

Balázs László

## Virtuális közös lekérdezés vagy valós központi adatbázis

### A MOKKA és a KözEIKat technikai rendszerének összehasonlítása

*Arra, hogy több adatbázisban egyszerre kereshessünk, két fő megoldás létezik: a közös keresőrendszer és a központi adatbázis. A közös keresőrendszerek olyan brókert alkalmazzanak, amely közös felhasználói interfészt bocsát a felhasználók rendelkezésére; ezen keresztül tehetik fel keresőkérdéseiket, és a találatokat is ebben a rendszerben kapják meg. A közös keresőrendszer a rekordokat nem tárolja, a keresést továbbítja az adatbázis-szolgáltatók felé, a rekordokat összegyűjti, rangsorolja és megjeleníti. Ilyen rendszer volt például az NIF KözEIKat rendszere a 90-es években. A központi adatbázisrendszereknek saját adatbázisuk van, ahová fizikailag is összegyűjtik az adatbázis-szolgáltatók rekordjait. A keresések a központi adatbázisban történnek, és a rekordok megjelenítésére is onnan kerül sor, bár a rekordok általában megőrzik kapcsolatukat a forrásul szolgáló adatbázissal. A MOKKA is ebbe a csoportba tartozik.*

Az utóbbi időben többször felmerült szakmai fórumokon, hogy jobb lenne a MOKKA jelenlegi megoldása helyett egy közös keresőt készíteni, ezért úgy gondolom, hasznos lehet a közös kereső és a központi adatbázisrendszerek összehasonlítása. Mivel részt vettem a KözEIKat<sup>1</sup>, a Vocal<sup>2</sup>, az ODR<sup>3</sup> és a MOKKA<sup>4</sup> informatikai rendszerének tervezésében, megvalósításában és üzemeltetésében, úgy vélem, kellő tapasztalattal rendelkezem e témában. Erre építve a cikkben összehasonlítom a rendszerek bonyolultságát, az adatbázis-szerverek felé közvetített terhelést, a keresési és a megjelenítési lehetőségeket, az authority kontroll lehetőségeit, a könyvtárakra gyakorolt hatást és kitérek a rendszerekkel elérhető felhasználói csoportokra is.

#### **A rendszerek bonyolultsága, karbantartási igénye**

A rendszerek megvalósíthatóságában a közös keresők a jobbak. Egyszerűbb ilyen rendszereket létrehozni, ezért olcsóbbak is. A megvalósítás könnyebb, gyorsabb és léteznek szabad forrású rendszerek. A KözEIKat indulásának idején még nem volt széles körben használt szabványos interfész a magyar könyvtárak integrált rendszereiben, illetve adatbázis-alkalmazásaiban, ezért akkor egy egyszerű, nem szabványos megoldást kellett megvalósítanunk. Ma már ez nem probléma, a legtöbb

online könyvtári katalógus Z39.50-es interfészen keresztül is lekérdezhető.

Ha azonban azt szeretnénk, hogy a közös keresőrendszer jól működjön, akkor igen jelentős fejlesztéseket kell elvégezni. Az egyes rendszerek esetében más-más keresőkérdést kell küldenünk a különböző katalogizálási gyakorlat és a könyvtári rendszer eltérő lehetőségei miatt (l. bővebben a keresési lehetőségekkel foglalkozó fejezetben). A MOKKA eddigi tapasztalatai azt mutatják, hogy idővel a katalogizálási gyakorlat és a használt rendszer is változik a könyvtárakban, ezért ennek követése jelentős folyamatos rendszerkarbantartást, esetleg folyamatos fejlesztést igényel a közös keresőrendszerekben is. Szintén a tapasztalatok alapján látszik, hogy nem működik az a modell, hogy a könyvtárak a közös katalogizálási gyakorlatnak megfelelő rekordokat küldenek.

