

Google rangsorolás: valóban az első találatok a legrelevánsabbak a felhasználók szerint?

Az internetes keresőrendszerek segítségével a felhasználók minél relevánsabb információkat szeretnének találni egy adott témában, minél kevesebb fáradtsággal, a lehető legrövidebb idő alatt. E rendszerek fejlesztői bonyolult algoritmusokat használva igyekeznek meghatározni, hogy egy weblap vagy más dokumentum tartalma mennyire illeszkedik egy adott keresőkérdéshez, és e fontossági érték alapján rendszerezik a megjelenített találatokat. Ha jó ez a sorrendezési technika, akkor a legfontosabb tételek a lista elejére kerülnek és a felhasználónak nem kell az első egy-két találati oldalnál többet átnéznie. Az iráni szerző cikkében bemutatott kutatás során azt vizsgálták, hogy a legnépszerűbb keresőgép, a Google első 40 találatát mennyire tartják relevánsnak a kísérletbe bevont egyetemi hallgatók, illetve hogy az első négy (egyenként 10 tételes) találati oldal között van-e számottevő különbség a releváns találatok aránya szempontjából? Más szavakkal: mennyiben egyezik meg a Google által szabadalmaztatott – a weblapok közti hiperlink-kapcsolatokat súlyozó – *PageRank* algoritmusra épülő és emellett számos egyéb paramétert is figyelembe vevő „objektív” sorrend a felhasználók „szubjektív” értékítéletével?

Korábbi kutatások

A keresőrendszerek találati pontosságát különféle módszerekkel már korábban is sokan vizsgálták – és meglehetősen vegyes eredményeket kaptak, illetve eltérő következtetésekre jutottak. Például egy 1999-es tesztben [1] az akkori öt legismertebb keresőgép teljesítményét hasonlították össze hat TREC (Text REtrieval Conference) rendszer információkeresési pontosságával, és akkor az előbbieket meglehetősen alulmaradtak. 2001-ben viszont már egy hasonló kísérletben [2] a TREC workshop keretében fejlesztett algoritmus mérhetően gyengébben teljesített, mint a négy legnagyobb nyilvános keresőgép. 2002-ben *Chowdhury* és *Soboroff*

[3] a Lycos, a Netscape, a Fast, a Google és a HotBot találati listáit elemezve nem talált statisztikailag jelentős eltérést közöttük relevancia szempontból. *Vaughan* 2004-es kutatása [4] viszont egyértelműen a Google fölényét mutatta ki az AltaVista és a Teoma nevű keresőgépekkel szemben. Azok a tesztek [5., 6., 7], amelyekben a felhasználók preferencia sorrendjeit vetették össze a keresőrendszerek által produkált rangsorral, többnyire alacsony fokú egyezést mutattak ki, vagyis az emberek nem igazán azt ítélték fontosnak, amit a gépi intelligencia. Az is bebizonyosodott, hogy nincs „átlagos felhasználó” – a tesztalanyok, még ha hasonló tudással rendelkeznek is egy témában, egyedi kontextusban értékelik a talált információkat, melyet kognitív, érzelmi és fizikai tényezők is befolyásolnak.

A jelen kutatás ismertetése

A cikkben ismertetett kutatás során két kérdésre szerettek volna válasz kapni: 1. Valós keresési feladatoknál valóban relevánsabb-e az a találat, amelyet a Google előrébb helyez a listában? 2. Az első 10 vagy 20 tétel között jelentősen több-e a releváns, mint a következő 3. és 4. találati oldalon?

