

Tudományos és Műszaki Tájékoztatás

Könyvtár- és információtudományi
szakfolyóirat

57. évfolyam
HU ISSN 0041-3917

2010/09

Tartalom

Ajánló

SZÁNTÓ Péter: **Új, innovatív fejlesztések**

Cikkek

- HORVÁTH György – HORVÁTH Zoltánné – PAPP Attila: **ADVISE adaptív automatikus kereső – miért más belül, mint kívül?**363
- BÁNKI Zsolt István: **Múzeumok az Europeanaban – az ATHENA projekt. Hozzáférés az európai kulturális örökség hálózataihoz**.....381

Beszámolók • Szemlék • Referátumok

- SHEPHERD, Peter: **COUNTER és SUSHI: a használói statisztika csúcsprogramjai** (Ref.: Bánhegyi Zsolt) 389
- APPLEYARD, Andy: **British Library – a dokumentumküldés útelágazás előtt** (Ref.: Burmeister Erzsébet) 390
- PORAT, Lynne – FINE, Sara: **A könyvtárközi kölcsönzés használatának és nem használatának jellemzése** (Ref.: Burmeister Erzsébet) 391
- CURRY, Elisabeth: **A könyvtárközi kölcsönzés kisebb változtatásainak is már nagy a hozadéka** (Ref.: Burmeister Erzsébet) 393
- PERRONE, Agnese: **Elektronikus könyvek szerzeményezése Olaszországban** (Ref.: Dancs Szabolcs)..... 394
- MAXYMUK, John: **Ingyenes online eszközök könyvtárosoknak** (Ref.: Dancs Szabolcs)..... 396
- ALIPOUR-HAFEZI, Mehdi – HORRI, Abbas – SHIRI, Ali – GHAEBI, Amir: **Digitális könyvtárak együttműködési modelljei. Szakirodalmi áttekintés** (Ref.: Drótos László)..... 398



Scientific and Technical Information

Hungarian journal of library and information science

Vol. 57 • 2010/09

Summaries

HORVÁTH, Gy. – HORVÁTH, Z. – PAPP, A.:

ADVISE – an adaptive search system 363

Adaptive search systems with an association network, concept visualisation and automatic concept generation increase the effectiveness of information search. The newly developed search engine named ADVISE relies on these tools. Its development started to serve business ventures, public administration and industrial companies. In their case everyday reporting is a general requirement, consequently, they need robust computerised systems which assist simultaneous and quick searches in various resources (data stores, databases, internet, e-mail systems, file systems etc.), and are able for automatic monitoring and reporting. The system should be adaptable with little human intervention, be able to learn and visualise. ADVISE as an adaptive system offers advantages also for those engaged in information management and for librarians, because of its capability to visualise concepts, create semantic networks and automatically set up categories. It facilitates service-provision by the actors of internet culture and those handling the mass of digitised materials. ADVISE can be integrated with any standardised system – the institutions are free to choose which systems they include into their special search service. ADVISE ensures authorised access as usual in the case of subscribed databases. The integration of the Web 2.0 tool IQportal and ADVISE provides complex possibilities for both librarians and users. A pilot project to introduce ADVISE in libraries has started, to test the system and tailor it to relevant needs, based on the integration of library resources.

BÁNKI, Zs. I.: Museums in Europeana – the Athena project..... 381

The Athena project is the museum part of the Europeana “family”. Athena will work in close cooperation with existing projects (Europeana: the European digi-

tal library network, and MICHAEL, EuropeanaLocal – all present in Athena) and develop intense clustering activities with other relevant projects. Athena is a Best Practice Network within the eContentplus Programme, initiated by the MINERVA network. It involves 20 EU Member States and 3 European observers, over 100 museums and other cultural institutions as direct partners or associated to the project, representing 20 European languages. It is coordinated by the Italian Ministry of Cultural Heritage. Athena’s main objectives: to reinforce, support and encourage the participation of museums and other institutions coming from those sectors of cultural heritage not fully involved yet in Europeana; to produce a set of scalable tools, recommendations and guidelines, focusing on multilingualism and semantics, metadata and thesauri, data structures and IPR issues, to be used within museums for supporting internal digitisation activities and facilitating the integration of their digital content into Europeana; to identify digital content present in the European museums; to contribute to the integration of the different sectors of cultural heritage, in cooperation with other projects more directly focused on libraries and archives, with the overall objective to merge all these different contributions into Europeana; develop a technical infrastructure that will enable semantic interoperability with Europeana. Its final aim is to bring together relevant stakeholders and content owners from all over Europe, to evaluate and integrate standards and tools for facilitating the inclusion of new digital content into Europeana, so conveying to users the original and multifaceted experience of the entire European cultural heritage. The Petöfi Museum of Literature, Budapest, Hungary takes part in the Athena project as a content provider. Its participation will bring experience for the Hungarian museum profession and acquaintance with implementations which enable its integration into the international scene, according to the museum field’s special needs.

ADVISE adaptív automatikus kereső – miért más belül, mint kívül?

Az ADVISE¹ innovatív keresőeszközt vállalati, államigazgatási, ipari és kulturális intézmények információs vagyónának automatikus keresésére, összefüggéseinek felismerésére, analitikus statisztikák előállítására, a kereséshez használható keresőnyelvek automatikus előállítására, fogalmi vizualizációra terveztük. Továbbfejlesztésében hangsúlyt kap a taxonómiák automatikus előállítása, tématerképek létrehozása, valamint a portálintegráció, hiszen az információs források, a vállalati adatvagyonok áttekintése és intelligens felhasználása sem kellően hatékony, sem kellően pontos nem lehet ezen eszközök alkalmazása nélkül.

Az információkereséshez alkalmazható fogalmi rendszerező eszközök fejlesztése és karbantartása általános problémaként jelentkezik a tudásalapú társadalmi környezetben. Az információmenedzsmenttel foglalkozók körében komoly szakmai erőfeszítések folynak az automatikus osztályozás, automatikus fogalomalkotás és az automatikus keresés megvalósítása, a szükséges leírónyelvek és szabványok készítése érdekében. Az információforrások rohamosan növekvő tömegéhez rendező elveket és nyelveket létrehozni, alkalmazni és szinten tartani nem kellően hatékony csak manuális módon és eszközökkel, ugyanakkor jogosan vitatott, hogy ez a tevékenység teljesen megoldható lenne az emberi intelligencia értékelő és elemző beavatkozása nélkül. Könyvtári környezetben alapvető probléma a különböző, heterogén adatforrások integrált keresésének és áttekinthetőségének megoldása. A feladatot tovább nehezíti az adatforrások egy részének távoli elérhetősége, valamint az alkalmazható adatlekérés és -áttöltés vegyesen szinkron és aszinkron lehetősége.

A nem információszolgáltatással foglalkozó intézmények, üzleti vállalkozások körében sem más a helyzet. Idővel ugyan megszületett a felismerés és a szándék az információtárolás szabályosságának és a metaadatok egységesítésének erősítésére, viszont a felismerést követően létrejöttek azok a robusztus információtechnológiai megoldások, amelyek az adatok tárolásának és kinyerésének hatékony, biztonságos megoldásait kínálják (adat-tárház, middleware-eszközök), valamint az utóbbi évek slágere, az üzleti intelligencia-rendszerek. Nem véletlen az „intelligencia” megnevezés, ugyanis a kulcsszó alapján történő keresés csak

felszíni eredményeket hoz, a mélyben rejlő és nem indexelt adatokat, az adatoknak a következtetéshez és a döntéshez szükséges összefüggéseit nem tárja fel. Ha kézi erővel történik a keresés, akkor a kulcsszó alapján kinyert információkat is ki kell egészíteni a találatok értékelésével (mennyiségi és tartalmi szűrés), a bennük rejlő információk esetleges szemléléseivel, analízisével, szintézisével – speciális esetekben a döntés-előkészítési szintig. Ezt a feladatot információkutatók, piackutatók, tudásmenedzserek, speciálisan felkészült könyvtárosok végzik, és a tevékenységnél rendkívül fontos az elemzőképesség, a számítástechnikai felhasználói szakértelem, valamint a kiszolgált terület ismerete és mindezen képességekhez szükséges ismeretek folyamatos fejlesztése. Az általánosan rendelkezésre álló szolgáltatásoknál azonban nem kívánhatjuk meg a felhasználóktól az ilyen szintű képességet és háttérismereteket, ezért ezeket egyre inkább a számítástechnikai rendszerektől várjuk. Az ún. „intelligens” kereső- és elemzőeszközök a keresést az adattárak „mély” rétegeiben is végzik, és az eredményeket táblázatos, grafikonos összeállításban, például egy portálon mutatják akár szinkron megjelenítéssel, változáskezeléssel, kiemelve a „veszélyes” mutatókat. A portálon megjelenő automatikus monitorozó és jelentéskészítő eszközök között vannak már üzleti szimulációs szoftverek is (pl. Oracle Essbase), amelyekben lehet kísérletezni esemény és következmény vizsgálatával.

Mindezek mellett mégis hiányzik egy „front-end” típusú keresőeszköz, amely az összes létező forrásban keres, legyen az e-mail, adatbázis, adat-tárház, fájlserver, internet vagy bármilyen más infor-

mációs vagy, és igényünk az, hogy a rendszer bonyolultságának megfelelő technológia a háttérben intézze a mély rétegek keresését, és az eredményeket felhasználóbarát környezetben kapjuk meg.

Az ADVISE kialakításánál a fentieket kiemelten kezeltük: az automatikus, szemantikai és vizualizációs módszerek alkalmazása során egyszerű felhasználói felület áll rendelkezésre, a program adaptív, tanuló rendszerként támogatja a keresést több (integrálható) forrásból, különböző adattárakból az együttes információkinyerés céljából. A háttérben zajlik a rendszerek integrációján alapuló tranzakciók sorozata, amelyet a felhasználó nem érzékel.

Az ADVISE ismertetése előtt teszünk egy kis kitérőt a keresés, a vállalati intelligens keresők, az internetes keresés és az internetes szemantikus keresők területén, amelyek értékeit és tapasztalatait a fejlesztés során felhasználtuk.

A döntések általában nem ott születnek, ahol az információ rendelkezésre áll

Az információk jelentős része (egy becslés szerint 80%-a) nem strukturált adatbázisokban jelenik meg, hanem különböző fájlokban (.doc; .ppt; .xls; .pdf; .mpp; .jpg; .html stb.), és bizonyos részük metabázisokban lévő strukturálatlan adat, amelyekre jellemző, hogy nincs egységes megjelenési felületük és közvetlen hozzáférésük (pl. adattárházak). A jelen tudásalapú és innovációvezérelt gazdasági környezetben a teljesítmény hatékonysága erősen függ az információk keresésétől, ezért flexibilis megoldásokra van szükség, alkalmazásuknál pedig kreativitásra. Az információkereséssel eltöltött – egyes felmérések szerint kb. 30% – munkaidő-hatékonysági tényező szempontjából rendkívül fontos, hogy ennyi idő alatt milyen eredményt tudunk felmutatni. Az üzleti intelligencia-eszközrendszer, technológia és eljárás a vállalatoknál az adattárakban, adatforrásokban, azok mély rétegeiben lévő információtartalom magas szintű kinyerését szolgálja, amelyek alapján következtetések vonhatók le – mára már automatikus, vizualizációs, adatbányászati, vagy értékelő, szintetizáló és analizáló megoldásokkal. Az ADVISE ezen eszközök körébe tartozik, mindamellett a hagyományos információszolgáltató intézmények számára is megoldást kínál.

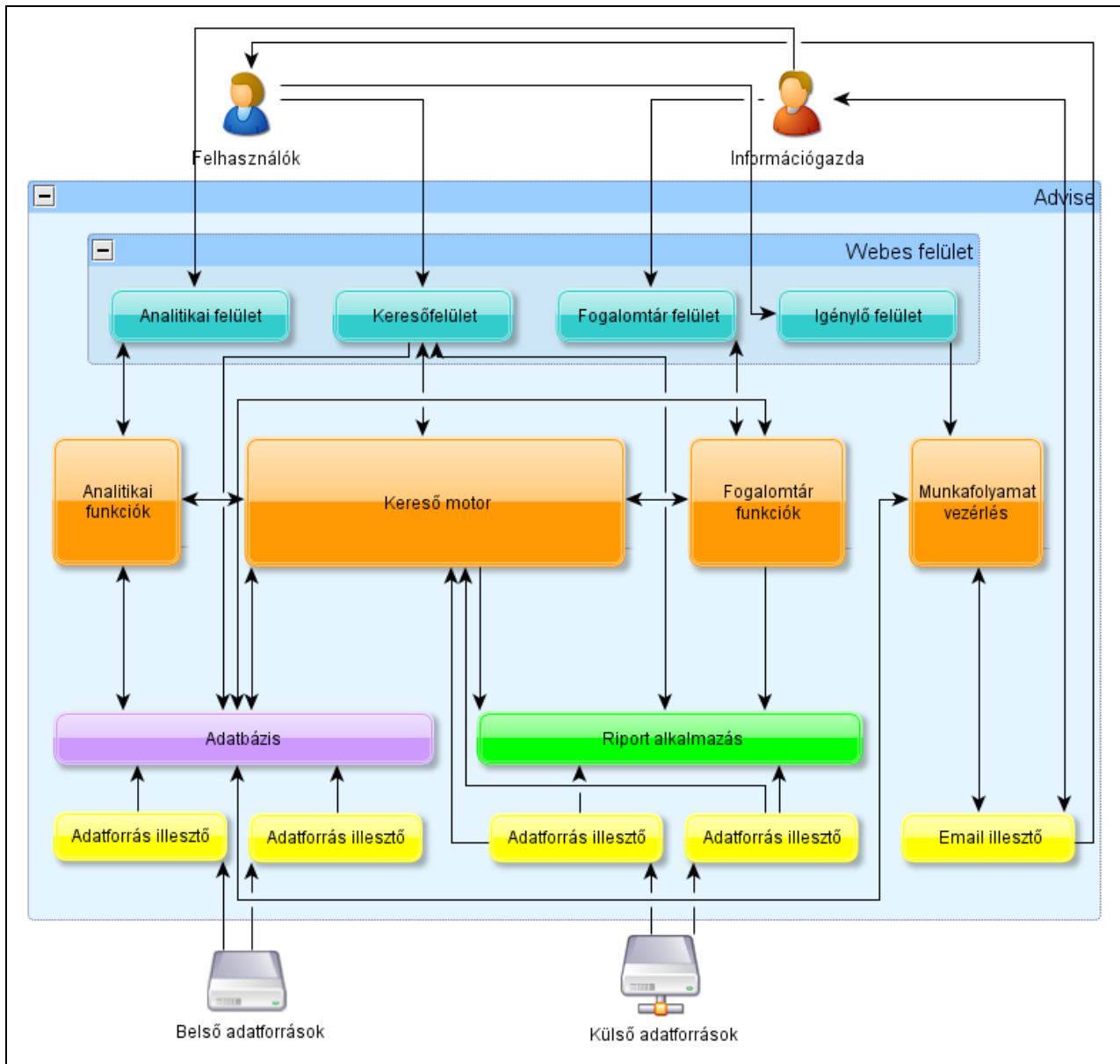
A teljesítménykényszer és az információmenedzsment

A teljesítményünk szervezése, növelése forradalmáról beszélnek egyes szakírók², akik szerint ezért fog egyre növekedni az elemző információszolgáltatók, „knowledge workerek” száma, akik lefedik majd a pénzügyi, az egészségügyi, a média és egyéb frekvenciált szakterületek munkavállalóinak 25%-át. Ha netán kételkednénk ebben a jóslatban, akkor is valószínű, hogy ezzel párhuzamosan, egymást erősítve egyre nagyobb figyelem hárul az innovatív információkereső megoldásokra. Az információforrások növekedését nem lehet megállítani, ezért az információ vagy jobban kihasználása egyre több erőforrást fog lekötni a kutatás-fejlesztés, és az emberi intelligencia szempontjából is. Nyilván nemcsak az elérés, hanem a források alapján végezhető műveletek lehetősége (analízis, szintézis, következtetések, transzformációk, döntés-előkészítés) jelentik azt az elméleti, tudományos, vagy üzleti előnyt, amelyből egyéb eredmények fakadnak.

A könyvtárak számára az üzleti intelligenciaeszközök a robusztus technológia miatt igen drágák, amelyek elterjedése a nagyvállalatokat követően csak néhány éve jellemző a közép- és kisvállalatoknál, azonban az ADVISE kereső reális lehetőséget kínál a könyvtári szektornak is arra, hogy a keresési metodika készségeinek a birtokában talán a könyvtárosok használhassák ki a leginkább a rendszer adottságait (1. ábra).

Keresés az automatikus és vizualizációs módszerek igénybevételével

Az internetkultúra környezetében a nagy tömegű elektronikus információk keresésénél a tartalomszolgáltatók és a felhasználók oldaláról egyaránt jelentkezik az automatikus keresés és az automatikus fogalomalkotás lehetőségének, az elkészült kereső- vagy tartalomosztályozó rendszerek gyors karbantartásának, és a rugalmas szerkezeti megoldásoknak, módosításoknak az igénye. Az egyes szakterületeken különböző fogalmi, osztályozási rendszerek készülnek és állnak rendelkezésre a hierarchia és a tudományos megalapozottság különböző szintjein. Megfigyelhető, hogy vállalati környezetben erős a pragmatikus megközelítés, amely a tudományos nyelvi elemzés és rendezettség helyett a teljes vállalat gyakorlati, közérthető szaknyelvi megközelítésére alapoz, és a fogalmi



1. ábra Az ADVISE igénylési munkafolyamatok támogatása

rokonságokat alacsony szintű struktúrában tükrözi, amelyet könnyebb átfordítani a számítástechnika nyelvére. Nagy jelentősége van a kombinált megoldásoknak, a rendszerek átjárhatóságának, a szemantikai keresőknek, és az összefüggéseket grafikusán is megjelenítő fogalmi vizualizációnak. A fogalmi vizualizációnak az összefüggések tükrözése, elemzése terén különös előnye, hogy a grafikus megjelenítés mellett grafikus eszközökkel módosíthatók az élek és a csomópontok, ezért a szerkezet könnyen karbantartható a használat során, ahogy ezt az ADVISE kereső esetében is tapasztaljuk.

A szövegösszefüggés szerepe – a szemantikai tér

Az információkeresésnél nem szavakat, hanem témákat keresünk, amelyek valamilyen szövegösszefüggésben, szerkezetben, nem „string”-ként jelennek meg, és „szemantikai térnek” is nevezhető³. Az információkereső segédeszközök kialakítása során ezért megnőtt a szerepe a szövegösszefüggések tükrözésének, azonban ez a módszer nem lehet annyira alapos, hogy ne legyen közzérthető, veszélyeztetve a használat gyorsaságát vagy könnyű elsajátítását. A szemantikai webhez kap-

csolódó kutatások, leírónyelvek is a szövegekben lévő szemantikai és szintaktikai összefüggések leírására törekednek, mert ezeket az összefüggéseket egy gépi rendszer számára formalizálni kell ahhoz, hogy az automatikusan felismerje. A leírónyelvek alkalmazása azonban messze áll jelenleg az általános használatbavétel lehetőségétől a szintaktikai és szemantikai formális nyelvi elemzés nehézsége, a leírás számítástechnikai átfordításának bonyolultsága miatt. A széles körű elterjedés megértésbeli gátjai rávilágítanak arra, hogy mennyire csodálatos az emberi intellektus, amely könnyedén mozog ebben a térben, ám nyelvezetünk formális visszatükrözése eléggé megoldhatatlannak látszik még akkor is, ha például a humor, a gúny, a metafora stb. tükrözésétől eltekintünk.

Az ADVISE mint szemantikus kereső újdonságértéke

A szemantikus keresőknek nevezett eszközök nem régen jelentek meg a piacon, és elsősorban az interneten megjelenő információk intelligens feltárását szolgálják. A Google és a Microsoft egymással versengve törekednek a keresés finomítására, a keresés vertikális lehetőségeinek kiterjesztésére. (L. *Google Universal Search, Google Analytics, Google Squared, Bing*). A keresőkben előre beállított kategóriákat adnak meg, amelyek a keresés típusa, az információ megjelenési formája, időbeli megoszlása és egyéb szempontok szerint szűrik a találatokat. A Google Squared táblázatokba tömöríti a találatokat, módosítható sorokkal és oszlopokkal, a táblázat elmentési lehetőségével. A kereső igen jól működik az ún. webmarketing-szolgáltatásokban, azonban a hazai tematikájú információk esetében meglepő és nehezen igazolható összefüggéseket mutathat a táblázatba foglalt információegyüttes.

Vannak statisztikai, vagy szövegkörnyezeti, jelentéstan, tudásalapú, vagy ontológiai összefüggésekre épülő keresők (*TextWise, Radar Networks Twine, Hakia, Wolfram Alpha, Jebol, WOWD* stb.). A Wolfram Alpha több beépített modellt használ, amelyek számos valós területhez kapcsolódnak, és nem a webmarketing, hanem a tényinformációk kereséséhez ajánlják.

Kutatások folynak különböző nemzetközi digitális könyvtári projektekben, ám ezek inkább az ontológiai nyelv szintjét célozták meg a keresőkatégoriák és keresőrendszerek fejlesztésével, ontológiai leíró nyelv alkalmazásával. Külön meg kell említeni az

Autonomy rendszert⁴, amelyhez a leginkább hasonlít az ADVISE. Mindkettő ún. „tanuló” rendszer, tudásportál-funkciókat támogat, megismertet az egyes témák gazdáival, automatikusan hozza létre a témacsoportokat, olyan adattárakban végzi a kereséseket, amelyek egyébként nem kommunikálnak egymással, jól kezeli a többnyelvű megoldásokat, feltárja a rendezett és strukturálatlan információk kapcsolatát, és automatikus taxonómia előállítását is lehetővé teszi.

Megemlíthető, hogy a szemantikus keresők és a szemantikus web kapcsolata nem következik a jelen keresők természetéből, így az ADVISE megoldása sem a szemantikus web eszközrendszerébe tartozik jelenleg. A rendszer filozófiájában a rugalmas és gyakorlati alkalmazás fejlesztését tartjuk követendőnek távlatilag is, és kiemelten fontosnak tartjuk az egyes szektorokhoz való alkalmazkodást.

Összességében megállapítható, hogy az újabban megjelent szemantikai alapú keresőrendszerek stratégiájában és megoldásban is jelentősen eltérnek az ADVISE rendszerétől, amely nem csupán a web keresésére, hanem több, különböző típusú információforrás egyidejű keresésére szolgál. Az eddig megismert hasonló célú keresők egyes elemeihez kimutathatók hasonlóságok (asszociációs, automatikus keresés, analitikus statisztika készítése stb.), azonban egyik említett keresővel sem rokonítható sem célját, sem az alkalmazott technológiát, sem a keresés módszereit tekintve.

Tématérkép előállítása automatikusan – az ADVISE innovatív eszközrendszerrel

A tématerképnél minden téma valójában egy szinonimacsoportot képez, amelyet egyetlen megnevezés képvisel. A témahely tárgyának formális deklarációjára és azonosítására a tárgyi osztályozás technikáját használja, de a tématerkép megfordítja az általunk ismert információfeltáró folyamatot; itt nem a dokumentumból, hanem a témától jutunk el az objektumhoz, az információforráshoz.

A megnevezés ad egy helyet (scope), ahol megjelenik a téma egy halmaza, amely egy tartalmat képvisel. A megnevezések lehetnek azonos alakúak is, azonban a típus, az előfordulás és a kapcsolat pontos értelmezést ad a megnevezésnek, például: Paris (mitológiai alak); Paris (város). A tématerkép a többnyelvű információk szolgáltatását is támogatja, a felhasználó saját nyelvén választhatja

ki a megnevezést, és a rendszer nem arra figyelmezteti, hogy egy másik, preferált kifejezést alkalmazzon, hanem belső kapcsolatai alapján „érti” a kérdést. A tématerkép háttérében a tudásintegrációra fejlesztett ISO-szabványt használunk.

Az ADVISE eszközzel célunk, hogy automatikusan állítsuk elő néhány szakmai terület tématerképét és megvizsgáljuk az azonos vagy hasonló elemek importját és integrációját más rendszerekhez.

