következtetésekre jutott arra vonatkozóan, hogy melyik szolgáltatás nyújtja a legrelevánsabb találatokat [Leighton-Srivastava97].

Ezenkívül több tanulmány kisebb tesztelési kísérletek eredményeire épült, és nem számolt be az általa használt módszertanról [Oppenheim-Morris-McKnight-Lowley2000]. Mivel ezek a tanulmányok kevés tesztelési eredménnyel rendelkeznek, ezért azok nem alkalmasak a mélyebb szintű statisztikai elemzések végzésére. A népszerű folyóiratokban közölt tanulmányok gyakran nem ismertetik azt sem, hogy hány keresést futtattak le mérés közben, és konkrétan milyen keresőkérdéseket alkalmaztak [Leighton-Srivastava97]. A megjelent értékelések főként rendszerjellemzőket vizsgáltak és azok összehasonlításait közölték több keresőszolgáltatás vonatkozásában [Harter97].

Su megállapítja, hogy hiányzik a szisztematikus megközelítés ezekből a tanulmányokból. Rámutat arra, hogy a kutatóknak nincs egységes kialakult véleménye arról, hogy mit mérjenek, és hogyan mérjenek egy szolgáltatást. Megjegyzi, hogy a legtöbb tanulmányból a felhasználók, mint aktív közreműködők kimaradnak. Általában az első 10 vagy 20 lekérdezett találat relevanciáját mérik. A relevancia ítéleteket többnyire a kutatók hozzák meg, nem pedig a felhasználók. A mérési eredmények rendszerint azt jelzik, hogy a legjobbnak minősített második vagy harmadik kereső között a teljesítménybeli különbség minimális [Su97].

A keresési folyamat sokféleképpen befolyásolható, például választható egy olyan tárgykör, amelyről köztudott, hogy az egyik kereső színvonalasabb szolgáltatást nyújt róla, mint a másik [Leighton-Srivastava97]. Fontos, hogy az értékelők ne legyenek elfogultak egyik keresővel szemben sem, és megőrizzék pártatlanságukat objektív módszerekkel. Ennek a szemléletnek tükröződnie kell a keresőkérdések megválasztásában is. Ez a kritika természetesen egyaránt vonatkozik a hagyományos információkereső rendszerekben végzett tesztelésekre is.

A keresők minőségi információszolgáltatásában problémát jelent a reklámok jelenléte, ami főként a hirdetőik érdekeit szolgálja, és távol esik a felhasználók tényleges elvárásaitól. A keresők értékelését nagymértékben nehezíti az a tény, hogy a keresők képesek befolyásolni a keresési találatokat rendkívül burkolt formában. Amennyiben nem feltűnő ez a befolyásolás, akkor azt maga a piac is képes tolerálni. A reklámozásból származó bevételek többnyire ösztönzik a keresőket a gyengébb minőségű találatok szolgáltatására [Brin-Page98].

Az értékeléseknél külön gondot jelent az internetes keresők állandó változása, mivel azok gyakran fejlesztik a keresési mechanizmusait és a felhasználói felületüket. Ehhez párosul még az a tény, hogy a világháló egy dinamikusan változó közeg. Ennek eredményeként a megjelenő

értékelések rendkívül rövid életűek, és csak a pillanatnyi helyzetképet tükrözik a keresőszolgáltatásokról. Mindezek ellenére igenis van értelme elemezni a használatban lévő keresőket, de szabványos értékelési módszereket erre a feladatra nem alkalmazhatunk. Általában megfigyelhető az a jelenség, hogy az elemzések egyéni értékelési módszerekről számolnak be, és többnyire elkerülik a szabványos értékelési módszerek használatát ([Oppenheim-Morris-McKnight-Lowley2000] p. 193-194).

Az információkeresés területén végzett kutatások rámutatnak arra, hogy rendkívül nehéz megfelelő mértékeket találni az értékelésekhez. Leggyakrabban a pontosságot és a teljességet használják mutatókként, amelyek érzékenyek arra, hogy a relevanciát hogyan definiáljuk és mérjük [Harter96]. A teljesség mérése megköveteli, hogy az értékelők a megtalált releváns találatok teljes halmazához hozzáférhessenek, vagy a releváns találatokat képviselő mintához. Ennek a követelménynek a kielégítése külön problémát jelent az internetes keresők értékelésénél [Oppenheim-Morris-McKnight-Lowley2000]. Sokan érveltek amellett, hogy nem lehet mérni a teljességet, mert nehéz meghatározni a (visszakeresett és a nem visszakeresett) releváns találatok összességét egy bizonyos keresésre a weben [Leighton95]. Az eddig megjelent tanulmányok minimális mértékben, vagy egyáltalán nem kísérelték meg a különböző keresők teljességének a mérését [Clarke-Willett97].

A pontosság mérése főként emberi relevancia ítéletektől függ, ezért az rendkívül szubjektív [Bar-Ilan2002a]. Azonban annak mérése egyértelmű, hiszen a keresési találatok megvizsgálása után azokat a releváns, illetve a nem releváns dokumentumok halmazába soroljuk [Oppenheim-Morris-McKnight-Lowley2000]. *Spink* és *Greisdorf* szerint jobb eredményeket nyerhetünk, ha a bináris mérés helyett a relevancia többféle szintjét definiáljuk [Spink-Greisdorf2001]. Ezeket a relevancia szinteket már korábban is felhasználták az internetes keresők pontosságának a mérésére [Clarke-Willett97]; [Ding-Marchionini96]; [Leighton-Srivastava99]. *Gordon* és *Pathak* hangsúlyozta, hogy a relevancia ítéleteket csak az eredeti információs igényekkel rendelkező egyének hozhatják meg [Gordon-Pathak99]. Mások a relevancia ítéletek meghozatalát szakértőkre, illetve szakértői csoportokra bízák.

A keresők lekérdezési teljesítményének mérésekor egyéb mértékeket is alkalmaztak, például elemezték a keresők gyűjteményének lefedettségét a weben, a felhasználói megelégedettséget, a felhasználók keresési viselkedését és a találatok megjelenítését [Bar-Ilan2002b].

Az eddig megjelent értékelésekben használt módszerek négy fő csoportba sorolhatók:

1. cranfieldi tanulmányok készítése egy szűk, behatárolt témáról, ahol a kutató ismeri egyenként a találatokat, és a releváns találatok kis halmazát vizsgálja meg.

2. A relatív teljességet elemző cranfieldi tanulmányok. Ebben az esetben a különböző keresőszolgáltatásokkal nyert releváns találatokat összeadják, amelyek a releváns találatok halmazát képviselik. A kutatók ezután egy adott internetes kereső esetében elemzik, hogy az mennyire képes a releváns találatok halmazát visszakeresni.
3. Statisztikai módszerrel mintát vesznek a webről, amelyben felmérik a releváns Web oldalak számát. Ezen a mintán vizsgálják a teljességet, és egy cranfieldi tanulmányt készítenek róla.
4. A teljességet figyelmen kívül hagyó tanulmányok. Ezek a kísérletek megpróbálják a teljességet más mérési paraméterekkel helyettesíteni, mint például: gyűjtemény lefedettség, indexelés, felhasználói felület, keresési technikák, találatok megjelenítése, stb. ([Oppenheim-Morris-McKnight-Lowley2000] p. 194).

Értékelési problémák származhatnak a következtelen relevancia ítéletekből, az automatizált technikák nem megfelelő használatából, és abból, hogy a téma másképpen szerepel a kezdeti kérdésfeltevésben, mint a keresőhöz ténylegesen elküldött keresőkérdésben. Az értékelésekben használt mutatók mérési sajátosságai további problémákat okozhatnak. Helytálló tehát az a megállapítás, mely szerint azok a mértékek, amelyek egy adott esetben jól működnek nem biztos, hogy egy másik helyzetben megfelelőek lesznek [Chignell-Gwizdka-Bodner99].

2.4. A keresőszolgáltatásokon végzett jelentősebb statisztikai vizsgálatok

A kilencvenes évek közepétől kezdődően számos statisztikai vizsgálat készült a keresők működésével kapcsolatban. Mivel a dolgozat kereteit meghaladná a korábbi vizsgálatok teljes körű áttekintése, ezért az általam legfontosabbnak tekintett vizsgálatok módszereit és eredményeit szeretném ismertetni ebben a részben.

I. Sroka elvégezte a helyi fejlesztésű keresők és az angol nyelvű keresők lengyel változatának értékelését. Összesen hat keresőszolgáltatást tanulmányozott, amelyek a következők voltak: Polski Infoseek, Polska AltaVista, Wirtualna Polska, NEToskop, Onet.pl., WOW. Mérésében a pontosságra helyeződött a hangsúly, amit relevancia ítéletek alapján állapított meg, azonban a weboldalak megbízhatóságát már nem elemezte. Összesen tíz keresőkérdést állított össze különböző témákban, ezek közül négy keresést Boole-operátorokkal (AND, AND NOT) fogalmazott meg. A lekérdezett első tíz találatot vizsgálta meg és kiszámolt egy pontossági értéket minden egyes keresésre vonatkozóan. Ezt úgy nyerte, hogy megnézte hány releváns találatot kapott a tíz találatból, és ennek a két értéknek a hányadosát vette. Ezután kiszámította, hogy egy kereső az első tíz találatból hány releváns dokumentumot kért le átlagosan.

Továbbá megvizsgálta, hogy egy adott keresésre a keresők átlagosan hány releváns dokumentumot találtak, ez alapján pedig következtetéseket vont le a keresőkérdések nehézségére vonatkozóan. Ezenkívül tanulmányozta a lekérdezett találatok közötti átfedést és a vizsgált keresők válaszüdejét. Keresőnként rögzítette az összes lekérdezett találat számát, azonban a teljességet, mint mértéket figyelmen kívül hagyta. Tesztelés közben a lengyel diakritikus jeleket használta. Kereséseket végzett azokkal, valamint azok nélkül. Megállapította, hogy azok használatával a releváns találatok legbővebb halmazát kapta a keresésekre ([Sroka2000] p. 92-95).

II. Clarke és Willett összesen harminc keresést futtatott le három angol nyelvű keresőn, amelyek az AltaVista, Excite, Lycos keresők voltak. A lekérdezett első tíz találat relevanciáját egy három fokozatból álló skála alapján határozták meg, amely a következőképpen definiálta a relevancia értékeket: a releváns weboldalak 1-es értéket kaptak, a részben releváns dokumentumok 0,5-ös értéket kaptak, a nem releváns dokumentumok pedig 0-s értéket kaptak. Meghatározták a pontosság, a teljesség és a gyűjtemény lefedettség átlagértékeit, azaz összeadták mindegyik keresésre a releváns, a visszakeresett és a releváns visszakeresett találatokat, ezeknek az összesített értékeknek az általános átlagértékeit számolták ki.

A keresők lekérdezési teljesítményében tapasztalt jelentősebb különbségeket a *variancia teszt Friedman-féle kettős analízisével* (angolul: *Friedman two-way analysis of variance test*) értékelték. Korábban *Leighton és Srivastava* használta ezt a tesztet a visszakeresett találatok elemzésére ([Clarke-Willett97] p. 187). Ennek az alkalmazott statisztikai módszernek a célja, hogy tesztelje azt a nullhipotézist (H_0 -t), hogy k különböző minta hozható létre azonos mediánnal rendelkező populációkból (ebben a mérésben a minták a három keresőszolgáltatásnak felelnek meg). Az adatok egy olyan táblázatban szerepelnek, amely N számú sorból és k számú oszlopból áll (itt a sorok azonosak a huszonhét kereséssel). Az adatok minden sorban 1- k -ig kerülnek rangsorolásra. A Friedman-féle statisztika, F_r , a keresők rangsor összesítésére épül [Siegal-Castellan88].

III. Leighton és Srivastava öt angol nyelvű keresőszolgáltatás (AltaVista, Excite, Infoseek, Hotbot, Lycos) lekérdezési hatékonyságát hasonlította össze. Egy tizenöt keresésből álló tesztsorozatot állítottak össze a keresők számára. Mérték az első húsz visszakeresett találat pontosságát, azaz megállapították a releváns találatok százalékos megoszlását az első húsz találat esetében. A találatok relevancia szerinti értékelése hat relevancia kategória alapján történt, ezek pedig a következők voltak:

<i>Ismétlődő találatok:</i>	Ebben az esetben az adott találat megegyezik egy korábbi weboldallal a találati listában, például az lehet inaktív, releváns találat. A tükrözött webhelyek nem minősülnek ismétlődő találatoknak.
<i>Inaktív találatok:</i>	A technikailag el nem érhető találatoknál jelennek meg a következő hibakódok: a „hozzáférés megtagadva” 403-as hibakód, a „szerver nem válaszol” 503-as hibakód, valamint az „állomány nem található meg” 404-es hibakód. A „hozzáférés megtagadva” 403-as hibakódnál, valamint az 503-as hibakódnál a hivatkozásokat többször újra ellenőrzik hosszabb időtartamon keresztül [BernersLee-Fielding-Frystyk96, RFC1945].
<i>0. kategória:</i>	A weboldal irrelevánsnak tekinthető, mert az nem elégíti ki a keresőkifejezés fontos aspektusát.
<i>1. kategória:</i>	A weboldal technikailag kielégíti a keresőkifejezést, azaz az tartalmazza a megadott keresőkifejezést, illetve annak szinonimáit. Azonban az oldal nem minősül relevánsnak, hogyha az nem kapcsolódik a megjelölt témához, illetve az terjedelmében túlságosan rövid, hogy használható legyen a felhasználó számára.
<i>2. kategória:</i>	A weboldal relevánsnak minősül és kielégíti a keresés által támasztott információs igények szűk körét. Ezek az oldalak potenciálisan relevánsak néhány felhasználó számára. Ebbe a kategóriába tartoznak azok a weboldalak is, amelyek hivatkoznak a 3-as relevancia kategóriába tartozó oldalakra.
<i>3. kategória:</i>	A weboldal relevánsnak tekinthető és kielégíti a keresés által támasztott információs igények széles körét, például az a hivatkozások gyűjtőoldalaként jelenik meg, illetve az rendkívüli alapossággal kezeli a témát.

1. táblázat. Leighton és Srivastava mérésében használt relevancia kategóriák

Öt különböző tesztelést végeztek a keresésekre kapott első húsz találat pontosságának a mérésére. Az első húsz találat pontossága általában egy szolgáltatás minőségét tükrözi. A pontosság kiszámításában súlyozást alkalmaztak, ezzel is megnövelve a találatrangsorolás hatékonyságának az értékét. A pontosságot és a találatrangsorolást egyetlen mérőszámban kapcsolták össze. Ez a mérőszám több tulajdonságot ötvözött magába, amelyek a következők voltak: először is az figyelembe vette, hogy a találat kielégítette-e a relevancia iránti követelményeket vagy sem. Bináris relevancia skálát használtak az értékelők, mert a relevancia kategóriákat különböző definíciók formájában határozták meg. Másodsor nagyobb súlyokat adtak a releváns weboldalak hatékony rangsorolásának. Harmadszor a statisztikának tükröznie kell azt a tényt is, hogyha egy kereső kevesebb találatot kérdez le, de ugyanannyi jó hivatkozást nyújt, akkor a felhasználó sokkal könnyebben találja meg a releváns találatokat a találathalmazon belül.

Ezután a kutatóknak el kellett dönteniük, hogy hogyan kezeljék az inaktív és az ismétlődő találatokat. A mérés első három tesztjében az ismétlődő találatokat csak a pontossági arányszám

számlálójából vették ki, és a szolgáltatásokat büntették emiatt. Az utolsó két tesztben az ismétlődő találatokat nem értékelték negatívan. A lekérdezett inaktív találatokat mind az öt kísérletben negatívan értékelték. Azonban nem vették figyelembe az ismétlődő találatok másik két típusát: a tükrözött weboldalakat és a weboldal klasztereket (azokat az oldalakat, amelyek ugyanabból a nyilvántartásból vagy szerverről származnak).

A keresőszolgáltatás lekérdezési teljesítményét meghatározó formulát 0 és 1 közötti értékben adták meg. Mindegyik vizsgált kereső esetében az első húsz visszakeresett találatot státusz – azaz eldöntötték a találatról, hogy az ismétlődő vagy inaktív találatnak minősül-e – és relevancia típus alapján csoportosították. A formula azzal kezdődött, hogy a relevancia kategóriákat 0 és 1 bináris értékekre konvertálták át. Ez azt jelentette, hogy amikor lefuttattak egy tesztet azokra a weboldalakra, amelyek minimálisan kielégítették a keresőkifejezést – ide tartoztak az 1-es, 2-es, valamint a 3-as relevancia kategóriákba sorolt weboldalak –, akkor azok 1-es értéket kaptak. Az összes többi weboldal pedig 0-s értéket vett fel.

A találatoknak a találatlistán belüli rangsorolt helyéhez egy értéket rendeltek hozzá egy húsz beosztásból álló lineáris érték skála alapján, ahol a legelső hely a legnagyobb értéket kapta, a legutolsó hely pedig a legalacsonyabb értéket vehette fel. Erre az érték hozzárendelésre azért volt szükség, hogy súlyozzák a találatoknak a találatlistán belüli rangsorolt helyét. Az első húsz találatot három csoportba sorolták. Mindegyik csoporton belül a találatokat egyformán súlyozták, még pedig ugyanazt a súlyt vehették fel, amit az első találat kapott az adott csoporton belül a lineáris érték skálán. Ennek értelmében az első három találat húszas súlyt, a következő hét találat tizenhetes súlyt, az utolsó tíz találat pedig tízes súlyt kapott. Ezeket a súlyozott értékeket összegezték és létrehozták a mérőszám számlálóját.

Például, hogyha egy kereső öt releváns oldalt keresett vissza, amelyek az első öt rangsorolt helyre kerültek a találati listában, akkor erre a formula $(3 \cdot 20) + (2 \cdot 17) = 60 + 34 = 94$ -es súlyozott értéket számolt. Ha ez az öt oldal a tizenegyedik és a huszadik rangsorolt helyek közé került a találati listában, akkor erre a formula csak $(5 \cdot 10) = 50$ -es súlyozott értéket kalkulált.

A mérőszám nevezőjét a lekérdezett találatok számából kiindulva hozták létre, amit összesen húsz találatig vettek figyelembe. Amennyiben egy kereső húsz vagy annál több találatot kért le, akkor a húsz találatra eső maximális súlyozott értékeket összegezték, ami: $(3 \cdot 20) + (7 \cdot 17) + (10 \cdot 10) = 279$ -nek felelt meg. Azonban húsznál kevesebb találat esetében a nevező értéke már megváltozott. Ha a kereső egyetlenegy oldalt sem keresett vissza, akkor a nevező 0 értéket vett fel definiálatlan mérőszámmal. A határfeltétel miatt a mérőszám nevezőjét úgy állították elő, hogy összeadták a húsz találatra eső maximális súlyozott értékeket, ami 279-

nek felelt meg, és ebből az összesített értékből kivontak annyiszor tízet amennyi találat hiányzott a húszból.

Például tizenöt visszakeresett találat esetében a nevező értéke $279 - (5 \cdot 10) = 229$ volt. Ha csupán egy találat került lekérdezésre, akkor a nevező értéke $279 - (19 \cdot 10) = 89$ volt. Végül pedig a számláló értékét elosztották a nevező értékével, hogy kiszámolják a tényleges mérőszámot. A Rube Goldberg gépet (angolul: Rube Goldberg Machine) használták függvényként, amely az alábbi formulával írható le:

$$\frac{(1-3 \text{ találat} \cdot 20) + (4-10 \text{ találat} \cdot 17) + (11-20 \text{ találat} \cdot 10)}{279 - \left[(20 - \min(20 - \text{lekérdezett találatok száma})) \cdot 10 \right]}$$

Mindegyik kísérlet mérőszámait értékelték, hogy megállapítsák az ANOVA-ból származó maradék értékek normális eloszlásúak-e. Csupán a negyedik kísérletnél voltak normális eloszlásúak a maradék értékek [Leighton-Srivastava99]. Mivel nem teljesült az ANOVA modell [Tinsley2000] számára szükséges normalitás feltétel, ezért a Friedmann-féle véletlenszerű blokktervezést (angolul: Friedmann's randomized block design) [Raghavarao88] használták a tesztelésekben, ahol a blokkok voltak a keresőkérdések és a kezelési hatások (angolul: treatment effects) pedig a keresők. A Friedmann-féle teszt inkább a populációs mediánokat vizsgálta, nem pedig az átlagértékeket az aszimmetria jelenléte miatt. Mind az öt kísérletben elutasították azt a H_0 -t, hogy a keresőszolgáltatások mediánjai egyenlők lehetnének egymással. A H_0 -ok elutasítása miatt páronkénti többszörös összehasonlításokat (angolul: pairwise multiple comparisons) végeztek az egyes keresőszolgáltatások között.

Mély statisztikai modellezéssel és módszerekkel megállapították, hogy az AltaVista, Excite és az Infoseek keresők több releváns találatot szolgáltatottak a HotBot és a Lycos keresőkhöz viszonyítva. Az első három helyezettnél megfigyelhető volt az, hogy az első húsz találat jóval gyakrabban tartalmazta a keresőkifejezéseket, mint a másik két kereső találathalmaza ([Leighton-Srivastava99] p. 875-877).

IV. Chignell, Gwizdka és Bodner két kísérletet hajtott végre az Excite, Hotbot és az Infoseek keresők lekérdezési teljesítményének a tanulmányozására. Az első kísérletben azt vizsgálták, hogy a napi idő és a keresési stratégia hogyan befolyásolta a keresés feldolgozási idejét mindegyik keresőnél. Kilenc keresőkérdést használtak, amelyek három különböző csoportba voltak sorolhatók: az általános témát leíró keresések, a nagy pontosságra és a nagy teljességre irányuló keresések.

A dokumentumok relevanciáját egy automatizált eljárással állapították meg, amely „consensus peer review” néven volt ismert. Ennek lényege, hogy kiválasztottak hat másik keresőszolgáltatást, amelyeken ugyanazokat a kereséseket futtatták le, mint a vizsgált keresőkön. A hat kereső (AltaVista, Lycos, Northern Light, Search.Com, WebCrawler, Yahoo) találatából bináris relevancia ítéleteket kaptak ugyanarra a keresésre vonatkozóan. Egy vizsgált kereső találatát akkor minősült relevánsnak, hogyha ugyanazt a dokumentumot a hat kereső valamelyike szintén megtalálta az adott keresésre. Ennek a módszernek a használata megkérdőjeleződött a találatok relevancia értékelésénél, mert a keresők találathalmazai között minimális átfedés volt tapasztalható. Mérték szintén a megszakadt és az ismétlődő találatok számát minden egyes kereséskor [Chignell-Gwizdka-Bodner99].

Az adatokat többváltozós variancia analízissel (MANOVA) [Bilodeau-Brenner99] vizsgálták. Találtak egy többváltozós hatást, ami a keresés típusának és a keresőknek a köztük kölcsönhatására irányult. Ezt a többváltozós hatást egy jelentős egyváltozós kölcsönhatás okozta a pontosságra vonatkozóan. A másik három függő változóra gyakorolt egyváltozós kölcsönhatások nem voltak számottevőek. Ezek a függő változók a vizsgált keresőknek feleltek meg.

A keresők lekérdezési idejének egy jelentős hatását vették észre. Az utólag elvégzett ún. Tukey teszt jelentős különbséget jelzett az Excite és az Infoseek keresők, valamint a Hotbot és az Infoseek keresők keresés feldolgozási ideje között.

A kísérletnek ebben a stádiumában a független változóknak az egyes függő változókra gyakorolt hatásait tanulmányozták. A keresők jelentős hatást gyakoroltak a megszakadt és az ismétlődő találatok számára, valamint a pontosságra. A napi idő nagy mértékben kihatott a keresés feldolgozási idejére. A keresés típusa jelentős mértékben befolyásolta a pontossági értékeket. A három keresőszolgáltatás az általános témával kapcsolatos kereséseket hajtotta végre a legeredményesebben, ezt követték a nagy pontosságra és a nagy teljességre irányuló keresések.

A lekérdezési teljesítménynek egy összetett, felhasználó-központú mérőszámát határozták meg a négy függő változó alapján. Az utólag elvégzett ún. Tukey teszt segítségével megkapták a három kereső rangsorát minden egyes függő változóra vonatkozóan. A rangsorolási információk segítségével egy egyszerű formulát hoztak létre: képezték az adott szolgáltatás első, második és harmadik rangsorolt helyeinek súlyozott összegét. Az első helyet 3-as együtthatóval, a második helyet 2-es együtthatóval, a harmadik helyet pedig 1-es együtthatóval vették figyelembe.

Így például az Excite kereső teljesítményére vonatkozó mutatót a következőképpen adták meg:

$$3 \cdot \frac{1}{4} + 2 \cdot \frac{2}{4} + \frac{1}{4} = \frac{2}{3} = 66,7\% .$$

Ennek az összetett mérőszámnak két hiányossága van. Egyrészt nem veszi figyelembe azt az esetet, amikor nincs jelentős statisztikai különbség a két keresőszolgáltatás között. Másrészt az összes függő változót egyenrangú félként kezeli.

A második kísérlet azt elemezte, hogy a földrajzi kiterjedés és a domain nevek hogyan befolyásolták a keresők lekérdezési teljesítményét. Három keresőt (Altavista, HotBot, Infoseek), hat domain nevet és négy keresőkérdést használtak fel a hetvenkét megfigyelés teljes faktoros megtervezésére. A négy keresőkérdést három különböző nyelvre fordították le. Az első húsz találat pontosságát vették figyelembe, amit emberi relevancia ítéletek alapján állapítottak meg. Egy négyes skálát használtak 0-3-ig a találatok relevanciájának meghatározására. Teljes faktoros többváltozós analízist (angolul: Full factorial multivariate analysis) hajtottak végre, ahol a keresők és a domain nevek független faktorként szerepeltek tizennégy mérőszám kíséretében [Chignell-Gwizdka-Bodner99].

V. A keresők tranzakciós naplójának az elemzése rávilágított arra, hogy az átlagos felhasználók maximum kétszavas keresőkérdéseket fogalmaznak meg, és ritkán használnak kereső operátorokat. Ezzel szemben a kereséssel foglalkozó szakértők bonyolultabb kereső operátorokat és összetettebb keresőkérdéseket használnak, mint az átlagos felhasználók [Silverstein-Henzinger-Marais-Moricz99]. *Lucas és Topi* átfogó tanulmányt készített arról, hogy a kereső operátorok és a keresőkifejezések kiválasztása milyen mértékben befolyásolta a keresési találatok relevanciáját. Összesen nyolcvanhét felhasználó vett részt a felmérésben, akik nyolc témában fogalmaztak meg keresőkérdéseket. Ezenkívül a kereséssel foglalkozó szakértők is keresőkérdéseket állítottak össze a megadott témákban, amelyeket ugyanúgy lefuttattak a nyolc vizsgált keresőn, mint a felhasználói kereséseket.

A weboldalak relevanciájának megítélésében egy tízes küszöbértéket állítottak fel, mivel a találati lista elején lévő első tíz találatot a felhasználók biztos, hogy megnézik. A visszakeresett találatok relevanciáját egy négyes skála alapján határozták meg. A kutatási modellben lévő változókat teljes többszörös regresszióval (angolul: full multiple regression) és lépcsős hierarchikus többszörös regresszióval (angolul: step-wise hierarchical multiple regression) vizsgálták meg [Lucas-Topi2002].

VI. *Radev, Libner és Fan* azt tanulmányozták, hogy a legnépszerűbb keresők mennyire pontosan válaszolták meg a természetes nyelven megfogalmazott kereséseket. Összesen hétszáz keresőkérdést futtattak le kilenc különböző keresőszolgáltatáson. A keresésekre kapott első

negyven találatot letöltötték és eltárolták. Megállapították, hogy mindegyik kereső legalább egy helyes választ nyújtott a feltett faktografikus keresések háromnegyed részére.

A kísérletben három hipotézist teszteltek, ezek pedig a következők voltak:

1. A keresőszolgáltatások hatékonyak a faktografikus keresések megválaszolásában.
2. A keresőkérdésnek bizonyos jellemzői előre jelzik a helyes válasz megtalálásának a valószínűségét a keresőknél.
3. Az egyedi sajátosságokkal rendelkező keresőkérdések nagyobb valószínűséggel helyes választ eredményeznek egyes speciális keresőknél.

A második és a harmadik hipotézis ellenőrzéséhez a hétszáz keresőkérdést négy faktor alapján elemezték, amelyek a következők voltak:

1. szükséges válasz típusa;
2. tulajdonnév jelenléte a keresésben;
3. keresőkérdés időbeli függősége;
4. keresőkérdésben előforduló szavak száma.

A három hipotézis tesztelésekor variancia analízist (ANOVA) használtak, egyrészt a kilenc keresőszolgáltatás általános értékeinek összehasonlítására, másrészt a fenti négy faktor fontosságát próbálták megállapítani ezeknek az értékeknek az előrejelzésében. Ezenkívül ezzel a statisztikai módszerrel mérték a keresők eltérő teljesítményét mindegyik faktorra vonatkozóan. A mérés eredményei rámutattak arra, hogy az összes vizsgált kereső helyes választ nyújtott az első negyven találatban a rendelkezésére álló idő több mint 75%-ában. Ebből adódóan az internetes keresőket eredményesen használták egy olyan rendszer részeként, amely alkalmas volt a faktografikus keresések megválaszolására [Radev-Libner-Fan2002].

VII. Bar-Ilan egy olyan mérés tapasztalatairól számolt be, ahol egyetlenegy keresőkérdést futtattak le hat keresőn (AltaVista, Excite, Fast, HotBot, Google, Northern Light) harminchárom alkalommal egy éven keresztül. Az értékelők egy jól körülhatárolt, stabil témát választottak visszakeresésre, ami a „patkányfarok kaktusz” volt. Úgy fogalmazták meg a keresőkérdést, hogy azt mindegyik kereső azonosan értelmezze. Ezáltal elkerülték a morfológiai elemzéssel és a többszavas keresőkérdések kezelésével kapcsolatos problémákat. A mérés elsődleges célja az volt, hogy megvizsgálják a keresők működését hosszabb időtartamon keresztül. A keresők találatsszolgáltatásának a stabilitását számos mérték irányából közelítették meg.

A vizsgálat során a következő módszerrel ellenőrizték, hogy a keresők helyesen távolították-e el az URL-eket a találatalmazukból. Mindegyik keresőt speciálisan kezelték a kereséseknél és a

kapott találatokat letöltötték. A találatokról leválasztották az URL-eket egy Visual Basic program segítségével. Excelben lefuttattak egy Visual Basic makró scriptet, ezáltal létrehoztak egy egyedi URL listát minden egyes keresésre vonatkozóan. A hivatkozásokat lekérdezték és a találatok szövegét helyben tárolták. Az aktuális keresés és az azt megelőző keresés eredményeit összehasonlították egymással. Ezután megkeresték a hiányzó URL-eket az aktuális találathalmazban [Bar-Ilan2002a].

Úgy vélem, hogy ennek a mérésnek legértékesebb részét képezi az alkalmazható mérési módszerek (mértékek) részletes bemutatása. A továbbiakban ismertetem ezeket a mértékeket a pontos meghatározásaikkal együtt.

2.4.1. Egy tesztelés során bevezetett új mértékek

A keresőszolgáltatások méréséhez a hagyományos értékelési paraméterek nem elegendőek. *Cleverdon* a következő hagyományos mértékeket javasolta használatra: gyűjtemény lefedettség, teljesség, pontosság, válaszidő, felhasználói erőfeszítés, találatmegjelenítési formátum. Ezeknek az értékelési paramétereknek az egyedüli hátránya, hogy azok nem veszik figyelembe a dokumentumhalmazban bekövetkező folyamatos változásokat, ezért új mértékeket kell meghatározni a keresők lekérdezési teljesítményének a mérésére egy dinamikusan változó környezetben ([Bar-Ilan2002a] p. 309).

1. Technikai pontosság keresőkérdésenként és a tanulmányozott időtartam alatt

A relevancia megítélése helyett egy objektív mértéket használtak, amely a megadott keresőszavakra keresett a dokumentumokban. Egy dokumentum akkor minősül technikailag relevánsnak, hogyha az kielégíti a keresés feltételeit. Tehát azok a keresőszavak és kifejezések, amelyekről feltételezzük, hogy jelen vannak egy dokumentumban, azok ténylegesen előfordulnak benne. Továbbá azok a keresőszavak és kifejezések, amelyekről feltételezzük, hogy hiányoznak egy dokumentumból, azok nincsenek jelen benne. Egy URL-t, akkor tekintünk technikailag relevánsnak, amennyiben az egy technikailag releváns dokumentumot tartalmaz.

A technikai pontosságot úgy határozhatjuk meg, hogy a lekérdezett dokumentumok összességéből kiszámítjuk a technikailag releváns dokumentumok százalékos megoszlását. A technikai relevancia előnye, hogy az gyorsan kiszámolható egy egyszerű minta-megfeleltető algoritmussal, továbbá az nagy dokumentumhalmazokra is használható. Ezenkívül az egyaránt alkalmas a “rossz dokumentumok”, azaz a nem létező URL-ek és a keresőkérdést ki nem elégítő dokumentumok kiszűrésére. Ennek a mutatónak az egyedüli hátránya, hogy az nem állapítja meg

a dokumentum fontosságát és annak megbízhatóságát. A fenti meghatározás egy URL technikai relevanciájára vonatkozik egy adott kereséskor.

Ez a definíció sokféleképpen kiterjeszhető az általunk vizsgált időtartamra: egy URL-t, akkor tekintünk technikailag relevánsnak az általunk vizsgált időtartamban, hogyha az annak minősült az első kereséskor is. Másik lehetséges megoldás számunkra az, hogyha összegyűjtjük a technikailag releváns URL-eket minden egyes kereséskor, illetve azokat az URL-eket, amelyek legalább egy kereséskor minősültek technikailag relevánsnak. Bármelyik megoldás választható számunkra.

2. Relatív gyűjtemény lefedettség keresőkérdésenként és a vizsgált időtartam alatt

Egy kereső relatív gyűjtemény lefedettségét keresőkérdésenként úgy számolhatjuk ki, hogy a vizsgált kereső technikailag releváns URL-jeinek a számát elosztjuk az összes vizsgált kereső technikailag releváns URL-jeinek a számával ugyanarra a keresőkérdésre.

Egy kereső teljes relatív gyűjtemény lefedettségét úgy kaphatjuk meg, hogy a vizsgált kereső technikailag releváns URL-jeinek a számát elosztjuk az összes vizsgált kereső technikailag releváns URL-jeinek a számával az összes keresésre.

Ez a két mérték durván becsli egy keresőszolgáltatás gyűjtemény lefedettségét, azaz egy kereső mennyire képes releváns találatokat szolgáltatni a megadott keresőkérdésre. A relatív gyűjtemény lefedettség kifejezi egy kereső gyűjtemény lefedettségét keresőkérdésenként egy adott időpontban. Ezzel szemben a teljes relatív gyűjtemény lefedettség tükrözi egy kereső relatív gyűjtemény lefedettségét egy adott időtartam alatt. Ez a két mutató jelentős mértékben függ a vizsgált keresők csoportjától.

