

# Tanulmány

## WEB 2.0, INFORMÁCIÓELÉRÉS, TAXONÓMIA

Veszelszki Ágnes

PhD, Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar  
veszelszki.agnes@gmail.com

„A digitális fogyasztó kifejezetten pragmatikus: véleménye szerint, ha egy információ nem található meg azonnal, egy helyen, akkor nem is érdemes keresni” – ez a 2008-as, David Nicholas–Ian Rowlands-idézet is lehetne a *folksonomy* vizsgálati terület mottója, hiszen éppen az információ létrehozásával, keresésével és elérésével foglalkozik.

Az internetkorszak általános és nyelvészeti jellemzője, hogy az információ mennyiség a korábban tapasztalható képest elképzelhetetlen mértékben és sebességgel nő („tudásrobbanás”, Pscheida, 2007, 9., 27.). Ez a tudásradar az átláthatatlan, befoghatatlan, feldolgozhatatlan információ mennyiség miatt egyrészt erőteljes bizonytalanságerzetet okoz, másrészt viszont azt az érzést is keltheti a magabiztos netfelhasználóban, hogy minden információ online elérhető – az *à la carte*-kultúrában (Origgi, 2006, 202.) ismerni kell a keresés és szelekció megfelelő módját. Az erre képes személy Christop Meinel és Harald Sack (2004) fogalmával a *homo surfens*. A keresni tudás pedig a legegyszerűbb megfogalmazása annak a tudásfajtának, amely a 2000-

es évek társadalmának egyik kulcsfogalma: a médiakompetencia. Az információs társadalom paradoxona, hogy ugyan a (kollektív) tudásmennyiség nő, ezzel egy időben viszont – a véges befogadási kapacitás és a tudáshalmaz átláthatatlansága és strukturátlansága miatt – az egyén tudása folyamatosan csökken (Pscheida, 2007, 29.). A tudásalapú társadalom folyamatos rugalmasságot, élethosszig tartó tanulás (*life-long learning*, *lebenslanges Lernen*), tudatos metakogníciót igényel az individuumból.

A mai technológia a jelennek szól. Karácsony András (2002) a jelen felértékelődését a modern világ általános jellemzőjének tekinti („a modern világ időtudatában [...] kitüntetett a jelen idődimenziója” [2002, 136.]), ám az internet ezt a jelenséget felerősíti, sokkal látványosabbá teszi. És hogy mit is értünk a „jelen”-en? Ismét Karácsony Andrásra hivatkozom, akinek a megfogalmazásában „a »jelen« értelme arra az időtartamra utal, amelyben a jövő múlttá válik. Ez az időtartam persze lehet egy pillanat, de lehet ennél jóval hosszabb idő is. [...] Inkább a jelenre vonatkozatható él-

ményünket, azaz a szubjektív időt fejezi ki” (Karácsony, 2002, 136.).

Az infokommunikációs korban tehát új keresési technikák alkalmazására van szükség: miközben korábban egy információt a megfelelő szótár, lexikon, enciklopédia vagy szakkönyv fellapozásával lehetett elérni, addig ma elegendő valamelyik keresőportált, például a Google-t használni. Jan Hodel (2010, 26.) szerint az új keresési mód egyelőre hibrid forma, hiszen a kereséshez, adatgyűjtéshez használt kompetenciáink (mint az előzetes tudás aktiválása, a keresés kivitelezése) még az analóg korból származnak, amelyeknek azonban már a digitális kor kihívásaihoz kell alkalmazkodniuk. Az előzetes tudás feltételül szükséges ahhoz, hogy a keresőszoftver által kiadott találatokat értékelni és rangsorolni tudjunk.