Különböző keresési módszerek és eszközök

Mikor van szükség keresésre? Az egyszerű válaszban: „amikor nem találok valamit” összetett feladatcsoportok határozhatók meg, amelyeknek minden összetevőjére figyelniük kell. (Válaszidőcsökkentés, adatszerkezet, adatforrás-indexelés, időskálán való elhelyezés, elavulás, kérdés időpontja, kulcsszókezelés, szinonimák, ragozott formák, információkereső nyelvek stb.)

A keresés során a válasz minősége növelhető azzal, ha tudjuk, ki kérdez. Egy keresőt a keresési szokásai alapján tudunk leírni. Ezt az információt használják ki az *adaptív keresők*, amelyek csoportjához az ADVISE is tartozik⁵. A következőkben a hagyományos szöveges alapú keresőtől a szemantikus keresőkön át az internetes keresők specialitásait érintve jutunk el a vállalati keresők világához. Mindegyik területnél áttekintést adunk az adott terület főbb kihívásairól, jellemzőiről, illetve kapcsolatáról az egyéb területekkel – amely tulajdonságokat az ADVISE innovációnál figyelembe vettük.

Szöveges keresés

A szöveges keresők a keresési problematikát a *hol* kérdésre összpontosítják. A keresés tárgyát szövegrészletek alkotják, amelyek előfordulását a rendelkezésre álló adatforrásokban nagy hatékonysággal meg tudják mondani. Ennek a megközelítésnek előnye az egyszerűség, a nagy teljesítmény, valamint a kiforrottság. Ugyanakkor kétségtelen hátrányként kell megemlítenünk a keresés többi tényezőjét, miszerint a fogalmi kapcsolatok hiányában, az idő és a kérdező ismerete nélkül a válaszok sok esetben irrelevánsak vagy pontatlanok lesznek. A szöveges keresőrendszerek döntő többségében a következő architektúris modulokat tartalmazzák:

- *Pásztázás* (ún. *crawling*)⁶: a rendelkezésre álló adatforrások bejárását vezérelni szükséges. En-

nek oka, hogy figyelembe kell venni az adatforrások hasznosságát, redundanciáját, valamint azt, hogy sok esetben nem is járható be a teljes halmaz, ezért szükséges algoritmizálni a bejárható szelet meghatározását.

- *Elemzés*: a bejárás során érintett adatforrásokat elemezni szükséges, hogy olyan reprezentációt készítsünk, amelyet egységesen és hatékonyan előkereshetően tudunk ábrázolni. Tipikus feladatok a formátumkonverzió, kis/nagybetűk kezelése, stopszavak kiküszöbölése, szótövezés, nyelvfelismerés, kivonatkészítés.
- *Tárolás*: az indexált adatok hatékony tárolása kulcsfontosságú, hiszen ez határozza meg döntő részben a keresés sebességét. Itt a relációs adatbázisok mellett nagy szerepet kapnak a speciális kiválmakokat is kezelő egyedi implementációk.
- *Keresés*: a keresés során a felépített adatszerkezet funkcióit használva meg kell határozni a találatokat, azok értékét-sorrendjét, valamint tipikusan valamilyen kivonatolt tartalmát.

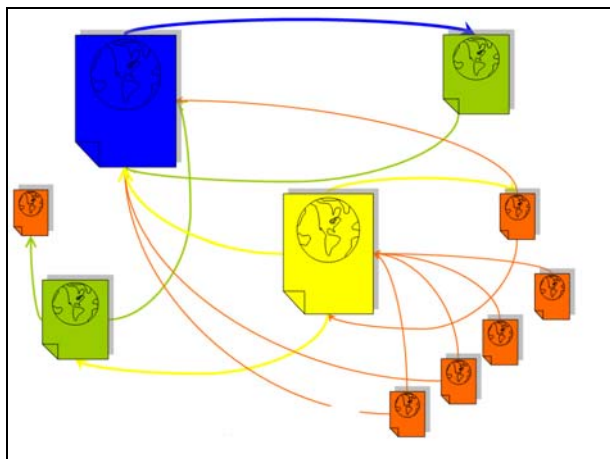
A pásztázó algoritmusok feladata, hogy adott számítási kapacitás mellett biztosítsák a megoldandó feladat által meghatározott optimumot a következő paraméterek esetében: mennyiség, aktualitás, pontosság. A fenti architektúris határok rugalmasak, a keresési funkcionalitás szempontjából kategorizálják egy rendszer komponenseit. Példaként gondoljunk arra, hogy egy adatbázisban történő keresés során is csak akkor tudunk hatékonyan lekérdezni, ha található index a kért információ tartalmazó oszlopokhoz. Ez esetben az indexelés természetesen nem ütemezetten, bejárás által vezérelve történik, hanem automatikusan az adatbázis-műveletek közben a háttérben.

PageRank algoritmus

A *PageRank* algoritmus⁷ (2. ábra) az egyik legismertebb módszer az internetes keresők körében. Alapötlete az, hogy rendeljünk minden oldalhoz egy rangot, amely azt tükrözi, hogy az adott oldal mennyire fontos. Ennek alapján már tudunk szelektálni a beláthatatlan mennyiségű oldal között, hogy melyeket érdemes indexálni. A kérdés csupán az, hogy az oldal fontosságát hogyan lehet megállapítani. A *PageRank* válasza az, hogy egy oldal annál fontosabb, minél több fontos oldal mutat rá. Formálisabban megfogalmazva: egy oldal rangja a rá mutató oldalak rangjának súlyozott összege.

A fenti definíciót alkalmazva egy *N* darab oldalból álló webrészletre meg lehet határozni az egyes oldalak rangját. A valós implementációkban ugyan-

akkor iteratív módszereket szükséges alkalmazni a rang meghatározására, hiszen az oldalak száma nagyobb annál, mint hogy direkt megoldó algoritmust lehetne alkalmazni. Ennek módszeréről az irodalom bőven ad tájékoztatást.



2. ábra PageRank működési séma

Elemzés

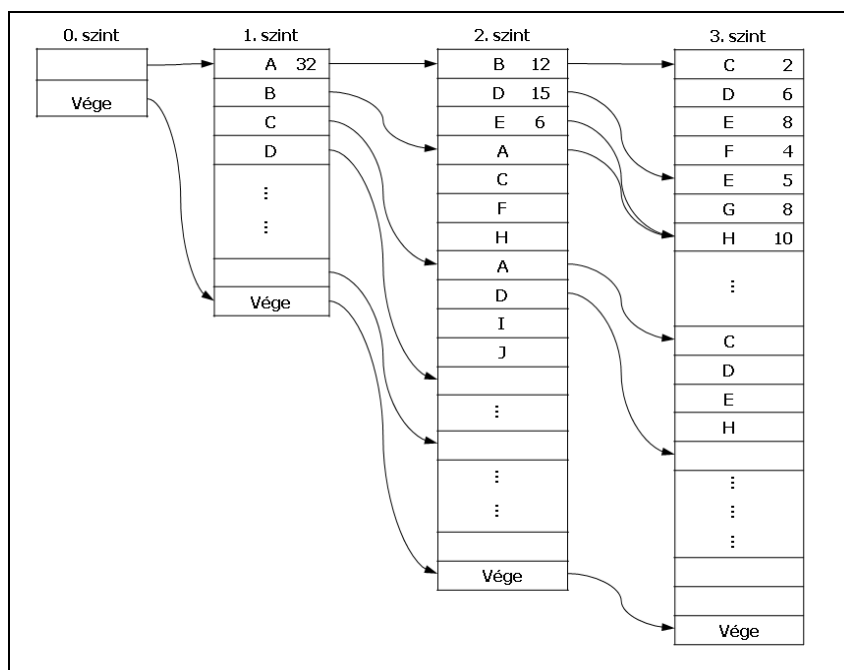
Az adatforrások elemzését az eredmény tekintetében a következő két csoportra bonthatjuk:

- *statikus elemzés*: az adatforrás tartalma a bejárás pillanatában lekérdezésre kerül, majd az elemzést ezen az információhalmazon végezzük el;
- *dinamikus elemzés*: az adatforráson keresztül elérhető információk leírása – ún. metaadatok – a bejárás pillanatában lekérdezésre kerül, azonban a tényleges információk lekérdezése és elemzése keresési időben történik.

Statikus elemzés

A lekérdezett adatokat több lépésben szükséges feldolgozni, hogy jól kereshető reprezentációhoz jussunk. A leggyakoribb feldolgozási lépések a teljesség igénye nélkül: normalizálás, stopszavak kiiktatása, szótövezés, nyelvi felismerés, szöveg-hasonlóság-elemzés, képi feldolgozás. A nyelvi detekció nem mindig történhet dokumentum-metainformációk alapján, mert azok sok esetben hiányosak vagy hibásak. Ezért szükséges magát a szöveges tartalmat alapul venni.

Az egyik legelterjedtebb módszer erre a *trigramstatisztika* készítése. A trigram egy betűhármast, a trigramstatisztika pedig ezen betűhármastak előfordulásának gyakoriságát egy szövegben (3. ábra).



3. ábra Trigramstatisztika készítése

Számos esetben ütközünk keresés során abba a nehézségbe, hogy adathibákból, elgépelésekből vagy akár pusztán marginális dokumentummódosításokból kifolyólag egy kérdésre helyes válasznak tekinthető dokumentum semmilyen formában nem tartalmaz szavakat a kérdésből, még szótő szintjén sem. Ilyenkor ad segítséget a szövegek hasonlóságának elemzése. Ezek a módszerek nagyrészt ki tudják küszöbölni a fenti okokból keletkező kis-mértékű eltéréseket a szövegekben.

Szövegek távolságának meghatározására számos gyors módszer ismeretes, ezek közül az egyik a *Levensthein-távolság*⁸. A 4. ábra ennek számítását illusztrálja:

| | B | A | N | Y | A | B | Á | N | Y | Á | S | Z | A | T |
|---|----|----|----|----|----|---|---|----|----|----|----|----|----|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| A | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| D | 2 | 2 | 2 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | 3 | 3 | 2 | 3 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 4 | 4 | 3 | 3 | 4 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 11 |
| B | 5 | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Á | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| N | 7 | 6 | 6 | 5 | 6 | 6 | 6 | 5 | 4 | 5 | 6 | 7 | 8 | 9 |
| Y | 8 | 7 | 7 | 6 | 5 | 6 | 7 | 6 | 5 | 4 | 5 | 6 | 7 | 8 |
| Á | 9 | 8 | 8 | 7 | 6 | 6 | 7 | 7 | 6 | 5 | 4 | 5 | 6 | 7 |
| S | 10 | 9 | 9 | 8 | 7 | 7 | 7 | 8 | 7 | 6 | 5 | 4 | 5 | 6 |
| Z | 11 | 10 | 10 | 9 | 8 | 8 | 8 | 8 | 7 | 6 | 5 | 4 | 5 | 6 |
| A | 12 | 11 | 10 | 10 | 9 | 8 | 9 | 9 | 8 | 7 | 6 | 5 | 4 | 5 |
| T | 13 | 12 | 11 | 11 | 10 | 9 | 9 | 10 | 10 | 9 | 8 | 7 | 6 | 5 |

B A N Y A B Á N Y Á S Z A T
A D A T B Á N Y Á S Z A T
+ = o o o = = = = = = = = =

4. ábra Levensthein-távolság számítása

Mint látható az ábráról, az algoritmus lineáris időben futtatható és kiküszöböli a fajlagosan kis eltéréseket két szövegrészlet között.

A szöveges dokumentumokban sok esetben hordoz kulcsfontosságú információt a kép. Természetesen a kép tartalmának általános meghatározása nem reális feladat, ugyanakkor számos alkalommal nyílik lehetőség hasznos információk felderítésére. Ehhez elegendő pusztán két kép hasonlóságának felismerése – melyre már léteznek hatékony algoritmusok.

Dinamikus elemzés

Dinamikus elemzést szükséges alkalmazni akkor, ha az elemzendő információhalmaz mérete irreáli-

san nagy és/vagy gyorsan változó. Ez tipikusan fennáll adatbázis-tartalmakra, melyekre például az internetes keresés témakörében a „mély web” terminológiát szokás alkalmazni, utalva arra, hogy az információ felszínre hozható ugyan az internetes felületen keresztül, de ehhez kéréseket kell specifikálni az adatbázis felé – legtöbbször valamilyen űrlap formájában. Felmerül a kérdés, hogy ha az információ közvetlenül nem indexálható, akkor mit lehet kezdeni az ilyen adatforrásokkal. A választ a metaadatokban találjuk, azokban az adatokban, amelyek az üzleti értéket hordozó adatokat írják le. Ezeket értelmezve és indexálva tudjuk megállapítani, hogy egy adott kérdést érdemes-e feltenni az adott adatforrásnak – keresési időben – vagy nem.

Keresési mátrix

A keresési feladatot a következőképpen lehet a legegyszerűbben matematikailag szemléltetni. Képzeljünk el egy nagy mátrixot – ezt a következőkben keresési mátrixnak fogjuk nevezni –, amelynek sorai a kérdések, oszlopai pedig a válaszok. A mátrix egyes celláiban egy mérőszám áll, amely azt fejezi ki, hogy az adott kérdésre az adott válasz mennyire jó. Könnyen belátható, hogy a mátrix általában igen nagy, ugyanakkor igen ritka. Előbbi adja a keresés egyik technikai nehézségét, utóbbi pedig a megoldást. A mátrixot a ritka mátrix-reprezentációnak megfelelően célszerű tárolni, azaz nem tárolunk le minden elemet, hanem minden sorból/oszlopból csak a nem nulla elemeket jegyezzük meg, pozíció szerint.

A feladatot nehezíti, hogy az indexálás során mindig egy oszlopban található adatok jelennek meg egyszerre, a keresés során pedig egy sor adataira vagyunk kíváncsiak. Mivel a célfüggvény az, hogy a keresés gyors legyen, az elemeket soronként csoportosítva kell tárolni, ami indexálási időben pontosan annyi egység módosítását jelenti, ahány kérdésre releváns választ találtunk. Az implementációkban ezért fontos szerepet kap a sorok elérési idejének minimalizálása.

Relációs modell

A keresési mátrix elemei egyszerűen betölthetők relációs adatbázisba például egy – *kérdés, válasz, relevancia* – adatszerkezetben (5. ábra). Indexet téve a *kérdés* oszlopra a lekérdezések hatékonyak lesznek.

| Kérdés/Válasz | http://iqwiki | file:///IQSYS-Mukodes/BIT | https://y-iqsyssps |
|----------------------|---|---|---|
| üzleti intelligencia | 0,5 | 1 | 0,3 |
| RDBMS | 0,8 | 0,5 | 0,4 |
| yEd | 0,9 | 0,6 | 0 |
| BI | 0,6 | 0,9 | 0,7 |

5. ábra Relációs modell

A megközelítés kiválóan alkalmazható kisméretű keresőrendszerek esetén. Nagyméretű rendszereknél az index mérete igen nagy lehet, ami performanciaproblémákhoz vezethet. Ezt orvosolják a következőkben említésre kerülő módszerek:

A *hashmap*⁹ egy olyan adatszerkezet, amely kulcsértékpárok között definiál hatékony leképezést a kulcsok alapján képzett ún. *hashértékek* felhasználásával. Hatékony működésének feltételei a következők:

- a *hashértékek* képzésére szolgáló *hashfüggvény* kellően homogén módon szórja szét a kulcsokat az értékészletben,
- a *hashterülethez* rendelkezésre álló tárhely összemérhető legyen a várható elempárok számával.

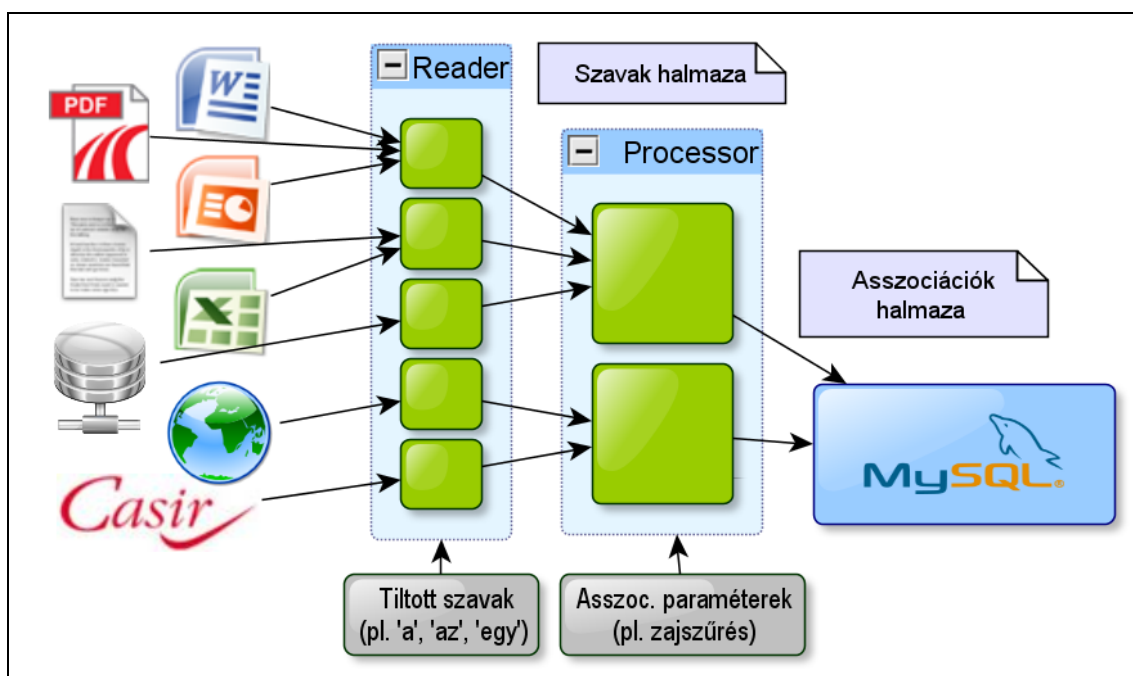
A fentieket biztosítva elmondható, hogy ez az adatszerkezet konstans időben tud választ adni a kérdésekre, egy kérdés-válasz kulcsértékpár men-

tén történő előzetes felépítés esetén. Alkalmazásának a *hashterület* nagysága tud határt szabni, amelyre az *elosztott hashmaptechnika*¹⁰ nyújt megoldást, amelynél az adatok több számítógépen vannak elosztva abból a célból, hogy a teljesítmény növelhető legyen.

Index a keresési mátrixhoz

A keresés alapja a keresési mátrixra felépített indexállomány. Ez teszi lehetővé, hogy a keresőrendszer méretezése során figyelembe vett adatforrás-mérettartományban a válaszdők egy előre meghatározott konstans alatt maradjanak. A keresőarchitektúra feladata az indexállományok karbantartása (6. ábra). Az új adatforrások frissítése, az elavultak öregítése. Adaptív rendszerek esetén itt szükséges figyelembe venni a megtanult információkat.

Adott adatforrás-mennyiség és hardverkapacitás mellett a keresési válaszdő tovább növelhető, ha a gyakori kérdésekre adott választ *gyorsítótárba* helyezük. Ezzel átlagos válaszdő-követelmény esetén erőforrást tudunk felszabadítani a rendszerben más feladatokra – például mélyebb elemzés, differenciáltabb keresés.



6. ábra Indexálás az ADVISE-ban

Szemantikus keresők

A szemantikus keresők működése a rendelkezésre álló kereshető tartalmak értelmezésén, jelentésének felderítésén alapul. Könnyű belátni, hogy ez a koncepció relevánsabb találatokhoz és gyorsabb keresési ciklusokhoz vezet, ha a háttérben álló tartalomértelmezés adekvátnak tekinthető. A fentiek következményeképpen a szemantikus keresők legfontosabb tulajdonsága a taxonómiaépítés módszere, amely alapvetően meghatározza a keresőrendszer használhatóságát.

Taxonómiaépítés

Az információk értelmezésének alapját az ún. taxonómiák adják¹¹. A taxonómia egy *fogalomrendszer*, amelyben a fogalmak között *relációk* vezetnek, ezzel hozva létre a szükséges kapcsolati rendszert a kereséshez. A fogalmak között húzódo kapcsolatok attribútumait a taxonómiaépítő módszertan határozza meg. Tipikus kapcsolatok a *szinonima*, *kategória*, illetve *tulajdonság*. Így könnyedén megfogalmazható, hogy például a „google” fogalom „*kategóriája*” a „*kereső*” fogalom.

A fogalmak és relációik meghatározása számos módon történhet a manuális – ember által végrehajtott – taxonómiaépítéstől kezdve a hibrid megoldásokon át a tisztán gépi hálózat kialakításáig. A manuális és az automatikus taxonómiaépítés tulajdonságainak összevetését mutatja az 1. táblázat.

1. táblázat

A manuális és az automatikus taxonómiaépítés tulajdonságai

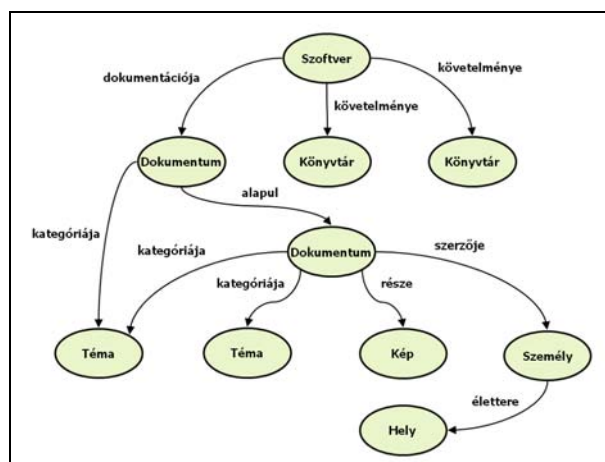
| | Manuális taxonómia-építés | Automatikus taxonómia-építés |
|--------------------------|---------------------------|------------------------------|
| Sebesség | Lassú | Gyors |
| Minőség | Magas | Közepes |
| Erőforrásigény | Nagy | Kicsi/közepes |
| Karbantarthatóság | Nehézkes | Triviális |
| Felhasználhatóság | Univerzális | Speciális |

A manuális és az automatikus taxonómiaépítést összevetve elmondható, hogy mind a mai napig kiegyensúlyozott a verseny és nincs általánosan kiválasztható „jó” irány. A megoldás általában a két módszer vegyítése, melynek módja erősen alkalmazás- illetve területfüggő. Példaként említhetjük, hogy a honlapokhoz kapcsolt metaadatok és a weboldalak kapcsolatai alapján könnyen lehet gépi

módszerrel taxonómiát építeni, azonban az adatok hiányossága miatt ezt sokszor további intelligenciával kell kiegészíteni: címek megállapítása, szótövezés, illetve végső esetben manuális korrekció segítségével.

Szemantikus web¹²

Tim Berners-Lee, a világháló atyja meg van győződve arról, hogy a jövő világhálója szemantikai alapokon fog működni. A jövőkép szerint a napon-ta több millió új oldal megjelenéséhez a későbbiekben ezzel összemérhető mennyiségű szemantikai információ fog társulni. Ahogy a világháló természetes nyelve a HTML (*Hyper Text Meta Language*), úgy a szemantikus információké az RDF (*Resource Description Framework*), illetve az OWL (*Web Ontology Language*). Az RDF erőforrások – esetünkben tartalmak – egyedi és relációs leírására alkalmas nyelv. Az OWL pedig ezt egészíti ki magasabb szintű osztályozási és relációs információk leírásával. Alkalmazásukról az irodalomban bőségesen találunk leírást¹³, itt csak példaként említhetjük meg egy személy nevének és a hozzá kapcsolódó információknak a kapcsolását elérhetőségek, naptár, honlap, illetve referenciákkal (7. ábra):



7. ábra Metaadatok szemantikus kapcsolatai

Visszaulva a manuális és az automatikus taxonómiaépítés összevetésére: a világháló méretéből kifolyólag a manuális taxonómiaépítés erős hátrányban van az automatikus módszerekkel szemben. Egyelőre nem látszik kellő mértékűnek az RDF és az OWL elterjedése ahhoz, hogy a szemantikus web elképzelése ilyen módon megvalósulhasson. Éppen ennek köszönhető, hogy olyan éles a verseny, és dinamikus a fejlődés az auto-

matikus taxonómiaépítő eszközök és a szemantikus keresők piacán.

