

Holl András

A repozitóriumi tartalmak felhasználásának megkönnyítéséről

Tárgyszavak: *repozitórium; metaadat; információkeresés*

Gyakran éri az a vád a repozitóriumokat, hogy adatsilók: a bennük tárolt tartalmak láthatósága csekély. Ebben van igazság, miközben mára már a repozitóriumok széles körben elterjedtek és hatalmas mennyiségű adatot tárolnak. Ezek az adatok többnyire szabadon hozzáférhetők és újrafelhasználhatók, ugyanakkor a potenciális alkalmazók viszont gyakorta nem találnak ingyenes, jogszerűen felhasználható tartalmakat. Jelen rövid írásunkban egy megoldási lehetőséget kívánunk felvázolni.

A repozitóriumok a tárolt digitális objektumok – dokumentumok, képek, videók, hanganyagok – leírására metaadatokat használnak. A használt Dublin Core metaadatséma – az adatcsere megkönnyítését célozva – egyszerű. A repozitóriumi szoftverek támogatják a metaadatokban és a teljes szövegben való keresést, és az OAI-PMH-protokoll alapján a gyűjtemények metaadatai aggregálhatók. Azonban a rendelkezésre álló metaadatok nem elég részletesek, és nem elég szakmaiak a számtalan lehetséges felhasználási terület számára. (Ugyanaz a dokumentum felhasználható lehet a közoktatásban vagy éppen a műtárgy-kereskedelemben, és más-más leírás szükséges a különböző területeken való kereshetőséghez.) Bonyolultabb metaadatséma alkalmazása sem jelent megoldást, hiszen a metaadatsémák az adattárak általános céljait szolgálják ki, minden terület metaadatigényét egyetlen sémába zsúfolni lehetetlen. A teljes szövegű keresés sem feltétlenül kínál megoldást. Amennyiben a különféle felhasználási területekhez alkalmazkodó címkékkel (tag) látnánk el a repozitóriumokban található objektumokat, megteremténk a hatékony, területspecifikus felhasználás lehetőségét. (Egy tételt több felhasználási területhez kapcsolódó címke is jelölhet, amelyek a különböző felhasználási területek számára készült címkékészletekből, teauruszokból származnak, különböző területeken használt rendszereknek felelnek meg.)

A repozitóriumokat fenntartó intézményeknek sem kapacitásuk, sem szakértelmük nincs a tételeik felcímkezésére. Csak a szakterületekről információkkal rendelkezők lehetnek képesek a tartalmakat címkékkel ellátni. Mivel a címkézés nagy munka, valószínűleg érdemes crowdsourcing megoldásokat alkalmazni. Itt a crowdsourcing a repozitóriumot működtető intézményen kívüli szakértők csoportjára való támaszkodást jelenti. A szakértői halmaz lehet zártabb, de lehet szó valódi, széleskörű, önkéntes munkáról is – de a résztvevőket mindenképpen meg kell válogatni.

Nem érdemes a címkézést repozitóriumonként végezni – a címkézés legmegfelelőbb helye a szakmai aggregátori szint lehet. Ugyanakkor nem zárjuk ki annak szükségességét sem, hogy a címkék az aggregátortól visszakerüljenek az egyes repozitóriumokba, illetve hogy esetenként a repozitóriumi szinten keletkezzenek. Külön metaadatmezőket kialakítani nem célszerű megítélésünk szerint – a szükségleteknek megfelelően strukturált címkéket meglévő Dublin Core mezőkben lehetne elhelyezni (téma/subject vagy tartalmi leírás/description).

Lehetséges sémák a címkék készítésére:

```
#felhasználási terület|kulcszó=érték||  
vagy  
#felhasználási terület|kulcsszó||  
vagy  
#felhasználási terület|alany/subject ; állít-  
mány/predicate ; tárgy/object||
```

az utóbbi példa RDF-tripleteket használ. A címkézési projektek tématerületeit az aggregátornál regisztrálni szükséges, a támogatott sémák valamilyékének megjelölésével, a címkékészletek megadásával.

Holl A.: A repozitóriumi tartalmak felhasználásának könnyítéséről

A felhasználók által (az aggregátori szinten) rendelkezésre álló keresőfelületek lehetőséget kell biztosítsanak a szakterület szerinti szűkítésre, és támogatniuk kell az adott területeken alkalmazható kulcsszókészleteket, tezauruszokat, ontológiákat.

Végül a fentieket egy példával illusztráljuk. A repozitóriumi tartalmak lehetséges felhasználási területe a közoktatás. A tananyagok készítői, válogatott vagy a feladatra jelentkező, és az adatgazdagítási jogosultságot megszerző pedagógusok a repozitóriumokban szisztematikusan kutatva, vagy repozitóriumi tartalmakra véletlenül rábukkanva a Nemzeti Alaptantervnek megfelelő, a tantárgynak, korcsoportnak, témának megfelelő címkéssel láthatnák el a repozitóriumokban tárolt, általuk felhasználásra érdemesnek minősített tételeket. A címkék alapján az órájára készülő tanár az aggregátorfelületen válogathat, vagy a címkék megadásával irányíthatja tanítványait a megfelelő tartal-

mak felé. Ugyanaz a repozitóriumi anyag – mondjuk *Bolyai János* Appendix-e – felhasználható a matematikaoktatásban, a tudománytörténetben, vagy éppen keresheti egy felhasználó az UNESCO Világemlékezet Programjában listázott magyar dokumentumok egyikeként. A megfelelő címkék használata elősegítheti, hogy a repozitóriumok valamelyikében található dokumentum egy adott célközönség számára megtalálható és újrafelhasználható legyen.

Beérkezett: 2019. VI. 17-én.



Holl András

az MTA KIK informatikai főigazgató-helyettese.

E-mail: holl.andras@konyvtar.mta.hu