

Így dolgozunk mi: Tanfolyamon itthon?

Elképzelni sem tudtam volna, amikor itthon maradásra ítélt minket a járvány, hogy online tanfolyamon fogok részt venni. Márpedig online tanfolyam indult a Könyvtárosok Dolgozószobája Facebook csoportban. Még hozzá nagyon nagy érdeklődés mellett, már az első napon több, mint 100-an vetünk részt a munkában. A kiemelt figyelem nem is csoda, hiszen nagyon időszerű kérdésről hallhatunk, a könyvtári feldolgozás jövőjét meghatározó szemantikus webről. **Hubay Miklós** indított előadássorozatot **Értsük meg a szemantikus webet!** – címmel.

Az előadássorozat célja, hogy megismertesse a hallgatóságot olyan manapság gyakran ismételt fogalmakkal, mint az RDA, RDF, Bibframe, ontológia stb. Célja továbbá megmutatni, hogy miként kapcsolódnak ezek az általunk korábban megszokott könyvtári feldolgozáshoz, azt is, hogy az új módon feldolgozott adatok milyen szerepet játszanak a webtechnológiában, milyen új online szolgáltatások nyújtására adnak lehetőséget. További cél a jó gyakorlatok bemutatása az adatkonverzió vagy az új típusú keresőszolgáltatások területéről. Ezenkívül megismerkedhetünk a témához kapcsolódó publikációkkal. Ez azért nagyon izgalmas, mert a téma gazdag angol nyelven megjelent irodalma mellett kevés a magyar nyelvű dokumentáció. Az első részben az **entitás** fogalmával ismerkedhetünk meg, s azzal hogyan kapcsolódik az entitás fogalma a könyvtári feldolgozó munkához.

Mi az az entitás, és hogy jön a könyvtári feldolgozó munkához?

Mióta beszélünk entitásokról a szakmában? – 1997 az FRBR (Functional Requirements for Bibliographic Records) **Bibliográfiai tételek funkcionális követelményei** megjelenése óta, amely dokumentum magyar nyelven *Berke Barnabás* né fordításában olvasható az OSZK honlapján.

<http://www.oszk.hu/sites/default/files/frbr.pdf>

Az FRBR az IFLA (International Federation of Library Associations and Institutions) Könyvtári Egyesületek és Szervezetek Nemzetközi Szövetsége megbízásából készített szakmai jelentés, amely azt vizsgálja, hogyan kell működnie egy rekordnak, miről kell információt adnia a használónak egy ilyen adatszerkezetnek. Amikor ezt elkezdtek vizsgálni, nem az egyes adatelemekre, adatokra (adatelem pl. a szerzőségi adat, címadat, kiadói adat stb.; adat Jókai Mór, Az aranyember, Kossuth kiadó stb. a leíró megjegyzése) figyeltek, hanem arra, amit az adatokkal leírhatunk, vagyis az **entitásra**. Tehát az entitás olyan valami, amit adatokkal, ismérvekkel lehet jellemezni, függetlenül attól, hogy létezik-e vagy sem, konkrét vagy fogalmi.

Ha megnézzük, hogy a rekordban milyen adatelemek szerepelnek, akkor ezeket a következőképpen csoportosíthatjuk: Vannak olyan adatelemek, amelyek bizonyos szellemi tartalomegységet írnak le. Ide sorolhatók a címadatok, a szerzőségi adat, ill. az osztályozási kifejezés is. Ezek magát a **művet** írják le. Vannak olyan adatelemek, amelyek a tartalom formai kifejezését, azaz a **kifejezési formát** írják le valamilyen formai eszközzel. A tartalom megjelenítésére szolgálhat pl. szöveg, zene, kép, mozgókép, de akár még a tánc is. A kifejezési formáról szól pl. a nyelvi adatelem, a forrástípus megnevezése. A **megjelenési forma**, a megvalósulás leírására szolgáló, a kiadásra, kiadóra, megjelenési helyre, időre vonatkozó adatelemek. A megvalósulás egy konkrét formája a **példány**. A példányadatok közé soroljuk a birtokló intézmény nevét, a vonalkódot, a possessor bejegyzést is. Entitás az a **személy** vagy a **testület**, azaz **ágens**, amely valamilyen módon kapcsolódik a szellemi alkotáshoz, például közreműködött a létrehozásában, róla szól a mű stb. Entitás a műhöz kapcsolódó **földrajzi hely** is, például ahol a mű megjelent, amely helyről szól, vagy ahol valamely ágens élt, működött. A szellemi alkotáshoz kapcsolódó **időtartam** is entitás, például amelyről szól a mű, amikor megjelent, ill. amikor valamely ágens élt vagy működött.

A legfrissebb összegző leírást az entitásokról a 2017-ben megjelent **LRM** (Library Reference Model) Könyvtári referenciamodell adja.