Szemantikus keresési metodika

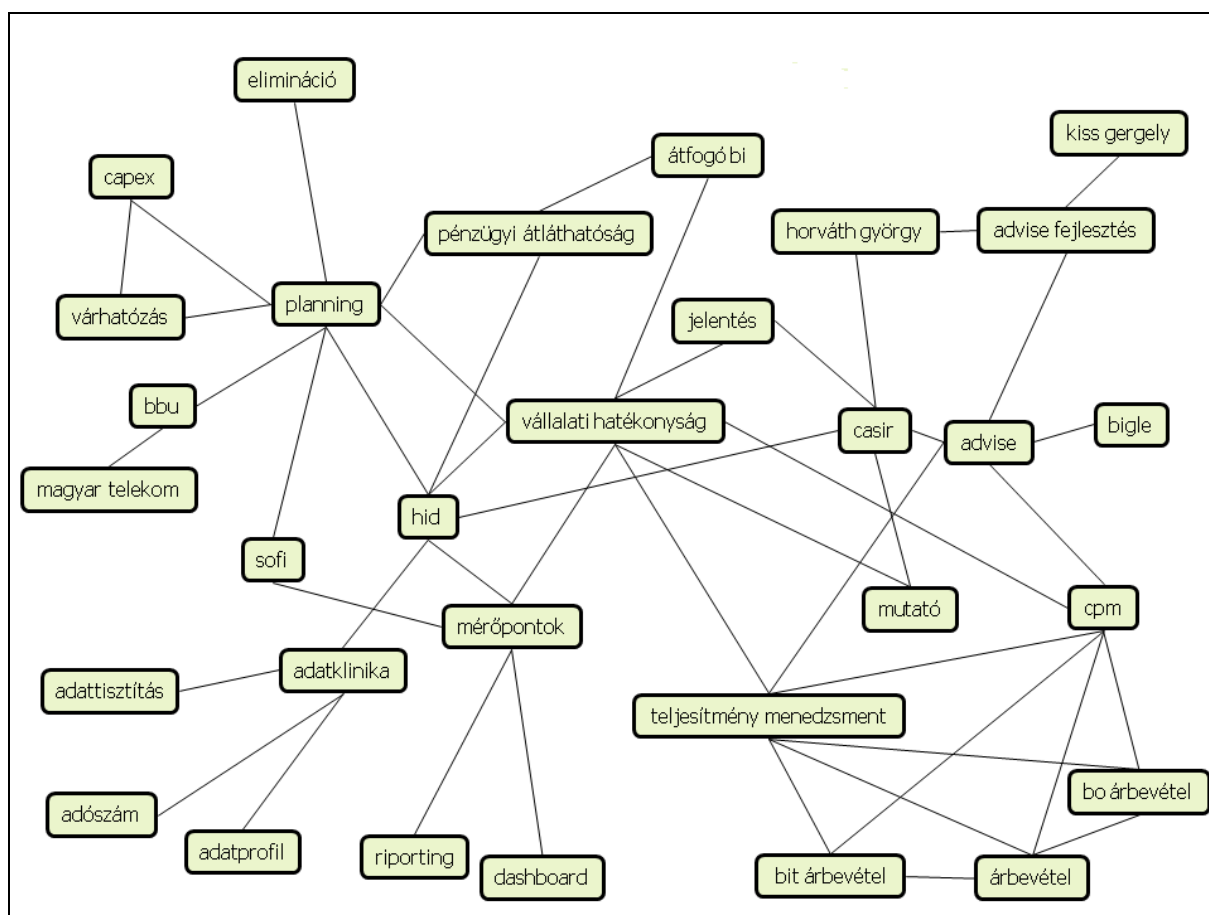
A szemantikus keresők jelentős része nemcsak adatforrásokat kínál fel találatként, hanem kapcsolódó kereséseket, fogalmakat és témaköröket egyaránt. Ezzel segítik, orientálják a felhasználót a kívánt eredmény irányába. Azaz ilyenkor a keresőmotor nem a szöveges egyezések alapján ad csak találatokat, hanem megpróbálja felderíteni azt, hogy a felhasználó mire gondolhatott és az hogyan, milyen formában található meg az adatforrásokban.

A felajánlások algoritmikus alapja sokféle lehet. Felépített taxonómiával rendelkező rendszer esetében természetesen a taxonómia adja a kapcsolódó fogalmakat és erőforrásokat, azokat pusztán

rangsorolni és megjeleníteni kell. Taxonómia hiányában a keresési szokások tanulása valamint a statisztikai, illetve szövegbányászati algoritmusok tudnak segítséget nyújtani.

Vizualizáció

Egy taxonómia igen jelentős méretű lehet, megjelenítésének módja és minősége már egy speciális szakterület esetén is kritikusá válhat, hiszen adott esetben ezen múlik, hogy a felhasználó kellő időben észreveszi-e a számára szükséges információt. Az XML-nek mint technológiafüggetlen információhordozó-formátumnak kiemelt jelentősége van az egyes tudásreprezentációs formák közötti átvitel szempontjából. Így nyílik lehetőség például a 8. ábrán egy diagramkészítő eszköz (yED)¹⁴ vizualizációs technikájának alkalmazására egy tetszőleges XML alapú taxonómialeírás esetében.



8. ábra Fogalmi vizualizáció yED eszközzel

A weben egyre gyakrabban megjelenő szemantikus kereső megoldások döntő része – mint ahogyan az ADVISE is – rendelkezik valamilyen vizualizációs technikával, melyben pozícióval, mérettel, színekkel és egyéb eszközökkel vezetik a felhasználó tekintetét – a rendszer által elképzelt – optimális irányba.

Klasszikus internetes keresők és a mély web

A klasszikus internetes keresők alapvetően szöveges alapú keresést végeznek. Ez tömören nem más, mint a keresőkifejezésben szereplő szavak előfordulásainak megkeresése az adatforrásokban – részlegesen és teljesen egyaránt. Emellett minden keresőnek szüksége van egy rangsorolási modellre, amelynek alapján sorba rendezik azokat a dokumentumokat, amelyekben a keresett kifejezések szerepelnek. Itt legtöbbször az előfordulás gyakorisága, illetve helye a döntő. Tekintettel arra, hogy a legnagyobb keresők sem képesek teljes mértékben lefedni a webes tartalmak teljes egészét, valamint a tartalmak egy része meglehetősen gyorsan változik, kulcskérdés a „fontos” oldalak meghatározása (l. PageRank algoritmus), azaz, hogy mely oldalakat érdemes indexálni, hogy a legtöbb kérdésre releváns választ tudjunk adni.

Az interneten keresztül elérhető tartalmak döntő része láthatatlan marad a keresők előtt, mert úrlapok kitöltésével érhetőek el. Az esetek többségében adatbázisból lekérdezett adatokról van szó. A web ezen – gép számára „láthatatlan” – részét nevezi az irodalom *mély web*nek. A mély web méretének becslése gyakorlatilag reménytelen feladat, hiszen a háttérben található adatbázisok szerkezete, mérete többnyire nem publikus, így azzal globális szinten nem lehetséges számolni.

Vannak azonban esetek, amikor a keresők – még ha kis számban is – fel tudják használni a mély web tartalmát. Ehhez speciális illesztőprogramokra, illetve metaadatokra van szükség az érintett adatforrásokhoz. Így lehetséges például, hogy a legnépszerűbb keresők az időjárást és a devizaárfolyamokat gond nélkül szolgáltatják – holott ezek az információk nyilvánvalóan nem HTML oldalak indexálásával álltak elő, hanem speciális adatbázis-hozzáférések által.

Vállalati keresés – miért más belül, mint kívül?

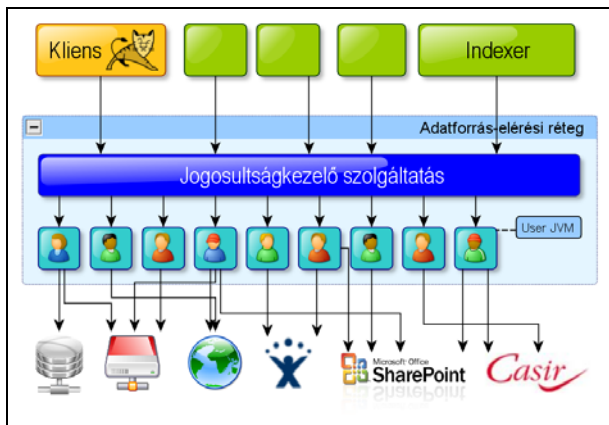
A vállalati keresők egészen más piacot képviselnek, mint az internetes keresők. Ennek oka az eltérő üzleti modell, technológiai háttér és a felhasználói kultúra. A vállalati keresőrendszereknek kisebb létszámú, egyértelműen azonosítható, hasonló érdeklődési körű, illetve általában kvalifikáltabb felhasználót kell kiszolgálni. Az azonosíthatóság nemcsak az információbiztonság szülte szükség, hanem előny is egyben a keresés szempontjából, hiszen a keresési szokások tipizálhatóak, az eredmények testre szabhatóak. Az adatforrások esetében is jelentős differencia mutatkozik. A vállalati rendszerek esetében rengeteg strukturált információ is rendelkezésre áll a strukturálatlan adatok mellett. A struktúrákhoz pedig az esetek többségében metaadatok is tartoznak, melyek segítik a keresést akár automatikus taxonómiát szolgáltatva. Vállalatok esetében a tevékenységi kör sok esetben jól körülhatárolható, ezzel specializálhatóvá téve a keresést és a találatok megjelenítését.

Mérhetőség, kontrollálhatóság, jogosultságkezelés

Vállalati keresőrendszer esetében az adatforrások birtoklása és a felhasználók azonosíthatósága révén a keresőrendszer az alap funkcionalitásához jelentős hozzáadott értéket tud előállítani a vállalat adatvagyonának, illetve a felhasználók, alkalmazottak munkaszokásainak feltérképezésével. Egy vállalati kereső üzemeltetése esetén képet kaphatunk az adatvagyon minőségéről, a hiányosságokról és a feleslegekről egyaránt. Ez visszacsatolást nyújthat a vezetésnek a fejlesztendő, vagy racionalizálendő adatterületek, illetve kompetenciák tekintetében.

A jogosultságkezelés alapvető kérdés vállalati közegben (9. ábra). Ennek megfelelően a vállalati keresőknek is igazodniuk kell ehhez. Ez adott esetben igen komplex feladatot is jelenthet, hiszen heterogén rendszerek esetében heterogén jogosultság-ellenőrzéssel állunk szemben, amelyet hibátlanul kell kezelni. További kihívást jelent a vállalati keresők számára, hogy a jogosultságokat valós időben – a keresés közben – szükséges vizsgálni, hiszen bármilyen gyorstárazási módszerrel biztonsági lyuk létrehozását kockáztatjuk.

A vállalati keresők jellemzését követően áttérünk az ADVISE bemutatására, amely vállalati keresőként indult, az IQPortál innovációs fejlesztését követően azonban könyvtári, illetve egyéb információmenedzsment-feladatokra is megkezdték alkalmazását.



9. ábra Jogosultságkezelés az ADVISE-ban

Mi is tulajdonképpen az ADVISE?

Az elnevezés az „Adaptive DataWarehouse Search Engine” játékos rövidítése alapján született. A megnevezés talán félrevezető lehet abból a szempontból, hogy azt sugallja: ez a keresőmotor nemcsak a dokumentumokban, hanem az „adat-tárházszerű” rendszerekben is keres. Miközben ez egyébként igaz, a következőkben láthatjuk, hogy sokkal többről van szó.

Az ADVISE innovációs termék eleinte elsősorban vállalati igények kielégítésére született, mert a vállalatoknál olyan ütemezett információ- és jelentéskényszer van, amely közvetlenül befolyásolja a gazdasági eredményt, vagyis kimutatható az eszköz közvetlen haszna. Az adattárházakban tényyszerű és számszerű adatok vannak, amelyek kinyerése tartalmi, minőségi, pontossági, teljességi és hatékonysági mutatószámokat eredményez. Összetett rendszerek keresése során az adattartalomnak legalább logikai szintű ismerete szükséges ahhoz, hogy a felhasználó meg tudja fogalmazni kérdéseit, illetve értelmezni tudja a kapott válaszokat. Szélesebb körben (pl. internethasználók vagy könyvtárhasználók esetében) már nem várható el a megfelelő szintű háttérismeretek. Az internetes keresőkhöz szokott felhasználók egy bonyolult háttérvilág egyszerű keresőjét használják az ADVISE alkalmazása során; az igen összetett

informatikai háttér egyetlen felszíni (front-end) megoldásban integrálja a strukturált és a strukturálatlan adatok kereshetőségét.

Az ADVISE egy keresésre tervezett webes felülettel rendelkezik, amely adaptív képessége révén a felhasználók keresési szokásai szerint javítja a találatok súlyozását. A felület és a rendszer „tanulóképesége” a felhasználó igényeihez történő alkalmazkodást és felhasználóbarát megoldást szolgálja. A felhasználók kereséseikhez és a találatokhoz egyaránt könyvjelzőket rendelhetnek, amelyek könnyedén megoszthatók más felhasználókkal. A rendszer tárolja a keresések történetét, a leggyakoribb kérdések egy gombnyomásra lekérdezhetők, és lehetőség van a felhasználók csoportosítására, amit a rendszer automatikusan képes szinkronizálni az elterjedt szolgáltatásokból (LDAP, Active Directory). A keresésnél látjuk, melyik forrásrendszerre várunk, módunk van a kiválasztott találatok rendezésére, értékelésére, jegyzetelésére. A keresés pontosságát fejlett idő- és típuszűrési funkciók támogatják (10. ábra).

The screenshot shows the search interface with several filter tabs: 'Időszűrés' (Time filter), 'Források' (Sources), 'Típusok' (Types), 'Nyelvszűrés' (Language filter), and 'Dol' (Jobs). Under 'Időszűrés', there are radio buttons for 'Bármikor' (Anytime), 'Egy éve' (One year), 'Egy hónapja' (One month), and 'Egy hete' (One week). Below this, there are input fields for 'Kezdő dátum' (Start date) and 'Záró dátum' (End date), with example values '2010.04.05.' and '2010.04.12.' respectively.

10. ábra Idő- és típuszűrési funkciók

Az ADVISE számos konceptuális elemet örökölt az alapvetően internetes keresésre kifejlesztett motoroktól – például tanulási képesség, adaptív logika alkalmazása, vagy a fogalmak közötti asszociációk építése és karbantartása –, számos területen viszont új megközelítést kellett kialakítani. Ilyen például a vállalati rendszerek esetében természetes jogosultságkezelés, és ezzel szoros összefüggésben a szerepkörvezérelt tanulási algoritmus. De az ADVISE szakít a hagyományos szekvenciális találati listával is, egy újszerűnek mondható, a fogalmak közötti szemantikai összefüggéseket hálószerűen ábrázoló megjelenítő felület bevezetésével (11. ábra). Nemcsak a belső hálózaton található strukturálatlan, szöveges tartalmakat kezeli, hanem a különböző rendszerekben, adatbázisokban és a kapcsolódó metabázisokban található információkat is összegyűjti. A megoldás rugalmas adatforrás-illesztéssel bír, amely – igény esetén –

lehetővé teszi további rendszerek bevonását is, ezért alkalmazható könyvtári környezetben a különböző adatbázisok, adatforrások integrálása nyomán a könyvtári információkeresésre, tudásmenedzsment-feladatokhoz és integrált portálkeresőként.

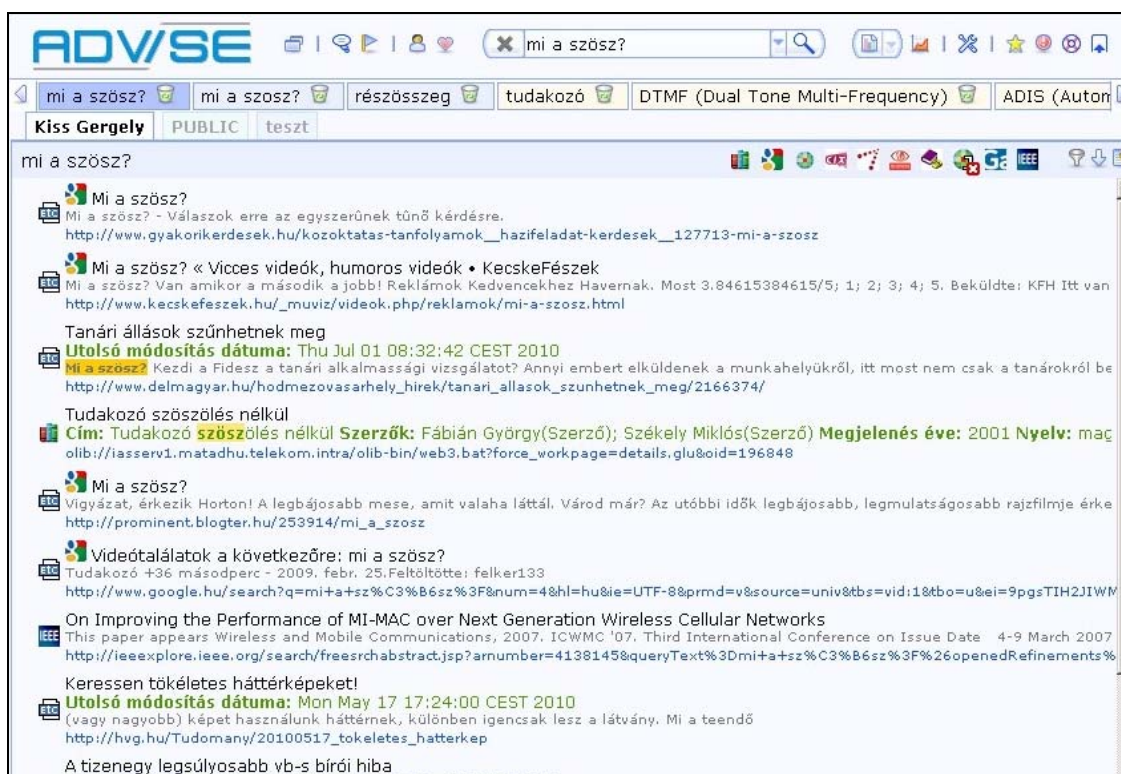
ADVISE – adaptív, tanuló, automatikus keresőrendszer

Az ADVISE fogalomalapú kereső automatikus tárgyszavazási folyamatot végez az adatforrások indexálása során. Természetesen ez a gépi algoritmus önállóan nem tudja azt a pontosságot elérni, mint amire egy ember képes. A felhasználók keresési szokásait azonban adaptálja a rendszer, ezáltal a fontos tárgyszavak köre behatárolható és azoknak a kapcsolati hálójá felépíthető. Ez lehetővé teszi, hogy egy riport definiálása előtt a rendszer már a felhasználó szokásai alapján a riportparaméterek döntő részét automatikusan meghatározza. Egy keresési folyamat gyakorlatilag az ad hoc riport fogalmához közelít, és meg is valósítja azt, amikor a felhasználó kéri az eredmények rendszerezését és formázását. Ezt az információt felhasználva a dokumentumokat újraindexálva jelentős pontosság érhető el az automatikus tárgy-

szavazás módszerében. További fontos tény, hogy számos ismeretterületen már rendelkezésre állnak tárgyszavazott vagy csupán tematikai besorolással rendelkező dokumentumok, melyek szintén felhasználhatók a kapcsolati háló felépítésére és finomítására. Ilyenkor a rendszer pontosan úgy viselkedik, mintha a tárgyszavazást végző felhasználó „annak idején” ezt az ADVISE tanulófelületén keresztül tette volna meg.

Hasonlóan a szemantikus webhez, képes témahelyek közötti meghatározott kapcsolatokra épülő automatikus akciók generálására. Szerkesztése a taxonómiák felső szintje, illetve azok kapcsolatai és előfordulásai szerint történik. Kapcsolódhat osztályozási rendszerekhez, emellett felhasznál(hat)ja a tezaurusz szemantikai ugrópontjait és keresési módszereit is.

A riportdefiniáláshoz, ha információt szeretnének kinyerni, az ADVISE a fogalmi háló vizualizációs képességével is támogatást nyújt. Az asszociációs kapcsolatok megjelenítése és szerkesztése könnyedén, intuitív módszerekkel megtehető. Kísérleti megvalósításunk alapján a megoldás egyszerűsége egy közönséges wiki vagy más web2-es technológiához hasonlítható.



11. ábra ADVISE keresőfelület különböző tartalmakból (projektkönyvtár, jogosultságellenőrzés, névadatok stb.) Az egyes adattípusok kiemelten jelennek meg.

Analitikai funkciók a lekérdezésekhez

Az ADVISE analitikai funkciói közül az alábbiakban néhányat felsorolunk:

Kereséshez kapcsolódó kimutatások

- Melyek a leggyakoribb keresések?
- Melyek az adott témakör legrelevánsabb fogalmai?
- Melyek a legkeresettebb témakörök?
- Melyek a hiánytémakörök?

Dokumentumok

- Melyek a leggyakrabban letöltött dokumentumok?
- Egy adott témakörhöz melyek a legrelevánsabb dokumentumok?
- Melyek a felhasználók szerint leghasznosabb dokumentumok?
- Melyek a felhasználók szerint haszontalan dokumentumok?
- Melyek azok a dokumentumok, amelyeket még senki nem használt?
- Milyen tipikus szűrőfeltételekkel található meg egy adott dokumentum?
- Mekkora a dokumentum eszmei értéke? (Hányan használják és milyen értékeléssel rendelkeznek?)

Forrásrendszerek, adatbázisok

- Milyen az egyes rendszerek sebessége, rendelkezésre állása?
- Mekkora az egyes rendszerek, illetve előfizetések kihasználtsága?
- Mely felhasználók vagy csoportok használnak egy adott adatforrást a legintenzívebben?
- Mely adatforrások adják a legrelevánsabb találatokat?

Felhasználók

- Mennyire aktívak a felhasználók (keresés, letöltés)?
- Mennyire elégedettek a felhasználók a szolgáltatással, a találatok minőségével?
- Milyen célcsoportok, kompetenciák vannak?
- Mely felhasználók, illetve csoportok kompetenciái azonosak?
- Milyen a csoportok közötti kollaboráció?
- Kik vannak feliratkozva egy-egy riportra?
- Ki milyen riportokra van feliratkozva?
- Mely szervezethez tartoznak a feliratkozottak?

Asszociációs jelleg – hogyan lehet riportot definiálni és elkészíteni?

A riportkészítés első lépése mindig annak a meghatározása, hogy mire vagyunk kíváncsiak, azaz milyen információra van szükségünk. Ehhez sok

esetben nem egyszer futunk neki a kérdésnek, és fokozatosan, iteratív módszerrel próbáljuk behatárolni, hogy mi lenne számunkra az igazán hasznos információ. Ezt a folyamatot tekintjük a riport definiálásának. Ennek a lépésnek a fontossága a strukturálatlan adatok esetében sem kisebb, mint a jelenlegi megoldásoknál. Kulcskérdés, hogy hogyan lehet megfogalmazni azt, hogy „milyen információra van szükség”? Ebben segít az asszociációs gondolkodás, mely az ADVISE-kereső motorjának és adatelemző rétegének kulcsfontosságú eleme. A fogalomalapú keresés során nem szövegrészleteket bocsátunk a rendszer rendelkezésére, hanem fogalmakat, amelyek között egy asszociációs háló írja le a kapcsolatokat, az agy alapvető működéséhez hasonlóan – azt természetesen lényegesen leegyszerűsítve. Az asszociációk mentén a rendszer az adott fogalomkörhöz legrelevánsabb információhalmazt tudja a forrásokból meghatározni, akkor is, ha a definiálás során a felhasználó az adatforrásokban található terminológiától eltérően fogalmazott.

A fentiek alapján már látható, hogy a strukturálatlan adatokra épülő riport nem más, mint egy jól meghatározott kritériumrendszer mentén végrehajtott adatvagyon-feltérképezés és -keresés majd -rendszerezés és formázott összegzés.

A végső jóváhagyást természetesen mindig az ember adhatja meg, az ADVISE fogalomvizualizációs felületén lehetőség van az automatikusan elkészített tárgyszóhalmaz megtekintésére és felülbírálására. Utóbbi esetben implicit módon ismét tanítottuk a rendszert – amely információ a következő indexálás során ismét felhasználható.

ADVISE – az automatikus fogalmi vizualizáció újszerű megoldása

A fogalmi vizualizáció automatikus előállításával az adott tárgykör fogalmi struktúrája érzékletesen mutatható be. Az ADVISE intuitív vizuális környezete egyszerűen módosíthatóvá teszi a taxonómiát, a későbbi terveink szerint „drag-and-drop” technikával is. A jelenlegi verzióháló export-importot tesz lehetővé, és a vizualizációt a yED eszköz szolgáltatja. A hálók és csomópontok korlátozás nélkül fejleszthetők.

