

Kardos András

Mi újság a MOKKA háza táján? – az UTCA1 projekt

A „Mi újság a MOKKA háza táján” című sorozatban ezúttal a MOKKA-ODR megújításának az UTCA projekt fejlesztői által elképzelt irányával ismerkedhetnek meg, amely innovatív megoldásokat keres – a jelen helyzet ismeretére alapozva. Cél a könyvtárakat – a közös katalógusra építve – a weben is ismert és elismert tudásbázissá tenni. Alapkérdések újrágondolása, a műszintű csoportosítás bevezetése, és a Könyvtárportállal való integrálás megvalósítása fontos eleme koncepcióknak. Javaslatokkal élünk a leendő közbeszerzési pályázat kiírásához és lebonyolításához is, az esélyegyenlőség növelése érdekében.

Itt az idő

A MOKKA bővítése és megújítása, az ODR-rel való összevonása évek óta elhangzó javaslat a szakmai fórumokon, sokat olvashattunk, hallhattunk róla, legutóbb a debreceni vándorgyűlésen. Most végre itt a lehetőség, hogy a fejlesztés európai uniós támogatásból megvalósuljon – de ha csak az valósul meg, amit a MOKKA indulásakor elképzelték, az ma már igencsak kevés. *„Mintegy 100 könyvtár egyesített katalógusa a kifejezetten helyi jellegű kiadványok egy kis részén kívül 100%-ig lefedi a magyar dokumentumokat... minden magyar könyvtár számára hiteles rekordok forrása... automatikusan váltja ki, sokkal magasabb szinten a központi katalógus(oka)t... rövid távon elektronikus könyvtárközi kölcsönzés alapjává is tehető...”* – írta Vajda Erik közel 10 évvel ezelőtt,² Mader Béla gondolataiból kiindulva.

A 2010-es közbeszerzési pályázattal (pályázatokkal?) dől el a projekt jövőbeli fejlesztési iránya, s ez hatalmas lehetőség, veszélyes kelepccékké válhatva. Ha a projekt egyszerűen csak „folytatódik”, vagyis stabilizálódik a működése (naprakész betöltések), itt-ott egy kis finomítás (allelőhelyek, statisztika stb.), és csatlakozik pár tucat könyvtár – azzal csak a régi célok valósulnának meg, s ez nagy lehetőség elszalasztása lenne a szakma részéről. Olyan rendszert működtetni, amely pusztán a közös katalógizálás és a könyvtárközi kölcsönzés kiszolgálója, ma már nem ésszerű, ugyanis a rendelkezésre álló anyagi erőforrásokból, a feladatok párhuzamosságoktól mentes megvalósításával, más projektekkel való együttműködéssel, és kész, nyílt forráskódú („open source”)

komponensek alkalmazásával sokkal több is kihozható.

Erről a pluszról szól az UTCA projekt 2006-ban indult munkája, illetve a pályázatokon indulni készülő konzorciumának elképzése. Mi a hatékony közös katalógizálás alapjain megvalósítandó új, felhasználóorientált szolgáltatásokra koncentrálnunk, miközben valós helyzetértékelésre alapozott, innovatív megoldásokat keresünk az alapproblémára. A „folytatásban” érdekelt MOKKA Egyesület, a projektvezetők, a többi fejlesztő, köztük a jelenlegi rendszer szállítója az e-Corvina, és az IMOLA koncepciót³ megalkotó SZTAKI/Mongúz konzorcium mellett (részben hozzájuk viszonyítva) szeretnénk mi is megmutatni, hogy hogyan képzeljük el a jövőt. Vázzuk, mi az, amit szerintünk most meg lehetne, és meg kellene valósítani a haladás érdekében. Projektünk vázlatos ismertetésével a nyílt verseny és az esélyegyenlőség erősítése a célunk – ezért a cikk végén magára a pályázat mikéntjére is kitérünk.

Helyzetkép

A „Mi újság a MOKKA háza táján?” sorozat előző részében Tóth Kornél alaposan összefoglalta a hazai közös katalógizálás „hivatalos” történetét, megemlítve a nagy projekteket, mint az OSZKÁR, a Közelkat 1 és 2, a Vocal, az ODR, illetve a MOKKA, valamint a kisebb rendszereket, így további részletezésre itt már nincs szükség. A közös katalógusokról, azok problémáiról már én is írtam egy korábbi, az UTCA projektet ismertető 2007-es cikkemben⁴. Érdemes a cikksorozat részeit alaposan elolvasni. Tanulságos lehet még Ung-

váry Rudolf tanulmánya az OSZK gépesítéséről⁵, illetve az azt beharangozó levele⁶ a KATALIST-en, melyben élesen mutat rá a szükségtelen párhuzamosságok létrejöttének emberi vetületeire. Bakonyi Géza nyíltan rámutatott a MOKKA fejlesztésének nehézségeire 2004-ben megjelent cikkében⁷, amelyben ezt írta: „az üzemszerű működés során felmerülő problémák prognosztizálhatók voltak”. A Könyvtárportál fejlesztéséből eredő tapasztalatok alapján ehhez még azt tenném hozzá, hogy az együttműködés a különféle rendszerek között sokkal inkább akarat, érdekérvényesítés és kommunikáció, mint technikai nehézségek leküzdésének kérdése. Szinte az összes hazai IKR-rel kapcsolatot tudunk teremteni⁸: több mint 1000 könyvtár állományát tettük kereshetővé, többet, mint eddig bármelyik hasonló projekt. Az adatbázisokból kölcsönzési adatokat, olvasójegy-lejáratokat is el tudunk érni, az olvasók önkéntes adatszolgáltatása alapján.

Közismert tény, bár keveset beszélünk róla, hogy a MOKKA fejlesztése igen rögzös úton haladt, és még ma is komoly gondokkal küzd. Mindig valamiféle bizonytalanság vette körül. Ki, miért és milyen forrásból fogja fenntartani? Mi van a nyilvánosan közzétett nehézségek (összeomlás, költözés, duplumszűrés, lassúsága stb.) hátterében? Milyen frissességgel érhető el a részt vevő könyvtárak állománya az adatbázisban és milyen az adatbázis kihasználtsága? Úgy érzem, nekem sikerült viszonylag jól átlátni a helyzetet: rengeteg utánajárás, beszélgetés, levelezőlista-archívumolvasás, valamint a MOKKA, egyébként példás rendben lévő irattárba való betekintés eredményeként. Az irattárba az eredeti tender anyagai után kutatva jutottam el.

Az említett kommunikációs gondok enyhítésére jó kezdeményezés a MokkaWiki⁹ létrehozása, a vándorgyűléseken tartott ODR-MOKKA fórumok, vagy a lassan-lassan beinduló párbeszéd és együttműködés a könyvtári szoftverek fejlesztői között, melynek előmozdításában több külső szereplő is részt vett, így az MKE FITT¹⁰ szekciója, a Könyvtári Intézet a Könyvtárportál révén, valamint a tatabányai ODR konferencia szervezői. Véleményünk szerint a nyílt kommunikáció és párbeszéd korábbi hiánya jelentősen hozzájárult a MOKKA és az ODR körüli kérdések megoldatlanságához, ezért az UTCA projekt elképzeléseinek ismertetése előtt röviden kitérek néhány, e sorozat korábbi cikkeiben érintett témára. Koltay Klára a sorozatban első cikkében¹¹ kifejti, hogy véleménye szerint a jobb duplumellenőrzéshez, a web2-es szolgáltatások

megvalósításához elengedhetetlen a MOKKA katalogizálási szabályzatának szigorú betartása, különös tekintettel a duplumkulcsban részt vevő MARC mezők kitöltése. A jó minőségű rekordok előállításának közös felelősségét szerinte „informatikai eszközökkel sem lehet pótolni”, az egyre automatizáltabb rendszer „káosszal büntet” a gyenge rekordokért, és az azokból épült adatbázisban való műszintű keresés „csak káoszt eredményez”. Koltay Klára feltehetően tisztában van azzal, hogy ha a könyvtárak egyszer szigorúan betartják a szabályokat, azt csak az új rekordoknál teszik meg, a meglévő sokmillió rekordállományt vélhetően senki sem fogja visszamenőleg javítani. Ebből viszont már-már az következik, hogy Koltay Klára szerint képtelenség a jelenlegi rekordállomány újszerű felhasználását megvalósítani. Mi nem értünk egyet ezzel, és persze látunk még kiaknázatlan lehetőségeket az informatika eszköztárában, amelyek segíthetnek a projekten. Sajnálatos, hogy azokon az előadásokon, amelyeken a fenti gondok megoldására, a „régik rekordok” felélesztésére, és újszerű csoportosításukra is képes prototípust mutattunk be az UTCA projekt keretében, többszöri próbálkozásunk ellenére sem alakult ki beszélgetés a fejlesztés ilyen irányba történő elmozdításáról. Bár az alternatív javaslatok felé való nyitás szándéka mindig megjelenik Bánkeszi Katalin előadásaiban, szomorú, hogy a témában tartott fórumokon valós párbeszéd nem alakult ki, a fejlesztők pedig sosem tudtak egy asztalhoz leülve érdemi vitákat folytatni a témában.