A kereséshez kötődő *folksonomy* információtudományi fogalmat magyarul pontosan, az angol nyelvi kontamináció megbontása nélkül nem lehet visszaadni. A terminust 2004-ben használta először egy konferencián Thomas Vander Wal, majd Gene Smith tanulmánya alapján vált ismertté (Vander Wal, 2004, Smith, 2004; Peters, 2009, 154.), és a *folk* ’nép’, illetve a *taxonomy* ’csoportosítás, osztályozás’ szavak vegyítéséből származik (Peters, 2009, 5.), ily módon a „nép”, vagyis a nem professzionális indexelők által létrehozott taxonómiára utal. A folksonomy szabadon választott, bármely információhoz köthető kulcsszavakból vagy címkékből (ún. tagekből) áll – e címkéket a post-itelek elektronikus megfelelőinek is szokás nevezni (Müller-Prove, 2008, 15.; Peters, 2009, 153.). A kollaboratív tagelés az a folyamat, amelynek során a felhasználó a megosztott tartalomhoz szabadon választott kulcsszavak formájában metaadatokat kapcsol. A tageléssel a tartal-

omlétrehozó indexelő is lesz egyben. A *folksonomy* fogalom szinte felsorolhatatlanul sok szinonimával rendelkezik; csak válogatásként a számos elnevezés közül: demokratikus indexelés (*democratic indexing* [Kellogg Smith, 2006]), közösségi klasszifikációs rendszer (*social classification system* [Feinberg, 2006]), etnoklasszifikáció (*ethnocalssification* [Star, 1996]), felhasználói metaadat (*user-generated metadata* [Dye, 2006]), tagszféra (*tagosphere* [Gruber, 2007]), tagszonómia (*tagsonomy* [Hayes et al., 2007]).

A folksonomy megjelenése a web 2.0 szociális hálózataihoz köthető. A web 2.0 – a fogalom 2006-os bevezetése óta – sokat használt információtudományi kulcsszóvá vált. Az ún. web 1.0 verziótól technológiailag, szerkezetileg és szociális szempontok szerint különbözik; egyesek szerint a korábbi objektumközpontú megközelítés helyét a személycentrikus hálózat veszi át (Peters, 2009, 15.). A web 2.0 információtechnológiai korszakban nem csupán professzionális tartalomfejlesztők (újságírók, írók, webdizájnerek, honlapfejlesztők stb.) hozhatnak létre és oszthatnak meg az interneten digitális tartalmakat, hanem gyakorlatilag bármely számítógéphasználó megteheti ezt. A szövegalkotási és -olvasási mód is megváltozott a gépi szövegbeviteli, valamint az internetes szövegkiadás következtében. A számítógép ugyan az írógép billentyűzetével rendelkezik, de a szöveget nem közvetlenül a papírra írjuk, a protoforma a gépben marad. Az anyagtalanság miatt lehetséges a nyomtalan törlés, a beszúrás, a nem lineáris szerkesztésmód. A korábbi lineáris olvasási mód helyett a cirkuláris olvasás a jellemző a hiperlinkes kapcsolatokkal létrejött szövegek befogadására.

A felhasználók nem csupán tartalmakat (szövegeket, képeket, videókat stb.), hanem

az ezekhez tartozó metaszintű információkat, címkéket, ún. *tageket* is létrehozunk; ezzel az indirekt kooperációval hozzájárulnak az információk megtalálhatóságához. Az információk metainformációval való ellátása nem új találmány: a könyvtár- és információtudomány egyik alapfoglatosságáé többek között a tárgykatalógusok, a különféle klasszifikációs rendszerek és tezauszok elkészítése. A professzionális osztályozási rendszerektől a folksonomy két szempontból különbözik: egyrészt megközelítési módja, másrészt a terminológiához való viszonya tér el attól. Míg a hagyományos taxonómia előbb kiválasztja az osztályozási kritériumokat, létrehozza a rendszer(ek)e)t, és azt tölti fel tartalommal, addig a folksonomy esetében ez éppen fordítva történik: a kritériumközpontú helyett a forrásközpontú strukturálás jellemzi. A másik, még erőteljesebb eltérés a terminológiát érinti: a hagyományos, ún. professzionális meta-információ-létrehozás során szigorúan szabályozott, előírt terminusokat lehet csak használni, ezzel szemben a folksonomy ellenáll a korlátozott terminológiahasználatnak. Így viszont el is jutottunk a folksonomy egyik gyenge pontjához: éppen szabályozatlansága miatt ki van téve a nyelvi idioszinkráziának (például a homonímiának és a szinonímiának), ami az információkeresés és -elérés során problémákat jelenthet.