Ennek alapján állíthatjuk, hogy egyetlen közös keresőrendszerben sem küldenének ilyen rekordokat, és nem egységesítenék a keresési lehetőségeiket. Ez pedig azt mutatja, hogy mindenképpen szükséges a közös keresőrendszerben is az a drágán implementálható és menedzselhető rendszer, amely a kérdéseket lefordítja az adatbázis-szerverek nyelvére, és a rekordok átalakítását is elvégzi. A KözEIKatban nem volt ilyen rendszer, ezért nem az olvasók, hanem inkább a könyvtár-

rosok használták, akik meg tudták tanulni a keresési trükköket is.

Ha központi adatbázist használunk, a rekordokat akkor is ellenőrizni kell, javítani, egységesíteni, és visszajelzést küldeni a könyvtáraknak. Ebben az esetben viszont erre több idő áll rendelkezésre, és a hibás rekordokat a rendszer üzemeltetői ki is zárhatják, hiszen ekkor a rekord felküldőjének lehetősége van a javításra, és így nem veszít találatot a felhasználó. Ha ilyen rendszert használ a központi rendszer, akkor kialakulhat egy közös katalógizálási gyakorlat. A MOKKA jelenleg ilyen rendszert használ, és megfigyelhető a közeledés a könyvtárak között.

### **Az adatbázisszerverek felé közvetített terhelés**

A közös keresőrendszerek az összes adatbázis-hoz továbbítják a kéréseket, így az összes keresés minden adatbázisszerveren megjelenik. A MOKKA napi terhelése 16-18 ezer keresés és időben erősen változik, az ODR szerver terhelése<sup>5</sup> 8-10 ezer keresés, ami szintén erősen változik. A keresések napon belüli eloszlása sem egyenletes, 60%-uk 10 és 16 óra között történik. Ezek a keresések egy közös keresőrendszerben minden adatbázisban végrehajtnának, jelentős többletterhelést okozva az adatbázisok kiszolgálóin<sup>6</sup>. Kisebb könyvtáraknál ez többszöröse lehet a helyi keresések számának, ami akár használhatatlanná is teheti az adatbázisszerveret, de a megfelelő szintű szolgáltatási színvonal kialakítása még nagy könyvtárakban is jelentős többletterheléseket igényel. Egy-egy különleges terhelést kiváltó esemény az egész hazai könyvtári rendszer leállítását vagy jelentős lassulását, esetleg a központi keresőszolgáltatás felfüggesztését okozhatja. A MOKKA indulásakor akkora volt a forgalom, hogy a szerveren le kellett állítani minden egyéb tevékenységet, hogy a keresések lehetőségét megoldjuk, de még így is a használhatatlanságig lelassult a keresés. Képzelnék el, ha ezt minden könyvtárhoz közvetítjük, vagy ha a könyvtárak leállítják a szolgáltatást a keresőrendszer felé. A szolgáltatás korlátozását nyilván a nagyobbak meg is tették volna, a kicsik pedig valószínűleg nem is értették volna, mi történik. Milyen visszhangja lenne ennek a felhasználók körében? Hogy ez az eset nem egyedi, elég az *Europeana*<sup>7</sup> indulására gondolnunk. Az egy központi adatbázist használó rendszer, kezdetben nem bírta a terhelést, de hamar tudtak olyan hardvert rendszerbe állítani, amely már ki tudja szolgálni az igényeket.

Ha a rendszer egy közös keresőrendszer lenne, akkor az összes partnernek új hardvert kellett volna beszereznie, ami valószínűleg nem történt volna meg ilyen hamar.

Tehát lehet, hogy a közös keresőrendszer készítése olcsóbb, mint egy központi adatbázisrendszer kialakítása, azonban az adatbázisszervereken szükséges beruházások ezt az előnyt kérdésessé teszik.

### **A keresési lehetőségek**

#### **A közös keresők előnyei**

- Az adatbázisokba bekerült rekordok azonnal kereshetők a közös keresővel, nincs késleltetés a közös adatbázisba kerülésig.

