

Miskolci Egyetemi Publikációs Adatbázis

*Kiss Andrea, konpinty@uni-miskolc.hu
Miskolci Egyetem, Könyvtár, Levéltár, Múzeum*

*Vitéz Gáborné, ildiko@swszl.szkp.uni-miskolc.hu
Miskolci Egyetem, Egyetemi Számítóközpont*

1. Igények egy új szolgáltatásra

Minden intézményben, ahol tudományos munka vagy kutatás folyik, születnek értékes eredmények, melyeket az alkotók, a szerzők közkinccsé kívánnak tenni és valamilyen formában megjelentetnek. Ezek a publikációk lehetnek folyóiratcikkek, konferencia-előadások, tanulmányok, disszertációk, kutatási jelentések, beszámolók, szabadalmak, szabványok és sok más is. Napjainkban egyre több munka nem nyomtatott, hanem elektronikus formában jelenik meg, és a publikációs listákba bekerülnek az online módon hozzáférhető írások mellett szoftverfejlesztések, interaktív oktatócsomagok, egyéb digitális kiadványok is. Mindezen publikációk mögött pedig többnyire jelentős tudásanyag és munka áll.

Fontosak ezek a publikációk mind az egyén, mind pedig azon intézmény szempontjából, ahol az eredmények megszületnek. Fontos az egyén szempontjából, hogy minél több és minél értékesebb publikációt jelentessen meg, hisz ez számára nem csak munkájának, eredményeinek elismerését jelenti, hanem ma már elengedhetetlen feltétele a szakmai előrelépésnek is. A fent említett intézményekben ma már egy új munkatárs kiválasztásakor, vagy vezetői kinevezéskor nagy súllyal esik latba, hogy a pályázó milyen hosszú és milyen értékes publikációkat felsoroló listát tud csatolni önéletrajzához és pályázati anyagához. Az intézményeknek is nagy súlyt kell fektetniük arra, hogy minél több tudományos eredmény lásson napvilágot munkatársaik révén, hiszen ez az adott intézmény súlyát, jelentőségét, elismertségét növeli, és nem elhanyagolható szempont ez akkor sem, amikor támogatásokról, pályázati forrásokról, esetleg felsőoktatási akkreditációról van szó. Ma már a kutatások, fejlesztések is leginkább csoportmunkában történnek és támaszkodnak az adott területen elért legfrissebb eredményekre. Ezeket az információkat és egy kutatáshoz lehetséges partnereket a témában megjelenő publikációkban lehet leginkább megtalálni.

De hogyan juthatunk hozzá ezekhez a publikációkhoz? Hol találhatjuk meg egy adott személy összes publikációját? Hol az a hely, ahol egy intézmény összegyűjti, tárolja, naprakészen frissíti és nyilvánosan hozzáférhetővé teszi a falai között született publikációkat vagy legalább azokat a bibliográfiai adatokat, melyek alapján hozzá lehet jutni az eredeti szövegekhez? Van-e olyan formájú megoldás, ahol nem csak megnézhetjük az egyes publikációk adatait, esetleg szövegét, hanem több szempont szerint kereshetünk is a felgyűlt anyagban?

Igény mindenképp van erre, ezt bizonyítja az a felkérés is, amely 2004 végén érkezett az Miskolci Egyetem vezetésétől az Egyetemi Számítóközpontba. A feladat egy központi adatbázis létrehozása volt az egyetemi publikációk számára, amely adatokat szolgáltat a különböző egyetemi beszámolók számára. A munka elvégzéséhez - mivel egy bibliográfiai adatokat tároló adatbázisról van szó - a Számítóközpont az Egyetemi Könyvtár segítségét kérte.

A közös munka első fázisa az volt, hogy utánanéztünk a már meglévő itthoni és külföldi megoldásoknak.

2. Célok és tervek

Sajnos a magyar gyakorlat meglehetősen szegényes képet mutat. Személyes honlapok mélyére dugott Word dokumentumokban, itt-ott egy intézmény honlapján összegyűjtve, szintén Word vagy statikus HTML formában lehet publikációs adatokat találni, elvéve linkkel a teljes szövegre.

A frissítések gyakorisága többnyire nem megállapítható, az adatok megjelenítése nem egységes, gyakran önkényes rövidítéseket és sorrendet alkalmaznak a publikációs lista készítői. Sok a pontatlan, hiányos adat, előfordul, hogy a rendelkezésre bocsátott információkból meg sem állapítható pontosan, hol és mikor jelent meg az adott írás.