Az informatika, az információtudomány, a szociológia, a kommunikáció- és médiatudomány, a nyelvészet és a könyvtártudomány metszéspontjában álló témakörrel 2006 óta foglalkozik a tudományos kutatás (ez nem véletlen; a folksonomy-alapú webes szolgáltatások 2002–3 környékén jelentek meg), ám összefoglaló, a *folksonomy* elméletét és a gyakorlati használati módokat, illetve előnyeit, hátrányait is taglaló monográfia nem született

Isabella Peters *Folksonomies* című 2009-es munkája előtt. E kötet fő célkitűzése, hogy feltárja azokat a területeket, amelyekben a bevett tudásreprezentációs és információelérési módszerek hasznosíthatják a folksonomies előnyeit (és *vice versa*, vagyis hogyan tehető hatékonyabbá a folksonomy például a félautomata egységesítő rendszerek használatával vagy a szabványosított nyelvhasználattal).

A közösségi információszolgáltatási rendszerekhez (collaborative information services) kötődő legfontosabb fogalmak a web 2.0, a közösségi szoftver, a kollaboratív információs rendszerek. E szolgáltatások (azaz a social bookmarking, e-kereskedelem és kereskedelmi információszolgáltatás, zene-, kép- és videómegosztó oldalak, könyvtár 2.0, közösségi hálózatok, blogok és blogkereső motorok, tagelő játékok) közé tartoznak az olyan online könyvjelzőgyűjtemények, mint a del.icio.us vagy a tudományos célokra használható Bibsonomy; továbbá többek között az online kereskedő Amazon. A Last.fm zene-, a Flickr fotó- és a Youtube videómegosztó kapcsán felmerül a kérdés, hogy kevés szöveges információval (címmel) ellátott, nem textuális forrásokat miképpen kell/lehet szöveges metaadatokkal ellátni.

A tudásreprezentáció (*knowledge representation*), valamint az információelérés és -kinyerés (*information retrieval*) módjai között elkülönítik egymástól a keresés (*searching*), a böngészés (*browsing*) és az információkinyerés (*retrieving*) magatartáshármasát (Peters, 2009, 287–293.). Sajátos fogalom a *tagging behaviour* (tagelő magatartás), amely az indexelés, azaz a tagek, címkék létrehozásának szociális és kognitív oldalát vizsgálja: miért és milyen elvek alapján készít egy felhasználó címkéket, illetve milyen kognitív erőfeszítés szükséges a tageléshez. A tagelés kognitív

mechanizmusa (Sinha, 2005 alapján) a konceptualizációra hasonlít: első lépésben a felhasználó összehasonlítja a címkézendő objektumot más, elméjében tárolt fogalmakkal, amelyek mentálisan aktíválnak számos, hasonló vagy kapcsolt fogalmat, majd kiválasztja ezek közül a legmegfelelőbbet, és elkészíti a címkét – a konceptualizáció utolsó lépése, a kategorizáció a tagelés esetében elmarad.

