

Így dolgozunk mi: A tanfolyam folytatódik...¹

Ebben a részben egy tanulmányt veszünk górcső alá, amely 2019 szeptemberében jelent meg a Cataloging & Classification Quarterly-ben. A kapcsolódó kutatási projekt weboldala itt található:

http://libdata.tab.ionio.gr/models/si-mapping/si_project.html

A tanulmányt három görög kutató készítette: *Sofia Zapounidou*, *Michalis Sfakakis* és *Christos Papatheodorou*.

Az RDA „származékos” típusú kapcsolatainak leképezése BIBFRAME 2.0-ban

A tanulmányban egy elméleti, adatmodellezési problémával ismerkedhetünk meg, amelynek gyakorlati jelentősége a különféle szótárak segítségével megfogalmazott gráfok egymásba alakításánál van. Az úgynevezett „származékos” kapcsolatok szerepét vizsgáljuk az RDA és a BIBFRAME adatmodelljében. Megnézzük, hogyan segíti mind ez a zökkenőmentes konverziót.

A bibliográfiai kapcsolatok elméletéről *Barbara Tillet* és *Richard P. Smiraglia* írt a *Relationships in the Organization of the Knowledge* (szerk.: *Carol A. Bean*) és az *Understanding FRBR – What is IT and How It Will Affect Our Retrieval Tools* (szerk.: *Arlene G. Taylor*) kötetekben. Tillet a tanulmányában **hét kapcsolattípust** definiált. Ide tartozik az **azonossági** (equivalent) kapcsolat-típus. Ezek egyike a **származékos** vagy ún. derivative kapcsolattípus, amelybe a fordítás, az illusztrált kiadás, a rövidített változat, a dramatizált változat, a forgatókönyv és az adaptáció is beletartozik. A származékos kapcsolattípus jellegzetessége, hogy bizonyos esetekben az eredeti mű és a származék kevéssé tér el egymástól, például: fordítások, rövidített változatok esetén **azonos műről** beszélünk. Ha az eredeti és a származék erősebben eltér egymástól, például: adaptáció, dramatizált változat, akkor ezeket két **külön műnek** tekintjük. Ezenkívül a **leírás** (descriptive), **rész-egész** (whole-part), **járulékos** (accompanying), **folytatólagos** (sequential), **kö-**

zös jellemző (shared characteristic) kapcsolattípus is ismert.

Smiraglia meghatározta a **bibliográfiai család** fogalmát. Mindig van egy **közös ő**s (progenitor), ebből erednek más művek, amelyek együttesen a bibliográfiai családot alkotják. Ezek nagyjából azonos eszmei – szellemi tartalommal bírnak. Az idő múlásával egyre terebélyesedik a család. Keletkeznek fordítások, eltérő kiadások (rövidített, illusztrált stb.) adaptációk (forgatókönyv, dramatizált változat stb.). Ezek mind a közös ő származékai.

Az **RDA** forrásleírási (katalogizálási) szabályzat, amely az AACR2-t (Anglo-American Cataloguing Rules – Angol-Amerikai katalogizálási szabályzat második kiadása) váltotta fel. Az RDA az **FRBR** fogalmi modelljét veszi alapul, a leírás folyamata entitásokra épül. Először leírjuk azt, amit a kezünkben tartunk a **megjelenési formát** és a **példányt**. Utána dolgozunk a **művel**, a **kifejezési formával** és egyéb entitásokkal. Az *RDA szabályai MARC-környezetben korlátozottan alkalmazhatók*. Ha azonban az adatokat gráfként tároljuk (RDF-ben, azaz háromelemű állításokként) az entitások adatai jobban szétválhatnak, és több kapcsolat rögzíthető. Ily módon a forrásleírási szabályzat követelményeinek jobban meg lehet felelni. Az állításokhoz azonosítók kellene, az elemek és a kapcsolattípusok azonosítói a következő portálon érhetők el: <http://rdaregistry.info>

Ezek az azonosítók tartozhatnak bizonyos elemekhez, például: a hordozótípusokhoz, nyelvekhez, közreműködői funkciókhoz stb. (szemantikus elemkészlet) és kapcsolattípusokhoz (szótár, ontológia).