A listák szerkesztési elve változó: hol megjelenési év szerint, hol dokumentumtípus szerint sorolják fel a publikációkat, de előfordul a cím vagy a szerzők neve szerinti sorrend is. Téma szerinti csoportosítással nem találkoztunk, viszont néhol nem csak a publikációkat, hanem a rájuk való hivatkozásokat vagy a hivatkozási indexet is feltüntetik.

Általános jellemző, hogy egy-egy személy publikációs listája teljesen elkülönül a többitől, így egy adott intézményben vagy szervezeti egységben születő publikációkról összegző képet kapni igen nehéz.

A Miskolci Egyetemi Publikációs Adatbázis készítésekor szándékunkban állt, hogy ezeket a hiányosságokat kiküszöböljük, és egy olyan egységes adatbázist hozunk létre az Egyetemhez kapcsolódó publikációkból, amely

- egy helyen gyűjti össze a szükséges anyagokat
- több szempontú keresést biztosít
- szabványos és egységes megjelenítésre alkalmas
- helyet biztosít nem csak a publikációs adatok, hanem tartalmi kivonatok, mellékletek, teljes szövegek számára is
- a publikációk META-adatait szabványos formában tárolja, lehetővé téve más rendszerekben való felhasználásukat is.

3. Problémák az előkészítés során

Terveink megvalósításához azonban igen sok akadályt kellett leküzdenünk, újra és újra dönteni kényszerültünk egy-egy megoldás előnyeinek és hátrányainak mérlegelésével.

A legnagyobb gondot, magának az adatbázis szerkezetének megtervezése jelentette.

El kellett döntenünk, mire terjedjen ki a felvett adatok köre. Nem elégedhettünk meg azokkal az adatokkal, melyeket a bibliográfiai hivatkozásokkal foglalkozó szabvány megjelöl. Az ugyanis messze nem terjed ki minden olyan dokumentumtípusra, amellyel nekünk dolgunk akadhat, ráadásul a több szempontú keresés biztosításához több adatra is szükség van, mint amit a szabvány kötelezően megadandóként feltüntet. Ilyen például a tárgyszó vagy szerzői kulcsszó, a dokumentum típusa, formátuma.

Különösen fontos lenne a téma szerinti keresés biztosításához, hogy minden publikáció rendelkezzen egy-két, a tartalmat leíró, jellemző kulcsszóval, amit mi az egyszerűség kedvéért tárgyszónak neveztünk. Valójában az lenne a leghatékonyabb, ha maguk a szerzők, a téma és a dokumentum szakavatott ismerői látnák el szerzői kulcsszavakkal publikációikat.

A típus és formátum meghatározásához az NDA által is támogatott qDC metaadatséma ajánlásokat vettük alapul, és a dokumentum tartalma alapján határoztuk meg a típusok körét (cikk, disszertáció, időszaki kiadvány, jegyzet, konferencia-anyag, kutatási jelentés,

monográfia, multimédia-anyag, szabadalom, szabvány, tankönyv), a dokumentum fizikai hordozója alapján pedig a formátumot (pl. nyomtatott, kézirat, CD/DVD, online, video, hangkazetta stb.).

Ezenfelül az egységes formátumú metaadatok létrehozása miatt a létrejövő kimenet, az adatok megjelenítése során kissé el kellett térnünk a szabványtól (a szerzőségi minőség sokfélesége miatt nem szószerint, hanem egységesített formában van megadva), mivel nem akartuk redundáns adatokkal megtölteni az épülő adatbázist.

Célunk volt, hogy a rögzített adatokat és a szolgáltatást minél több szempontból használhatóvá tegyük, ezért a szerzői, közreműködői adatok felvételekor is sokkal bővebb információhalmazt kívántuk rögzíthetővé tenni, mint a puszta név, esetleg év és kiegészítő adatok. Szerettük volna megkönnyíteni a kapcsolatfelvételt a publikációk szerzőivel, ezért külön táblát hoztunk létre az adatbázisban a szerzői adatok nyilvántartásához, ahol az illető e-mail és posta címét, személyes honlapjának URL-jét is képesek vagyunk rögzíteni azon kívül, hogy tagja-e egyetemünknek, s ha igen, melyik szervezeti egységhez tartozik. Ez a fajta tárolási mód viszont megbonyolította a szerzők és az egyes publikációk közötti kapcsolat megteremtését szervezési szinten.

