



Az öt legnépszerűbb webes kereső teljesítményének értékelése és összehasonlítása

Az internetezők többsége elsősorban keresőgépek segítségével igyekszik releváns információkat találni a weben. A keresési iparág nagyon gyorsan fejlődik, de a keresőgépeket értékelő kutatások nem tartanak ezzel lépést. Pedig ezeknek a vizsgálatoknak kétféle előnyük is van: egyrészt segítik a felhasználókat a legjobb szolgáltatások kiválasztásában, másrészt hasznos információkkal szolgálhatnak a fejlesztőknek. A cikk szerzői viszonylag nagy számú mintán végzett statisztikai elemzéssel próbálták értékelni az öt legnépszerűbb keresőgép teljesítményét és hatékonyságát. A Google, a Yahoo!, a Live, az Ask és az AOL keresőjét tesztelték 2007 júliusában, majd a lekérdezéseket egy hónap múlva megismételték, hogy az időbeli változást is mérni tudják.

Korábbi kutatások

A kilencvenes évek közepén jelentek meg az első publikációk a keresőgépek értékelésével kapcsolatban, és az azóta eltelt időben többen is megpróbáltak olyan módszerek kidolgozásával, amelyekkel objektív és egymással összevethető mérési eredményeket lehetne elérni. Az elsők között kell megemlíteni *Chu* és *Rosenthal* [1] 1996-os kutatását, melynek során az AltaVista, az Excite és a Lycos összehasonlítását végezték el: mindegyiknél az első tíz találat pontosságát határozták meg egy háromfokú skálán. *Gordon* és *Pathak* [2] 1999-ben nyolc szolgáltatást (AltaVista, Excite, Infoseek, Open Text, HotBot, Lycos, Magellan és Yahoo!) tesztelt egy jóval nagyobb vizsgálat keretében: 33 kérdést futtattak le, és az első 200 találatot értékelték egy négy szintű relevanciaskálát használva. *Hawking* és szerzőtársai [3] 2001-es publikációjukban már húsz keresőgépet értékelték a Text REtrieval Conference (TREC) egyik workshop-jában, mindegyiknél a legjobb 20 találatot elemezték. *Shang* és *Longzhuang* [4] 2002-ben automatikus módszerekkel 3 ezer keresést végzett az

AltaVista, a Fast, a Google, a Go, az iWon és a NorthernLight adatbázisában, majd a találatok pontosságát meghatározva statisztikai módszerekkel mérték össze a hat keresőgép teljesítményét. 2005-ben a *Dogpile.com* [5] folytatott le egy nagyobb kutatást, összehasonlítva a Google, a Yahoo!, az MSN és az Ask Jeeves találati halmazait. Az eredmények jelentős különbségeket mutattak az akkori négy legnépszerűbb kereső között és arra is meggyőző bizonyítékokkal szolgáltak, hogy a keresőgépekkel nem mindig találják meg a felhasználók azt, amit szeretnének. *Spink* és munkatársai [6] 2006-ban publikálták egy jelentős méretű kutatás eredményeit: több mint 10 ezer keresés találati listáinak első oldalát elemezték az MSN Search, a Google, a Yahoo! és az Ask Jeeves esetében. Születtek már bibliográfiák és honlapok is a témában, utóbbiak közül a legfontosabbak a SearchEngineWatch, a SearchEngineLand és a SearchEngineJournal.

