

## Keresőgépek: valós idejű keresés

### Bevezetés

A „valós idejű keresés” divatos téma lett, de hogy pontosan mit is értünk alatta, az attól függ, hogy a kifejezés melyik tagjára helyezzük a hangsúlyt. Az egyik értelmezés szerint [1] olyan tartalmak kereséséről van szó, amelyeket gyakorlatilag azonnal közzétesznek, tehát alig telik el idő a létrehozásuk és a publikálásuk között. Vagyis például egy olyan blogbejegyzés visszakeresése, amelyet mondjuk öt órán át írt a szerzője, nem „valós idejű”, még akkor sem, ha a *post* megjelenése után néhány másodperccel már megtalálható valamelyik keresőgéppel. Egy másik definíció [2] így határozza meg ezt a fogalmat: a megfelelő válasz megtalálása a felhasználót éppen érdeklő témában az éppen elérhető tartalmak között. Jelen cikk szerzője az első értelemben vizsgálja a *real-time search* helyzetet, vagyis hogy hogyan lehet olyan írott, képi, vagy videodokumentumokat megtalálni, amelyeket az elmúlt percekben hoztak létre és tettek fel a webre, tehát arról szólnak, ami pont most történik a világban. Ezért ebben az írásban elsősorban a mikroblog-szolgáltatások (Twitter és társai) kereséséről esik szó, és nem az olyan hírkeresőkről, mint amilyen például a Google News.

A valós idejű tartalmaknál mindig felmerül a *spam* problémája és a hitelesség kérdése. A különösen divatos vagy „forró” témákat egyesek megpróbálják meglovagolni és a saját oldalukra csalogatni a téma iránt érdeklődőket „Érdekes hír <az aktuális üggyről>, de látogasd meg az én hogyan-legyünk-milliomosok weblapomat is” típusú üzenetekkel. Ami pedig a hitelességet illeti: a hagyományos webes műfajoknál megvan arra a lehetőség, hogy átnézzük a honlap egyéb részeit, visszafelé is elolvassuk a blogbejegyzéseket, vagy ellenőrizzük, hogy kik linkelik az adott oldalt. A valós idejű tartalmak között való tallózáskor nincsenek ilyen módszerek a közlő megbízhatóságának ellenőrzésére, legfeljebb az életrajzi adatait tudjuk megnézni, ha megadott egyáltalán valamit magáról. Vagyis bízunk kell benne, hogy az illető valóban ott van,

ahol most zajlik valami, nem pedig egy olyan hamis tudósítást olvasunk, amely igazából meg sem történt.

További probléma lehet az így termelődő információ mennyisége. Ha valaki egy aktuális *breaking news* típusú eseményt követ egy blog- és mikroblog-keresővel, akkor néhány percenként százszámra kap újabb találatokat. Ehhez képest egy hagyományos, statikus weblapkereső által adott ötezer találat, amiket kényelmesen át lehet nézegetni, egészen ideális helyzetnek számít.

### Valós idejű keresők

A legfontosabb mikroblog-szolgáltatónak, a *Twitternek* (*twitter.com*) természetesen van saját keresője. Sajnos a Twitter nem az elvárható gondossággal őrzi a rajta megjelenő tartalmat, rendszeresen elvesznek üzenetek, így a keresője sem tekinthető teljes körűnek. De az alapfunkciókon kívül van néhány érdekes lehetőség az összetett keresőoldalon (*search.twitter.com/advanced*), érdemes ezért kipróbálni, ám mivel a Twitter honlapjáról nincs erre az úrlapra hivatkozás, valószínűleg kevesen használják. Pedig lehet vele földrajzi helyre és időintervallumra szűkíteni, pozitív és negatív üzeneteket keresni (az alapján, hogy :) vagy :(hangulatjelek vannak-e a szövegben), sőt kérdésekkel is (ilyenkor a ? karakter előfordulását veszi figyelembe). Hogy mennyire valós idejű a Twitter keresője, azt a szerző egy egyedi teszttüzenettel (*asd123poi456*) próbálta ki: a délelőtt 11.50-kor elküldött szöveg 25 másodpercen belül jelent meg a keresőben.

A *Sency* (*sency.com*) egy új kereső, nemrég lépett túl a béta-állapoton. A keresősor alatt két oszlopban láthatók a pillanatnyilag, illetve az utóbbi időben aktuális témák. A próbakeresés alapján feltételezhető, hogy csak a Twitter adatbázisát használja, és ez is egész gyorsan megtalálta a fenti teszttüzenetet. Nincsenek azonban további funkci-

ói, és nincs hozzá súgó vagy „Gyakran Ismételt Kérdések” oldal. Hasznos viszont, hogy kérhetünk egy HTML kódot a keresőkérdésünkhöz, ami egy weblapba vagy blogba beépítve hírcsatornaként fog működni.