Az LRM megjelenése után 2017-ben frissítették az **RDA** (Resource Description and Access) Forrásleírás és Hozzáférés **katalogizálási szabályzatot**. Ez a szabályzat hagyományokkal szakítva a feldolgozás folyamatát az entításokra építi. Minden egyes entitás esetén megadja, milyen adatelemek tartoznak hozzá. Egyre nagyobb az elégedetlenség a MARC-kal szemben. A MARC csak korlátozottan alkalmas az RDA rendelkezéseit mindenben betartó leírások tárolására, mivel nem teszi lehetővé a fontos entítások adatainak elkülönítését. A bibliográfiai leírás több entitás adatait tartalmazza egyszerre a megjelenési formáét, a műét és a kifejezési formáét. Az entítások adatait elkülönítő bibliográfiai leírások adatait más **adatformatum**-ban kell tárolni. Ennek az adatformatumnak támogatnia kell az **RDF** (Resource Description Framework) logikát, amely háromelemű állítások segítségével történő jellemzés.

Egy hagyományos MARC struktúrában a hívójel szerepelhet **bibliográfiai rekordban**, **besorolási rekordban** vagy **példány rekordban**. A bibliográfiai rekordban szereplő adatelemek típusát hívójellekkel és almező jelekkel azonosítjuk. Végül az adatelemekhez társítjuk az adott műre vonatkozó adatokat, azaz almező értékeket. Az adatelemeket másféle struktúrában is elrendezhetjük, mégpedig annak megfelelően milyen entitást jellemez az adott adatelem (mű, ágens, kifejezési forma, megjelenési forma, példány stb.), az adatelem típusát milyen MARC hívójel és almezőjel azonosítja. Végül milyen almező értéket (adatot) kapcsolunk az adatelemhez. Ily módon háromelemű állításokat kapunk. Ezekből az állításokból összeáll egy lista, amelyben az entítások adatai különválogatva szerepelnek. Nincs szükség különálló rekordokra, az összes entitás jellemzése egyetlen szerkezetben történik, egy gráf segítségével. Az adatelemek szervezésének elve nem a rekord (bibliográfiai, besorolási, példány), így az entítások külön-külön jellemezhetővé válnak. Ilyen adatszerkezetet a MARC nem tud kezelni, tehát más háromelemű állításokat leképezni tudó formátumot kell alkalmazni. Ez a szemantikus webtechnológia alkalmazásának egyik indoka. A többről később.

II. BIBFRAME-ről MARC-ra

A folytatásban: „BIBFRAME-ről MARC-ra? Hát nem arról volt szó, hogy elérkezett a szemantikus

web Kánaánja a könyvtárak számára? Az epizódban megvizsgáljuk a Kongresszusi Könyvtár május 1-jén közzétett bejelentésének hátterét, és megtudjuk, mi áll az elsőre furcsán hangzó hírek hátterében...”

Elkészült a megoldás, melynek segítségével **BIBFRAME gráfok**at lehet MARC-rekordokká konvertálni. A Kongresszusi Könyvtár javaslata alapján lehetőség lesz a fordított irányú adatkonverzióra, annak ellenére, hogy első ránézésre a BIBFRAME előnyösebbnek látszik mind a könyvtárosok, mind a használók számára. Miért érdemes mégis egy látszólag értelmetlen konverziós folyamat kidolgozására energiát fordítani?

Mindezt először is a **BIBFLOW** nevű kutatási jelentéssel magyarázzuk, amely azt tárgyalja, hogyan állhat át a könyvtár a **linked data** (kapcsolt adat) alapú katalogizálásra. Ezt 6 lépésben tervezik megvalósítani. A második fázis első lépése, hogy a könyvtárosok által kitöltendő űrlapok nem a bibliográfiai rekordokon, hanem az entításokon alapulnak. A könyvtártudomány 11 entitást használ. A feldolgozás az **RDA** szabályaira és az ebben definiált adatelemekre épül. A kitöltött űrlapok adatainak tárolása nem MARC-rekordokban, hanem **háromelemű állításokban** történik. Az entítások adatai egymásután egyetlen adatszerkezetben vannak felsorolva, ezt nevezzük gráfnak. Több szoftver is segíti az entitás alapú feldolgozást. (BIBFRAME Editor, BIBFRAME Scribe, Synopia, Libris)

Az új szerkesztőfelületek párhuzamosan működnek más, még MARC-alapú alrendszerrel. Azonban ha minden alrendszert egyszerre állítunk át, akkor hogyan történik a **kölcsönzés** vagy a **leltározás**? A BIBFLOW válasza, hogy nem is érdemes egyszerre minden alrendszert átállítani, először csak a **feldolgozással** érdemes foglalkozni. Az a megoldás, hogy a gráffal együtt egy csökkentett adat tartalmú MARC-rekordot is létrehozunk a hagyományos funkciók támogatására. A Kongresszusi Könyvtár munkatársai jelenleg kétszer írják le ugyanazt a forrást. Elkészül az **entitás alapú leírás**, majd a **MARC-rekord** is. Elkészül az azon gondolkodni, hogyan lehet az elvégzendő munkát csökkenteni. Így merült fel az igénye a MARC-rekord automatikus elkészítésének.