Az UTCA koncepció

Az UTCA projekt¹² abból indul ki, hogy ma már nem elégséges olyan rendszert üzemeltetni, amely az olvasóknak szánt funkciókat csak mellékesen, vagy jövőbeli tervként tartalmazza, és amely elsősorban a közös katalogizálás és a könyvtárközi kölcsönzés kiszolgálója. Véleményünk szerint, élve a megújulás – vagyis az EU támogatás – lehetőségével, a MOKKA és az ODR adatbázisának, felületének és funkcionalitásának összevonását csupán alapnak kell tekinteni, amire építve olyan könyves témájú tudásbázist és szolgáltatási csomagot kell létrehozni, amely a weben is elismert tényezővé teszi a könyvtárakat. Ehhez az alapok megváltoztatására is szükség van: nagyon fontos például a duplumkezelésben a művek szerinti csoportosítás bevezetése, részben az FRBR-re alapozva¹³, mely a keresés és megjelenítés mellett a más rendszerek felé való nyitást – az UTCA koncepció fontos elemét – is támogatja. Az azonos

integrált könyvtári szoftvert alkalmazó könyvtárak csoportjaiból kiindulva a bemeneti oldalon radikálisan növelni kell a feltöltő könyvtárak számát, valamint elérhetővé kell tenni más könyves forrásokat is, például antikváriumokat, hiszen ők is „előhelyei” a könyveknek, ők is írnak róluk ismertetőket. Egyszerűsíteni és teljesen automatizálni kell a feltöltéseket. A könyvtárak érdekeltségét növelő, a könyvtárosok könnyebb munkáját elősegítő változtatásokat kell bevezetni, hogy a közös katalógizálás számukra több haszonnal és kevesebb nyúggal járjon. A kimeneti oldalon adatforrássá kell válni: a katalógusaink tartalmát a Google webes keresője felé kell közvetíteni, ha már ezt szeretik használni az olvasók, – a már meglévő Google Scholarba való betöltést pedig műalapúvá kell átalakítani. A könyvtári adatokat elérhetővé kell tenni a könyvtárak népszerűségének, hasznátságának növelése érdekében más kezdeményezések számára is, mint amilyen például a népszerű könyves közösségi oldal, a „Moly”, vagy a nagy látogatottságú, könyvkritikákat közlő a „Könyvesblog”. Erre több megoldás is elképzelhető: keresődobozon, könyvlistán, könyvismertető panelen vagy valamilyen programozói interfészen keresztül, amelynek segítségével azok úgynevezett mashupokat¹⁴ hozhatnak létre. A jogi viszonyok tisztázásához elegendhetlen az évek óta „készítés alatt” álló „MOKKA igénybevételei szabályzat” megalkotása¹⁵.

A közös katalógus webes felületét olyan ráépülő eszközökkel kell kiegészíteni, amelyek azt a felhasználók számára napi szinten teszik használhatóbbá – olvasójegyeik egyetlen regisztráció mögötti kezelése, virtuális polcok létrehozása, kedvenc könyvtárak, könyvtártípusok megjelenítése, kölcsönzési történet kezelése, és a keresési és kölcsönzési történet alapján könyvajánlatok generálása. Az „új MOKKA” gyakorlatban is alkalmazhatja a web 2.0 vagy a könyvtár 2.0 címkével fémjelzett irányzatokat. Egyrészt, a divatos értelmezés mentén közösségi térré kell válnia, ahol az olvasást egymás között is népszerűsítik az olvasók, könyveket, hozzászólásokat, ajánlásokat osztanak meg egymással, együtt játszanak a könyvtárak által indított játékokban. Másrészt, s ez az újdonság az UTCA koncepciójában, a könyvtárosok szélesebb körének bevonására van szükség: lehetővé kell tenni, hogy ők (és az olvasók is, megfelelő kontroll mellett) érdemi visszajelzésekkel segítsék a katalógus javítását. Gondoljunk csak el, milyen hatalmas erő rejlik abban, ha lehetővé tesszük, hogy az ország kb. 10 000 könyvtárosa minden nap egy-

egy dokumentum leírását ellenőrizze, javítsa, mindössze pár percnyi idő befektetésével!

A MOKKA, az ODR kéréstovábbító és a Könyvtárportál jelenlegi funkcióit elképzelésünk szerint egyetlen weboldalon kell egyesíteni, akkor is, ha azokat különböző helyen fejlesztett, különálló alrendszerek szolgálják ki a háttérben – e nélkül ugyanis egyes funkciókat több helyen kéne megvalósítani, ami párhuzamosságok létrejöttét vagy fennmaradását eredményezné. Logikus-e, ha egy keresett mű megtalálása után a találat személyes polcra tételéhez, az ahhoz való hozzászóláshoz vagy könyvismertető írásához, annak kölcsönkéréséhez, vagy saját katalógusunkba emeléséhez más és más felületre kell átugranunk? Ha ezen feladatok elvégzése előtt több helyen kell regisztrálnunk és belépnünk, sőt, esetleg már a keresést is más-más felületen kell elindítanunk? A válasz nyilvánvalóan: nem.

Az UTCA fejlesztése 2006 óta szinte kizárólag ingyenes, nyílt forráskódú szoftverek felhasználásával történik. Ez a tudatos – a HunTéka fejlesztéséhez hasonló – választás tette lehetővé, hogy a fejlesztés minimális anyagi ráfordással történjen. Teljes ingyenességről szó sincs, a teljes rendszer mindig is igényel majd olyan ráfordítást, saját munkánkon kívül, amire érdemes anyagi forrásokat áldozni. A LEGO-szerűen felhasznált komponensek, mint a Linux operációs rendszer, a MySQL adatbázis-kezelő, a PHP nyelv vagy a Lighthttpd webservert alkalmazása viszont segíthet abban, hogy ugyanakkora anyagi ráfordítás mellett több „hasznos” funkciót valósítsunk meg. Ezeket a komponenseket számtalan fejlesztő használja, így a létező fizetős support mellett segítőkész és tapasztalt alkalmazói közösséggel is rendelkeznek. Minőségük is bizonyított: olyan weboldalak háttérét adják, mint a Wikipédia, a Facebook, a YouTube és a Flickr, amelyek forgalma, terhelése percenként akkora, mint a MOKKA rendszerének egy egész évben. Igaz, a felsorolt szolgáltatások a hatalmas terhelés kiszolgálásához erős szerverparkot és nagy létszámú fejlesztőgárdát tartanak fenn, a lényeg itt a „fizetős” szoftverekkel, például az Oracle adatbázis-kezelővel való egyenrangúságon van. Mára a nyílt forráskód körül kialakult mozgalom eljutott a könyvtárig is: Z39.50 kliens és szerver, MARC szerkesztőfelület és konverziós eszközök érhetőek el ingyen, úgy, hogy mögöttük széles felhasználói bázis áll. Az UTCA konzorcium tagjai az egyes elemek használatában több éves tapasztalattal rendelkeznek, ezért a további fejlesztést is hasonló alapokon képzeljük el.

Működési modell

Az UTCA részben ugyanazt a működési modellt alkalmazza, amit jelenleg a MOKKA is: fizikailag központi katalógus, amely a könyvtárak saját rendszereibe felvitt rekordokat egyesíti közös katalógussá, és szolgáltatja a többi könyvtár felé. Közvetlenül a központi adatbázisban nem folyik katalogizálás, mert bár ez előnyökkel járna, bevezetése sem az 1997-es MOKKA pályázat idején – amikor pályázóként ezt ajánlotta az Aleph szállítója – sem most nem reális a könyvtárak eltérő egyéni érdekei miatt. Az UTCA ezt az alapmodellt kibővítve, az ún. hibrid közös katalogizálási modellt választotta. A fizikai összetöltés mellett egyes kiegészítő adatbázisokat képes lesz majd virtuálisan bekapcsolni a keresésekbe, valamint egyes adatokat, mint a példány szintű lelőhely- és kölcsönzési státuszinformációkat valós időben lekérdezni. A fizikailag is összetöltött, az országos dokumentumállomány nagy részét lefedő, „helyben” tárolt bibliográfiai adatokra építve az ismert előnyökön túl (sebesség, névváltozatokra keresés stb.) azért is szükség van, hogy a virtuális módon lekérdezett forrásokból jövő rekordokat viszonylag hatékonyan lehessen bekapcsolni a műszintű csoportosításba – ahogy az a tisztán virtuális modellben véleményünk szerint nem lenne megvalósítható.