A „relevancia” fogalmát kétféleképpen is szokás értelmezni. Az első esetben a visszaadott dokumentumok témáját vagy esetleg egyéb jellemzőit vetik össze a keresőkérdéssel, és ennek alapján határoznak meg egy közelségi értéket. A második értelmezés szerint a relevancia azt jelenti, hogy egy adott dokumentum mennyire értékes vagy hasznos a kereső személy számára. Ebben a kutatásban az első értelmezést használták, vagyis a tesztben részt vevő 34 egyetemi hallgatót arra kérték, hogy a 34 keresés során kapott találatok mindegyikét egységes mércével értékeljék, azokat is, amelyeket ismernek, és ezért már érdektelenek számukra. Minden egyes tételt egy háromfokú

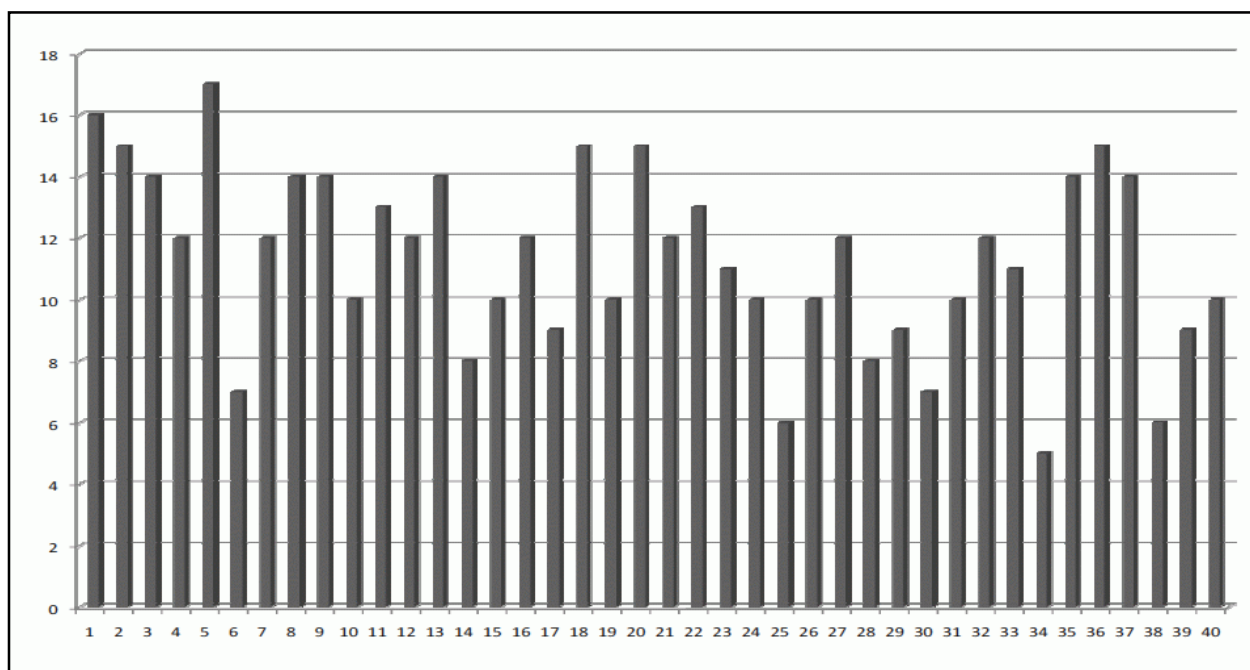
skálán minősítettek: „A” osztályzatot kaptak a legrelevánsabb találatok, „B”-t a csak részben pontosak, és egy „-” jelet azok, amelyek nem kapcsolódtak eléggé a keresett témához. A keresőkérdések valós kutatási témákra vonatkoztak (pl. *globalisation + "Islamic countries"*, *transportation + "air pollution"*, *vandalism + children*, *metadata + indexing*, *librarians + burnout*) és a Google angol nyelvű űrlapjába írták be őket. A tesztelés 2008 októberé és 2009 szeptemberé között zajlott.

Az első kérdés megválaszolása céljából az első 40 találatot értékelték és meghatározták minden találati pozícióhoz az „A” valamint az „A+B” minősítések számát, illetve ezek százalékát. A második kérdést pedig az első 4 találati oldal összehasonlításával döntötték el: kiszámolták mindegyik oldalon a releváns találatok arányát, és *t*-próbával határozták meg, hogy a középértékek közötti eltérések statisztikailag szignifikánsak-e.