Mindeközben a következő kérdések merültek fel a konverziós szoftver kidolgozása során: Mindjárt elsőként az, hogy a **forrásleírásból**, a mű adataiból milyen MARC-rekord készüljön? Készülhet

egységesített cím besorolási rekord, bibliográfiai rekord, de mindkét esetben lesznek olyan adatelemek, amelyekre nincs megfelelő bibliográfiai **adatmező**. Ezért vagy újabb adatmezők bevezetésére lesz szükség a MARC-ban, vagy bizonyos adatelemeket ki kell hagyni. A következő kérdés, hogyha egy mű (pl. Rómeó és Júlia) megjelenési formájához (pl. szöveges dokumentum, dráma) több kifejezési forma (pl. nyomtatott, elektronikus) is tartozik, egy vagy több bibliográfiai rekord készüljön-e? Újabb kérdésként merült fel, hogyan őrizzük meg az egyes adatelemekhez tartozó **állandó azonosítókat, URI-kat**? Ezek az állandó azonosítók a MARC21 rekordban a \$0 azonosítóval ellátott almezőbe kerülnek szabvány szerint. HUNMARC-ban ezek az almezők nincsenek definiálva. Rögzített hosszúságú mezők esetén nincsenek meghatározva további almezők, ebben az esetben az állandó azonosító csak a rekord valamely más pontján tárolható. A MARC-rekord bizonyos adatelemei nem generálhatók a gráfból, akkor ezek kihagyásra kerülnek?

A MARC-ban előfordul, hogy bizonyos adatelemek (pl. a dokumentum nyelve) több mezőben is rögzítésre kerül. Egy adatelem többszöri ismétlődése esetén csak egy mezőben kerül rögzítésre a rekonverzió során. Kivétel ez alól az az eset, amikor az adatelem rögzített hosszúságú mezőhöz csatlakozik. Ilyenkor az ismétlődés az azonosító elhelyezhetőségét szolgálja.

Előfordul a MARC-ban, hogy egy mezőn belül több almező kerül rögzítésre különböző értékekkel. (Pl. a 650-es mező esetén a tárgyszólánccsal elemek külön-külön almezőben vannak rögzítve). A BIBFRAME-re történő konverzió során mindegyik almezőérték külön azonosítót kap. Visszaíráskor a rekord tartalma kiegészül a \$0 almezőbe kerülő azonosítóval. Ilyenkor nagyon gondosan kell eljárni, hogy minden egyes almezőérték a megfelelő azonosítóhoz kapcsolódjon. Ellenkező esetben **adathiba** lép fel.

A fenti néhány bekezdésben bemutattuk a BIBFRAME-ről MARC-ra történő rekonverzió vagy visszaírás nehézségeit. Most nézzük milyen új tartalmak kerültek fel a BIBFRAME weboldalára:

<http://www.loc.gov/bibframe/>

Elkészültek a BIBFRAME 2.0-ról MARC 21-re történő konverziós folyamat specifikációi az Index Data szoftverfejlesztő cég közreműködésével. Ezenkívül készítettek egy konverziós segédeszközt is, amelynek letöltésével ki-ké magam is próbálkozhat BIBFRAME állományok visszaírásával MARC-rekordokba. Készült egy összehasonlító segédeszköz, ahol egyszerre nézhetem meg ugyanazt az adattartalmat BIBFRAME gráf formájában, illetve az abból generált MARC-rekordot is. A kongresszusi könyvtár rekord azonosítóját kell megadni, ez alapján a segédeszköz megkeresi a rekord BIBFRAME gráfját XML formában, majd elkészíti a hozzá tartozó MARC-rekordot. Érdemes összehasonlítani a rekonvertált MARC-rekordot az eredeti MARC-rekorddal is. Lehetőség nyílik az oda-vissza átalakításokra. A különféle átalakítások során eltérések lehetnek a más-más módon keletkezett rekordok illetve BIBFRAME gráfok között. Ha hibát találunk, akkor a konverziós folyamatot finomítani kell.

Az előadó ezt követően a svéd Libris közös katalógus bemutatásával folytatja. Ennek a közös katalógusnak az a különlegessége, hogy ide a feldolgozó könyvtárosok nem beküldik az elkészült rekordokat, hanem a közös felületen történik a forrásleírás, azaz entitás alapú feldolgozás történik. Az adatbázis gráfokat tárol. A feldolgozás végzetével több RDF-formátumban is megtekinthető a gráf. Ezután a feldolgozó könyvtáros már csak, mint export formátumot használja a MARC-ot. A BIBFRAME gráf alapján készült MARC rekord kerül be az egyes könyvtárak adatbázisába, ahol ez alapján kölcsönöznek, leltároznak stb.. A MARC-rekord előnézete attól függ, melyik könyvtár fogja majd azt használni. Az egyes könyvtárak eltérő mezőket definiálnak. (Ilyesmit a MOKKA-ban is tapasztalunk időnként. – a leíró megjegyzése) Ezzel zárult az előadássorozat második része.

Ezzel zárult az előadássorozat második része. Az online tanfolyamon résztvevők nevében is nagyon köszönöm „kedves Tanár Úr, Miklós” a tanultakat, remélve, hogy átmentem a vizsgán.

Prokné Palik Mária
(BME OMIKK)