Az egyes publikációk közötti kapcsolatok kidolgozása is problematikusnak bizonyult. Egy szaklapban megjelent cikk esetén még viszonylag egyszerű a cikk és az azt tartalmazó időszaki kiadvány adatait egymáshoz csatolni, de mit tegyünk, ha például egy olyan konferencia-előadásról van szó, melynek kiadványa szekciónként külön kötetekben jelenik meg? A többszörös kapcsolat bonyolultsága miatt végül kénytelenek voltunk azt a megoldást használni, hogy a tartalmazó dokumentum címében tüntetjük fel a közös címet és az adott kötet jelzését és címét is. Emiatt viszont minden egyes kötetet külön fel kell vennünk, ami egyrészt megint redundanciához vezet, másrészt többletmunkát követel. Ezzel egyenes út vezetett ahhoz az igényhez, hogy az egyes rekordok másolhatóak legyenek, így az azonos adatokat ne kelljen többször manuálisan felvenni.

Egy másik jelentős problémát az olyan dokumentumok okozták, mint például az elektronikus dokumentumok, a szabványok, a disszertációk, ahol speciális adatokat is célszerű lenne rögzíteni. Ilyen többlet információ például elektronikus dokumentumnál a fájl típusa, mérete, a verziószám, a működési követelmény. A bibliográfiai leírás kitér ezekre, de a hivatkozási szabvány nem. Nem elég, hogy el kellett döntenünk, milyen adatokat vegyünk fel ezek közül és az adatbázisban megfelelő táblákat és mezőket kellett létrehoznunk, de az adatbeviteli űrlap megjelenítésekor is le kellett programoznunk, hogy a típusnak megfelelő űrlapmezők jelenjenek meg.

A legnagyobb és igazából máig sem teljesen megoldott problémát az Egyetem szervezeti egységei jelentették. Az Egyetem, miközben igyekszik lépést tartani a változó igényekkel, maga is állandóan átalakul. Új szakok indulnak és szűnnek meg mások, s ennek megfelelően az intézményi háttér is változik. Néha csak a neve módosul egy-egy tanszéknek, karnak, vagy intézetnek, de időnként teljesen átalakul vagy máshová kerül a hierarchiában. Sajnos jelenleg nincs az intézményben egyetlen olyan számítógépes nyilvántartás sem, amely naprakészen és folyamatosan, megbízhatóan frissítve tárolná a szervezeti egységek adatait, a meglévő nyilvántartások különböző sorrendben és módon (teljes névvel vagy rövidítve) tartják nyilván a szervezeti egységeket. Igyekeztünk ezek közül a legpontosabbat nyilvántartást kiválasztani, de még így is szükség volt változtatásokra, hogy beépíthessük az adatbázisunkban az ezzel kapcsolatos adatokat.

Sajnos csak a jelenlegi helyzetet tudtuk tükrözni, nincs módunk arra, hogy egy-egy oktató publikációinak felvételekor meg tudjuk határozni, mely tanszék dolgozója volt (vagy egyáltalán az egyetem dolgozója volt-e), amikor az adott publikációt írta. A továbbiakban

viszont az okozhat problémát, ha egy újabb átalakulás, névváltoztatás vagy átszervezés miatt egy adott oktató már egészen más szervezeti egységhez fog tartozni, viszont korábbi publikációi még jelenlegi munkahelyéhez lesznek kapcsolva.

Természetesen ezek az adatok módosíthatóak, de nem látjuk biztosítékát annak, hogy egy átszervezés után értesítést kapjanak az adatbázis gondozói ezekről a változásokról. Így viszont statisztikai adataink egy része (pl. egy adott szervezeti egységen belül hány publikáció jött létre, illetve ezek átlaga munkatársakként) torz képet mutathat.

4. Eredmények

Voltak tehát problémák bőven, melyeket le kellett küzdenünk a tervezés és a programozás során, de végül létrejött egy olyan rendszer, amely képes a különböző típusú adatok befogadására, ugyanakkor egyszerű felületen sokoldalú visszakeresést tesz lehetővé.

Természetesen lehet *böngészni*

- a felvett publikációk címének betűrendes listájában,
- a szerzők és közreműködők nevének betűrendes listájában,
- valamint a szervezeti egységek vagy karok szerint is lekérdezhető a publikációk listája.

Rendelkezésre áll ezen felül az *egyszerű keresés*, amellyel

- a publikációk címében előforduló szavak,
- szerzői, közreműködői nevek,
- tárgyszavak között kereshetünk.

Ugyanezek a lehetőségek rendelkezésre állnak az *összetett keresés*nél is, de itt

- a keresési metódust is beállíthatjuk (and, or kapcsolat, vagy a kifejezés egészének keresése),
- a személynevek mellett testületi nevekben is kereshetünk,
- szervezeti egységek szerint,
- és azonosítók (pl. ISBN, ISSN, szabványszám, stb.) között is megfogalmazhatunk lekérdezést.

Ugyanakkor az összetett keresés minden formájában *szűkíthető* a találati halmaz

- megjelenési év,
- a dokumentum típusa,
- a dokumentum nyelve szerint is.