Az UTCA alapja olyan adatbázis, amelyben minden olyan könyvtár rekordjai szerepelnek, az összes lelőhellyel, allelőhellyel és példányazonosító kóddal egyetemben, amelyek valamilyen módon beszerezhetők. Az eredeti rekordok téra ez, amely később, a megjelenítési felületek kiszolgálásakor is elérhető, így nem szükséges minden begyűjtött információt egyetlen MARC rekordba sűríteni. Erre az adatbázisra épül egy másik, amely a forrásrekordokból kinyert érdemi információk, tehát nem bizonyos MARC mezők, hanem személyek és kiadók, szerepeik és névváltozataik, címek, helyek és tárgyszavak és azok kapcsolatai köré rendeződik – informatikus fogalommal élve: relációs adat-szerkezetben. A második adatbázis az elsőnek a duplumszűrt és a lehetőségekhez mérten művek szerint csoportosított leképezését adja. Az ezt megvalósító technológia prototípusát több konferencián bemutattuk¹⁶. A művek szerinti tárolásnak – ellentétben a megjelenítés pillanatában való csoportosítással – több előnye is van: gyorsabbá teheti a keresést, mélyebb összefüggések meglátását teszi lehetővé, segíti a könyvtárközi kölcsönzési kérések feladását és a könyvek Google indexbe kerülését is – hiszen ezeknél többnyire egy-egy mű az érdekes, nem egy-egy kiadás.

Alapvető eltérés a jelenlegi MOKKA működési modelljéhez képest, hogy az UTCA minden bejövő rekordot eltárol a későbbi felhasználás céljából. A tárolás normalizált formában, Unicode karakterkészletben¹⁷ történik. Az eredeti rekordok megőrzésének több előnye is van. Megszűnik az a döntéskényszer, hogy a közös katalógusba kerülő rekordok adattartalmát bevigyük-e egy közös rekordba vagy eldobjuk, hiszen az említett második, relációs szerkezetű bibliográfiai adatbázis sokkal szabadabb teret ad az egyes adatelemek kezeléséhez. A jelenlegi módszer vélhetően teljesen MARC rekordokra hagyatkozik a bibliográfiai adatok tárolása terén, ami véleményünk szerint erősen korlátozza a mozgásterét. A MARC-nak mint adatcsere-formátumnak kizárólag a be- és a kimeneti interfészeknél (pl. Z39.50) van „előnye”.

Az UTCA modelljének eredményeképpen nem kell például mellőzni az azonos dokumentumot különbözőképpen leíró (a tartalmi keresést egyébként nagymértékben segítő) ETO-jelzeteket, csak azért, mert azok túlságosan megnövelnék a később letöltött rekordok méretét. A sokat emlegetett possessorbejegyzések és más, példányspecifikus adatok kezelése is tisztábbá válhat. Ezen adatoknak külön helye lehet az UTCA relációs adatbázisában, s így nem kell őket erőltetetten az egységesített rekordban tárolni vagy onnan kizárni. Később, ha a megjelenítésnél szükség lesz rájuk, akkor könnyedén elővehetők lesznek az eredeti rekordokból. Az eredeti rekordok a központi adatbázis naprakészségét biztosító frissítések alapját is képezhetik, hiszen hozzájuk hasonlítva az újonnan beérkező adatokat, könnyen felfedezhetők a törölt, vagy lelőhelyet váltott példányok is.

A „művek szerinti csoportosítás” kapcsán, melyre az UTCA projekt keretében törekszünk, fontos kitérni az FRBR-re is. Az FRBR, vagyis „Functional Requirements for Bibliographic Records” nem szabvány, hanem ajánlás. Modell, amely hierarchikus rendbe próbálja szervezni az általánosabb értelemben vett *műtől*, mondjuk a „Rómeó és Júliától” az azt „megvalósító” könyvön, filmen, operán és a többin át, egészen a konkrét példányokig nyúló ívet. Az FRBR körül élénk vita alakult ki – számos példa hozható fel ugyanis, amely nehezen illeszthető e szigorú rendszerbe. A kérdés alaposabb kifejtése messze túlmutatna e cikk keretein, röviden ezért csak annyit mondanánk, hogy az UTCA projektben, részben e viták tanulságát levonva, egy egyszerűbb modellt szeretnénk alkalmazni, amely a szigorú kategóriák határait fellazítja, de részletesebben foglalkozik az egyes entitá-

sok tulajdonságaival és kapcsolataival. Számunkra Rómeó és Júlia története, legyen az akár könyv, film, vagy mese, nyelvtől függetlenül egyetlen mű lenne – melyhez közelebb lépve megtaláljuk az egyes dokumentumokat, az őket megkülönböztető tulajdonságokkal egyetemben (pl. kétféle fordítás). A feladat nem oldható meg sem 100%-osan, sem egyetlen lépésben, ezért fokozatosan jutunk el erre a szintre, újabb és újabb eszközöket vetve be a művek azonosítására és az őket tartalmazó dokumentumok megkülönböztetésére.

A hibrid közös katalógus jelleget a magot kiegészítő további szolgáltatás valósítja meg, amely olyan forrásokat kapcsol a rendszerbe, melyek rekordjai előre begyűjthetetlenek. Ezek lehetnek hazai – a fizikailag begyűjtött adatbázisokban már szereplő dokumentumok újabb lelőhelyei, vagy külföldi források, az adatbázisban még ismeretlen dokumentumokkal. Ezeket a találatokat a keresések során „röptében” soroljuk be a saját adatbázisból jövő művek mellé. A műszintű tárolás előnye itt az, hogy többnyire elegendő információ áll majd rendelkezésre a hiányos adatokkal rendelkező rekordok azonosításához – ha nem is kiadás, de legalább mű szinten. Tehát, ha egy olyan rekordot kapunk, amely összesen annyiból áll, hogy „Toldi / Arany János”, akkor is meg tudjuk majd mondani, hogy az 1817-ben született Arany Jánoshoz kapcsolható, ha ez az adott rekordban nem szerepel. Ugyanez a megoldás teszi lehetővé az aktuális státusz- és példányadatok elérését, tehát azt, hogy egy adott mű, esetleg adott kiadása vagy példánya éppen kölcsönözhető-e, s ha igen, milyen feltételekkel, vagy ha nem, várhatóan mikortól válik elérhetővé. Az ezt szolgáló technológia már rendelkezésre áll – az UTCA katalógus korábbi prototípusának bemutatóján látható is volt. Használatát IKR-ek, vagy más alrendszerek, például a leendő ODR kérésstovábbító számára API-kon keresztül lehetővé tudjuk tenni.

Adatok fogadása: inicializálás, import, frissítés

A befogadott adatokkal kapcsolatban az UTCA projekt alapelve, hogy a lehető legkevesebb elvárást szabad csak támasztani a kapcsolódó könyvtárak felé, és központilag oldjuk meg az esetleg szükséges formátum- és kódkonverziókat. A lényeg a bibliográfiai adatok és a lelőhelyek pontos beazonosíthatóságán van, hasonlóan ahhoz, ami felé a MOKKA már jelenleg is elmozdult. Elképzelésünk szerint a rekordok tartalmával kapcsolatban

alacsonyabbra kell tenni a mércét. Érdemesebb olyan szűrőket helyezni a bemenet elé a közös katalógusban, amelyek eldöntik, hogy egy rekord betölthető, mert teljesnek látszik; azonosítható, vagyis hiányos leírást tartalmaz ugyan, de az adatbázisban már meglévő rekordok alapján azonosítható, hogy milyen dokumentumot ír le, tehát csak a lelőhelyadatokat érdemes átvenni; vagy értelmezhetetlen, és ezért elutasítandó. A rekordok dokumentumtípus szerinti szűrésével sem a könyvtárakban kell bajlódni (például a KSH jelentős, időszaki kiadványként leírt állományával). Ha azokat a könyvtár látni szeretné valaha a közös katalógusban, küldje csak el, legfeljebb a kérdés megoldásáig mellőzik őket, és ha születik majd megoldás, már nem kell a rekordszolgáltatás átalakításával bajlódni. A rekordok módosításait, a dokumentumok új vagy törölt lelőhelyeinek figyelmét is a központi rendszerben képzeljük el, mindig a könyvtár már ott meglévő, eredetiben tárolt rekordjaival való összehasonlítás révén.