3. Új és teljesen új URL-ek

Egy URL, akkor minősül újnak egy keresőszolgáltatás számára az adott keresőkérdésre, hogyha azt a kereső egyetlenegy korábbi keresés alkalmával sem kérdezte le. Ezt a mértéket csak a második kereséstől kezdődően lehet kiszámolni, azaz csak az utolsó $n-1$ keresésben kapott URL-eket kell figyelembe vennünk. Tehát az első keresés mindig alapul szolgál, amihez viszonyítjuk a következő keresést. Az új URL-ek mutatója azt jelzi számunkra, hogy a kereső milyen mértékben bővíti az adatbázisát a témával kapcsolatban.

$$\text{új URL-ek}_{q,i} = |\{ \text{az } i. \text{ keresési alkalom során kapott technikailag releváns URL-ek} \}| - |\{ \text{a } j. \text{ keresési alkalom során kapott technikailag releváns URL-ek, ahol } j < i \}|$$

Ebben az egyenletben és a továbbiakban $\{.\}$ a $\{.\}$ halmaz elemeinek számát jelenti.

Az új URL-ek halmaza további két csoportra osztható: a teljesen új URL-ekre és az újonnan felfedezett URL-ekre. Egy URL, akkor minősül teljesen újnak egy kereső számára az adott kereséskor, hogyha azt korábban egyetlen kereső sem kérdezte le. A teljesen új URL-ek halmazát olyan URL-ek halmazának is tekinthetjük, amelyeket először találtak meg egy adott kereséskor.

teljesen új URL-ek $_{q,i,s} = \{ \text{az } i. \text{ keresési alkalom során az } s \text{ kereső által visszakeresett technikailag releváns URL-ek} \} - \{ \text{az } i. \text{ keresési alkalom előtt lekérdezett URL-ek} \}$

Újonnan felfedezett URL-eknek minősülnek azok a korábban létező és régóta releváns weboldalak, amelyeket a kereső csak mostanában fedezett fel és adott hozzá a saját adatbázisához.

újonnan felfedezett új URL-ek $_{q,i,s} = \text{új URL-ek}_{q,i} - \text{teljesen új URL-ek}_{q,i,s}$

4. Elfelejtett, újra megtalált, elveszített, jól kezelt és rosszul kezelt URL-ek

A keresők sokszor tévesen távolítják el az adatbázisukból azokat az URL-eket, amelyek továbbra is léteznek a weben és technikailag relevánsak. Az ilyen URL-eket *elfelejtett* URL-eknek nevezzük. Az elfelejtett URL-ek meghatározhatók keresőnként és keresőkérdésenként egyaránt. Azonban azok definiálhatók a keresőtől függetlenül is keresőkérdésenként. Az elfelejtett URL-eket a következőképpen határozták meg $i > 1$ -re:

elfelejtett URL-ek $_{q,i} = \{ i-1. \text{ keresési alkalom során kapott technikailag releváns URL-ek, amelyek továbbra is léteznek a weben és technikailag relevánsak, azonban az } i. \text{ keresési alkalom során mégsem kerülnek lekérdezésre} \}$

Az elfelejtett URL-ek újra felbukkanhatnak egy későbbi kereséskor. Egy URL csakis a legelső alkalommal számít újra megtaláltként, miután az tévesen kimaradt a kereső adatbázisából a korábbi keresésekkor. Ezeket az URL-eket újra megtalált URL-eknek nevezzük. Az újra megtalált URL-ek meghatározhatók keresőnként és keresőkérdésenként egyaránt. Azonban azok definiálhatók a keresőtől függetlenül is keresőkérdésenként. Az újra megtalált URL-eket a következő módon határozták meg $i > 1$ -re:

újra megtalált URL-ek $_{q,i} = \{ \text{a } j. \text{ keresési alkalom során kapott technikailag releváns URL-ek, amelyek tévesen maradtak ki a kereső adatbázisából az } i. \text{ keresési alkalom során, ahol } i < j \text{ és nem kerültek lekérdezésre a } j-1. \text{ keresési alkalom során} \}$

Elveszített URL-eknek nevezzük azokat az URL-eket, amelyek tévesen maradtak ki a kereső adatbázisából. Továbbá azok tartalmilag nem helyettesíthetők semmilyen más URL-el az aktuális kereséskor. Tehát ezen URL-ek esetében az információ teljesen elvész a felhasználó

számára az adott kereséskor. Az elveszített URL-ek meghatározhatók keresőnként és keresőkérdésenként egyaránt.

elveszített URL-ek_{q,i} = $\{ \text{azok a kereső adatbázisából tévesen eltávolított URL-ek, amelyeknek nincs más tartalmi megfeleltetésük a } q \text{ keresőkérésre az } i. \text{ keresési alkalom során} \}$

A jól kezelt URL-ek halmazába tartoznak azok az URL-ek, amelyeket a kereső folyamatosan megtalál az első keresés óta, továbbá azok az URL-ek, amelyeket a kereső nem őriz meg a későbbiekben, mert azok nem találhatók meg a weben, illetve technikailag nem relevánsak többé.

jól kezelt URL-ek_q = $\{ \text{azok a } q \text{ keresőkérésre kapott technikailag releváns URL-ek, amelyeket tévesen sohasem távolítottak el a kereső adatbázisából a vizsgált időtartam alatt} \}$

A rosszul kezelt URL-ek halmazába tartoznak azok az URL-ek, amelyek legalább egyszer tévesen maradtak ki a kereső adatbázisából a vizsgált időtartam alatt. A rosszul kezelt URL-ek halmaza két további csoportra osztható: a rosszul kezelt, újra megjelenő URL-ek körére – ez megegyezik az újra megtalált URL-ek halmazával, de annál bővebb halmazról van szó – és a rosszul kezelt, elveszített URL-ek körére. A rosszul kezelt URL-ek esetében csak az $n-1$. kereséskor talált URL-eket vesszük figyelembe, hiszen egy URL-t nem tekinthetünk rosszul kezeltnek, hogyha azt az utolsó kereséskor fedeztük fel először.

rosszul kezelt URL-ek_q = $\{ \text{azok a kereső adatbázisából tévesen eltávolított URL-ek, amelyeknek nincs más tartalmi megfeleltetésük a } q \text{ keresőkérésre valamely } i. \text{ keresési alkalom során, ahol } i > 1 \}$

5. Az adatbázis változékonysága (angolul: self-overlap=önátfedés)

Ezt a mértéket úgy számolhatjuk ki, hogy az egy adott keresésre kapott technikailag releváns URL-ek számát elosztjuk a kereső által megtalált összes technikailag releváns URL számával a vizsgált időtartam alatt. Ez a mutató a kereső adatbázisának a változását tükrözi egy bizonyos időtartam alatt, hogyha annak magas a számértéke az összes keresésre, akkor ez azt jelzi számunkra, hogy a kereső találatai stabilak. Azonban hogyha annak túl magas a számértéke, akkor a találatok stabilitása mellett ez azt is jelenti számunkra, hogy a kereső adatbázisának a tartalma elavulttá válik. Ennek a mértéknek az optimális értékei se nem túl magasak, se nem túl alacsonyak.

adatbázis változékonysága_{q,i} = $\{ \text{az } i. \text{ keresési alkalom során visszakeresett technikailag releváns URL-ek} \} / \{ \text{az adott } q \text{ keresőkérésre kapott összes technikailag releváns URL a vizsgált időtartam alatt} \}$

Megjegyzem, hogy a nevezőben lévő érték egy virtuális halmazt jelöl, mert az tartalmazhat olyan URL-eket is, amelyek sohasem léteztek egy időben a weben.

6. Azon keresések száma, amellyel az adott URL-t lekérdezték

Ez a mérték a kereső adatbázisának a változékonyságához kapcsolódik, amely azt jelzi, hogy egy adott URL-t pontosan hány kereséssel kérdezték le. A mérték meghatározható keresőnként, valamint a vizsgált keresők teljes halmazára is.

7. Állandósult URL-ek

Állandósult URL-eknek tekintjük azokat a technikailag releváns URL-eket, amelyeket az összes keresés során is technikailag releváns URL-ekként kapunk meg.

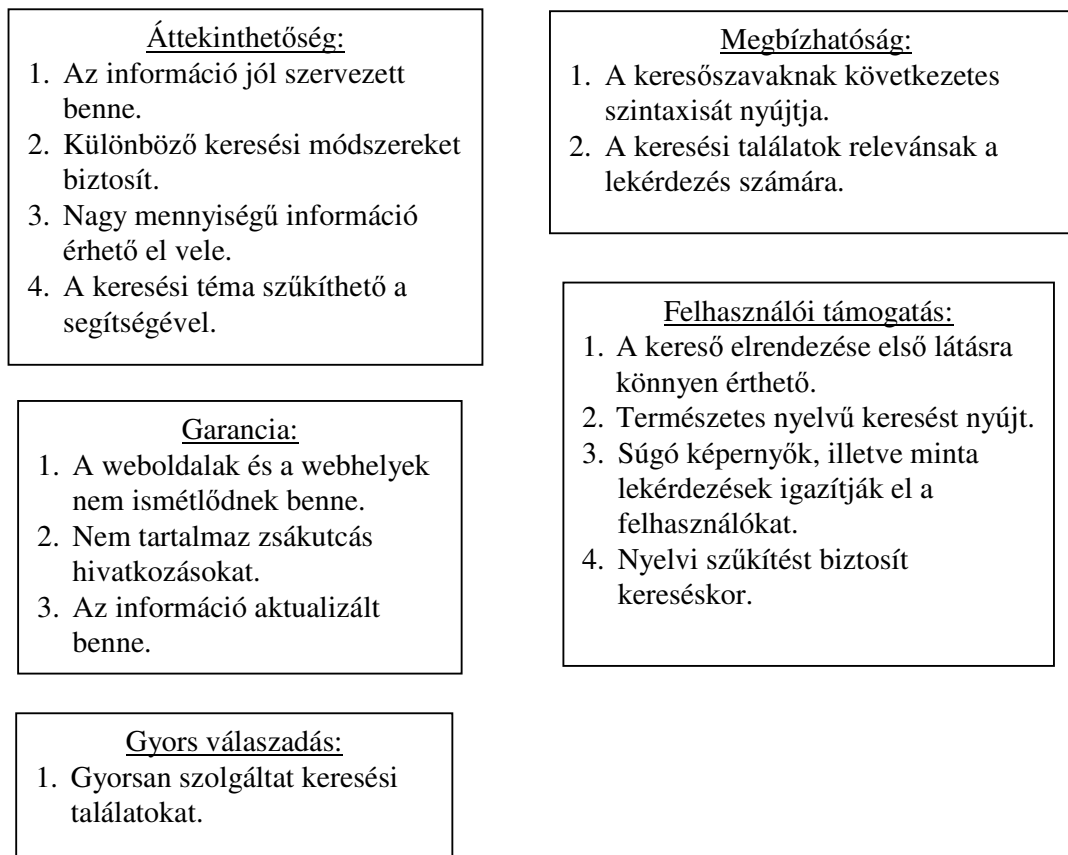
A fenti mértékek többsége kiszámolható az állandósult URL-ek halmazára is. Ezen URL-ek halmaza lehetővé teszi, hogy tanulmányozzuk a keresők változásait egy bizonyos időtartam alatt, még pedig úgy, hogy azok változásait elkülönítjük a világháló dinamikus változásaitól. Keresési témánktól függően ez a halmaz túlságosan kicsi lehet az elemzés számára ([Bar-Ilan2002a] p. 310-312), [Bar-Ilan2002b].

Az előzőekben tárgyalt mértékek tükrözik a keresőszolgáltatások stabilitását, illetve azok következtelenségét találatok kezelésében. Nagyon fontos, hogy olyan követelményeket határozzunk meg, amelyek mérik a lekérdezett találatok számában történő változásokat. Próbáljuk megérteni, hogy a találatok számában jelentkező ingadozások ténylegesen tükrözik-e a világhálón végbemenő változásokat. Továbbá ezekre a találatszám-beli ingadozásokra hatnak-e az adatbázis frissítések, a pillanatnyi üzemszünetek, valamint a keresőszolgáltatás üzleti modelljében, irányelveiben bekövetkező változások. Ezeknek a mutatóknak a megfelelő használata kiküszöböli a méréseknél felmerülő szubjektivitást [Bar-Ilan2002a].

2.4.2. A SERVQUAL minőségbiztosítási modell használata a keresők értékelésére

Ebben a részben egy olyan vizsgálat eredményeit foglalom össze, amely a minőségbiztosítás irányából közelíti meg és válogatja ki a keresők értékelésére irányuló felhasználói elvárásokat. Később ezek a felhasználói elvárások fontos változókként jelennek meg egy-egy mérés során. A vizsgálat érdeme, hogy felhasználói észrevételekre támaszkodik a változók csoportosításában és azoknak a megfelelő faktorhoz történő sorolásában. *Xie, Wang és Goh* egy minőségbiztosítási modellt adaptált az internetes keresők minőségének a vizsgálatára. A SERVQUAL modellt *Parasuraman, Zeithaml és Berry* fejlesztette ki azzal a céllal, hogy mérni lehessen egy szolgáltatás minőségét [Parasuraman-Zeithaml-Berry91]. Kezdetben egy szolgáltatás

minőségének a jellemzését tíz dimenzióban javasolták, később az empirikus kutatások ezt lecsökkentették öt dimenzióra. Az öt dimenzió összesen tizennégy változót használ az internetes keresők minőségének a mérésére, amelyek besorolását az 1. ábra szemlélteti ([Wang-Xie-Goh99] p. 500).



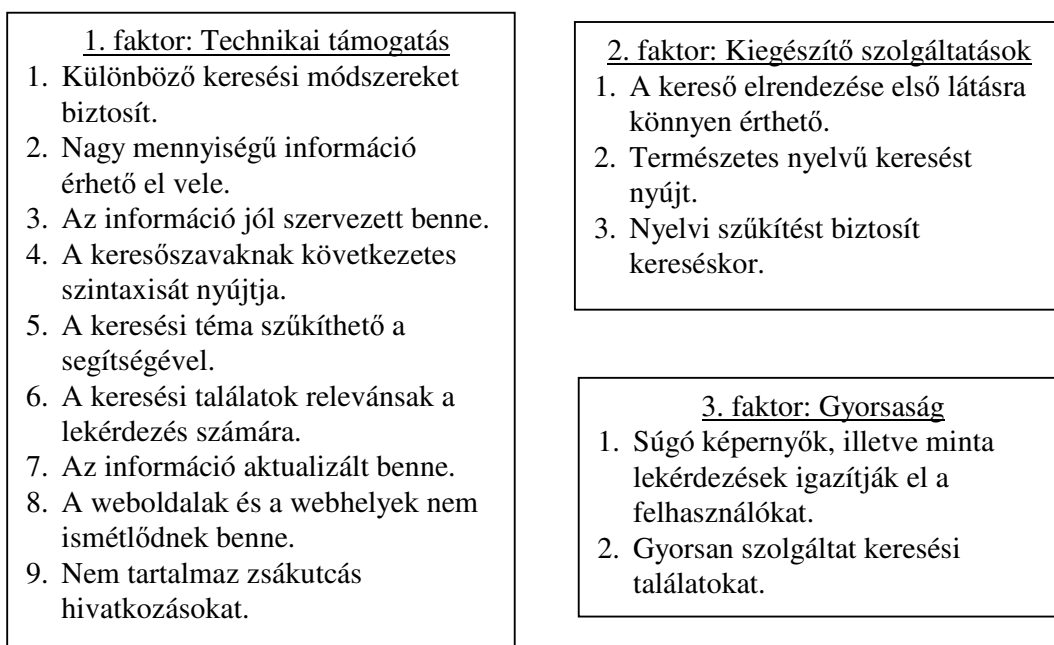
1. ábra. Az internetes keresőkkel szemben támasztott felhasználói elvárások tizennégy változója
Forrás: ([Wang-Xie-Goh99] p. 500)

Egy kérdőíves felmérést végeztek a felhasználók körében, amely két részből állt. Az első rész általános információkat közölt a felhasználókról és a kedvelt keresőszolgáltatásaikról. A második rész az internetes keresőkkel kapcsolatos felhasználói elvárásokat és a tényleges, érzékelt állapotokat mérte.

Általában egy szolgáltatás alacsony, illetve magas minőségének megítélése attól függ, hogy a felhasználók hogyan érzékelik egy szolgáltatás aktuális teljesítményét a saját elvárásaik függvényében. Ezért a felhasználói elvárásokat és a felhasználók által érzékelt tényleges állapotokat különbség értékekkel (angolul: gap) hasonlították össze egymással. Ezt a módszert különbség elemzésnek (angolul: gap analysis) hívjuk [Parasuraman-Zeithaml-Berry90]. A pozitív különbség érték azt jelzi számunkra, hogy a szolgáltatás minősége kielégíti, vagy esetleg túllépi a felhasználói elvárásokat. A negatív különbség érték pedig azt

jelenti, hogy a szolgáltatás minősége nem tesz eleget a felhasználói elvárásoknak. A SERVQUAL különbség értékek kicsik voltak, $-1,44$ és $0,66$ között mozgott a nagyságuk. Az elvárási szintre és az érzékelési szintre vonatkozó megbízhatósági becslésre hasonló megfontolásokat és eredményeket találhatunk [Cronbach51]-ben.

Az eredeti öt dimenziót mások sem alkalmazták egységesen a megjelent tanulmányokban. Ezért faktoranalízist használtak annak érdekében, hogy megállapítsák hány faktor, illetve főbb dimenzió fejezi ki a legjobban a változók kezdeti halmazai között megfigyelt korrelációkat [Gorsuch83]; [Rummel70]. „Varimax” forgatásos módszerrel választották el a faktorokat egymástól és maximalizálták a betöltéshez szükséges szórásnégyzet összegeket [Hair-Anderson-Tatham-Black98]. Az öt dimenzió helyett összesen hármat azonosítottak be, amelyek már faktoroknak minősültek. A három faktorhoz kapcsolódó tizennégy változót másképpen rendezték el, mint a korábban kifejlesztett ötdimenziós eszköz esetében. Ez a változás nyomon követhető a 2. ábrán [Wang-Xie-Goh99].



2. ábra. A faktoranalízis után kapott tizennégy változó újabb elrendezése

A SERVQUAL modellben található áttekinthetőséget, megbízhatóságot és garanciát a felhasználók nem tekintették különálló faktornak. Forgatás után a betöltéshez szükséges szórásnégyzet összegek a technikai támogatás esetében 5,362-nek feleltek meg, ami a teljes variancia 38,3%-át tette ki. Emiatt tekintjük a technikai támogatást a legnagyobb és egyben a legfontosabb faktornak a három faktor közül. Három változót a felhasználók a kiegészítő szolgáltatások faktorhoz soroltak a korábbi felhasználói támogatás faktor helyett. Ez az új faktor

a teljes variancia 12,9%-ának felelt meg. A gyorsaság faktor csak két változót tartalmazott és a teljes variancia 11,4%-át nyújtotta.

Megfelelési (correspondence) analízist (angolul: correspondence analysis) [Greenacre84] használtak a felhasználók foglalkozása és az általuk keresett információ közötti kapcsolat vizsgálatára, továbbá a felhasználók oktatási háttere és az általuk keresett információ közötti kapcsolat vizsgálatára. Ennek az elemzésnek az eredményei a következők voltak: a technikai személyzet és a diákok inkább a tudományos témákra kerestek. Ezzel szemben az egyházi, kereskedelmi, szolgáltató személyzet, valamint a felsőfokú végzettségűek az üzleti élettel és a munkával kapcsolatos információkra kerestek főként. Az eltérő oktatási háttérrel rendelkező felhasználók azonosan koncentráltak mindegyik témára. A diplomások és a posztgraduális képzésben résztvevők a tudományos és a szórakoztató témákra kerestek az Interneten ([Wang-Xie-Goh99] p. 502-506); ([Xie-Wang-Goh98]).

Az internetes keresők értékelésével foglalkozó kutatás nem egységes az alkalmazott módszerek és megközelítések tekintetében, ezért egyre növekvő igény jelentkezik egy megfelelő szabvány kidolgozására. Tehát a mutatóknak egy általánosan használható szabványát kellene kidolgozni a keresők teljesítményének a mérésére, amely révén a kutatók összehasonlíthatják a keresők lekérdezési hatékonyságát, és hosszabb időtartamon keresztül tanulmányozhatják azokat. Mindezt pedig olyan szintű következetességgel valósíthatják meg, amire eddig még nem volt példa. Ez a fejezet áttekintést nyújt számunkra a korábbi vizsgálatokban használt statisztikai módszerekről, ezáltal pedig útmutatóként szolgálhat egy mérés megtervezéséhez. A vizsgálat elején többek között el kell döntenünk azt is, hogy melyik statisztikai módszer felelne meg leginkább a kutatási célkitűzéseinknek, ezután pedig ki kell alakítanunk egy statisztikai modellt a megfelelő változókkal.

Irodalom:

[Bar-Ilan2002a] 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.

[Bar-Ilan2002b] 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, 7 May 2002.

<http://www.dcs.bbk.ac.uk/webDyn2/onlineProceedings.html> 2007.03.02.

[BernersLee-Fielding-Frystyk96, RFC1945] Berners-Lee, T.–Fielding, R.–Frystyk, H.: Hypertext Transfer Protocol - HTTP/1.0. May 1996.

<http://www.rfc-editor.org> 2007.08.10.

[Bilodeau-Brenner99] Bilodeau, M.–Brenner, D.: Theory of multivariate statistics. Springer, 1999.

[Brin-Page98] 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.

[Chignell-Gwizdka-Bodner99] 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.

[Clarke-Willett97] Clarke, S. J.–Willett, P.: Estimating the recall performance of web search engines = Aslib Proceedings, vol. 49. no. 7. (1997), p. 184-189.

[Cronbach51] Cronbach, L. J.: Coefficient alpha and the internal structure of tests = Psychometrika, vol. 16. (1951), p. 257-334.

[Ding-Marchionini96] 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.

[Froehlich2001] Froehlich, T. J.: Case study 5.1: Developing search engine evaluation criteria. In.: Library evaluation. Libraries Unlimited, 2001. p. 185-200.

[Gordon-Pathak99] 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.

[Gorsuch83] Gorsuch, R. L.: Factor Analysis. Hillsdale: Lawrence Erlbaum Associates, 1983.

[Greenacre84] Greenacre, M. J.: Theory and applications of correspondence analysis. London: Academic Press, 1984.

[Hair-Anderson-Tatham-Black98] Hair, J. F.–Anderson, R. E.–Tatham, R. L.–Black, W. C.: Multivariate Data Analysis. Prentice-Hall, 1998.

[Harter96] 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.

[Harter97] 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.

- [Horváth-Sütheő2003] Horváth T.–Sütheő P.: A tartalmi feltárás. In.: Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés. Szerk. Horváth T.–Papp I. Bp, 2001. Osiris
- [Landoni-Bell2000] Landoni, M.–Bell, S.: Information retrieval techniques for evaluating search engines: A critical overview = *Aslib Proceedings*, vol. 52. no. 3. (2000), p. 124-129.
- [Leighton95] 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.
- [Leighton-Srivastava97] 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.
- [Leighton-Srivastava99] 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.
- [Lucas-Topi2002] 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.
- [Oppenheim-Morris-McKnight-Lowley2000] 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.
- [Parasuraman-Zeithaml-Berry90] Parasuraman, A.–Zeithaml, V. A.–Berry, L. L.: *Delivering quality service: Balancing customer perceptions and expectations*. New York: Free Press, 1990.
- [Parasuraman-Zeithaml-Berry91] 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.
- [Radev-Libner-Fan2002] 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.
- [Raghavarao88] Raghavarao, D.: *Constructions and combinatorial problems in design of experiments*. Dover Publications, 1988.
- [Rummel70] Rummel, R. J.: *Applied Factor Analysis*. Evanston: Northwestern University Press, 1970.
- [Schwartz98] Schwartz, C.: Web search engines = *Journal of the American Society for Information Science*, vol. 49. no. 11. (1998), p. 973-982.
- [Siegal-Castellan88] Siegal, S.–Castellan, N. J.: *Nonparametric statistics for the Behavioral Sciences*. Singapore: McGraw-Hill, 1988.
- [Silverstein-Henzinger-Marais-Moricz99] 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.
- [Spink-Greisdorf2001] 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.

[Sroka2000] 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.

[Su97] 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

[Tinsley2000] Tinsley, H. E. A.: *Handbook of applied multivariate statistics and mathematical modeling*. Academic Press, 2000.

[Tóth2004] 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.

[Ungváry2001] 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.

[Wang-Xie-Goh99] 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.

[Xie-Wang-Goh98] 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.

3. Az internetes keresők nyelvi lehetőségeinek vizsgálata

3.1. A tesztelés célja és motivációja

Dolgozatomban egy olyan tesztelés eredményeiről szeretnék beszámolni, amely *Bar-Ilan* és *Gutman* 2002 novemberében végzett vizsgálatához kapcsolódik, akik elsődlegesen morfológiai szempontból elemezték az orosz, francia, magyar és a héber nyelvű keresések értelmezését az angol nyelvű keresőknél és a helyi fejlesztésű keresőknél. Kutatásuk során mindketten arra koncentráltak, hogy a keresők mennyire képesek figyelembe venni a kiválasztott nyelvek egyedi sajátosságait és milyen hatékonyan válaszolnak a nem angol nyelvű keresésekre. Eredményeikben a keresők találatainak a számát közölték a lefuttatott keresőkérdésekkel együtt [Bar-Ilan-Gutman2005].

Ezen szakirodalmi előzmény alapján végeztem egy tesztelést 2005 júliusa és szeptembere között. Vizsgálatom célja az volt, hogy feltérképezzem az angol és a magyar nyelvű keresők nyelvi lehetőségeit a különböző megadott keresőkérdésekre. Az angol nyelvű keresőkön angol és magyar nyelvű keresőkérdéseket egyaránt lefuttattam, a magyar nyelvű keresőkön azonban csak magyar nyelvű kereséseket végeztem, mert azok főként magyar nyelvű weboldalakat indexelnek, kivéve az Origo-Vizsla és az Altavizsla szolgáltatásokat. Arra a kérdésre kerestem a választ, hogy ezek a keresőszolgáltatások tettek-e bármilyen erőfeszítést a keresőkérdések pontosabb nyelvi értelmezése érdekében, vagy továbbra is megoldásra váró feladatok maradnak a számukra. 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 tesztelés eredményei teljes mértékben alátámasztották az említett szerzők korábbi megállapítását, mely szerint: a magyar nyelvű keresők csak bizonyos mértékig veszik figyelembe a magyar nyelv egyedi sajátosságait. Ezzel szemben az angol nyelvű keresők komoly hiányosságokat mutatnak ezen a téren, hiszen főként pontos szóalakra és kifejezésre keresnek, és egyszerű minta megfeleltetést végeznek [Bar-Ilan-Gutman2005]. Egy érdekes magyar kutatási projektre szeretném felhívni a figyelmet, amelynek keretében morfológiai elemző programokat hoztak létre kisebb uráli nyelvekre, valamint elektronikus korpuszokat készítettek ezekre a nyelvekre vonatkozóan [Prószéky-Novák2005].

A magyar nyelvben kiemelt fontossága van annak, hogy a keresők megfelelő nyelvi elemzést végezzenek a keresésekre, különben azok releváns információkat veszíthetnek el a témával kapcsolatban. A keresők sokszor nem ismerik fel az összetartozó szóalakokat és nem képesek azokat együtt kezelni. Egy keresőkérdés toldalékolt alakjainak a megtalálásánál a leggyakoribb probléma az, amikor egy toldalék megváltoztatja a szó eredeti alakját. Erre példaként említhetők: a *bagoly-baglyok*, *kő-kövek*, *sátor-sátrak* kifejezések a magyar nyelvben. A keresők automatikus csonkolási algoritmusaitól elvárható lenne, hogy azok megtalálják egy keresőkérdés ragozott szóalakjait. Azonban ez nehezen teljesíthető követelmény. Amennyiben a keresők nem képesek automatikusan felkutatni ezeket a ragozott szóalakokat, akkor a felhasználók releváns weboldalakot hagyhatnak figyelmen kívül a visszakeresett témával kapcsolatban. Később látni fogjuk ezt a problémát a „májmetely” keresőkérdés példáján keresztül is. A magyar nyelvben az ékezetek és a stopszavak helyes kezelése nagyon fontos felhasználói elvárás. Ha egy kereső nem tesz eleget ezeknek a követelményeknek, akkor az sok zajos találatot szolgáltat a magyar felhasználók számára. Egy keresőkérdés szinonimáinak a sikeres felkutatása nagymértékben elősegíti számunkra, hogy még több releváns találatot kapjunk egy lefuttatott keresésre. A keresők úgy képesek eleget tenni ennek az elvárásnak, hogyha egy tezaurusz használatára támaszkodnak [Bar-Ilan-Gutman2005]; ([Tóth2006a] p. 39).

3.2. A tesztelés módszere

A tesztelés mindkét nyelv esetében a megfelelő szavak, kifejezések felkutatásával kezdődött. Körültekintően választottam ki a tesztelésre szánt keresőkérdések halmazát. Ezután próbakereséseket végeztem 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. Természetesen a tulajdonneveket figyelmen kívül hagytam, hiszen azok nem fordulnak elő más morfológiai alakban az angol nyelvben. 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 keresőkifejezésként egy szűkebb találathalmazt eredményező szakkifejezést adtam meg a „májmetely”, ami egy parazitát, illetve férget jelöl. Megjegyzem, hogy erre a keresőkifejezésre egy elektronikus növény- és állatrendszertani tezauruszban találtam

rá [Taxaurusz2000]. Csonkolásnál rögzítettem az egyes keresők által lekérdezett találatok számát. Elsődlegesen a találatok leírását vettem alapul, de ahol indokolt volt azok tartalmába is belenéztem. A kereséseket 2005 júliusa és szeptembere között hajtottam végre.

3.3. A tesztelésben részt vevő keresők

A tesztelésbe három angol nyelvű keresőt vontam be: a Google-t [Google], az AltaVista-t [AltaVista] és az AlltheWeb-et [AlltheWeb]. A választásom azért esett erre a három angol nyelvű keresőre, mert azok lehetővé teszik a magyar nyelvű weboldalakra történő keresést. Ezenkívül a Google kereső magyar nyelvű lekérdezési felületet biztosít a felhasználók számára. Öt jelentősebb magyar nyelvű keresőt teszteltem, amelyek a következők voltak: Heuréka [Heuréka], Origo-Vizsla [Origo-Vizsla], Kurzor [Kurzor], Góliát [Góliát] és az Altavizsla. Ezek közül egyedül az Altavizsla kereső esetében nem találtam semmiféle tájékoztató segédletet. A keresők kiválasztásánál fontos kritérium volt számomra, hogy azok rövid válaszidő után szolgáltatassanak megfelelő mennyiségű találatot a keresésekre.

Az egyértelműség kedvéért fontos, hogy meghatározzam a *keresőkulcs*, illetve a *keresőkulcs-érték* fogalmát, mivel azok gyakran fordulnak elő később a dolgozatban. Keresőkulcs vagy keresőkulcs érték alatt azokat a felhasználók által megadott releváns kulcsszavakat értem, amelyekre a kereső a weboldalak visszakeresését végzi. A továbbiakban a keresőkulcsokat, illetve a keresőkulcs értékeket relációs jelek között dőlt betűvel jelölöm.

3.4. A kiválasztott vizsgálati szempontok ismertetése

A teszteléshez *Bar-Ilan* és *Gutman* közös tanulmányában [Bar-Ilan-Gutman2005] található szempontokat vettem alapul, amelyek az alábbiak voltak: morfológiai elemzés (stemming), stopszavak és ékezetek kezelése, csonkolás, szinonimákra történő keresés.

- Morfológiai elemzés (stemming) alatt azt értem, amikor a kereső megtalálja egy keresőkérdés toldalékolt alakjait. Morfológiai elemzésnél vizsgáltam tehát, hogy az adott kereső megtalálja-e egy keresőkulcs többes számú alakját vagy sem (vagyis keres-e a keresőkérdés többes számú alakjára). Keresőkulcsaim a következők voltak: <dog-dogs>, <ház-házak>, <kocsi-kocsik>, <kutya-kutyák>. Az első két magyar példánál megfigyelhető, hogy a tővégi 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 végez-e valamilyen morfológiai elemzést erre a kifejezésre.

- A stopszavak kezelésénél azt ellenőriztem, hogy a keresőkulcs megjelenik-e a határozott és határozatlan névelőkkel együtt a találatokban. A keresésekre kapott találatokban azt tanulmányoztam, hogy a keresők külön keresnek-e a megadott névelőkre vagy sem. Ennek a kérdésnek az eldöntése egyértelmű volt, mivel a megadott keresőkulcs a névelővel együtt többnyire ki volt emelve a találatleírásokból, amikor az adott kereső keresett a névelőkre. Azonban a másik esetben, amikor a névelők ki voltak zárva a keresésből csupán a keresőkulcs szerepelt kiemelve a találatleírásokban. A következő keresőkulcs értékeket 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> keresőkulcsokat használtam. A stopszavak esetében gyakori kérdőszavakra, illetve számjegyekre nem kerestem.
- Az ékezetek kezelését kizárólag a magyar nyelvben vizsgáltam meg, hiszen az nem releváns vizsgálati szempont az angol nyelv esetében. Itt nevezetesen két keresőkérdést használtam a <kertem> és az <alma> keresőkulcs értékeket. Arra a kérdésre kerestem a választ, hogy ezeknek a keresőkulcsoknak az ékezetes változatait (*kértem*, *álma*) is visszakeresi-e a kereső a találatok között vagy sem.
- Csonkolásnál gondot jelentett számomra, hogy a magyar nyelvű keresők közül egyedül a Heuréka kereső állította magáról, hogy képes csonkolni egy keresőkifejezésre. Az összes többi magyar nyelvű kereső tájékoztató segédletében nem találtam erre vonatkozó információt. Minden esetben a * karaktert használtam csonkolásra a keresőkifejezés után. Az angol nyelv esetében az <Olympi*> keresőkulcsot 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* (=olimposzi), stb. A magyar nyelv esetében szándékosan törekedtem egy szűkebb, behatároltabb találathalmaz vizsgálatára annak érdekében, hogy a keresőkérdés 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*> keresőkulcs értékkel. Továbbá megvizsgáltam a különböző keresőknél, hogy a <májmétely*> keresőkulcsra kapott találathalmazok ténylegesen tartalmazzák-e a keresőkifejezés ragozott szóalakjait vagy sem. Végeredményül azt kaptam, hogy ezek a találathalmazok nem tartalmazták a keresőkérdés ragozott szóalakjait egyik kereső esetében sem. Ekkor szükségesnek láttam külön egyenként, manuálisan is keresni a <májmétely> keresőkifejezés ragozott szóalakjaira, hogy ellenőrizhessem léteznek-e olyan ragozott szóalakok, amelyeket a keresőknek vissza kellett volna keresniük, de azt mégsem tették meg.

