

Ralf Thomas Göllner

20. századi kapcsolattörténeti szöveggyűjtemény az MMI adatbankjában¹

A nyolcvanas évek táján elindult egy olyan digitális, informatikai és technológiai forradalom, amelynek eredménye a tudás Gutenberg óta nem tapasztalt mértékű megsokszorozódása és elterjedése. De ez a forradalom nemcsak a tudás megsokszorozódásának egyengette az útját, hanem új kutatási, oktatási és tanulási módszereknek is. Az internet, az adatbázisok és az információs szolgáltatások napjainkban fontos és a kommunikáció, a kutatás és a könyvtárak ma már nélkülözhetetlen eszközei. Ezek a lehetőségek gyorsan, majdnem mindig és mindenhol rendelkezésre állnak, és nemcsak a tudományos munkát gyorsítják fel. Ezen fejlődés során létrejött a huszonegyedik századi információs társadalom és a hozzá tartozó ipar, amelyik – mint Google, Yahoo és más nemzetközi és pénzügyi szempontból kiválóan működő cég – információkat közvetít, de nem mindig biztosít. Jelenleg ezek a cégek (főleg a Google és a Wikipedia) nemcsak arra készülnek, hogy öncélúan, az eredeti tudományos, könyvtári és levéltári szakértelmet megkerülve orvvadász módra közvetítsenek információkat, hanem hogy feldolgozott és strukturált ismeretanyagot kínáljanak a használóknak. Jó példa erre a Google könyvtári digitalizációs kezdeményezése.

E rövid helyzetfelmérés rámutat arra, hogy a tudományos, könyvtári, levéltári és múzeumi szakértelmek éppúgy, mint a tudásközvetítés, amely sokkal összetettebb, mint az információközvetítés és a tudáshitelesítés, nagy intellektuális, technikai és persze finanszírozási kihívások előtt áll. Ha a könyvtári vagy más tudományos célú gyűjtemények nem akarják ezeket a főkompetenciákat és feladatokat a technológiai fölényrel bíró, profitorientált cégeknek átengedni, akkor az új technikákat és az új lehetőségeket mindenképpen fel kell használni tudományos célokra. Másképpen elveszítik a versenyt még akkor is, ha a tudományos eredmények jobbak, alaposabban, mint például a Wikipedia-cikkek, amelyek újabbban még egyetemi munkákban is szerepelnek. Nem akarom azt állítani, hogy ezek a kommerciális ajánlatok haszontalanok. Bizonyos fokú hasznuk van, de a tudományos feldolgozást és a tudományos közvetítést nem tudják pótolni – ez még a tudomány világához tartozik. Néha az a benyomásom, hogy két világról van szó: egy technikai lehetőségeket

¹ Eredeti változata elhangzott 2008. március 26-án a budapesti Közép-európai Kulturális Intézet „Magyarország Bajorországból nézve. 45 éves a Münchener Magyar Intézet” című rendezvényén.

felhasználó és mozgékony szféráról és egy tudományos, de lassú és a technikával szemben szkeptikus szféráról (németül az úgynevezett *Elfenbeinturm*-ról, 'elefántcsonttorony'-ről). Pedig a tudomány nagyon jól fel tudja vagy fel tudná használni a modern technológiát saját céljaira, de oktatási és könyvtári célokra is; akárcsak a saját tevékenységének hatékonyabb bemutatására és a tudomány hasznának minden lehetséges fórumon való bizonyítására. Ez a mindenütt zsurgorodó finanszírozási bázis miatt is fontos szempont.

E gondolatok jegyében kezdte meg 1996/1997-ben a Münchener Magyar Intézet a koncepcionális munkát a digitális adatbázisok terén, az intézet „Magyarország a Kárpát-medence kapcsolattörténeteiben” című főtémáját illetően. Ezek a digitális szolgáltatások céljuk szerint kiemelik a tudományos kutatást a technológiai sötétségből, de nem asszimilálják a felületes nem-tudományos és profitorientált módszerhez, hanem a tudományos módszert és eredményeiket gazdasági és technikai eszközökkel terjesztik.