A közös katalógusnak szó szerint naprakésznek kell lennie, hogy hiteles tájékoztatást tudjon nyújtani, ezért érdemes a frissítéseket teljesen automatizálni, kiiktatva minden emberi beavatkozást. Tapasztalatunk azt mutatja, hogy szinte nincs olyan rendszer, amelyből ne lehetne automatikusan információkat kinyerni, és nincs olyan, amelynél ez egy-két hétnél több munkát igényelne, akkor is, ha ehhez a szoftver fejlesztője nem tud, vagy éppen nem akar segítséget nyújtani – hiszen a rendszerek mögött többnyire szabványos SQL adatbázisok állnak. Így van ez az integrált könyvtári rendszerekkel is, ezért a megfelelő jogi vonatkozások tisztázása után készíthető olyan program is, amely közvetlenül a szoftver adatbázisába nyúlva szolgáltatja a szükséges adatokat. A Könyvtárportál közös keresőjének fejlesztési tapasztalatából látszik, hogy szinte mindegyik rendszer elérhető valamilyen többé-kevésbé szabványos interfészen keresztül. Ezek csekély módosításával – a módosítási dátumra való szűréssel – könnyedén kialakítható az elképzelt, a központból a könyvtárak felé forduló teljesen automatikus frissítési mechanizmus. Idő- és költségkímélő megoldás, ha nem írjuk elő újabb szabványok, például OAI implementálását, ha a már meglévőkkel, pl. Z39.50 segítségével is megoldható a frissítés. Az OAI implementálása inkább a kimeneti oldalon lehet érdekes. Véleményünk szerint a MOKKA katalogizálás céljára való használatának gördülékenyebbé tétele érdekében az újonnan megjelenő könyvek leírásait, vélhetően a könyvkereskedőktől vagy kiadóktól, már a megjelenés pillanatában be kell szerezni, és úgy kell

elérhetővé tenni őket a könyvtárak számára, hogy minimális pluszmunkával kerülhessenek át rendszereikbe. Kiemelt fontosságú a dokumentumok könyvtári lelőhelyeinek mielőbbi beszerzése is – akár csak a minimális, az azonosítást lehetővé tevő alapadatok formájában. A beszerzett és a könyvtárak rendszereibe bekerülő „minimalista” rekordok későbbi javításait, tartalmi vagy analitikus feltárását pedig a rekordokban elhelyezett azonosítók alapján a központi katalógusból később közvetíteni kell a könyvtárak felé. Ez utóbbi technológia az UTCA koncepció egyik kulcsfontosságú eleme, amely erősen épít az IKR fejlesztők együttműködésére.

A digitális könyvtárak kiemelt támogatása ugyancsak fontos eleme az UTCA koncepciójának, ami két szempontból is fontos lehet. Egyrészt segít a felhasználóknak, ha „elérjük tesszük” a digitális változatot olyan művekből, amelyekre nagyon nagy igény van, ahogy azt a Könyvtárportál is teszi a MEK, a DIA, vagy az MTDÁ állományával. Másrészt a digitális tartalmakról szóló információ segít a könyvtáraknak, hogy elkerüljék a többszöri digitalizálást. Üdvözlendő kezdeményezés, hogy az ODR továbbfejlesztésének része a már egyszer digitalizált dokumentumok tárolása további szolgáltatás céljából. Érdekes a digitalizált részdokumentumok (cikkek, tanulmányok stb.) nyilvántartását is a központi katalógusban tárolni, és jelezni a felhasználóknak, hogy fennáll a gyorsabb vagy olcsóbb beszerzés lehetősége számukra. Hasznos, ha a főlépcsőként megjelölt dokumentumok adatai is eljutnak a központi adatbázisba, ugyanis így azok hatékony elosztórendszerévé válhat a MOKKA.

A betöltések kapcsán érdemes pár szóval kitérni a kétszintű közös katalógizálásra, amely mostanság több szakmai fórumon is felmerült. E modellben az egyes könyvtárak és a központi katalógus között közvetítő katalógusok jelennek meg. Véleményünk szerint ez csak abban az esetben lehet hatékony megoldás, ha a köztes szintet az azonos IKR-t használó könyvtárak közössége, vagy viszonylag elkülönülten működő könyvtárak csoportja hozza létre, mint például az egyházi könyvtárak. Az IKR-alapú közösségeknél is csak akkor érdemes ilyet kialakítani, ha valamilyen okból a könyvtárak maguk nem érhetők el stabilan a weben. Ilyen lehet például a Szirén vagy a Szikla rendszer felhasználói tábora. Egyéb esetekben a köztes szint beiktatása véleményünk szerint felesleges. A TextLib és a HunTéka könyvtárak virtuális közös katalógusát használni a MOKKA felé menet ugyancsak szük-

ségtelennek tűnik, hiszen az ilyen rendszert használó könyvtárak közvetlenül is elérhetőek – ezzel értékes másodperceket nyerve a keresések futtatásakor.

Duplumellenőrzés, művek, többkötetese

A duplumok megtalálása és összevonása a közös katalógizálás egyik kulcskérdése, hiszen autonóm rendszerekből építkezünk, nem egyetlen nagy katalógust építünk, mint azt teszik például az izlandi könyvtárak. A duplumkérdés mind a fizikailag, mind a virtuálisan közös katalógusoknál felmerül, a célok is hasonlóak lehetnek, de a technikai korlátok, s így a módszerek is eltérőek. A virtuális keresőkben csak korlátozott erőforrás és idő áll rendelkezésre – ennek „gyengeségét” láthatjuk a Könyvtárportál keresőfelületén is. A cél lehet az, hogy az azonos dokumentumok különböző könyvtárakból érkező leírásait társítsuk, de lehet az azokban rejlő műveket is összevonni – az UTCA ez utóbbira törekszik. A műszintű előzetes csoportosítás több előnnyel is jár, szemben a manapság elterjedt, megjelenítéskor történő csoportosítással.

A MOKKA a duplumkulcsos módszert alkalmazza: a MARC rekordok egyes mezőinek, almezőinek értékét egyetlen hosszú szöveges kulccsá alakítják, jellemzően a szerző-cím-kiadás-formátum adatokat felhasználva, majd azon rekordokat tekintik azonos dokumentumot leíróknak, amelyek egyforma kulccsal rendelkeznek. A MOKKA kulcsképzési módszerének leírása elérhető a MokkaWiki-ben¹⁸, így ha a pályázati kiírásban ez kötelező elemként szerepel, akkor az immár nyilvános dokumentáció alapján könnyedén megvalósítható. Az UTCA itt más eljárásokat (is) ajánlana és alkalmazna, de ezeket részletekbe menően nem tárgyaljuk, hiszen ez az egyik „ütőkártyánk” – az eredményt több konferencián, rendezvényen már bemutattuk, s mire ez a cikk megjelenik, egy szabadon kipróbálható prototípus is elérhető lesz a <http://konyvtar.info> weboldalon. A megoldás alapja, hogy a feldolgozott adatokat egyrészt a duplumkulcsos módszernél alaposabban, közelebbről megvizsgálva, a MARC szerkezettől elvonatkoztatva használjuk fel, másrészt a rekordokra nemcsak egyesével, hanem tágabb összefüggéseket keresve úgymond madártávlatból is ránézünk. Olyan, részben könyvtáros, részben informatikus módszerek ezek, amelyek egy gyakorlott könyvtáros hozzáállását próbálják a gépeknek megtanítani. Fontos megemlíteni, hogy az UTCA a jelen helyzet reális megítéléséből indul ki: milliányi kész,

sokszor egyedi leírási szokásokat tükröző rekord fogadására – tehát arra lett felkészítve, hogy a „gyengébb” rekordokból is kihámozza a releváns adattartalmat, amelynek alapján az ilyen rekordok beküldőit is képesek legyünk lelőhelyként szerepeltetni a megfelelő műnél. Nem várhatjuk ugyanis el, hogy a könyvtárak egyszer csak feladják saját szokásaikat, vagy hogy nekiálljanak korábbi rekordjaikat a MOKKA érdekében javíttatni. A sokszínűséget központilag kell lekezelni.

Az UTCA projekt keretében megpróbáltunk feloldani egy régi dilemmát is, ami a bibliográfiai, vagy lelőhely-adatbázis kérdése kapcsán szokott felmerülni: szigorú vagy lazább duplumszűrésre van szükség? A válasz: is-is, a megoldás itt a megjelenítésben rejlik. A mű szerinti csoportosítás lehetővé teszi, hogy az adatbázist láthassuk úgy, mintha szigorúan szűrt lenne – erre akkor van szükség, ha nem lényeges, hogy melyik kiadást kapjuk meg –, de úgy is láttathatjuk, mintha lazábban szűrnénk, például, ha éppen konkrét kiadásra vagy formátumra van szükségünk. Egyszerűen egy második csoportosítási szintet vezetünk be a művek alatt. Ha a találati halmaz elsősorban műveket tartalmaz, de van lehetőség mélyebbre is ásni, megnézni a művek egyes kiadásait is – egy csapásra feloldottuk az ellentmondást!

Egy másik örök vita a valós, illetve a virtuális közös katalógizálás előnyeiről-hátrányairól szól. Az előbbi előnyei nyilvánvalóak: gyors keresés, a kereséshez való előzetes segítségnyújtás lehetősége (authority állományoknak, tezauruszoknak köszönhetően). Az UTCA által használt elsődleges duplumkezelési módszer is csak fizikailag összetöltött adatbázison működik, és ez a MOKKA alapkonceptiója is. Ha a naprakész betöltést sikerül is elérni, és így semlegesíteni a virtuális közös katalógus e téren fennálló előnyét, akkor is lesznek olyan katalógusok, amelyeket csak élő lekérdezéssel lehet a rendszerhez kapcsolni. Az UTCA koncepció ezért a két megoldás előnyeit próbálja kombinálni a hibrid modellben. A művek szerint csoportosított rekordok adattartalmára építve fogja azokhoz az élőben lekérdezett rekordokat hatékonyabban illeszteni (hiszen pl. ismeri annak kiadásait).