#### **A közös keresők hátrányai**

- Az adatbázisok adattartalma erősen eltérhet. Például: nincs tárgyszó a rekordokban.
- Az adatbázisok indexkészlete eltérő. Például: egyes adatbázisokban nem lehet tárgyszóra keresni, bár van ilyen adatelem az adatbázisban.
- Az adatbázisok indexelési stratégiája eltérő. Azonos nevű indexekben más adatelemek találhatóak. Például: a címindexben nem szerepel a sorozati cím.
- Nincs közös besorolási adatokat tartalmazó fájl, így a besorolási rekordok utalóinak használata nem lehetséges, vagy csak több egymás utáni keresésre bontva.
- A különböző rendszerek eltérően kezelhetik a szavas kereséseket. Egyes rendszerek, ha két szót írunk be, akkor azt két külön szónak tekintik, míg mások egy kifejezésnek. Ennek a hatását tompítani lehet a találatok megfelelő rendezésével, de teljesen eltüntetni nem.
- A különböző rendszerek eltérően kezelhetik a csonkolást.

Az említett különbségek azt okozhatják, hogy a keresés néhány könyvtárban sikeres lesz, másoknál nem, pedig ott is lenne megfelelő rekord, csak másképpen kellene keresni. Erről a félsikerről a felhasználó nem értesül, hiszen vannak találatok. Joggal gondolhatja, hogy megfelelően keresett, mert ha nem így lett volna, nem lett volna találat. Emiatt a felhasználó nem keres tovább, és így esetleg nem találja meg a keresett dokumentumot, vagy csak egy másik, távolabbi könyvtárban. Ez a félrevezető helyzet annál gyakrabban fordul elő, minél nagyobbak az eltérések az adatbázisok között.

A fenti hátrányok egy része erős és jól konfigurált közös keresővel kiküszöbölhető. Például: az indexelési stratégia különbsége esetleg elkerülhető több indexben történő egyidejű kereséssel. Ez viszont annyira bonyolult, és olyan kevés esetben kivitelezhető, hogy elhanyagolható hatású.

### A központi adatbázisok előnyei

1. Az azonosan indexelt rekordok egyszerűbben, könnyebben kereshetők. Nő az esély a rekordok megtalálására. Csökken a karbantartási igény, mert nem kell figyelni a helyi rendszerek indexelési stratégiájára.
2. A közös adatbázisba kerüléskor ellenőrizhető az adattartalom megfelelősége, a nem megfelelő rekordokat el lehet utasítani. Ez hosszabb távon a katalogizálási gyakorlat egységesedéséhez vezet.
3. Közös besorolási rekordokat használhatunk, ezzel megkönnyítve a keresést és a böngészést.

### A megjelenítési lehetőségek

A több adatbázist átfogó keresések szükségképpen duplumokat eredményeznek a találatok között. Ennek a megoldására közös adatbázis esetén a feltöltéskor is kínálkozik lehetőség. Ekkor elvégezhető a duplumszűrés és a rekordok összeolvasztása, ami (a rekordok minőségétől függően) jórészt eltünteti a duplumokat a találati halmazokból. Ekkor lehetőségünk van a beérkező rekordot az adatbázis összes rekordjával összehasonlítani, így megtalálhatjuk az összes duplumot. Közös kereső esetében erre nincs lehetőség, csak a találati halmazokban lehet duplumellenőrzést végezni. Ekkor azonban (a keresési stratégiától függően) lehetséges, hogy a találati halmazban nincs benne minden duplum, és így elveszítünk találatokat.