Az 1. ábrán az látszik, hogy a 34 keresés során a találati listák egyes helyein hány esetben volt „A” minősítésű dokumentum, vagyis hogy az első 40 helyen milyen gyakorisággal voltak valóban releváns találatok. A legtöbb (17) esetben az 5. pozícióban volt releváns tétel, ezt követte 16 esetben az 1. pozíció. Érdekes megfigyelni, hogy még a 4. oldalon is vannak olyan helyek (a 35., a 36. és a 37. pozíció), ahol igen gyakran (14, illetve 15 esetben) jelentek meg fontos dokumentumok. Ha az

„A” és a „B” minősítésű találatok számát összeadjuk, akkor a 4. hely érte el a legmagasabb értéket: ott 28 esetben, vagyis a keresések 82 százalékában volt többé-kevésbé releváns találat.

A négy találati oldalt összehasonlítva a kutatók azt az eredményt kapták, hogy jelentősebb különbség csak az 1. és a 3. között volt az „A” jelzésű tételeket figyelembe véve, az „A+B” esetben pedig az 1. és a 4. lap között. A negyedik lapon – a 40-49. pozícióba került találatok között – több volt az „A” minősítésű releváns tétel, mint a harmadikon, vagyis ha a felhasználók csak az első egy-két oldalt nézik át a listában, akkor fontos dokumentumokat hagynak figyelmen kívül. Célszerű tehát tovább lapozni olyankor, amikor a top 10 vagy 20 találat között kevés az igazán releváns. Az „A+B” együttes kategóriát vizsgálva már jobban kirajzolódott az a trend, hogy az első oldal legmagasabb találati pontossága (72%) után a következő oldalakon egyre alacsonyabb volt ez az érték, ami azt mutatja, hogy a Google a felhasználói szempontokkal nagyjából összhangban rendezi a valamelyest releváns találatokat, csak az igazán fontosak esetében nem egyezik a gépi és az emberi rangsor. Ez az eredmény megerősíti a korábbi kutatásokból leszűrt véleményeket, melyek szintén azt fogalmazták meg, hogy a korreláció nem jelentős, és hogy a jelenlegi nyilvános webes keresőgépek sorrendezési algoritmusain van még mit javítani.



1. ábra Az „A” minősítésű találatok száma az első 40 listapozícióban (34 db. keresés alapján)

Hivatkozások

- [1] HAWKING, D. – CRASWELL, N. – THISTLEWAITE, P. – HARMAN, D.: Results and challenges in web search evaluation. = *Computer Networks*, 31. köt. 11–16. sz. 1999. p. 1321–1330.
- [2] SINGHAL, A. – KASZKIEL, M.: A case study in web search using TREC algorithms. *Proceedings of the 10th International Worldwide Web Conference, 2001. május*
<http://www10.org/cdrom/papers/pdf/p317.pdf> (letöltve: 2011. április 20.)
- [3] CHOWDHURY, A. – SOBOROFF, I.: Automatic evaluation of world wide web search services. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland. New York, ACM Press, 2002. p. 421–422.*
- [4] VAUGHAN, L.: New measurements for search engine evaluation proposed and tested. = *Information Processing and Management*, 40. köt. 4. sz. 2004. p. 677–691.
- [5] NOWICKI, S. A.: Students' success with world wide web search engines: retrieving relevant results with respect to end-user relevance judgments. PhD thesis, Nova Southeastern University, 2002.
- [6] PATIL, S. – ALPERT, S. R. – KARAT, J. – WOLF, C.: THAT's what I was looking for: comparing user-rated relevance with search engine rankings. *Human-computer Interaction, INTERACT 2005 International Conference Proceedings, Lecture Notes in Computer Science, 3855. köt. Berlin, Springer, 2005. p. 117–129.*
- [7] BAR-ILAN, J. – LEVENE, M. – MAT-HASSAN, M.: Methods for evaluating dynamic changes in search engine rankings: a case study. = *Journal of Documentation*, 62. köt. 6. sz. 2006. p. 708–729.
- /HARIRI, Nadja: Relevance ranking on Google: are top ranked results really considered more relevant by the users? = *Online Information Review*, 35. köt. 4. sz. 2011. p. 598–610./**

(Drótos László)