A többkötetes és a gyűjteményes dokumentumok kezelésével kapcsolatban egyetértünk a MOKKA jelenlegi fejlesztési irányával: a különböző forrásokból származó, eltérő leírási móddal készült rekordokat egységes formára kell hozni, a rekordszerkezetet és nem a feladó könyvtárat figyelembe

véve. Csak ezután juthatunk el a duplumok, illetve a művek felismeréséig.

Az UTCA katalógus tökéletesítésébe – több szem többet lát alapon – be kívánjuk vonni a katalógus használóit is, ha első körben nem is a közvetlen javításba, de a hiányosságok feltérképezésébe mindenképpen. Egy-két kattintás a használó részéről a milliónyi rekord közül a megfelelőre irányíthatja figyelmünket. Ez a módszer elsősorban – ugyan a „long tail” effektus¹⁹ sem elhanyagolható – azoknak az állományrészeknek a megjelenítését fogja javítani, amelyeket a legtöbben használnak, ami végül is pont jó választás. A feldolgozási módszerünk ezen visszajelzések és a saját további kutatásaink alapján később fokozatosan finomítható, így idővel egyre pontosabban tudjuk majd azonosítani a műveket, további duplumokat a helyükre illesztve. Ez azért sem jelent különösebb gondot, mert a MOKKA és az ODR összevont rekordmennyiségére, 5-10 millió rekordra vetítve egy alapfeldolgozás (inicializálás, a MOKKA terminológiájában) várhatóan kb. egy-két hetet, egy-egy újabb újracsoportosítás kb. egy-két napot vesz igénybe a fejlesztés során használt – egyébként az eBay internetes kereskedelmi portálon kb. 140e forintért vásárolt – középkategóriás, kétprocesszoros HP DL380 G3 típusú szerveren.

Webes felület: a katalógus

A MOKKA eredetileg a szakmai közösségnek készült, Z39.50 eléréssel, de ma már nyilvánvaló, hogy az olvasók tömegeinek kell használni ahhoz, hogy érdemes legyen fenntartani. A MOKKA és vele együtt az egész könyvtári rendszer „arca” tehát a webes felület, ezért is fontos, hogy ez a lehető legtöbbet nyújtsa. A dokumentumállomány kihasználtságában véleményünk szerint igen nagy szerepet játszik a webes katalógusok milyensége – mind pozitív, mind negatív irányban. Az idei *Internet Librarian* konferencián *Dave Pattern* tartott a témáról igen érdekes előadást²⁰. Ők a kölcsönzési és a keresési adatokat használják fel a katalógus előzékenyebbé tételéhez, s ezzel a használati mutatók növeléséhez. Ha ugyanis a katalógus nehézkes, lassú, vagy nem elég segítőkész, akkor az olvasó esetleg nem találja meg, amit keres, sőt visszaretten a katalógus használatától. Nem szerencsés, ha elvárjuk az olvasótól, hogy úgy írja a neveket, ahogyan a könyvtári szabvány megkívánja („Merle, Robert” illetve „Szilvási Lajos”), hogy tudja milyen jelet – *, \$ vagy éppen % – használunk a „csonkolás” jelölésére (már a szó is ijesztő),

és a dokumentum nyelvének megadásakor sem szerencsés ISO nyelvkódok ismeretét elvárni. Az előzékenység hiánya az is, ha csak a megfelelően kötőjelezett, vagy éppen kötőjel nélküli ISBN-re adunk találatot, vagy ha egy könyv szerzőjét csak azért nem jelenítjük meg azonnal a találati listában, mert az nem a MARC 100-as mezőjében szerepel, vagy ha a keresés finomításához, egyes információk megtekintéséhez feleslegesen sok kattintás szükséges, az mind-mind potenciális olvasókat riaszt el.

De nézzünk meg egy valós példát a MOKKA WebPAC-ban! Keressünk *Schiller Erzsébet* névére a „szerző” mezőben. A 19 találat közül – amelyből csak tízet látunk lapozás nélkül – az első háromnál nem derül ki, miért jelentek meg. Közülük kettőben tényleg közreműködő a keresett szerző, de a harmadik találat egy tipikus hibára mutat rá, abban ugyanis csak *Schiller Alfréd* és Székács *Erzsébet* a közreműködő. Négy további találat is hasonló, *Friedrich Schiller a szerző*, ezek a találatok *Gergely Erzsébet* révén kerültek ide. A tájékozottabbak azt gondolnák, „Schiller Erzsébet” nevét így, idézőjelek közé kellett volna beírni a helyes találatokhoz, mi mégis úgy véljük, hogy egy intelligens keresőnek e nélkül is előbbre kéne sorolnia a releváns találatokat. Sajnos, az idézőjel sem befolyásolja a MOKKA keresőjét, ugyanazt a 19 találatot jeleníti meg. Az „Arany János” keresésnél a „túl sok találat (2167 darab)” üzenetet írja ki – nem a vélhetően kevesebb, de Arany Jánoshoz tényleg kapcsolódó találatot. A „szerző” mező túl szigorúan vett értelmezése miatt az egyébként helyes találatok közül is csak kettőnél látjuk *Schiller Erzsébet* nevét. Az, hogy mire kerestünk, a találati oldalon már nem látható, és a keresés sem finomítható itt tovább. A dátum szerinti rendezés is hibás, a „cop.” jelzésű, vagy a kapcsos zárójelbe tett dátumok rossz helyen szerepelnek. A jelenlegi MOKKA felület kritikája lényegében a működtető e-Corvina szoftver kritikája, de az észrevételek más rendszerekre is állnak. Nem olyan hibák ezek, amelyek lehetlenné teszik egy-egy dokumentum megtalálását – mégis érdemes rajtuk kisebb-nagyobb munkával segíteni, hiszen ez még csak az alap – az ilyen hibáktól mentes katalógusban is sok lehetőség lenne még arra, hogy országos szinten növelje a könyvtárak állományának kihasználtságát.

Szeretnénk olyan webes keresőfelületet alkotni az UTCA számára, amely mentes ezektől a hibáktól, sőt, előzékenységgel, számtalan apró kis finomítás bevezetésével, valamint a nemzetközi trendekben megjelenő ötletek kiaknázásával javítja annak

használhatóságát. Mivel ezeket igazán csak „előben” lehet kipróbálni, szeretnénk egy prototípust elérhetővé tenni, várhatóan e cikk megjelenése környékén. Hiszen pusztán leírva nehezen értékelhető, hogy a katalógusnak viszonylag kevés, de releváns találatot kell hoznia (lásd a fenti kritikát, illetve a mű szerinti csoportosítás tárgyalását), gyorsan (2-3 másodpercen belül), és egy felületen elrendezve minden szükséges információt (pl. a dokumentumok aktuális kölcsönzési státuszát is, vagy a művek kiadásokra bontását). Talán egy apró példa: a jelenlegi felületen egy könyvtárban vagy az összesben egyszerre lehet keresni, a használó tehát egyesével végigkeresi kedvenc könyvtárait, vagy az összesben keres, így viszont sok irreleváns találatot kap. Egy lehetséges megoldás: az összesben keresést szorgalmazzuk, mert így kiderül, hogy valahol máshol azért lehet példány az adott könyvből, viszont kedvenc könyvtárainak találatait külön megjelöljük.

Számomra meghatározó olvasmányélmény volt a 90-es években – ne feledjük, informatikus volnék – az „Apple Human Interface Guidelines”²¹ című könyv, az Apple, akkori munkaadóm útmutatója a felhasználói felületek tervezéséhez. Hosszú fejezeteken át, sok szempontból elemezgették, hogy miként működjön, és hogyan nézzen ki egy felhasználóbarát alkalmazás. Milyen elrendezésük legyenek az ablakok, milyen szövegezésük az üzenetek, milyen színeket alkalmazzunk, hogyan segítsük a felhasználót a tájékozódásban, és mi az, amit kerülni illik? Olyan részletekre is kitértek, amelyeket a felhasználók tudatosan nem is érzékelnek, de amelyektől komfortérzetük erősen függ. Apró finomságokra, figyelemre, emberismeretre, hétköznapi pszichológiára tanítottak – ez az a hozzáállás, amit az UTCA katalógus felületének megalkotásakor képviselni szeretnénk.