A web 2.0-s indexelés tipikusan a kollektív intelligencia (*collective intelligence*, más megfogalmazásokban: *the wisdom of crowds* „a tömegek bölcsessége”, *crowdsourcing*) használatához kötődik: olyan különböző emberek működnek decentralizáltan együtt, akiknek sem közös tudásuk, sem (valódi) közös céljuk nincsen, de együtt intelligensebbek, mint külön-külön (vö. Surowiecki, 2005). Itt érdemes megemlíteni a hálózatkutató Barabási Albert-László 2012 nyaratól folyó projektjét (URL1), amelyben az általa vezetett kutatás eredményeit folyamatosan közzéteszik az interneten, és a hozzászólások alapján változtatnak rajta. Az idegen nyelvekre való fordítást szintén a *crowdsourcing* módszerrel tervezik, egy próbafordítás után gyakorlatilag bárki fordíthat fejezeteket a készülő kötetből.

A tömegek bölcsessége azonban gyakorta problémákhoz, hibás címkékhez, ismétlődésekhez vezet. Megfigyelések szerint minél nagyobb a csoport, annál több hiba fordulhat elő a címkék között. Ennek tudatos elkerülésére szolgál a darwini metaforával *tag gardening*-nek, „címkékertészkedésnek” nevezett folyamat, a címkék – elsősorban utólagos – rendezése, rendben tartása, strukturálása, a kétértelműségek kiküszöbölése (Peters, 2009, 235–247.).

A felhasználó generálta digitális metaadatok, a tagek legfontosabb előnyei közé tartozik, hogy friss információkat jelenítenek meg,

nem szükséges elkészítésükhöz előzetes standardizáció, a minőségértékelő tagek hozzáadott értéként jelennek meg, és a tagek nyelvi szempontból rugalmasak. Éppen ez utóbbi jelenti azonban a legfőbb hátrányukat: az emberi nyelv variabilitása (például szemantikai szempontból a homonimák, szinonimák, alá- és fölérendeltségi viszonyok; írástechnikai szempontból a különböző írásváltozatok, esetleg helyesírási hibák) és a használt szókincs ellenőrizetlensége megnehezítik és lelassítják a keresési folyamatot. Mindezen problémák feloldására és a használhatóság, a keresési hatékonyság növelésére félautomatikus vagy manuális javításra, a címkerendszer tudatos újrastrukturálására van szükség. A tageknek kiemelkedő szerepük van a multimédiás digitális tartalmak megtalálásában: hiszen a minimális információt tartalmazó cím mellett a keresőrendszerek számára használható verbális adatokat nyújtanak, vagyis új elérési utakat nyitnak.

A könyvtár az egyik alapvető forrásgyűjtő és metaadat-létrehozó intézmény, amelynek szintén szembe kell néznie az internet hozta kihívásokkal: ezt a koncepciót könyvtár 2.0-nak szokás nevezni. Isabella Peters (2009, 56.) meglátása szerint a 2.0-s könyvtárak két stratégiát követhetnek: 1. munkatársaik a különböző közösségi hálózatokban az információkinyerés, az értékes információforrás kiválasztásának szakértőiként vesznek részt; 2. a másik módzat bevonja a könyvtárhasználókat a metaadatok létrehozásába, vagy a könyvtár saját (OPAC-)rendszerébe címkékészítő funkciók integrálásával, vagy külső, internetes tagelő platformok használatával (például: LibraryThing, BibSonomy).

Kulcsszavak: *digitális kommunikáció, indexelés, információtudomány, keresés, netnyelvészet*