A vizualizáció a lexikai egységek közötti relációkat áttekinthetően mutatja. Rendelkezik intuitíve használható interfésszel, hogy ösztönözze a felfedezést. Érdekessége, hogy olyan elemeket rendel

egymáshoz, amelyeket nem lehet számszerűsíteni, vagyis a kifejezések jelentéseit és kapcsolatait.

A vizuális háló megfigyelései során rugalmasan változó lehet a háló, megfigyelhető és ábrázolható a használók egy adott oldalhoz kapcsolódó internet-használati magatartása is. A vizualizációval elemezhetők akár a rejtett szerkezetek, például az üzleti struktúrákban. Ha például kompetenciátérképben készül vizualizáció, látható a hiány vagy a telítettség, amely döntési, stratégiai információ az irányítás kezében. A kompetenciainformációkhoz tartozó és személyekre vonatkozó háló szemléletesen mutatja egy-egy munkatárs tevékenységi struktúráját, kapcsolati rendszerét, vagy tudásának, tevékenységének irányultságát, színvonalát.

Portál és ADVISE-integráció automatikus kategorizáló, lekérdező és tartalomszolgáltató feladatokhoz

Az érdeklődőkkel folytatott konzultációk alkalmával a szakterületen jártas kollégák számára a szemantikus keresésnek és az automatikus taxonómia építésének előnyei könyvtári környezetben pillanatok alatt nyilvánvalóvá váltak, ezért előzetes egyeztetéseket végeztünk mind üzleti, mind architekturális témában az IQPortál és az ADVISE integrációjáról.

Az ADVISE mint kifejezetten heterogén adatforrásokra tervezett szemantikus kereső, az automatikus fogalomépítés mellett biztosított fogalmi hálóbetáplálás képességével a portál heterogén információforrásaiból egy keresési folyamatban képes az információkat kinyerni – ha szükséges, jogosultságokhoz kötve. Az integráció révén a portál nemcsak információszerepet tud betölteni, hanem tudásportál-funkciókat is. Az ADVISE alkalmas az explicit információk sokféle formájából a kompetenciák felderítésére és a kompetenciákhoz tartozó tartalmak kinyerésére, anélkül, hogy a terület szakértőjének vagy művelőjének közreműködését kellene kérni. (Ezt a szabadságot korlátozhatja a jogosultság limitálása.)

Az integrált motor támogatni fogja az OAI-PMH protokollt, a keresőfelület pedig az IQPortál felületébe illeszkedő módon fog megjelenni. A taxonómiaépítés támogatásához a felületen lehetőséget adunk a fogalmi háló megjelenítésére és szerkesztésére is. A könyvtári szakma számára az ADVISE kereső automatikus fogalmi hálóépítési képessége és vizualizációja, a rendszer adaptív, asszociációs

képessége és rugalmas módosítási lehetősége jelenti a fő vonzerőt.

Az integráció a fentiekben megfogalmazott célok felhasználói felületét képezik az alábbi megoldásokkal:

- a. Felület funkcionalitásának differenciálása felhasználói tapasztalat szerint.
- b. Kompetenciamenedzsment-felület.
- c. Exchange Server illesztő.
- d. SAP BW illesztő (főként vállalati környezetben fontos).
- e. Lokális keresés támogatása.
- f. SSO támogatás (Single-Sign-On – egyszeri bejelentkezés).
- g. OpenSearch illesztés.
- h. Windows tálcakomponens.
- i. Dokumentumkezelő illesztés.

Technológia

Az ADVISE alapja egy elosztott architektúrára tervezett skálázható keresőmotor. A motor biztosítja a rendszer adaptív funkcióinak integrálását a klasszikus indexálási feladatokon keresztül az ún. okos indexáló bővítményekhez, amelyek metaadatokból, adatbázis/tábla adatokból, szöveges adatokból, dokumentációból, illetve minden olyan tevékenységből származnak, amelyet a felhasználók a keresőrendszerrel végeznek. A keresőmotor a legkorszerűbb lineáris hálózati analízisen alapul, számos specialitással kiegészítve. Ez ad lehetőséget arra, hogy a gyakorlatban fokozatosan módosuló adatbázis változását rentábilisan le lehessen követni algoritmusokkal. A fenti apparátus az igen elterjedt *Hibernate* eszközön keresztül kapcsolódik a JDBC-kompatibilis adatbázisokhoz a legrobustusabb elosztott gyorstárolási és kapcsolatkezelési megoldások támogatása mellett. A rendszer használata nem igényel fejlesztői beavatkozást, üzemeltetése minimális IT-erőforrást köt le. Lehetőség van a moduláris bevezetésre, a többszálú, többgépes működésre, amely a terheléselosztással javítja a teljesítményt. Rugalmas konfigurációt nyújt JVM-en belül / JVM-ek (gépek) között, és lehetővé teszi a runtime (üzemidő alatti) újrakonfigurálást. Folyamatosan megoldott a teljesítménymérés, a memória-nyomkövetés és a távoli hibaelhárítás-funkció.

Az ADVISE automatikus fogalmi hálóépítési mechanizmusa támogatja fogalmi háló importálását, illetve exportálását. A rendszer alapja a platform- és adatbázis-független Java technológia, mely lehetőséget biztosít a megannyi forrásrendszerhez

és adatbázishoz való illesztéshez, valamint kiválóan támogatja a vállalati webes alkalmazások fejlesztését. A rendszer fejlesztése során törekszünk az adatbázis-függetlenségre, hogy a teljesen szabad forráskódú és ingyenes szoftverkomponensektől (pl. MySQL adatbázis, Apache webservert) a nagyvállalati méretekig (Oracle adatbázis és OC4J webservert) minden fontosabb és szabványos adatbáziseszközt lefedjünk a közbülső lépcsőkről sem megfélekezve (pl. Microsoft SQL Server).

Az architektúra tervezésénél kiemelkedő fontosságot tulajdonítunk a skálázhatóságnak. A rendszer moduláris, valamint klaszterezett felépítésű. Bővítésre lehetőség van már telepített rendszer esetében is számottevő költség nélkül.

A webes felületek AJAX technológiát alkalmaznak, amely letisztult, interaktív tájékozódást tesz lehetővé a keresések és a riportok elkészítése, valamint böngészése során. A rendszer többi részéhez hasonlóan itt is törekszünk a platformfüggetlenségre, ezért gyakorlatilag minden kurrens böngészőt támogatunk (Explorer, Firefox, Chrome, Opera, Safari).

A rendszer architektúráját az alábbi komponensek alkot(hat)ják, de ettől részben eltérő is lehet:

- *Adatforrás-illesztő.*
- *E-mail-illesztő.*
- *Adatbázis:* a rendszer fogalmi hálóját valamint a felhasználói és az analitikai adatokat tároljuk itt.
- *Riportalkalmazás.*
- *Analitikai funkciók.*
- *Keresőmotor.*
- *Fogalomtár-funkciók.*
- *Webes felület (analitikai felület a rendszer analitikai funkciói és jelentései eléréséhez, keresőfelület, fogalomtár-felület, igénylőfelület).*

Az ADVISE mint termék

Az információmenedzsment területén folyamatosan fennálló költségelvonás és -hiány miatt különösen fontos az integrációban rejlő szinergiák kihasználása, és a fenti fejlesztés termékszerű megjelenítése az egyes szakterületek számára (könyvtár, levéltár, múzeum, üzleti vállalkozás, MOKKA stb.) – mindazon informatikai lehetőségekkel, amelyet az ADVISE automatikus szemantikus kereső és az IQPortál integrációja kínál. Ezért tervezzük különböző intézménytípusok számára a rendszer terméként való értékesítését, amely a meghatározott funkciócsoportokhoz standard

megoldásokat kínál – főként az alábbi modulokra bontva:

- *Automatikus egyidejű keresés elektronikus dokumentumokban, fogalmi rendszerezés a szemantikus keresés pontosítása érdekében (automatikus osztályozás és keresés, taxonómiák és ontológiák, tématerképek, fogalmi háló stb.).*
- *Integráció az IQPortál termékkel és a preferált rendszerek illesztésével (adatbázisok, fájlrendszerek, digitális könyvtárak, távoli adatbázisok, megvásárolt adatbázisok, internet, kompetenciamenedzsment-felület, SSO támogatás stb.).*
- *Használat során megfogalmazott igények (OpenSearch és Windows tálca bővítmény, dokumentumkezelő illesztés, saját webfelület-tervezés stb.).*

oooOOOooo

Az ADVISE és az IQPortál együttesen teszi megoldhatóvá az elektronikus tartalmak egyidejű keresését a költséges keresőnyelvek előállításának kötelezettsége nélkül. Ugyanakkor a keresések során a rendszer „tanulja” és tárolja az adott szervezet által használt fogalmakat és kapcsolataikat, ezáltal a keresőnyelv mégis automatikusan létrejön a használat során, és kis közreműködéssel további előnyös automatikus rendezési lehetőségeket tesz lehetővé (fogalmi vizualizáció, tématerkép, ontológia támogatása stb.).

A könyvtári szakma számára az ADVISE kereső automatikus fogalmi hálóépítési képessége és vizualizációja jelenti a fő vonzerőt, e képességek beépítése az IQPortál termékbe a legfontosabb eddig megismert igény. Meglátásunk szerint a létrejövő MOKKA rendszer összetett keresési igényeihez is jelentős támogatást nyújthatna az ADVISE, tekintettel arra, hogy bármely szabványos alapokon álló rendszerrel képes integrációra.

Jegyzetek és hivatkozások

- ¹ Az ADVISE mozaikszó az *Adaptive DataWarehouse Search Engine* összetételből ered, a márkanév „ADVISE” formában használatos.
- ² DRESNER, Howard: *The performance management revolution; Business results through insight and action.* New Jersey, Wiley, 2008. 231 p.
- ³ CSIK Tibor – VARGA Katalin: *A tudás és az információfeldolgozás.*
= http://tmt.omikk.bme.hu/show_news.html?id=4007&issue_id=464

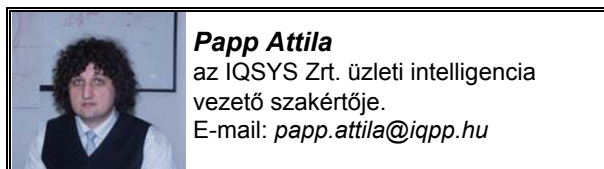
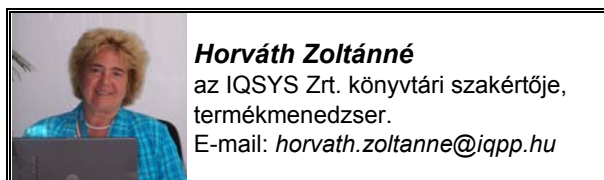
- ⁴ Autonomy, vö.: BÁNHEGYI Zsolt: Vállalati-üzleti információszerzés: a szoftveripar újdonságai. = TMT, 55. köt. 5. sz. 2008. http://tmt.omikk.bme.hu/show_news.html?id=4894&issue_id=493
- ⁵ Adaptivitas: Alkalmazkodóképesség, tanuló rendszerek jellemzője. Kiemelt fontosságú tulajdonság olyan környezetekben, ahol a megoldandó feladat algoritmikusan nem kódolható előre.
- ⁶ Crawling: http://en.wikipedia.org/wiki/Web_crawler ill. http://en.wikipedia.org/wiki/Distributed_web_crawling; Carlos CASTILLO: Effective Web Crawling http://www.webir.org/resources/phd/Castillo_2004.pdf
- ⁷ PageRank: <http://en.wikipedia.org/wiki/PageRank>
- ⁸ Levenshtein-távolság: http://en.wikipedia.org/wiki/Levenshtein_distance; GILLEAND, Michale: Levenshtein Distance, in Three Flavors. <http://www.merriampark.com/ld.htm>
- ⁹ Hashtechnika: Értékek hasítása, olyan hatékonyan (gyorsan) végrehajtható függvényrel, amely során az értelmezési tartomány elemei egyenletesen szórnak a hashértékkészlet tartományában. Alkalmazásával gyorsan kikereshetők értékek, ez adatbázisoknál és a keresés területén egyaránt fontos technológiai kitétel.
- ¹⁰ Elosztott hash-tábla: http://en.wikipedia.org/wiki/Distributed_hash_table; http://en.wikipedia.org/wiki/Hash_function; <http://www.prototypejs.org/api/hash>
- ¹¹ Taxonómiákra nem térünk itt ki, mert több típusról kellene szólni. Ugyancsak nem térünk ki a taxonómia és a tezausz viszonyának taglalására. A vállalati taxonómiák a gépi tartalomrendszerezés céljára készülnek, vagy internetes tartalom rendezésére. L. bővebben: HORVÁTH Zoltánné: Taxonómia – az egyezményes nyelvek szerepe és rokonságai – útközben a szemantikus webhez. = http://tmt.omikk.bme.hu/issue.html?issue_id=472
- ¹² Szemantikus web: Értelmező/intelligens web, olyan világháló, amelyen az információk számítógépes értelmezésre is felkészített formában állnak rendelkezésre, ezzel elősegítve a gépi keresés és egyéb intelligens szolgáltatások készítését.
- ¹³ Az RDF és az OWL nyelvről: [OWL] Web Ontology Language = http://en.wikipedia.org/wiki/Web_Ontology_Language; [RDF] Resource Description Framework = http://en.wikipedia.org/wiki/Resource_Description_Framework
- ¹⁴ yEd: szabad forráskódú Java alapú, hálózati vizualizációs (diagramkészítő) alkalmazásról: http://www.yworks.com/en/products_yed_about.html

Irodalom

- BARÁTNÉ HAJDÚ Ágnes: A percepció és megjelenítés jelentősége az információkereső nyelvekben. 2007. = http://tmt.omikk.bme.hu/show_news.html?id=4785&issue_id=487
- BOGNÁR Katalin: Tudásalapú rendszerek és technológiák. 2006. = http://www.inf.unideb.hu/~bogнар/mestint4/mestint_konyv.pdf
- BRODER, Andrei: A taxonomy of web search. = <http://www.sigir.org/forum/F2002/broder.pdf>
- FAJSZI Bulcsú – CSER László – FEHÉR Tamás: Üzleti haszon az adatok mélyén. Az adatbányászat mindennapjai. Budapest, Alinea, 2010. 414 p. Tartalomjegyzék. = http://www.alinea.hu/pages/uzletihason/adatbanyaszat_tartalom.pdf
- Information retrieval on the WWW and active logic. A. A. Barfoursh et al. = <http://www.lib.umd.edu/drum/bitstream/1903/1153/1/CS-TR-4291.pdf>
- KISS Gergely: Skálázható, intelligens megoldások fejlesztése Java technológiával. = IQSYMPOSIUM, 2009. október 7. Összefoglaló. = <http://www.iqsys.hu/web/guest/iqsymposium-operativ-informaciotecnologia-2009>
- KISS Gergely: Adaptív adattárház újdonságok. ADVISE 1.2. = IQSYMPOSIUM, 2010. április 14. (pdf). Összefoglaló. http://www.iqsys.hu/c/document_library/get-file?44i...
- KOVÁCS László – MICSIK András: Szemantikus webszolgáltatások tervezése és megvalósítása. = http://www.hiradastechnika.hu/data/upload/file/2006/2006_1/HT_0601-4.pdf
- LEHNMANN Miklós: Vizualitás. A képek szerepe a tudományban. = http://www.tofk.elte.hu/tarstud/filmuvtort_2001/lehmann.htm
- Integration and verification of semantic constraints in adaptive process management systems. DADAM, Peter et. al. = <http://www.informatik.uni-ulm.de/dbis/01/dbis/downloads/LRD07.pdf>
- REEVE, Larry: Information retrieval on the semantic web using ontology-based visualisation. = <http://www.pages.drexel.edu/~lhr24/courses/Info780-06Paper.pdf>
- Semantic data integration for the Enterprise. Oracle White papers. = http://www.oracle.com/technology/tech/semantic_technologies/pdf/semantic11g_dataint_twp.pdf
- Taxonomies. Frameworks for Corporate Knowledge. Second ed. by Jan WYLLIE, Trand Monitor... David SKYRME., ed. Simon LELIC, Ark Group. - London, ARK Group, 2005. 80 p. ISBN 0-9549674-1-0

UNGVÁRY Rudolf: Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása. = http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450

Beérkezett: 2010. VII. 11-én.



Jelentkezési felhívás segédkönyvtáros tanfolyamra

A Budapesti Műszaki és Gazdaságtudományi Egyetem Országos Műszaki Információs Központ és Könyvtár (BME OMIKK) emelt szintű OKJ-s segédkönyvtáros tanfolyamot hirdet.

A végzett hallgató munkaköre: segédkönyvtáros.

Az oktatás elsősorban gyakorlati jellegű, amely a vizsgakövetelményekben is érvényesül.

A tanfolyam **2011. januárban**, keresztféléves képzési formában indul.

A képzés időtartama két félév.

A foglalkozásokat heti egy alkalommal csütörtökönként, valamint minden hónap utolsó hetében szerdán és csütörtökön 8-tól 17 óráig tartjuk.

Részvételi díj a két félévre

150 000,- Ft, a vizsgák költsége 50 000,- Ft.

Felvételi vizsga nincs, a beiratkozás feltétele az érettségi bizonyítvány bemutatása.

A tanfolyam jegyzeteit, segédkönyveit kölcsönzés formájában biztosítja a szervező intézmény.

A képzésre azoknak a jelentkezését várjuk, akik a könyvtári munka gyakorlatát rövid idő alatt kívánják elsajátítani, és a számítógép használatában négy ECDL modul megismerésével jártasságot akarnak szerezni.

Jelentkezni az alábbi címre eljuttatott (kitöltött, kinyomtatott) jelentkezési űrlappal lehet:

BME OMIKK
segédkönyvtáros képzés
1111 Budapest, Budafoki út 4-6.

A jelentkezési űrlap a BME OMIKK honlapjáról letölthető

<http://www.omikk.bme.hu/main.php?folderID=1159&articleID=1816&ctag=articlelist&iid=1>

Jelentkezési határidő: **2010. december 15-ig**

További felvilágosítás **463-3534**-es telefonszámon és a gylengyel@omikk.bme.hu e-mail címen Lengyel Gyöngyitől kérhető.

Bánki Zsolt István

Múzeumok az Europeanában – az ATHENA projekt

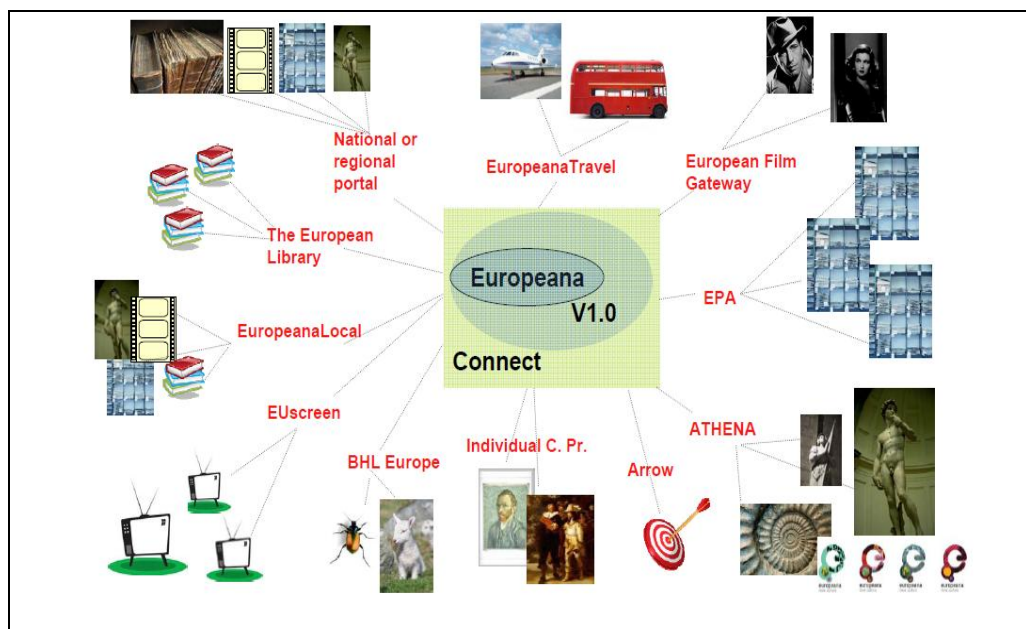
Hozzáférés az európai kulturális örökség hálózataihoz

A TMT 2007/9. számában még borúlátó hangon írtam a múzeumok informatikai állapotáról, hiányoltam a nemzetközi fejlesztéseket, projekteket. Jelen tanulmány azonban már örvendetes fejleményekről kíván beszámolni, mind a hazai, mind a nemzetközi trendeket illetően. Az eContentPlus projektcsalád új tagjaként az Athena a múzeumok digitális tartalmainak integrációját célozza az Europeanába, új szabványosítási és szervezési ajánlásokat téve a muzeális gyűjtemények számára. Magyarországi résztvevőként a Petőfi Irodalmi Múzeum (PIM) mint tartalomszolgáltató (content provider) saját digitalizálási gyakorlatában és technológiai fejlesztéseiben meghonosítja az Athena ajánlásait, s ezáltal – a magyar múzeumok közül elsőként – tartalmat juttathat el az Europeanába.

Az ATHENA az uniós projektek családjának tagjaként

Az Europeana és az ATHENA¹ is az Európai Bizottság által támogatott projekt. Mindkettő annak érdekében tevékenykedik, hogy az Európa-szerte működő és az európai kulturális örökséget feldolgozó intézmények ezreiből származó digitális tartalmak azonnal hozzáférhetővé váljanak a felhasználók számára szerte a világon.

Az ATHENA az Europeana család múzeumi tagja (1. ábra). Olyan európai múzeumi hálózat, amelynek célja, hogy múzeumi tartalmakat szüreteljen és továbbítson az európai digitális könyvtárba. A vállalkozás két, kormányzati keretek között 2001-ben indított projektből fejlődött ki: a MINERVA² és a MICHAEL³ hálózatokból. Mindkét hálózat továbbra is támogatja a múzeumokat azon munkájukban, hogy tartalmakat szolgáltatassanak az Europeana számára.



1. ábra Az Europeana-család felépítése

A MINERVA projekt céljait az OSZK (Országos Széchényi Könyvtár) vonatkozó honlapja⁴ így határozza meg: „A MINERVA az európai uniós tagállamok minisztériumai közötti együttműködés, mely a kulturális és tudományos tartalmak digitalizálását hangolja össze. A program célja egy közösen elfogadott európai platform kialakítása a digitalizálásra, metaadatokra és a hosszú távú megőrzésre vonatkozó ajánlások és irányelvek megfogalmazása formájában. A nemzeti programok koordinálása mellett az európai országok, nemzetközi szervezetek és projektek közti kapcsolatok kiépítését is segíti. A program ezen felül tanácsadó szerepet lát el, mely a már működő jó példák összegyűjtését, valamint a lundit alapelvek elfogadását és terjesztését szorgalmazza.” A MINERVA több területen egyfajta modellként szolgál az ATHENA számára, aminek egyik példája a projekten belül alkalmazott munkamegosztás, vagyis a nemzeti munkacsoportok tevékenységére épített európai együttműködés. A különféle szakterületekről érkező, különféle kompetenciákkal rendelkező szakértőkből épülnek a nemzeti Athena-munkacsoportok.

A MICHAEL projektben 17 európai ország szervezetei készítik el saját kulturális gyűjteményi nyilvántartásukat, s kapcsolják össze az egyes adatbázisokat. A MICHAEL-portálon 16 nyelven kereshető mintegy 4000 európai kulturális intézmény 10 000 digitális gyűjteménye. A szolgáltatás elsősorban oktatási, kutatási és turisztikai célokra jött létre. A nemzeti MICHAEL-portálok feladata, hogy feltérképezzék a digitális kulturális örökséget, új tartalmakat és intézményeket fedezzenek fel, s átláthatóvá tegyék a szabványok alkalmazását. A MICHAEL-szolgáltatás révén lehetővé válik, hogy egyre több, digitális gyűjteménnyel rendelkező múzeum kapcsolódjon az ATHENA-hálózathoz is.