- Végül pedig azt elemeztem, hogy a keresők visszakeresik-e egy keresőkérdés szinonimáit vagy sem. A szinonimák hogyan jelennek meg a kapott találatokban, például azok ki vannak-e emelve a találatleírásokból vagy sem, azok a megadott keresőkulccsal együtt fordulnak-e elő a találatleírásokban, vagy azok attól függetlenül is szerepelnek bennük. Az angol nyelv esetében az alábbi keresőkulcs értékeket vizsgáltam meg ilyen szempontból: <car> (=autó), <glasses> (=szemüveg). A magyar nyelvben pedig a <kutya> és a <vetélkedő> keresőkulcsok szinonimáira kerestem.

Tesztelés közben felfigyeltem egy hibára az AltaVista és az AlltheWeb esetében. Amikor a keresést leszűkítettem kizárólag a magyar nyelvű találatokra, ekkor angol, spanyol és francia nyelvű oldalak is megjelentek az első 100 találat között az <alma> keresőkulcsra. Itt a weboldalakra történő keresést nem domain név szerint, hanem nyelv szerint szűkítettem le. Ebben az esetben a találatok nyelvi szűkítése nem működött kielégítően ennél a két keresőnél. Ugyanez a hiba nem fordult elő a Google és a többi magyar nyelvű kereső esetében.

A továbbiakban az angol és a magyar nyelvű keresők nyelvi lehetőségeit tekintem át a korábban ismertetett vizsgálati szempontok alapján.

3.5. Az angol nyelvű keresőknél talált nyelvi megoldások

3.5.1. Morfológiai elemzés

Ez a visszakeresési művelet nem működött a Google és az AlltheWeb keresők esetében az angol és a magyar nyelvű példákra. Az angol nyelvű keresők közül egyik sem végzett morfológiai elemzést a <tisztességtelen> keresőkulcsra, hanem csak a megadott keresőkérdés pontos szóalakjára kerestek. Egyedül az Altavista kereső találta meg automatikusan a <dog> keresőkulcs érték többes számú alakját, a többi vizsgált keresőről ez nem mondható el. Azonban a morfológiai elemzés ugyanúgy nem működött az Altavista esetében, mint a másik két vizsgált keresőnél sem a magyar nyelvű keresőkulcsokra. Azaz a három kereső nem kereste vissza a <ház>, a <kocsi>, valamint a <kutya> keresőkulcsok többes számú alakját a találataiban.

3.5.2. Stopszavak kezelése

A Google kereső jeleskedik a stopszavak kezelésében a többi keresőhöz képest, mivel az teljes mértékben kizárja a keresésből az *a*, *an*, *the* névelőket az angol nyelvű keresőkulcsokra. Ezenkívül kizárja az *a* határozott névelőt a keresésből, de nem zárja ki az *az*, *egy* névelőket a keresésből a magyar nyelvű keresőkérdésekre. Az Altavista és az AlltheWeb keresők nem zárják ki a határozott és a határozatlan névelőket a keresésből az angol és a magyar nyelvű példákra.

3.5.3. Ékezetek kezelése

Ennek a követelménynek a teljesítése problémás mind a három keresőnél, mert azok megtalálták a *kértem* kifejezést tartalmazó weboldalakat a <*kertem*> keresőkulcsra, valamint az *álma* kifejezést tartalmazó oldalakat az <*alma*> keresőkulcs értékre.

3.5.4. Csonkolás az angol nyelvben

Mind a három kereső megvalósította ugyan a csonkolást a megadott keresőkulcsra, de hibás működésnek tekinthető, hogy a csonkolt <*Olympi*> kifejezés önmagában is megjelent a találatleírásokban egyéb, hosszabb változataival együtt.

A Google kereső a csonkolt <*Olympi*> keresőkifejezést az alábbi hosszabb szavakban fedezte fel: *Olympic* (=olimpiai), *Olympia* (=földrajzi név), *Olympiad* (=olimpiász). Találatai között a csonkolt keresőkifejezés az alábbi hosszabb változatokkal együtt fordult elő: *Olympics* (=olimpia, olimpiai játékok), *Olympian* (=olimposzi).

Az Altavista szolgáltatás a csonkolt <*Olympi*> keresőkifejezést az alábbi hosszabb szavakban találta meg: *Olympic* (=olimpiai), *Olympics* (=olimpia, olimpiai játékok), *Olympia* (=földrajzi név), *Olympiad* (=olimpiász), *Olympian* (=olimposzi). Találataiban a csonkolt keresőkifejezés az alábbi hosszabb változatokkal együtt jelent meg: *Olympia* (=földrajzi név), *Olympic* (=olimpiai), *Olympics* (=olimpia, olimpiai játékok).

Az AlltheWeb kereső a csonkolt <*Olympi*> keresőkifejezést az alábbi hosszabb szavakban fedezte fel: *Olympic* (=olimpiai), *Olympics* (=olimpia, olimpiai játékok), *Olympiad* (=olimpiász), *Olympia* (=földrajzi név), *Olympian* (=olimposzi). Keresési eredményeiben a csonkolt keresőkifejezés az alábbi hosszabb változatokkal együtt jelent meg: *Olympic* (=olimpiai), *Olympics* (=olimpia, olimpiai játékok), *Olympiad* (=olimpiász), *Olympia* (=földrajzi név), *Olympian* (=olimposzi).

3.5.5. Csonkolás a magyar nyelvben

Ez a lekérdezési művelet nem működik a vizsgált keresőknél a megadott példára. Megfigyelhető volt mindegyik esetben, hogy a <*májmétely**> keresőkérdésre kapott találatalmazok nem tartalmazták a keresőkulcs ragozott szóalakjait, ezért külön egyenként, manuálisan kellett azokra keresnem. Ezekben a keresőeszközökben a következő ragozott szóalakokat találtam meg: *májmételynek* (ahol a ”-nek” a részeshatározós eset és a birtokos eset ragja), *májmételyt* (ahol a ”-t” a tárgyeset ragja), *májmételyek* (ahol a ”-k” a többes szám jele), *májmételyről* (ahol a ”-ről” az előljárós eset ragja). A <*májmétely*> keresőkulcs érték csonkolt és csonkolatlan változataira

kapott találatszámokat, valamint a ragozott szóalakokra nyert találatszámokat az 1. táblázatban ismertetem.

	Google	Altavista	AlltheWeb
Csonkolás <i>májmetely*</i>	106 találat ¹	100 találat	62 találat
<i>májmetely</i>	140 találat	94 találat	62 találat
<i>májmetelynek</i>	1 találat	1 találat	1 találat
<i>májmetelyt</i>	6 találat	5 találat	4 találat
<i>májmetelyek</i>	8 találat	6 találat	1 találat
<i>májmetelyről</i>	3 találat		

1. táblázat. A májmetely keresőkérdésre kapott találati eredmények

Ebben a táblázatban az Altavista keresőnél minimális eltérés figyelhető meg a keresőkérdés csonkolt és csonkoltalan változataira lekérdezett találatszámokban. A Google szolgáltatásnál a keresőkifejezés csonkolt változatára kapott találatszám alacsonyabb a csonkoltalan változat találatszámához képest. Az AlltheWeb keresőnél ugyanaz a találatszám fedezhető fel a keresőkérdés csonkolt és csonkoltalan változataira. Ez utóbbi két eset azt tükrözi számunkra, hogy a csonkolás egyáltalán nem működik ezeknél a keresőknél a magyar nyelvű keresőkulcsra, mert alapértelmezés szerint a keresőkifejezés csonkolt változatának nagyobb találathalmazt kell eredményeznie a csonkoltalan változat találathalmazához képest. Továbbá az is észrevehető a táblázatban, hogy mindegyik szolgáltatás csak kevés találatot nem keresett vissza a témával kapcsolatban. Ezek a figyelmen kívül hagyott találatok a keresőkérdés ragozott szóalakjainak a segítségével voltak megtalálhatóak.

3.5.6. Szinonimák visszakeresése az angol nyelvben

Az angol nyelvű keresők mindegyike keresett a <car>, <glasses> keresőkulcsok szinonimáira, azonban ennek a feladatnak a megvalósításában találtam hibás nyelvi értelmezésből eredő hibákat. A <car> (=autó) és a <glasses> (=szemüveg) keresőkulcs értékekre kapott találatokat mindegyik kereső vonatkozásában a 2. táblázatban foglalom össze:

¹ Gondolkodtam utólag azon, hogy mi okozhatta azt, hogy a Google kereső kevesebb találatot keresett vissza a <májmetely> keresőkérdés csonkolt változatára, mint ahogyan az alapesetben elvárható lett volna. Valószínűleg a Google csonkolást végez a háttérben és csonkolással kiszűri az ismétlődő találatokat.

	Google	Altavista	Alltheweb
Lekérdezett találatok	pl.: ~ <i>car</i> – automobile (USA), automotive (USA), auto (USA), motor (nem szinonima!), vehicle (nem szinonima!), racing (nem szinonima!); pl.: ~ <i>glasses</i> – eyeglasses, glassware, sun-glasses (nem szinonima!), goggles (nem szinonima!).	pl.: <i>car</i> – automobile (USA), automotive (USA), auto (USA), vehicle (nem szinonima!); pl.: <i>glasses</i> – spectacles, eyeglasses, glassware, sun-glasses (nem szinonima!), reading-glasses (nem szinonima!), goggles (nem szinonima!).	pl.: <i>car</i> – automobile (USA), automotive (USA), auto (USA), vehicle (nem szinonima!); pl.: <i>glasses</i> – eyeglasses, glassware, sun-glasses (nem szinonima!), reading-glasses (nem szinonima!), goggles (nem szinonima!).

2. táblázat. A *car* és a *glasses* keresőkérésekre kapott találatok

A <*car*> keresőkulcsnak az alábbi szinonimái fordultak elő a találatokban: *automobile (USA)* (=autó, automobil, gépkocsi, gépjármű), *automotive (USA)* (=autó, gépkocsi, automobil, gépjármű), *auto (USA)* (=autó, kocsi).

Továbbá a *motor* (=motor, jelzői értelemben is használatos kifejezés, aminek a jelentése autó(s)-, motoros), *racing* (=verseny-, versenyzés) és a *vehicle* (=jármű, közlekedési eszköz) kifejezések is szinonimaként jelentek meg, azok azonban jelentésükből adódóan nem tekinthetők a <*car*> keresőkérdés tényleges szinonimáinak. (Bár a *motor* kifejezés előfordulhat ilyen értelemben bizalmas jellegű társalgásnál, azonban a *vehicle* szó csak tágabb értelemben használatos kifejezés). A <*glasses*> keresőkulcsnak a következő szinonimái jelentek meg a találatokban: *eyeglasses* (=szemüveg), *spectacles* (=szemüveg, pápaszem). Ezenkívül a *sun-glasses* (=napszemüveg), *reading-glasses* (=olvasószemüveg) és a *goggles* (=motor-szemüveg, védőszemüveg) kifejezésekre is kerestek az angol nyelvű keresők, ezek inkább szemüvegfajtáknak tekinthetők és nem pedig a <*glasses*> keresőkérdés szinonimáinak. Mind a három kereső megtalálta a <*glasses*> keresőkulcs egyes számú alakjának a szinonimáját is a *glassware*-t, aminek a jelentése „üvegáru”. Itt fontos megjegyezni, hogy az egyes számú alak, azaz a *glass* kifejezés jelentései a következők: „üveg, pohár, üvegedény, üvegáru”.

A táblázatban látható az, hogy mindegyik kereső megtalálta a <*car*> keresőkulcs amerikai angolban használatos szinonimáit. Azonban a Google kereső három nem szinonimának minősülő kifejezést is lekérdezett. Az Altavista és az AlltheWeb külön-külön egy nem szinonimának

tekinthető kifejezést fedezett fel. A másik esetben az Altavista kereső két szinonimát talált meg a <glasses> keresőkulcsra és három nem szinonimát az egyes számú alak szinonimáján kívül. A Google és az AlltheWeb egyetlenegy szinonimát kérdezett le az egyes számú alak szinonimáján kívül. A Google két nem szinonimát talált meg, míg az Alltheweb három nem szinonimát fedezett fel.

A Google keresőnél külön a ~ (tilde karakter) segítségével kerestem egy adott keresőkérdés szinonimáira. Itt a megtalált szinonimák kiemelve szerepeltek a találatok leírásában és azok többnyire együtt fordultak elő a keresőkulccsal. Ezzel szemben a másik két keresőnél a szinonimák nem voltak kiemelve a találatok leírásából és azok mindig együtt fordultak elő a keresőkérdéssel.

3.5.7. Szinonimák megtalálása a magyar nyelvben

Mind a három keresővel kerestem a <kutya> és a <vetélkedő> keresőkulcsok szinonimáira, ezekre a keresésekre azonban nem kaptam értékelhető szinonimákat. A vizsgált keresők csupán a keresőkérdés pontos szóalakjára kerestek, és nem vették figyelembe annak szinonimáit a magyar nyelvben ([Tóth2006a] p. 42-45).

3.6. A magyar nyelvű keresőknél talált nyelvi megoldások

3.6.1. Morfológiai elemzés

A magyar nyelvű keresők közül kizárólag a Heuréka keresőről valószínűsíthetjük azt, hogy végez morfológiai elemzést a vizsgált példákra, mert a <tisztességtelen> keresőkérdésre az alábbi szóalakokat fedeztem fel a találatok leírásában, pl.: *tisztesség, tisztességes, tisztességért, tisztességgel, tisztességtelen, tisztességtelenül*. Ennél a keresőnél a <ház>, a <kocsi>, a <kutya> keresőkulcsok többes számú alakja mindig együtt fordult elő azok egyes számú alakjával a találatleírásokban. A megadott keresőkérdések nem voltak kiemelve a találatleírásokból, ami megnehezítette annak az eldöntését, hogy a Heuréka kereső keres-e külön a keresőkulcs többes számú alakjára vagy sem. Az összes többi kereső esetében nem működött a morfológiai elemzés, vagyis azok nem kerestek automatikusan a <ház>, a <kocsi>, a <kutya> keresőkérdések többes számú alakjára. Ezenkívül a <tisztességtelen> keresőkulcsra egyikük sem végzett morfológiai elemzést, mivel azok csak a megadott keresőkérdés pontos szóalakjára kerestek.

3.6.2. *Stopszavak kezelése*

Ennek az elvárásnak a teljesítése eléggé megosztott képet mutatott a vizsgált keresőknél, mert az Origo-Vizsla és a Kurzor keresők nem zárták ki a határozott és a határozatlan névelőket a keresésből *<a ház>*, *<az ember>*, *<egy kocsi>* keresőkulcsokra. Ez azzal volt magyarázható, hogy a keresőkulcsok a határozott, illetve a határozatlan névelőkkel együtt szerepeltek kiemelve a találatleírásokban. A Heuréka kereső esetében nem lehetett egyértelműen megállapítani, hogy a határozott és a határozatlan névelők ki voltak-e zárva a keresésből, mert a keresőkulcsok a névelőkkel együtt nem voltak kiemelve a találatok leírásából. A kapott találatoknál megfigyelhető volt az is, hogy azokban a keresőkérdések sokszor a megadott névelők nélkül fordultak elő. Ebből arra lehetett következtetni, hogy a Heuréka kereső igyekezett kizárni a névelők használatát a keresésből a megadott példákra. A Góliát és az Altavizsla keresők kizárták az *a*, *az* határozott névelőket a keresésből, de nem zárták ki az *egy* határozatlan névelőt a keresésből a vizsgált keresőkérdésekre. A keresőkulcsok és a névelők kiemelésének a problémája egyedül a Heuréka keresőnél jelentkezett, a többi keresőnél ez nem okozott gondot.

3.6.3. *Ékezetek kezelése*

Erről a követelményről elmondható, hogy azt a keresők nagy része sikeresen oldotta meg a vizsgált példákra. Az egyedüli kivételt képezte az Origo-Vizsla kereső, ahol nem működött az ékezetek kezelése. Tehát az megtalálta a *kértem* kifejezést tartalmazó weboldalakat is a *<kertem>* keresőkulcsra és az *álma* kifejezést tartalmazó oldalakat is az *<alma>* keresőkulcsra. A Heuréka keresőnél az ékezetek kezelése csak az ékezhelyes keresési lehetőség kiválasztásánál működött jól a vizsgált keresőkulcsokra. Azonban az automatikus ékezetkezelési lehetőség kiválasztásánál az már nem működött megfelelően a *<kertem>* és az *<alma>* keresőkulcsokra. A Kurzor, Góliát és az Altavizsla keresők esetében hatékonyan működött az ékezetek kezelése a megadott példákra.

3.6.4. *Csonkolás*

Egyik kereső esetében sem működött a csonkolás a vizsgált példára. Sokszor a *<májméty>* keresőkifejezés csonkolt változatára nem kaptam egyetlenegy találatot sem. A Heuréka, Góliát és az Altavizsla keresők esetében ugyanazt a negatív eredményt kaptam. Ennek magyarázata az lehet, hogy a kereső nem képes értelmezni a csonkolást, mint lekérdezési műveletet. Csonkolásnál fennállt egy másik eset is, amikor ugyanannyi találatot kaptam a keresőkérdés csonkolt és csonkoltatlan változataira egyaránt. Ez megint csak azt tükrözi, hogy a csonkolás nem működik egyáltalán, hiszen alapértelmezés szerint a keresőkérdés csonkolt változatának bővebb

találathalmazt kell eredményeznie a csonkoltan változat találathalmazához képest. Ez utóbbi eset állt fenn az Origo-Vizsla és a Kurzor keresőknél. A <májmétely> keresőkulcs csonkolt és csonkoltan változataira kapott találatok, valamint a ragozott szóalakok a rájuk kapott találatoktal együtt megtekinthetők a 3. táblázatban.

	Heuréka	Origo-Vizsla	Kurzor	Góliát	Altavizsla
Csonkolás <i>májmétely*</i>	0 találat	35 találat	30 találat	0 találat	0 találat
<i>májmétely</i>	36 találat	35 találat	30 találat	30 találat	30 találat
<i>májmételyek</i>	1 találat	3 találat	1 találat		
<i>májmételyt</i>		1 találat	2 találat		
<i>májmételynek</i>		2 találat			

3. táblázat. A májmétely keresőkérdésre kapott találati eredmények

A Góliát és az Altavizsla keresőknél nem kaptam egyetlenegy találatot sem a keresőkérdés ragozott szóalakjaira. Ezek a szolgáltatások valószínűleg nem indexelnek olyan weboldalakat, amelyekben szerepelnek a keresett szóalakok.

3.6.5. Szinonimák megtalálása

A <kutya> és a <vetélkedő> keresőkulcsok szinonimáira történő keresések csupán a Heuréka és az Origo-Vizsla keresők esetében voltak eredményesek. Az összes többi kereső nem keresett a megadott keresőkérdések szinonimáira. Ezzel a két keresővel az alábbi szinonimákat kaptam a <kutya> keresőkulcsra: *eb, öleb*. A <vetélkedő> keresőkérdésre a következő szinonimákat találtam meg: *kvíz, kvízzjáték, verseny, agytorna*. Mindkét keresőnél a szinonimák nem voltak kiemelve a találatok leírásából, ami nem teszi egyértelművé az ilyen típusú kereséseket. A szinonimák többnyire együtt fordultak elő a keresőkulccsal a találatok leírásában, és azok csak néhol szerepeltek önállóan. A magyar nyelvű keresők közül kizárólag a Heuréka keresőnél találkozunk olyan beépített tezaurusszal, amely fogalmi relációival támogatja a keresési témával kapcsolatos további lekérdezési lehetőségeket ([Tóth2006b] p. 309-310).

3.7. A vizsgálat eredményeinek összefoglalása

Eredményeimet a korábban felállított hipotéziseim alapján mutatom be.

3.7.1. 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.

Ezenkívül megfigyelhető az is, hogy azok hiányosságokat mutatnak hasonló területeken, például nem kezelik helyesen az ékezeteket, ami fontos felhasználói elvárás a magyar nyelvű keresések vonatkozásában. Ez a hiányosság egyben tükrözi a hátrányukat abban, hogy nem képesek releváns információkat szolgáltatni a magyar felhasználók számára.

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 küzdött meg ezzel a problémával 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.

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

A magyar nyelvű keresők többsége helyesen kezeli az ékezeteket, ezáltal 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.

Irodalom:

[AlltheWeb] AlltheWeb tájékoztató segédlete

<http://www.alltheweb.com/help/index> 2007.03.10.

[Altavista] Altavista tájékoztató segédlete

<http://www.altavista.com/help/search/default> 2007.03.10.

[Bar-Ilan-Gutman2005] 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.

[Google] Google tájékoztató segédlete

<http://www.google.com/support> 2007.03.10.

[Góliát] Góliát tájékoztató segédlete

<http://info.goliat.hu/haszn.htm> 2007.03.10.

[Heuréka] Heuréka tájékoztató segédlete

<http://www.heureka.hu/heureka/tippek.html> 2007.03.09.

[Kurzor] Kurzor tájékoztató segédlete

<http://www.kurzor.hu/help/help.htm> 2007.03.10.

[OrigoVizsla] Origo-Vizsla tájékoztató segédlete

http://www.vizsla24.hu/pages/sugo_tartalomjegyzek.html 2005.08.11.

[Prószéky-Novák2005] 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.

[Taxaurusz2000] Taxaurusz: Növény- és állatrendszertani fogalmak makrotezausa: 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.

[Tóth2006a] Tóth, E.: Exploring the capabilities of English and Hungarian search engines for various queries = Libri, vol. 56. no. 1. (2006), p. 38-47.

[Tóth2006b] Tóth, E.: Kísérlet az internetes keresők nyelvi lehetőségeinek felmérésére = Könyvtári Figyelő, 53. évf. 2. sz. (2006), p. 303-311.

4. A magyar nyelvű keresők vizsgálata

4.1. A vizsgálat célja

Vizsgálatom elsődleges célkitűzésének tekintetem, 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. Más szóval feltételezik, hogy ez a mértéke a rendszer azon képességének, hogy megtalálja a releváns dokumentumokat, s ugyanakkor visszatartja az irrelevánsakat. 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).

4.2. A vizsgálat koncepciója

A magyar nyelvű keresőknél a pontosságot úgy elemeztem, hogy csupán az első hatvan találat relevanciáját tanulmányoztam alaposan. Ezt a döntésemet a Google kereső fejlesztőinek azon észrevételére alapoztam, mely szerint: a Google fejlesztésénél elsődlegesen arra törekedtek, hogy minden egyes keresőkérdésre nagyon releváns találatok jelenjenek meg a találati lista elején, mert a felhasználók általában csak az első néhány tíz találatot tekintik át kereséskor. Azonban ez a pontosság a teljesség hátrányára is érvényesülhet ([Brin-Page98] p. 3). Megjegyzem, hogy a pontosság és a teljesség mindig fordított arányban van egymással, ezért az ideális állapot nem érhető el. Minél teljesebb egy keresés, annál pontatlanabb, azaz a teljesség növelésével a pontosság csökken és viszont: a pontosság növelése a teljesség csökkenését vonja maga után ([Horváth-Sütheő2003] p. 180).

A pontosságot a használat és a hatékonyság szempontjából vizsgáltam. 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. A találatok relevanciájának a meghatározása minden esetben a közölt találati lista sorrendje alapján történt. Kiszámítottam egy százalékkértéket a találati pontosságra vonatkozóan, ami azt fejezte ki, hogy hány százalékban

kaptam releváns találatokat a vizsgált első, második és harmadik hús találatból az adott keresőkérdésre. Ezután megvizsgáltam, hogy lehetséges-e meggyőző következtetést meghatároznom a releváns találatok számára vonatkozóan a találathalmaz hátralévő részében.

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). A csiperkegomba témára irányuló vizsgálatomban a releváns találatok halmaza összesen huszonegy releváns dokumentumból állt. Ezt a halmazt a vizsgált keresőknek az összes visszakeresett találatából nyertem, továbbá nyomon követtem a témára vonatkozó hiperhivatkozásokat és a kapcsolódó szinonimákat is. A kiválasztott releváns weboldalak tulajdonképpen referenciaként kerültek bele ebbe a halmazba és bárholnan felkutathatók voltak a weben. Ezt követően azt tanulmányoztam, hogy az egyes keresők mennyire voltak képesek megtalálni ennek a halmaznak a releváns dokumentumait a saját találathalmazukban. Ezáltal a relatív teljességre vonatkozóan kiszámoltam egy százalékkértéket, ami azt tükrözte, hogy az egyes keresők a vizsgált halmaz releváns dokumentumainak hány százalékát találták meg.

4.3. A weboldalak relevanciája

A pontosság és a teljesség vizsgálatokor szükséges meghatároznom, hogy mikor tekintek egy találatot relevánsnak.

A vizsgálat során egy weboldal akkor minősült relevánsnak, hogyha:

- tartalmazta a megadott keresőkulcsot és/vagy annak más egyéb szinonimáit;
- technikailag bármikor elérhető volt;
- magyar nyelven jelent meg;
- tartalmában és teljes terjedelmében szorosan kapcsolódott a megjelölt témához. Ennek az utóbbi feltételnek a teljesülését az biztosította, hogyha a weboldal tartalmát minden egyes esetben alaposan ellenőriztem és meggyőződtem annak megbízhatóságáról és hitelességéről formai szempontok alapján. Amennyiben ezek a definiált feltételek együttesen nem álltak fenn, akkor egy weboldal nem minősülhetett relevánsnak a vizsgálat számára.

Egy keresőkulcs szinonimáival kapcsolatban megjegyzem, hogy a szemantikus keresők elméletileg megtalálják egy téma szűkebb vagy tágabb értelemben vett szinonimáit. Az azonban már vitatott kérdés számomra, hogy mikor tekintek egy szinonimát a téma szempontjából relevánsnak és mikor nem. Ezért a relevancia értelmezése körültekintő megközelítést igényel az ilyen típusú vizsgálatoknál.

A továbbiakban ismertetem az általam végzett vizsgálat egyes lépéseit és részletezem az azok során felmerülő módszertani kérdéseket.

4.4. A vizsgálatban szereplő keresőeszközök

Összesen öt magyar nyelvű keresőt vontam be a vizsgálatba, amelyek a következők voltak: Heuréka, Origo-Vizsla, Kurzor, Góliát és az Altavizsla. Azért választottam ezeket a keresőket, mert a vizsgálat ideje alatt azok bizonyultak a legjelentősebb keresőeszközöknek a magyar felhasználók számára. Találatszolgáltatásukat megbízhatónak találtam, mert azok rövid válaszdíőn belül megfelelő mennyiségű találatot nyújtottak a lefuttatott keresésekre. Kritikaként megfogalmazható az általam vizsgált keresőkkel kapcsolatban, hogy azok nem szűrték ki az ismétlődő találatokat a találathalmazaikban. Ezek a duplum oldalak mindig növelték a zajos találatok számát az egyes keresésekkor, ebből adódóan azokat irreleváns dokumentumnak tekintettem a vizsgált esetekben.

Ebben a fejezetben gyakran használom a *témakör* fogalmat, ezért szükségesnek tartom meghatározni annak pontos jelentését. Mindig egy speciális keresőkulcs érték jelölhet egy témakört, például a <szökőár> keresőkulcs érték fejezi ki a „szökőár” témakört. Ebben a fejezetben a témaköröket, a keresőkulcsokat, illetve a keresőkulcs értékeket relációs jelek között dőlt betűvel jelölöm.

4.5. A vizsgálat során használt keresőkulcsok

Vizsgálatomat 2006 januárja és júniusa között végeztem el manuális úton, ezzel is kiküszöbölve az automatizálásból eredő téves értelmezéseket és hibákat. Összesen öt témakörben végeztem kereséseket a fent nevezett keresőkön, amelyeken kizárólag magyar nyelvű keresőkérdéseket futtattam le. Választott témaköreim a következők voltak: <csiperkegomba>, <Rákóczi-szabadságharc>, <evészavar>, <kegyhely>, <szökőár>. Az öt témakör közül egyetlenegy irányult a teljesség vizsgálatára (<csiperkegomba>), a fennmaradó négy témakör pedig a pontosság tanulmányozására helyezte a hangsúlyt. A teljesség vizsgálatára egy olyan összetett keresést fogalmaztam meg, amely teljes mértékben figyelembe vette a megadott témakör különböző sajátosságait. A <csiperkegomba> témakörhöz kapcsolódóan a következő keresőkulcs értékeket határoztam meg: <sampinyon>, <termesztett csiperke>, <kétspórás csiperke>, <agaricus bisporus>. Vizsgálatomban elsődlegesen tesztkulcssor létrehozására törekedtem. A keresőkulcsok kiválasztásánál arra ügyeltem, hogy azok különböző témákat képviseljenek.

A pontosság vizsgálatokor rögzítettem keresőnként a különböző keresőkérdésekre kapott összes találat számát. Ennek a figyelembevételére azért volt szükség, mert az összes találat száma

alap esetben befolyásolja azt, hogy pontosan hány releváns találatot kapok a találathalmaz hátralévő részében, amikor egy következtetést állapítok meg a pontosság alakulására vonatkozóan. A lekérdezett dokumentumokat teljes weboldalként mentettem el a saját gépem meghajtójára az Internet Explorer böngészőn keresztül. Mivel a pontosságot csak az első hatvan találat esetében tanulmányoztam, ezért értelemszerűen csak ezt a találatmennyiséget mentettem el. Létrehoztam egy logikus felépítésű könyvtár struktúrát, ami alapján a mentéseket végeztem. Ebben a struktúrában a könyvtárak elnevezései a vizsgált keresőknek feleltek meg és az azokon belüli alkönyvtárak nevei pedig a keresőkérdéseknek.

A találatok tartalmát mindig elemeztem, mert a relevancia megállapítása megköveteli ezt az alapos ellenőrzést. A találatok relevanciájának megállapítására nem hoztam létre külön relevancia kategóriákat, hanem a találatok teljes terjedelméből kiemeltem azt a szövegrészt, ami alapján egy találatot relevánsnak minősítettem. A dolgozat függelékében közlöm az általam relevánsnak tekintett találatokat mind az öt témakörre vonatkozóan, ahol feltüntettem, hogy melyik kereső kereste vissza az adott találatot a vizsgálat során. Továbbá *<csiperkegomba>* témakörben jelölöm, hogy melyek azok a releváns, bárhol felkutatott hiperhivatkozások, amelyeket egyik vizsgált kereső sem talált meg. (Részletekkel kapcsolatban lásd 1. mellékletet.)

4.6. A keresőkulcsok kiválasztása

A vizsgálat megtervezésekor arra törekedtem, hogy megtaláljam azokat a keresőkifejezéseket, amelyek kielégítik a vizsgálat célkitűzéseit. Kezdetben próbakereséseket futtattam le és az azok során nyert tapasztalataimra támaszkodtam a későbbiekben. Ezek a próbakeresések rávilágítottak arra, hogy olyan keresőkérdéseket célszerű használnom a vizsgálatához, amelyek pontosan behatárolják a témát, például ilyen volt a *<szökőár>* keresőkulcs is, ami sok releváns találatot eredményezett. Ennek megfelelően a témát pontosan és konkrétan meghatározó keresőkulcsokat választottam keresőkérdésnek, elkerültem a témát túl általánosan leíró keresőkérdések használatát, amelyek több ezer irreleváns találatot eredményeztek volna. Ilyen volt például a *<rajzfilm>* keresőkulcs érték, amire sok zajos találatot kaptam. Ezért figyelmen kívül hagytam ennek a keresőkulcsnak a használatát a pontosság vizsgálatakor.

Próbakereséseket végeztem többszavas keresőkérdésekre is különböző kereső operátorok segítségével („...” jel, AND, + jel). Kiválasztottam a *<táplálkozási zavar>* keresőkérdést, amely ugyanabban az értelemben használatos az *<evészavar>* keresőkulcs helyett a pszichológia területén. Annak ellenére, hogy a megfelelő operátorokkal kerestem erre a keresőkulcs értékre, számos irreleváns találatot kaptam rá, például sok egészséges életmóddal, étrenddel kapcsolatos oldalt szolgáltatottak a vizsgált keresők. Ezért elvettem a *<táplálkozási zavar>* keresőkulcs

használatát és helyette az <evészavar> keresőkérdést alkalmaztam, amely sokkal több releváns találatot eredményezett a témával kapcsolatban. Kipróbáltam még a <testkép zavar> keresőkulcs értéket is, amely ugyanabban az értelemben használatos, mint az előző két keresőkérdés a pszichológia területén. Ez a keresés ugyan pontos, témába vágó találatokat eredményezett, de azok kis száma miatt nem vettem figyelembe azt. Jelen esetben kis találatszámnak tekintettem a hatvan találat alatti találathalmazokat. Mivel ennek a találatmennyiségnek a megléte a pontosság tanulmányozásának fontos kritériuma volt, ezért ahhoz ragaszkodnom kellett a vizsgálat összeállításakor.

A fenti keresőkulcsokat informális úton nyertem a pszichológia szakterület egyik ismerőjétől, aki a megfelelő szakmai háttérrel rendelkezett és ezt a témát javasolta figyelmembe. Ilyen szempontból a <Rákóczi-szabadságharc> keresőkulcs érték egyértelműen meghatározott egy magyar történelmi eseményt, ezért az sok releváns dokumentumot nyújtott számomra. Összetett alakja ellenére elmondható róla, hogy az egzakt keresőkérdésnek minősült.

A pontosság vizsgálatakor lényeges követelményként fogalmazódott meg bennem, hogy csakis olyan témákat vegyek figyelembe, amelyekre elegendő számú releváns találatot kapok a weben. Ennek értelmében az <organikus építészet> keresőkulcs nem felelt meg a vizsgálat célkitűzésének, mert az kevés releváns weboldalt eredményezett. Ezért annak használatát mellőztem a pontosság vizsgálatakor. Ezzel szemben a <kegyhely> keresőkulcs értékre gazdag és releváns dokumentumanyagot fedeztem fel a weben, tehát azt megfelelő keresőkérdésnek tekintettem.

A teljesség vizsgálatakor nehezebb dolgom volt, hiszen utána kellett nézmem az adott téma különböző sajátosságainak. A <csiperkegomba> témakör esetében csakis olyan keresőkérdéseket választhattam, amelyek azt pontosan körülírták. A keresőkulcsokból – egy kivételével – kizártam a csiperkegomba különböző fajait és azoknak a latin nyelvű megnevezéseit. Elektronikus növény- és állatrendszertani teauruszban és egy gombákkal foglalkozó szakkönyvben utánajártam annak, hogy milyen más egyéb szinonimákat használnak a szakértők a csiperkegomba kifejezésre ([Albert-Locsmándi-Vasas95] p. 20); [Taxaurusz2000]. Ekkor találtam rá a „sampinyon” szinonimára, továbbá rájöttem arra, hogy ezt a megnevezést csupán a kétspórás csiperkére és a természetett csiperkére használják. Itt szükséges megjegyezmem, hogy a csiperkegomba legismertebb természetett fájának a kétspórás csiperkét tekintik általában. Ezért a <sampinyon>, a <kétspórás csiperke> és a <termesztett csiperke> keresőkulcsokat releváns keresőkérdéseknek vettem a téma szempontjából. Ezenkívül megtaláltam a kétspórás csiperkének az <agaricus bisporus> latin nyelvű megnevezését is, amit szintén releváns keresőkulcs értéknek tekintettem. Ezeket a keresőkérdéseket tulajdonképpen a téma előzetes, körültekintő

tanulmányozásával nyertem. Mindegyik keresőkulcsot külön-külön lefuttattam az egyes keresőkön a csiperkegomba témával kapcsolatban. A keresőkérdésekre kapott találatokból alapos és megbízható tartalmi ellenőrzés után kiválasztottam azokat a releváns találatokat, amelyeknek a visszakeresését később ellenőriztem az egyes keresők vonatkozásában.