Egyre több adatbázisban van lehetőség az FRBR<sup>8</sup> szerinti megjelenítési formára. Ez a jövőben elvárható lesz az országos szolgáltatásoktól is. Az FRBR szerinti megjelenítés két módon képzelhető el:

1. *A találati halmazban lévő rekordokat rendezzük össze, és azokon nyújtunk FRBR szerinti megjelenítést.* Ekkor természetesen csak a halmazban lévő rekordokat csoportosíthatjuk FRBR formába. A halmazban lehetnek az absztrakt művek egyes megjelenési formái, de (a keresési stratégiától függően) nem biztos, hogy mind-

egyik ott van. Így pedig az egyes rekordok FRBR szerinti megjelenítése a keresési stratégiától függ majd. Ez a megoldás kevésbé jó, de mindkét rendszerben megvalósítható.

2. *A közös adatbázisban eltárolunk absztrakt rekordokat, amelyek a műről szólnak, és nem kötődnek a konkrét megjelenési formához.* A konkrét rekordok kapcsolódnak az absztrakt rekordokhoz, így a keresések az absztrakt rekordokon történnek, és csak a találati listába rendezéskor kapcsoljuk hozzá a művek konkrét megjelenési formáit. Erre nyilván csak a közös adatbázis lehet képes. A közös adatbázisban a feltöltéskor futó duplumellenőrzés kialakíthat absztrakt rekordokat, amelyeket kézzel javítva juthatunk egy olyan adatbázishoz, amely egy lényegesen pontosabb, a nemzetközi trendeknek megfelelő FRBR szerinti megjelenítést tesz lehetővé.

Tehát mind a duplumellenőrzés, mind az FRBR szerinti megjelenítés pontosabb lehet, ha központi adatbázist használunk.

### Az authority kontroll lehetőségei

Közös kereső esetében nincs lehetőség közös authority kontrollra. Közös besorolási rekordokat használhatnak, de az informatikai rendszer ezt nem segíti.

A központi adatbázisban ez a lehetőség adott, és jelentős munkát spórolhat meg az ezt használó könyvtár. A közös authority kontroll azt jelenti, hogy a közös adatbázisban egységes besorolási rekordokat használhatunk, azok a bibliográfiai rekordokhoz kapcsolódnak, és együtt, vagy külön is átvehetők helyi rendszerekbe. A központi adatbázis lehet az egységes besorolási rekordok elsődleges forrása. A közös besorolási rekordok a központi adatbázisban is karbantarthatók, illetve a javított rekordok felkerülhetnek a helyi rendszerekből. Az egyre jobb rekordok automatikusan áttölthetők a helyi rendszerekbe. Például: ha egy névtípusú besorolási rekord az egyik rendszerben kiegészül a szerző születési és halálozási idejével, a központi adatbázis összes kapcsolódó bibliográfiai rekordja kiegészül ezekkel, illetve a könyvtárak döntése szerint, a más helyi rendszerekbe is automatikusan megtörténhet ez a javítás. Illetve, ha a központi rendszerben aktualizáljuk a *Köztaurusz* rekordjait, azok automatikusan aktualizálhatják a helyi rendszerekben tárolt *Köztaurusz*-rekordokat, és a helyi bibliográfiai rekordokat is. De ha az au-

tomatizmus nem elfogadható a könyvtárban, akkor kézzel mindenképpen átvehető az új rekordok. A közös adatbázisban a tárgyszórendszerek egy listában böngészhetővé válnak, a különböző tárgyszórendszerek között megfeleltetés készíthető. Ekkor a közös keresőben olyan tárgyszó alapján is megtalálható lehet egy rekord, amely nincs is benne a bibliográfiai rekordban. Például: a Köztaurusz alapján is megtalálható egy csak LCSH<sup>9</sup> szerint tárgyszavazott rekord (ha létezik megfeleltetés). Vagy csak ETO-t tartalmazó rekord is megtalálható a Köztaurusz alapján.

Még nagyon sok lehetőség elképzelhető, amelyek azonban erősen függenek a rendszer megvalósításától; ha jó központi adatbázist használunk, jelentősen nőnek a lehetőségeink.