Könyvtárportál-integráció

A „könyvtár 2.0” fogalom, amit sokszor már-már divatból is használunk, valójában olyasmit takar, ami a MOKKA alapvető feladatai közé tartozik, és amit valahol a könyvtárosok (és sokan mások is) régóta művelnek: közösen, együtt építeni egy tudásbázist. Véleményünk szerint már az is komoly dolog lenne, ha ebben a munkában több könyvtáros vehetne részt, egyenrangú félként. A Könyvtárportál²² és a MOKKA szemléletmódja nagyon is hasonló: gyűjtjük és szolgáltatassuk egyetlen központi adatbázisból az adatokat, a láthatóság növelése mellett munkát takarítva meg minden részt

vevő fél számára. A MOKKA bibliográfiai adatok, a Könyvtárportál felhasználók, olvasójegyeik, könyvismertetőik és hozzászólásaik, könyvtárak és hírek „közös katalógizálását” és szolgáltatását végzi. Jól kiegészíthetik egymást, ezért is örömteli, hogy a „Könyvtárportál integráció” már több helyen felmerült. *Bánkeszi Katalin* is többször jelezte ennek fontosságát, és *Tóth Kornél* is említi az IMOLA koncepcióról szóló cikkében. A *konyvtar.hu* regisztráció mögött ott van a felhasználó kedvenc könyvtárainak és olvasójegyeinek listája – tehát tudjuk, hol kíván elsősorban keresni, és ezt felhasználhatjuk a MOKKA kereséseknél. Az „új MOKKA” adatai, ha megfelelő minőségű bibliográfiai adatokkal tudnak szolgálni, kitűnő háttérrel adnak a portál további funkcióihoz, például: könyvlisták (virtuális polcok) kezeléséhez, könyvismertetők hozzárendeléséhez, vagy a Google találati listáiba való bekerüléshez. A portál katalógusa ugyan több mint 1000 könyvtár állományát teszi kereshetővé, és többet, és többféle rendszerből, mint eddig bármelyik efféle hazai próbálkozás, de ezt meglehetősen „fapados” felületen teszi, ráadásul a virtuális keresési jelleg miatt gyenge duplumszűréssel és viszonylag lassan. A MOKKA ennél sokkal jobb háttér lehetne. Elképzelésünk szerint a két szolgáltatás ideális esetben egyetlen felületre kerülhetne, tovább erősítve a könyvtárak egységes megjelenését a weben.

Statisztika

Ha arról van szó, hogy a MOKKA használói között sok-sok olvasót szeretnénk látni, egy dologra érdemes mindenképpen kitérni, ez pedig a statisztikák kérdése. Nyilvánvaló és jogos aggodalom a könyvtárak részéről, hogy a közös katalógusban keresők az ő statisztikáikból „hiányoznak” – ezzel rontva megítélésüket fenntartóik szemében. Lehetővé kell tenni a könyvtáraknak, hogy jelentéseikben elismertethessék azokat a kereséseket is, amelyeket egy közös katalógusban indítottak, és amelyben a könyvtár releváns találatként, vagy legalább forrásként megjelent. Ha rákeresünk egy témára, s egy könyvtár néhány dokumentummal szerepel a találati listában – úgymond „rávetült egy pillantás”, angol webstatisztikai fogalommal élve „page impression”-t ért el –, azt mindig fel kell jegyezni, és később összesítve továbbítani a könyvtárak felé. Ez az adat a könyvtárak számára azért is nagyon hasznos lesz, mert mutathatja az állományuk egyes részei iránti érdeklődés mértékét. Ugyancsak érdekes lehet a könyvtárak számára az

őket érintő, de tőlük találatokat nem hozó keresések figyelése is – vagyis az, hogy esetleg mit lenne érdemes beszerezniük.

Adatszolgáltatás: interfészek, használati jogok

A könyvtárosok akkor lesznek valóban érdekelték a közös katalógus használatában és építésében, ha abból valódi, napi szinten megjelenő előnyük származik, ha annak használata egyszerű, és mentes a gépies rutinfeladatoktól. A duplumkérdés csak az egyik, amely ma a gördülékeny munkát gátolja. Komoly gondot jelent az is, hogy a rekordletöltés után megfigyeléseink szerint túl sok időre van szükség az adatok saját rendszerbe illesztéséhez. Mint egy nemrégiben tartott MOKKA műhelytalálkozón²³ is szembesülhettünk vele, a katalógizáló munkatársak sokszor hosszasan alakítgatják a rekordokat a letöltés után: például (pl.?) rövidítések létrehozásával vagy feloldásával, központosítási jelek manipulálásával, analitikus feltárás mezőinek átalakításával, feleslegesnek ítélt almezők törlésével – ahogyan azt a helyi szokások és a használt könyvtári rendszer megkívánja. Gépies, de kézzel elvégzett munka: időpocsékolás és nyűg. Ezeknek a nehézségeknek az orvoslása véleményünk szerint kulcspontja lehet a MOKKA hagyományos könyvtári célra való alkalmazásának népszerűsítésében. Az UTCA projekt keretében már kidolgoztunk egy elgondolást ennek technikai megoldására. A megoldás három elemből áll. Lehetővé kell tenni, hogy a könyvtárosok személyes preferencia-sorrendet állítsanak fel a használt rekordok forrását illetően, és az általuk közölt szabályrendszerek alapján az említett „fazonírozásokat” a MOKKA oldalán, még a letöltés előtt a rendszer végezze el helyettük, valamint az egyes integrált rendszerekben kötelezően tárolni kell a letöltött rekordok központi adatbázisban használt azonosítóját is, amelynek segítségével később követni lehet a rekord változásait (bővülését), és azt vissza lehet vezetni a használó könyvtár adatbázisába. Ez utóbbihoz már természetesen az IKR-fejlesztők aktív közreműködése is szükséges.

A hatékony adatcsere érdekében a webes felület, a Z39.50 interfész és a Corvina könyvtárak által használt, de egyébként nem dokumentált CCL felület²⁴ mellett olyan programozói interfészeket kell létrehozni, amelyek az adatbázisnak a bibliográfiai adatokon kívüli elemeit is elérhetővé teszik, hiszen azok ugyanúgy, mint a bibliográfiai rekor-

dok, hasznosak lehetnek az egyes könyvtárak számára: a felépülő, névtérként is használható authority állományok, a várhatóan integrált tezauruszfunkció, vagy akár az egyes lelőhelyek státuszadatai. Ezek az interfészek lehetnének a MOKKA adatbázis és a ráépülő ODR funkciók elválasztásának határai is. Ezeket más rendszerek is elérhetnék, hiszen az ODR számára szükséges lelőhely-információ nemcsak a majdani kérdéstovábbító rendszer, hanem az olvasók, sőt az egyes IKR-ek számára is hasznos lehet. Mint már említettük, újabb protokollok bevezetése inkább a kimeneti oldalon érdemes, ott, ahol a MOKKA mint adatszolgáltató lép fel. Az adatokat minden olyan interfészen keresztül elérhetővé kell tenni, ahol az segít bekerülni a nemzetközi vérkeringésbe, a különféle európai szintű közös kezdeményezésekbe. Itt fontos igazán az OAI, az SRU, az SRW, a Z39.50, az ILSDI²⁵, valamint további HTTP alapú felületek minél jobb minőségű megvalósítása, hiszen a MOKKA interfészeként alkalmazva ezeket, egy egész ország kulturális vagyonát tehetjük elérhetőbbé. Külön kiemelném a Google Scholarba való exportálást, amely ugyan már évek óta megvalósult, de amelyen a művek szerinti csoportosítás ugyancsak sokat segíthetne, hiszen oda sem egyes kiadásokat, hanem inkább műveket kell közvetítenünk.

Véleményünk szerint a könyvtárak katalógusában lévő adatok önmagukban ma már nem képviselnek értéket, hiszen a MOKKA megvalósítása értelem-szerűen megszünteti azt a piacot, amelyen a könyvtárak esetleg egymás közötti fizetős szolgáltatásként adhatnak-vehetnek rekordokat. A katalógusadatok minél szélesebb körű felhasználásának támogatása ma már inkább a könyves világra való figyelemfelkeltés egyik eszköze lehet. Igen, a tapintható, szagolható, fizikai könyvek világára. Ezért szükséges – és ez az UTCA koncepció egyik fontos eleme –, hogy a könyves adatokat minél könyvben beágyazhatóvá, kereshetővé tegyük az interneten, saját katalógusainkon túl is. Ennek érdekében újra kell gondolni azok licencelését, és sürgősen meg kell fogalmazni a „MOKKA igénybevételi szabályzat” című dokumentumot, mely a MOKKA weboldalán közel egy évtized óta „Készítés alatt” státuszban van.