Az ATHENA projekt – építve a fent bemutatott előzményekre – az eContentplus program keretében egy úgynevezett „Legjobb Gyakorlat Hálózatot” (*Best Practice Network*) kíván megvalósítani. Szorosan együttműködik már létező Europeana projektekkel, és intenzív, klaszter jellegű kooperációt alakít ki más releváns területekkel (pl. EuropeanaLocal). A projekt végrehajtása 2008. november 1-jén indult, s 30 hónapig tart.

Résztvevők

A projekt indulásakor, 2008-ban, 109 intézmény vett részt az ATHENA munkájában (20 EU-tagországból, valamint Izraelből, Oroszországból

és Azerbajdzsánból). A 26 partnerszervezet között – közvetlen vagy társult partnerként – hat kulturális minisztérium, több mint 100 múzeum és egyéb kulturális intézmény található, amelyek 20 európai nyelvet reprezentálnak. A projekt koordinációját az olasz Kulturális Örökség Minisztériuma végzi.

Magyarországot az Athena indulásakor az OKM Közgyűjteményi Főosztályának felkérése alapján egy múzeum (Petőfi Irodalmi Múzeum) és egy archívum (*Magyar Rádió Archívuma*) képviselte. 2010 elején csatlakozott a projekthez a Szépművészeti Múzeum és a Magyar Nemzeti Múzeum is.

Célkitűzések

A projekt kiemelt feladata egy olyan műszaki infrastruktúra kifejlesztése, amely lehetővé teszi az Europeanával való szemantikus együttműködést. Ehhez testre szabható segédeszközöket, ajánlásokat és irányelveket kíván létrehozni, amelyek a többnyelvűségre, szemantikára, metaadatokra, tezaurszokra, adatstruktúrákra vonatkoznak, s ezáltal segítik a múzeumok belső digitalizáló tevékenységét, valamint előmozdítják digitális tartalmaik integrálását az Europeanába.

Az ATHENA közös szabványok alapján kíván hozzájárulni az Europeana felépítéséhez, katalizálva a múzeumi szabványosítás folyamatát, és a már létező szabványok közötti kommunikációs platformok létrejöttét. (Ebben a tekintetben a MINERVA munkáját folytatja a digitalizáció és a metaadatolás szabványainak és irányelveinek harmonizációja területén.) Kiemelten épít a MICHAEL és a nemzeti hálózatok által összegyűjtött tapasztalatokra, de tovább lép az együttműködési képesség, a szemantikus web és a minőségi szolgáltatások irányába.

Tevékenységek

2009-ben megtörtént a metaadatsémák, szabályozott terminológiák és tezaurszok feltérképezése (1. táblázat).

A felmérés a következő kérdéskörökhöz szerveződött:

- A gyűjtemény leírása a Michael mintájára (mit tartalmaz, milyen nyelven).
- Mivel járul hozzá az adott intézmény az Athenához: digitális objektumokkal és/vagy nézőképekkel?

- A digitális objektumok milyen metaadatkörrel rendelkeznek?
- Szöveg, kép, audió, videó fájlformátumok.
- Milyen metaadatsémát, szabványokat alkalmaz az intézmény?

1. táblázat

A felmérésben felsorolt szabványok

| | | |
|-------------|---|--|
| CDWA | Categories for the Description of works of Art | http://www.getty.edu/research/institute/standards/cdwa/index.html |
| CIDOC-CRM | CIDOC Conceptual Reference Model | http://cidoc.ics.forth.gr |
| Dublin Core | The Dublin Core Element Set Version 1.1 | http://dublincore.org/documents/1999/07/02/dces/ |
| EAD | Encoded Archival Description | ftp://ftp.loc.gov/pub/ead/ead.dtd (DTD) http://www.loc.gov/ead/ead.xsd (W3C schema) |
| FRBR | Functional Requirements for Bibliographic Records | http://www.ifla.org/VII/s13/frbr/frbr.pdf |
| ISAD(G) | ISAD(G): General International Standard Archival Description, Second Edition | http://www.ica.org (ica website) |
| MAB2 | Maschinelles Austauschformat für Bibliotheken | http://www.d-nb.de/standardisierung/txt/titel/mab.txt http://www.d-nb.de/eng/standardisierung/formate/mab.htm |
| MARC | MAchine-Readable Cataloguing | http://www.loc.gov/marc/bibliographic/ecbdhome.html |
| METS | Metadata Encoding and Transmission Standard | http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/standards/mets/mets-schemadocs.html (METS website) |
| MIDAS | MIDAS Heritage | http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_One.pdf http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_Two.pdf http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_Three.pdf http://www.fish-forum.info (FISH website) http://www.heritage-standards.org.uk (FISH Interoperability Toolkit) |
| MODS | Metadata Object Description Schema | http://www.loc.gov/standards/mods/ |
| Museumdat | Fachgruppe Dokumentation im Deutschen Museumsbund | http://museum.zib.de/museumdat/museumdat-v1.0.xsd http://museum.zib.de/museumdat/museumdat-v1.0-en.pdf (documentation) http://www.museumdat.org/index.php?ln=de&t=home (German website) |
| Object ID | | http://www.object-id.com/guide/guide_index.html http://www.object-id.com/checklist/check_eng.html (checklist) |
| SPECTRUM | SPECTRUM: The UK Museum Documentation Standard, 3 rd Edition (Version 3.1) | http://www.collectionstrust.org.uk/spectrum (UK version) http://www.collectionstrust.org.uk/spectrum-nl (Netherlands version) http://www.collectionstrust.org.uk/spectrum-nl-be (Flanders version) http://www.collectionstrust.org.uk/schema (XML schema) |
| TEI | TEI P5: Guidelines for Electronic Text Encoding and Interchange | http://www.tei-c.org/P5X/ http://www.tei-c.org (Text Encoding Initiative Consortium website) |
| VRA | VRA Core (Version 4.0) | http://www.vraweb.org/projects/vracore4/index.html (website) http://www.vraweb.org/projects/vracore4/vra-4.0.xsd (XML schema) |

Szellemi jogok, szerzői jogok (IPR)

A gyűjtemény felhasználási szempontú jogi státusa: public domain / licensed. License methods: Creative Commons; European Model EDI Agreement; Open access licenses; Orphan works (árva művek).

Használt ontológiák, teauruszok, azok szabványossága: Földrajzi névterminológia és koordináták. Dátum- és korszak-terminológia. Tárgyszó-terminológia. Személy- és intézmény-terminológia.

Hogyan történik az Europeana-hozzjárulás?

Egyéni URL, garantált időtartam / más azonosító (katalógusszám, NBN, DOI, PURL, ARK, Open URL) / FTP hozzáférés / OAI-PMH harvest / keresőprotokoll (pl. Z39.50, SRU vagy SQL) / Integrált múzeumi rendszer.

Ki látja el az aggregátor-funkciót?

A munka eredményeként elkészült dokumentumok⁵ összefoglalóan elemzik az európai múzeumokban a kulturális örökség tartalmi feltárására, leírására alkalmazott ontológiákat, teauruszokat, a technikai és adatcsereszabványokat és protokollokat. A felmérés eredményeképpen ajánlásokat és irányelveket dolgoztak ki az intézmények technológiai fejlesztésére.

Az ATHENA a múzeumok tekintetében kiemelten foglalkozik a szerzői jogokkal kapcsolatos problémák elemzésével és a publikációs lehetőségek jogszerű megoldásával⁶. Technikai infrastruktúrát alakít ki az adatok Europeanaba való integrálására. Aggregációs feladatot vállal a múzeumi tartalmak összegyűjtése és továbbítása céljából, és ezzel a technológiailag elmaradottabb intézmények számára is megteremti a lehetőséget, hogy Europeana-partnerekké válhassanak. Az összegyűjtött tudás továbbadása céljából kompetenciaközpontokat alapít minden részt vevő országban, és aktív disszeminációs tevékenységet folytat. Magyarországon is több múzeumi-informatikai szakmai napot, előadást tartottunk, a múzeumi szakemberekkel megismertettük a célkitűzéseket és a várható eredményeket.

Munkacsoportok

Az Athena, az európai uniós programok metodikájának megfelelően, munkacsoportokban végzi tevékenységét:

- WP1 Projektmenedzsment, monitoring és értékelés / MiBAC (Olaszország).
- WP2 Figyelemfelkeltés és disszemináció: a hálózat kiterjesztése és a szolgáltatás promóciója / SPK (Németország) – ICIMSS (Lengyelország).
- WP3 Sztenderdek azonosítása és ajánlások kidolgozása / RMAH (Belgium) – Collections Trust (Egyesült Királyság).
- WP4 A meglévő adatstruktúra integrációja az Európai Digitális Könyvtárba / Michael-Culture (Aisbl).
- WP5 Tartalmak koordinációja / MiBAC (Olaszország) – ICCS-NTUA (Görögország).
- WP6 Jogi kérdések elemzése a szellemi tulajdonjog (IPR) területén, lehetséges megoldások meghatározása / Packed (Belgium) – UP (Görögország).
- WP7 Plugin-ek fejlesztése az Európai Digitális Könyvtárba való integráció számára ICCS-NTUA (Görögország) – MKRS (Szlovénia).

Az Athena a Petőfi Irodalmi Múzeumban

A Petőfi Irodalmi Múzeum részt vett a korábban említett felmérésekben, valamint kiemelten a WP2, WP3 és a WP4 munkacsoportok munkáiban, és ellátja a „nemzeti kontakt pont” feladatát, koordinálva a társintézmények bekapcsolódását az Athenába. Saját gyűjteményének áttekintésével adatokat szolgáltat a digitális gyűjtemények, valamint a metaadatszabványok és ontológiák, teauruszok felméréséhez.

Az Athena egy „magteauruszt” kíván fejleszteni, 300-500 lexikai egység kiinduló állománnyal a múzeumi objektumok tartalmi feltárására. Ez a teaurusz az Europeana elveinek megfelelően, nagy hangsúlyt helyez a többnyelvűségre. Kezdeti formájában az angol és az olasz nyelv mellé a magyart választották a „work package” vezetői. A munkafolyamat keretében – az Országos Széchényi Könyvtár jóváhagyásával – a fejlesztők rendelkezésére bocsátottuk a KÖZTAURUSZ SKOS formátumát, és részt veszünk a három nyelvi verzió megfeleltetésében.

A szabványosítást kiemelt területként kezeljük, mivel a magyar múzeumokban alkalmazott rendszerek döntő többségükben nem alkalmaznak nemzetközi szabványos formátumokat. Az Athena elősegítette, hogy ismereteink jelentősen bővüljenek a létező nemzetközi múzeumi szabványok terén. Ahhoz, hogy megfelelő adatszerkezetben legyünk képesek tartalomszolgáltatásra, saját in-

tegrált múzeumi rendszerünket – amely alkalmas HUNMARC adatcserére – fel kellett készíteni az Athena által preferált szabványos formátum előállítására.

Ehhez kiindulópontnak az egyszerű, *Dublin Core* (DC) alapuló *Europeana Semantic Elements (ESE)*⁷ tekinthető, de jelenleg a projekt az ennél lényegesen bővebb adatstruktúrájú *LIDO harvest*⁸ formátumot alkalmazza.

A múzeumi szabványosítás helyzete, hazai fejlesztések; az Athena hozadéka

Bár a projekt indulásakor a WP4 vezetői még a Dublin Core (DC) alkalmazásában gondolkodtak, hamar felismerték, hogy a DC nem elég részletes és alapos a múzeumi adatcseréhez. Ezért szükségessé vált egy új szabványos formátum kidolgozása – ez lett a LIDO. A formátum kidolgozásánál az alábbi, már meglévő standardokat vették figyelembe: CDWA Lite; CIDOC CRM; Museumdat; Spectrum.

CDWA Lite

A *CDWA Lite (Categories for the Description of Works of Art)*⁹ egy XML forma OAI PMH adat aratásához. Elsősorban művészeti és egyéb, emberek által készített tárgyak leírására szolgál. Megkülönböztet indexelt (kereshető) és csak megjelenítésre szolgáló elemeket.

Cidoc CRM

A *Cidoc CRM (Comité International pour la Documentation – Conceptual Reference Model)*¹⁰ fogalmak közötti kapcsolatok modellje, egy speciális CRM a múzeumi dokumentáció számára. Döntően ez is emberek által alkotott objektumok leírására szolgál. Célja egy szemantikusan definiált hajlékony keretrendszer kialakítása, amelyre bármilyen adat felfűzhető. Lehetőséget kíván adni többféle forrásból származó információk összekapcsolására. Maga a CIDOC CRM nem egy adatleíró forma katalógusok kialakítása számára, hanem egy leíró nyelv, amelynek segítségével adatleíró formákat lehet definiálni. Nem tekinti feladatának a természetes nyelvi szövegek géppel feldolgozható formába való konvertálásának támogatását, a keresőkérdések megfogalmazását viszont elősegíti. A CIDOC CRM a lehető legszélesebb körben támogatja a különféle múzeumi adattartalmak integrálását és azonosítását, mind a valóság-

gosan összetöltött adatbázisok, mind a virtuális adatok kereshetősége tekintetében. További előnye, hogy lehetővé teszi nagy adattömeg kezelését is. Útmutatást ad CRM-kompatibilis adatszerkezetek tervezésére, melyek veszteségmentesen tudnak előállítani és fogadni metaadatokat a CIDOC CRM keretben leírtak szerint.

Museumdat

A *Museumdat (LIDO XML)*¹¹ elsősorban német nyelvterületen elterjedt szabványalkalmazás. A CIDOC CRM és a CDWA Lite formátumot egyaránt használja. A kulturális örökség tárgyai mellett a természetben fellelt objektumok leírására is alkalmas. Fontos jellemzője a többnyelvűségre (szótárak), adatkonverzióra és a besorolási állományokra való kiemelt odafigyelés. Az eseményszemléletű adatmegközelítés (osztályozás, azonosítás, leírás, esemény, kapcsolat) a CIDOC CRM-logikára alapszik. A múzeumi adatelemek számára harvest formátumot alkalmaz a webservice támogatására.

LIDO

A LIDO szüretelési séma, nem katalógus-szerkezet; nem szolgálja a tárgymozgatás, beszerzés és nyilvántartás céljait. Minden eseményt önálló információs egységnek tekint: a létrehozást, a gyűjtést, a használatot. A hozzájuk kapcsolódó adatokat, valamint az adatforrásokat egyaránt leírja, és egy kapcsolati-fogalmi hálóban kezeli.

Spectrum

A legjelentősebb brit múzeumi gyűjteménykezelői szabvány¹². Két alapvető adattípust különböztet meg: a procedúrákat és az információs anyagokat. A procedúrákon belül 21 aktivitási típust definiál. Az információs elemek esetében minden olyan adatot ideért, amely informatív az objektummal kapcsolatban. Ezekből információs csoportokat képez. Fő erénye, a munkafolyamat lekövetése, valamint, hogy az információs elemeket kapcsolati hálóban kezeli, és a hozzáférési jogok lehető legteljesebb kezelésére törekszik.

Múzeumi szabványok alkalmazása a Petőfi Irodalmi Múzeumban

A PIM, mivel HUNMARC alapú rendszere van, azt a célt tűzte ki, hogy az Athena ajánlásainak megfelelő szabványkörből választva szabványos múzeumi outputot hoz létre. Mivel a projekt LIDO-ban

fogadja a metaadatrekordokat, előnyös volt olyan szabványt választani, amely már rendelkezik LIDO mappinggel. Ebből a helyzetből kiindulva a Museumdat mellett döntöttünk, s a Huntéka fejlesztőivel együttműködve készül a *HUNMARC–Museumdat–LIDO* megfeleltetés. A fejlesztés eredményeképpen megvalósul az első magyar szabványos múzeumi rendszer, amely standard outputon, OAI-PMH harvesten keresztül képes adatcserére.

Az Athena mint aggregátor

Az Athena, felismerve azt, hogy a múzeumok Európa-szerte rendkívül heterogén felkészültségi állapotban vannak az Europeanaba való tartalomszolgáltatás tekintetében, és gyakorta saját digitális tartalmaik archiválására és hosszú távú megőrzésére is alkalmatlanok, önmagát is aggregátorként definiálja. Erre a célra megteremtette a szükséges IT infrastruktúrát, és saját tartalomfeltöltési rendszert fejlesztett ki. A partnerek ezen a szolgáltatáson keresztül juttathatják el tartalmaikat az Athenába, ahol archiválják, majd továbbítják őket az Europeanaba. A feltöltés logikája figyelembe veszi a tartalomszolgáltatók sokszínűségét, vagyis a tartalomszolgáltatás lehetősége az offline adatszolgáltatástól egészen az online szüretelésig terjed.

Az Athenában felépített modell felveti a hazai múzeumi aggregáció helyzetének kérdését. Ha a teljes közgyűjteményi szférát vizsgáljuk, akkor azt láthatjuk, hogy a magyarországi aggregációs feladatok megfogalmazására az elmúlt két-három évben került sor.

Korábban a nagytömegű digitalizációs projektek tervezése, és a digitalizálandó dokumentumok körének meghatározása állt a közgyűjtemények figyelmének középpontjában. Ezek a törekvések döntően a könyvtári és az audiovizuális gyűjteményekre korlátozódtak, s csak az elméleti előkészítés szakasza tekinthető sikeresnek, mivel nem jutottak el a megvalósítás fázisához. A közgyűjteményi szféra múzeumi és levéltári szegmensében a feladatmeghatározás és a követendő technológiák tekintetében is lemaradás tapasztalható.

Az intézményi felkészültség területén az Országos Széchényi Könyvtár és a *Neumann Kht.* tett konkrét lépéseket a digitális anyagok gyűjtésére és szolgáltatására. Az OSZK-ban a *Magyar Digitális Képkönyvtár*, a Neumann Kht.-ban az *Nemzeti Digitális Adattár (NDA)* és a *Nemzeti Audiovizuális*

Archívum (NAVA) működése tekinthető aggregációs elven alapuló szolgáltatásnak.

A múzeumok és a levéltárak a mai napig csak a felkészülés fázisában járnak. A kijelölt intézmények (Magyar Nemzeti Múzeum, *Magyar Országos Levéltár*) ugyan vállalták az aggregációs feladatot, technológiai felkészültségük azonban még nem teszi őket alkalmassá a szolgáltatás elindítására.

A hazai szakmai közösségekben az elmúlt években kiérlelődött az az álláspont, hogy az aggregációt ne egyetlen országos aggregátor lássa el, hanem közgyűjteményi áganként egy-egy kijelölt intézmény. Hangsúlyos törekvésként jelent meg az a szándék, hogy az aggregációs modellbe a közgyűjteményi szakmai szereplők mellé kívánatos volna bevonni egy technológiai partnert is, amely rendelkezik a megfelelő *know-how*-val, IT eszközpárral, ugyanakkor nem piaci szereplő. Erre a feladatra az *NIIIF (Nemzeti Információs Infrastruktúra Fejlesztési Program)* tűnik a legalkalmasabbnak. Egy ilyen modell megvalósíthatja a digitális objektumok, illetve a rájuk vonatkozó metaadatok szakterületi gyűjtését, biztonságos, elosztott rendszerben való tárolását, valamint a gazdaságos megvalósíthatóság és a működtethetőség követelményét is.

A Magyar Nemzeti Múzeum az elmúlt évben létrehozta a *Múzeumi Digitalizációs Bizottságot*, amelynek elsődleges feladata kialakítani a magyar múzeumi aggregációs tevékenységet az Athena-modell tanulmányozásával. A Nemzeti Múzeum csatlakozása az Athenához kiváló alapot teremt ehhez. Már születtek döntések a megvalósítás technológiai és szabványos alapjait illetően, hogy a létrejövő szolgáltatás alkalmas legyen a hazai és a nemzetközi együttműködésre. Remélhetőleg az alapszolgáltatás megvalósulhat az Athena futamideje alatt.

Europeana update – a Rajna- és Duna-szakasz

Az Europeana tartalma jelenleg 5,9 millió tétel – az összes európai tartalomszolgáltatót figyelembe véve. Ennek dokumentumtípusonkénti megoszlása:

- 4 100 000 kép: fotó, festmény, grafika, képeslap, plakát;
- 1 700 000 szöveg: könyv, időszaki kiadvány, részdokumentum (cikkek, tanulmányok), kézirat (levelezés, analekta);

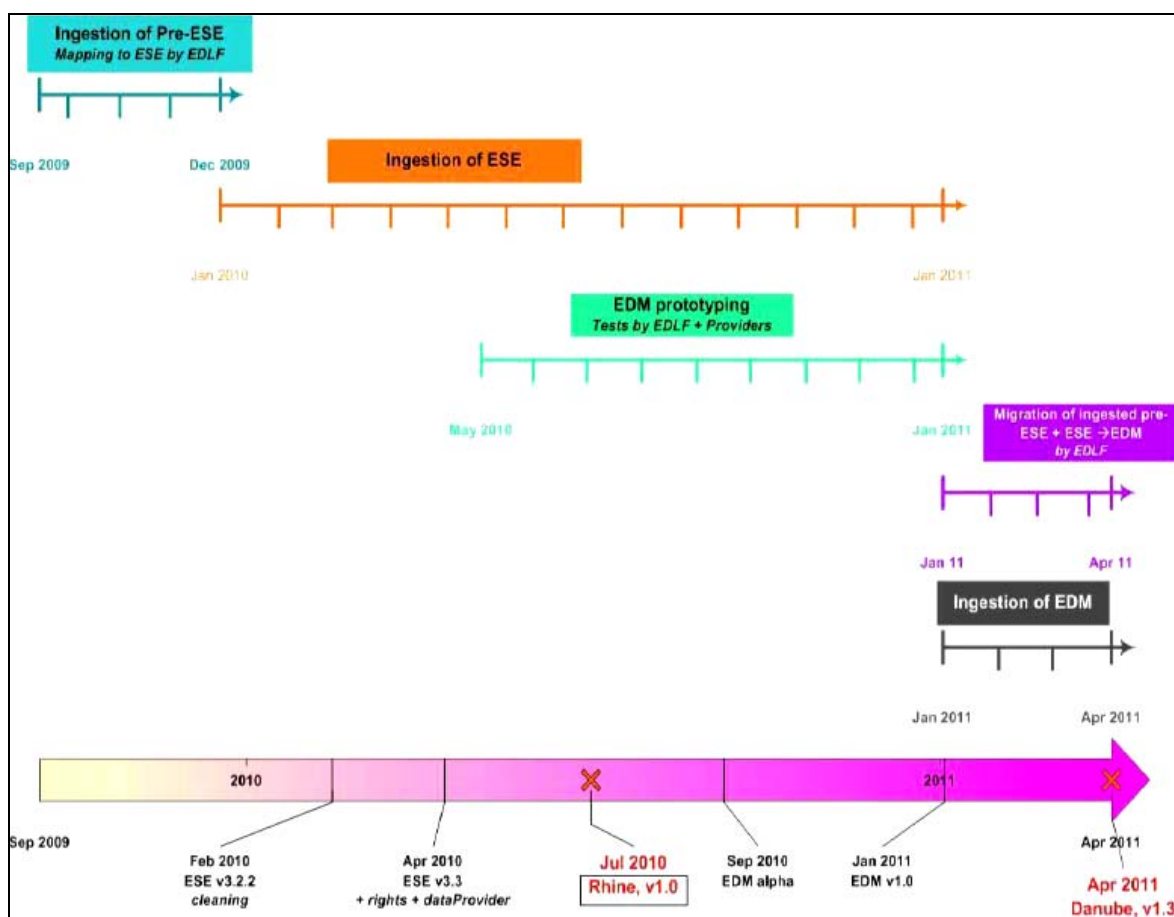
- 92 000 mozgókép: játék- és dokumentumfilm, TV-felvétel, ismeretterjesztő film;
- 22 000 hangfelvétel: fonográf-henger, hanglemezek, rádiófelvétel, stúdió kívüli felvételek.

Az *Europeana Foundation*¹³ célkitűzése, hogy a tartalomfejlesztés a korábinál intenzívebb lendületet kapjon. Ennek érdekében két feltöltési periódust határoztak meg (2. ábra). A két szakaszt Európa két nagy folyamáról Rajna- és Duna-szakaszok keresztelték.

A *Rajna-szakaszban* rövidtávú célként a tartalomfejlesztésre helyezik a hangsúlyt, mintegy 10 millió tételre növelve az Europeana gyűjteményét, erőteljesebben reprezentálva ezzel az európai kulturális örökséget, mélyebben és arányosabban integrálva az európai nemzeteket, nyelveket és a fent említett

dokumentumtípusokat. A *Rajna-szakasz* megvalósításának időpontja 2010 nyara. A *Petőfi Irodalmi Múzeum* a *Rajna-szakaszban* az Athenan keresztül juttatja el az *Europeana*-ba a *Digitális Irodalmi Akadémia* gyűjteményét.

