

**Webkettő és webhárom. Okosodó hálózatok;**  
**Természet Világa, 2012 ápr., 183. o.**



## **Web 2.0 - Web 3.0**

*Holl András*

Ahhoz, hogy a Web 2.0-át és Web 3.0-át bemutathassuk - az előbbi esetében inkább csak új szemszögből való bemutatásról beszélhetünk - vissza kell tekintsünk a Web „1.0”-ára. Tim Berners-Lee 1989-ben fejlesztette ki a CERN-ben, a tudósok közötti együttműködés megkönnyítésére, elosztott, egymáshoz kapcsolódó dokumentumok kezelésére. 1993-ban a CERN szabaddá tette a web technológiát, és megszülettek az ugyancsak szabadon elérhető és a gyakorlatban alkalmazható web-böngészők és szerverek. Hamar felfedezték a webet azok a tudományos szolgáltatások, amelyek eddig korábbi Internet-technológiákat használtak, vagy éppen születőben voltak. Ilyen volt a SPIRES, a Stanford Egyetem fizikai, majd az ADS, a NASA/SAO csillagászati bibliográfiai adatbázisa, az xxx (ma már arXiv néven ismert) fizikus különnyomat lerakat, a Strasbourgi Observatóriumban épített SIMBAD adatbázis, vagy az Information Bulletin on Variable Stars (IBVS) az MTA Csillagászati Kutatóintézetében. (A fizikusok és a csillagászok élen jártak, ám más területekről is lehetne példákat hozni. A szerző a hozzá közel állókból válogatott.) 1994 végére a web-szolgáltatók száma elérte a tízezret.

A Webet úgy jellemezték, hogy demokratikus: csekély beruházás kellett egy webszerver elindításához, így sok intézmény megengedhette magának. Arra, hogy az intézményeken túl magánemberek is részt vehetnének a web építésében valószínűleg nem sokan gondoltak ekkor. A Web hamar kinőtt tudományos bölcsőjéből, iparaggá vált. Mára webszerverek működnek a nyomtatókban, lassan talán a hűtőszekrényben is. Felhasználói közül sokak számára ez az Internet.

## Web 2.0

A Web 2.0-át nem szükséges bemutatnunk: a közösségi oldalak, tartalommegosztás, wikipédia, blogok bizonyára minden olvasónak ismerősek. Amíg az eredeti Web esetében a tartalom létrehozói és olvasói jól elválaszthatóak voltak, a Web 2.0-ánál mindenki részt vehet a tartalom létrehozásában. Magát a Web 2.0 kifejezést Tim O'Reilly egy 2004-es konferencián a „dotcom” válság után újjáéledő webre alkalmazta.

A megújult „2.0-ás” világháló a tudományban is szerepet kap - de ez esetben a viszony gyakorta fordított: a széles körben elterjedt, civil alkalmazástípusok számára keresnek akadémiai felhasználást. A Wikipédia az elterjedt közösségi oldalakat felsoroló táblázatban négyről írja, hogy tudományos, kutatók számára hozták létre. 2011 szeptemberében a Kutatók Éjszakája rendezvénysorozat része volt a Kutatói Blog és Weboldal Találkozó.

A kutatók is emberek, barátkozni, fecsegni akarnak egymással, vagy éppen szeretnék kötetlenül megosztani gondolataikat az intézeti kávézón kívül is. De lehet-e szerepe a Web 2.0-ának a tudományos kutatásban, az oktatásban? Erre is akad példa szép számmal. A repozitóriumokba - az arXiv-ba és utódaiba - a szerzők töltik fel cikkeiket. A nagyobb tudományos projektek gyakran wikik segítségével építik tudásbázisukat - ahogyan a Web kezdetén ezt elképzelték. A már említett ADS csillagászok és csillagász-könyvtárosok önkéntes segítségét veszi igénybe a beszkenelt folyóiratkötetek cikkekre való szétdarabolásához.

A kutatóknak már az egyetemen hozzá kell szokniuk a szakirodalom rendszerezéséhez, kijegyzeteléséhez, hivatkozáslisták kezeléséhez, jegyzetek és tankönyvek másolásához és kipreparálásához, szövegrészek kiemeléséhez. És ki más segíthetne ebben jobba, mint egy szorgalmasabb évfolyamtársnő (vagy évfolyamtárs)? Ma már nem csak a folyosón, a kollégiumban vagy a szomszéd irodában találhatunk segítségre, hanem a Világhálón is. A Mendeley segít. Ingyenes hivatkozáskezelő, szakirodalom-rendszerező és tudományos kapcsolati hálózat, majd 130 millió cikkel és nagyjából századannyi felhasználóval.