A találatok kétféle formában jeleníthetjük meg. A *rövid megjelenítés* nagyjából a hivatkozási szabványnak felel meg, míg a *részletes megjelenítés*, melyet minden publikációnál külön kérhetünk, az adatok bővebb halmazát mutatja meg címkés formában. Itt van lehetőség további információkra szert tenni az egyes közreműködőkkel kapcsolatban, pl. a személyes honlapok innen érhetők el.

Ha az adott publikáció elérhető teljes szöveggel is, a link minden esetben megjelenik már a rövid megjelenítés során. Ha több helyről is elérhető a szöveg, a részletes megjelenítésnél a további URL-ek is látszanak.

A teljes szöveget magában az adatbázisban is lehet tárolni (ezt szorgalmazzuk is, mivel így biztosíthatjuk, hogy a felvett URL címek élők is maradjanak), a feltöltés egyszerű webes formában történik.

Már most, a teszt fázisban van olyan dokumentumunk, amely először a mi adatbázisunk révén vált online elérhetővé, kiadóként ezért a Miskolci Egyetemi Publikációs Adatbázis lett feltüntetve.

A rendszer alkalmas arra is, hogy még készülöben lévő publikációk adatait is felvegyük úgy, hogy a megfelelő státusz beállításával jelezzük: ez még nem megjelent dokumentum.

5. Tapasztalatok és távlati célok

A tesztelés során már eddig is sok tapasztalatot gyűjtöttünk és szembesültünk a használói igényekkel.

Ezek közül a legszomorúbb, hogy képtelenek vagyunk az adatokat a legpontosabb forrásból, magukból a dokumentumokból megszerezni. Az oktatók saját maguk által összeállított, többnyire Word-ben leírt listáikat továbbították nekünk. Ezek azonban nem csak pontatlanok, de gyakran olyannyira hiányosak, hogy hosszas nyomozás során sem tudunk megszerezni olyan alapvető adatokat, mint pl. egy tanulmányt tartalmazó kiadvány pontos címe, vagy a konferencia pontos megnevezése, ahol elhangzott egy előadás. Különösen akkor vagyunk gondban, ha csak elhangzott, de meg nem jelent előadásról van szó. Mennyiben tekinthető ez publikációnak, ha nem maradt nyoma - vetődik fel a kérdés a bibliográfus szemszögéből, noha tudjuk, hogy a szakmai elismerés az ilyen előadásokért is joggal jár. Az adatok rögzítése azonban egy olyan rendszerben, amely megjelent publikációkra van tervezve, és kiadót, megjelenési helyet, terjedelmet adatot stb. követelne, ilyenkor még inkább nehézségbe ütközik.

A listákból való adatbevitel másik hátránya, hogy hiányoznak a tárgyszavak, vagy szerzői kulcsszavak és a dokumentum hiányában, pusztán a címből az adatfeldolgozó nem tudja ezt pótolni.

Ezenfelül a listák általában nem tartalmazzák az összefoglalót, a tartalmi kivonatot, sem pedig a teljes szöveg helyét, még ha ez rendelkezésre is áll. Ez utóbbiak nem csak a felhasználók miatt lennének fontosak, hanem azért is, mert szerepelnek az NDA és qDC ajánlásokban, amelyekhez igazodni kívántunk, elsősorban azért, hogy adatainkat az NDA is átvehesse. A tesztadatok küldésére felkért tanszékeknek jeleztük, hogy ilyen adatokat is várunk, erre azonban eddig nem reagáltak. Persze a teljes szövegek esetében a szerzői jogi problémákba is rögtön beleütközünk.

Úgy tűnik, a hivatkozások feltüntetése az egyes publikációk mellett több tanszéken igényként jelenik meg, és el kell gondolkoznunk azon, hogy milyen formában tudnánk hozzákapcsolni a meglévő adatstruktúrához ezeket az információkat.

Az adatbázis gyarapodásával feltétlenül szükség lesz a megjelenítés során az adatok rendezésére. A több oldalas találati listák, a típus és év szerinti rendezés feltétlenül kidolgozásra fog kerülni. Ezenkívül a különböző statisztikák gyors elérését is biztosítanunk kell.

Terveink között szerepel még a bevitel egyszerűsítése (pl. a rekord-másolás lehetőségének biztosítása) is.

A legfontosabb azonban az adatok, még hozzá a pontos és a jelenleginél bővebb adatok (tárgyszavak, absztraktok, teljes szövegek) beszerzése, illetve az adatgyűjtés megszervezését látjuk a legfontosabbnak - ehhez viszont felsőbb szintű segítség szükséges.