4.7. A pontosság vizsgálata

4.7.1. „Rákóczi-szabadságharc” témakör

Ebben a témakörben sok releváns forrást találtam az Interneten, ennek részben az volt az oka, hogy nemrégem emlékeztünk meg kitörésének 300. évfordulójáról. E neves évforduló kapcsán számos kiállítás került megrendezésre, ezek közül csak azokat a weboldalatokat tekintettem relevánsnak, ahol részletes leírást találtam a szabadságharc eseményeiről, illetve annak vezéralakjairól. A tudományos igényességgel készült szócikket, tanulmányokat, konferencia-előadásokat relevánsnak tekintettem mindegyik esetben. Ezek a források különböző aspektusból mutatták be a szabadságharcot, például annak adó- és pénzügypolitikáját, egészségügyi, valamint külpolitikai vonatkozásait stb. Nagy számban fordultak elő érettségi tételek a témával kapcsolatban, ezek elsődlegesen az iskolai tanulás céljából íródtak. Találtam olyan írásokat, amelyek a szabadságharc egyes kiragadott eseményeit mutatták be részletesen egy adott településen. Ezeket az oldalakat helyismereti szempontból relevánsnak minősítettem, mert azok elősegítik a megtörtént helyi események jobb megismerését a szabadságharc vonatkozásában. Itt azonban vigyáznom kellett arra, hogy olyan weboldalatokat ne tekintsek relevánsnak, amelyek csak néhány mondattal hivatkoznak a Rákóczi-szabadságharcra, mint a település életében bekövetkezett eseményre. Relevánsnak ítélem meg a Magyar Elektronikus Könyvtárban található digitalizált kézikönyvek vonatkozó fejezeteit, valamint az elektronikus lexikonok (például Magyar Virtuális Enciklopédia, Wikipédia) releváns szócikkeit is. Minden esetben kizártam az angol nyelvű weboldalatokat a releváns dokumentumok köréből.

4.7.2. „Szökőár” témakör

Ezzel a témakörrel kapcsolatban sok híryanagot, cikket és ismeretterjesztő szintű írást találtam. Azonban nem tekintettem relevánsnak azokat az oldalakat, amelyek az egyes országok, és neves cégek jótékonyági célú adományozásairól szóltak a katasztrófa sújtotta területeken. Relevánsnak vettem viszont azokat az oldalakat, amelyek beszámoltak a szökőár bekövetkezéséről, annak következményeiről, valamint a nemzetközi humanitárius szervezetek különböző segélyezési akcióiról. Ezenkívül relevánsnak fogadtam el azokat az oldalakat, amelyek ismeretterjesztő

szinten tárgyalták ennek a természeti katasztrófának a jellegzetességeit, kialakulásának az okait, valamint azt, hogy milyen kifinomult technikai megoldásokkal lehet azt előre jelezni stb. Azonban egyetlen esetben sem tekintettem relevánsnak a témáról szóló nagyon rövid terjedelmű híryanagokat.

4.7.3. „Evészavar” témakör

Az <evészavaroknak> két fő fajtáját ismerjük az „anorexia nervosa”-t és „bulemia nervosa”-t, ezek a betegségek főként fiatalkorú lányoknál jelentkeznek. A téma szempontjából tehát relevánsnak minősítettem azokat a dokumentumokat, amelyek mindkét betegséget együttesen vagy külön-külön ismertették. Ebből adódóan egyértelmű volt annak a megállapítása, hogy a dokumentum a megadott témáról szól-e vagy sem. Számos orvostudományi szakfolyóiratban megjelent cikket és ismeretterjesztő szintű weboldalt fedeztem fel a témával kapcsolatban, amelyek bemutatták ennek a két betegségnek a tüneteit, valamint azok terápiás kezelését. Továbbá a megtalált források rávilágítottak azokra a társadalmi, kulturális, biológiai és pszichológiai tényezőkre, amelyek az evészavarok kialakulásához vezethetnek.

4.7.4. „Kegyhely” témakör

Ezzel a témakörrel kapcsolatban sok értékes, ismeretterjesztő szintű dokumentumot fedeztem fel az Interneten. Több központi gyűjtőoldal állt rendelkezésemre ebben a témában, amelyeknek az egyes weboldalait a vizsgált keresők visszakeresték. Azonban ezeket az oldalakat nem tekintettem releváns dokumentumnak. Ezenkívül számos újságcikk és katolikus plébánia honlapja tájékoztatót engem erről a témáról. Relevánsnak minősítettem azokat az oldalakat, amelyek egy adott kegyhely történetét, illetve az ott meghonosodott búcsújárás több évszázados fejlődését tekintették át. A keresők többször találtak meg olyan oldalakat is, amelyek a „búcsújáróhely” szinonimát tartalmazták a kegyhely keresőkérdésre. Azonban nem fogadtam el releváns dokumentumnak azokat a híryanagokat, amelyek csak érintőlegesen szóltak a témáról. Például a római katolikus pápa vagy egy zarándokcsoport látogatásáról szóltak egy adott időpontban a kegyhelyen.

4.8. A teljesség vizsgálata

Visszaültnék a korábban elmondottakra először, tehát a releváns találatoknak egy halmazát kellett létrehoznom a vizsgált keresők által visszakeresett összes találatból <csiperkegomba> témakörben. Ezután pedig azt tanulmányoztam, hogy az egyes keresők milyen arányban találták meg ennek a halmaznak a releváns dokumentumait.

4.8.1. „Csiperkegomba” témakör

A halmazba kerülő releváns dokumentumok kiválasztásakor a következő esetek merültek fel számomra, amikor megfelelő döntéseket kellett hoznom. Releváns dokumentumnak tekintettem minden olyan tudományos kutatást, szakmai beszámolót, amely valamilyen speciális szempontból vizsgálta a csiperkegomba-termesztést. Doktori tézisek, disszertációk, szakdolgozatok, valamint szakmai jellegű pályázati anyagok reprezentálták a tudományos kutatást ebben a témakörben. Nem tettem különbséget a különböző tudományterületek megközelítései között, mindegyiket egyenrangúnak tekintettem. Ezenkívül releváns dokumentumnak minősítettem minden olyan írást, amely ismeretterjesztő szinten mutatta be a csiperkegomba-termesztést, illetve az tartalmazta a csiperkegombának, mint növényfajnak az ismertetőjegyeit.

A releváns dokumentumok kiválasztásakor automatikusan figyelmen kívül hagytam a csiperkegomba egyes fajaira vonatkozó weboldalakat, mert azokra külön nem kerestem a keresőkérdések során. Egyedüli kivételt képezett a kétspórás csiperke gombafaj, valamint annak a latin nyelvű megnevezése, amelyekre a téma szempontjából indokolt volt keresnem. Nem tekintettem releváns dokumentumnak azokat a weboldalakat, amelyek csak nagyon röviden – 1 vagy 2 mondattal hivatkozva – szóltak a témáról. Amennyiben a dokumentum több gombafajt vizsgált meg egy adott aspektusból és csak röviden utalt a csiperkegombára, akkor azt kizártam a releváns dokumentumok köréből, például a gombaszagokról vagy a gombamérgezésről szóló dokumentumokat. Ha a dokumentum átfogó képet nyújtott a gombatermesztési ágazat helyzetéről a magyar mezőgazdaságban, akkor azt szintén nem tekintettem releváns dokumentumnak, hiszen az nemcsak a csiperkegomba termesztésével foglalkozott, hanem más egyéb gombafajok termesztésével is. Nem releváns dokumentumnak minősítettem az angol nyelvű weboldalakat és a 404-es nem létező oldalra történő hivatkozással [BernersLee-Fielding-Frystyk96, RFC1945] megjelent weblapokat.

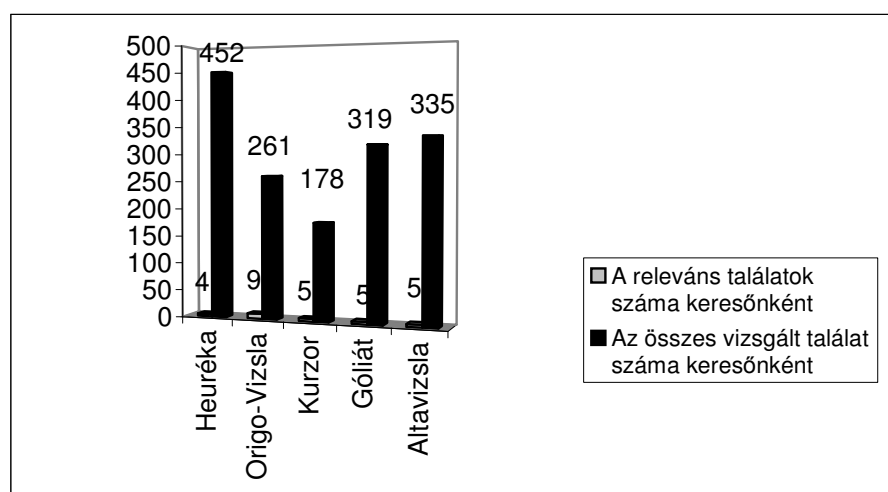
Amikor azt állapítottam meg, hogy az egyes keresők pontosan hány találatot kerestek vissza a releváns dokumentumok halmazából, a következő esetekkel találkoztam. Ha a halmazban lévő weboldal ugyanazon az URL-en volt lekérdezhető, mint ahol azt korábban a kereső felkutatta, akkor azt megtalált dokumentumnak minősítettem. Természetesen ez az eset fordult elő a leggyakrabban. Ha az oldal párhuzamosan több URL-en volt elérhető, a kereső pedig ezek közül csak az egyik helyen találta meg a dokumentumot, akkor azt felkutatott releváns találatnak vettem. Azonban, hogyha a kereső a dokumentumnak egy sokkal korábbi URL-jét indexelte, ahol az már ténylegesen nem volt elérhető és csupán a „cache”-ben tárolt változatát szolgáltatta, akkor

azt már nem tekintetem megtalált weboldalnak. Döntésemnek oka az volt, hogy a releváns dokumentumok halmazát létező dokumentumokra mutató halmaznak tekintetem.

4.9. A teljességre kapott eredmények értékelése

A <csiperkegomba> témakör esetében rögzítettem, hogy összesen hány találatot nyújtottak az egyes keresők a különböző megadott keresőkulcsokra. A keresőkérdésekre kapott összes találatszámokat összeadtam, ezáltal jó gyakorlati mérőszámot kaptam az egyes keresők által szolgáltatott összes találat számára a megadott témakörrel kapcsolatban (a közösen előforduló duplum weboldalak számát az egyes témákra megközelítőleg azonosnak vettem). Ezt követően az öt kereső összes találatának a számát és az általam ténylegesen relevánsnak minősített találatok számát elemeztem, amit az 1. diagram szemléltet számunkra.

4.9.1. A releváns találatok száma és az összes találatok száma



1. diagram. A releváns találatok száma és az összes találatok száma „csiperkegomba” keresőkérésre

A diagramban nyomon követhető, hogy az összes találatszámoknak milyen elhanyagolhatóan kis töredéke minősült relevánsnak a téma szempontjából. Megfigyelhető egy érdekes összefüggés a diagramon, még pedig az, hogy egy nagyobb összes találatból álló halmaz nem feltétlenül eredményez több releváns dokumentumot számunkra. Ez a megállapítás különösen érvényes a Heureka keresőre, hiszen a többi kereső több releváns találatot nyújtott hozzá képest, habár azok összes találatainak a száma sokkal kisebb volt.

A következő lépésben megvizsgáltam, hogy az egyes keresők pontosan hány dokumentumot kerestek vissza az általam létrehozott releváns találatok halmazából, ami huszonegy weboldalt tartalmazott. Az 1. táblázat közli a vizsgálat során kapott eredményeket.

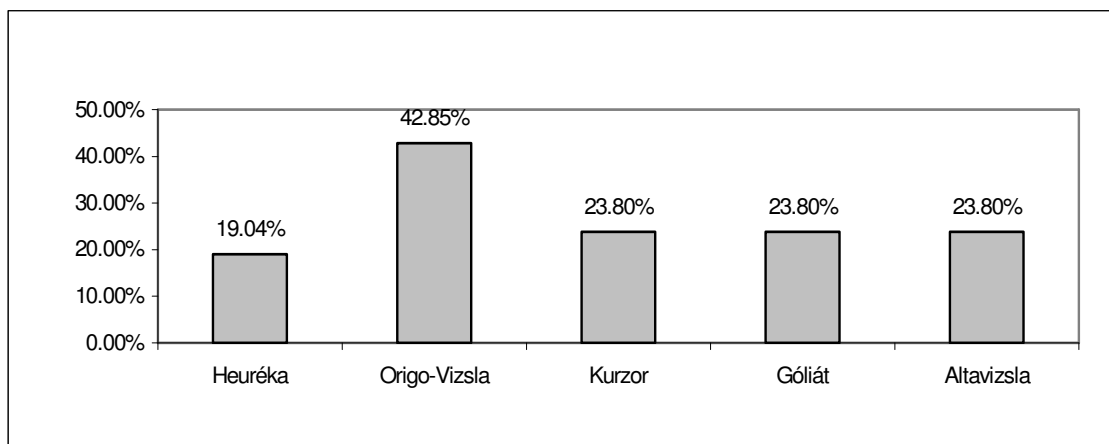
4.9.2. A relatív teljesség vizsgálata

Keresők	Megtalált releváns találatok
Heuréka	4
Origo-Vizsla	9
Kurzor	5
Góliát	5
Altavizsla	5
Átlaga:	5,6

1. táblázat. A keresők által visszakeresett releváns találatok száma és azok kiszámított átlagértéke

A táblázat azt mutatja, hogy az Origo-Vizsla kereső találta meg a legtöbb dokumentumot a releváns találatok halmazából. A többi kereső közel azonos módon teljesített ebben a tekintetben. Tehát a keresők átlagosan 5,6 releváns dokumentumot kutattak fel a weben <csiperkegomba> témakörrel kapcsolatban, ami a vizsgált halmazban lévő huszonegy találat 26,66%-ának felelt meg. Megjegyzem, hogy a bárhonnan felkutatott releváns hiperhivatkozásokat – melyeknek száma összesen 8 volt a vizsgált halmazban – egyik magyar nyelvű kereső sem kereste vissza. (Részletekkel kapcsolatban lásd a Függelék 1. mellékletét.)

Ezután kiszámoltam egy százaléértéket a relatív teljességre vonatkozóan, ami azt fejezte ki, hogy az egyes keresők ezen találathalmaz releváns dokumentumainak hány százalékát keresték vissza. A 2. diagram összefoglalja számunkra a relatív teljességre kiszámított százaléértékeket az egyes keresők vonatkozásában.



2. diagram. A keresők relatív teljeség szerinti százalékos megoszlása

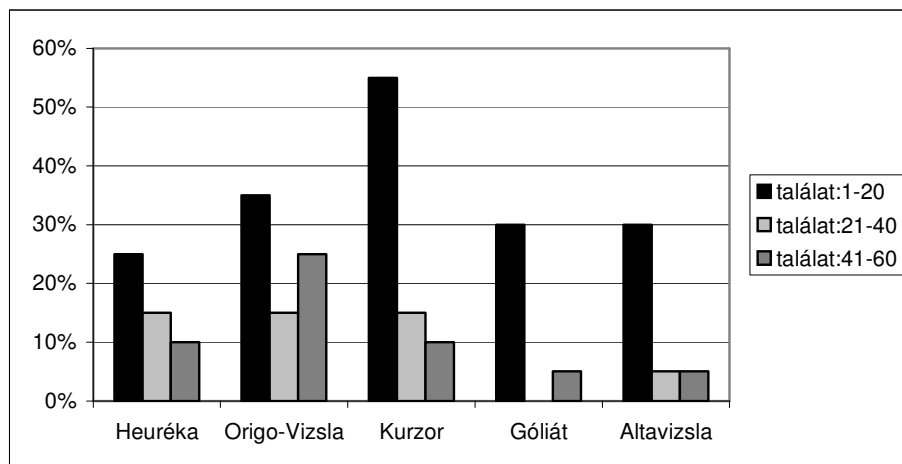
A diagramból leolvasható, hogy az Origo-Vizsla kereső találta meg a legeredményesebben a vizsgált találathalmaz releváns dokumentumait, hiszen azok 42,85%-át kutatta fel a weben. A rangsorban ez után következtek a Kurzor, Góliát és az Altavizsla keresők 23,80%-kal, mivel azok azonos mennyiségű releváns dokumentumot kerestek vissza ebből a találathalmazból. Végül pedig a Heuréka kereső került az utolsó helyre 19,04%-kal, mert az kissé lemaradt az előző három kereső lekérdezési teljesítményétől.

4.10. A pontosságra kapott eredmények értékelése

Négy különböző keresőkulcsot futtattam le a keresőkön a pontosság tanulmányozására irányulóan. A keresőkérdésekre kapott találati listákban csupán az első hatvan találat relevanciáját vizsgáltam. A találatok pontosságára vonatkozóan kiszámítottam egy százalékkértéket, ami azt tükrözte, hogy hány releváns találatot kaptam az első, a második és a harmadik húsz visszakeresett találatból.

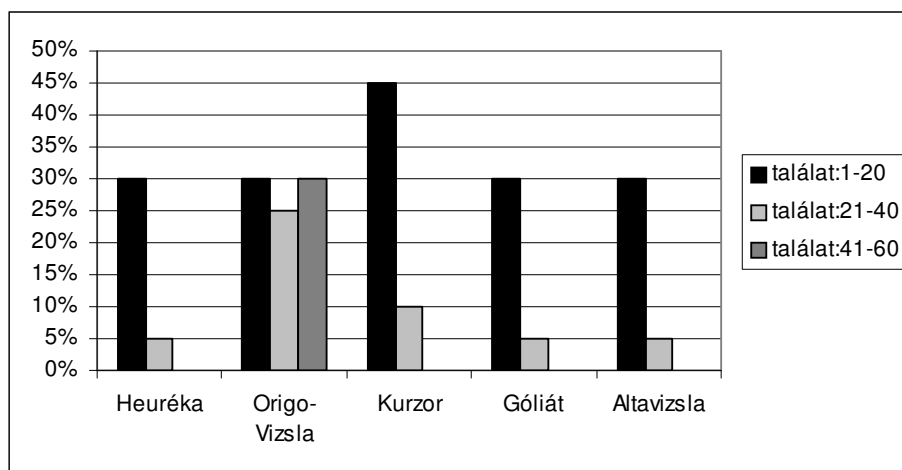
A továbbiakban közlöm az egyes keresőkulcsokra kalkulált találati pontosság értékeket keresőnként az első, a második és a harmadik húsz találat esetében.

4.10.1. A találati pontosság változása keresőkulcsenként



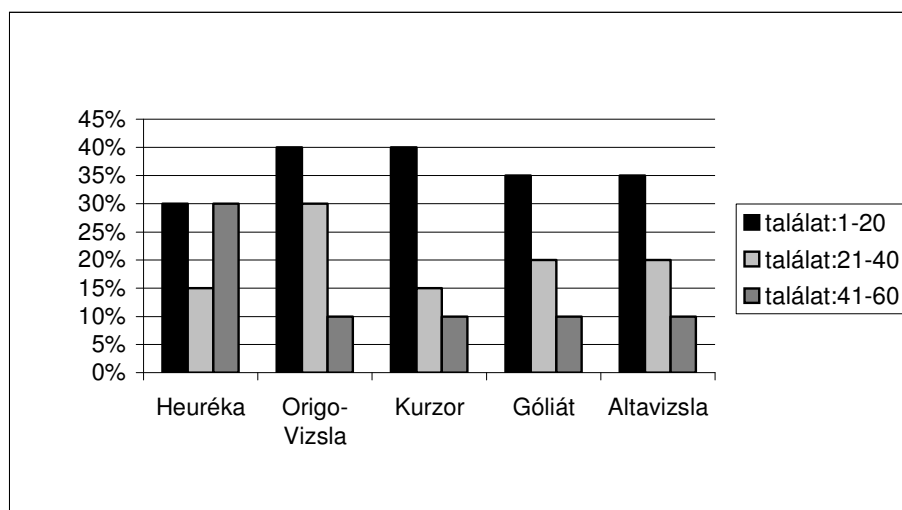
3. diagram. „Rákóczi-szabadságharc” keresőkulcs – találati pontosság

A diagram azt mutatja, hogy mindegyik kereső az első húsz találatban szolgáltatotta a legtöbb témába vágó találatot a <Rákóczi-szabadságharc> keresőkulcsra. A Kurzor keresőnél kiugróan magas volt a releváns találatok száma az első húsz találatban, majd azt követte az Origo-Vizsla kereső. A Góliát és az Altavizsla keresők azonos mennyiségű releváns találatot nyújtottak az első húsz találatban, végül pedig a Heuréka kereső kutatta fel a legkevesebb releváns találatot ugyanezen a helyen. Ezenkívül megfigyelhető volt az is, hogy a Heuréka és a Kurzor keresők szigorúan monoton csökkenő tendenciában hozták a releváns találataikat az első hatvan találatban erre a keresőkérdésre. Ezzel szemben az Origo-Vizsla és a Góliát keresőknél nem monoton tendenciáról beszélhetünk a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkulcsra. A Góliát és az Altavizsla keresők között minimális különbség jelentkezett a releváns találatok számában, ami a második húsz találat esetében volt tapasztalható. Az Altavizsla kereső azonos számú releváns találatot keresett vissza a második és a harmadik húsz találatban. Míg a Góliát kereső egyáltalán nem szolgáltatott releváns találatot a második húsz találatban. Azonban az általa visszakeresett releváns találatok száma a harmadik húsz találatban megegyezett az Altavizsla keresőnek ugyanezen a helyen lévő releváns találatainak a számával. Az Altavizsla kereső monoton csökkenő tendenciát valósított meg a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkulcsra.



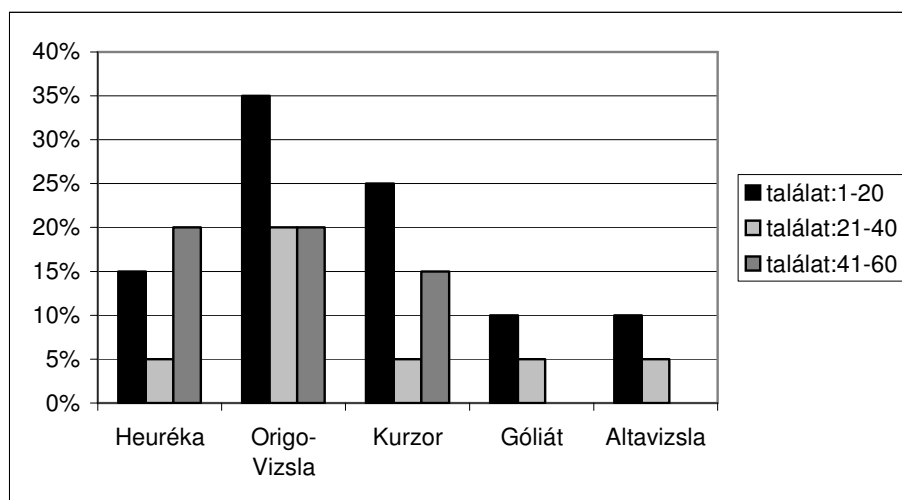
4. diagram. „Evészavar” keresőkulcs – találati pontosság

A diagram alapján elmondható az, hogy a keresők többsége a legtöbb releváns találatot kereste vissza az első húsz találatban az <evészavar> keresőkulcsra. Ez alól az Origo-Vizsla kereső volt az egyedüli kivétel, mert az ugyanannyi releváns találatot szolgáltatott az első húsz találatban, mint a harmadik húsz találatban erre a keresőkérdésre. A Kurzor kereső nyújtotta a legtöbb releváns találatot az első húsz találatban, a többi kereső pedig azonos számú releváns oldalt hozott ugyanezen a vizsgált helyen. Az Origo-Vizsla kereső kivételével az összes többi keresőnél ugyanaz a szigorúan monoton csökkenő tendencia volt jelen a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkérdésre. Ráadásul ezen keresők közül egyik sem keresett vissza egyetlenegy releváns találatot a harmadik húsz találatban. Az Origo-Vizsla keresőnél viszont nem monoton tendencia volt megfigyelhető a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkulcsra.



5. diagram. „Kegyhely” keresőkulcs – találati pontosság

A diagramról leolvasható, hogy a keresők nagy része a legtöbb releváns találatot nyújtotta az első húsz találatban a <kegyhely> keresőkulcsra. A Heuréka kereső viszont eltérően viselkedett a többi keresőhöz képest, mert az ugyanannyi releváns találatot keresett vissza az első húsz találatban, mint a harmadik húsz találatban erre a keresőkérdésre. Az Origo-Vizsla és a Kurzor keresőknél magas volt a releváns találatok száma az első húsz találatban, amely megegyezett egymással. A Góliát és az Altavizsla keresők szintén azonos számú releváns találatot szolgáltatott az első húsz találatban, de azok már a második helyre kerültek a rangsorban. A Heuréka kereső kevésbé ugyan, de mennyiségileg lemaradt az előző két kereső releváns találatainak a számától az első húsz találatban. A Heuréka keresőt leszámítva az összes többi keresőnél észrevehető volt egy szigorúan monoton csökkenő tendencia a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkérdésre. Azonban a Heuréka kereső nem monoton tendenciát valósított meg a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkulcsra.



6. diagram. „Szökőár” keresőkulcs – találati pontosság

A diagram azt szemlélteti, hogy a keresők többsége a legtöbb releváns találatot kereste vissza az első húsz találatban a <szökőár> keresőkulcsra. A Heuréka kereső volt az egyedüli kivétel, amely ettől eltérően viselkedett, mert az több releváns találatot szolgáltatott a harmadik húsz találatban, mint az első húsz találatban erre a keresőkérdésre. Az Origo-Vizsla kutatta fel a legtöbb releváns találatot az első húsz találatban. Azt követte a Kurzor kereső alacsonyabb releváns találatszámmal ugyanezen a helyen. A Heuréka kereső került a harmadik helyre a rangsorban a releváns találatok száma alapján az első húsz találatban. Végül pedig a Góliát és az Altavizsla keresők lettek az utolsók a rangsorban, továbbá azok azonos számú releváns találatot kérdeztek le az első húsz találatban. A Góliát és az Altavizsla keresőknél szigorúan monoton csökkenő tendencia jelentkezett a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkérdésre. Ráadásul azok azonos számú releváns találatot nyújtottak az első hatvan találatban erre a keresőkulcsra. Az Origo-Vizsla kereső monoton csökkenő tendenciát mutatott a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkulcsra. A Heuréka és a Kurzor keresők pedig nem monoton tendenciát valósítottak meg a releváns találatok lekérdezésében az első hatvan találatban erre a keresőkérdésre.

4.10.1.1. A diagramokból levonható konklúziók

A diagramok egybehangzóan azt tükrözik, hogy a vizsgált magyar nyelvű keresők többsége arra törekedett, hogy a legtöbb releváns találat jelenjen meg az első húsz találatban mind a négy keresőkulcsra. Ezenkívül a kapott eredmények alapján megfigyelhető, hogy a Kurzor kereső az első két keresőkulcs értékre a legtöbb releváns dokumentumot kereste vissza az első húsz találatban. A harmadik <kegyhely> keresőkulcsra az holtversenybe került az Origo-Vizsla keresővel, mivel mindketten azonos számú releváns találatot szolgáltatott ugyanazon a vizsgált

helyen. A negyedik <szökőár> keresőkulcsra azonban a Kurzor kereső már elveszítette az első helyét, mert azt megelőzte az Origo-Vizsla kereső a visszakeresett releváns dokumentumok száma terén az első húsz találatban. Amennyiben ezeket az eredményeket a felhasználó szempontjából közelítjük meg, akkor megállapíthatjuk, hogy a Kurzor és az Origo-Vizsla keresők bizonyultak a legjobbnak a használat szempontjából, hiszen azok nyújtották a lehető legnagyobb számban a releváns oldalakat az első húsz találatban.

Megjegyzem, hogy a Góliát és az Altavizsla keresők minimális eltéréssel ugyan, de közel azonos releváns találatmennyiséget kerestek vissza az első hatvan találatban mind a négy keresőkulcsra. Továbbá azok többnyire – egy keresőkérdés kivételével – megegyező tendenciában hozták a releváns dokumentumokat az első hatvan találatban a négy keresőkulcs értékre. Azonban ennek a két keresőnek a keresőkulcsokra visszakeresett összes találataik száma már különböző volt. Ebből adódóan komolyabb következtetést nem vonhatok le az adatbázisaik megegyezésére vonatkozóan.

4.10.2. A találatok pontosságára vonatkozó következtetés vizsgálata

A következő lépésben tanulmányoztam az első, a második, valamint a harmadik húsz találatban kapott releváns találatok számát mind a négy keresőkulcsra az egyes keresőknél. Annak érdekében, hogy meggyőző következtetést határozzak meg a releváns találatok számára vonatkozóan a találati lista további részében, figyelembe kell vennem azt is, hogy a keresők külön-külön összesen hány találatot keresnek vissza az egyes keresőkulcs értékekre. Az alábbi táblázatban összefoglalom az első, a második, és a harmadik húsz találatban lévő releváns találatok számát, valamint az összes találat számát, amit a keresők külön-külön az egyes keresőkérdésekre nyújtottak. A következő táblázatban azonban hely hiánya miatt nem részletezem, hogy melyik kereső pontosan mely keresőkulcsra szolgáltatta az összes találat számát és a releváns találatok számát. (A táblázat teljes terjedelmét lásd a 7. mellékletben. Az egyes keresőkulcsokra kapott releváns és összes találatok számai megtekinthetők a 6. mellékletben.) A táblázat utolsó sorában közlöm az oszlopokban lévő adatoknak az összesített számértékeit.

Releváns találatok száma az első 20 találatban	Releváns találatok száma a második 20 találatban	Releváns találatok száma a harmadik 20 találatban	Összes találat száma	
5	3	2	836	
7	3	5	4 240	
11	3	2	157	
6	0	1	16 641	
6	1	1	60 894	
6	1	0	125	
6	5	6	1 070	
9	2	0	1 422	
6	1	0	5 296	
6	1	0	5 296	
3	1	4	446	
7	4	4	6 560	
5	1	3	12 766	
2	1	0	247	
2	1	0	247	
6	3	6	255	
8	6	2	2 790	
8	3	2	3 469	
7	4	2	5 286	
7	4	2	5 286	
Összesen:	123	48	42	133 329

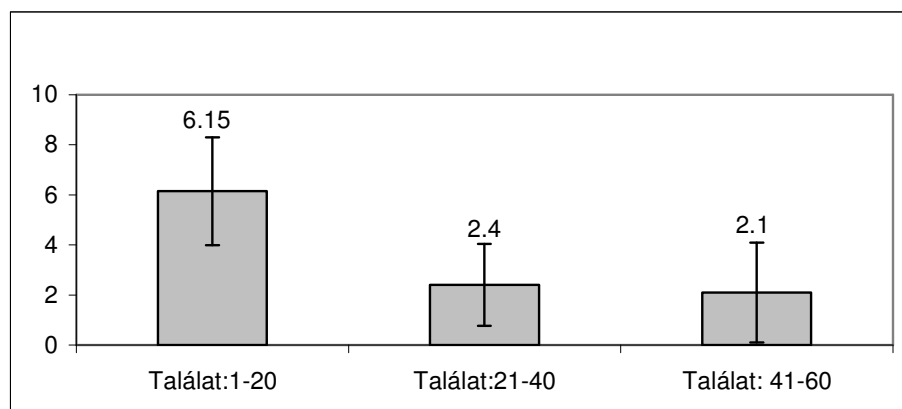
2. táblázat. A keresőkulcsokra lekérdezett összes találatszámok és releváns találatszámok

A táblázatban lévő adatokból kiderül, hogy képtelenség konkrét összefüggést találni a releváns találatok száma és az összes találat száma között. Ez többek között abban is megnyilvánul, hogyha például közel azonos nagyságrendű összes találatszámokat veszek alapul, akkor nem kapok egymáshoz hasonló nagyságrendű releváns találatszámokat azokra, csupán a releváns találatszámoknak különféle variációit nyerem. Továbbá a táblázatból az is leolvasható, hogy a keresők összesen hat esetben egyetlen releváns találatot sem kerestek vissza a harmadik húsz találatban a megadott keresőkulcsokra. Ezért ezekben az esetekben nem határozhatok meg meggyőző következtetést a releváns találatok számára vonatkozóan a találathalmaz további részében.

Összességében tehát megállapíthatom, hogy nem fedezhetek fel semmilyen összefüggést az első hatvan találatban lévő releváns találatok száma és az összes találat száma között, ezért nem adhatok meg meggyőző következtetést a releváns találatok számára vonatkozóan a találathalmaz hátralévő részében.

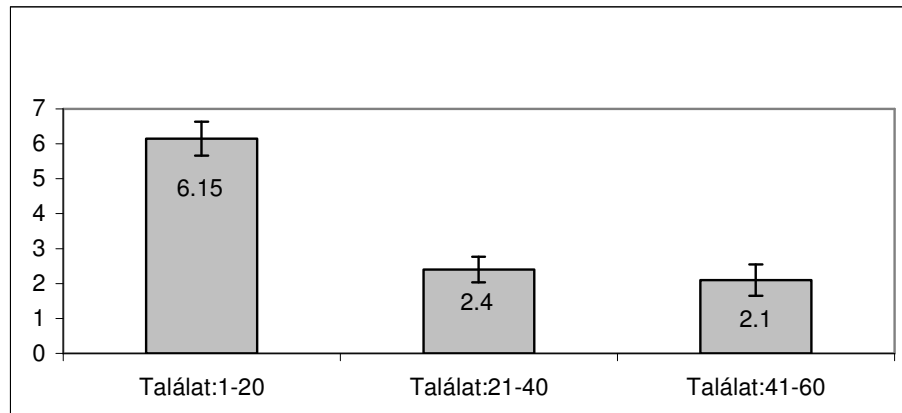
4.10.3. A szórás és a várható érték elemzése

Ebben a részben a szórást és a várható értéket különböző megközelítésekből mutatom be, amelyeket számos diagrammal illusztrálok. Mindegyik esetben az oszlopdiagramok felett közlöm a várható értéket, valamint függőleges vonallal jelölöm a tapasztalati szórást és a mintaátlag szórását pozitív és negatív értéktartományban. A diagramok elemzésével kapcsolatos megjegyzések a minta nagysága miatt természetesen nem tekinthetők statisztikai következtetéseknek, azonban határozottan kijelölhetik egy jövőbeli statisztikai elemzés célkitűzéseit.