Nem csak a profik, az amatőrök is részt vehetnek a tudományban az Internet, a Web 2.0-a körébe tartozó eszközök segítségével. Számos tudományos program toboroz önkéntes segítőköt, végeztet számításokat kalákában a résztvevők számítógépein (lásd: Amatőr kibertudomány, *Természet Világa*, 2011 december).

## Web 3.0

A 3.0-ás Web nem a 2.0-ás továbbfejlesztett változata, hanem a számítógépek, az adatok hálója. A végső felhasználó az ember, de az egyes weboldalakat már az érdekében eljáró elektronikus ügynökök, robottitkárok olvassák helyette. Tim Berners-Lee 2001-ben szemantikus webnek nevezte a Scientific American-ben megjelent cikkében. Berners-Lee példája szerint a szemantikus weben működő elektronikus ügynökök képesek kell legyenek arra, hogy önállóan felderítsék a weben elérhető információkat és szolgáltatásokat, meggyőződjenek a megbízhatóságukról, és megértsék azokat, vagy éppen szót értsenek velük, gazdájuk ügyeit intézve. Noha a szemantikus web kiépítéséhez már 2001-ben rendelkezésre állt néhány technológia (XML, RDF), és azóta továbbiak születtek, még ma sem beszélhetünk elterjedt, működő rendszerről. A tudományban viszont számos olyan szolgáltatás

működik, van legalábbis tervezés alatt, ami a szemantikus web tulajdonságainak egyikével vagy másikkal már rendelkezik, akár a szemantikus web technológiáit, vagy valamilyen egyszerűbb, szűkebb körben használatos, éppenséggel egyedi megoldást alkalmazva. Ezek közül mutatunk be néhányat.

Az MIT Media Lab-jában Nicholas Negroponte már húsz éve robottitkárokon gondolkodott, akik az elektronikus napisajtóból gazdájuk számára összeollózzák az ízlésének megfelelő cikkeket. Ma „mashup”-nak nevezzük ezt (webkollázs talán?). A webkollázsok automatikusan, a keresés pillanatában állítják elő a kért oldalt, különböző internetes forrásokból összeollózva. A tudomány területén is szükség van ilyen kollázsokra: a csillagászok felfedeznek egy változócsillagot (vagy mindjárt tízezret egy program keretében), alapvető adatokat és ábrákat közölnek róla egy cikkben (vagy egy az adott programhoz kapcsolódó adatbázisban). Szükség van viszont egy nagy katalógusra, ami minden ilyen csillagot tartalmaz, és mindegyikhez a megfelelő ábrákat is elérhetővé teszi, ne kelljen a következő kutatónak, aki a csillagba „belebotlik”, előlről kezdenie mindent, vagy átnyálaznia az egész szakirodalmat. A már említett hazai folyóirat, az IBVS felkészült erre az igényre, és az egyszer talán megvalósuló kollázs-katalógus számára megfelelő formában is közreadja a cikkekben megjelent ábrákat. A jelenlegi szolgáltatáshoz azonban tudni kell, az újság melyik számában, hányadik ábráról van szó. Igazi Web 3.0-ás szolgáltatás akkor lesz ebből, ha a kollázs-katalógus automatikusan fel tudja majd deríteni, hol található ábra az adott csillagról a weben. Egy webkollázs-kísérlet az Ispecies.org. Egy állat- vagy növényfaj latin nevét begépelve fotókat, cikkeket, elterjedési térképet kapunk. A webkollázs enciklopédistái nem maguk gyűjtögetik az információkat, inkább programoznak. Hiszen rengeteg információ található a weben, és a keresők egy pillanat alatt megtalálnak mindent... A jó webkollázshoz persze, több kell: ismerni kell a megbízható forrásokat, és megfelelő metaadatok kellene. (A Web 3.0-ás webkollázs automatikusan győződik majd meg a megbízhatóságról, valahogy úgy, ahogy az elektronikus aláírás működik: tanúsítók láncolatán keresztül, ahol csak a lánc végén lévő kell megbízhatónak tudni. A metaadatok nélkül pedig azt kapjuk, amit a népszerű keresők a képernyőkre öntenek: tüt, szénakazlastul.)