Az viszont már ennyiből is látszik, hogy egy központi adatbázis erős authority kontrollal jelentősen egyszerűsítheti a helyi rendszerekben a katalogizáló munkát, illetve jelentős megtakarítások érhetőek el úgy, hogy a rekordok minősége is javul.

### A könyvtárakra gyakorolt hatás

A KözEIKat tervezésekor derült ki számunkra, hogy az egyes könyvtárakban mennyire különböznek a keresési lehetőségek és az adattartalom. Nagyon kevés adatelemre lehetett keresni minden könyvtárban. Ezért maradt a szerző, a cím és a tárgyszó szerinti keresés, bár tárgyszó sem volt minden részt vevő könyvtárban. Az első fontos tanulság az volt, hogy még két azonos rendszert használó könyvtár sem tudta átvenni egymás rekordját a helyi katalogizálási szokások eltérő volta miatt. A rekordok átírása több időt vett igénybe, mint egy eredeti leírás elkészítése. A résztvevők számára világossá vált, hogy az együttműködés gátja az eltérő katalogizálási gyakorlat. A rendszer viszont nem adott lehetőséget a rekordok ellenőrzésére, az eltérő gyakorlat jelzésére, és persze szervezet sem volt, amely közös gyakorlatot alakíthatott volna ki. A MOKKA indulásakor ezekre a tapasztalatokra is építve terveztük meg a rekordok átalakítására szolgáló rendszert. Ez a rendszer képes a hibás rekordok kizárására, a kisebb hibák jelzésére, ezzel erősítve a közös gyakorlat kialakulását. Míg kezdetben egyes könyvtáraknak gondot okozott a formailag helyes MARC rekordok szolgáltatása, ma már ez senkinek nem jelent nehézséget, és a rekordok adattartalmában sincs akkora különbség, mint a kezdetekkor. A szintaktikai ellenőrző folyamatos szigorítása lehetővé teszi az

adattartalom folyamatos és egyre magasabb fokú egységesítését.

### A célközönség

A KözEIKat rendszert eredetileg az olvasóknak terveztük, majd kiderült, hogy a keresés az eddig leírtak miatt mégsem annyira egyszerű, így inkább a könyvtárosok kezdték használni a könyvtárközi kérések lelőhely-információinak megállapítására. A már felsorolt hiányosságok miatt a közös katalogizálási használatra tett kísérletek nem voltak sikeresek. A legtovább és legaktívabban a profi keresők, a könyvtárközi keresést végző könyvtárosok használták. Ők megtanulták a lehetőségeket, és a saját keresési stratégiájukkal tudták pótolni a rendszer hiányosságait. Ennek a tudásnak egy része beépíthető a közös keresési rendszerbe, de ebben az esetben a közös kereső legfontosabb előnye az egyszerűbb implementálhatóság vesz el.

### Nemzetközi trendek a keresésben

- A *The European Library*<sup>10</sup> első változatában közös keresőnek indult, ma már létezik központi adatbázis is, igaz megtartották a közös keresés lehetőségét is.
- Az *EBSCO*<sup>11</sup> 2009-ben mutatta be közös keresőjének, az *EBSCO Hostnak*<sup>12</sup> béta-változatát, ami túllép a korábbi közös keresőn és egy hibrid megoldást valósít meg. Ha lehetséges, beépíti a rekordokat egy közös adatbázisba, és csak azokat kérdezi online, ahol a beépítés nem lehetséges.
- A nagy könyvtári rendszerszállítók a könyvtári federated search<sup>13</sup> rendszerek megvalósítására is a hibrid rendszereket javasolják, és csak azt kérdezik online, amit nem lehet egy közös adatbázisba illeszteni.
- Az open source federated search<sup>14</sup> rendszerek is összegyűjtik a rekordokat és a szoftveren belül oldják meg a keresést.