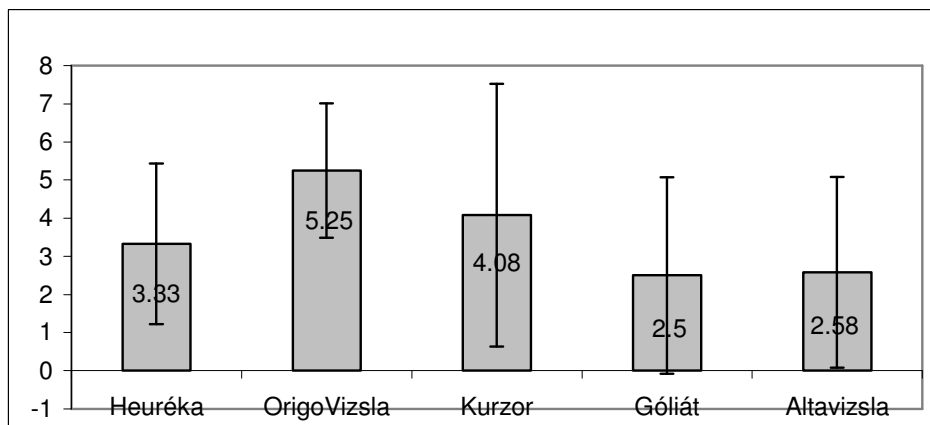
7. diagram. Az első, a második és a harmadik húsz találatban lévő releváns találatok átlaga és tapasztalati szórása a keresőkulcsokra és a keresőkre összesítve

A diagram azt mutatja, hogy a vizsgált keresők releváns találatainak az átlagértéke mind a négy keresőkulcsra szigorúan monoton csökkenő tendenciát követ az első hatvan találat esetében. A releváns találatok tapasztalati szórása az első húsz találatban azt jelzi, hogy a valószínűségi változó értéke nagyobb valószínűséggel tér el az elméleti várható értéktől, mint a második és a harmadik húsz találatban jelentkező tapasztalati szórás esetében. A vizsgált keresők hatékonysága az első húsz találatban mind a négy keresőkulcsra a legnagyobbnak tekinthető, mivel a releváns találatok átlagértéke itt bizonyul a legmagasabbnak. Azonban nagyon alacsony a keresők hatékonysága a második és a harmadik húsz találatban mind a négy keresőkérdésre, amit a releváns találatok rendkívül alacsony átlagértékei tükröznek. Továbbá minimális különbség figyelhető meg a keresők hatékonyságában a második és a harmadik húsz találatban mind a négy keresőkulcsra, mert csekély eltérés jelentkezik a releváns találatok átlagértékei között ezeken a helyeken.



8. diagram. Az első, a második és a harmadik húsz találatban lévő releváns találatok átlaga és mintaátlag szórása a keresőkulcsokra és a keresőkre összesítve

A diagramról leolvasható, hogy egy véletlenszerűen választott minta átlagértéke csekély mértékben tér el a populáció elméleti várható értékétől az első, a második és a harmadik húsz találatban mind a négy keresőkulcsra és az összes keresőre vonatkozóan. Amennyiben a mintaválasztást, a mérést és az átlagszámítást többször is elvégezzük, akkor a kiszámított mintaátlagok csak minimális mértékben fognak szórni a populáció elméleti várható értékétől. Tehát ez a diagram a mintaátlag hibáját jellemzi számunkra.

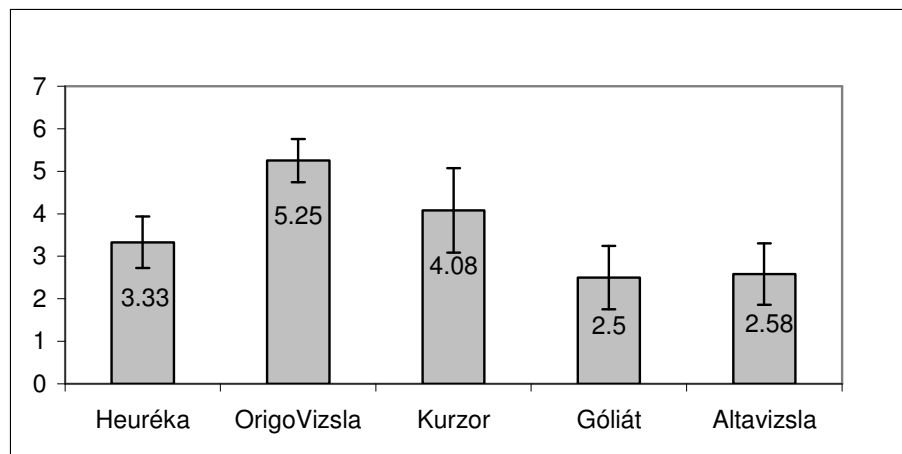


9. diagram. Az első hatvan találatban mért releváns találatok átlaga és tapasztalati szórása a négy keresőkulcsra összesítve keresőnként

A diagram alapján elmondható az, hogy a Kurzor keresőnél a legnagyobb a valószínűsége annak, hogy a valószínűségi változó értéke eltér az elméleti várható értéktől az első hatvan találatban mind a négy keresőkulcsra. Ez a valószínűség azonban nagy mértékben csökken a Góliát és az Altavizsla keresőknél az első hatvan találatban a négy keresőkérdésre. Az pedig kisebb mértékben ugyan, de tovább csökken a Heuréka keresőnél az első hatvan találatban a négy keresőkulcsra. Az Origo-Vizsla keresőnél viszont a legkisebb a valószínűsége annak, hogy a valószínűségi változó értéke nagy mértékben eltérjen az elméleti várható értéktől az első hatvan

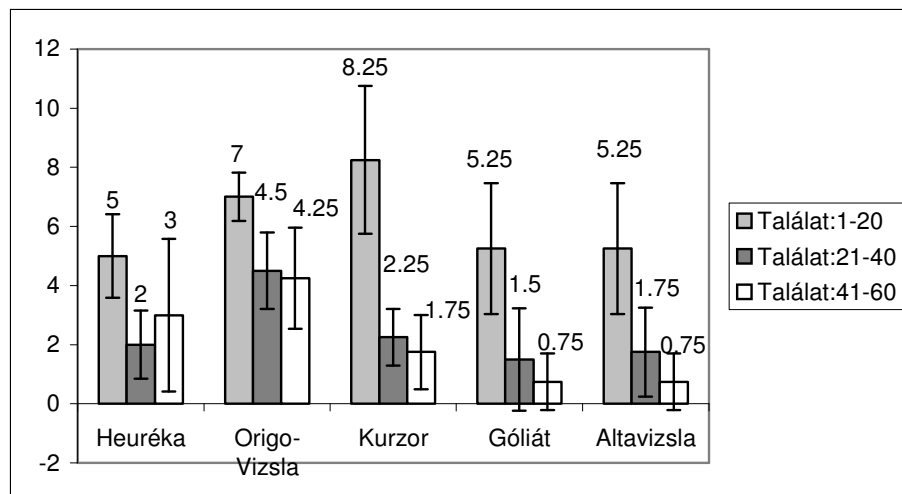
találatban mind a négy keresőkulcsra. Továbbá ez a kereső tűnik a leghatékonyabbnak az első hatvan találatban mind a négy keresőkulcsra, hiszen a releváns találatainak az átlagértéke a legmagasabb a többi vizsgált keresőéhez viszonyítva. Ugyanebből a szempontból a Heuréka kereső a második helyen szerepel, bár releváns találatainak az átlagértéke sokkal kisebb az Origo-Vizslához képest az első hatvan találatban a négy keresőkérdésre. A hatékonyságot illetően a Kurzor kereső kerül a harmadik helyre, hiszen a releváns találatainak az átlagértéke alacsony az első hatvan találatban a négy keresőkulcsra, amit az itt található rendkívül magas tapasztalati szórás érték eredményez. Végül pedig a hatékonyság szempontjából az utolsó helyre sorolhatók a Góliát és az Altavizsla keresők. Fontos hangsúlyozni, hogy nincs eltérés ez utóbbi két kereső hatékonysága között, mivel azok tapasztalati szórása közel azonos nagyságrendű, valamint minimális különbség jelentkezik a releváns találataik átlagértékei között az első hatvan találatban mind a négy keresőkulcsra.

Feltételezésem szerint a diagramban látható számottevőbb tapasztalati szórás értékek abból adódhatnak, hogy a keresőknél az új weboldalak indexelésének az átfutási ideje meglehetősen lassú, ami jelentős mértékben kihat a keresők hatékonyságára. Tudomásom szerint a Heuréka [Heuréka] és az Origo-Vizsla [Origo-Vizsla] keresőknél a keresőrobot átlagosan 2-3 hetente gyűjti be az új oldalakat. A Kurzor keresőről [Zádori2007] elmondható, hogy annak csak néhány hetente frissül az adatbázisa. Azonban a Heuréka és a Kurzor keresők a híryanagokat napi rendszerességgel követik nyomon. Hiteles információk [Hódi2007] szerint a jelenleg publikusan elérhető Góliát kereső egyáltalán nem frissíti az adatbázisát. Továbbá az Altavizsla kereső [Bocsán2007] a Góliát keresőt használja a háttérben, saját indexet külön nem hoz létre. Ez utóbbi értesülést támasztják alá a vizsgálat eredményei is, mert a Góliát és az Altavizsla keresők ugyanazokat a releváns dokumentumokat keresték vissza mind az öt témakörre vonatkozóan egyetlenegy weboldal kivételével. (Részletek megtalálhatók a Függelék 1-5. mellékleteiben.)



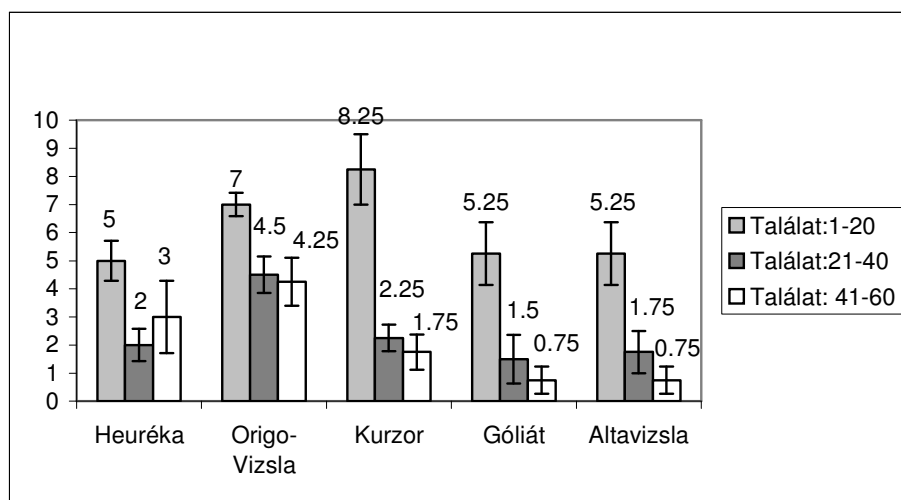
10. diagram. Az első hatvan találatban mért releváns találatok átlaga és mintaátlag szórása a négy keresőkulcsra összesítve keresőnként

A diagram azt tükrözi, hogy a Kurzor keresőnél egy tetszőlegesen választott minta átlagértéke jelentős mértékben eltér a populációban lévő elméleti várható értéktől. Tehát a mintaátlag hibája ennél a keresőnél a legnagyobb a többi keresőhöz viszonyítva az első hatvan találatban mind a négy keresőkulcsra. Ez után következnek a Góliát és az Altavizsla keresők kicsit kisebb, de közel azonos nagyságrendű mintaátlag hibával az első hatvan találatban a négy keresőkérdésre. A Heuréka és az Origo-Vizsla keresőknél a mintaátlag hibája tovább csökken ugyanezen a vizsgált helyen mind a négy keresőkulcsra.



11. diagram. A keresők első, második és harmadik húsz találatában lévő releváns találatok átlaga és tapasztalati szórása mind a négy keresőkulcsra

A diagram azt szemlélteti, hogy a Heuréka kereső kivételével az összes többi keresőnél a releváns találatok átlagértékei szigorúan monoton csökkenő tendenciát követnek az első, a második és a harmadik húsz találatban mind a négy keresőkulcsra. A Heuréka keresőnél a releváns találatok átlagértékei nem monoton tendenciát mutatnak az első, a második és a harmadik húsz találatban mind a négy keresőkérdésre. Továbbá a Heuréka és az Origo-Vizsla keresőknél a harmadik húsz találatban a legnagyobb a valószínűsége annak, hogy a valószínűségi változó értéke eltér az elméleti várható értéktől ugyanezen keresőknek az első és a második húsz találatához képest. Míg a Kurzor, a Góliát és az Altavizsla keresőknél ez a nagy valószínűség az első húsz találatban fordul elő ugyanezen keresőknek a második és a harmadik húsz találatához képest. Mindez pedig a releváns találatok tapasztalati szórás értékei alapján állapítható meg az első, a második és a harmadik húsz találatban mind a négy keresőkulcsra. Ezenkívül jól látszik a diagramon, hogy mindegyik kereső az első húsz találatban arra törekszik, hogy a leghatékonyabb legyen mind a négy keresőkulcsra. Ez azzal magyarázható, hogy mindegyikjük releváns találatainak az átlagértéke sokkal magasabb az első húsz találatban, mint a második és a harmadik húsz találatban. A Góliát és az Altavizsla keresők között nincs eltérés a hatékonyság tekintetében, hiszen a releváns találataik közel azonos átlagértéket és tapasztalati szórás értéket tükröznek az első, a második és a harmadik húsz találatban mind a négy keresőkérdésre.



12. diagram. A keresők első, második és harmadik húsz találatában lévő releváns találatok átlaga és mintaátlag szórása mind a négy keresőkulcsra

A diagram azt mutatja, hogy a Heuréka és az Origo-Vizsla keresőknél a harmadik húsz találatban egy tetszőlegesen választott minta átlagértéke nagyobb mértékben tér el a populáció elméleti várható értékétől ugyanezen keresőknek az első és a második húsz találatához képest. Ez a jelentősebb mértékű eltérés figyelhető meg a Kurzor, a Góliát és az Altavizsla keresők esetében az első húsz találatban ugyanezen keresőknek a második és a harmadik húsz találatához képest. Ezenkívül a Góliát és az Altavizsla keresőknél nagyon hasonló mértékű eltérésről beszélhetünk egy tetszőlegesen választott minta átlagértéke és a populáció elméleti várható értéke között az első, a második és a harmadik húsz találatban. Tehát a mintaátlag hibája közel azonos nagyságrendűnek tekinthető ennél a két keresőnél.

4.10.4. A vizsgálat eredményeinek összefoglalása

Ebben a részben vizsgálatom eredményeit a korábban felállított hipotéziseim alapján tárgyalom.

4.10.4.1. *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 húsz találatban mind a négy keresőkérdésre. (Részletekkel kapcsolatban lásd a 4.10.1. alfejezet 3., 4., 5., 6. diagramjait.) Ezenkívül mindkét kereső megegyező indexet használt a visszakeresésre, amit a vizsgálat eredményei is megerősítettek, hiszen ugyanazokat a releváns dokumentumokat keresték vissza mind az öt témakörre vonatkozóan egyetlenegy weboldal kivételével. (Részletek megtalálhatók a Függelék 1., 2., 3., 4., 5. mellékleteiben.)

A vizsgálat során nyert adatokból kiderült, hogy nem fedeztem fel evidencia szintű összefüggést az első hatvan találatban lévő releváns találatok száma és a keresőkulcsokra kapott összes találat száma között. Ebből adódóan nem határozhattam meg meggyőző következtetést a releváns találatok számára vonatkozóan a találathalmaz hátralévő részében. Tehát ezt a hipotézist el kellett vetnem. (Részletekkel kapcsolatban lásd a 4.10.2. alfejezet 2. táblázatát.)

4.10.4.2. 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 szolgáltatssa az első húsz találatban mind a négy keresőkulcsra.

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. (Részletek megtalálhatók a 4.10.1. alfejezet 3., 4., 5., 6. diagramjainál.)

A hatékonyság szempontjából elemezve a keresőket azt állapíthatom meg, 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 keresőkulcsra, 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 keresőkulcsra, hiszen csekély különbség jelentkezett a releváns találataik átlagértékei között ezeken a vizsgált helyeken. (Részletekkel kapcsolatban lásd a 4.10.3. alfejezet 7. diagramját.)

4.10.4.3. A vizsgálat eredményeként megkapom azokat a magyar nyelvű keresőket, amelyek a pontosságot a legnagyobb, illetve a legkisebb mértékben valósítják meg a találatszolgáltatásukban.

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. (Részletek megtalálhatók a 4.10.1. alfejezet 3., 4., 5., 6. diagramjainál.) Amennyiben a keresők találatszolgáltatásának az egyenletes minőségére helyezem a hangsúlyt az első hatvan találatban, és nem csupán a visszakeresett releváns találataik mennyiségére fókuszálok ugyanezen a helyen, akkor ez utóbbtól eltérő eredményeket kapok. Megjegyzem, hogy egy kereső találatszolgáltatásának az

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

Az Origo-Vizsla kereső bizonyult a leghatékonyabbnak az első hatvan találatban mind a négy keresőkulcsra, 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 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 a négy keresőkulcsra.

Végül pedig az utolsó helyen osztozkodtak a Góliát és az Altavizsla keresők, azonban azok hatékonysága nem különbözött egymástól, mert a releváns találatok átlagértékei és a tapasztalati szórás értékeik közel azonosak voltak az első hatvan találatban a négy keresőkérdésre. (Részletekkel kapcsolatban lásd a 4.10.3. alfejezet 9. diagramját.)

4.10.4.4. 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. (Részletek megtalálhatók a 4.9.2. alfejezet 2. diagramjánál.)

Ehhez kapcsolódóan egy fontos mutatóra hívnám fel a figyelmet, mely szerint: a magyar nyelvű keresők átlagosan 5,6 releváns dokumentumot találtak meg a weben <csiperkegomba> témakörrel 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> témakörben, ami meglehetősen szerény teljesítménynek minősül.

Irodalom:

- [Albert-Locsmáncsi-Vasas95] Albert, L.–Locsmáncsi, Cs.–Vasas, G.: Ismerjük fel a gombákat! Gabó, 1995. 191 p.
- [BernersLee-Fielding-Frystyk96, RFC1945] Berners-Lee, T.–Fielding, R.–Frystyk, H.: Hypertext Transfer Protocol–HTTP/1.0. May 1996.
<http://www.rfc-editor.org> 2007.08.10.
- [Bocsán2007] Bocsán Róbert [Személyes közlés] 2007.09.04.
- [Brin-Page98] 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.
- [Heuréka] Heuréka tájékoztató segédlete
<http://www.heureka.hu/heureka/tippek.html> 2007.03.09.
- [Hódi2007] Hódi Imre [Személyes közlés] 2007.08.13.
- [Horváth-Sütheő2003] Horváth T.–Sütheő P.: A tartalmi feltárás. In.: Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés. Szerk. Horváth T.–Papp I. Bp.: Osiris, 2001.
- [Oppenheim-Morris-McKnight-Lowley2000] 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.
- [OrigoVizsla] Origo-Vizsla tájékoztató segédlete
http://www.vizsla24.hu/pages/sugo_tartalomjegyzek.html 2005.08. 11.
- [Ungváry2001] 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.
- [Taxaurusz2000] 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.
- [Zádori2007] Zádori Janka [Személyes közlés] 2007.08.10.

5. Összegzés és további kutatási feladatok

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

5.1. 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 küzdött meg ezzel a problémával 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. tézis: Az angol nyelvű keresőknél a csonkolás és a szinonimákra történő keresés hatékonyan működik az angolban, azonban az nem eredményes a magyar nyelvben.

2. tézis: 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 keresőkulcsokat 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.

5.2. 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.

5.3. 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 húsz találatban mind a négy keresőkérdésre. (Részletekkel kapcsolatban lásd a 4.10.1. alfejezet 3., 4., 5., 6. diagramjait.) Ezenkívül mindkét kereső megegyező indexet használt a visszakeresésre, amit a vizsgálat eredményei is megerősítettek, hiszen ugyanazokat a releváns dokumentumokat keresték vissza mind az öt témakörre vonatkozóan egyetlenegy weboldal kivételével. (Részletek megtalálhatók a Függelék 1-5. mellékleteiben.)

A vizsgálat során nyert adatokból kiderült, hogy nem fedeztem fel evidencia szintű összefüggést az első hatvan találatban lévő releváns találatok száma és a keresőkulcsokra kapott összes találat száma között. Ebből adódóan nem határozhattam meg meggyőző következtetést a releváns találatok számára vonatkozóan a találathalmaz hátralévő részében. Tehát ezt a hipotézist el kellett vetnem. (Részletekkel kapcsolatban lásd a 4.10.2. alfejezet 2. táblázatát.)

5.4. 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 szolgáltatssa az első húsz találatban mind a négy keresőkulcsra.

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. (Részletek megtalálhatók a 4.10.1. alfejezet 3., 4., 5., 6. diagramjainál.)

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 keresőkulcsra, 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 keresőkulcsra, hiszen csekély különbség jelentkezett a releváns találataik átlagértékei között ezeken a vizsgált helyeken. (Részletekkel kapcsolatban lásd a 4.10.3. alfejezet 7. diagramját.)

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

6. tézis: Továbbá azok a leghatékonyabbnak bizonyultak az első húsz találatban mind a négy keresőkulcsra, 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.

5.5. A vizsgálat eredményeként megkapom azokat a magyar nyelvű keresőket, amelyek a pontosságot a legnagyobb, illetve a legkisebb mértékben valósítják meg a találatsszolgáltatásukban.

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. (Részletek megtalálhatók a 4.10.1. alfejezet 3., 4., 5., 6. diagramjainál.) Amennyiben a keresők találatsszolgáltatásának az egyenletes minőségére helyezem a hangsúlyt az első hatvan találatban, és nem csupán a visszakeresett releváns találatok mennyiségére fókuszálok ugyanezen a helyen, akkor ez utóbbtól eltérő eredményeket kapok. Megjegyzem, hogy egy kereső találatsszolgáltatásának az egyenletes minőségét a hatékonyság szempontja fejezi ki számomra, amit a visszakeresett releváns találatok átlagértékei és tapasztalati szórás értékei figyelembevételével kapok meg.

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

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

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

Végül pedig az utolsó helyen osztozkodtak a Góliát és az Altavizsla keresők, azonban azok hatékonysága nem különbözött egymástól, mert a releváns találatok átlagértékei és a tapasztalati szórás értékeik közel azonosak voltak az első hatvan találatban mind a négy keresőkérdésre. (Részletekkel kapcsolatban lásd a 4.10.3. alfejezet 9. diagramját.)

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

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.

5.6. 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. (Részletek megtalálhatók a 4.9.2. alfejezet 2. diagramjánál.)

Figyelemreméltó mutató számomra, hogy a magyar nyelvű keresők átlagosan 5,6 releváns dokumentumot találtak meg a weben <csiperkegomba> témakörben, 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> témakörben, 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> témakörben.

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.

5.7. 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. Érdeemes 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.)

6. A study on the analysis of search engines' retrieval effectiveness

Summary and further research tasks

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

6.1. 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 conclude 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 draw the conclusion 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.

6.2. 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.

6.3. 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. (For further details see subchapter 4.10.1., diagrams 3, 4, 5, 6.) 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. (See details in appendices 1-5.)

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 search keys. 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. (For further details see subchapter 4.10.2., table 2.)

6.4. 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 relevant results in the first twenty hits for all the four search keys.

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. (See details in subchapter 4.10.1., diagrams 3, 4, 5, 6.)

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 search keys, 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 search keys, since a slight difference appeared between the means of their relevant results at these places. (For further details see subchapter 4.10.3., diagram 7.)

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 search keys, 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.

6.5. 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. (See details in subchapter 4.10.1., diagrams 3, 4, 5, 6.) 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 search keys, 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 mean 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 search keys.

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. (For further details see subchapter 4.10.3., diagram 9.)

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 search keys.

Efficiency here reflects for me how equally the analysed search engines can retrieve 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.

6.6. 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. (See details in subchapter 4.9.2., diagram 2.)

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.

6.7. 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).

7. Felhasznált irodalom

- [1] Albert, L.–Locsmáncsi, 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/tippe.html> 2007.03.09.

- [31] Hódi Imre [*Személyes közlés*] 2007.08.13.
- [32] Horváth T.–Sütheő P.: A tartalmi feltárás. In.: *Könyvtárosok kézikönyve. 2. 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, GKIE NET 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.
- [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.: *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.

8. 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 = Problemi ekonomiesno ta szosalnava rozvitku regionu i praktika naukovava ekszperimentu, Kijev-Uzgorod-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.

1. melléklet

„Csiperkegomba” – releváns találatok

1. Mellékletek I.: melléklet. FFV-24 számú ENSZ/EGB szabvány. (UN/ECE STANDARD FFV-24) Termesztett csiperkegomba. (Ajánlott nemzetközi kereskedelmi szabvány) 1. ENSZ/EGB szabványok friss gyümölcsre és zöldségre

... Ez a szabvány a friss állapotban forgalomba kerülő fogyasztásra szánt *Agaricus bisporus* termesztett fajtáinak termőtestére vonatkozik, és azokat a követelményeket határozza meg, amelyek az értékesítésre előkészített és csomagolt csiperkegombára vonatkoznak. Nem tárgya e szabványnak az ipari felhasználásra szánt csiperkegomba. ...

<http://www.omgk.hu/MGUT6/mel1.html> (Origo-Vizsla, Kurzor, Góliát, Altavizsla)

2. Dr. Rácz László: Talaj és termesztett csiperkegomba fő- és nyomelemmigrációjának vizsgálata. Doktori értekezés összefoglalója. Debreceni Egyetem Kémiai Tudományok Doktori Iskolája, Debrecen 1998.

... Vizsgáltam a zárt térben termesztett csiperkegomba fejlődését, illetve a termésidőszak alatt annak nyomelemtartalom-változását, amikor a termőtalajba, illetve az azt fedő felső rétegbe (ún. takarótalajba) ismert és változó mennyiségben juttattuk be a gomba fő elemeit (K, Na, Mg, Ca), az esszenciális nyomelemeket (Mn, Zn, Cu, Co, Ni), valamint a toxikus nyomelemeket (Cd, Cr, Hg, Pb). Az analitikai vizsgálatokat elsősorban azért tartjuk fontosnak, mert a környezetből (levegőből, vízből és talajból) egyre gyakrabban juthatnak mérgező nyomelemek ezen élelmiszerekkel szervezetünkbe ...

<http://www.doktorijegyzek.hu/jeg.php?subid=1&pid=228&back=3> (Heuréka, Góliát, Altavizsla)

3. Dr. Rácz László: Makro- és mikroelemek felszívódása a talajból és azok hatása néhány termesztett gombafajtára. Doktori (Ph.D.) értekezés tézisei.

... A kontroll- és fémsóoldatokkal kezelt táptalajon termelt csiperkegomba mintavétele nagyrészt a négy terméshullámban történt. Általában 3-3, illetve 5-5, egyes esetekben 6-6 párhuzamos termesztőzsákból megfelelő számú 2, 3, illetve 5 gombamintát szedtünk. Műanyag késsel szűrőpapíron negyedelve, esetleg nyolcadolva átlagmintavétel alapján 20 g körüli mennyiséget mértünk be analitikai mérlegen ...

<http://kemia.ektf.hu/racz.htm> (Heuréka, Kurzor)

4. Csiperkegomba: környezeti igénye, a termesztés feltételei, a termesztés módszerei. In: Zöldségtermesztők kézikönyve/ szerk. Balázs Sándor. Bp.: Mezőgazda Kiadó, 2004.

... Hő. A csiperkegomba hőigénye az egyes termesztési ciklusokban eltérő: más az átszövetési (1) időszakban, más a lappangás (2) idején és ismét más a termesztési (3) időszakban. Az átszövetési időszakában 22–25°C-ot igényel optimális fejlődéséhez. ... -

<http://www.hik.hu/index.asp?a=r&r=90%2C261&oid=57> (Kempelen Farkas Hallgatói Információs Központ, Kempelen Farkas Digitális Tankönyvtár honlapja) (Origo-Vizsla)

5. Kétspórás csiperke – *Agaricus bisporus*

... Kalap: 5-10 cm átmérőjű; fiatalon félgömb alakú, domború, majd ellaposodik, végül benyomottá válhat; fehéres, piszkos-, sötétbarna, széle világosabb; felületét barnás, lenyomott, sugarasan szálás pikkelyek díszítik; széle sokáig begöngyölt, gyapjas szálak, pelyhek lóghatnak le róla. Lemezek: sűrűn és szabadon állók, először rózsaszínűek, majd feketésbarnára sötétednek; éle fehéresen csipkézett. ...

<http://www.gombanet.hu/gomba.php?gid=186> (GombaNET.hu – a Magyar Mikológiai Társaság honlapja) (Heuréka, Origo-Vizsla)

6. Szarvas József: A bazídiomos gombák szaporodása, életciklusa II.: a kétspórás csiperke (*Agaricus bisporus*) = Magyar Gombahíradó, (2003)

... A kétspórás csiperke fejlődésmenetének ismerete nélkülözhetetlen a nemesítői munkához. A fejlődési ciklusa a legtöbb "bazídiomos nagygombához" képest eltérő. Szaporodásukat korábban heterotallikusnak vélték, azonban ma tudjuk, hogy másodlagosan homotallikus fajról van szó (Fritsche, 1991). A sejtjei sokmagvúak, nincs egyértelmű különbség a monospórás és a multispórás eredetű micélium között és mind a monospórás mind a multispór micélium képes termőtestképzésre anélkül, hogy más hifával anasztomizálnának. ...

<http://free.x3.hu/korona/other/bazid02.pdf> (Heuréka, Origo-Vizsla)

7. Kocsner Nóra: A gombaszúnyogok elleni védekezés entomopatogén nematodákkal. TDK dolgozat összefoglalója. Veszprémi Egyetem, Georgicon Mezőgazdaságtudományi Kar Növényvédelmi Állattani Tanszék, Keszthely 2004.

... A termesztett csiperke védelme nehézkes, mivel termesztéséhez sok szerves anyagot tartalmazó és megfelelően érlelt komposzt szükséges, amely nemcsak a csiperkegombának, hanem más patogén mikroorganizmusoknak is táptalaja, illetve életciklusát tekintve gyors lefutású szervezet, emiatt a vegyszeres védekezés a termesztés idején korlátozott. ...

<http://www.vein.hu/tdk/ITDK2004/AT/AT-KocsnerN.doc> (Origo-Vizsla, Kurzor)

8. Rovarpatogén fonálféreg fermentatív termelése, alkalmazása. Szerződésszám: OMF-01766/00 (BIO-012/00).

... Elsősorban a csiperkegomba (*Agaricus bisporus*) termesztésében a Diptera, Sciaridae taxonokba tartozó *Lycoriella solani*, L. *melii* fajok lárvái a termőtestek megrágásával, elpusztításával főként a tavaszi, nyári és őszi időszakokban nagy termés kiesést okoznak. Magyarországon a gombalegyek ellen leggyakrabban használt szer a Dimilin 25WP (Diflubenzuron). Korlátozott azoknak a vegyszereknek a száma, melyeket a Sciarid-lárvák ellen használhatunk, mivel a legyek egy része rezisztenssé válik az aktív hatóanyaggal szemben ...

<http://www.nkth.gov.hu/letolt/k+f/Biotech2000.doc> (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

9. Rózsai Gábor: A blokkos csiperkegomba termesztéstechnológiája. Kecskeméti Főiskola Kertészeti Kar. Oktatási segédanyag (MSPPT prezentáció)

... Csiperkegomba termesztés története Magyarországon: az 1800-as évek második felétől már nagyobb mennyiségben termesztünk. 1926-ban dr. Makó László: első magyar szintenyészetű gombacsíra. 1930-tól mind a csíragyártás, mind a gombatermesztés gyorsan fejlődik. 1930-ban 200.000 m²-en évi 600 tonna terem. A világon a harmadik helyen, Európában a második helyen áll Magyarország. ...

<http://iatt.ttk.pte.hu/> (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

10. Kétspórás csiperke. Familia: agaricaceae. Család: csiperkefélék

... Kalap: 4-10-(12) cm, gömb vagy félgömb alakból domború, ritkán kiterülő; széle sokáig begöngyölt, túlnő a lemezeken; felszíne változatos: fiatalon sima, később vagy sima, vagy kisebb-nagyobb koncentrikus pikkelyekkel; színe változatos, a fehértől a világosbarnáig vagy kakaóbarnáig. Lemezek: keskenyek, sűrűk, rózsaszínek, öregeen bíborbarnák, szabadok. ...

http://www.terra.hu/gomba/html/agaricus_bisporus.html (Origo-Vizsla, Kurzor, Góliát, Altavizsla)

11. Mészáros Gábor Sándor: Csiperkegomba zsákos termesztés-technológiájának értékelése, különös tekintettel a minőségbiztosításra. Szakdolgozat. Veszprémi Egyetem, Georgicon Mezőgazdaságtudományi Kar Növényteni és Növényélettani Tanszék, Keszthely 2001.

... A kétspórás csiperke a természetben leginkább trágyás, bolygatott talajon, korhadó növényi maradványokon, komposzton található meg. A lebontó szervezetek könnyen kiszorítják, és ezért nem kap életteret. Viszont ha valamilyen környezeti hatás folytán ezek a szervezetek visszaszorulnak vagy elpusztulnak, és életképes micélium kerül az adott táptalajba és a környezeti feltételek is adóttak, akkor elindul a micélium-fejlődés benne. ...

<http://www.georgikon.pate.hu/lib/diploma.htm> (Origo-Vizsla)

12. Horváth Balázs: A CO² koncentráció és a hőmérséklet termésmennyiségre gyakorolt hatásai a lappangási szakaszban, a csiperketermesztésben. Szakdolgozat. Veszprémi Egyetem, Georgicon Mezőgazdaságtudományi Kar Növényteni és Növényélettani Tanszék, Keszthely 2002.

... A túl korán megkezdett szellőztetés hatására kis gombák mélyen képződnek, és szinte feltúrják a takaróanyagot. A borsónagyság elérésekor a szellőztetés mértéke az addigi kb. harmadára csökkenthető. Ebben a szakaszban adódik utoljára lehetőség a takaróanyag megöntözésére, mivel a szedés előtti napokban kiadott víz hatására a gomba megnyúlik, és érzékenyebbé válik a mechanikai hatásokra, így szedés közben könnyen elszíneződik. ...

<http://www.georgikon.pate.hu/lib/diploma.htm> (Origo-Vizsla)

13. Gömöri Zsolt: Pótlólagosan adagolt dúsítóanyag szerepe a csiperketermesztésben. Szakdolgozat. Veszprémi Egyetem, Georgicon Mezőgazdaságtudományi Kar Növényteni és Növényélettani Tanszék, Keszthely 2003.

... A komposzthoz adagolt dúsítót, vagy a csírázással egy menetben, vagy a csírázástól számított 10-12. napon adagolják. Általában 0,3%-os formalinnal kezelt, granulált szójalisztet alkalmaznak 1 kg/100kg mennyiségben. Ilyen dúsítóanyag például a Millichamp 6000, amelynek külföldi és magyar kutatóintézetek vizsgálata alapján – mint a Horsti Gombatermesztési Kutatóintézet, vagy a Budapesti Növényegészségügyi Állomás – termésmenvelő hatása van. ...