Az IBVS, más fejlett tulajdonságokkal is rendelkezik. Ma már minden tudományos szaklapnál követelmény, hogy az irodalomjegyzék a webes változatban hiperszöveges legyen, az olvasó egy (két) kattintással továbbléphessen a hivatkozott cikkekre. A legtöbb lapnál ezeket a hiperhivatkozásokat kézzel kell a szerkesztőségnek beilleszteni - nem így az IBVS-nél! A Strasbourg-i CDS-ben fejlesztett technológiát felhasználva a referencia-hivatkozások reptében, automatikusan készülnek el a szerző által megadott, az (emberi) olvasó számára szerkesztett irodalomjegyzékből. A kereső is többet tud a megszokottnál. Ha az „Augusta Ada Byron” karaktorsorozatot írjuk be egy keresőbe, nem találjuk meg azokat az oldalakat, ahol az első számítógépes algoritmus szerzőjét „Ada Lovelace”-ként említik. Az IBVS szerzői is többféle néven, nevüket különféle módon írva közöltek cikkeket. Egy-egy csillagnak pedig akár többtucatnyi különböző elnevezése lehet. Az IBVS olvasója nem csupán egy karakterláncra kereshet, hanem egy szerzőre, egy csillagra. A különböző elnevezéseket ráadásul nem csupán egy helyi szótárból tudja a keresés tárgyára kigyűjteni a rendszer, de külső szolgáltatásoktól is segítséget kérhet. A szerzők különböző névváltozatait a NASA/SAO ADS, a csillagok különböző neveit a strasbourgi CDS és a Moszkvában készített elektronikus GCVS adja meg az IBVS keresője számára. Szerző vagy csillag, bármilyen néven keresik is, megtalálják. Az olvasónak az eredmény a fontos, ám itt most fontosabb, hogy a kis szaklap nem mindent maga próbál megoldani, hanem internetes szolgáltatások hálózatára támaszkodik olvasói kiszolgálásában.

Az IBVS nem csak a keresőnél vesz igénybe külső segítséget. Térképi információk megjelenítésére és manipulálására a CDS Aladin szolgáltatását használja. A - mondjuk bostoni - olvasó a cikkben elhelyezett hiperhivatkozásra kattint, és a két (budapesti és strasbourgi) szolgáltatás között lezajlott

gépi párbeszéd után a képernyőjén felbukkan az Aladin ablaka az IBVS-ből származó adatokkal a CDS-ből érkező térkép háttére előtt. A kiszolgálók közötti adatsere a Virtuális Observatórium (VO) XML-es adatformátumában (VOTable) történik. Az Aladin egyébként nem csupán a weben áll szóba más alkalmazásokkal: a PLASTIC/SAMP protokoll segítségével könnyedén cserél adatokat a felhasználó gépén futó más, VO-képes programmal.

The screenshot shows the CDS Aladin software interface. The main window displays a star field with two objects labeled **NOMAD1\_1281-0377168** and **NOMAD1\_1281-0377481**. The interface includes a menu bar (File, Edit, Image, Catalog, Overlay, Tool, View, Interop, Help), a toolbar with various tools (select, pan, zoom, etc.), and a search bar. Below the main window is a table of object data.

Name	RA	DEC	V	B-V	V-Rc	V-Ic
<input type="checkbox"/> NOMAD1 1...	288.3583...	+38.2699...	12.158	0.691	0.412	0.808

At the bottom of the interface, there is a copyright notice: (c) 2010 UDS/CNRS - by CDS - Distributed under GNU GPL v3 and a status bar: 1 sel / 4 src 16Mb.

### *IBVS adatok megjelenítése a CDS Aladin segítségével*

A csillagászok Virtuális Observatórium projektje az adatözön hullámaint lovagolja meg. Lehetővé teszi, hogy a kutató a világ különböző részein található adatokat együtt elemezze, adatbányászatot végezhesen az óriási katalógusokban. Már az egyik prototípus VO alkalmazás próbája egy új barna törpe csillag felfedezéséhez vezetett 2003-ban, az SDSS és a 2MASS katalógusok összevetésével. VO alkalmazások segítik majd a néhány év múlva üzembeállító LSST adatainak feldolgozását is. Ha a z eget pásztázó óriástávcső érdekes objektumot talál, az adatfeldolgozó programlanc egy VOevent üzenetet - a célra fejlesztett XML formájú üzenetet - küld, és valahol egy másik observatóriumban egy teleszkóp megszakítja beütemezett megfigyelési programját, és rááll az új, érdekes objektumra. Az

észlelő csillagász csak nézi az önálló életre kelt távcsövet, és az értesítést a monitoron, valahogy így: „Megfigyelési programját rövid időre megszakítjuk magas prioritású célobjektum váratlan felbukkanása miatt.”