## IRODALOM

- Dye, Jessica (2006): Folksonomy: A Game of High-tech (and High-stakes) Tag. *EContent*. April, 38–43. • <http://www.columbia.edu/cu/libraries/inside/units/bibcontrol/osmc/dye.pdf>
- Feinberg, Melanie (2006): An Examination of Authority in Social Classification Systems. In: *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*, Austin, Texas • [https://www.google.hu/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDQQFjAA&url=http%3A%2F%2Fjournals.lib.washington.edu%2Findex.php%2Facro%2Farticle%2Fdownload%2F12490%2F10989&ei=uMAMUZmwDcGYhQfvkIDYQC&usq=A FQjCNEbtpB4GzE6Zp54OvyjG2\\_7BaKBHQ&sig2=mthnJGTAukZ5\\_H2RrMo8Bw](https://www.google.hu/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDQQFjAA&url=http%3A%2F%2Fjournals.lib.washington.edu%2Findex.php%2Facro%2Farticle%2Fdownload%2F12490%2F10989&ei=uMAMUZmwDcGYhQfvkIDYQC&usq=A FQjCNEbtpB4GzE6Zp54OvyjG2_7BaKBHQ&sig2=mthnJGTAukZ5_H2RrMo8Bw)
- Gruber, Thomas (2007 [2005]): Ontology on Folksonomy: A Mash-up of Apples and Oranges. *Int'l Journal on Semantic Web & Information Systems*. 3, 2., Originally published in 2005, this is a rebuttal to a popular anti-ontology blog, with a constructive call to action. • <http://tomgruber.org/writing/ontology-of-folksonomy.htm>
- Hayes, Conor – Avesani, P. – Veeramachaneni, S. (2007): An Analysis of the Use of Tags in a Blog Recommender System. In: *Proceedings of IJCAI 2007 Hyderabad, India*, 2772–2777. • <http://ijcai.org/papers07/Papers/IJCAI07-445.pdf>
- Hodel, Jan (2010): Recherche: Google – and Far Beyond. In: Gasteiner, Martin – Haber, Peter (Hrsg.): *Digitale Arbeitstechniken für die Geistes- und Kulturwissenschaften*. Böhlau, Wien–Köln–Weimar, 25–37.
- Karácsony András (2002): Individualitás a nomádok földjén. Pillanatképek. In: Nyíri Kristóf (szerk.): *Mobilközösség – mobilmegismerés. Tanulmányok*. MTA Filozófiai Kutatóintézete, Budapest, 127–138.
- Kellog Smith, Martha (2006): Viewer Tagging in Art Museums: Comparisons to Concepts and Vocabularies of Art Museum Visitors. In: *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*, Austin, Texas • <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.3019>
- Meinel, Christoph – Sack, Harald (2004): *www. Kommunikation, Internetworking, Web-Technologien*. Springer
- Müller-Prove, Matthias (2008): Modell und Anwendungsperspektive des Social-Taggings. In: Gaiser, Birgit – Hampel, T. – Panke S. (eds.): *Good Tags – Bad Tags: Social Tagging in der Wissensorganisation*. Waxmann, Münster–New York–München–Berlin, 15–22. • <http://www.waxmann.com/fileadmin/media/zusatztexte/2039Volltext.pdf>
- Nicholas, David – Rowlands, Ian (2008): In Praise of Google. *Library & Information Update*. December, 44–45.
- Origgi, Gloria (ed.) (2006): *text-e. Text in the Age of the Internet*. Palgrave Macmillan, Houndmills–New York, 14–15.
- Peters, Isabella (2009): *Folksonomies. Indexing and Retrieval in Web 2.0. (Knowledge & Information. Studies in Information Science)* De Gruyter Saur, Berlin • <http://books.google.hu/books?id=HBYg36gbnegC&printsec=frontcover#v=onepage&q&f=false>
- Pscheida, Daniela (2007): *Internetkompetenz von Erwachsenen. Medienpraxis der Generationen*. Krämer, Hamburg
- Sinha, Rashmi (2005): *A Cognitive Analysis of Tagging (or How the Lower Cognitive Cost of Tagging Makes It Popular)*. • <http://rashmishinha.com/2005/09/27/cognitive-analysis-of-tagging/>
- Smith, Gene (2004): *Folksonomy: Social Classification*.
- Star, Susan Leigh (1996): *Slouching toward Infrastructure*. Digital Libraries Conference Workshop. • [http://is.gseis.ucla.edu/research/dig\\_libraries/star.html](http://is.gseis.ucla.edu/research/dig_libraries/star.html)
- Surowiecki, James (2005): *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Anchor Books, New York
- Vander Wal, Thomas (2004): *Feed on This*. • <http://vanderwal.net/random/entries/id.php?blog=1562>  
URL: [http://network.blog.hu/2012/07/30/fordisuk\\_kozosen\\_barabasi\\_albert-laszlo\\_uj\\_konyvet\\_kozosseg\\_kollaboracio\\_crowd\\_sourcing](http://network.blog.hu/2012/07/30/fordisuk_kozosen_barabasi_albert-laszlo_uj_konyvet_kozosseg_kollaboracio_crowd_sourcing)