<http://www.georgikon.pate.hu/lib/diploma.htm> (Origo-Vizsla)

14. Gombatermesztés az ókori Európában

... A termesztéssel foglalkozók némelyike valószínűleg felfigyelt a "termeszthető csiperke" tulajdonságára: emberi segédlet nélkül megkapaszkodó képességére a lótrágyán. Athénaiosz a Nikandroszt követő "gondolkodó" embereket említi. Közöttük minden bizonnyal akadtak olyanok, akik például a saját trágyagyűjtőhelyükön a spontán megjelent csiperkék gombafonalaktól "penészes" lótrágyáját fáik lombosára alá telepítették át. Lehettek, akik az utak mellett talált csiperketelepekből kapartak vagy ástak ki részeket, és vitték azokat haza. ...

<http://www.hhrf.org/uh/2002/uh13079.htm> (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

15. Geml József: Molekuláris filogenetikai vizsgálatok és termesztési kísérletek vadon termő csiperke (Agaricus) taxonokkal. Doktori disszertáció. Budapesti Corvinus Egyetem Kertészettudományi Doktori Iskolája, Budapest 2005.

... A kétspórás csiperke életciklusát széleskörűen tanulmányozták már az 1970-es években is (RAPER et al. 1972, ELLIOTT 1972). A legtöbb csiperkénél, mint a bazídiomos gombák nagy részénél általában, a termőtest bazídiiumain képződő négy bazídiospórába, a kariogámia és a meiózis után, egy-egy sejtmag kerül, így ezekből a spórákból kihajtó homokarionok csak az egyik párosodási típus genetikai hátterét hordozzák. ...

www.lib.uni-corvinus.hu/phd/geml_jozsef.pdf (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

16. Pallas Nagy Lexikona

... Csiperke vagy cseperke, csöpörke-gomba, pecserke, csiperke-galóca (*Agaricus campestris* L., vagy *Psalliota c.* Pers.; franc. champignon, l. a gombák képén), valamennyi ehető, kalapos azaz igazi gomba közt a leghasználatosabb. Földünk északi mérsékelt égövének nyirkos legelőin, kertekben, réten, erdőben és megtrágyázott gyepen májustól októberig igen közönséges. Tönkje a kalap középpontjával függ össze, 1,3-5 cm. hosszú, 0,6-2,5 cm. vastag, tömött. Spórája a kalap alsó színén levő, sugaras helyzetű lemezekon képződik. ...

<http://mek.oszk.hu/00000/00060/html/023/pc002380.html> (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

17. Vízhányó Tünde: Colour Analysis for Mushroom Disease Assessment. Doktori értekezés összefoglalója. Szent István Egyetem Élelmiszertudományok Doktori Iskolája, Gödöllő 2000.

... A csiperkegombát számos mikrobás támadás érheti, melyek hatására a gomba felületén a betegség barnás foltok, illetve szöveti elváltozások formájában jelentkezik és teszi a gombát értékesítésre alkalmatlanná. A kétféle eredetű barnulás (természetes és betegség okozta) megkülönböztetése volt a doktori munkám feladata. Az elkülönítéshez csak a színinformációt használtam, alapul véve a barnulások spektrális jellemzőit, illetve a képfeldolgozás kínálta lehetőségeket. ...

<http://www.doktorijegyzek.hu/jeg.php?subid=1&pid=383&back=3> (Góliát, Altavizsla)

18. A gombatermesztés fejlesztése a molekuláris biológia módszereivel. Szerződésszám: OMF-00352/02 (BIO-00046/01)

... Vadon termő *Agaricus bisporus* törzseket gyűjtünk be hazai, ill. külföldi génbankokból és hazai természeti környezetből. Ezekből szövettenyészeteket, ill. szaporítóanyagokat készítünk termesztési, nemesítési és molekuláris genetikai kísérletek céljából. Többek között az USA-ban (*Agaricus Resource Program*) és Franciaországban (INRA) begyűjtött törzsek igen értékes tulajdonságokat hordoznak. Magyarországon is célszerű egy ilyen génbank létrehozása a már meglévő (gén-)gyűjtemény kibővítésével. ...

<http://www.nkth.gov.hu/letolt/k+f/fitotechnologia2001.doc> (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

19. Szabó G., Rajkó R., Hodur C.: Csiperkegomba konvekciós és mikrohullámú szárításának összehasonlító vizsgálata. Szegedi Tudományegyetem, Szegedi Élelmiszeripari Főiskolai Kar Élelmiszeripari Műveletek és Környezettechnika Tanszék. Konferencia előadás. 4. Magyar Szárítási Szimpózium, Mosonmagyaróvár, 2001. október 18-19.

... Kutatásunk célkitűzése volt olyan kíméletes eljárás kidolgozása, amelynek eredménye új típusú élvezeti (gyorsan oldódó - instant és/vagy könnyű - snack-szerű) tulajdonságokkal rendelkező termék előállítására, ún. váltakozó rendszerű, konvektív és mikrohullámú szárítás kombinálását magába foglaló technológiával. Kritériumként tekintettük, hogy a termék a visszanedvesítés után az eredeti friss gombához – annak fogyaszthatósági tulajdonságait tekintve – a lehető legteljesebb mértékben hasonlítson. ...

<http://www.szef.u-szeged.hu/~rajko/pdfs/SzaboRajkoHodur4MSzSz.pdf> (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

20. Zárójelentés. OTKA nyilvántartási szám: F 030634.

... A kutatás elsődleges célja módszer kidolgozása volt csiperkegomba (*A. bisporus*) látható minőségi jellemzőinek számítógépes látórendszerrel történő mérésére, a módszert alkalmazva a jellemzők mérésére az egyes minőségi osztályokkal összerendelve. A feltárt összefüggések alapján a csiperkegomba minőségének becslése. A számítógépes képfeldolgozás segítségével automata ellenőrző robotok építhetőek, amelyek minimális felügyelettel és beavatkozással képesek a betegségeket korán azonosítani, valamint az automatizált betakarítás során a minőségi osztályokba sorolást elvégezni. ...

www.uni-corvinus.hu/~blaszlo/data/pdf/kutatas_otka_f030634.pdf (Releváns hiperhivatkozás, amit egyik magyar nyelvű kereső sem keresett vissza.)

21. Viczián Zsolt: Kertészeti-élelmiszeripari minőségbiztosítási rendszerek informatikai támogatása. Doktori értekezés összefoglalója. Budapesti Műszaki és Gazdaságtudományi Egyetem, Budapest 2000.

... A gombanövekedés folyamatának nyomon követésére olyan mérés technikát határoztam meg, melynek kivitelezése olcsó, ugyanakkor a növekedési folyamatra gyakorolt hatása elhanyagolható. A mérés technikát, mérési értékelést támogató szoftverek megtervezésével és elkészítésével egészítettem ki. A készített szoftverek alkalmasak a zsákon természetesen csiperkegomba felülnézeti fényképei alapján a kalapátmérő eloszlás automatikus, illetve félautomatikus meghatározására. ...

<http://www.chemonet.hu/hun/food/phd/bme/viczian.html> (Kurzor, Góliát, Altavizsla)

2. melléklet

„Evészavar” – releváns találatok

1. Dr. Túry Ferenc: Anorexia és bulimia: az evés zavarai a mindennapos gyakorlat szemszögéből. = Hippocrates, 4. évf. 2. sz. (2002)

... Az evészavarok két alapvető típusa az anorexia nervosa és a bulimia nervosa. Jellegzetesen pszichoszomatikus zavarok, amelyek kialakulásában biológiai, pszichológiai (egyéni és családi), valamint társadalmi-kulturális tényezők egyaránt helyet kapnak. Előtérbe kerülésüket a magas morbiditási mutatók és a magas anorexiás mortalitás indokolja. Gyakoriságuk fiatal nők között 1-4%, ezt hazai adatok is megerősítik. A tünetek gyakran rejtettek, különösen a bulimia esetén. Terápiájukban a gyógyszeres kezelésnek kisebb a szerepe ...

<http://www.medlist.com/HIPPOCRATES/IV/2/110.htm> (Heuréka, Origo-Vizsla)

2. Az elhízástól való túlzott félelem betegségnek számít

... A kóros soványság rendkívül veszélyes. Az éhezés miatt a testhőmérséklet, a pulzusszám és a vérnyomás alacsony szintre esik le, gyakran alakul ki szívritmuszavar, bőrszárazság, gyakori a hajhullás. A csontok a mézshiány miatt törékennyé válnak, különféle anyagcserezavarok lépnek fel, felborul a folyadék- és hormonháztartás, és emiatt elmarad a menstruáció. A fizikális leromlás gyakran jár máj- és veseműködési zavarokkal, fehérjevesztéssel. A betegség súlyosságát jelzi, hogy még a legjobb evészavar-klinikákon kezelt betegek csaknem 10 százaléka is tíz éven belül meghal vagy a nagyfokú testi leromlás, vagy öngyilkosság következtében ...

<http://www.origo.hu/noilapozo/egeszseg/20010910azelhizastol.html> (Heuréka, Origo-Vizsla, Kurzor)

3. Dr. Balázs János: Evési zavarok: enni vagy nem enni

... A bulimia lelki háttere kicsit egyszerűbb. Az a nő lehet rá hajlamos, akit gyermekkorában minden sírására, bajára etetéssel vigasztaltak. Így általános feszültségcsökkentő, önvigasztaló módszerként tanulja meg az evést, és később is evéssel reagálhat szorongásokra, rossz hangulatra, kudarcra, dühre, agresszivitásra. Itt az evés örömszerzés, a falánkság által okozott testsúlynövekedés hányással megelőzhető, és egyben feszültségoldó hatású. Mindezek azonban rohamszerűen zajlanak, sokszor szinte önkívületi állapotban ...

<http://hetilap.hetek.hu/index.php?cikk=34265> (Heuréka, Origo-Vizsla)

4. Szilágyi Gabriella: Lelki eredetű lesoványosodás

... A kutatók és a gyakorló orvosok, pszichológusok egyetértenek abban, hogy az anorexia nem szervi eredetű betegség. Habár bizonyos jelek kétségtelenül az éhség és a jóllakottság szabályozását végző, köztiagybeli hipotalamusz működési zavarára utalnak, ez közvetetten, lelki hatásokra következik be. Ebben nem kis szerepe van a társadalmi-kulturális környezetnek, például annak, hogy manapság a vékony testalkatú nő a szépségideál, s ebből a szempontból az étkezésre nagyon is oda kell figyelni. Emellett szól az a megfigyelés is, hogy azokban a fejlődő országokban, ahol ez a „nyugati típusú” szépségideál terjed, egyre több az anorexiában szenvedő nő ...

<http://www.sulinet.hu/eletestudomany/archiv/1997/9737/sovany/sovany.html> (Heuréka, Origo-Vizsla, Kurzor)

5. Bulimia nervosa

... A bulimia minden tápláltsági állapotban megjelenhet, általában normális testsúllyal jár. A bulimiás tünetek gyakran rejtettek maradnak. Az epizódokat szénhidrát-preferencia jellemzi, néha éjszaka jelentkeznek: ekkor félálomszerű, disszociatív állapot lehet jellemző. Az epizód akár naponta többször is felléphet, több ezer kalória elfogyasztásával járhat, rövid ideig tart. A

hányás során a gyomorsav a fogak belső felszínén zománceróziót okoz, ezért a rejtett zavar felismerésében a fogorvosok szerepe jelentős. A hányások vagy a vízajtózás miatti kálium- és egyéb ionvesztés életveszélyes szívritmuszavarok vagy görcsös roszullétek forrása lehet ...

<http://lefogyni.starfish.hu/index.php?f=lefogyni%2Feveszavarok%2Fbulimia> (Heuréka)

6. Anorexia nervosa

... Az anorexiások bőre száraz, lanugóval fedett, hajuk hullik, vérnyomásuk alacsony, vitális funkcióik lelassultak. Hormonzavarok, emésztési zavarok lépnek fel. A vitális funkciók lassulása ellenére mozgásos hiperaktivitás gyakori. A betegek személyiségére a perfekcionizmus, teljesítmény-orientáció, túlkontrolláltság, kényszeresség jellemző. Kerülik az érzelmeket, a szexualitást is. Személyiségük részben meg is változik, visszahúzódó, dacos, gyakran depressziós lesz a beteg, betegségbelátása nincs ...

<http://lefogyni.starfish.hu/index.php?f=lefogyni/eveszavarok/anorexia> (Heuréka, Origo-Vizsla)

7. Étkezési zavarok II.

... Nincs egységesen megfelelő eljárás az anorexia nervosa gyógyításában, mert az anorexiások különböző betegségi előtörténetekkel és tapasztalatokkal rendelkeznek. A kezelés első lépésében a biológiailag megfelelő testsúly visszanyerése a cél. Ez rendszerint kórházban történik, tapasztalattal rendelkező szakemberek felügyelete alatt. Ritkán intravénás vagy az orron át a gyomorba vezetett szondán keresztül táplálásra is szükség lehet. A kezelés következő szakaszában, ha a beteg tápláltsági állapota elfogadhatóvá válik, hosszú távú pszichoterápiás kezelésekre kerül sor, amelyet gyógyszeres kezelés is kiegészíthet ...

http://parabogyo.nolblog.hu/?post_id=9676 (Heuréka, Origo-Vizsla)

8. Lányok csapdában – az evés zavarai

... A bulimia nervosa-tól szenvedő az átlagosnál sokkalta nagyobb mennyiségű ételt töm magába, naponta többször is, hogy aztán a frissen elfogyasztott táplálék a lehető leggyorsabban távozzon a szervezetéből. Erre a célra hánytató szereket, hashajtókat használ, vagy egyszerűen ledugja az ujját a torkán. A betegség évekig maradhat rejtve még akár a közeli hozzátartozók előtt is, mert a beteg leginkább egyedül szeret enni, gyakran elrejtje az elfogyasztandó ennivalót, az alkalmas pillanatra várva ...

<http://www.bioitworld.hu/rovat/eletm/cikk.php?id=21739> (Origo-Vizsla, Góliát, Altavizsla)

9. Dr. Túry Ferenc: A depresszió és az evészavarok = Házi orvos Továbbképző Szemle 1. sz. (1996), p. 83-85.

... A depresszió tünetei olyan gyakoriak evészavarokban, hogy bulimia nervosában például a kórkép korábbi kritériumai között szerepelt is a depressziós hangulat és önvádolás a falási epizód után. Anorexia nervosában a malnutritio okozta általános testi gyengeség is hozzájárul a depressziós tünetekhez, bulimia nervosában a sok hányás miatti hypokalaemia lehet hasonló hatású. Az általában titkolt evészavar (ez főleg bulimia nervosára igaz) következményes izolációhoz vezet, ami szintén oka lehet a depressziós hangulatnak. A mai pszichiátria egyik kihívása éppen a rejtett zavarok felkutatása (ilyen a bulimia nervosa, a kényszerbetegség, a szociális fóbia stb.).

<http://sunrise.sote.hu/htsz/turey.htm> (Origo-Vizsla, Kurzor)

10. Kamaszfiúk diétaőrülete

... Az egyik elmélet szerint a fiúk evési problémái azért jelentkeznek, mert a kamasz így fedi el a férfivá válással kapcsolatos konfliktusait. A test gyorsan változik, a serdülő nem tud ezzel mit kezdeni, megpróbál tehát ott tenni valamit, ahol módja van rá. A kor pedig azt szuggerálja: légy olyan, mint egy atléta! Míg a lányok elsősorban a súlyukkal foglalkoznak, a fiúk az alakjukat és testméreteiket figyelik. Úgy vélik, ez a kulcsa a sikernek, nem utolsósorban a szerelem terén. A különféle evési problémák mögött mindkét nem esetében ugyanazok az okok fedezhetők fel: az

önértékelés zavarai, depresszió, félelem attól, hogy elvesztik a kontrollt (evésben), identitás-problémák ...

<http://www.patikamagazin.hu/index.php?cikk=4522> (Origo-Vizsla, Kurzor)

11. Új divatbetegségek: szégyellik bevallani a média ideálja a hajszálvékony nőtípus

... Sem a bulimia, sem az anorexia nem vezethető vissza egyetlen okra, csupán rizikófaktorok léteznek, amelyek megléte hajlamossá teszi az egyént a betegségre. Vagyis az evési zavar nem örökölhető. De ha valakinek fiatalon problémái voltak a családban, gyermekkorában fizikailag bántalmazták, esetleg szexuálisan zaklatták, akkor nagyobb az esélye, hogy testképzavara és önértékelési problémái lesznek. Elégedetlen lesz a testével, úgy érzi, nem elég csinos. Ehhez társul még a média által közvetített kép az eszményi nőről: az elmúlt évszázadok során a nőideál egyre vékonyabb lett, ez a folyamat az utóbbi évtizedekben még inkább felgyorsult ...

<http://www.tebolyda.hu/faliujsag/sajtofigyelo/200111/nszb200111.htm> (Origo-Vizsla, Kurzor)

12. Szumska Irena, dr. Túry Ferenc, dr. Jakabfi Péter: Az evészavarok prevenciója – kérdések és ellentmondások. = Lege Artis Medicinae, 13. évf. 1. sz. (2003), p. 24-28.

... Az intervenciók felében csak fiatal lányok és nők vettek részt. Ez egyrésztől ésszerűnek tűnhet, hiszen náluk az evészavarok tízszer gyakoribbak, mint a férfiak között. Ha viszont hatékony környezeti intervenciókra törekszünk, hibás felfogásnak tűnik, hogy kihagyjuk a férfiakat, mivel ők is részei annak a szociális környezetnek, amely alakítja és fenntartja az egészségtelen diétázási normákat. Néhány kutatás során azt találták, hogy a fiúk a lányoknál jóval negatívabb attitűdökkel rendelkeznek a túlsúlyos emberekkel szemben ...

<http://www.lam.hu/folyoiratok/lam/0301/5.htm> (Origo-Vizsla, Kurzor, Góliát, Altavizsla)

13. Kempf Zita: Szépség és szörnyeteg

... Az anorexia egy évszázada számít önálló betegségegységnek. Egészen a közelmúltig nem pszichiátriai, hanem belgyógyászati problémának tartották, ennek megfelelően áldozatai belgyógyászati vagy gyermekosztályokra kerültek. Később kezdték felismerni, hogy a rendellenességnek lelki okai vannak. Lehet, hogy egy drasztikus fogyókúrával kezdődik a folyamat, de aztán a diétázó elveszíti a kapcsolatot mind a realitással, mind az eredeti céljával. A fogyás önálló értéket képvisel, egyfajta teljesítménnyé válik ...

http://www.anyaklapja.hu/02_10/a1078.htm (Origo-Vizsla, Kurzor)

14. Anorexia és bulimia

... A bulimia nervosa szintén evészavar. Csakhogy ebben a helyzetben nem koplalásról van szó, sőt! Evésrohamok jellemzik ezt a fajta betegséget, méghozzá általában egy-két óra, vagy még kevesebb idő alatt elfogyasztott hatalmas mennyiségű étel. Ezután erős büntudat jelentkezik és az ételtől való azonnali megszabadulás kényszere, amit hánytatással, hashajtókkal, koplalással old meg a beteg. Ezek a falásrohamok hetente többször is előfordulhatnak. A bulimiások sokszor napi 15-20 hashajtót is beszednek ...

<http://www.starfish.hu/index.php?f=lefogyni/eveszavarok/anorexiabulimia&p=2> (Origo-Vizsla)

15. Túry Ferenc: A bulimia – modern szenvedély

... A diéta és a bizonytalanságérzés evészavart indíthat el más kockázati tényezővel is rendelkező egyéneknél. A diétázást elősegítheti mások megjegyzése is, vagy például az, hogy az egyén saját magát kövérnek érzi. A diétázás mindig megelőzi a bulimiát. Ha az anorexia és bulimia egyszerre fordul elő, az anorexia sokkal valószínűbben előzi meg a bulimiát. Ez is a diétázás fontos szerepére utal a bulimia kialakulásában. Az evés korlátozása fokozza az evési késztetést normális étvágy esetében: ez a diétázók dilemmája ...

http://www.behsci.sote.hu/hungarostudy2002/st_bulimia.htm (Origo-Vizsla, Góliát, Altavizsla)

16. Kun J. Viktória: Dacból nem evők: a népesség harmada táplálkozási zavarokkal küzd. = Népszabadság, 2004. szeptember 14.

... A páciensek nagy részének kettős tüneteinek vannak, az anorexiások felében bulimiás tünetek is kialakulnak, illetve a kezdetben bulimiás betegek jelentős részében anorexiás tünetek fejlődhetnek ki. Bármelyik csoportba tartozó páciensek naponta és órákon át végezhetnek testgyakorlatokat, gyakran öndestruktívak, szexuális érdeklődésük csökken, teljesítményük romlik, állandóan fáradtak, hangulatuk nyomott, deprimált - részletezte Túry Ferenc ...

<http://www.nol.hu/cikk/332673> (Origo-Vizsla)

17. Persányi Norina: Evési rendellenességek, testkép és média II.

... Az anorexiások célja a súlycsökkentés, sokszor annak ellenére, hogy feltűnően soványak, mégis félnek a súlynövekedéstől vagy akár az elhízástól. Testkép zavaruk miatt, önmagukat szinte mindig túlsúlyosnak látják, de érdekes módon ez a zavar csak saját testformájuk megítélésére vonatkozik, tehát mások testsúlyát reálisan mérik fel. Ez a testképük megítélésében jelentkező torzulás másban is megnyilvánul: önértékelésükre aránytalanul nagy hatást gyakorol a testsúly és az alak, vagy tagadják soványságuk súlyosságát ...

http://www.terasz.hu/main.php?id=enciklopedia&cikkid=2844&gyoker=6922&page=cikkek&P_HPSESSID=053f8ad1b298f29163c94f1cf246d7b6 (Origo-Vizsla, Góliát, Altavizsla)

18. Rathner, Günther dr., Túry Ferenc dr. és Szabó Pál dr.: Evészavarok a rendszerváltás utáni, az ún. keleti blokk országokban. = Orvosi Hetilap, 142. évf. 25. sz. (2001), p. 1331-1333.

... 1988-1989-ben, mielőtt a nagy társadalmi változásokat előre láthattuk volna, közös kutatást végeztünk Magyarország, az NDK és Ausztria közreműködésével, melyben összehasonlítottuk a táplálkozási magatartásokat és az evészavarok prevalenciáját. A vizsgálat célja az volt, hogy elemezze a két különböző társadalmi rendszer evészavar-morbiditására gyakorolt hatását. Adataink egyedülállóak és nagy jelentőséggel bírnak a kulturális változás társadalmi és pszichológiai vonatkozásainak megértésében. Mind a kelet-, mind a nyugat-európai diákok testsúly adatai nagyon hasonlóak, a kívánt súly pedig szinte teljesen megegyezik mindhárom országban. Az evészavarok prevalenciáját a "szimulált" DSM-III-R diagnózis (azaz interjú nélküli kérdőíves felmérés) alapján állapítottuk meg ...

<http://vitalitas.hu/olvasosarok/online/oh/2001/25/4.htm> (Origo-Vizsla, Kurzor)

19. Barbie középkorúan is hódítani akar

... Az 1980-as évek elején kezdték alkalmazni evészavarok esetére is az úgynevezett viselkedés- és kognitív terápiát, melynek célja a hibás viselkedés - az evészavar - korrigálása. A viselkedésterápiában igen fontos a páciens tüneteinek és a tünetekhez tartozó helyzetek részletes elemzése. Ebbe beletartozik a beteg motivációjának elemzése is, annak feltárása, hogy mit szeretne elérni kóros viselkedésével. Továbbá felméri, mi várható, ha megszűnik az evészavar. A módszerek közé tartozik a szorongás oldása, valamint a különböző önkontroll-gyakorlatok ...

http://www.hazipatika.com/articles/Barbie_kozepkoruan_is_hoditani_akar (Kurzor)

20. Dr. Fűtő László: Jómódú társadalmak betegsége: az anorexia

... Az anorexia elnevezés, vagyis az étvágytalanság nem felel meg a valóságnak, hiszen ezeknek a betegeknek az étvágya tulajdonképpen mindig is megmarad. Ettől függetlenül az anorexiások súlyosan lesoványodnak. A megbetegedés tulajdonképpen étkezési zavarként fogható fel. Már a klinikai tünetek megjelenése előtt feltűnő lehet, hogy az érintettek viselkedése megváltozik. Jellemző rájuk a nagyfokú pedánság, aprólékoság, precizitás; rendkívül intelligensek, az iskolában igen jó tanulók, önmaguk számára magas teljesítménymércét állítanak fel ...

<http://www.origo.hu/noilapozo/egeszseg/20020521anorexia.html> (Kurzor)

21. A gyerek tünete, a család betegsége

... Az evészavaros fiatalok édesanyja gyakran túlságosan óvó, domináns személyiség, aki gyermeke életének minden területét befolyásolni szeretné. A családtagok sokszor nem eléggé önállóak, mintha lelki életük egy közös térben zajlana, bármelyiküket éri valami baj vagy váratlan esemény, mindenkit egyformán felzaklat, mindenki egyformán izgalomba jön. Ezeknek a megfigyeléseknek a nyomán a szakemberek először azt gondolták, leghelyesebb a beteg fiataalt mihamarabb kiragadni e megbetegítő környezetből, és egyedül kezelni őt ...

<http://www.sulinet.hu/tart/fcikk/Kfed/0/11632/1> (Kurzor)

22. Róth I. Zsófia: Kortünet az evészavar

... A kóros falánkságot (bulimia) sokszor a fogorvos veszi észre először, mert a sok hányás következtében a gyomorsav a fogzománcot és az ínyt is kikezdi. A bulimiás – hasonlóan az anorexiához – igyekszik a saját maga által "belőtt" súlyt elérni, de akaraterő hiányában, elveszítve a kontrollt, időszakonként elképesztő zabálásba fog, majd közvetlenül ezután a mellékhelyiségben hánytatja magát. A fogazat romlása mellett a magas vízvesztés következtében kiszáradás, különböző vese- és húgyúti fertőzések, szívritmuszavarok léphetnek fel. Gyakoribb, mint az anorexia, de könnyebben gyógyítható ...

<http://hetilap.hetek.hu/index.php?cikk=9121> (Góliát, Altavizsla)

23. Persányi Norina: Evési rendellenességek, testkép és média I.

... Az evészavarra vonatkozóan több pszichológiai magyarázat is létezik. Az analitikus megközelítés szerint akiket gyerekkorukban súlyos családi vagy egyéb lelki megrázkódtatás, trauma ért, például gyermekkorában fizikailag bántalmazták, esetleg szexuálisan zaklatták, azoknál nagyobb az esély a testképzavar és az ehhez kapcsolódó önértékelési problémák kialakulására. Sokan a negatív gyerekkori tapasztalatok miatt elégedetlenek a testükkel, vagy akár úgy tagadják szexualitásukat, hogy kórosan lefogynak -megszabadulva ezzel nőies formájuktól ...

<http://www.terasz.hu/main.php3?id=enciklopedia&gyoker=6922&cikkid=2811&page=cikkek>

(Góliát, Altavizsla)

24. Az evészavarok – az anorexia nervosa és a bulimia nervosa I. (Távoktatási tananyag)

... A testkép a leggyakrabban és a legkifejezettebben evészavarokban változik meg. Hilde Bruch a testkép perceptuális és konceptuális zavaraként definiálta az AN-t. Szinte mindegyik betegségosztályozási rendszer tartalmaz olyan kritériumot, amely vagy közvetlenül utal a testképzavarra, vagy többé-kevésbé szoros kapcsolatban áll vele. A BN-t önálló betegségként leíró Russell (1979) gyakran idézett közleményében úgy mutatja be a BN-s beteget, mint akinél a testsúllyal és a testméretekkel való törődés túlértékelt eszmévé válik, aki veszélyesnek és utálatosnak tartja a kövérséget ...

<http://www.moktav.hu/index.php?p=/tan/book-36/content> (Góliát, Altavizsla)

3. melléklet

„Kegyhely” – releváns találatok

1. Péliföldszentkereszt kegyhely és zarándokház

... Kegytemplom: A műemlék barokk templomot 1735-ben Eszterházy Imre hercegprímás építtette. Ő nyilvánította hivatalosan is Péliföldszentkeresztet búcsújáráhellynek. A templomnak kitűnő akusztikája van. A főoltár a szentségi trónusig vörös márványból készült 1763 körül. Az oltár előlapján a jellegzetes embléma: a kereszt négy mezőjében INNR (Jesus Nasareus Nasarenorum Rex-Jézus, a nazarénusok királya). A főoltár építménye és a templomban található összes szobor fából készült. A főoltár legmagasabb pontján az Atya, alatta a Szentlélek és a Keresztrefeszített erőteljes megjelenítése ...

<http://www.szentkereszt.try.hu> (Heuréka, Góliát, Altavizsla)

2. Fatima

... 1916. május 13-tól, 1917. október 13.-ig, először három alkalommal angyal-jelenés, majd hat alkalommal a Boldogságos Szűzanya megjelenése nyomán alakult világméretű kegyhellyé, az eredetileg meglehetősen kietlen portugáliai 'Cova d'Iria (az Iriai völgy). Három egyszerű kis pásztorgyereknek, Santos Lucia-nak, (10 éves), s két unokatestvérének, Marto Ferenc (9 éves) és Jácintának (7 éves) megjelent a béke angyala és imára buzdította őket, majd mint Portugália őrzőangyala az áldozatra buzdította őket. A Szent Szűz, először 1917. május 13.-án jelent meg a 'Béke Völgyében' és a rózsafüzér imádkozását kérte a gyermekektől, és azt, hogy 6 hónapon át ugyanezen időben jöjjenek el hozzá ...

<http://www.parochia.hu/Segedanyagok/Segedanyagok/fatima.htm> (Heuréka, Kurzor)

3. A (máriabesnyői) kegyhely története

... A kegyhely híre egyre terjedt, sok ima-meghallgatás történt, amit az oltár háta-mögött látható sok kis ezüst szív (offer - fogadalmi tárgy) is mutat. A 20. században Máriabesnyő az ország második leglátogatottabb Mária-kegyhelyévé vált. A kapucinus rend 1912-ben, majd 1942-ben felújította a templomot, csak 1950-ben, a szerzetesrendek feloszlásakor voltak kénytelenek elhagyni azt. 1989. augusztus 15-én térhettek vissza a nagyon megfogyatkozott, és elöregedett kapucinusok. Máriabesnyőn kívül Budán a Fő utcában, Mórton és Tatán telepedtek meg az erősen lepusztult egykori épületeikbe ...

http://www.mariabesnyo.hu/historia_kegyhely.php (Heuréka, Origo-Vizsla, Kurzor)

4. A búcsúszentlászlói templom története

... A ferences szerzetesek 1694-ben jöttek ide, amikor Széchenyi Pál, a veszprémi püspök rájuk bízta az addigra már közismertté vált zarándokhelyet, „amely helyet a nép Szentlászlónak nevezte”. „Számos csodát” említ a püspök a megbízó levelében, amely csodák Szűz Mária közbenjárására történtek. Ettől kezdve a hívő emberek mind a nagyobb Mária-ünnepeken, mind pedig Szent László király ünnepén nagy számban érkeztek a kegyhelyre, ahol a ferences barátok fogadták őket ...

<http://www.bucsuszentlaszlo.koznet.hu/> (Heuréka, Origo-Vizsla)

5. Az andocsi kegytemplom

... A már évek óta ismert és nevezetes kegyhely plébániatemplomát Padányi Bíró Márton veszprémi megyéspüspök szentelte fel. Ettől az időponttól kezdve a kegyszobrot felöltöztetik. A Mária-szobor első palástját Széchenyi Katalin grófnő ajándékozta, példáját azóta többen követték. A Mária-szobornak a világ minden tájáról több mint 240 ruhája van. Egy-egy palást két

hétig van a kegyszobron, színük a liturgia színeihez alkalmazkodik. A Mária-szobrot péntek délelőtt öltöztetik fel az andocsi asszonyok ...

<http://kincsestar.radio.hu/harangok/dhandocs.php> (Heuréka, Kurzor, Góliát, Altavizsla)

6. Lourdes – Bernadette Soubirous látomásai

... Március 25-én, Gyümölcsoltó Boldogasszony napján Bernadette kérdésére a megjelenő Úrnő így mutatkozott be: "Én vagyok a Szeplőtelen Fogantatás!" A Szent Szűz kérését teljesítve 1876-ban pompás háromszintű templomot építettek a jelenés helyére, amely azóta is világhírű Mária-kegyhely: évente átlagosan több mint 5 millió zarándok látogatja, 400.000 zarándok fürdik meg évente a gyógyító forrás vizében, s mintegy 3 millió gyertyát égetnek el a barlang előtt. A lourdes-i látnok egész életét az egyháznak szentelte: 1866. július 7-én lépett be a nevers-i zárda kapuján, s ott is hunyt el 1879. március 16-án ...

<http://merengo.hu/forum/?t=358> (Heuréka, Origo-Vizsla)

7. A katolikus egyház Celldömölkön

... Idővel a templom zarándokhely lett, 1446-ban az egyházi hatóság a dömölki Mária-szobrot csodatevő szobornak ismerte el. Ezeknek az adatoknak Eszterházy Pál nádor és Lancsics Bonifác dömölki jószágkormányzó voltak feljegyzői, melyek igazolják: már a török dúlás és a mohácsi vész előtt búcsújáróhely volt a dömölki apátsági templom. "Itt ünnepezt a nép a mikor Nagy Lajos uralma hitet és hazát mentett, amikor Hunyadi szentolvasós karja megállította az ozmán hadakat." Majd, amikor az ország nagy részében a török lett az úr, a földönfutó magyarokkal együtt elmenekült Dömölkéről a Szűzanya szobra is ...

<http://www.celldomolk.hu/start.php?nyelv=magyar&menu=92&poz=o> (Heuréka, Origo-Vizsla)

8. A Csíksomlyói búcsú - a pünkösdi búcsú eredetéről

... Az 1661-es tatár-török pusztítás alkalmával a kegyszobrot nem sikerült elrejtetni, amit bizonyít a következő legendás történet: a tatár pusztítók behatolva a kolostor falai mögé, felgyújtották a templomot és a vele egy földél alatt lévő kolostort. A kolostor udvarába menekült népet halomra gyilkolták. A templomban elrejtőzött nép a kegyszoborhoz menekült, és könyörgött a szent Szűzhez, hogy védje meg őket. A tatár vezér látta, hogyan bújnak a kegyszoborhoz az emberek, gondolta, hogy ez a szobor valami nagy kincs lehet. Elhatározta elszállítását. De amikor szekerre tetette, négy, hat, de nyolc ökör sem tudta megmozdítani a szekeret. ...