Korábban szót kerítettünk már az elektronikus szakfolyóiratok irodalomjegyzékére. Az olvasó néhány kattintás után már olvashatja is a hivatkozott cikket - már amennyiben joga van hozzá, mondjuk, az intézménye internetes címtartományára érvényes előfizetéssel rendelkezik. Ha nem, csak egy üzenetet kap, miszerint megvásárolhatja az adott cikk letöltési jogát. Ám előfordulhat, hogy a cikket saját kollégája írta, és egy példány megtalálható a helyi cikkadatbázisban, az intézményi repozitóriumban. Az OpenURL rendszer segítségével a kiadói weblapról elért cikkben a hiperhivatkozás egy helyi feloldó szolgáltatás közbeiktatásával készül el, a helyi elérési lehetőségek figyelembevételével. Az olvasó kattint, és ha lehet, az intézményi repozitóriumból vagy éppen a nemzeti előfizetési konzorcium által előfizetett archívumból jut hozzá a cikkhez.

A tudományos szakcikket érdemes repozitóriumban elhelyezni, nem csupán kitenni a tanszéki weboldalra. A repozitórium leíró adatokat (metaadatokat) tartalmaz, és ezeket az OAI-PMH protokollon keresztül terjeszti is. A PhD dolgozatokat kínáló repozitóriumokat a DART-Europe portálon keresztül kereshetjük. Ha majd a DART automatikusan fedezi fel a PhD-dolgozat gyűjteményeket, és győződik meg megbízhatóságukról (valóban akkreditált doktori iskolához tartoznak-e?), igazi Web 3.0-ás szolgáltatás lesz majd. Az automatikus, többnyelvű plágiumkeresés sem merész álom ma már, egyszerű talán ez is a szolgáltatás része lehet majd.

A tudományos adatözön publikációs áradatot kelt. Egy kutató sem tud ma már minden cikket elolvasni (tágabb) kutatási területén. Kinek-kinek meglehetnek a módszerei a tájékozódáshoz, ám a például a csillagászok már segítséget is kaphatnak az ADS-től. Ha találnak egy cikket a szakirodalmi adatbázisban, rögtön listát kaphatnak a hasonló cikkekről. (Ha a cikkeket vektorokként ábrázoljuk egy sokdimenziós térben - annyi dimenzió, ahány szó van a szótárban, a vektor minden koordinátája pedig az adott szó előfordulásainak száma -, könnyen megtalálhatjuk a hasonló cikkeket, azaz hasonló irányba mutató vektorokat). De feltehetünk olyan kérdéseket is, melyik a legfontosabb cikk az adott témában? (Amelyiket a legtöbben idéznek.) Vagy melyikkel kell kezdeni a témával való ismerkedést? (Amelyiknek az irodalomjegyzékében a legtöbb szerepel a fontos cikkek közül.) Ezek a másodrendű operátorok - a Google Page Rank algoritmushoz hasonlóan - a dokumentumok közötti kapcsolatokat elemezve nyújtanak segítséget az olvasónak.

A hazai kutatók egyre inkább rászorulnak az MTMT szolgáltatásaira pályázatírás, habilitálás, jelentéskészítés közben. Ha a kutató pályázik, az OTKA elektronikus rendszerében meg kell adnia legutóbbi publikációi közül a legfontosabbakat. A doktori iskolákban oktatók publikációs listáját is megtekinthetik az érdekelt hallgatók. Az MTMT már most kapcsolatban áll ezekkel az adatbázisokkal, de a tervezés alatt álló új MTMT szoftver mindezeket az információkat már webszolgáltatásokon keresztül nyújtja majd - egy kis lépéssel közelebb a Web 3.0-hoz.

A 3.0-ás web létrehozásához szükséges technológiák mára jobbára már rendelkezésre állnak, ám szükség van adatokra is. A nyílt, strukturált, értelmezhető adatok előállítását a Linked Open Data (LOD) projekttől várhatjuk. Ahogyan az IBVS feltárja és adatbázis-szerűen kereshetővé, összekapcsolhatóvá teszi a cikkekben megjelent ábrákat, a járulékos adatállományokat, a DBpedia a Wikipédia bizonyos adataival teszi ugyanezt - csak éppen sokkal általánosabb módon. A Web 3.0-hoz felhasználható adatok mennyisége is egyre nő.