A 2011 tavaszára tervezett *Duna-szakasz* hangsúlyosan a technológiai és a szolgáltatásfejlesztésre koncentrál a tartalomfejlesztés mellett. Új funkciókat tervez a többnyelvű és a szemantikus keresés terén, olyan, a felhasználók által személyre szabható szolgáltatásokkal, mint a multimédia-tartalmak annotálhatósága, vagy az *e-book on-demand* szolgáltatás. Ebben a szakaszban a PIM múzeumi gyűjteménye kerül az Athenába, illetve az *Europeana*-ba.



2. ábra Az Europeana fejlesztési ütemezése

Hivatkozások

- 1 ATHENA = Access to Cultural Heritage Networks Accross Europe / <http://www.athenaeurope.org/index.php?en/1/home>
- 2 Minerva = Ministerial Network for Valorising Activities in Digitisation / <http://www.minervaeurope.org/>
- 3 Michael = Multilingual Directory of Cultural Heritage in Europe / <http://www.michael-culture.org/hu/home>
- 4 <http://www.mek.oszk.hu/minerva/>
- 5 ATHENA Deliverables and documents / <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>
- 6 WP6 - Analysis of IPR issues and definition of possible solutions / <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>
- 7 http://version1.europeana.eu/web/guest/provide_content/
- 8 LIDO = Light Information Describing Objects / <http://www.athenaeurope.org/getFile.php?id=535>
- 9 http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html
- 10 <http://www.cidoc-crm.org/>
- 11 <http://www.museumdat.org/index.php?ln=en&t=home>

¹² <http://www.collectionstrust.org.uk/stand>

¹³ <http://version1.europeana.eu/web/guest/edl-foundation/>

Irodalom

<http://www.athenaeurope.org>

<http://vod.niif.hu/player/index.php?q=1052/500K>

http://tmt.omikk.bme.hu/show_news.html?id=4776&issue_id=486

<http://www.digitmus.hu/>

(A tanulmány elkészítése kapcsán köszönetet mondok Simon Andrásnak, aki múzeumi szabványok feltérképezésében úttörő tevékenységet folytat. Segítsége nélkül a szabványok bemutatása nem valósulhatott volna meg.)

Beérkezett: 2010. VII. 5-én.



A blogokat Oroszországban egyszerűen csak ötödik hatalomnak nevezik

Az elnevezés utal a blogok kiemelkedő befolyására. *Rusztsem Rinatovics Adagamov* az egyik legismertebb blogger, oldalát naponta 600 ezren olvassák. Több olvasója van, mint sok moszkvai napilapnak, amilyen például a *Vedomosztija*, az *Izvesztijja* vagy a *Kommerszant*. A titok nyitja, hogy megpróbálja a világ eseményeit fotómontázsokkal megjeleníteni. Ezek lehetnek saját felvételek vagy a *Reuters* képei. Az emberek szeretik ezeket a beszámolókat, hiszen gyorsan megértik a hírek lényegét. Sokaknál Adagamov blogja helyettesíti a napilapokat.

Oroszországban több mint 7,5 millió blog van. Ennek oka, hogy régen, még a Szovjetunióban kizárólag a konyhában lehetett szabadon beszélni. Ma a webnek hasonló szerepe van. Még mindig nincsenek hangsúlyosan jelen a politikai pártok, illetve csak a fikciók szintjén léteznek.

„*Alekszej Navalnyij* barátom rendszeresen blogol az állami cégeknél és hivataloknál tapasztalható korrupcióról. Bátor, félttem is őt. Nálunk bizonyos embereket sokkal kisebb dolgok miatt is eltettek már láb alól. Nem vagyok pesszimista, csupán realista. Bizonyos dolgokba egész egyszerűen nem szeretem beleütni az orromat. Úgy gondolom, hogy ugyan nem fogjuk követni Kína példáját, de sokáig nem hagyják majd a szabad véleménycserét. Ez az utolsó szabad territórium, de már nem sokáig. Amíg viszont egy olyan ember ül a Kremlben, mint *Medvegyev*, aki szereti az internetet, addig talán lelassítható ez a folyamat” – jelentette ki Adagamov.

/SG.hu Hírlevél, 2010. szeptember 2., <http://www.sg.hu/>

(SzP)



COUNTER és SUSHI: a használói statisztika csúcspogramjai

Bízhatunk-e a szemünk előtt lévő tartalomban? Ez a kérdés mindig is életbevágó volt a tudományos kommunikáció szereplője számára – ám egyszerű válasz nincs rá. A tartalom iránti bizalmat hagyományosan a szerző hírneve, a kutatásának helyet adó intézmény reputációja, a tanulmányát közlő folyóirat státusza, illetve az adott kiadó tekintélye garantálta. Újabban a citációs adatok, legfrissebb fejleményként pedig a használói statisztikák váltak a megbízható tartalom népszerű indikátorává. Ez nem véletlen, hiszen az adat ma a döntési folyamat központi eleme, s az adatok idézése és használata komoly tényező a tartalom hatásának, értékének és megbízhatóságának felmérésében.

A hálózati elektronikus források online használatának számlálóeszköze, a *COUNTER* (az angol mozaikszó kibontva: *Counting Online Usage of Networked Electronic Resources*) (<http://www.projectCounter.org>) és a szabványos használati statisztikákat begyűjtő kezdeményezés, a *SUSHI* (*Standardized Usage Statistics Harvesting Initiative*) (<http://www.niso.org/workrooms/sushi>) egymást kiegészítő programok a használói statisztika megbízhatóságának javításához. A *COUNTER* garantálja a használói statisztika hitelességét, az *Amerikai Szabványügyi Hatóság (NISO)* által támogatott *SUSHI* pedig a statisztikákhoz való szabványos, egyszerű hozzáférést teszi lehetővé.

A *COUNTER* folyóirat- és adatbázis-üzemelési kalauzájának 3. kiadása 2009 szeptemberében váltotta le a korábbi kiadást. A friss változatban kifejezetten a könyvtári konzorciumok számára fokozták a *COUNTER* használói jelentések megbízhatóságát, egyszerűsítették a periodikaarchívumok használati követését, könnyebbé vált a statisztikák kezelése és elemzése; a *SUSHI* protokoll pedig már integrált elemként tartozik bele. Tudnivaló, hogy a *SUSHI* előtt nem létezett mechanizmus a különböző forrásokból származó *COUNTER* használói adatok automatikus visszanyerésére és táro-

lására: a *SUSHI* épp a statisztikabegyűjtés automatizálásának szabványos modelljévé vált.

A szolgáltatóvállalatok nyomban honosítani kezdték a *COUNTER* 3. kiadását. 2010 elejére a zökkenőmentes alkalmazást nyújtó vendorok száma több mint 100: ezzel mintegy 16 ezer teljes szövegű folyóirat online használatának mérése már a *COUNTER* legfrissebb verziójával történik. A megbízható használati statisztikákat kínáló *COUNTER* és *SUSHI* programokat a könyvtárosok a szerzeményezés segédeszközeként, a folyóiratok és az adatbázisok hasznosságának feltérképezésére használják, a kiadók pedig azokkal demonstrálják gyűjteményeik értékét. Mára rutinszerű mérőeszköz lett a letöltések szerinti költség (cost per download), illetve a használók száma szerinti költség (cost per FTE).

A két projekt sikere nyomán a szerzők, a kutatóintézetek és a kutatástámogató ügynökségek egyaránt felismerik a használói statisztikák értékét. A megbízható statisztikai mérés tökéletesítése érdekében a *COUNTER* két, jelenleg futó brit irányítású projektben is szerepel. 1.) A *Journal Usage Factor Project* (Folyóirat-használati faktor projekt) (<http://uksg.org/usagefactors>) a citációalapú impaktfaktor kiváltását szorgalmazza, annak használati alapú megfelelőjét igyekszik kifejleszteni. A globális használói mérőeszköz nem csupán a folyóirat impaktfaktor alternatívája lehetne, hanem az impaktfaktorral le nem fedett folyóiratok számára mennyiségi mutatóként is működhetne. 2.) A *PIRUS* projekt (Publisher and Institutional Repository Usage Statistics) (<http://tiny.cc/PIRUS>) célja egy, a *COUNTER*-rel kompatibilis szabvány előmozdítása az egyes cikkek regisztrálásához és használati méréséhez.

A tartalom megbízhatóságára vonatkozó kérdésre végül is azt a választ adhatjuk, hogy a használói adatok önmagukban nem tudják garantálni a tartal-

lom minőségét, de számos más tényezővel karöltve, egy mátrix elemeként igenis fontos indikátora lehet a tartalom hasznosságának és megbízhatóságának.

/SHEPHERD, Peter: COUNTER and SUSHI: Providing accessible usage statistics deserving your trust. = Library Connect, 8. köt. 1. sz. 2010. p. 6./

(Bánhegyi Zsolt)

British Library – a dokumentumküldés útelágazás előtt

2011-ben ünnepli a *British Library* (BL) a dokumentumküldés 50. évfordulóját. Az ünneplés válással jár. Számos kihívással áll ugyanis szemben a szolgáltatás az utóbbi évek változásai miatt, például mások a piaci elvárásai, a *CISTI* és az *Infotrieve* bejelentette együttműködését, megváltozott a szerzői jogi szabályozás Németországban.

Egyre inkább az elektronikus tartalom felé fordulnak a kutatók, új üzleti modellek jelennek meg: nyílt hozzáférés (open access), használat szerinti fizetés (pay to view, más forrásokhoz hozzáférés a Google révén. Nem a könyvtári katalógusokkal kezdik a kutatást, sőt a katalógusok szerepe egyre inkább háttérbe szorul.

A BL dokumentumküldő szolgálata ismeri a jelenségeket és egyetért Darwin kijelentésével: nem a legerősebb és legintelligensebb egyedek a túlélők, hanem azok, amelyek legjobban tudnak alkalmazkodni. Épp ezért 2006-ban a BL áttekintette a helyzetet és megalapította a *Dokumentumküldés jövője csoportot* (*Document Supply Futures = DSF*) a megújulás előkészítésére, mivel a dokumentumküldés statisztikai jelentős csökkenést mutattak, ugyanakkor a költségek változatlanok maradtak.

Mit jelentett a megújulás?

- A szolgáltatás „helyes méretének” kialakítása.
- A hatékonyság javítása a működési költségek egyidejű csökkentésével.
- Új szolgáltatások bevezetése.

A *Hat szigma* (vállalati folyamatok javítására és fejlesztésére szolgáló eljárások összessége) és a *kaizen* (az összes alkalmazott bevonásával a folyamatos tökéletesítést támogató vállalati kultúra létrehozása) vállalatirányítási módszerekhez folyamodtak. Többek között a munkatársakat más munkaterületekre helyezték át (raktár, katalógizálás, adminisztráció) és mindig csak a fennálló igényeknek megfelelő munkatársi gárdát foglalkoztatják a dokumentumküldésben. A *kaizen* keretében

tréningeket tartottak a problémák megoldásához szükséges tudás és tapasztalatok megszerzése érdekében.

A változtatások eredményei hamar látszóttak:

- az igények csökkenő tendenciája lelassult,
- a gyors dokumentumküldés az átlagos 3 óra helyett 1 órára csökkent, a szokásos szolgáltatás időtartama az átlagos 45 óra helyett 30 órára redukálódott,
- a nap végén még folyamatban levő munkák száma háromszor kevesebb lett,
- a munkatársak munkához való viszonyulása javult.

A BL vezetőségét annyira meglepték a látványos eredmények, hogy a teljes szervezetre kiterjesztették a módszereket.

Új szolgáltatások

A következő új szolgáltatásokat vezették be a dokumentumküldéshez kapcsolódóan:

Egyesült Királyság kutatási tartaléka (UK Research Reserve = UKRR)

A BL átvállalja a programhoz csatlakozó felsőoktatási intézmények (eddig 15 jelentkezett) kevésbé használt nyomtatott folyóiratainak tárolását és megőrzését a dokumentumküldés műszaki infrastruktúrájának javítása és bővítése mellett. Az intézmények régi állományuk alacsony kihasználtságú részét átadják a BL-nek. Így náluk más célokra használható hely szabadul fel, a BL pedig megőrzi és szolgáltatja a régi, ritkán használt folyóiratokat. Minden folyóiratszámból a teljes felsőoktatási közösség még két példányt tart meg valamely csatlakozó tagnál.

A közös tárolási program 5 év alatt valósul meg 10 millió font támogatással, 100 polckilométer hely

szabadul fel és a becsült megtakarítás 29 millió font.

Disszertációk online szolgáltatása (Electronic Theses On-Line Service = EThOS)

Szintén a felsőoktatási intézményekkel közös program a brit disszertációk online szolgáltatása. A brit disszertációk mikrofilmalapú használata egyre csökkent, mert a mikrofilmek előállítás és megtekintése bonyolultabbá vált.

Az EThOS célja:

- az Egyesült Királyságban lévő disszertációk egy helyen szolgáltatása,
- a disszertációk nyomtatott formátumáról elektronikusra való átállítás,
- a meglévő disszertációk digitalizálása,
- az Egyesült Királyságbeli kutatás bemutatása.

Az EThOS az elektronikus disszertációkat rendszeresen leartja az intézményi repozitóriumokból. A papíralapú disszertációkat kérésre digitalizálják, melyért a tulajdonos intézmény fizet, és aztán a dokumentum nyílt hozzáférésű lesz, vagyis a felhasználók ingyenesen letölthetik. Az így modernizált szolgáltatás iránt tízszeresére nőtt az érdeklődés a mikrofilm-szolgáltatáshoz képest. Az EThOS teljesen önfenntartóvá vált a költségek szempontjából. A kezdetektől számítva 80 000 felhasználót szolgált ki 25 000 digitalizált disszertációval.

Tömeges digitalizálás

Több mindenre összpontosít a tömeges digitalizálási tevékenység:

- kiszolgálja az EThOS igényeit,
- megalapozza az újságprogramot, vagyis 100 millió 19. századi újságoldal digitalizálását,
- működteti a kiadói digitalizáló szolgáltatásokat (*Publisher Digitisation Service = PDS*), amelyben a BL – gyűjteményét felhasználva – a kiadók régi folyóiratszámait digitalizálja, hogy a kiadók azo-

kat honlapjukon a kurrens számokkal együtt szolgáltatassák.

Egyéb szolgáltatások:

- a felsőoktatási intézmények számára tananyag-digitalizálás és a hallgatók számára elektronikus letöltés lehetővé tétele;
- teljes könyvtári szolgáltatás (Total Library Service): ha valamely anyag nincs meg a BL gyűjteményében, azt harmadik féltől megkeresik, átveszik és szolgáltatják;
- folyóiratadatok szolgáltatása cikk- és folyóiratcím szinten.

A BL teljesen új üzleti modelleket hozott létre, mert felismerte a változtatások szükségszerűségét. A további nagyra törő terveket jelenleg alakítják ki és beágyazzák a szervezet stratégiai programjába az alábbi fő irányvonalak mentén:

- a munkafolyamatok átalakításának szükségessége,
- az e-tartalmon, a papíralapú tartalom és a nyílt hozzáférésen alapuló dokumentumszolgáltatás közötti egyensúly létrehozása,
- több együttműködés és integráció más intézményekkel, kölcsönös egyezmények révén,
- a licencszerződések felülvizsgálata és a szerzői jogi szabályozások megváltoztatása melletti lobbizás,
- a szerzői jog által már nem védett dokumentumok tömeges digitalizálásában rejlő lehetőségek nagyobb mérvű kihasználása,
- a közös tárolási modellek kiaknázása és partneri viszony a felsőoktatási intézményekkel,
- az ügyfelek szükségleteinek középpontba helyezése és könyvtárból információszolgáltatóvá válás.

/APPLEYARD, Andy: *British Library document supply – a fork in the road. = Interlending & Document Supply*, 38. köt. 1. sz. 2010. p.12–16./

(*Burmeister Erzsébet*)

A könyvtárközi kölcsönzés használatának és nem használatának jellemzése

Bár ma már rengeteg információ érhető el az interneten is, mégis sok értékes anyag csak könyvtárközi kölcsönzéssel szerezhető be.

A dokumentumszolgáltatás világméretben, különösen a természet- és műszaki tudományok területén a 2000-es évek eleje óta csökkenő tendenci-

át mutat. De a könyvek kölcsönzésére sok felsőoktatási könyvtárban, különösen az Egyesült Államokban még mindig igen nagy az igény, 3700 amerikai tudományos fokozat megszerzését lehetővé tevő intézményben 1998 és 2004 között a könyvek kölcsönzése 26%-kal nőtt. Ennek számos oka van:

- a kiadott könyvek száma folyamatosan nő,
- a könyvek felfedezését segítő Amazon, Google Books szolgáltatásokat egyre többen ismerik és használják,
- felhasználóbarát könyvkérő programok terjednek, például a Borrow Direct,
- kevesebb könyvet tudnak a könyvtárak beszerezni a folyóiratok árának emelkedése miatt,
- több a kutatás új, integrált tudományokban, például: kriminológia, gerontológia, információs rendszerek.

Az 1970-es évek óta számos kutatást végeztek a könyvtárközi kölcsönzéssel kapcsolatban: az elektronikus folyóiratok elterjedésének hatásai, a felhasználói elégedettség, a könyvtárközi kölcsönzés és a gyűjteményfejlesztés kapcsolatai. De nem történt átfogó kutatás arra vonatkozóan, hogy milyen különbségek vannak a könyvtárközi kölcsönzést használók és nem használók között.

A felmérések amerikai és izraeli felsőoktatási és kutató intézményekben azt mutatják, hogy aránylag alacsony a könyvtárközi kölcsönzést igénybe vevők száma. Az izraeli *Haifai Egyetemen* például a doktoranduszoknak csak a 23%-a használ könyvtárközi kölcsönzést.

Hagyományosan a könyvtárközi kölcsönzés használatát négy tényezőhöz kapcsolják:

- a helyi könyvtári állomány nagysága,
- a potenciális felhasználók mennyire érzik kényelmetlennek a könyvtárközi kölcsönzést,
- tudnak-e a létezéséről,
- van-e rá pénzügyi keret.

A könyvtárközi kölcsönzést használók és nem használók közötti különbségek

Az említett izraeli tanulmány azzal foglalkozik, hogy a kutatók egy része miért nem használja a könyvtárközi kölcsönzést, bár az anyagok számukra relevánsak lennének, és saját könyvtárukban nem férnek hozzá. Annak a kutatása is cél, hogy milyen különbségek vannak a használók és a nem használók információkeresési szokásaiban. Ebből a célból 330 kérdőívet küldtek ki a Haifa Egyete-

men (37% válaszolt) ahol a természettudományok, társadalomtudományok a fő szakterületek és 1090 kérdőívet (18% válaszolt) a Technionban, amelyre a természet-, a műszaki és az orvostudományok művelése a jellemző.

Az első kérdés arra kereste a választ, hogy mi a különbség a használók és a nem használók között az alábbi szempontokat tekintve:

- könyvtárhasználat gyakorisága,
- információkeresés stílusa,
- demográfia,
- akadémiai irány, életkor, fokozat, termékenységi szint és tudományág.

A második kérdés a kutatók könyvtárközi kölcsönzéssel való elégedettsége és a következők közötti kapcsolatok:

- szekunder információforrások tanulmányozásának haszna,
- tájékoztató könyvtáros segítségnyújtása,
- indikatív és informatív címek választása, melyek a kutatási projekt módszerét és eredményeit felvázolják,
- a kért példányok időbeli szállítása.

Megállapítások

A cikk legfontosabb megállapítása, hogy a könyvtárközi kölcsönzést használók az idősebbek, produktívak és a bölcsészettudományok művelőiből kerülnek ki, akik gyakran kölcsönöznek könyveket és információkeresési szokásaikra az alaposság jellemző. A könyvtárközi kölcsönzést nem használók kevésbé idősek, kevésbé produktívak, a természettudományi karokon tevékenykednek és információkeresésük felületesebb.

További következtetés, hogy a nem használat fő oka az, hogy jobban szeretnek szabadon hozzáférhető dokumentumokat letölteni. A nem használat néhány fő oka a kérdőívek alapján, hogy a megkérdezett:

- számos izraeli egyetemi könyvtár anyagához hozzáfér,
- szívesebben megveszi a számára nagyon fontos könyveket,
- rengeteg folyóirathoz van hozzáférése, mert felsőoktatási intézményben tagja,
- számára kényelmetlen a könyvtárközi kölcsönzés,
- számára hosszadalmas a könyvtárközi kölcsönzés,
- számára költséges a könyvtárközi kölcsönzés,

- fórumok tagja és egymásnak elküldik az anyagokat,
- nem ismeri a könyvtárközi kölcsönzést.

A kutatási eredmények gyakorlati haszna az, hogy a használók és a nem használók jellemzőinek ismerete megkönnyíti a könyvtáros számára, hogy felismerje a nem használókat és bátorítsa őket a könyvtárközi kölcsönzés használatára.

Egy jövőbeni kutatás témája lehet annak kiderítése, hogy a könyvtárközi kölcsönzést használók kutatásainak minősége magasabb vagy alacsonyabb, mint a könyvtárközi kölcsönzést nem használóké.

Következtetés

A könyvtárközi kölcsönzés jövőjére vonatkozó pesszimista jóslatok ellenére ez a tanulmány azt mutatja, hogy a könyvek kölcsönzése a közeljövőben nem valószínű, hogy eltűnik, sőt emelkedni fog, különösen a bölcsészettudományok területén, a ritka és a nem angol nyelvű dokumentumok között. A könyvtárközi kölcsönzés a komoly tudományos kutatáshoz alapvető szolgáltatásként járul hozzá.

/PORAT, Lynne – FINE, Sara: Factors and characteristics of interlibrary loan use and non-use. = Interlending and Document Supply, 37. köt. 1. sz. 2009. p. 20–27./

(Burmeister Erzsébet)

A könyvtárközi kölcsönzés kisebb változtatásainak is már nagy a hozadéka

Az elmúlt öt évben az USA-beli *Hilbert College* (New York állam, Hamburg város) *McGrath Könyvtár*ában számos újítást vezettek be, köztük néhány a könyvtárközi kölcsönzést érintette. A könyvtárnak megközelítőleg 30 000 kötetes az állománya, a munkatársak száma hat és az intézménynek 1000 hallgatója van. 2004-től a bevezetett módosítások következtében a könyvtárközi kölcsönzések száma fokozatosan nőtt, a teljesített kérések száma 30%-kal emelkedett. Melyek voltak az apró, de eredményes változtatások?

- Áttérés a papírról az online megrendelésre: a könyvtár weboldalán online kérelmet hoztak létre. Ezzel lényegesen csökkent a kérések teljesítésének ideje, hiszen az oktatók vagy a hallgatók bárhol kitölthetik a kérelmet, azt azonnal elküldhetik, nem kell bejőniük a könyvtárba. A kitöltés után azonnal e-mail megy a könyvtárközi kölcsönző könyvtárosnak a szükséges információkkal.
- Az EBSCO adatbázisokban való keresés közben könyvtárközi kérés küldése: ha az oktatók vagy a hallgatók olyan cikket találnak az EBSCO adatbázisokban, amelyhez nem tartozik teljes szöveg, akkor lehetőségük van a cikket azonnal könyvtárközi kölcsönzéssel megkérni. A megtalált, teljes szöveg nélküli cikkeknel rögtön megjelenik számukra egy a könyvtárközi kölcsönzést felkínáló üzenet, és ha ezt a lehetőséget választják, akkor máris megjelenik a kívánt cikk adatai-

val kitöltött kérelmet, amelyhez már csak saját személyes adataikat kell hozzáadniuk. Ugyanúgy, mint a webes űrlapnál, ennek elküldése után is automatikus e-mail megy a könyvtárosnak, aki máris elkezdheti végrehajtani a tennivalókat. Ez a módosítás kényelmes a felhasználók számára, mert a keresésüket nem kell megszakítani és átlépni a könyvtári kérelem űrlapra, hanem folytathatják az adatbázisban a kutakodást.