<http://www.consultationmagazin.hu/index.php?menu=cikk&id=371> (Heuréka, Góliát, Altavizsla)

9. A (csíksomlyói) kegyszobor

... A templom legértékesebb tárgya a Szűz Mária kegyszobra. A XVI. század elején (1510-1515) reneszánsz stílusban készült, hársfából. Alkotója ismeretlen. Magassága 2,27 m. A világon ismert kegyszobrok között a legnagyobb. A napba öltözött Asszonyt ábrázolja, akinek lába alatt van a hold, fején a tizenkét csillagból álló koszorú. Királynőként is ábrázolja Szűz Máriát: fején korona, jobb kezében jogar, bal karján tartja szent Fiát, a világ Megváltóját. 1798-ban, Batthányi Ignác erdélyi püspök idején az egyház "Csodatevőnek, segítő szent Szűznek" nevezi el ...

http://www.csiksomlyo.ro/index.php?option=com_staticxt&Itemid=46&xt_item=1&staticfile=hu/kegyszobor.html (Origo-Vizsla, Góliát, Altavizsla)

10. A máriapócsi kegyhely

... A pócsi kegyképet 1676-ban Csigri László máriapócsi bíró készíttette fogadalomból és hálából, a török rabságából történt szabadulásának emlékére. A képet Papp Dániel helybéli görög katolikus lelkész öccse, Papp István festette 6 magyar forintért. Ezt a bíró nem tudta kifizetni, így az ikont egy tehetősebb polgár, Hurta Lőrinc vásárolta meg és ajándékozta a templomnak.

Valószínűleg ő írta a kép aljára orosz nyelven a következő szöveget: “Én, Isten szolgája állítottam fel a képet vétkeim bocsánatára.” ...

<http://www.atanaz.hu/mp/mptort.htm> (Origo-Vizsla, Heuréka)

11. A máriapócsi kegyhely

... Századunk elején harmadszor is bebizonyította könnyezésével az Istenszülő, hogy Máriapócsot választotta kegyhelyéül. 1905. december 3-án Gávris Kelemen bazilita atya, a kegykép őre zarándokokat vezetett a templomba. Amikor az ikont közelről megmutatta nekik, észrevette, hogy a Szűzanya arca a szokottnál sötétebb, jobb szeméből pedig könnycsatorna húzódik le az arcon, és egy könnycseppben végződik. A könnyezés december 19-éig folyamatosan tartott, majd december utolsó két napján is, összesen 18 napig. A könnyezés tényét egyházi és világi vegyesbizottság vizsgálta, s megállapították a könnyezés valódiságát ...

<http://www.c3.hu/~klio/klio993/KLIO9339.html> (Origo-Vizsla, Kurzor, Góliát, Altavizsla)

12. Máriavölgy kegyhely rövid története

... Mint ismeretes, Nagy Lajos király (1342-82) nagy Mária-tisztelő volt. Ő is ellátogatott a csodás gyógyulások helyére, és 1377-ben az egyetlen - magyar alapítású szerzetesrendnek, a pálosoknak adományozta a kegyhelyet, majd kolostort és templomot építtetett itt számukra. (Néhány forrás a kolostor temploma építésének időpontját 1471-re helyezi, s Rozgonyi Lászlót jelöli meg építtetőként.) A pálosok - akik megkapták a királytól a kegyhely melletti falut is, ahol a mai napig szlovák lakosság él - egészen 1786-ig éltek itt, amikor is II. József feloszlatta a szerzetesrendeket, így nekik is el kellett hagyniuk a kies völgyet ...

<http://www.catholic.ro/?to=box&cat=zarandokhely&id=9> (Origo-Vizsla, Kurzor, Heuréka)

13. Történelmi kegyhelyeink: Bába – a Vérző Ostya kegyhelye

... A zarándokokat a kegyhely Vérző Ostyája vonzotta ide. IV. Eugén (Jenő) pápának 1434. január 21-én kelt, búcsút engedélyező levele "Krisztus Urunknak az Oltáriszentségből kiömlő csodálatos vérét" említi Bába rendkívüli értékeként. A pápa Zsigmond király kérésére adta meg a búcsú kiváltságokat. (Zsigmond király 1433-ban járt Rómában, ekkor koronázták német-római császárrá, és ez alkalommal valószínűleg Hunyadi János is ott volt kíséretében.) Petrus Ransanus - 1475 és 1478 között a nápolyi király követe Mátyás király udvarában - Magyarország nevezetességeiről írt művében Bábát is megemlíti ...

http://web.axelero.hu/kesz/jel/01_11/szemle3.htm (Origo-Vizsla, Kurzor)

14. Bornár Emese: Kőbe vésett üzenetek. = Új Szó, 2005. 03. 17.

... Gróf Szapáry Béláné 1881-ben rendkívül mély benyomásokkal tért haza Pozsonyba az első magyar lourdes-i zarándoklatról. 1892-ben arra az elhatározásra jutott, hogy Pozsonyban kifaragtatja a lourdes-i barlang mását. A kegyhely ma is létezik a Kálvária-hegyen álló katolikus templom alatt. A kis völgy falait mára szinte teljesen beborították a kegytáblák. Az első 1892-ből származik, a legújabbak 2004-ből valók. Nagyon sok hívő ember jár oda, hiszen a „lourdes-i” barlang máig Pozsony zarándokhelyeként él a köztudatban. ...

http://www.ujso.com/clanok_tlac.asp?cl=119958 (Origo-Vizsla, Kurzor)

15. Engesztelésül Anglia megtéréséért: a nemzeti kegyhely és egy bátor hölgy története. = Új Ember, LVIII. évf. 30. sz. (2002), p. 2817.

... Charlotte Boyd (1837-1906) sírján, a londoni Kensal Green városrész katolikus temetőjében egy kőtábla található a következő felirattal: Ő szerezte vissza a walsinghami Papucsos Kápolna katolikus tulajdonjogát. Anglia római katolikus nemzeti kegyhelyének helyreállítója a legnagyobb tiszteletet érdemli meg. Charlotte 1837. március 21-én született Kínában. Szülei brit hajótulajdonosok voltak. Charlotte-ot az anglikán egyház tagjaként keresztelték meg. Jómódban nőtt fel ...

<http://ujember.katolikus.hu/Archivum/2002.07.28/0901.html> (Origo-Vizsla, Heuréka, Góliát, Altavizsla)

16. Helytörténet: Péliföldszentkereszt

... Péliföldszentkereszt első lelkészei remeték. Amikor azonban 1763-ban a szegényes remetelak megüresedett, Barkóczy Ferenc prímás a szigorúságukról híres, mezítláb és hajadonfőn járó nazarénusokat telepítette az erdőkoszorúta szép völgybe. Noha csak rövid ideig működhettek Péliföldszentkeresztben - az 1765-ki törvénycikk a Szentszék jóváhagyásával nem bíró rendeket, II. József 1782-ben minden remeteszervezetet eltörölt - bűnbánó, vezeklő életmódjuknak nagy része volt benne, hogy a kegyhely az egész környék lelki oázisává lett ...

<http://www.ehf.hu/~ej/szt/hely.htm> (Origo-Vizsla, Kurzor, Góliát, Altavizsla)

17. A régi és az új máriabesnyői kegyhely. = Új Ember, LXI. évf. 13-14. sz. (2005), p. 2954-2955.

... A felsőtemplomban helyezték el a loretoi Mária-szobor fekete cédrusfából faragott másolatát, amelyet a kapucinus szerzetesek gyalogosan hoztak el Loretóból. Az ország legkisebb kegyszobra (tizenegy centiméter magas, négy centiméter széles) csodás megtalálásának híre gyorsan elterjedt, és Máriabesnyő az ország egyik leghíresebb zarándokhelyévé vált. Ma is évente több százezer zarándok keresi fel, kérve a Szent Szűz anyai pártfogását. Az első időktől kezdve nagyon sok imameghallgatás és csoda történt itt. Erről tanúskodnak a kegyoltár környékén elhelyezett kis hálaajándékok. ...

<http://ujember.katolikus.hu/Archivum/2005.03.27/2302.html> (Kurzor, Origo-Vizsla)

18. Történelmi kegyhelyeink: Vác-Hétkápolna

... A törökök kiűzése után, 1690-ben, ennek a forrásnak közelében helyezte el egy váci asztalosmester fogadalmi ajándékát - Szűz Máriát a kis Jézussal ábrázoló képét - hálából azért, hogy ő is és családja is meggyógyult egy súlyos betegségből. A "Kuti Képhez" hamarosan megindult a búcsújárás. A kép előtti imameghallgatások és csodák hírére egyre többen keresték föl a helyet. A nagyarányú zarándoklatok hatására építtette 1714-ben gróf Kollonich Zsigmond váci püspök a mai kápolna elődjét, és ezt Migazzi Kristóf püspök, későbbi bécsi érsek a XVIII. század közepén kibővítette. A templomot Mária nevére szentelték föl, és a kegyképet főoltárára helyezték ...

http://web.axelero.hu/kesz/jel/01_09/szemle3.htm (Kurzor, Góliát, Altavizsla, Heuréka)

19. Római Katolikus Egyházközség - Celldömölk

... A kegykápolnában lévő főoltár fölött van a Boldogságos Szűz csodatevő szobra, Mária karján a kisdéd Jézus. A kápolna homlokzatán a búcsújáróhely nagy jótevőinek, gróf Erdődy Györgynek és nejének, herceg Eszterházy Teréziának címere látható. A kápolnát kupola fedi, melynek közepén kis torony emelkedik ki. A homlokzat két sarkán pedig harsonás angyalok térdelnek. A kegykápolna előtt két hatalmas barokk stílusú angyalszobor látható. A kegykápolna mögött áll Szent István-oltára. Oltárképe a XVIII. századból való, mely felső felén Szent Henriket és Szent Istvánt, alsó részén pedig Szent Istvánnak Boldog Gizellával történő eljegyzését ábrázolja ...

<http://kis-mariacell.uw.hu/> (Kurzor, Origo-Vizsla, Góliát, Altavizsla)

20. Mátraverebély története

... A kegyhely első fénykora a 14-15. sz.-ra tehető. A török hódoltság alatt a zarándoklatok ritkultak. 1700-ban XI. Kelemen pápa a csodás gyógyulásokat kivizsgáltatta, s elfogadta azok hitelességét. Az első kőkápolna a forrás szomszédságában 1705-ben épült fel. 1710-ben a pápa már ehhez a kápolnához kötötte a búcsú teljes elnyerését. A mai kegytemplomot 1758-63 között emelték. A templomba az első kápolnából átkerült a Szűzanya kegyszobra, kezében a királyi jogarral, karján a kis Jézussal, aki bal kezében az országalmát tartja ...

http://www.matrainfo.hu/telep_verebely.php (Kurzor, Origo-Vizsla)

21. Mátraverebély-Szentkút, a Palócföld leglátogatottabb Mária kegyhelye
... Egyébként Vereb (Egyházasvereb, ma Mátraverebély) az 1220-as évektől szerepel oklevelekben, templomának az apostoli Szentszéktől már 1258-ban búcsú kiváltsága volt! 1331-ben Vereb községben már pap is volt, s a templomba özönlő zarándokok az erdőben lévő szent kutat is felkeresték. A verebélyi és szentkúti kegyhely fénykora a 14-15. század fordulóján volt. Vereb Péter 1370 körül kolostort is létesített (a templom mellett, a mai temető helyén), ahová fehér ruhás szerzeteseket minden bizonnyal pálosokat telepített. IX. Bonifác pápa 1400. november 9-én kelt oklevelében Nagyboldogasszony ünnepével kapcsolatban tizenkét gyóntató pap alkalmazását engedélyezte a verebi plébánosnak ...
<http://www.nordtour.hu/1-373.html> (Kurzor, Origo-Vizsla, Góliát, Altavizsla)

22. Udvarhelyi Olivér: Loretói gondolatok
... “A loretói Szent Ház a Szűzanya első igazi nemzetközi kegyhelye, évszázadok óta a keresztény Mária-kultusz lüktető szíve...” – olvashatjuk abban a levélben, melyet II. János Pál pápa a kegyhely 700 éves jubileumára írt. A Magyar Nagylexikon⁷ Loreto szócikkének szerzője szerint a Szent Ház legendájának nincs történelmi alapja, bár hozzáteszi, hogy a kegyhelyen már 1196-ban templom állt. Az ősrégi hagyomány szerint a loretói szentélyben valóban Szűz Mária názáreti házáat találjuk meg. Ezt a tényt megerősítik az újabb történeti, régészeti, művészettörténeti kutatások eredményei is ...
http://web.axelero.hu/kesz/jel/03_05/loretoi.html (Heuréka, Góliát, Altavizsla)

23. A magyar Máriacell. = Új Ember, V. évf. 5. sz. (2004)
... A kegytemplomnak három homlokzati tornya van. Két tömzsi barokk, és közte egy égbe törő kilencven méteres gótikus. Ez ösit a bazilika főhajójával egyetemben Nagy Lajos királyunk építtette 1360 táján, fogadalma teljesítéseképpen, mert a celli Boldogasszony megsegítette a túlerőben lévő törökkel szemben. Máriacell immár hat évszázada Közép-Európa legkedveltebb búcsújáró helye. A 14. század óta szinte folyamatosan érkeznek ide más népekhez hasonlóan a magyarok is, hogy örömeiket és fájdalmukat, kéréseiket és hálájukat a celli Szűzanyával megosszák, gondjaira bízzák. A kegyhely kicsiny, 12. századból való kegyszobra egyszerre Magna Domina Hungarorum (Magyarok Nagyasszonya) ...
http://magazin.ujember.katolikus.hu/Archivum/2004.05/09_p.html (Heuréka, Góliát, Altavizsla)

24. Jövők búcsút jární ... = Új Ember, LVI. évf. 24. sz. (2000), p. 2707.
... A mostani barokk oltár helyén 1746-ig gótikus stílusú állt. Egy év múlva szentelte fel a templomot Padányi Bíró Márton püspök. Az andocsi kápolna a hódoltság idején ismert búcsújáróhely volt, a ferencesek megjelenésével a vidék lelki központja lett: Mária-ünnepeken, Szent Ferenc és Antal ünnepén ezrével keresték fel a hívek. Századunk elején a Jézus Szíveteletnek is méltó keretet adtak, az egyik mellékoltár ennek a tanúsítója. Aztán persze Trianon: a ferences rendtartomány nagyszombati klerikus és malockai novíciusháza Csehszlovákiához került ...
<http://ujember.katolikus.hu/Archivum/000611/0901.html> (Heuréka, Góliát, Altavizsla)

4. melléklet

„Rákóczi-szabadságharc” – releváns találatok

1. Fördös Annamária: A török kiűzése és a Rákóczi szabadságharc

... A szabadságharc első szakasza: Esze Tamás és Kis Albert vezetésével újra fegyverkezni kezd a parasztság, ez lesz a Tiszaháti felkelés. A két vezető elmegy Rákóczihoz és megkérlik, hogy álljon a seregek élére. Rákóczi boldogan vállalja a feladatot, és követekkel elküldi a breznai kiáltványt, és egy zászlót, felirattal: "Cum Deo pro patria et libertate" (Istennel a hazáért és a szabadságért). Rákóczi azonban átlátja, hogy csak a nemességgel együtt lehetséges a győzelem. Rákóczi híre miatt egyre többen csatlakoznak a felkeléshez (6000 fő összesen). Elfoglalják, a későbbi központi vármegyéket, Szatmárt, Bereget, Ugocsát. A nemesek azt hiszik, hogy a szervezkedés ellenük folyik ...

<http://www.palya.hu/dolgozat/dolgozat.cfm?id=1557> (Origo-Vizsla)

2. A Rákóczi-szabadságharc és Európa. = História, 3. sz. (2003)

... 1703. május 6. Rákóczit a breznai várban kuruc felkelők keresik fel és kérik, álljon a Habsburg-ellenes felkelés élére. Rákóczi kiáltványban szólítja fel Magyarországon nemes és nemtelen lakóit: fogjanak fegyvert hazájuk szabadságáért a Habsburgok ellen. Esze Tamás ezerekapitányt megbízza, gyűjtsön sereget. 1703. május I. Péter orosz cár visszafoglalja a svédektől a Finn-öböl vidékét, megalapítja Szentpétervárt (1712-től lesz birodalma központja). 1703. június 16. Rákóczi Naménynál magyar földre lép, és a már korábban kirobbant népi felkelés élére áll. 1703. november XIV. Lajos francia király tudatja Rákóczival, hogy az erdélyi trón jogos örökösének ismeri el, és havi pénzügyi támogatást biztosít számára. ...

<http://www.historia.hu/archivum/2003/0303kronol.htm> (Origo-Vizsla, Kurzor)

3. Érettségi tételek: a Rákóczi szabadságharc okai és menete

... Katonai események részletezése: 1703. május: Breznai - Esze Tamás zászlóadás II. Rákóczi Ferencnek. Tarpai - Tiszaháti felkelés. Dolha - Károlyi Sándor leveri azt. 1703. június: II. Rákóczi Ferenc hazatérése és a Tiszántúli elfoglalása. A Duna-Tisza közének visszafoglalása (október). Vetési pátens: - a harcoló jobbágyok mentesek a földesúri szolgáltatás alól, az otthon maradtak kötelesek teljesíteni azt. 1704. II. Rákóczi Ferenc kiáltványa Európa népeihez. 1705. Bottyán János vezetésével: Dunántúli offenzíva. 1706. őszi hadjárat: Erdély felszabadítása szilézia, morva, magyar konföderáció ...

<http://bel.freeweb.hu/erettsegi/som.htm> (Origo-Vizsla, Kurzor)

4. Szomolányi Gábor: Adópolitika és pénzügypolitika a Rákóczi szabadságharc idején

... Rákóczi a kurucok kezére került pénzverésben Körmöcbányán és Nagybányán már a szabadságharc elején beindíttatja a pénzverést. A megnövekedett pénzverési igények miatt Kassán és Munkácsra is megszervezik a pénzverést. Meg kell említeni, hogy Rákóczi részére mint erdélyi fejedelemnek Kolozsváron is vertek pénzeket aranyból, melyeknek a pénzforgalomban nem volt jelentőségük. A Rákóczi szabadságharc alatt vert magyar típusú nemesfémpénzek három fő típusát ismerjük: az aranyforintokat, az ezüst-polturákat és az ezüstforintokat. ...

<http://jam.nyirbone.hu/konyvtar/web-konyv/nuiz-konf/Szomola.html> (Origo-Vizsla, Góliát, Altavizsla)

5. A Rákóczi-szabadságharc kitörésének 300. évfordulója. Tóth Ferenc: Bujdosó vagy száműzött? A Rákóczi-emigráció a történelem tükrében. = Magyar Tudomány, 6. sz. (2003)

... A 19. század romantikus, nemzeti függetlenségi lelkesedéstől áthatott Rákóczi-kultusza a Rodostóban elhunyt fejedelem hamvainak hazahozatala idején érte el tetőpontját, és vezette be örökre II. Rákóczi Ferencet a legnevesebb magyar uralkodók és nemzeti hősök előkelő Panteonjába. Az idilli képet egy addig kevésbé ismert történész, Szekfű Gyula kritikusabb hangvételű munkájának megjelenése zavarta meg. A száműzött Rákóczi 1913-as megjelenése fordulópontot jelentett a magyar történetírás menetében. Addig feltáratlan levéltári források alapján Szekfű rámutatott az emigrációban élő Rákóczi tevékenységének árnyoldalaira is, és kiemelte személyiségének gyengeségeit. ...

<http://www.matud.iif.hu/03jun/toth.html> (Origo-Vizsla)

6. Barta János, ifj.: Nyugat-Európa és Rákóczi. = História, 3. sz. (2003)

... A szabadságharc záró szakaszában Rákóczi a pápai udvarba is küldött követet, a bécsi udvarral jó kapcsolatokat tartó XI. Kelemtől azonban érdemi segítséget vagy közbenjárást nem kapott. A pápa a magyar katolikusokat a császár iránti hűségre való visszatérésre biztatta. A többi európai államtól pedig Rákóczi még ennyi támogatást sem remélhetett. Anglia tartózkodását az is kifejezhette, hogy amikor a száműzetésbe induló fejedelem hajója Hullban kikötött, az angol hatóságok nem engedték partra szállni. Hollandia követe pedig a béketárgyaláson nyilatkozott úgy, hogy más kérdés a protestantizmus védelme és más a politika - azaz Erdély ügye -, amit nem kívántak támogatni. ...

<http://www.historia.hu/archivum/2003/0303barta.htm> (Kurzor, Góliát, Altavizsla)

7. Történelem tételek: II. Rákóczi Ferenc

...1705-ben a Szécsényben tartott országgyűlésen az új államformát rendi konföderációban jelölték meg. II. Rákóczi Ferencet Magyarország vezérlő fejedelmévé választották, s megfogalmazták a Habsburgokkal való kiegyezés feltételeit is. A fejedelem mellett 28 tagú szenátus működött. Hatásköre csak belpolitikai kérdésekre terjedt ki. I. József császár (1705-1711) hajlandó volt tárgyalásokba bocsátkozni, de főként időnyerés céljából. 1707-ben az ónodi országgyűlésen sikerült Rákóczinak anyagi támogatást szerezni a harcokhoz. Megszavazták a Habsburg-ház trónfosztását és a szabadságharc tartalmára a nemesi adózást ...

<http://k2.jozsef.kando.hu/~guczi/TORI/rakoczi.htm> (Heuréka, Kurzor)

8. A Rákóczi-szabadságharc kitörésének 300. évfordulója. Köpeczi Béla: II. Rákóczi Ferenc külpolitikája. = Magyar Tudomány, 6. sz. (2003)

... Rákóczi két célt tűzött ki diplomáciája számára. Az egyik az volt, hogy visszaállítva a régi rendet Magyarország független állam legyen, amely elszakad a Habsburg-háztól és új királyt választ, feltehetőleg Miksa Emanuel, bajor választó személyében. Ez szolgálta az ónodi országgyűlés trónfosztása, amelyet nem követett királyválasztás. Amikor 1704-ben Rákóczit erdélyi fejedelmévé választották, a maga számára új célt tűzött ki: az erdélyi önállóságot kívánta megvalósítani. A nagyszombati béketárgyalásokon ezt a követelést a császár elutasította, bár a fejedelem letett esküjére és az erdélyiek kívánságára hivatkozva végig kiállt mellette, a bujdosás idején is. ...

<http://www.matud.iif.hu/03jun/kopeczi.html> (Origo-Vizsla, Kurzor)

9. A Rákóczi-szabadságharc kitörésének 300. évfordulója. R. Várkonyi Ágnes: "AD PERPETUAM REI MEMORIAM" - Rákóczi államáról. = Magyar Tudomány, 6. sz. (2003)

... Montmejan abbé, a varsói lazarista rendház főnöke 1701-1702 folyamán bizalmas segítőjük volt, Rákóczi már mint államfő azt írta neki, hogy minden segítséget nélkülözve azért ragaszkodott az eredeti tervéhez, mert együtt érzett a meggyötört néppel, az országot polgárháború fenyegette. "Úgy kellett tehát tennem, ahogy tettem, azért, hogy megindíthassam

ezt a nagy ügyet. Megvettem azokat a veszélyeket, amelyek fenyegettek, amikor az alsó nép egy maroknyi csapata élére álltam, pénz nélkül, sereg nélkül, a szükséges tisztek nélkül, akik megteremthették volna az alapot és regularizálhatták volna a hadakat...

<http://www.matud.iif.hu/03jun/varkonyi.html> (Góliát, Altavizsla)

10. Asbóth Miklós: Kalocsa és környéke a Rákóczi szabadságharc idején: az ordasi menlevél és az imsósi hadihíd

... Rákóczi délvidéki hadjárata és az egyesült angol és birodalmi német csapatok július eleji győzelme a franciák felett lélegzetvételhez juttatta a bécsi udvart, elmúlt Dunántúl elvesztésének közvetlen veszélye. Többé nem volt szükség Széchényi Pál amúgy is eredménytelen béketárgyalási kísérleteire. A háború folyt tovább. 1705-ben Bercsényi Miklós felvidéki sikerei és Károlyi Sándor eredménytelen dunántúli hadjárata után Rákóczi legfontosabb feladatának tekintette Dunántúl visszaszerzését. Károlyi kudarca után Dunántúlon csak Dunakömlőd közelében az ú.n. kömlődi hídfő maradt kuruc kézen. (A március végi (Balaton) kiliti vesztes csata fáradalmait Károlyi Kalocsán pihente ki.) ...

<http://ordas.freeweb.hu/knint/neplap/kn2004/KN040528.html> (Heuréka, Kurzor)

11. Rákóczi Szegeden

... A kuruc hadak először 1704 májusában jelentek meg Szegeden, elfoglalták a Palánkot, de fegyverek híján a várat a császáriak kezében kellett hagyniuk. Rákóczi Ferenc 1704. július 20-án érkezett ide seregével. A vár ostromát azonban augusztus 12-én ő is föladta, mert megbetegedett és fejedelemmé választása miatt el kellett távoznia. A szegedi vár "erős falakkal, jó karban levő tornyokkal, árokkal és fedett úttal volt ellátva. Különben a parancsnok (Globitz János Frigyes báró) sem látszott oly féle embernek, aki 400 emberével 50 középszerű bombám durranására megadja magát"- írja II. Rákóczi Ferenc Emlékiratai a magyar hadjáratról 1703-1711 című művében. ...

<http://www.sk-szeged.hu/kiallitas/rakoczi/szeged.html> (Origo-Vizsla, Góliát, Altavizsla)

12. Kovács Sándor: Kárpátalja – a kuruc szabadságharc bölcsője

... A szabadságharc kibontakozása idején, 1703. szeptember 17-én a nagybereznai ruszin Beca Iván az Ung völgyében toborzott „gubás hada” elfoglalta Ungvár városát, és félévi ostrom után, 1704. március 16-án a várat is bevette. Így vált Ungvár és a vár a szabadságharc egyik legfontosabb támaszpontjává. A fejedelem 1707-ben az év nagy részét az ungvári várban töltötte. Intézte a kuruc szabadságharc diplomáciai ügyeit, fogadta a külföldi követeket, a hadvezéreit. Ungvár várának a kuruc szabadságharcban betöltött szerepére utal az a tény is, hogy csupán Munkács vára volt képes tovább ellenállni a császáriaknak. ...

http://web.axelero.hu/kesz/jel/03_06/karpatalja.html (Kurzor, Góliát, Altavizsla,)

13. Első fejezet: a békekötés. Rákóczi Lengyelországban. In: Asztalos Miklós: II. Rákóczi Ferenc és kora. Bp.: Dante, [1934]

... Károlyi ekkor már nem igen várt Rákóczitól instrukciót, hanem ment a maga útján. Holott Rákóczi a február 18-iki salánki levelében, mielőtt Lengyelországba kiment, az alkudozások kikapcsolásával bízta a felkelés ügyének vezetését Károlyira. Károlyi ugyan még ezekben a napokban is így írt Rákóczinak: "abban pedig bizonyos lehet Felsőged, hogy egyszer adott resolútióm szerént particularis békességre nem lépek s Felsőgedtől el nem szakadok"... de már jelezte azt is, Bercsényire célozva, hogy "viszont particularis emberekért egész nemzetünket elveszteni s reménység fejében az bizonyost elbocsátani képes-e Felsőged, bölcs itiletire hagyom. ...

http://www.mek.iif.hu/porta/szint/tarsad/tortenei/mo_1526/rakoczi/html/04.htm

(Heuréka, Kurzor)

14. Vörös Béla: A Rákóczi szabadságharc tábornokai, vitézei, hősei=Bolyai Szemle, XII. évf. 4. sz. (2003)

... Károlyi Sándor tábornagy, gróf: Ősi arisztokrata családból származik. 1689-ben lett Szatmár megye főispánja Ebből adódóan szembeszállt a bujdosókkal, Dolhánál egy csatában meg is verte őket. Később 1703-ban jelentkezett Rákóczinál, aki generális főstrázsamesternek tette meg. Egy évvel később altábornagy. Sikeresen vezetett hadjáratot a Felvidéken, Dunántúlon, de a császári túlerővel szemben később nem tudta megtartani Dunántúlt. Harcolt még Tiszántúlon, Erdélyben. Katonai összeomlás előtt kapcsolatba lépett Pálffy Jánossal, a császári csapatok akkori parancsnokával, és a vele kötött békében kiharcolta, hogy a Rákóczi szabadságharc résztvevői szabadon elvonulhassanak, és büntetlenséget kapjanak....

<http://www.bjkmf.hu/bszemle2003/egyeb120401.html> (Origo-Vizsla, Kurzor)

15. Niderhauser Emil: Rákóczi és Kelet-Európa. = História, 3. sz. (2003)

... Az orosz cár egy ideig nem lépett, pedig tudni lehetett, hogy az ő jelöltjének komoly esélyei lennének Szaniszlóval szemben. Péter számára sem volt érdektelen a kérdés. Az északi háború kellős közepén mindenképpen Franciaországgal szeretett volna szövetségre lépni, vagy legalábbis valamelyest kapcsolatot teremteni, ami biztonnal növelte volna nyugati hírnevét. És tudni lehetett, hogy Rákóczinak van valamiféle kapcsolata XIV. Lajos francia királlyal. [...] Péter felajánlotta Rákóczinak a lengyel trónt. Varsóban szerződést is kötött Rákóczival (1707. szeptember 14.). [...] A varsói szerződés számos pontot tartalmazott.

<http://www.historia.hu/archivum/2003/0303nieder.htm> (Heuréka, Kurzor)

16. Történelem tételek: a Rákóczi szabadságharc 2.

... Rákóczi gazdaságpolitikája a szabadságharc szolgálatában állt. Az élelmezés megkönnyítésére Rákóczi tárházakat állított fel. A textilipar fejlesztését manufaktúrák alapításával kellett megerősíteni. A fejedelem ösztönözte a vashámor-tulajdonos polgárokat, támogatta a kézműveseket, a mestereket fölmentette a katonáskodás alól. Modern tisztviselő szervezetet alakított ki. Terveket dolgoztatott ki az ipar és a kereskedelem átfogó fejlesztésére. A legnagyobb gondot mégis a pénzügyi fedezet hiánya okozta. A bányák, vámok, fejedelmi birtokok jövedelme kevésnek bizonyult. A francia segélypénz pedig külföldi fegyvervásárlásra kellett. ...

<http://x3.hu/freeweb/frameset.x3?user=/eagle&page=/tort.htm> (Heuréka, Kurzor)

17. Levél az örök igazsághoz. In: II. Rákóczi Ferenc: Emlékiratok: részletek. [Szentendre]: Interpopulart, 1993.

... Így ilyen szerencsétlenül kezdődött a magyarországi háború, amelybe - önként bevallom - az okosság minden szabálya ellenére fogtam. Csak egy fiatalember hevesége és a haza szeretete lelkesített. Még visszavonulhattam volna, és volt is rá nagy okom. De bátorított és erősített az a szándék, hogy megérdemeljem a nép bizalmát és szeretetét, és kitűzött célomban szilárdan bízva elküldtem Kálnássy Istvánt Wiśniowiecki herceghez és Potocki kijevei palatínushoz, hogy sürgesse a várt segítséget. Elhatároztam tehát, hogy folytatom utamat, s a szétszórt népet újra összegyűjtve, a lengyel határon rejtőzve várom meg a segélycsapatok érkezését, és nem hagyom kialudni a nép szívében lángra kapott tüzet. ...

<http://www.mek.ro/01000/01019/01019.htm> (Heuréka, Kurzor)

18. A Rákóczi-szabadságharc

... A felkelő jobbágysághoz a nemesség is csatlakozott Rákóczi zászlaja alatt. A nemzeti összefogás gyors sikereket hozott. Hamarosan az egész ország a kurucok birtokába került. A nemesség és a jobbágyság érdekeinek egyeztetése Rákóczi érdeme volt. Vetési táborában kiadta a „jobbágyságról való” rendeletét. Ebben a fegyvert fogott parasztokat és családjukat mentesítette a földesúri terhek alól. Rákóczi hadserege főleg jobbágyokból, végvári vitézekből, Thököly

kurucokból állt. Ezek ún. mezei (nem állandó) hadsereget alkottak. Később egészült ki reguláris (állandó, képzett, egységesen felszerelt) hadsereggel – Rákóczi udvari ezredeivel. ...
<http://www.netmozaik.hu/tortenelem/rakoczi.html> (Altavizsla, Góliát)

19. Rákóczi-szabadságharc. In.: Magyar Virtuális Enciklopédia

... Rákóczi-szabadságharc (1703-1711) - A török kiűzése után a bécsi kormány intézkedései súlyosan sértették a magyar társadalom különböző rétegeit. 1703-ban az elégedetlen parasztok, a szélnek eresztett végvári katonák, valamint Thököly bujdosó kurucok a Tiszaháton felkeltek a Habsburg kormányzat ellen, s mozgalmuk vezetésére hazahívták a Lengyelországba menekült II. Rákóczi Ferencet. A felkelőkhez egyre nagyobb számban csatlakoztak nemesek, sőt arisztokraták is. ...

http://www.enc.hu/lenciklopedia/fogalmi/torttud_magy/rakoczi-szabadsagharc.htm (Origo-vizsla, Heuréka)

20. Rákóczi-szabadságharc. In.: Wikipédia

... Az ónodi országgyűlés: A kialakult helyzet megoldására Rákóczi országgyűlést hirdetett 1707. május 1-jére, az Ónod határában lévő mezőre. Az országgyűlésnek három lényeges programpontja volt: a gazdasági nehézségek leküzdése, a hadsereg és államszervezet megerősítése és – egyelőre titokban tartott programként – a Habsburg-ház trónfosztása. A május 31-én megnyitott, Rákóczi jelenlétében tartott országgyűlésen június 6-án véres inzultusba fajuló vita kezdődött. A Túróc megyei követek – a megyei nemesség soraiból szerveződő békepárt szószólói – a rézpénzzel kapcsolatos vita során önző érdekek követésével gyanúsították meg a fejedelmet. ...

<http://hu.wikipedia.org/wiki/R%C3%A1k%C3%B3czi-szabads%C3%A1gharc> (Origo-Vizsla, Kurzor)

21. (Hatvan története) a Rákóczi-szabadságharc idején (1703-1711.)

... II. Rákóczi Ferenc maga is többször megfordult Hatvanban, így például 1704 áprilisában. Bercsényi Miklós pedig. 1706 szeptemberében innen követte figyelemmel Rabutin Kassa felé vonuló seregét. A hatvaniak közül számosan beálltak a kuruc seregbe. Közülük elsőként lehet említeni Szabó Mátét, aki korábban részt vett a török ellenes felszabadító harcokban, a császár szolgálatában. A kuruc seregben már 1703-ban lovas ezreskapitány lett. 1705-ben a fejedelem kinevezte a hatvani vár főkapitányának. Ugyancsak a kuruc hadseregben szolgált a fia, Szabó András vicekapitányként, de ő hősi halált halt. ...

<http://www.hatvan.hu/varos/tortenelem/szabharc6.htm> (Heuréka)

22. A Rákóczi-szabadságharc

Az első katonai sikereket a Tiszántúl elfoglalásával aratták. A sikerek hatására a nemesség egy része és a hajdúvárosok Rákóczi hűségére esküdtek. Közöttük volt a Bécsben mellőzött Károlyi Sándor báró is. 1703. októberére a Duna-Tisza köze és a Felvidék jelentős része Rákóczi kezébe került. XIV. Lajos elismerte Rákóczit az erdélyi trón jogos örökösének, harcát havi tízezer tallérral támogatta. Decemberben a kurucok bevonultak Lőcsére. Károlyi Sándor átkelt a Morván, és rohammal bevette a Bécs védelmére épült sáncokat.