A *Collecta* (*collecta.com*) szélesebb körből merít: a Twitter mellett blogokat és azok kommentjeit, továbbá a *Jaikua*, az *Identi.ca*, a *Flickr*, a *TwitPic*, a *yFrog* tartalmát, valamint a *YouTube* és a *Ustream* videóit is leindexeli. Háromoszlopos megoldást használ, ahol is bal oldalt van a keresőmező, az éppen legfontosabb témák, a különböző opciók és egyebek; középen a találati lista, és jobb szélén egy adott találat részletesen. Az eredmények ötféle kategória szerint szűrhetők: Stories, Comments, Updates, News, Video. Ha például a friss üzenetekre vagyunk kíváncsiak, akkor az Updates opciót kell csak kipipálni, így a keresés a Twitter, a Jaiku és az *Identi.ca* tartalmára korlátozódik. A keresési eredmények megoszthatók különféle közösségi oldalakon (pl. Facebook, Delicious). A korábbi keresőkérdéseink megmaradnak addig, amíg a *Collecta* honlapján vagyunk, így rájuk kattintva megnézhetjük, hogy milyen újdonságok jelentek meg az eredeti lekérdezés óta. A rossz hír az, hogy nincs RSS csatorna és minimális a segítség a keresőkifejezések lehetséges szintaxisához.

A *OneRiot* (*oneriot.com*) szintén a valós idejű web keresőjeként hirdeti magát. Leírása szerint robotja azokat a linkeket járja végig, amelyeket az emberek megosztanak egymással a Twitteren, a Digg-en és más közösségi helyeken, majd ezek tartalmát indexeli le másodpercek alatt. A találati listákban tehát azok a tartalmak szerepelnek, amelyek éppen az internetes kommunikáció középpontjában vannak. Így érthető, hogy az *asd123poi456* üzenetet még 45 perc múlva sem találta meg a *OneRiot*. Kétfajta keresési opciója van: Web és Video; és a találati lista is két szempont szerint rendezhető: Realtime és Pulse. Az első a legutóbb megosztott tartalmakat helyezi előtérbe, a második pedig azoknak ad nagyobb súlyt, amelyekre a legtöbben hivatkoztak eddig.

A *Scooper* (*scooper.com*) szintén olyan kereső, amely többféle forrást indexel: Twitter, Flickr, Digg, Delicious és egyebek. Hasonlóan két csoportra osztja a találatokat, mint a *OneRiot*: ám itt két oszlopban egymás mellett látjuk a „legnépszerűbb” és a „valós idejű” találatokat. Az eredmények között képek és videók is vannak, de ezeket kiszűrhetjük, ha akarjuk. A My Searches menüpont alatt a korábbi kereséseinket újra lefuttathatjuk, a Hottest Topics alatt pedig az aktuálisan legnépszerűbb

témákat találjuk. A keresések szűkíthetők nagy témakörök szerint, például Technology, World Business, Sports. A *Scooper* az elmúlt percben közzétett tartalmakat is megtalálja, és szépen visszaadja a tesztüzenetet is időbélyegzővel, a szerzője avatárjával (az őt jelképező képecskével) és Twitter-oldalának címével együtt.

A *CrowdEye* (*crowdeye.com*) egy béta-állapotú Twitter-kereső, de a találati listája már így is nagyon informatív. A keresési eredmények időszakok szerint jeleníthetők meg, és egy grafikon mutatja, hogy mely órákban hány üzenet született az adott témában, így jól kirajzolódnak a csúcsidezőszakok. A találatok pontosítására a rendszer címkefelhőt, saját szavak beírására alkalmas szűkítő mezőt, valamint *hashtag*-alapú (# karakterrel jelölt kulcsszavakra vonatkozó) kapcsolódó kereséseket kínál fel. A középső oszlopban linklistát kapunk a keresett témában legnépszerűbb blogokról és hírforrásokról, és itt jelennek meg a releváns Twitter-üzenetek is, melyeket fontosság vagy idő szerint rendezhetünk. Érdekes specialitás, hogy minden szerző avatárja mellett egy szám is szerepel, ami az illető befolyását mutatja a Twitter-közösségen belül. Például *Stephen Fry* pontszáma 86, a *Daily Telegraph* újságé 50. A képernyő jobb szélén pedig az aktuálisan „forró” témák, és a legnépszerűbb 20 webhelyre mutató hivatkozások sorakoznak. A *CrowdEye* egy jól megtervezett, hatékony és könnyen használható kereső. Ha a Twitteren kívül további valós idejű forrásokat is leindexelne, szinte verhetetlen lenne ebben a műfajban.

A fura nevű *Stinky Teddy* (*stinkyteddy.com*) egy metakereső, mely a Bing, a Yahoo!, a Videosurf, a Twitter és a *Collecta* keresőit kérdezi le, és megadja, hogy hány üzenet, weboldal, videó és kép található az adott témában, valamint egy *buzz-ometer* nevű grafikon segítségével még arról is informál, hogy percenként vagy másodpercenként hány új üzenet jelenik meg a Twitteren ezzel kapcsolatban. A találatok típusok szerint vannak rendezve és szűkíthetők is e kategóriák szerint: Web, News, Video, Images, Real-time. A teszteléshez használt egyedi üzenetet a *Stinky Teddy* először látszólag megtalálta, de utána magának ellentmondva közölte, hogy nincs megjeleníthető találat.