### Összefoglalva, a rendszerek erősségei

#### A központi adatbázisrendszereknél:

- Könnyebb és hatékonyabb keresés.
- Az FRBR és a duplumellenőrzés lehetőségei lényegesen jobbak.
- Elősegíti az egységes katalogizálási gyakorlat kialakulását.

- A megfelelően kialakított központi adatbázisra országos szolgáltatások együttműködése épülhet. Például az ODR rendszer is feltételez központi bibliográfiai adatbázist.
- A nemzetközi kapcsolatokban is fontos szerepe lehet, például a TEL jelenleg is a MOKKA adatbázisát kérdezi le, és az Europeana számára is megfelelő lesz.
- A központi adatbázis is tárolja a könyvtárak rekordjait, így azok onnan is visszaállíthatók, ha szükséges.
- Az authority kontroll jelentősen megkönnyítheti a katalogizáló munkát.

#### A közös keresőknél:

- Olcsóbban implementálhatók (ez az előny hamar elvész, ha jól használható rendszert akarunk építeni).
- Nagyon sok adatbázis kérdezhető le egyszerre.

Véleményem szerint közös katalogizálás megvalósítására, elsődleges rekordforrásnak a központi adatbázis a megfelelő, míg forrás- és lelőhelykeresés esetén a közös kereső is lehet jó választás.

#### Megjegyzések

- <sup>1</sup> Közös Elektronikus Katalógus, az NIIF projektje volt 1996-ban, sajnos ma már nem működik. A rendszert ismertető előadás az 1997-es Networkshopon hangzott el:  
<http://www.niif.hu/rendezvenyek/networkshop/97/tartalom/NWS/3/2/index.htm>
- <sup>2</sup> A Vocal a Corvina rendszert használó könyvtárak közös katalógusa, a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár üzemelteti 1998-tól. Az ODR ezt használja bibliográfiai adatbázisnak, ebben hajtódnak végre a keresések.  
<http://vocal.lib.unideb.hu>
- <sup>3</sup> Országos Dokumentumellátó Rendszert, a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár üzemelteti 2001-től. <http://odr.lib.unideb.hu/>
- <sup>4</sup> Magyar Országos Közös Katalógus,

<http://www.mokka.hu>

- <sup>5</sup> Forrás: Az ODR rendszer statisztikai oldala,  
<http://odr.lib.klte.hu/stat/>
- <sup>6</sup> A könyvtárak adatbázisszerverei, amelyeken könyvtári rendszerek futnak.
- <sup>7</sup> Europeana, <http://www.europeana.eu/portal/>
- <sup>8</sup> FRBR, <http://www.oszk.hu/hun/szakmai/frbr/frbr.pdf>
- <sup>9</sup> LCSH, Library of Congress Subject Headings,  
<http://www.loc.gov>
- <sup>10</sup> The European Library (TEL):  
<http://search.theeuropeanlibrary.org/portal/hu/index.html>
- <sup>11</sup> EBSCO, <http://www.ebsco.com/home/>
- <sup>12</sup> EBSCO HOST: <http://search.ebscohost.com/>
- <sup>13</sup> Ilyen rendszer például az ExLibris csoport Primo szoftvere:  
<http://www.exlibrisgroup.com/?catid={6FA08552-67DA-4192-8E77-8BBE75241395}>
- <sup>14</sup> Ilyen rendszer például a vufind: <http://vufind.org/>

#### Irodalom

A bibliográfiai tételek funkcionális követelményei (Functional Requirements for Bibliographic Records). = <http://www.oszk.hu/hun/szakmai/frbr/frbr.pdf>

COUSINS, Shirley: Virtual OPACs versus union database: two models of union catalogue provision. = The Electronic Library, 17. köt. 2. sz. 1999. p. 97–103.

KOLTAY Klára: Mi újság a MOKKA háza táján? 2. A MOKKA és a tagkönyvtárak. = TMT, 55. köt. 10. sz. 2008. p. 455–460.

Beérkezett: 2010. I. 29-én.