- Cikkek e-mailes küldése az olvasóknak: a teljesített kéréseket közvetlenül elküldik az oktatóknak vagy hallgatóknak e-mail mellékleteként, a cikkeket nem nyomtatják ki, nem kell értük bemenni a könyvtárba. Ez az újítás is jelentősen csökkentette a teljesítések idejét és még kényelmesebbé tette a szolgáltatást, mert a felhasználóknak nem kell ezért felkeresniük a könyvtárat.
- A könyvtárközi kölcsönzés marketingje a könyvtári oktató kurzusok keretében: a könyvtári órák anyagában mindig szerepel a könyvtárközi kölcsönzés mint könyvtári szolgáltatás bemutatása. Az órákon a hallgatók a keresések gyakorlása közben azonnal meg is rendelhetik a számukra fontos dokumentumokat; így a gyakorlatban is tapasztalják a könyvtárközi kölcsönzés lehetőségeit és hasznát.
- Marketing az információkeresési órákon: a könyvtári órákon kívül az információkeresést oktató kurzuson is tanítják a könyvtárközi kölcsönzést (ezekre évente egyszer kerül sor).

- Marketing a tájékoztató pultnál: a tájékoztató könyvtárosok is bátorítják a hallgatókat a könyvtárközi kölcsönzés igénybe vételére. Ha egy hallgató kérdéssel fordul a tájékoztató könyvtárhoz, az mindig felhívja a figyelmét a könyvtárközi kölcsönzés lehetőségére is. Ha a hallgató éppen akkor nem is szeretné igénybe venni a szolgáltatást, a későbbiekben erre még sor kerülhet.
- Marketing az oktatóknak a kapcsolattartó program keretében. A könyvtárközi kölcsönzések száma 2004-től igen megemelkedett az oktatók körében. A könyvtár akkor vezette be a kapcsolattartó programját, vagyis az oktatókkal szorosabb kapcsolatot épített ki. Ez azt jelenti, hogy gyakrabban kommunikálnak az oktatókkal speciális szakterületük szakirodalmi igényeinek kielégítését segítve.

Az apró módosítások jelentős emelkedést eredményeztek a könyvtárközi kölcsönzésben. Az állománygyarapításra szánt összegek csökkenése együtt jár a könyvtárak közötti kölcsönzési forgalom növekedésével. Az olvasók többször használnak adatbázisokat is, ami szintén hozzájárul a kérések emelkedéséhez. Az McGrath Könyvtár könyvtárközi kölcsönző könyvtárosai továbbra is keresik a még jobb szolgáltatási módszereket, eljárásokat.

/CURRY, Elisabeth: Little Changes to Interlibrary Loan Services make a Big Difference at a Small Academic Library. = Journal of Interlibrary Loan, Document Delivery & Electronic Reserve, 19. köt. 4. sz. 2009. p. 287–290./

(Burmeister Erzsébet)

Elektronikus könyvek szerzeményezése Olaszországban

Az elektronikus könyvek szerzeményezése egyfelől újfajta szakmai megközelítést, másfelől az üzleti és a hozzáférési modellek szerteágazó arzenáljának ismeretét igényli az állománygyarapítással foglalkozó könyvtárosoktól. A tanulmány egy 2008-as vizsgálatra alapozva igyekszik kideríteni, hogy milyen módszereket követnek az egyes könyvtárak az e-bookok beszerzésekor, illetve mennyiben vált a tudatos gyűjteményfejlesztési stratégia részévé az elektronikus könyvek szerzeményezése.

Egy-egy termék megvásárlásakor a gyarapító könyvtárost számos tényező befolyásol(hat)ja. Ilyenek például a használati statisztikák, az oktatói és hallgatói visszajelzések, az árinformációk, a termék licencére vonatkozó információk, a digitális jogkezelési problémák, avagy a beszerzés konzorciális jellegéből fakadó szempontok. A nyomtatott művek fizikai birtokbavétele és örökös tulajdonlása mellett a könyvtárakban szakadatlanul nő az e-források jelenléte, amely többnyire csupán virtuális tulajdonlást és ideiglenes birtokbavételt jelent. A licenclapú, előfizetéses beszerzés következtében a könyvtárosnak az olvasói igények, a minőségi szempontok és az árak mellett azt is mérlegelnie kell, hogy a későbbiek során milyen szintű ellenőrzést gyakorol a könyvtár az előfizetett termék fölött. A szerző által vizsgált olaszországi felsőoktatási könyvtárakban az elektronikus könyvek beszerzése jobbra még nem tudatos szerzeményezési stratégia alapján történik. Ezt bizonyít-

ja, hogy az utóbbi években Itália-szerte csekély számú, az e-könyvgyűjtemények könyvtári állományba való integrálását célzó projekt valósult meg.

Az egyik legnagyobb olaszországi műszaki egyetem, a 15 ezres hallgatói létszámú *Politecnico di Milano* könyvtára 2004-ben egy kísérleti projekt keretében elemezte a kiadói ajánlatokat, értékelve azok tartalmát, valamint a kínált hozzáférési és licencmodelleket. Úgy találták, hogy igen sok az átfedés a csomagokban kínált elektronikus könyvek címei között. Problémát jelentett az egyes kiadók által rendelkezésre bocsátott különböző formátumú és adattípusú használati statisztikák kiértékelése és egymással történő összevetése is. Miután az értékelés elsősorban ezekre az adatokra épített volna, a felhasználói statisztikák szabványosítása területén lévő hiányosságok jelentős mértékben szűkítették az elemzés lehetőségeit. A könyvtárvezető 2007-ben adott projektértékeléséből látható, hogy ekkor még nem volt kidolgozva az elektronikus könyvek beszerzésére vonatkozó külön szerzeményezési stratégia.

Egy másik projekt a *Bolognai Egyetem* mérnöki karán valósult meg 2002–2004 között. A cél egy elektronikus referenzgyűjtemény létrehozása volt. A beszerzendő e-könyveket maguk a könyvtárosok választották ki az oktatói gárda közreműködése nélkül. A projekt második fázisa az olvasók széles körű tájékoztatását célozta meg: speciális, olvasó-

ra szabott elektronikus irodalmi jegyzéket állítottak össze; a könyvtár weboldalán és szórólapon, valamint e-mailekben tájékoztattak a beszerzett e-bookokról. A jelentős számú hallgatói és oktatói gárda miatt még ezek a módszerek is kevésnek bizonyultak az információk hatékony áramoltatására, ezért a könyvtárosok úgy döntöttek, hogy katalogizálják a műveket, az egyes e-könyvcímek leírásait integrálják az online katalógusba. Ez hatásosnak bizonyult, hiszen ezt követően jelentős számú, a WebOPAC-ból indított adatbázishozzáférést regisztráltak. A használói statisztikákból kiderült, hogy egyes művek iránt nagyobb igény mutatkozik, és egyre inkább szükségessé válik egy olyan e-book-értékesítési stratégia elterjedése, amely lehetővé teszi, hogy előre összeállított csomagok és adatbázisok beszerzése helyett a könyvtáros címenként dönthessen, és maga választhassa ki egyesével az általa megvásárlásra ítélt műveket.

A piacenai egyetemen, csakúgy, mint a bolognain, a könyvtárosok maguk vezényelték le a 2005-ben indult projektet, amelynek során mindekenélőtt kiválasztották a szükséges címeket, majd a címek száma alapján (egy 30 napos próbához-záférést követően) megegyeztek egy egyéves előfizetési díjban a szolgáltatóval, amelytől megkapták az elektronikus könyvek MARC rekordjait. További kritérium volt: párhuzamos hozzáférés megoldása; teljes szöveges keresés és címböngészés; megfelelő, a szövegrészek kiemelését vagy jegyzetekkel való ellátást megengedő eszközök nyújtása az e-bookokhoz; használati statisztikák lehívásának lehetősége. A könyvtárosok jelentős energiát fektettek az e-bookgyűjtemények propagálásába: speciális linket helyeztek el az intézmény weboldalán, szórólapokat osztottak szét az olvasók között, valamint kérdőívek segítségével begyűjtötték a visszajelzéseket. Ez utóbbiak, valamint a használati statisztikák alapján tovább pontosították az előfizetendő elektronikus könyvek körét.

Az esettanulmány, amelyre alapozva a szerző felvázolja az olasz felsőoktatási könyvtárak tapasztalatait az e-book-szerzeményezés terén, egy egyetemközi konzorcium szakértőjével, egy könyvtári vezetővel, illetve a *Casalini kiadó* tulajdonosával készített interjúra és egy kérdőívre épül. A kérdőív kitöltésére az elektronikus könyvek beszerzésében érdekelt könyvtárosokat kért fel a cikkíró.

Az adatok elemzéséből kiderült, hogy az e-bookokat némi fenntartással kezelik a könyvtárosok, és bár az elektronikus források beszerzésének irányelveit a

legtöbb intézményben már lefektették, az e-folyóiratok beszerzése jóval rutinszerűbb, mint az e-könyveké. Problémát jelent, hogy az e-könyveket a terjesztők csomagokban értékesítik, s ez kikényszerítheti a gyűjteményfejlesztési politika újragondolását. Ehhez a következő kérdéseket kell megvizsgálni: a nyomtatott és elektronikus formában egyaránt elérhető gyűjteményrész megfelelő arányának meghatározása; a használói igények értékelése a költségvetési korlátok betartásával; az elektronikus tartalmak kezelésének a könyvtári munkafolyamatba (deziderálás, rendelés, katalogizálás) való integrálása, illetve az integrálási munkafolyamatot lehetővé tevő elektronikus forráskezelő rendszer (ERMS) kifejlesztése/implementálása. Az interjúkból kiderült az is, hogy általában nincsenek megelégedve az olaszországi könyvtári konzorcium tevékenységével: több aktivitást várnak el a részt vevő könyvtárak szükségleteinek, valamint a kiadói kínálatnak az elemzése, illetve a licencfeltételek tisztázása terén.

Jóllehet az e-bookok kiválasztása során a könyvtárosok igyekeznek tartalmi szempontok alapján dönteni, igencsak megköti a kezüket, hogy előre összeállított csomagok közül kell választaniuk. Ennek következménye, hogy az elektronikus könyvgyűjteményt nem tudják a saját gyűjtőkori szempontjaik alapján alakítani, és a könyvek kiválasztásakor követett irányelvek is sérülnek. Ezen irányelvek: a felhasználói igények és szükségletek értékelése-elemzése; a gyűjtemény potenciális és valóságos használatának mérlegelése; tartalmukat tekintve minőségi és naprakész művek kiválasztása.

Az esettanulmány arra is fényt derített, hogy a könyvtárosok komoly elvárásokat fogalmaznak meg az e-bookok kapcsán, ezek például: tartalomfrissítések, egyszerű hozzáférés és használhatóság, a hatékony keresést támogató platform stb. E kívánalmaknak gyakorta nem felelnek meg a szolgáltatók kínálatai. A *Casalini kiadó* *Editoria Italiana Online (EIO)* projektje a felhasználói igények elemzése alapján egy korrekt értékesítési modellt dolgozott ki, amelynek keretében minőségi olasz tudományos munkákat kínál eredeti nyelven. A modell egyfelől lehetővé teszi, hogy a könyvtáros tartalmi szempontok alapján kialakított e-könyvcsomagra fizessen elő, valamint testre szabható e-bookgyűjteményt kínál: a címek 20%-a lecserélhető, s újabb pénzügyi megszorítások életbe lépésekor kisebb csomagra lehet váltani. A kiadó igyekszik az új technológiai kihívásoknak és piaci trendeknek megfelelő szolgáltatást nyújtani, így az EIO folyamatos fejlesztés alatt áll. A kiadói megközelítés találkozik

a könyvtárosi elvárásokkal, amelyek: állandó hozzáférés a megvásárolt tartalmakhoz; a rekordok beilleszthetősége az OPAC-ba és más forrásokkal való összekapcsolhatósága; párhuzamos hozzáférés lehetősége; színvonalas tartalom. Mindennek ugyanakkor megvan a maga ára: a könyvtáraknak még a megvásárolt tartalmakért is hozzáférési díjat kell fizetniük, amely gyakran gondot okoz, hiszen nem tudják garantálni a pénzforrásokat évekre előre. A színvonalas gyűjteményfejlesztés rovására mehet továbbá az a kérdőívekből kiolvasható gyakorlat, amelynek lényege, hogy a könyvtárosokat elsődlegesen anyagi megfontolások vezérik az EIO-s előfizetések megújításakor, s kevésbé a felhasználói igények és statisztikák. A szerző szerint szükség volna mindenekelőtt egy újfajta megközelítés meghonosítására: a gyarapító könyvtárosnak tevékenyen részt kell vennie a konzorciumi szerződések előkészítésének tárgyalási szakaszában és az olyan eszközök kifejlesztésében, amelyek alkalmasak a kiadók által kínált új tartalmakat a könyvtár küldetését és a gyűjtemény érdekeit szem előtt tartva kiértékelni.

Összefoglalásképpen

A vizsgálatból kiderül, hogy az olasz könyvtárosok idegenkedése az elektronikus könyvekkel szemben a következő hiányosságokra vezethető vissza: nincs egy versenyképes e-book-értékesítési modell, hiányoznak a szabványos formátumok, valamint az e-bookok szerzeményezését segítő

elektronikus forráskezelő rendszer (ERMS). A tanulmány feltárta továbbá, hogy hiányoznak az olasz nyelvű tudományos elektronikus könyvek. Ezen a téren jelentős előrelépés a Casalini cég flexibilis gyűjteményformálást lehetővé tevő, a kisebb kiadók kiadványaival is számoló EIO projektje, amelyet elsősorban a felhasználói felületet illetően érnek kritikák. Jóllehet a könyvtárosok igyekeznek körültekintően eljárni az e-bookgyűjtemények gyarapításakor, a szűkös pénzügyi források miatt az anyagi szempont gyakorta jobban érvényesül, mint a felhasználói igények szem előtt tartása. Ezért is igyekeznek nyomást gyakorolni a kiadókra, hogy rugalmasabb értékesítési modellek keretében kínálják az elektronikus könyveket.

Az e-könyvek terjedését gátló tényező, hogy a könyvtárosok elvárásai gyakran nem találkoznak a kiadói kínálattal, amely sokszor nem elég naprakész, nem elég standardizált, nem az elvárt minőséget képviseli, és túlságosan is korlátozó jellegű digitális jogi megszorítások sújtják. Fontos szerep jut ezen a téren az olasz könyvtári konzorcium érdekérvényesítő képességének, aminek egyik előfeltétele, hogy új, a konzorciumi tagok érdekeit szem előtt tartó tárgyalási irányelvek kerüljenek kidolgozásra.

/PERRONE, Agnese: Electronic book collection development in Italy: a case study. = IFLA Journal, 35. köt. 4. sz. 2009. p. 305–312./

(Dancs Szabolcs)

Ingyenes online eszközök könyvtárosoknak

A könyvtáros sokrétű tevékenységet folytat, s ezek javát ma már a virtuális térben. A cikk szerzője olyan ingyenesen elérhető alkalmazásokat gyűjtött egybe, amelyek könnyebbé teszik az „online könyvtáros” munkáját.

Elsőként lássuk, mi lehet segítségünkre a *könyvtári munka tervezési szakaszában*. Előnyös mindenekelőtt egy olyan alkalmazás beszerzése, amely lehetővé teszi számunkra a jegyzetelést. Ilyen a *StickIt*, amellyel rövid megjegyzéseinket közvetlenül az asztali háttérre (más szóval: tapétára) „ragaszthatjuk”, s azok mindaddig láthatóvá válnak – még a számítógép újraindítása után is –, amíg be nem csukjuk őket. Megjegyzéseinket a *StickIt* révén másokkal is megoszthatjuk. A program letölthe-

tő innen: <http://www.singerscreations.com/StickIt.asp>

Egy másik hasonló célú ingyenes alkalmazás az *Evernote* (www.evernote.com). A különböző platformokkal (Windows, Macintosh, iPhone stb.) együttműködő program segítségével nem csupán megjegyzéseket tehetünk: elmenthetünk weboldalakát (vagy azok egyes részeit), hangemlékeztetőket vehetünk fel, egy helyre gyűjthetjük a telefonunkkal lefotózott névjegykártyák képfájljait, elraktározhatjuk repülőgépünk e-ticket számát stb. Ezeket azután a könnyebb visszakereshetőség végett címkékkel (tagekkel) láthatjuk el, vagy hagyatkozhatunk az alkalmazás által kínált automatikus indexelésre.

A *Remember The Milk* (www.rememberthemilk.com) a *Gmail*-el és a *Google Naptárral* együttműködve az elvégzendő feladataink számon tartásában, illetve napi teendőink megszervezésében lehet leginkább a segítségünkre. Feladataink listáját másokkal is megoszthatjuk.

Említést érdemel a *Jott* szoftver (<http://jott.com>), amely ma már sajnos nem tölthető le ingyenesen. A program a hangzó beszédet számítógépen megjeleníthető szöveges formába konvertálja. Ezzel gépbe diktálhatunk feljegyzéseket, emlékeztetőket, e-mailek szövegeit, naptári bejegyzéseket stb.

Hasznunkra lehet még az ingyenes *Wordle* (www.wordle.net) program, amelynek segítségével begépelte szövegünkéből kedvünkre formázott címkéfelhőket hozhatunk létre.

A könyvtári munka során ugyancsak hasznos lehet, ha alkalmas eszköz áll rendelkezésünkre ahhoz, hogy a weben található tartalmakat *elmentsük* és *megosszuk* másokkal. Nyilakkal, feliratokkal, kiemelésekkel ellátott képernyőképeket vagy a képernyőn folyó tevékenységünket rögzítő ötperces videót készíthetünk a Windows és a Macintosh platformon is működő *Jing* alkalmazással (www.jingproject.com).

A *Clipmarks* (<http://clipmarks.com>) ingyenes alkalmazás a Firefox vagy az Explorer böngészők részeként működik, és a weboldalról kiválasztott szöveg, kép vagy akár videó elmentésére, illetve e-mailbe, webnaplóba ágyazására szolgál.

A *Blinklist* (www.blinklist.com) a könyvjelzőink kezelésében áll rendelkezésünkre. A böngészőben megjelenő „blink!” feliratú gomb segítségével könnyedén elmenthetünk weboldalakat. Ezek azután kereshetővé válnak, és a felhasználó gépén is tárolhatjuk őket. A linkek elmentése és megosztása egyetlen kattintással történik.

A szerző figyelmünkbe ajánlja még a *Slideshare* (www.slideshare.net) programot, amellyel megoszthatjuk prezentációs anyagainkat, dokumentumainkat. A feltöltött tartalmakat az internetes keresőmotorok indexálják, gondoskodva visszakereshetőségükről.

Míg az előbb ajánlott szoftver a prezentációs anyagaink megosztását szolgálja, addig a *Prezi* nevű webalkalmazással (<http://prezi.com>) *elkészíthetjük magukat a prezentációkat*. Ha nem előadást, hanem *felmérést szeretnénk készíteni*, ezt megte-

hetjük a *SurveyMonkey* (www.surveymonkey.com) és a *PollDaddy* (<http://polldaddy.com>) programokkal. A felmérés megtervezése mellett ezekkel az alkalmazásokkal összegyűjthetjük és kiértékelhetjük a válaszokat.

Ha a könyvtáros *saját weboldal készítésére* szánja el magát, kínálózó megoldás a különböző widgetekkel és modulokkal, ún. *Flake*-ekkel (lemezekkel/rétegekkel) építkező *Pageflakes* (www.pageflakes.com) használata. A kész weboldalt megoszthatjuk barátainkkal vagy akár a világháló széles közönségével. Ilyen, „előre gyártott lemezekből” felépült oldalakkal gyakorta találkozhatunk például a Facebook és a MySpace közösségi portálokon is. Egy másik alkalmazás, a *Ning* (www.ning.com) segítségével *saját közösségi hálónkat is felépíthetjük*.

Ha kreatitásunkat *képfájlok szerkesztésében* szeretnénk kiélni, és nem rendelkezünk Photoshoppal, néhány ingyenes megoldás akkor is kínálkozik. Míg a FotoFlexer alkalmazás (<http://fotoflexer.com>) csak fényképek szerkesztésére szolgál, addig a *Pixlr* (www.pixlr.com), a *Splashup* (www.splashup.com) és a *Picnik* (www.picnik.com) programok teljes funkciók képszerkesztők, amelyek új képfájlok kreálására is alkalmasak. Ezek a szoftverek tehát a képszerkesztőknél megszokott minden funkcióval rendelkeznek a képrészlet kivágásától kezdve a szövegek vagy egyéb elemek (alakzatok, vonalak) hozzáadásáig. Az ingyenes regisztrációt igénylő *Aviary* (<http://aviary.com>) mindezen túl hangzó anyagok szerkesztését is lehetővé teszi.

Ha a könyvtárosok *az olvasók vagy a munkatársaik oktatását* tűzik ki célul, akad néhány ingyenesen letölthető szoftver, amely a segítségünkre lehet. A *Moodle* (<http://moodle.org>) egy nyílt forráskódú kurzuskezelő rendszer (*course management system*), amely dinamikus tartalmú, oktatási célú online weboldalak szerkesztésére szolgál. A webszerveren futó rendszer a tanfolyamok tartalmának szabványos, SCORM-csomag formájában történő közvetítésén túl a dolgozatok és tesztek kiértékelését is lehetővé teszi, valamint a tanulás élményét különböző modulok (fórum, wiki, adatbázis stb.) révén teszi még teljesebbé.

A *WiZiQ* programmal (www.wiziq.com) egy virtuális osztályteremben érezhetjük magunkat, ahol különböző típusú dokumentumokat (szöveges dokumentumok, képek, videók) oszthatunk meg egymással. A Moodle-lel együttműködtetve webkonferencia

tartására is alkalmas. Két további ingyenes szoftver, amely szóba jöhet, ha webkonferenciáról van szó: a *Dimdim* (www.dimdim.com) és a *Yugma* (<https://www.yugma.com>).

A felsorolt alkalmazások jól mutatják, hogy egy alacsony költségvetésű intézmény is sok mindent oldhat meg, ha rendelkezik jó internetkapcsolattal,

a munkatársakban pedig megvan a megfelelő nyitottság az új dolgok iránt.

/MAXYMUJ, John: Online tools. = The Bottom Line: Managing Library Finances, 22. köt. 4. sz. 2009. p. 135–138./

(Dancs Szabolcs)

Digitális könyvtárak együttműködési modelljei

Szakirodalmi áttekintés

Bevezetés

A digitális könyvtárak a rend szigetei a web kaotikus információs tengerében. Fontos szerepük van a civilizáció előrehaladásában, mert igen hatékonyan, földrajzi, nyelvi és kulturális határokat átlépve, bármikor és bárhol hozzáférhetővé teszik a felhalmozott emberi tudást. Az még azonban kevés, hogy rájuk bukkanunk egy keresőgép vagy egy tematikus webkatalógus segítségével: ezután be kell lépni valamelyikbe, ott keresgélni vagy böngészni, majd továbbmenni a következőre, ha nem találtuk meg, amire szükségünk van. Felhasználói szempontból fontos lenne tehát, hogy ezek a digitális gyűjtemények „együttműködjenek”, vagyis hogy a különböző számítógépek, hálózatok, operációs rendszerek és alkalmazások „értsék egymást”. Ahogy *Suleman* [1] fogalmazott a diszszertációjában: az „interoperabilitás” azt jelenti, hogy egy e-könyvtár együttműködik más hasonló gyűjteményekkel annak érdekében, hogy minél jobb szolgáltatást tudjon nyújtani felhasználóinak. *Paepcke* és szerzőtársai [2] pedig már 1998-ban arra hívták fel a figyelmet, hogy az együttműködés a digitális gyűjtemények fejlesztésének, illetve továbbfejlesztésének kulcskérdése. Ezt erősítette meg évekkel később *Warren* és *Alsmeyer* [3] is, akik kimutatták, hogy a szakterület egyik legtöbb vitatott témája az egyes szolgáltatások közti kooperáció.