Néhány további hasonló kereső: *Twazzup* (*twazzup.com*), *Twingly* (*twingly.com/microblog-search*), *Topsy* (*topsy.com*) ... és akár tucatnyit lehetne még találni, ami jól jelzi, hogy a valós idejű keresés erőteljes fejlődést mutató terület lett az elmúlt néhány évben.

## Hagyományos keresőgépek

Természetesen a nagy, általános keresőgépek is igyekeznek a valós idejű tartalmakat az adatbázisukba integrálni. A Google például gond nélkül megtalálta a teszttüzenetet; a *site:twitter.com* keresőkérésre pedig 332 millió találatot adott – ezek persze tovább szűkíthetők keresőszavakkal és a Google hagyományos keresőopcióival.

E cikk írásakor, 2009 októberében – szinte egy időben – jelentette be a két nagy kereső: a *Google* és a *Bing*, hogy bekötik oldalukra a Twitter „tűzoltófecskeendőjét”, vagyis a rövid üzenetek áradatát élőben követhetővé teszik. A Bing próbaverziója már működik is a *www.bing.com/twitter* címen, de használni csak akkor tudjuk, ha *United States*-re állítjuk a földrajzi helyzetünket, vagy ha eleve olyan országban vagyunk, ahol angol a hivatalos nyelv. (Rejtély, hogy miért nem tudja a Microsoft minden nemzeti verzióban egyszerre megjelentetni az új funkciókat.) A felhasználót a keresőmezőn kívül itt is egy címkefelhő fogadja a pillanatnyilag legnépszerűbb témákkal. Nem teljesen ugyanazok a *tag*-ek láthatók itt, mint amelyek a Twitter kezdőlapján vannak – a Bing láthatóan másképp súlyozza az átvett adatokat. És más a frissítés gyakorisága is: a *Windows 7* témára keresve a Bing két perccel korábban megjelent találatokat adott, miközben a Twitter saját keresőjében a legfrissebb üzenet kevesebb mint 30 másodperces volt. Furcsa az is, hogy miközben a Twitter néhány perc alatt 376 új üzenetet jelenített meg, a Bing találati listája nem frissült (annak ellenére, hogy a Pause gomb nem lett lenyomva), csupán az eltelt percek száma változott az üzenetek mellett. Hogy a szolgáltatás még erősen béta-állapotban van, azt az is mutatja, hogy bár elvileg a rendszer a rövidített URL-eket (mint pl. a *bit.ly* kezdetűek) feloldja, ez a valóságban nem mindig történik meg, így a gyanútlan felhasználó esetleg spamoldalakra téved. A Bing ugyanis az üzenetek mellett a bennük megosztott legnépszerűbb linkeket is kilistázza, sőt ezek a listák tovább szűrhetők és rendezhetők olyan szempontok szerint, mint a hivatkozás óta eltelt idő, a visszahivatkozások száma és ezek

íróinak fontossága. Gyermekbetegségei ellenére a Microsoft Twitter-keresője jó előrelépés, miközben a Google még nem tudott megjeleníteni egy demóval sem.

A kisebb vetélytársak közül az *Icerocket* (*icerocket.com*) bezzeg gyorsan átpozícionálta magát valós idejű keresővé és egész jól teljesít. Szépen megoldotta a tesztfeladatot és egy percnél nem régebbi Twitter-üzeneteket is megtalált. A Big Buzz opciót választva egyszerre tudunk keresni blogokban, a Twitteren, a video- és képmegosztó oldalakon, valamint a hírek között. A találati listáknál kérhető automatikus frissítés és a keresőkérések elmenthetők.

## Következtetések

A valós idejű keresés még gyerekcipőben jár, nem alakultak ki a „kötelező” funkciók, különféle megoldásokkal folyik a kísérletezés. Azok a keresők, amelyek csak a Twitterre korlátozzák magukat, nem lesznek túl népszerűek, mert a felhasználók inkább a Twitter saját keresőjét használják – ha csak nem kínálnak olyan extra lehetőségeket, mint például a CrowdEye. A hagyományos keresők közül pedig még csak egy-kettő látott hozzá, hogy szolgáltatásába integrálja a valós idejű tartalmakat és funkcionalitást. Érdemes lesz figyelemmel kísérni a versenyt ezen a területen.

## Irodalom

- [1] SULLIVAN, Danny: What Is Real Time Search? Definitions & Players. = Search Engine Land, 2009. július 9.  
<http://searchengineland.com/what-is-real-time-search-definitions-players-22172>
- [2] OneRiot.com – Realtime Search for the Realtime Web. =  
<http://www.oneriot.com>

/BRADLEY, Phil: Search Engines: Real-time Search. = *Ariadne*, 61. sz. 2009. október 30.  
<http://www.ariadne.ac.uk/issue61/search-engines/>

(Drótos László)