Persze számos módszer, anyag és lehetőség létezik a Kárpát-medence kapcsolattörténetének feldolgozásához, mégis az egyik legfontosabb alapot a dokumentumok és a források adják. Ezért az intézeti digitális projektum ebben az irányban fejlődött, bár kezdetben csak mint belső koncepció. A digitális könyvtári katalógus, amely a német mellett diakritikus és cirill betűs tételeket is tartalmaz, valószínűleg az első Münchenben – hozzácsatolva a Bogyay Tamás hagyatékának egy részét felölelő dokumentumtárat, valamint egy cikk- és egy recenzió-adatbázist (s mindegyiket keresőfunkcióval is ellátva): ezek voltak az első, először csak az intézetben használható eredmények.

1999 végén és 2000 elején tárgyalások kezdődtek az intézet részvételével, amelyek eredménye a bajor Kelet-Európai Kutatási Hálózat /FOROST/ létrehozása volt. 2000 áprilisában a Magyar Intézet és más bajor egyetemi és kutatóintézetek bemutatták a Bajor Tudományügyi, Kutatási és Művészeti Minisztériumnak az egyes projektterveket finanszírozásra. Decemberben a minisztérium által kinevezett szakértőcsoport pozitívan értékelt az intézeti projekttervet, amelyet a minisztérium jóvá is hagyott, úgyhogy a projekt 2001-ben el is indult a következő hosszú név alatt: *„Quellen und Materialien zur Beziehungs- und Integrationsgeschichte in Ostmittel- und Südosteuropa im 20. Jahrhundert unter besonderer Berücksichtigung Ungarns, der Minderheitenfrage und der europäischen Integration.”*

A projektum célja dokumentumok és más anyagok digitális és teljes szövegű feldolgozása volt; olyan forrásoké, amelyek fontosak a kapcsolattörténet ismeretéhez: például törvények, szerződések, ügyiratok, politikai előadások, emlékiratok, statisztikai adatok és – mint külön súlypont – kisebbségtörténeti anyagok. A módszer és a feladat tudományos volt, de a gyűjtés, az adminisztráció és a prezentáció modern technikai alapokra épült.

De mi is a különbség egy hagyományos dokumentum-könyvkiadvány, egy tradicionális adatbázis, valamint az intézeti információs szerver között?

1. A dokumentumok elérhetősége független a nyitvatartási időtől és a lakóhelytől. A dokumentumokat mindig és mindenhol használni lehet; csak egy számítógép és egy internetkapcsolat kell hozzá. Így azt a közönséget is el lehet érni, amelyik nem tud vagy nem akar könyvtárba járni. Ez a tudáshitelesítési módszer azokat az érdeklődőket is rá tudja bírni a dokumentumok használatára, akik eddig nem voltak ehhez szokva. A tudományos szférán kívül a levéltári vagy könyvtári munka kevésbé elterjedt, így ez az információs szerver pedagógiai célokat is szolgál.
2. A szerver rendszere több kereső módot kínál, úgyhogy sokkal több használati módszer teszi teljessé, mint például egy linklista vagy egy egyszerű adatbázis esetében. A használó rá tud keresni egy dokumentum címére vagy egy címrészre, le tudja szűkíteni a vizsgálni kívánt időszakot, vagy címszavakra tud keresni. E három kereső módszer kombinációja is lehetséges. Ez az összetett módszer egy könyvtári katalógus használatához hasonló. Azon kívül integráltunk egy keresőlehetőséget, amely indexbe veszi az összes dokumentumot, s ekként egyes szavakra is rá lehet keresni. Ez az index a *Google* keresővel működik, tehát költségmentesen, de mégis megbízhatóan.