# INFORMÁCIÓÁRADAT ÉS HULLÁMLOVAGLÁS

Holl András

csillagász, informatikus  
MTA Csillagászati és Földtudományi Kutatóközpont  
Konkoly Thege Miklós Csillagászati Intézet, MTA Könyvtára  
[holl.andras@konyvtar.mta.hu](mailto:holl.andras@konyvtar.mta.hu)

## Bevezetés

A tudományos szakirodalom exponenciális növekedését Derek de Solla Price ismerte fel a 20. század közepén, a *Philosophical Transactions of the Royal Society* majd két évszázad során publikált köteteit időrendben feltornyozva (lásd Ekers, 2009). Michael J. Kurtz és Edwin A. Henneken (2012) évi 3,5%-os növekedést említ: az évente publikált cikkek mennyisége húsz év alatt megduplázódik. Philip Young (2009) szerint a növekedés üteme évi 3%. A kutatói elme kapacitása feltehetően nem növekszik, ám a kutatók száma igen. Ennek következménye az egyre erőteljesebb specializálódás. A növekvő ismeretanyag kezelése – ha nem fogadjuk el a növekvő elszigetelődést, pontosabban a csupán kevésbé kapcsolódó szakterületek kialakulását, ha nem adjuk fel a „teljes kép” áttekintésének igényét – egyre nagyobb kihívást jelent mind a kutatóknak, mind a kutatást irányítóknak.

A nehézségek már régóta nyilvánvalóak a tudományos szakkönyvtárak számára. A tudományos folyóiratok előfizetési díjai gyorsabb ütemben növekszenek, mint amekkorát az infláció indokolna (Peine, 2010) – tegyük hozzá, az infláció és a terjedelmnövekedés

együttese által indokoltnál is.<sup>1</sup> A tudományos folyóiratok kiadását nem szabályozza hatékonyan a piac: mivel az egyik folyóirat nem helyettesítheti a másikat, minden cikk egyedi (Young, 2009). Nem a költségvetés jelenti a könyvtárak egyetlen gondját: azokat a folyóiratokat, amelyeket meg tudnak rendelni, egyre nehezebb a polcokon elhelyezni.

A tudományos adatok növekedése is exponenciális, de a növekedési ütem a szakirodalom növekedésénél sokkal gyorsabb. Ennek oka, hogy a kutatók számának növekedésénél sokkal gyorsabban emelkedik az adatokat gyűjtő műszerek és számítógépek száma, ráadásul az adatgyűjtő képességük is rohamosan nő.<sup>2</sup>

Az adatokra is igaz, amit a szakirodalomra nézve leírtunk: a tárolás, feldolgozás megoldható a felhasznált számítógépek számának növelésével – a sok gépre, sok merevlemezre elosztva gyorsabb ütemben növekedhet a

<sup>1</sup> A 2007-es árnövekedés 7% Young (2009) szerint, Mike Peine (2010) az 1989 és 2011 közötti árnövekedésre hasonló adatot közöl az USA-ban kiadott folyóiratokra.

<sup>2</sup> A mindennapi életben is megfigyelhető jelenség a CCD-csipek növekedése a fényképezőgépekben és a telefonokban, ami együtt jár a képek méretének növekedésével.