<http://bdeg.sopron.hu/~imo/tetel2/Rakoczi.htm> (Origo-Vizsla, Altavizsla)

23. Takáts László: A Rákóczi-szabadságharc egészségügye

... Az ütközet színhelyéül, még vesztes csata után is, igyekeztek sebesültjeiket elszállítani. Rákóczi Bagossy Pál brigadéroszt megdicsérte, mert ... Sebeseit és Szekereit elhozta, avval én előttem vitézségnek nagyobb jelét mutatta, mint hogy személyében hús németet ölt volna, mert ... Olyan szoros állapotban ... egész bátorságát mutatta. Dul Mihály azt jelenti, hogy 1703 augusztusában a Várad körüli harcokból ... „csak Debreczenbe is vittek tizenkét szekér

sebeseket ...”. - Nem sokkal előbb szenvedett két ízben is vereséget Bóné az egyesült szegedi-váradai rákocitól, nagy valószínűséggel innen származik a Wienerisches Diarium értesülése ...

<http://www.neumann->

[faz.hu/muvek/tudomanytortenet/3_Orvostudomany_tortenete/Rakoczi/rakoczi_01.pdf](http://www.neumann-haz.hu/muvek/tudomanytortenet/3_Orvostudomany_tortenete/Rakoczi/rakoczi_01.pdf) (Origo-Vizsla, Heuréka)

24. A Rákóczi szabadságharc érmészete. In.: Huszár Lajos: A kuruc-kor érmészete, Rákóczi emlékkönyv

... Rákóczi aranyainak pénzlábra vonatkozólag két egykorú adatunk van. Az egyik Hellenbach levele a fejedelemhez 1707. aug. 9-én, melyben hangsúlyozza, hogy a márkához 23 karát 10 gramm színarany s két gramm réz vétetik a egy Thaly által közölt adat szerint ebből 70 db került kiverésre, tehát egy db súlya ezek szerint 3.50 g. Hellenbachnak ez a levele azzal kapcsolatban íródott meg, hogy Nagybányán valamivel több rezet kevertek a színaranyhoz, s félt esetleges pénzverési zavarok keletkezésétől. Egy másik adat szerint 309 márka pénzfémből 21630 db arany került kiverésre, ami megint csak a fenti eredményt hozza ki, tehát meg lehet állapítani, hogy Rákóczi aranyforintjai 3. 50 g súlyban kerültek kiverésre. ...

<http://www.numismatics.hu/cikkek/rakoczi1.htm> (Origo-Vizsla, Kurzor)

25. Érettségi tételek: a Rákóczi-szabadságharc társadalmi, katonai és külpolitikai története

... 1704-ben Rákóczit az erdélyi rendek fejedelemmé választották. A kuruc seregek legénységét főleg a nem nemesi rétegek alkották. A tisztikar viszont a nemesekből került ki. Rákóczi óriási erőfeszítéseket tett arra, hogy képezze seregét, hogy fel tudják venni a harcot a császári csapatok ellen. Központosító, abszolutizmus vonásait hordozó államépítő elképzeléseit a kuruc oldalra állt mágnásokkal szemben is igyekezett megvalósítani. Udvari Tanácsát széleslátókörű, művelt nemesekből állította össze, akik az államigazgatás ágai szerint szakosodva készítették elő a fejedelem döntéseit. ...

<http://bel.freeweb.hu/e3/a15.html> (Origo-Vizsla)

26. A kuruc-kor nyomdászata. In.: Fitz József: A magyar könyv története 1711-ig. Bp.: Magyar Helikon, 1959.

... Debrecen is 1703-ban, július végén került kuruc kézre. Egy héttel később már közkézen forgott Rákóczi ott készült, díszes kiállítású oltalomlevele. A nyomda vezetője, Vincze György, jól értette a szakmáját. Ugyanez évben kiadta Rákóczi Imádságának első kiadását, 6 levélnyi magyar és latin szöveggel, csinos magyar motívumú záróvignettával. A következő évben pedig a Recrudescunt-manifestumot, s 1705-ben a kuruc hadirendszabást. Meg kellett értük lakolnia. Október végén Herbeville nagy labanc serege tört rá a városra. A lakosság menekült, vele Vincze is, kinek azóta nyoma veszett. A város jegyzője "Promemoria"-ban siratta meg az elpusztított főiskolai könyvtárt és a városi nyomdát. ...

<http://www.mek.iif.hu/porta/szint/tarsad/muvtort/fitz/html/fitz31.htm> (Origo-Vizsla, Kurzor)

27. A Rákóczi-szabadságharc kitörésének 300. évfordulója. Bánkúti Imre: Rákóczi államának gazdasági problémái. = Magyar Tudomány, 6. sz. (2003)

... A sárospataki országgyűlés (1708. december) kisebb, de így is irreális mértékű adót rótt ki a hadélelmezésen felül: a vezetők tiszteletdíjával 800 ezer forintot és 12 ezer hajdú kiállítását. 1710-ben pedig már csak a Gazdasági Tanács vetett ki 100 ezer forintot, és a konföderáció vezető tisztviselői és főtisztjei egyénileg ajánlottak fel 215 150 forintot, a megyék pedig a nemesség részéről 10 ezer, a parasztság részéről 31 ezer forintot ígértek. Rákóczi ekkor már csak annyi pénzt remélt, amivel az idegen zsoldosokat kifizetheti. (De az adó teljesítése szinte teljesen marhával történt.) ...

<http://www.matud.iif.hu/03jun/bankuti.html> (Heuréka)

28. Várkonyi Ágnes: Közép-Európa II. Rákóczi Ferenc politikájában. = Magyar Szemle, Új folyam XII. évf. 11-12. sz. (2003)

... Rákóczi lengyel-svéd-porosz-magyar konföderációs tervének alapgondolatát a fejedelmi instrukció 3. pontja foglalja magában: "A mostani háborúban akár a francia akár a Habsburg győz, felborul az európai hatalmi egyensúly, és a győztes elnyomja majd a németalföldieket, meg a német birodalmi fejedelmeket. ... Magyarország ahogy a múltban igen erős fékezője volt mind a Porta, mind az ausztriai ház törekvéseinek, a jövőben is bástya lenne hatalmi törekvések ellenében, ugyanakkor biztos sarkköve, az európai hatalmi egyensúlynak. A Manifestum három kulcsszó: a hatalmi egyensúly, a konföderáció és a nagyhatalmak törekvései. A hatalmi egyensúly teóriája iránti érzékenység és érdeklődés a magyar politikai kultúrában az ország egységének, önállóságának és békéjének igényére vezethető vissza. ...

http://www.magarszemle.hu/archivum/12_11-12/varkonyi.html (Kurzor, Altavizsla, Góliát)

5. melléklet

„Szökőár” – releváns találatok

1. Havas Péter: Szökőár és pedagógia

... Száz turista életét mentette meg a dél-ázsiai katasztrófa pillanataiban egy tízéves brit diáklány, aki nem sokkal korábban tanulta földrajzórán, hogy melyek a tengeri szökőár fenyegető előjelei. Az újévi első számában The Sun brit napilap beszámolója szerint Tilly Smith szüleiivel töltötte a karácsonyi szünetet a thaiföldi Phuket közelében, a tengerparton, amikor arra lett figyelmes, hogy az óceán kavarni kezd, majd hirtelen messze visszahúzódik a partvonalától. Azonnal eszébe jutott, amit a szökőárról és annak előjeleiről földrajztanára két héttel a szünet előtt tanított, és szólt szüleinek, hogy szerinte a tenger perceken belül el fogja árasztani a partot ...

http://korlanc.ngo.hu/szokoar_es_ped.html (Heuréka)

2. Bennszülött törzsek a szökőár fő áldozatai

... Szakértők szerint a kis szigetek őslakosai már eddig is kevesen voltak, és a halálos hullám veszteségeinek felmérése után úgy tűnik, hamarosan végleg kihalnak. Szakértők szerint a külvilágtól elszigetelt törzsek megmentésére indított expedíciók eleve reménytelenek. Antropológusok leginkább a Sentinel és Nikobár szigetek vadászó-gyűjtögető, félnomad törzseit féltik a végső megsemmisüléstől. A törzsek eddigi elszigeteltségéhez az is hozzájárul, hogy a szigetcsoporton több indiai katonai támaszpont is található, ahová külföldieket nem engednek ...

<http://www.tte.hu/?page=hirek&id=1227> (Heuréka)

3. Szökőár: 30 százalékra emelkedhet a munkanélküliség

... Közvetlenül a szökőár sújtotta megyékben a NMSZ szerint a munkanélküliség a megelőző 9,2 százalékról 20 százalékra emelkedhet. E két országban az érintett emberek 50-60 százaléka az év végére találhat magának megint valamilyen megélhetési forrást, és a "munkahelyek 85 százalékát két éven belül újjáteremthetik" - olvasható a NMSZ jelentésében. A szervezet azt javasolja, hogy a "munkahelyek létrehozása az újjáépítési és az emberességi alapon nyújtott segítség szerves része legyen" ...

<http://www.hrportal.hu/index.phtml?page=article&id=34267> (Heuréka)

4. A szökőár bűvárszemmél

... A Kínában élő, de francia állampolgárságú Stephanie Adams a szerencsésebbek közé tartozott, mivel vele "mindössze" annyi történt, hogy 20 méteres mélységben búvárkodva a látótávolság egyik pillanatról a másikra nullára csökkent, és így elszakadt a többi búvártól, végül azonban biztonságosan a felszínre tudott emelkedni. Egy másik bűvárcsapat tagjai egy szemvillanás alatt 30 méter helyett 4 méteres mélységben találták magukat, majd váratlanul újra 20 méterre zuhantak. Aki tisztában van a bűvárookra ható fizikai törvényekkel, sejtheti, milyen veszélyt jelentett ez a bűvárookra, ám szerencsésükre egyiküknek sem esett komolyabb baja ...

<http://www.divecenter.hu/modules.php?sid=172&file=article&name=News> (Heuréka, Origo-Vizsla)

5. Képriport – pánik és félelem, újabb szökőár és földrengés fenyeget?

... A szakértők egyelőre nem látják a pontos okát annak, miért nem kísérte gyilkos erejű szökőár a hétfői indonéziai földrengést, csak sejtéseik vannak róla. A tudósok a lehetséges okok között említik, hogy a hétfői rengés epicentruma a tengerfenék alatt 30 kilométerrel volt, míg a december 26-án bekövetkezett földmozgásé jóval közelebb volt a tengerfenékhez. Allen Clarknak, a Maui-szigeten lévő csendes-óceáni katasztrófa-előrejelző központ igazgatójának

magyarázata szerint ha túlságosan mélyen van az epicentrum, akkor a rengés jelentősen veszít erejéből, még mielőtt eléri a tengerfeneket és hullámokat kelt ...

<http://ma.hu/page/cikk/ab/0/116867/1> (Heuréka)

6. Amikor meglódul a tenger = Cunami!

... Ezt tenger alatti földrengések és vulkánkitörések okozhatják, mégpedig oly módon, hogy a felszabaduló gígászi erők hatására a tenger aljzatában – rendszerint egy-egy törési sík mentén – óriási tömegáthelyeződés megy végbe. A megbomlott egyensúly következtében a terület fölött gyűrű alakú hullám keletkezik, amely aztán nagy sebességgel terjed tova. A nyílt óceánon az esetek többségében mindössze 1-2 méteres hullámok akár 800–850 kilométert is befuthatnak óránként, s ilyenformán a nagyobbak 15–20 óra alatt akár az egész Csendes-óceánt átszelhetik ...

<http://www.sulinet.hu/eletestudomany/archiv/1998/9832/cunami/cunami.html> (Heuréka, Origo-Vizsla, Kurzor)

7. Simonyi Gyula: Természetesek-e a természeti katasztrófák? = Evangélikus Élet, 5. sz. (2005)

... A szökőár Srí Lanka keleti és déli partvidékén mintegy ötszáz kilométer hosszan pusztított. Közel másfél százezren estek áldozatául, kétszázezer családot – közel egymillió embert – kellett kitelepíteni, több mint százezer épület megrongálódott, több tízezer halászbárka megsemmisült. Az első becslések szerint 1,3 milliárd dollár kár keletkezett. A Magyar Ökumenikus Segélyszervezet nyomban vízzel, élelemmel és orvosi segítséggel sietett a helyszínre, a Magyar Baptista Szeretetszolgálat árvaházat hoz létre, a Katolikus Karitászközpont iskolákat hoz rendbe, a Magyar–Srí Lanka-i Baráti Társaság pedig falut és benne iskolát épít ...

<http://www.evelet.hu:8080/ujzagok/evelet/archivum/2005/05/111> (Heuréka)

8. Illyés András: Szökőár: gígászi mozgás a tenger mélyén

... A pusztító erejű hullámokat egy hatalmas, a tengerfenék alatt hirtelen bekövetkező kéregmozgás váltotta ki. Ezen a helyen a hirtelen felszabaduló feszültség hatására az Indiai-közetlemezzel nagyjából tízenöt métert mozgott Indonézia irányába, a Burmai-mikrolemezzel alólódva. Ez a folyamat indította útjára a környező partokat elöntő tengerár nagy erejű hullámait. Bill McGuire (University College London) geofizikus szerint a földmozgás láthatóan befolyásolta a környező szigetek (Simeulue, Nicobar- és Andamán-szigetek) és Szumátra földrajzát. A változások szó szerint néhány percen belül zajlottak le ...

<http://www.origo.hu/tudomany/fold/20050104ujrarajzolta.html> (Heuréka, Origo-Vizsla, Kurzor)

9. A halászatot elvitte a szökőár

... A szökőár sújtotta országokban az egész halászati ipart súlyos veszteségek érték: hajók tízezrei mentek tönkre, felszerelések tűntek el, halászati telepeket mosott el a szökőár. Egyes becslések szerint a halászbárkák 70-80 százaléka a hullámok martaléka lett. Csupán Srí Lankán - amelynek nemzeti össztermékében két százalékot képvisel a jórészt a belső piacot ellátó halászat - 13 ezer halász halt meg, de az érintett régió part menti övezetében halászfalvak és kikötők százai pusztultak el. A FAO még csupán a károk becslésénél tart ...

<http://www.agrarkapu.hu/sajto.php?cikk=5026&hcs=13> (Origo-Vizsla)

10. Béres Zsolt: Szökőár-előrejelzés SMS-ben

... Pontosan 5 hónappal a délkelet-ázsiai cunamikatasztrófa után az Intellipaxx software-ház egy új szolgáltatást vezet be utazók számára. A www.mobipaxx.net oldalon keresztül megrendelhető szolgáltatás SMS-ben értesíti az előfizetőket a megjelölt régiókban várható tengerrengésekről. A veszély közelebb van, mint gondolnánk: a tengeri mozgásokat vizsgáló kutatók szerint a szökőárak több mint 80%-át földrengések okozzák, melyekből közel 12% a Földközi tenger vidékét fenyegeti ...

<http://www.sg.hu/cikk.php?cid=37414> (Origo-Vizsla)

11. Zágoni Miklós: Szökőár: kérdések és válaszok

... A szumátrai tengeralatti földrengés észlelése után 8 perccel az amerikai tengerészeti hivatal által koordinált Csendes-óceáni Cunami Megfigyelő Rendszer (PTWC) fogta a lökéshullám első jeleit, melyről az információt három perc múlva szétküldték a Csendes-óceánon lévő többi megfigyelő állomáshoz. További négy perc múltán azon államok nemzeti szolgálatai, amelyek részt vesznek a PTWC-ben, az alábbi táviratot kapták: „A Csendes-óceánon romboló erejű cunami nem várható.” Azonban India, Indonézia, Thaiföld, Sri Lanka és a Maldiv-szigetek nem részesei az egyezménynek, így az Indiai-óceánra nem telepítettek megfigyelő bójarendszert ...

<http://hps.elte.hu/zagoni/Cunami.htm> (Origo-Vizsla)

12. Online visszaélések a dél-kelet-ázsiai szökőár katasztrófa nyomán

... IT-biztonsággal foglalkozó szakértők egy csoportja arra hívja fel a figyelmet, hogy alaposan nézzük meg, milyen szervezet(ek) számára utalunk pénzt, hiszen vannak közöttük valójában nem is létező csoportosulások, akik csak a jószándékú emberek zsebét akarják kiforgatni. Tulajdonképpen ideális terepet kínál a most kialakult helyzet a phishing támadásoknak: Paul Wood, a Messagelabs egyik elemzője szerint csak idő kérdése, hogy megjelenjenek az ismert segélyszervezetek oldalait másoló csaló weblapok, melyek a gyanútlan internetezőkből kiszedik a pénzt, amit aztán természetesen nem utalnak tovább a károsultak számára ...

<http://www.terminal.hu/newsread.php?id=07201401054011> (Origo-Vizsla)

13. Lakossági tájékoztató a szökőár által sújtott egyes délkelet-ázsiai országok 2005. január 29-i közegészségügyi helyzetéről az Egészségügyi Világszervezet jelentése alapján 2005. január 31.

... A jelentések szerint a szökőár következtében összesen kb. 214500 személy vesztette életét, további 142000 főt eltűntként tartanak nyilván. Az egészségügyi személyzet kb. 35000 sérültet látott el. A szökőár által sújtott térségben közel 2 millió ember vált hajléktalanná. A szökőár idején a fertőződések kockázata alacsony volt, ezért az érintett területet akkor vagy közvetlenül azt követően elhagyó külföldiek esetében semmilyen szűrővizsgálat nem volt indokolt, ilyet sem az Egészségügyi Világszervezet, sem az Európai Unió nem javasolt. ...

http://efrirk.antsz.hu/portal/page?_pageid=240,83684&_dad=portal&_schema=PORTAL (Origo-Vizsla)

14. Thaiföldön történelem a Cunami

... Thaiföldön a kormány elsődleges célként kezeli az ország déli részének helyreállítását, hiszen a turisztikai iparág bevételeinek negyven százalékát ez a körzet adja. Phuket frekvenciált részein a turisztikai kapacitás a közeli jövőben eléri a szökőár előtti szintet és jelenleg a legnagyobb probléma csak az, hogy a szokásos főszezonhoz képest a turisták száma a tizedére csökkent. Az árak hihetetlenül alacsonyak és a turizmus mélyrepülése miatt a helyi lakosság körében komoly foglalkoztatási gondok várhatók ...

<http://index.hu/politika/kulfold/thama0129> (Origo-Vizsla, Kurzor)

15. Turistacsalogató roncs

... A vonaton mintegy kétezer ember utazott tavaly karácsonykor, amikor az ázsiai szökőár elérte Sri Lanka partjait. A hullámok ereje olyan nagy volt, hogy még a síneket is feltépték. Az utasok szinte mind meghaltak. Az áldozatok emlékére a vonat három megmaradt kocsiját érintetlenül hagyták a szerencsétlenség helyszínén. A mementónak szánt roncsokhoz az elmúlt egy év alatt özönlöttek a turisták. "Úgy kerestünk eddig pénzt, hogy a nélkülözésünkről beszéltünk a turistáknak, akik megszántak némi apróval" - mondta egy asszony. Hozzátevé, hogy a vonat romjai így körülbelül 35 családnak adnak megélhetést.

<http://www.online.rtklub.hu/hirek/hatter/?id=0512214488> (Origo-Vizsla)

16. Szegő Iván Miklós: A legnagyobb földrengés 40 éve – szökőár Ázsiában = National Geographic Magyarország, (2004)

... A Szumátra északi csúcsán fekvő Aceh tartományt érintette a legsúlyosabban a cunami, ez a térség egyébként a kormányerők és a szeparatista gerillák egyik fő harcterének számít. Banda Acehben sok ház is megsemmisült, és a város legnagyobb szállodája is részben összeomlott a katasztrófa nyomán. A szomszédos Malajziából is több tucat halálos áldozatot jelentettek a hírügynökségek. Indiában Tamil Nadu és Andhra Pradesh szövetségi államokat sújtotta a természeti csapás. 250 halászt még most is keresnek e térségben a mentőosztagok ...

<http://www.geographic.hu/index.php?act=napi&id=3878> (Origo-Vizsla)

17. Grawátsch Péter: Pokol a paradicsomban: egy évvel a gyilkos szökőár után

... Azon a helyen, amelyet tavaly december végén elsöpört a gyilkos szökőár, a magasabb területekre vezető utak és az előrejelző rendszer megépítése jelentik a túlélés reményét az embereknek egy újabb katasztrófa esetén. Az előrejelző rendszert szeizmikus érzékelők, víz alatti szenzorok és műholdak alkotják, és persze vannak még a nem éppen korszerű menekülő utak, valamint a biztonsági házaknak nevezett menedékek. A riasztó rendszert, amelyet német és helyi tudósok telepítettek a part mentén, néhány nap múlva üzembe helyezik. Azt ígérik, 10 percen belül már a megfigyelő állomáson tudni fogják, ha a földrengés cunamit vált ki ...

<http://www.online.rtlklub.hu/hirek/hatter/?id=0512214458> (Origo-Vizsla)

18. Rejtélyes az állatok hatodik érzéke

... A szökőár után ugyanis szinte egy állati tetemet sem találtak, robbant ki újra egy régi vita: vajon az állatok tényleg rendelkeznek egy úgy nevezett "hatodik érzékkel", mely megsejteti az állatokkal a katasztrófákat? Amikor a tengervíz visszahúzódott a Yalta Nemzeti Parkból, halotti csend lett. Sri Lanka legnagyobb természet rezervátuma teljesen elpusztult. Mikorra a segítők végre odaérkeztek körülbelül 200 emberi hullát találtak, de állati tetemet egyet sem, habár a rezervátum krokodiloknak, több, mint 200 elefántnak, vaddisznónak, vízilónak, majmok, leopárdok és még sok más fajnak az otthona ...

http://www.mcd.hu/pub_med/2005/05_01/005.htm (Origo-Vizsla)

19. Szegő Iván Miklós: Megacúnamik fenyegetik a tengerparti városokat? = National Geographic Magyarország, (2004)

... Ám furcsa dolog, hogy míg a földrengések által előidézett cúnamikat a tudósok gyakran megfigyelik születésüktől fogva a pusztító végkifejletig, a földcsuszamlások által keltett nagyobb szökőárakat, nyílt óceánokon lévő cúnamikat még nem sikerült megvizsgálni. Ezeknek a megacúnaminknak a létezésére vannak bizonyítékok, s az is sejthető, hogy körülbelül százezer évenként következhetnek be ilyen katasztrófák - legalábbis így vélekedik Gary McMurtry, a University of Hawaii szakértője ...

<http://www.geographic.hu/index.php?act=napi&id=3837> (Origo-Vizsla)

20. Az állatok előre érezték a cunami érkezését

... "Valószínű hanghullámokról van szó, amik a vízben nagyon gyorsan terjednek, még az sem biztos, hogy magát a földrengést érezték meg, mert akkor nem egy irányba menekülnének, hanem össze vissza. Valószínű, hogy azokat a lökéshullámokat érezték meg, amik a földrengés során keletkeznek és magát a cunamit hozzák létre" - mondta Dr. Bárdos György biológus. Bizonyos állatok, főleg a föld közelében élő rágcsálók, vagy a hullók, hamarabb érzékelik a szokatlan rengéshullámokat, mint más nagytestű állat, vagy éppen az ember ...

<http://www.tv2.hu/magellan/cikk/17841/> (Origo-Vizsla)

21. Ezreket ölt meg a szökőár Ázsiában

... A rengés nyomán fellépő szökőár Srí Lankától Indonéziáig végigpusztította a partvidéket és a szigetvilágot, egyelőre még megbecsülhetetlen halált és rombolást okozva. A tegnapi esti médiaösszegzések nyolc országban tizenegyezerre tették a már bejelentett halálesetek számát, könnyen lehet azonban, hogy ez köszönő viszonyban sincs a valós pusztítás mértékével. Az érintett országok nagyban különböznek kommunikációs felkészültségük és hajlandóságuk, szervezettségük, hatóságaik hatékonysága tekintetében. ...

<http://www.skymagic.hu/455.0.html> (Kurzor)

22. Mi az a szökőár?

... A tenger alatti földrengések az egész világon megszokottak, azonban csak ritkán ekkora erejűek. Roger Musson földrengéskutató (British Geological Survey) egész eddigi pályafutása során még sosem találkozott a mostanihoz fogható földrengéssel. A 9-es erejű rengés egész pontosan három egymást követő rengésből állt össze, amelyek néhány másodperc elteltével követték egymást. A kezdő rengés, amely Szumátra északi csücskétől nyugatra következett be, két további rengést indított el északi irányban ...

<http://sulinet.hu/tart/fcikk/Kjc/0/24506/1> (Kurzor)

23. Csomán Gábor: Szökőár után

... A szökőár visszahúzódása után a nemzetközi közvéleményt szinte sokkolta, hogy egyes bűnbándák a hozzátartozó nélkül maradt gyerekeket egyszerűen összegyűjtötték, és elszállították (különös figyelmet szentelve a fehér bőrű, nyugati gyerekeknek, akik ára sokkal magasabb a helyiekénél). Az árván maradtakat átmenetileg nevelőszülőkhöz adták, akik egy része azonban csak a támogatásért vállalta a kiskorúak gondozását. Egy lehetséges megoldás a gyerekek csoportos külföldre utaztatása lenne, ahol garantáltan nem kerülnének bűnözők kezébe ...

<http://www.europeer.hu/vilag/cunami050112.html> (Kurzor)

24. Fuldokló korallak a szökőár után

... Az ENSZ 1 millió dollárt különített el egy szakértőcsoport felállítására, melynek feladata, hogy felmérje a szökőár által okozott környezeti károkat. A korábbi környezeti katasztrófák (pl. hurrikánok) által okozott pusztítások felmérései alapján szinte biztosra vehető, hogy a tengeri élővilág itt is hatalmas és tartós károsodást szenvedett, hiszen a szökőár jóval komolyabb rombolást okozott, mint korábban bármely hurrikán. Az amúgy is minden környezeti változásra fokozottan érzékeny korallzátonyok a katasztrófaövezetben darabokra töredezték és megfojtotta őket a vastag iszap és törmelék ...

<http://zold.silent.hu/belso/cunami4.html> (Kurzor)

25. Nagypénteki földrengés. In.: Wikipédia

... A nagypénteki földrengés (másnéven nagy alaszakai földrengés) 1964. március 27-én a mérések kezdete óta feljegyzett második legerősebb földrengés volt, a Richter-skála szerint 9,2-es erősségű. 131 áldozatot követelt. Epicentruma a Prince William-szorosban, közelebbről a College-fjordnál, Dél-Közép-Alaszka partjaitól nem messze volt. Helyi idő szerint délután 5 óra 36 perckor kezdődött és három-öt percig tartott. A csendes-óceáni és az észak-amerikai lemez találkozásának törésvonalánál a tengerfenék elmozdulása hatalmas szökőárt indított el, a szárazföldön pedig a talajszint néhol 11,5 métert is süllyedt ...

http://hu.wikipedia.org/wiki/Nagy_alaszakai_f%C3%B6ldreng%C3%A9s (Kurzor)

26. A szökőár és az információs technológiák

... A szökőár az internetes naplók, vagy más néven blogok területén is hatalmas forgalmat indukált. Ezt elsősorban az információk megosztásának gyorsasága és szerteágazó volta táplálta. A Nielsen//Netratings felmérése szerint a blogger.com látogatottsága 73 százalékkal növekedett a katasztrófa hatására, míg a híroldalagnál ez az érték 27 százalék volt. Részben a blogokhoz

kapcsolódva számos más változást is nyomon követhetünk, amelyek leginkább a különböző technológiák összekapcsolódásáról, és a felhasználók tudatosságának növekedéséről szólnak ...
<http://www.ittk.hu/infinit/2005/0113/index.html> (Kurzor)

27. Pusztító hullámok a vulkánsírból: százméteres szökőár fenyeget

... A valaha látott legnagyobb cunamit, vagyis szökőárt a Kanári-szigetek egyik tagján, a La Palmán fekvő Cumbre Vieja vulkán kitörése fogja kiváltani – állítja Simon Day, a Londoni Egyetem Benfield Greig kutatóintézetének munkatársa. A vulkánkitörés révén bekövetkező földcsuszamlás óriási hullámok kialakulásához vezet majd, amelyek hangsebességhez közeli lendülettel, mintegy 800-900 km/óra sebességgel fognak a Brit-szigetek partjaihoz érkezni. "A kérdés nem az, hogy megtörténik-e, hanem az, hogy mikor történik meg. Elképzelhető, hogy tíz éven belül sor kerül rá, de az is lehet, hogy csak száz év múlva ...

<http://hetilap.hetek.hu/index.php?cikk=33564> (Góliát, Altavizsla)

28. Szökőár pusztított az angol partokon

... Cornwall turisták által is nagyon kedvelt tengerparti településeiről és a környék festői faluiból több, mint száz embert kellett evakuálni kedden rendőrségi források szerint. A vízözönt a hétfő délutáni, több mint hat centiméternyi csapadék váltotta ki, a szökőár a helyi lakosok beszámolója szerint közel egyméteres magasságú hullámban öntötte el a környék településeit. Hivatalos források szerint eddig nyolc embert kellett orvosi ellátásban részesíteni, többnyire töréses sérülések és kiszáradás miatt. A környékről kimentett lakosokon túl akár további ezer ember légi mentése válhat szükségsszerűvé ...

<http://www.nol.hu/cikk/329588/> (Góliát, Altavizsla)

29. Bohus Péter: Egy vulkán, amely letarolhatja az USA keleti partvidékét = National Geographic Magyarország, (2004)

... Ha La Palma szigetén a Cumbre Vieja vulkán kitör, egy kisebb szigetnyi kősziklát robbantva a vízbe, akkora erejű szökőárt, más néven cunamit indít el, hogy Amerika összes erőműve termel fél év alatt annyi energiát. A hullám több mint 6000 kilométert megtéve az Atlanti óceánon, kilenc óra alatt elérné a Karib-szigeteket, Kanadát és az Egyesült Államok keleti partját, ahol betörve a szárazföldre egy 30 kilométer széles sávban, mindent eltörölne, ami az útjába akad. Az amerikai Benfield Grieg Katasztrófakutató Központ munkatársai úgy vélik, valódi a veszély ...

<http://www.geographic.hu/index.php?act=napi&id=2920&PHPSESSID=61ac61e7ea513c6b5028664ce85b4fa0> (Góliát, Altavizsla)

6. melléklet

Az egyes keresőkulcsokra visszakeresett releváns és összes találatszámok

<i>Rákóczi-szabadságharc</i>	Releváns találatok száma az első 20 találatban	Releváns találatok száma a második 20 találatban	Releváns találatok száma a harmadik 20 találatban	Összes találat száma
Heuréka	5	3	2	836
Origo-Vizsla	7	3	5	4 240
Kurzor	11	3	2	157
Góliát	6	0	1	16 641
Altavizsla	6	1	1	60 894

1. táblázat. „Rákóczi-szabadságharc” keresőkulcsra kapott találatok száma

<i>Evészavar</i>	Releváns találatok száma az első 20 találatban	Releváns találatok száma a második 20 találatban	Releváns találatok száma a harmadik 20 találatban	Összes találat száma
Heuréka	6	1	0	125
Origo-Vizsla	6	5	6	1 070
Kurzor	9	2	0	1 422
Góliát	6	1	0	5 296
Altavizsla	6	1	0	5 296

2. táblázat. „Evészavar” keresőkulcsra kapott találatok száma

<i>Szökőár</i>	Releváns találatok száma az első 20 találatban	Releváns találatok száma a második 20 találatban	Releváns találatok száma a harmadik 20 találatban	Összes találat száma
Heuréka	3	1	4	446
Origo-Vizsla	7	4	4	6 560
Kurzor	5	1	3	12 766
Góliát	2	1	0	247
Altavizsla	2	1	0	247

3. táblázat. „Szökőár” keresőkulcsra kapott találatok száma

<i>Kegyhely</i>	Releváns találatok száma az első 20 találatban	Releváns találatok száma a második 20 találatban	Releváns találatok száma a harmadik 20 találatban	Összes találat száma
Heuréka	6	3	6	255
Origo-Vizsla	8	6	2	2 790
Kurzor	8	3	2	3 469
Góliát	7	4	2	5 286
Altavizsla	7	4	2	5 286

4. táblázat. „Kegyhely” keresőkulcsra kapott találatok száma

7. melléklet

A keresőkulcsokra kapott releváns és összes találatszámok

Releváns találatok száma az első 20 találatban	Releváns találatok száma a második 20 találatban	Releváns találatok száma a harmadik 20 találatban	Összes találat száma	Rákóczi - szabadságharc	Evészavar	Szökőár	Kegyhely	Heuréka	Origo-Vizsla	Kurzor	Góliát	Altavizsla
5	3	2	836	1	0	0	0	1	0	0	0	0
7	3	5	4 240	1	0	0	0	0	1	0	0	0
11	3	2	157	1	0	0	0	0	0	1	0	0
6	0	1	16 641	1	0	0	0	0	0	0	1	0
6	1	1	60 894	1	0	0	0	0	0	0	0	1
6	1	0	125	0	1	0	0	1	0	0	0	0
6	5	6	1 070	0	1	0	0	0	1	0	0	0
9	2	0	1 422	0	1	0	0	0	0	1	0	0
6	1	0	5 296	0	1	0	0	0	0	0	1	0
6	1	0	5 296	0	1	0	0	0	0	0	0	1
3	1	4	446	0	0	1	0	1	0	0	0	0
7	4	4	6 560	0	0	1	0	0	1	0	0	0
5	1	3	12 766	0	0	1	0	0	0	1	0	0
2	1	0	247	0	0	1	0	0	0	0	1	0
2	1	0	247	0	0	1	0	0	0	0	0	1
6	3	6	255	0	0	0	1	1	0	0	0	0
8	6	2	2 790	0	0	0	1	0	1	0	0	0
8	3	2	3 469	0	0	0	1	0	0	1	0	0
7	4	2	5 286	0	0	0	1	0	0	0	1	0
7	4	2	5 286	0	0	0	1	0	0	0	0	1
<i>Összesen:</i> 123	48	42	133 329	5	5	5	5	4	4	4	4	4

AZ INTERNETES KERESŐK LEKÉRDEZÉSI HATÉKONYSÁGÁNAK VIZSGÁLATA

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
a Matematika és Számítástudományok tudományágban

Írta: Tóth Erzsébet okleveles informatikus könyvtáros – angol szakos tanár

Készült a Debreceni Egyetem Matematika és Számítástudományok Doktori Iskolája
(Informatika Programja) keretében

Témavezető: Dr. Fazekas Gábor
(olvasható aláírás)

A doktori szigorlati bizottság:

elnök: Dr. Dömösi Pál
tagok: Dr. Kormos János
Eszenyiné Dr. Borbély Mária

A doktori szigorlat időpontja: 2006.

Az értekezés bírálói:

Dr.
Dr.
Dr.

A bírálóbizottság:

elnök: Dr.
tagok: Dr.
Dr.
Dr.
Dr.

A nyilvános vita időpontja: 200... .