Arms 2002-es cikkében [4] az együttműködés három szintjét vagy modelljét különböztette meg: szövetkezési (*federated*), aratásos (*harvested*) és összegyűjtési (*gathering*). Természetesen a valóságban ezek kombinációi is előfordulnak, vagyis egy digitális könyvtár többféle módon is kapcsolódhat más gyűjteményekhez. Az interoperabilitásnak ezen kívül műszaki, tartalmi és

szervezési szintjei is vannak: az elsőhöz tartoznak például a formátumok, a protokollok, a biztonsági megoldások; a másodikhoz a metaadatok és azok jelentésének értelmezése; a harmadikhoz pedig olyan kérdések, mint például a hozzáférés szabályozása, a felhasználók azonosítása, a fizetés módja. A jelen tanulmányban az iráni szerzők az első kérdéssel, vagyis a digitális könyvtárak együttműködésének technikai megoldásaival foglalkoznak, áttekintve a témában 2008 februárjáig megjelent szakirodalmat és bemutatva a különböző modelleket, illetve az azokat alkalmazó projekteket. A feldolgozandó publikációkat a *digital library*, az *interoperability*, a *meta-searching* és a *cross-searching* kifejezések kombinálásával keresték ki különböző nagy szakirodalmi adatbázisokból.

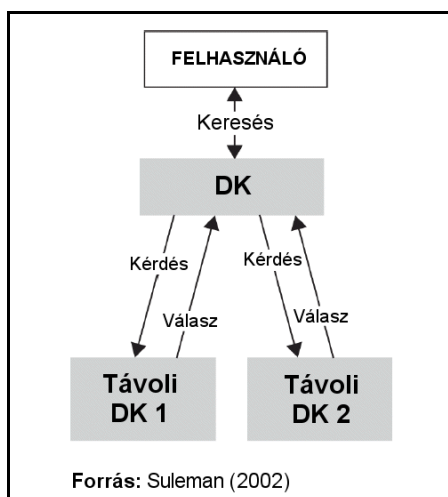
Együttműködési modellek

Több tényező befolyásolja, hogy a különféle technológiákra épített digitális archívumok milyen megoldást választanak az együttműködéshez. Függ például attól, hogy mikor fejlesztették ki a rendszert, mennyi pénz volt erre a célra, és hogy milyen hatékonyságot akartak elérni a fejlesztők. Léteznek általánosan elterjedt egyszerű szabványok, amelyek esetleg kevésbé hatékonyak az adott célra, és vannak újszerű megoldások is (pl. Java-alapú alkalmazások), amelyek célszerűbbek ugyan, de lehet, hogy kevesebben tudják őket használni. Néha elegendő egy gyors, egyszerű, szélesebb kört lefedő, de felületes megoldás, és természetesen vannak olyan digitális gyűjtemények is, amelyeknél megéri egy költséges, de professzionálisabb eredményt adó technológiát választani. Az említett három modell közül a szövetkezési megoldás jelenti a legnagyobb terhet a partnereknek, míg az összegyűjtési módszernél

gyakorlatilag nincs is feladatuk a tartalomgazdáknak – igaz, ez adja a leggyengébb végeredményt. Az aratásos megoldás a kettő között van a megvalósítás nehézsége és a hatékonyság szempontjából. Az sem ritka – ahogy a következő példák mutatják –, hogy egyes projekteknél több modellt is alkalmaznak, különböző partnerek felé vagy eltérő célokra: például Z39.50 és OAI alapon egyaránt folytatnak adatcserét.

Szövetkezési modell

Ez az együttműködések leghagyományosabb formája, amikor is a partnerek megegyeznek, hogy közös megoldásokat és szabványokat fognak alkalmazni. Ilyen szövetkezés esetén a keresés menete az 1. ábrán látszik: a digitális könyvtár (DK) elküldi a felhasználó keresőkérdését több távoli repozitóriumnak is (DK 1, DK 2), majd ezek válaszait begyűjti, összefésüli és továbbítja a kérdezőnek. A föderált modellnél a leggyakrabban használt szabvány a *Linked Systems Project* keretében kifejlesztett Z39.50 protokoll, melynek első változata még 1984-ben jelent meg, és amelyet régóta használnak a könyvtárak bibliográfiai adatok lekérdezésére. A Z39.50 kommunikáció kliens/szerver elven működik: a kliens program tartja a kapcsolatot a felhasználókkal, illetve a szerverrel, a szabványban meghatározott keresőnyelv és kommunikációs protokoll segítségével. Ezt a modellt köztes program (*middleware*) segítségével is meg lehet valósítani – ilyenkor a közvetítő szoftverréteg kapcsolódik a szerverekhez és a felhasználó ennek köszönhetően hozzáfér más gyűjtemények tartalmához is.



1. ábra A szövetkezési modell folyamata

A föderált keresés néhány konkrét megvalósítása:

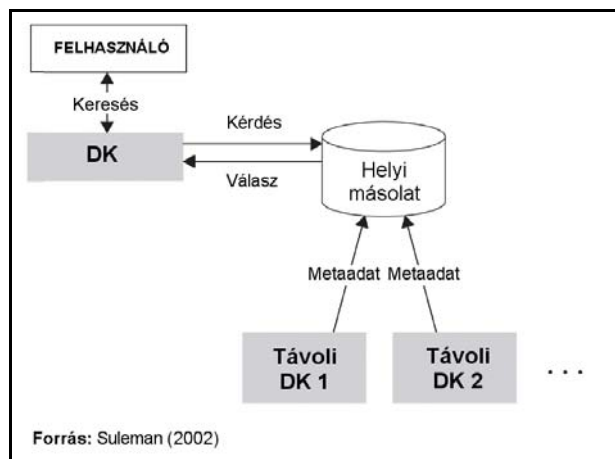
- A *Columbia University* PERSIVAL nevű projektje az *SDLIP* (*Simple Digital Library Interoperability Protocol*) és a *STARTS* (*STANford protocol proposal for Internet ReTrieval and Search*) ajánlások kombinálásával valósítja meg különféle dokumentumgyűjtemények közös lekérdezését. Az egyesített protokoll: a *SDARTS* meglehetősen egyszerű, de mégis kifinomult metakeresőket lehet készíteni erre alapozva.
- A *Greenstone* digitális gyűjteménykezelő is támogatja a Z39.50 szabványt. Erre a célra egy új szervert fejlesztettek ki a rendszer programozói, amely a nyílt forráskódú *YAZ* programkönyvtárat használja előtérnek (*front end*), vagyis ez fogadja a Z39.50-es kliensektől érkező kéréseket, majd lefordítja őket a *Greenstone* saját protokolljára. A *Greenstone* válaszait pedig szintén a *YAZ*-on keresztül alakítják vissza Z39.50 nyelvű kifejezéseké és küldik vissza a kliensnek.
- A *Colorado State Library* által 1998-ban indított digitalizálási programban szintén a Z39.50-es protokollt használják a gyűjtemények közös lekérdezésére. Továbbá egy *DC Builder* alkalmazást is létrehozta azért, hogy a különféle rendszerű és formátumú metaadatokat egységesen Dublin Core-ra konvertálják.
- Illinois állam 12 könyvtára a *ZMARKO* projektnek köszönhetően teszi elérhetővé a bibliográfiai rekordjait. Egy olyan adatszolgáltatást fejlesztettek ki, amely a Z39.50 szerverrel lekérdezhető MARC rekordokat OAI-PMH protokoll szerint is tudja szolgáltatni (vagyis egy OAI-PMH/Z39.50 gateway). Így ez a rendszer a *federated* és a *harvested* modellt egyaránt támogatja.
- A *Mountain West Digital Library* program Utah állam egyetemi könyvtárainak együttműködése 2002 óta. A metaadatok cseréje itt az *Open Archives Initiative* protokoll alapján zajlik, de 2003-ban egy *ZContent* nevű, nyílt forráskódú kiegészítő modult is kifejlesztettek Perl nyelven, ami lefordítja a Z39.50-es kérdéseket, vagyis ez is egy kéttípusú rendszer.

Aratásos modell

Sokszereplős szövetségeket nehéz létrehozni és fenntartani, ezért újabban már inkább lazább együttműködési formákat választanak a digitális könyvtárak. Ezekben az esetekben a résztvevőknek elég csak néhány közös alapfunkcióban megegyezni és nem kell egy egész sor részletkérdésben megállapodniuk. Az *Open Archives Initiative* is ezt az elvet valósítja meg a metaadatok aratásához. Ennél a modellnél a digitális könyvtárak egy-

szerű csereformátumban teszik elérhetővé az adataikat, melyeket azután különféle metaadat-szolgáltatók begyűjtenek egy szabványos protokollt használva. A metaadatok aratásának ötlete már az 1990-es évek elején felmerült, de akkor még nem terjedt el. 1998-ban azután egy *Universal Preprint Server* elnevezésű prototípusban valószínűsítették meg. Ez az alkalmazás a különféle preprintarchívumok közös lekérdezését célozta, és ebből a kísérletből alakult azután ki az OAI, amely egyszerű és olcsó megoldás azoknak a digitális könyvtáraknak, amelyek meg akarják osztani a metaadataikat másokkal.

A 2. ábrán látható ez a folyamat: az együttműködő könyvtárak felállítanak egy szervert, ahová mindegyik résztvevő egy közös HTTP kapcsolaton át elküldi a metaadatait. A felhasználók ezután ezekben a felmásolt metaadatokban tudnak keresni, vagyis egyben látják a különböző gyűjtemények anyagát. Olyan megoldás is elképzelhető, hogy kétirányú kapcsolat van a *harvester* szerver és a partnerkönyvtárak között: ez esetben mindegyik partner letükrözi magának a szerveren levő integrált adatbázist. Az is lehetséges – bár elég komplikált – megvalósítás, hogy a résztvevők szintén kétirányú kapcsolatokon át, közös szerver fenntartása nélkül, közvetlenül egymásnak küldik át a metaadataikat.



2. ábra Az aratásos modell folyamata

Néhány példa az aratásos metaadat-gyűjtésre:

- A *SIGAL* (*Système d'Information Géographique et Agent Logiciel*) projektnél a *C-FOAL* (*Conception par Frameworks Orientés-Agents Logiciels*) módszer segítségével hoztak létre együttműködési környezetet georeferált digitális gyűjteményekhez. A fejlesztésnél három keretrendszerből (szerver,

kliens és logikai forrás) alakították ki a rendszer architektúráját.

- Az *NDLTD* (*Networked Digital Library of Theses and Dissertations*) elnevezésű együttműködésnek 13 tagja van a világ különböző pontjain. A közös lekérdezéshez OAI protokollt használnak, az alkalmazott metaadatszabvány pedig az *ETDMS* (*Electronic Theses and Dissertation Metadata Set*). A *National Science Digital Library* szintén az OAI-t választotta ahhoz, hogy egy központi adatbázisba gyűjtse a metaadatokat.
- A *University of Michigan* OAIster rendszere OAI aratást végez: a DC metaadatokat XML formátumban gyűjti be, majd ezeket XSLT segítségével a projekt saját bibliográfiai formátumára: DLXS-re alakítja át, leindexeli őket és az XPAT keresőmotorral lekérdezhetővé teszi az adatbázist a felhasználók számára.
- A *NASA* is az OAI-t használja ahhoz, hogy hozzáférést nyújtson a világ legnagyobb űrkutatási és űrtechnológiai információforrásaihoz az USA-ban levő 10 központjában, illetve parancsnokságán, valamint több mint félszáz egyéb országban. A bibliográfiai adatok szabványa a Dublin Core, a teljes szövegű dokumentumok pedig PDF formátumban érhetők el.
- Ugyanezekre a szabványokra alapozták a *CNDLTD* (*China Networked Digital Library of Theses and Dissertations*) projektet, amely még 1996-ban indult és a *CALIS* (*China Academic Library and Information System*) konzorcium irányítja. Az együttműködésben részt vevő kínai felsőoktatási könyvtárak egy központi adatbázisba szolgáltatják be a tudományos fokozatok eléréséhez szükséges értekezések metaadatait.
- A tajvani *National Science Council* 2002 januárjában indította el az *NDAP* (*National Digital Archives Program*) projektet. Ebben az együttműködésben számos egyetem és kutatóintézet vesz részt. Az OAI-PMH segítségével gyűjtik be a DC metaadatokat XML formátumban.
- A *Hong Kong University of Science and Technology* hasonlóan jó példa az OAI-PMH használatára: a DSpace-re alapozott gyűjteményük adatait az OAIster és a Scirus is learatja, így az egész világ számára megtalálhatókká válnak. Ők minősített Dublin Core-t (qDC) használnak a repozitórium metaadataihoz, maguk a dokumentumok pedig PDF-ben tölthetők le.
- A CDS/ISIS-alapú rendszerek számára a CDSOAI projekt teremti meg az interoperabilitás lehetőségét. A CDSOAI egy köztes szoftverréteg az adatforrások és a keresőszolgáltatást biztosítók között, és hasonlóan nyílt forráskódú fejleszt-

tés, mint az ARC nevű, OAI-PMH-t használó Java-alapú programcsomag.

- Bell és Lewis [5] 2006-ban publikált cikkükben az OAI-PMH és a METS használatát vizsgálták a különböző rendszerű repozitóriumok közötti metaadat- és digitális objektumcsere szempontjából. Háromféle alkalmazást elemeztek: a *University of Wales Aberystwyth* könyvtárában működő DSpace-re alapozott repozitóriumot, a *National Library of Wales* Fedora-alapú gyűjteményét és az *ETHOS (Electronic Theses Online Service)* által használt UKETD rendszert, s végül az OAI-PMH-t javasolták az elektronikus formában levő disszertációk cseréjére a három repozitórium között. Shen [6] pedig, szintén 2006-os PhD dolgozatában, a digitális gyűjtemények számára kidolgozott 5S keretrendszer alkalmazhatóságát vizsgálta meg abból a szempontból, hogy hogyan lehetne ezen az elven egy integrált digitális könyvtárat kialakítani a régészeti témájú elektronikus archívumokból.

Összegyűjtéses modell

Ha semmilyen együttműködés, közös alap nincs a különféle tartalomszolgáltató szervezetek között, akkor is el lehet érni egy minimális interoperabilitást úgy, hogy egy webes keresőgéppel összegyűjtjük a nyilvánosan elérhető információkat. Tekintve, hogy ebben az esetben a gyűjtemények részéről semmilyen feladat nincs, nagy mennyiségben lehet így összehozni különböző digitális könyvtárak tartalmát, de persze a közös kereső által nyújtott szolgáltatás színvonala alacsonyabb lesz, mint amikor a partnerek közvetlenül együttműködnek. Mindenesetre az új webes fejlesztések egy része éppen az interoperabilitás javítását célozza, úgyhogy, bár a szemantikus web teljes körű megvalósítása még csak szép álm, a jövőben várhatóan egyre jobb szolgáltatásokat lehet majd építeni a keresőrobotokkal begyűjtött adatokra is. A *ResearchIndex* (korábbi nevén: *CiteSeer*) kiváló példa a nyilvános információk teljesen automatizált összegyűjtésére alapozott digitális könyvtári szolgáltatásra. Megemlíthető még, hogy az ingyenes *Greenstone* gyűjteménykezelő rendszerben is megvalósították ezt a *gathering* modellt – a *federated* és a *harvested* mellett – az *Exchange Center* nevű szoftverkomponensben.

Következtetések

A könyvtárak mindig és mindenhol törekednek az együttműködésre, hogy más gyűjteményekhez is hozzáférést tudjanak nyújtani és ezáltal minél jobban kiszolgálhassák a használóikat. Az együttműködés konkrét formái a mindenkori fejlettségi szintől függenek. A digitális könyvtárak viszonylag új szolgáltatások és a világméretű web virtuális környezetében működnek. Ebben a világban olyan gyors változások zajlanak, hogy a legtöbb eszköz és protokoll élettartama nagyon rövid. Amelyeknek sikerül hosszabb ideig fennmaradni, azok válnak általánosan elterjedté és a használók elkezdik támogatni őket. Elképzelhető, hogy vannak az itt ismertetetteken túl még egyéb interoperabilitási technológiák is, de azok vagy nagyon újak, vagy nem használják ezeket a könyvtárak, vagy még nem publikálták őket a fontosabb szaklapokban.

A témába vágó publikációkat áttekintve megállapítható, hogy az első együttműködésekéről szóló beszámolók 1998 körül jelentek meg, tehát nem sokkal a digitális könyvtárak megszületése után. A szakcikkek száma 2003 óta nőtt meg jelentősen, ami azt jelzi, hogy egyre nagyobb jelentőséget kap a kooperáció a digitális gyűjtemények fejlesztésében. Az ilyen jellegű projekteknek mintegy fele az Egyesült Államokhoz kötődik, és itt dolgozták ki az olyan fontos szabványokat is, mint az OAI és a Z39.50. Megfigyelhető az is a szakirodalmat elemezve, hogy bár a Z39.50-nek nagy múltja volt a könyvtári szférában, és széles körben használják hagyományos dokumentumok bibliográfiai adatainak közös lekérdezésére, a digitális könyvtáraknál a több lehetőséget nyújtó és egyszerűbben megvalósítható OAI protokoll terjedt el. Bár a jelen tanulmány a metaadatok szintjén zajló együttműködések vizsgálata, meg kell említeni még, hogy egyre gyakrabban merül fel feladatként maguknak a teljes szövegű dokumentumoknak a cseréje is az egyes digitális gyűjtemények között, amihez szintén kellően szabványosított formátumokra és eljárásokra van szükség.

Hivatkozások

- [1] SULEMAN, H.: Open digital libraries. Dissertation for the degree of Doctor of Philosophy in Computer Science and Applications. Blacksburg, VA., Virginia Polytechnic Institute and State University, 2002.

- [2] PAEPCKE, A. – CHANG, C. – GARCIA-MOLINA, H. – WINOGRAD, T.: Interoperability for digital libraries worldwide. = *Communications of the ACM*, 41. köt. 4. sz. 1998. p. 33–43.
- [3] WARREN, P. – ALSMEYER, D.: Applying semantic technology to a digital library: a case study. = *Library Management*, 26. köt. 415. sz. 2005. p. 196–205.
- [4] ARMS, W. Y.: A spectrum of interoperability. = *D-Lib Magazine*, 8. köt. 1. sz. 2002.
<http://www.dlib.org/dlib/january02/arms/01arms.html> (letöltve: 2008.04.10.).
- [5] BELL, J. – LEWIS, S.: Using OAI-PMH and METS for exporting metadata and digital objects between repositories. = *Program: electronic library and information systems*, 40. köt. 3. sz. 2006. p. 268–276.
- [6] SHEN, R.: Applying the 5S framework to integrating digital libraries. Dissertation for the Doctor of Philosophy in Computer Science and Application. Blacksburg, VA., Virginia Polytechnic Institute and State University, 2006.
- /ALIPOUR-HAFEZI, Mehdi – HORRI, Abbas – SHIRI, Ali – GHAEBI, Amir: Interoperability models in digital libraries: an overview = *The Electronic Library*, 28. köt. 3. sz. 2010. p. 438–452./

(Drótos László)

A Wikipedia és a Facebook a tájékozódás fő eszközei

A *Telekom Austria Group* az osztrák, a szlovén és a horvát internetezők szokásait vizsgálta. Kiderült, hogy a népszerű online lexikont és a közösségi portált használják a leggyakrabban az emberek.

Ausztriában, Szlovéniában és Horvátországban a megkérdezett felhasználók 75 százaléka fontosnak tartotta, hogy szélesávú internet-hozzáférése legyen. Az osztrákok 50 százaléka internetelés nélkül úgy érzi magát, mintha el lenne vágva a külvilágtól, és erősen korlátozva lennének a kommunikációs lehetőségei. Szlovéniában ez az arány még magasabb, 65 százalék, míg Horvátországban 80 százalék.

„Az internetezőknél nagyon magas azoknak az aránya, akik a Web 2.0-s alkalmazásokra mint hiteles információforrásokra tekintenek. Úgy is mondhatjuk, hogy a Wikipedia az új televízió és az új napilap, s a megkérdezettek több mint 50 százaléka tartja hitelesnek” – összegezte a tapasztalatokat prof. dr. *Rudolf Bretschneider*, a *GfK Austria* ügyvezetője. Emellett kiderült az is, hogy a fiatalok gyakorlatilag képtelenek meglenni a közösségi oldalak nélkül. Ausztriában a 19 évesnél fiatalabbak 85, Szlovéniában 90, míg Horvátországban 100 százaléka regisztrálta magát a Facebookon.

Mivel a felhasználók bíznak ezekben a honlapokban, így ott könnyen megosztják másokkal a velük kapcsolatos adatokat. A horvátok 67 százaléka adta ki a Facebookon a valódi nevét, 69 százaléka a fotóját és 63 százaléka az e-mail címét. Szlovéniában a válaszadók 29 százaléka még a postai címét és 20 százaléka a telefonszámát is megosztotta a Facebookon lévő ismerőseivel. Az osztrákok és a szlovén fiatalok egyharmada egyáltalán nem foglalkozik a magánélete védelmével, a horvátoknál ez az arány pedig 50 százalék.

Az internet vált a mindennapi tájékozódás első számú eszközévé, ezzel szemben a rádió és a televízió szinte kizárólag kikapcsolódási célokat szolgál. Emellett egyre jobban összefonódik a világháló és a mobil készülékek világa. Ezt támasztja alá, hogy az osztrák és a szlovén megkérdezettek 30 százaléka már a mobiljáról lép be a közösségi oldalakra, Horvátországban pedig ugyanígy tesz a válaszadók 40 százaléka.

/SG.hu Hírlevél, 2010. szeptember 2., <http://www.sg.hu/>

(SzP)

Science of Synthesis Houben-Weyl Methods of Molecular Transformations®

- The most effective and reliable methods for group transformations
- Covers all fields of organic chemistry
- Continuously updated (4 releases per year)



Pharmaceutical Substances

Syntheses | Patents | Applications

- All approved active pharmaceutical ingredients (APIs)
- Continuously updated
- Structure searchable

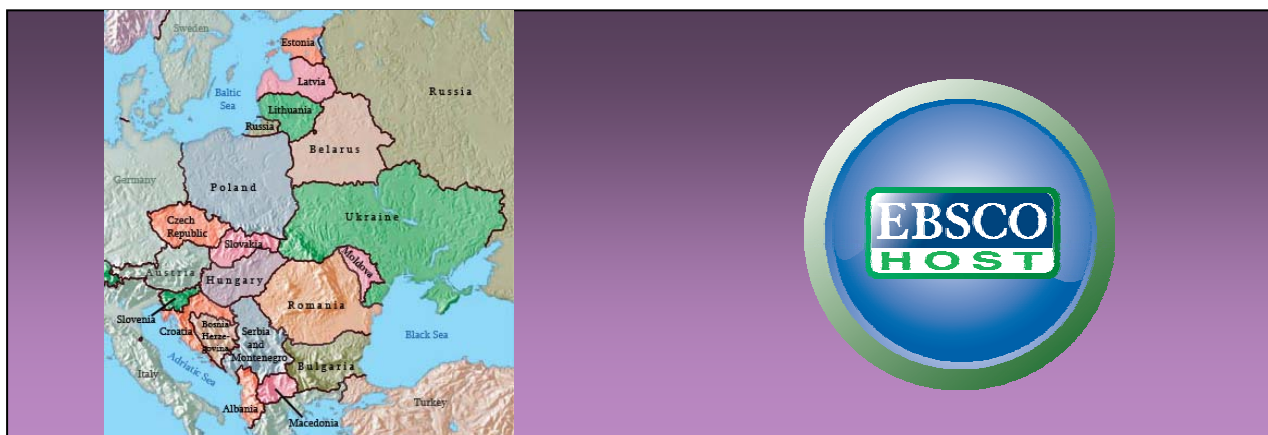
Science of Synthesis:
Organic and organometallic synthetic methods critically evaluated by experts – an authoritative information resource

Pharmaceutical Substances:
Your reliable source for the syntheses of active pharmaceutical ingredients.

For more information and a free trial please contact:

Thieme Institutional Sales
E-mail: eproducts@thieme.de
Tel.: +49-7 11-89 31-407





Central & Eastern European Academic Source™ (CEEAS)

az EBSCO új, teljes szövegű adatbázisa

A Central & Eastern European Academic Source™ (CEEAS) multidiszciplináris, többnyelvű adatbázis, mely Közép- és Kelet-Európában publikált teljes szövegű folyóiratokat tartalmaz.

A gyűjtemény több mint 435 teljes szövegű publikációja lefedi a régióhoz kapcsolódó akadémiai kutatások valamennyi területét.

Az adatbázisban megtalálható tudományos területek:

- ✧ Közgazdaság, üzleti tudományok
- ✧ Történelem
- ✧ Jog
- ✧ Nyelvtudomány
- ✧ Könyvtártudomány és informatika
- ✧ Irodalom
- ✧ Orvostudomány
- ✧ Politológia
- ✧ Szociológia

További információ: EPPrague@ebscohost.com

www.ebscohost.com