Vizsgálati módszerek

A jelen cikkben ismertetett vizsgálat során az online piacot elemző cégek (*ComScore*, *OneStat*, *Nielsen/NetRatings*, *Hitwise*) által évek óta az élmezőnybe sorolt Google, Yahoo!, Live, Ask és AOL keresőket tesztelték indiai szakemberek. A tesztkérdéseket könyvtár- és információtudományi folyóiratok publikációjának kulcsszavaiból állították össze: egy nagyobb, 350 fogalmat tartalmazó listából választottak ki végül 50 reprezentatív szakkifejezést, és ezeket írták be a keresőmezőkbe (idézőjelek között). Mivel a korábbi vizsgálatok azt mutatták, hogy a felhasználók kb. 80%-a sosem néz meg többet az első 10 találatnál, ezért a kutatók is csak az első tíz – nem szponzorált – linkre kattintottak rá, és a megjelenő weboldalakat elmentették egy adatbázisba, majd kielemezték őket négy szempontból: relevancia, hibás linkek, duplumok, stabilitás. Az egyes keresőgépek teljesít-

ményét ezek után többféle vonatkozásban is értékelték: lefedettség, pontosság, linkek, stabilitás, egyedi találatok. A statisztikai átlagok összevetésénél az egyszempontos varianciaanalízist (*one-way ANOVA*) használták, illetve a Tukey-féle HSD (*honestly significantly different*) próbával ellenőrizték a csoportok közötti eltérések szignifikanciáját.

Eredmények

Lefedettség

A legtöbb találatot minden esetben a Google adta, őt követte a Yahoo!, a Live pedig általában lényegesen kisebb halmazokat adott vissza. Az AOL teljesítményét nem lehetett értékelni ebből a szempontból, mert az nem jelzi a találatok számát. Azt, hogy a Google sokkal több weblapot indexel le a társainál, más kutatások is megerősítik.

Pontosság

Hogy egy találat hová kerül egy relevancia szerint rendezett listában, azt sok szempont befolyásolja: például az adott weblap népszerűsége, a keresőszavak helye, egymáshoz viszonyított pozíciója és előfordulási gyakorisága az oldalon. Az egyes keresőgépek találati listáinak pontosságát a kutatók úgy mérték, hogy az 50 kérdésre adott első 10 találat mindegyikéről eldöntötték, hogy releváns-e vagy sem az adott téma szempontjából, majd ezek százalékos arányát vetették össze. Itt is a Google érte el a legjobb eredményt: tízből átlagosan 7,90 releváns találatot adott mindkét tesztelési fázisban, míg a Yahoo! 6,98-as, illetve egy hónappal később 7,16-os értéket ért el. Az Ask és a Live teljesített a leggyengébben: 5,92 és 6,34 közötti értékekkel. Az ANOVA és a Tukey-féle HSD elemzések kimutatták, hogy jelentős eltérések vannak az öt keresőgép által adott releváns találatok között, és hogy páronként összehasonlítva őket, a legnagyobb különbség a Google és az Ask, illetve a Live találati halmazai között van.

Ismétlődő linkek

Mindegyik keresőgépnél előfordultak időnként azonos oldalra való többszörös hivatkozások ugyanazon találati listában, vagyis egyiknél sem tökéletes a duplumszűrés. A legkisebb számban (1,03% – 1,65%) az Ask listáiban lehetett ilyeneket találni, a legtöbbet (2,07% – 4,16%) pedig a Yahoo! és az AOL produkálta.

Hibás linkek

Több oka is lehet annak, ha egy hivatkozás nem működik egy találati listában: például törlődött vagy már nem publikus az adott weblap, a szerver nem válaszol vagy nem sikerül feloldani a doménnevét. A kutatás első fázisában az Ask listáiban volt a legtöbb problémás link (6,20%), a legkevesebb pedig a Google találatai közt (2,46%); de az egy hónappal később megismételt kísérletben a sorrend megváltozott: ekkor az Ask nyert (3,90%), míg a Google a középmezőnybe került (4,35%).