A harmadik – tudományos szempontból legfontosabb és ebben a kombinációban valószínűleg első – ajánlat az úgynevezett virtuális gyűjtemények, illetve virtuális forrásgyűjtemények (németül: *die virtuellen Sammlungen oder virtuelle Quellensammlungen*). Jelenleg 14 virtuális gyűjtemény létezik. Ez a rész nemcsak a kutatókhoz szól, hanem diákokhoz és más érdeklődőkhöz is. Ezért mindegyik virtuális gyűjtemény előtt nagyon rövid bevezetés szerepel, amelyet még bővíteni lehetne. De ez miért is virtuális gyűjtemény? Az adatbázisban sok dokumentum kóddal van ellátva, és ez a kód az egyes dokumentumot egy virtuális gyűjteményhez csatolja. Ahogy növekszik az adatbázis és a dokumentumok száma, úgy nőnek a virtuális gyűjtemények is. A gyűjtemények számát is nagyon egyszerű bővíteni. Ehhez csak új dokumentumok vagy új gyűjteményi témák kellene, egy új kód és egy új kereső parancs a honlapon. Így akár csak pár nap alatt is jelentősen kiépíthető és bővíthető az információs ajánlat.

Ez a három felhasználói módszer – a sima címkeresés, a szókeresés és a tudományos feldolgozás a virtuális gyűjtemények alapján – tudományos szempontból valószínűleg a legalaposabb digitális forrásgyűjtemény Münchenben.

3. A technikai és tudományos lehetőségek jelenleg mindegyik hozzáférési és keresési módszert biztosítják. De technikailag és szakmailag mégis nyitott

rendszer ez, tehát jövőendő fejlesztések egyszerűen beépíthetők. Az intézeti információs szerver nyitott és változtatható berendezés, technikai szempontból is mindig fejleszhető.

4. Finanszírozási szempontból a Magyar Intézeti információs szerver egy *win-win*-szituációban működik. A szerver működése nem kerül plusz pénzbe, mert a dokumentumok azon a szerveren kaptak helyet, amelyen az intézeti honlap is elérhető. Az adatbázis jelenleg ingyen egy partnercég szerverén működik; átköltöztetését az intézeti szerverre még ebben az évben tervezzük. Ez az új elhelyezés is ingyenes működést biztosít, de a flexibilitás sokkal nagyobb lesz, mint a mostani megoldásban.

A technikai és szakmai kérdések mellett szerkesztési és koncepcionális kérdések vártak megoldásra. Egy ilyen ajánlat csak akkor hasznos a tudományos szférában, ha az idézhetősége biztosított, és megbízhatóan prezentálja az anyagot. Ezért a digitalizációs folyamatban a minőségre és nem a mennyiségre helyeztük a hangsúlyt, s biztos és kötelező munkalépéseket dolgoztunk ki. Az első lépésben a munkatársak megbízható publikációkban vagy eredeti lelőhelyeken megtalálható, tudományos szempontból fontos forrásokat és dokumentumokat választottak ki. Ebben a lépésben már az első szakmai strukturálás is megtörtént. A második lépésben a dokumentumok digitalizációja, illetve a beszkenelés és az optikai karakterfelismerés következett. Szkenelés után az anyagot elmentettük JPG-formátumban, majd az optikai karakterfelismerés után RTF-formátumban, így a dokumentumok most két típusban állnak rendelkezésre. Azt mondhatjuk, hogy mindkét formátumnak van előnye és hátránya. A JPG-formátum előnye, hogy az eredeti dokumentumot mint fényképet mutatja, de hátránya, hogy a tartalom nem kereshető. Az RTF-formátum a *Google* kereső funkcióval elérhető, de nem mutatja az eredeti anyagot; az eredeti forma – nem az eredeti tartalom (!) – elveszett. Az optikai karakterfelismerés pontossága körülbelül 95 százalékos, ha elég jó az eredeti beszkenelt dokumentumkép. Ahhoz, hogy az RTF-formátum megbízhatóan idézhető legyen, egy harmadik munkaszakaszban három korrektúra következett három munkatárssal. Ez fontos volt, mert csak így tudtuk az anyag idézhetőségét biztosítani. Fontos volt ebben a lépésben a hivatalos helyesírás, a diakritikus betűk korrekt megjelenítése, amelyek sok más digitális dokumentumgyűjteményben (főleg angol nyelvű országokban) elvesznek. A korrektúrákat követően az anyag PDF-formátumba került, amely az utolsó és most elérhető formátum. A dokumentumtár adminisztrációja egy saját *Microsoft Access*-adatbázisban történik, amely a duplamunkákat is ellenőrzi. Ez az adatbázis a végleges MySQL alapja is. 2005. május 4-én az intézeti információs szerver az internethez kapcsolódott, és azóta elég sokszor használt ajánlattá fejlődött.