¹ Az írás Hubay Miklós: Értsük meg a szemantikus webet! – címmel a Könyvtárosok Dolgozószobája Facebook csoportban tartott előadás-sorozata alapján készült. A TMT 2020-ban megjelent 67. évfolyama 5. számában Így dolgozunk mi: Tanfolyamon itthon? – címmel megjelent beszámoló folytatása.

A BIBFRAME relációsztár (egy a sok közül) kapcsolattípusok azonosítóit tartalmazza. A BIBFRAME egyszerre azonosító lista és adatmodell. A kapcsolattípusokat entitásokként határozza meg, megköti az alkalmazhatóságukat. El kell gondolkozni azon, hogy minek (milyen entitásnak) van oldalszáma, minek van vonalkódja, minek van szerzője stb. Oldalszáma a megjelenési formának van. Ezért az oldalszámot csak a megjelenési forma jellemzésénél használhatjuk fel. A vonalkódot a példány entitásjellemezésére alkalmazhatjuk. A szerző pedig csak a műhöz kapcsolódhat.

A BIBFRAME és az RDA adatmodellje eltér egymástól. Az **RDA** az **FRBR** entitásait veszi alapul, a művet, a kifejezési formát, a megjelenési formát és a példányt. A **BIBFRAME**-ben nincs kifejezési forma, ily módon a kifejezési forma jellemzőit is a mű jellemzőiként kell megadni. Az RDA négy alap entitással dolgozik, a BIBFRAME pedig csak hárommal, ennek erőteljes hatása van a bibliográfiai kapcsolatok megőrzésére.

A **származéki kapcsolat** az RDA-ban *művek és kifejezési formák között* állhat fenn. Fennállhat a származékos kapcsolat egy mű két kifejezési formája között (pl.: az eredeti nyelvű művet lefordítjuk valamilyen más nyelvre). A származékos kapcsolat fennállhat *két különböző mű között* is, ebben az esetben nem tudjuk melyik kifejezési forma volt a kiindulópont a származék előállításakor. Fennállhat a származékos kapcsolat *két mű egy-egy kifejezési formája között* is, (pl.: egy mű eredeti formájából készítünk egy dramatizált változatot, ez nagyon erős eltérés, tehát új mű keletkezik). A származékos kapcsolat jelölhető általános kapcsolatmegnevezéssel (X Y-nak a származéka), de konkretizálható is ún. kapcsolati altípusokkal (pl. *X fordítása Y-nak*, *X rövidített változata Y-nak* stb.). A szótárban rögzített információk miatt a gép számára is kikövetkeztethető, hogy mindegyik altípus az általánosabb származéki viszonyt is jelenti (pl.: ha *X rövidített változata Y-nak*, ez egyben azt is jelenti, hogy *X a származéka Y-nak*). Az emberi agy számára ezek a következtetések magától értetődők, de a szemantikus web éppen azt jelenti, hogy ezeket a gép számára is egyértelművé tesszük, azáltal, hogy a kapcsolatokat hierarchizáljuk. Azt mondjuk például, hogy a fordítási viszony a származéki viszonynak egy altípusa.

A származéki kapcsolat a BIBFRAME-ben kizárólag művek között állhat fenn. Ebben az esetben ismerjük a kiinduló és a származtatott kifejezési formát is. Itt az általános kapcsolatmegnevezés a

relatedTo, azaz *X kapcsolódik Y-hoz*. Ennek van egy specifikusabb változata: *X-nek származéka Y* azaz *Y származik X-ből* (**hasDerivative** azaz **derivativeOf**). Ezen belül kevesebb kapcsolattípus fejezhető ki konkrét relációval, mint az RDA-ban. Ez azt jelenti, hogy van az RDA-ban olyan specifikus kapcsolattípus, amit a BIBFRAME-ben csak a **relatedTo** általános kifejezéssel írhatunk le. Elvesztjük a speciális kapcsolattípusokat.