Stabilitás

A kutatók megnézték azt is, hogy az egyes keresőkérdéseknél a találatok száma mennyit változott 2007 júliusa és augusztusa között, illetve hogy az első tízbe került weblapok hány százaléka maradt ugyanaz ez alatt az egy hónap alatt. Ami a találati számot illeti: ez a Yahoo! esetében kiugróan megnőtt, de a Live-on kívül a többinél is növekedés mutatkozott. Ha csak a releváns találatok százalékos arányát nézzük, akkor a Google bizonyult a legstabilabbnak, nála mindössze 0,90%-kal változott ez az érték, míg az Ask esetében 3,93%-kal nőtt a relevancia. A top 10 találat stabilitását vizsgálva a Yahoo! 93,14%-os eredményt ért el, ezt követte az AOL 89,40%-os és a Google 86,34%-os értékkel – vagyis egy hónap elteltével ilyen százalékban maradt ugyanaz az első találati oldaluk összetétele. Az egyszempontos varianciaanalízis ezen a téren is jelentős különbségeket mutatott ki az egyes keresőgépek között.

Egyedi találatok

Ha egy kereső első tíz találati között volt olyan, amely a többinél nem fordult elő, akkor ezt egyedinek minősítették a kutatók, és kiszámolták ezek arányát mindegyik keresőgépnél. Ebben a tekintetben a Google bizonyult a legjobbnak: 10-ből átlag 1,6 találat volt egyedi, őt követte a Yahoo! fele ennyi egyedi találattal.

Ez a kutatás is igazolta, hogy az eltérő technológiát használó keresőgépek nagyon eltérő eredményeket adnak mindenféle szempontból. Összességében a Google teljesített a legjobban, míg a második helyre a Yahoo! került. Természetesen ennek a vizsgálatnak is megvoltak a maga korlátai: például csak az első 10 találatot elemezték és csak idézőjeles kifejezéseket kerestek, így nem lehetett felmérni sem azt, hogy összesen mennyi releváns találatot adnak az egyes keresők, sem

azt, hogy komplexebb kérdések (pl. a NOT operátor használata bizonyos szavak kizárására) hogyan befolyásolnák a találatok pontosságát. Az eltérő felhasználói felületek és keresési opciók is hatással lehetnek a keresés eredményességére, ezek vizsgálatára sem tért ki ez a kutatás. Továbbá, mivel a tesztek egy hónap múlva ismételték meg, ez kevés volt ahhoz, hogy érdemi következtetéseket lehessen levonni a találati listák időbeli változásával kapcsolatban. És végül: érdemes volna a metakeresők hatékonyságát is megvizsgálni és összehasonlítani az egyedi keresőgépek teljesítményével.

Hivatkozások

- [1] CHU, H. – ROSENTHAL, M.: Search engines for the world wide web: a comparative study and evaluation methodology. ASIS Annual Conference Proceedings, Baltimore (MD), 1996. október 19–24., p. 127–135. <http://www.asis.org/annual-96/ElectronicProceedings/chu.html> (letöltve: 2007.01.20.)
- [2] GORDON, M. – PATHAK, P.: Finding information on the world wide web: the retrieval effectiveness of search engines. = Information Processing & Management, 35. köt. 2. sz. 1999. p. 141–180.
- [3] HAWKING, D. – CRASWELL, N. – BAILEY, P. – GRIFFITHS, K.: Measuring search engine quality. = Information Retrieval, 4. köt. 1. sz. 2001. p. 33–59.
- [4] SHANG, Y. – LONGZHUANG, L.: Precision evaluation of search engines. = World Wide Web, 5. köt. 2. sz. 2002. p. 159–173.
- [5] Dogpile.com: Different engines, different results. 2005. <http://www.infospaceinc.com/onlineprod/Overlap-DifferentEnginesDifferentResults.pdf> (letöltve: 2007.07.01).
- [6] SPINK, A. – JANSEN, B.J. – BLAKELY, C. – KOSHMAN, S.: A study of results overlap and uniqueness among major web search engines. = Information Processing & Management, 42. köt. 5. sz. 2006. p. 1379–1391.

/DEKA, Sanjib K. – LAHKAR, Narendra: Performance evaluation and comparison of the five most used search engines in retrieving web resources. = Online Information Review, 34. köt. 5. sz. 2010. p. 757–771./

(Drótos László)