Google-faktor

Tudjuk, az olvasók manapság a Google-ban keresnek először, akkor is, ha szerzők, könyvek után kutatnak. A profitorientált szférában, és nem csak

a közvetlenül a webből élők körében köztudott, hogy a Google találatában való jó helyezése szinte létkérdés. Hogy ez a könyvtárak számára is fontos dolog, jól mutatja, hogy a Könyvtárportál indulása óta látogatóinak több mint a fele jön a Google felől. A *konyvtar.hu* oldalai esetenként más könyves témájú oldalakat, vagy éppen a *Népszava* online kiadását is megelőzik az első találati oldalon. Kevesebb „versenyző” közül kell az élbolyba, a találatok élére kerülnünk, hiszen a magyar nyelvet viszonylag kevesen használják (ellentétben például az angollal vagy a franciával), van tehát esély, hogy ezt a lehetőséget még jobban kihasználjuk, és ezzel hozzásegítsük a könyvtárakat ahhoz, hogy a weben is „látható” tudásbázissá váljanak. Vetélytársaink persze vannak. A könyvkereskedők, antikváriumok már régen ott vannak a Google találati listában, némelyek biztosan céltotán is foglalkoznak a felhasználók e téren való elérésével. Fontos, hogy maga a Google érdekelt is abban, hogy a felhasználók számára releváns helyi tartalmat jelenítsen meg – a „könyvtár” szóra keresve nekem rögtön egy térkép is megjelenik a találati listában, rajta néhány nagy budapesti könyvtárral.

Az UTCA projekt elképzelése szerint a fenti cél eléréséhez a leendő közös katalógus kitűnő alap, mivel nagy forgalmú, sokat hivatkozott oldal lesz, a találatként kapott oldalak pedig releváns információkat tartalmaznak majd: a könyv adatai mellett esetleg ismertetőjét, olvasói hozzászólásokat, és ami az egyes könyvtárak számára fontos: lelőhelyeket országszerte. Nem érdemes illyesmivel egy-egy könyvtár OPAC-jában próbálkozni; ez csak felesleges vetélkedést jelentene a „csapaton belül”. Mint már több helyen említettük, ehhez is elengedhetetlen a művek szerinti csoportosításra való törekvés, mely az UTCA projekt része – hiszen a találatokba csak egyetlen „Aranyember”-t akarunk bejuttatni, nem tucatnyi, egymással versengő, a mű valódi lelőhelyeinek csak részhalmaival bíró kiadását.

A pályázatról

Sarkalatos pont a hazai közös katalógizálás megújításakor a pályázattal rendszer és annak előkészítése, melynek során végül eldől, milyen irányba lépünk tovább. Három tényező kiemelten fontos: az, hogy az innováció megfelelő teret és elismerést kapjon, hogy az érdekelt szoftverfejlesztők között szigorú esélyegyenlőség legyen, és hogy az értékelés minél inkább valós eredményekre tá-

maszkodhasson. Az UTCA projekt fejlesztőinek véleménye szerint ezek nélkül a pályázat csak igen kétes értékű eredményeket hozhat.

Csak emlékeztetésképpen: az eredeti, 1997-es MOKKA tender nyertese, a Dynix, rövid időn belül alkalmatlannak bizonyult, holott a kiírást alapos előkészítés előzte meg, és vitákkal tarkított, de körültekintő értékelés követte. A gond az volt, hogy a funkciók nagy része csak papíron létezett – a pályázók sokszor csak jelezni tudták, igen, képesek lesznek majd megvalósítani a kért funkciókat. Kész lehetett ugyan a katalogizáló modul vagy az OPAC – leginkább ezzel rendelkeztek a pályázó rendszerek –, de számtalan, később kemény fejfájást okozó elem, mint a betöltés, a konverzió, a duplumellenőrzés, csak koncepcióként létezett, és bizony ezek voltak az igazán nehéz feladatok az itthoni, igencsak heterogén szoftveres környezetben. A könyvtárak már akkor is ragaszkodtak ahhoz, hogy továbbra is saját rendszerükben, saját szokásaik alapján katalogizáljanak, így azok a megoldások, amelyek a véleményünk szerint előremutatóbb, az egyetlen közös rendszerben való katalogizálást preferálták, elbuktak. A hozzáállás sem sokat változott azóta, sőt, a helyzet tovább romlott, még több „egyéni” rekord keletkezett, amit közös katalógusba kéne szervezni. A feladat tehát nehéz, nagyon nehéz, és más azt papíron, és megint más a valóságban megoldani.

Az első tényező az innováció elősegítése, amire jó példa a duplumellenőrzés: elő lehet írni a jelenlegi módszer megvalósítását is, de teret kell adni az új elképzeléseknek, amelyeknek, mint korábban megmutattuk, számos előnyük van. Ha a kiírás ragaszkodik a jelenlegi duplumkezelési módszerekhez, vagy azok enyhén módosított változatához, azzal elveti az esetleg hatékonyabb megoldások lehetséges előnyeit. Tágabb értelemben, a rendszer egészére tekintve sem feltétlenül a jelenlegi – egyébként a MokkaWiki részeként végre precízen és áttekinthetően dokumentált – módszerek az üdvöztöek, és ezért nem célravezető, ha pontosan ezek megvalósítását várjuk el a pályázóktól. Véleményünk szerint az újító elképzeléseknek és megoldásoknak, a kialakult jövőképnek teret és megfelelő sújt kell adni a pályázat értékelésekor. Nem várható ugyanis el, hogy az érdekelt felek ötleteiket vetélytársaik tudomására hozzák a pályázat előkészítő fázisában, az viszont nem valószínű, hogy e nélkül minden hasznos újítási javaslat felmerül.

A második tényező az esélyegyenlőség. Az esélyegyenlőséget több eszközzel lehet segíteni, és ezek közül első a nyilvánosság. Az elkészült szakértői javaslatok kerüljenek a szakmai nyilvánosság elé, lehessen róluk szakmai vitát nyitni, bármilyen rövid határidővel is. Ha ez nem történik meg, komoly esélyt adunk arra, hogy ne új, emelt szintű, korszerű megoldás szülessen, hanem a jelenlegi meglévő rendszer csinosítása készüljön el. A jelenlegi és az előzetesen bevont fejlesztők olyan helyzeti előnnyel rendelkeznek, amit alapos tájékoztatással kéne ellensúlyozni. Az alternatív javaslatoknak, mint az IMOLA vagy az UTCA, azonos eséllyel kell indulniuk. A nagyobb nyilvánosság valamennyi potenciális fejlesztő, illetve közvetve a felhasználók érdekében rendkívül fontos.

A MOKKA és az ODR összevonása ésszerű és szükséges – különálló működésük eddig is nehezen védhető érvekre épült, mint például a funkcióból eredő eltérő adatbázis-szerkezet, vagy az eltérő duplumkezelés igénye. Az összevonás konkrétumairól, a fejlesztésekre a TÁMOP keretében külön-külön forrásokat elnyerő OSZK, DEENK és a digitális dokumentumküldéssel foglalkozó *Pannon Egyetem Egyetemi Könyvtár és Levéltár* együttműködésének részleteiről, a feladatok elosztásának mikéntjéről viszont a nyilvánosság szinte semmit nem tud, holott ezek kritikus kérdések, újabb felesleges párhuzamosságok megjelenésének terepei lehetnek. A 2008-as szombathelyi ODR-MOKKA fórum anyagaiban²⁶ például feladatként jelenik meg „a MOKKA bibliográfiai adatbázishoz kiegészítő példányadatbázis kiépítése, üzemeltetése és folyamatos bővítése”, illetve külön könyvtárosi és olvasói „ODR portál kialakítása”, vagy olyan megjegyzést olvashatunk, hogy „mivel az ODR adatbázis bibliográfiai alapja a MOKKA lesz” – ami, ha nem csak félreérthető fogalmazásról van szó, arra enged következtetni, hogy az „egyesítés” nem feltétlenül jelent egységes webes megjelenést vagy valóban egységes adatbázist. Mivel várhatóan nem egyetlen pályázat keretében valósul meg a MOKKA és az ODR továbbfejlesztése, ezért ha a pályázatokat nem egyszerre írják ki, szükséges, hogy már jóval az első megjelenése előtt tisztázott legyen a projektek együttműködésének minden technikai részlete. Az egymással szorosan összefüggő funkciók megvalósítására csak így alkothatók koherens, a párhuzamosságoktól mentes rendszertervek. A mi koncepcióinkban szervesen együttműködő, egy felületen megjelenő, egységes regisztrációs rendszert és adatbázist használó komponensek szolgálják ki a közös katalogizálást és a könyvtárközi kölcsön-

zést. A könyvtárközi kérések könyvtárosok általi kezelése, digitális példányok bejelentése szinte az egyetlen részlet, amit ténylegesen le lehet választani azokról a funkciókról, és azokról a felületekről, amelyeket az olvasók is használnak, tehát például a lelőhelyek és a példánystátuszok pontos ismerete, kedvenc források bejelölése, vagy találati listák elmentése. Mivel az OSZK és a DEENK külön-külön nyert jelentős anyagi erőforrásokat a megvalósításra, úgy tűnik – bár logikus lenne, de sosem hallottunk említést róla – nincs szó közös közbeszerzési pályázat kiírásáról.