Most egy konkrét példa *Homérosz Odüsszeia* című műve kapcsán hasonlítsuk össze az RDA-t és a BIBFRAME-et. Az RDA értelmezésben vett mű, csak egy bizonyos szellemi tartalom az *Odüsszeia*. Ennek az RDA szerinti értelmezésben példánkban van három kifejezési formája: egy ókori görög nyelvű szöveges kifejezési forma, egy angol nyelvű szöveges kifejezési forma, és egy angol nyelvű Braille írással készített kifejezési forma. Ha mindezt BIBFRAME formában próbáljuk megfogalmazni, akkor a **három megjelenési formából három különálló mű** keletkezik, azaz *Odüsszeia görög nyelvű szöveg*, *Odüsszeia angol nyelvű szöveg* és *Odüsszeia angol Braille szöveg*. Annyi különálló BIBFRAME-mű keletkezik, ahány kifejezési formája volt a műnek az RDA szerint. Ezeket azután össze kell kapcsolni, a **hasExpression**, (~van [más] kifejezése) kapcsolattípussal.

Az adatmodellezés nehézségét az adja, hogy ha két RDA segítségével leírt mű kapcsolódik, akkor a származék kifejezési formáival képzett BIBFRAME-művek **minden**, az eredeti műből és kifejezési formáiból létrejött BIBFRAME-mű származékai lesznek. Ez pedig nem igaz. Valójában a származék egyetlen mű segítségével jön létre. Most nézzük mindezt egy példán: válasszuk ki *Emily Brontë Üvöltő szelek* című regényét, ami az eredeti RDA-értelmezésben vett mű, azaz szellemi tartalom. Ennek három kifejezési formája az eredeti angol nyelvű szöveg, a francia fordítás és az örmény fordítás. Az eredeti mű származéka *Alison Case Nelly Dean – Visszatérés Szelesdombra* című regénye. Ez egy az eredeti műtől oly mértékben különböző tartalomegység, hogy két külön műről beszélünk. Ez utóbbi mű egyik kifejezési formája az eredeti angol nyelvű szöveg, a másik a német fordítás szövege. Ha ezt az RDA-relációinak segítségével megfogalmazzuk, majd ezt átalakítjuk BIBFRAME-mé, akkor létrejön egy olyan kapcsolati háló, amely azt sugallja, hogy Alison Case művének mindkét formája származékos kapcsolatban áll Emily Brontë regényének mindhárom kifejezési formájával. Ez azonban tévedés, hiszen Alison Case nyilvánvalóan csak az eredeti angol nyelvű

szöveget használta fel. A probléma az, hogy ha egy eredeti RDA gráfot, amelyben származékos kapcsolat fordul elő BIBFRAME gráffá alakítok, akkor megjelenik ez a hiba. A görög kutatók azt mondják, csak olyan származékos kapcsolat menthető át BIBFRAME-be **művek közötti kapcsolatként**, amely az RDA-értelmezésben **kifejezési formák között** áll fenn. Két RDA-mű származékos kapcsolata nem azonos két BIBFRAME-mű származékos kapcsolatával. Két RDA kifejezési forma származékos kapcsolata feleltethető meg két BIBFRAME-mű származékos kapcsolatának. Az **RDA-szótár segítségével mű és mű között leírt származékos kapcsolatot figyelmen kívül kell hagyni a BIBFRAME gráf elkészítésekor.**

Az RDA-ban kifejezett konkrét kapcsolattípusok egy részét a BIBFRAME-ben csak általánosabb kifejezéssel tudjuk helyettesíteni. Például ha az RDA-ban tudjuk jelölni a fordítása viszonyt, akkor a BIBFRAME-ben is tudjuk jelölni, hogy az egyik mű fordítása a másiknak. Ugyanakkor az RDA-ban leírható dramatizált változata ill. rövidített változata kapcsolatjelölőt a BIBFRAME-ben csak az általánosabb származéka kapcsolatjelölővel tudjuk helyettesíteni. Ez csak úgy orvosolható, ha a BIBFRAME-szótárat kiegészítjük az RDA-szótárból vett relációkkal a saját alkalmazásunkban.