Végezetül a következő rövid statisztikai adatok jól mutatják a szerver elfogadottságát.

Látogatók

Látogatók 2005-ben:	22.688
Látogatók 2006-ban:	30.849
Látogatók 2007-ben:	58.322
Látogatók 2008. március 12-ig:	12.417
Összesen:	124.276

Oldalnézettség

Oldalnézettség 2005-ben	41.182
Oldalnézettség 2006-ban:	49.311
Oldalnézettség 2007-ben:	82.695
Oldalnézettség 2008. március 12-ig:	18.723
Összesen:	191.911

Honnan jöttek a látogatók?

.net oldalak:	11.834	26,69%
.com oldalak:	25.962	43,60%
.de oldalak:	6.314	15,81%
.at oldalak:	1.439	3,76%
.ch oldalak:	879	1,99%
.pl oldalak:	794	1,79%
.hu oldalak:	450	1,38%

Különleges országok látogatói

Argentína, Brazília, Chile, Kína, Dél-Afrika, Fidzsi-szigetek, Japán, Kolumbia, Malajzia, Marokkó, Mexikó, Nicaragua, Pakisztán, Peru, Szingapúr, Tajvan, Togo, Tuvalu

Melyik oldalról jöttek a látogatók?

google.de	51,87%
google.com	11,54%
google.at	5,61%
google.pl	2,26%
wikipedia.org	1,65%
herder-institut.de	0,86%

Legtöbbször megnyitott dokumentumok

Dokumentumok:	Megnyitott:
1. Friedensvertrag von Brest-Litowsk vom 3. März 1918	1.463
2. Österreichisch-ungarisches Ultimatum an Serbien vom 22. Juli 1914	1.343
3. Armistice Convention with Austria-Hungary. Protocol of Conditions of the Armistice between the Allied and Associated Powers and Austria-Hungary. Signed at Villa Giusti, November 3, 1918)	1.337

4. Beneš-Memoranden zur Pariser Friedenskonferenz 1919/1920. Memorandum Nr. 10: Probleme der Berichtigung der Tschechoslowakischen und deutsch-österreichischen Grenzen	1.074
5. Treaty of Peace between Poland, Russia and the Ukraine, signed at Riga, March 18th, 1921	1.074

Göllner, Ralf Thomas

Corpus of 20th century interstate and intercultural relations in the HIM database

Today the internet is an all important device of research, yet it primarily mediates information rather than ensure it. The scientific, library, archival and museum expertise has the mission of knowledge transfer and knowledge authentication, which is a great intellectual, technical and financial challenge. The Hungarian Institute Munich (HIM) started the conceptual work concerning digital databases in 1996. The article describes in detail the work on an information server, which was first introduced in 2001. The server presents digitally established documents as well as full text documents, which are important for Central-Eastern European history focused on mutual relationships between states and nations: laws, contracts, documents, political talks, life recollections, statistical data, and – as a separate main focus – minority history material.