Az esélyegyenlőség érdekében – véleményünk szerint – a pályázóknak már a jelentkezéskor vállalniuk kellene, hogy ha vesztenek, akkor is gördülékenyen, és már a pályázat kiírásakor elfogadott rendszerterv alapján, fix költségek és határidők mellett együttműködnek a nyertessel. Például a projektindítás első napján a nyertes rendelkezésére bocsátják (vagy akár előre letétbe helyezik) a rendszerüket használó együttműködő könyvtárak rekordjait a betöltések teszteléséhez, majd később tevékenyen részt vesznek a szükséges új funkciók implementálásában. Ez a feltétel megelőzné, hogy az IKR fejlesztők között fennálló feszültségek, az erős piaci verseny miatt akár meg is érthető érdekellentétek – melyek a TÁMOP keretében készülő portálok előkészítése kapcsán már a felszínre törtek – a MOKKA megvalósításának rovására menjenek.

A harmadik tényező az értékelés nehézségeivel függ össze. Mint már említettük, ez az 1997-es pályázat esetében is gondot okozott. Akkor az ár 70%-os súllyal vett részt a pályázatok pontozásában, ami a szűkös anyagi források miatt érthető volt. A mostani helyzet egészen más. Az OSZK a TÁMOP keretében feltehetően elegendő, s egyben fix forrást nyert el a MOKKA megújítására. Véleményünk szerint a rendelkezésre álló összeg nagyságát nyilvánosságra kell hozni, és ezzel – örömteli, hogy erre már van utalás a projektvezetők előadásaiban – el kell kerülni az árversenyt. Talán helyes lett volna a TÁMOP-hoz szükséges árajánlatkérés tartalmát is nyilvánosságra hozni, vagy legalábbis minden érdekeltnek eljuttatni, mivel e nélkül az árajánlatot adó cégek látszólag helyzeti előnyhöz jutottak.

Érdekes megoldás lenne, ha nem a cégek „szándéknyilatkozatait” bírálná a közbeszerzési bizottság – mint azt tette részben 1997-ben –, hanem kétlépcsős pályázatot írnának ki. Az első lépcsőben a szakértők által készített és a szakmával

megvitatott anyagok alapján a pályázók egy viszonylag kis összeget – 1-2 millió Ft-ot – kapnának egy prototípus készítésére, amely bemutatná a pályázó működő megoldását. Ezzel egy időben, tehát a munka elkezdésekor, a második lépés pályázati anyagát is be kéne adniuk, így kerülve el azt, hogy később mások prototípusaiból „tanulva” finomítsanak megoldásukon. A különböző tervek a nagy nyilvánosság előtt is megmérkőzhetnének, segítve ezzel a felelős döntést. Mindegyik résztvevőnek meghatározott összeg állna rendelkezésére a prototípus fejlesztésére, amelyet mindenképpen megkapnának a határidő lejártakor. A második lépésben e prototípusok, az első lépéskor leadott dokumentáció alapján, természetesen a kiírás egyéb értékelési szempontjaival egyetemben, megalapoznák a pályázók reális és biztonságos megítélését. Ez a nyitott és teljesen átlátható megoldás kiküszöbölné az egyébként óhatatlanul felmerülő jól értesült suttogásokat, pletykákat a „várható” fejlesztőkről, nyertesekről.

Zárszó

Reméljük, elképzeléseink vázlatos ismertetése is élénkíti majd a MOKKA körüli párbeszédet, a változás szükségességéről és módjairól való gondolkodást, és ezzel a könyvtári rendszer épülését szolgálja majd – hiszen mi más lehetne mindannyiunk érdeke? Konzorciumunk tagjai, többek között a Szikla-21 rendszert fejlesztő NetLib és az UTCA eredeti csapata, minden igyekezetünkkel ezen leszünk.

Irodalom és jegyzetek

- 1 Univerzális Tartalomfeltáró és Csoportosító Alkalmazás
- 2 VAJDA, Erik: Közös (osztott) katalogizálás – közös (központi) katalógus. (Terminológia, tipológia, stratégia.) = Könyv, Könyvtár, Könyvtáros, 9. köt. 2. sz. 2000. p. 28–39.
http://www.ki.oszk.hu/3k/19972006/valcikkek/valcikk_ek0002/vajda_e.html
- 3 TÓTH Kornél: Mi újság a MOKKA háza táján? Az IMOLA (Integrált MOKKA, ODR, OLA) koncepció. = TMT, 56. köt. 8. sz. 2009. p. 351–370.,
<http://konyvtar.hu/wiki/IMOLA>
- 4 KARDOS András: UTCA: egy fejlesztésben lévő „közösségi katalógus”. = TMT, 54. köt. 7. sz. 2007. p. 329–333.
http://tmt.omikk.bme.hu/show_news.html?id=4751&is_sue_id=484
- 5 <http://www.ki.oszk.hu/kf/kfarchiv/2003/1/ungvary.html>

- ⁶ <https://listserv.niif.hu/pipermail/katalist/2003-July/004766.html>
- ⁷ Közös katalogizálás Magyarországon. Sopron, Nyugat-magyarországi Egyetem, 2007. 85 p. ISBN 963-9364-60-6
- ⁸ Aleph, e-Corvina, HunTéka, LibriVision/Amicus, Szikla, Szirén, TextLib, SLIB, OLIB, továbbiak előkészítés alatt.
- ⁹ <http://wiki.mokka.hu>
- ¹⁰ Fejlett Információs Technológiák és Társadalom – <http://fitt.klog.hu>
- ¹¹ KOLTAY Klára: Mi újság a MOKKA háza táján? 2. A MOKKA és a tagkönyvtárak. = TMT, 55. köt. 10. sz. 2008. p. 455–460.
- ¹² Univerzális Tartalomfeltáró és Csoportosító Alkalmazás.
- ¹³ <http://www.loc.gov/cds/FRBR.html>
- ¹⁴ [http://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)).
- ¹⁵ A <http://www.mokka.hu> oldalon elindulása, 2001. óta szerepelt eme hivatkozás.
- ¹⁶ Networkshop: 2007, 2008. MKE Vándorgyűlés: 2008, Közgyűjteményi Napok: 2007.
- ¹⁷ A Föld szinte valamennyi nyelvének karaktereit tárolni képes szabvány, l. <http://hu.wikipedia.org/wiki/Unicode>
- ¹⁸ <http://www.mokka.hu/wiki/Duplumkulcs>
- ¹⁹ http://en.wikipedia.org/wiki/The_Long_Tail, illetve http://www.kithirlevel.hu/index.php?kh=the_long_tail_hogyan_magyaritsuk
- ²⁰ <http://www.daveyp.com/blog/archives/1317>
- ²¹ Apple Computer Inc., illetve Addison-Wesley Publishing Company: <http://portal.acm.org/citation.cfm?id=573097>
- ²² <http://konyvtar.hu>
- ²³ 2009. június 23., OSZK – meghívó: <https://listserv.niif.hu/pipermail/katalist/2009-June/018639.html>
- ²⁴ A Z39.50-hez hasonló funkciójú, rekordok átvitelére alkalmas, egyszerű kommunikációs protokoll, mely a MOKKA mögötti Corvina rendszer része.
- ²⁵ ILS Discovery Interfaces – <http://www.diglib.org/architectures/ilsdi/>
- ²⁶ <http://www.mokka.hu/?q=mokka/odr>

Beérkezett: 2009. XI. 25-én.



Menekülnek a szerzők a Wikipédiáról

Egyre több ügy bizonyítja, hogy közel sincs minden rendben a népszerű online lexikon körül. A jelenségről a *The Wall Street Journal* számolt be, és a gazdasági lap két videót is megjelentetett a témában. Ezek szerint szabályosan menekülnek a szerzők a Wikipédiáról. Csak az 2009 első negyedévében több mint 49 000 szerző fordított hátat az online lexikonnak. Összehasonlításként: 2008 elején csak 4900 felhasználó döntött hasonlóképpen.

A *Wikimedia Foundation* illetékesei a hírt kénytelenek voltak megerősíteni a *The Wall Street Journal* munkatársainak, ugyanakkor közölték, hogy a jelenség főleg az angol nyelvű bejegyzések készítői körében figyelhető meg. Bármennyire is aggasztó a helyzet, a Wikimédiánál egyáltalán nincs pánik, és hallani sem akarnak a szolgáltatás megszűnéséről vagy szüneteltetéséről.

Az okokról egyelőre megoszlanak a vélemények. Sok szakértő szerint elsősorban az áll a jelenség hátterében, hogy egyre kevesebb érdekes témáról lehet írni. A mind több előírás és publikálási feltétel szintén hozzájárult a helyzet kialakulásához.

/SG.hu Hírlevél, 2009. november 25., <http://www.sg.hu>

(SzP)