A görög kutatók vizsgálatokat végeztek. Leválogatták gazdag kiadástörténettel rendelkező művek MARC-rekordjait (pl.: Száz év magány 14 rekord, Bűn és bűnhődés 24 rekord és így tovább). Egy bibliográfiai családba a származékaikkal együtt 15-30 rekord tartozott. Ezekből a MARC-rekordokból az RDA-szótár segítségével háromelemű állításokat hoztak létre. Az így előálló halmazt RDA GOLD-halmaznak nevezték. A MARCXML-rekordokat RDF-gráfokká konvertálták, majd ezt kétféle módon is BIBFRAME gráffá alakították. Az RDA-relációkat megfeleltették BIBFRAME-relációknak, a kutatók által megállapított szabályok szerint. Az egyik esetben figyelték a művek és a kifejezési formák közötti kapcsolatot, a második esetben csak a kifejezési formák kapcsolatait vizsgálták. Végül megnézték miben tér el egymástól a két BIBFRAME gráf. A legelső lépésben az RDA gráffá alakítás során 48 mű, 195 kifejezési forma és 236 megjelenési forma képződött. Amikor ebből BIBFRAME gráfot képeztek, akkor 195 BIBFRAME-mű képződött (ez megfe-

lel 195 RDA kifejezési formának), és létrejött a 236 megjelenési forma is. Tehát az entitások kezelése helyes volt. Az RDA-ban az egy műhöz tartozó kifejezési formák közötti kapcsolatot BIBFRAME-ben is meg kell tudni jeleníteni. Erre való a hasExpression tulajdonság. **A gráfok alapos tanulmányozása igazolta, hogy nem szabad figyelembe venni azt, ha az RDA gráfban két mű között van származékos kapcsolat, csak azt, ha két kifejezési forma között áll fenn származékos kapcsolat.** Az RDA-művek származékos kapcsolatainak figyelembe vétele zajt eredményezett a BIBFRAME-művek származékos kapcsolatainak kialakításánál. A projekt weboldalán hozzáférhetők a nyílt kutatási adatok.

A leíró kérdése a tanár úrhoz: miért van szükség kétféle adatmodellre, ezek miért nem kompatibilisek egymással? Az RDA gráfról BIBFRAME gráffá alakítás egy izgalmas játék, de mi a haszna a gyakorlatban? Ezzel zárult is a harmadik előadás.

Hubay Miklós válasza: az átalakításnak igazából akkor van haszna, ha valaki ezt kéri a könyvtártól – elég sok könyvtári rendszer képes a MARC-rekordokat BIBFRAME-mé konvertálni. Ezzel kiszolgálhatja mindazokat, akik BIBFRAME relációs gráfokra építettek valamilyen szolgáltatást, például közös keresőt. RDA-relációkra alapuló gráfot is lehet szolgáltatni, ha van rá valakinek szüksége. A gyakorlatban még nem sok olyan rendszer létezik, amely az RDA-szótárra épülne, legalábbis teljesen, de egy-egy elemét felhasználják, mást kiegészítve. Ennek igazából akkor lesz majd nagy jelentősége, ha egy könyvtár úgy dönt, hogy az adattörzsgépzést az RDA-szótárra alapozva végzi (tehát már elrugaszkodott a MARC-tól), de szeretne BIBFRAME-et is adni, hogy ne maradjon ki a közös keresőkből. Úgy tűnik, hogy a Kongresszusi Könyvtár kifejezetten azt szeretné, hogy a könyvtárak BIBFRAME-et használjanak, tehát legyenek olyan egységesek, mint a MARC esetében. Minél több könyvtár készíti el a konverziót, annál nagyobb lesz a közösen kereshető halmaz. Nem lesznek duplumrekordok, mint a MOKKA-ban, hanem egymás mellé fűződnek az állítások.

Prokné Palik Mária
(BME OMIKK)