

A metaadat sémák és a Szemantikus Web: egységesítés és specializáció a metaadatok világában

Fülöp Csaba, Kovács László, Micsik András
{csaba.fulop, laszlo.kovacs, micsik}@sztaki.hu
Elosztott Rendszerek Osztály
MTA SZTAKI
Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutató Intézet

Abstract

The ability to semantically relate various data and metadata is a basic criterion of the Semantic Web concept. The most obvious and most urgent subtask in this area is to share and align metadata schemas used in digital libraries, collections and databases. As a result of international research activities, an RDF based framework is evolving which provides a common metadata schema definition format and the possibility to reuse existing schemas and schema elements. The Department of Distributed Systems of MTA SZTAKI tries to introduce these results in Hungary. Our open service provides metadata schema registration, overview of registered schemas and their interconnections. It can also support the creation of new schemas using elements also from existing schemas. Such service could facilitate the reuse and standardization of metadata schemas, which is of public utility, as it can make data management and data retrieval more economical.

Bevezetés

A metaadatok alkalmazásának a Dublin Core indukálta terjedése során nagy számban jelennek meg új metaadatsémák illetve létező sémák kombinációi, variációi. Ezen a területen az újrafelhasználás, rendszerezés még megoldatlan problémák. A Szemantikus Web eszközei viszont kiválóan használhatók a metaadatok és sémáik ábrázolására és kezelésére. Nemzetközi kutatások eredményeképpen kialakulóban van egy RDF alapú modell a metaadatsémák pontos és teljeskörű leírására, amely lehetővé teszi a sémák egységes kezelését, a meglévő sémákból és sémaelemekből történő építkezést.

Az MTA SZTAKI Elosztott Rendszerek Osztálya részt vett a CORES¹ európai projektben, melynek célja volt, hogy segítséget nyújtson a metaadat szerkesztőknek a metaadat szabványok és megoldások együttműködésének növelésével és a különféle metaadat sémák nyilvántartásával. A projekt során elkészítettünk egy metaadat nyilvántartást és egy metaadat sémák létrehozására és újrahaznosítására szolgáló kliens alkalmazást, amely a metaadat sémák újrahaznosításához a nyilvántartásban már szereplő adatokat használja. Mind a nyilvántartás és a kliens alkalmazás Szemantikus Web technológiákra épül. Az alábbiakban röviden bemutatjuk a Szemantikus Web és a metaadatok világának jelenlegi státuszát, majd ismertetjük a projekt eredményeit és lehetséges hasznosításait.

A Szemantikus Web²

A Szemantikus Web alapkonceptiója szintén a World Wide Web atyjához, Tim Berners-Lee-hez kötődik. A World Wide Web (WWW) lehetővé tette, hogy rengeteg információ és dokumentum könnyen elérhetővé váljon a világhálón. Ezek az anyagok azonban egyáltalán nem rendezettek, és ez az ad hoc információhalmaz egyre nehezebben kezelhető. Az Interneten való keresést például már-már tanítani kell, és magasszinten, eredményesen keresni egyfajta művészetnek is tekinthető. A Szemantikus Web célja, hogy a számítógépeket, egyszerű gépi intelligenciákat hívja segítségül ezeknek a problémáknak a megoldására. Ehhez azonban az információkat számítógépek által értelmezhető formára kell hozni. Ez a folyamat már 1999 előtt megkezdődött, de hatása máig alig érezhető a világban. A későbbiekben ennek okaira is rávilágítunk.

A Szemantikus Web módszere, hogy különállóan vagy internetes erőforrásokhoz rendelt gépileg is feldolgozható állításokat helyezünk el az Interneten. Ennek alapeszköze az RDF nyelv, mellyel egységesen reprezentálhatók olyan állítások mint például:

- Ezt a dokumentumot Zöld Béla készítette
- Zöld Béla az MTA SZTAKI munkatársa
- Az MTA SZTAKI egy kutatóintézet

Látható, hogy ezen állítások egy része a metaadatok körébe tartozik, míg más állítások egyfajta globális tudásbázis morzsáiként foghatók fel. Ezeket az állításokat összekapcsolva azután különféle logikai következtetéseket lehet elvégezteni a számítógépekkel, amelyek a felhasználókat különböző módokon segíteni tudják. Egyes kérdésekre automatikusan választ kaphatunk (pl. Hol dolgozik ennek a dokumentumnak a szerzője? Mikor indul legközelebb vonat Pécsre?), újfajta kereséseket végezhetünk (pl. a magyarországi kutatóintézetekben készült dokumentumokat keressük), vagy hasznosabb eredménylistát kaphatunk a kereséseinkről (pl. kiemelve, hogy kollégánk mely dokumentumokat talált érdekesnek). A Szemantikus Web eszközkészlete az emberi kommunikációban is alkalmazható: állításainkat, megjegyzéseinket ezzel a módszerrel tetszőleges mások által írt weblaphoz is hozzárendelhetjük.

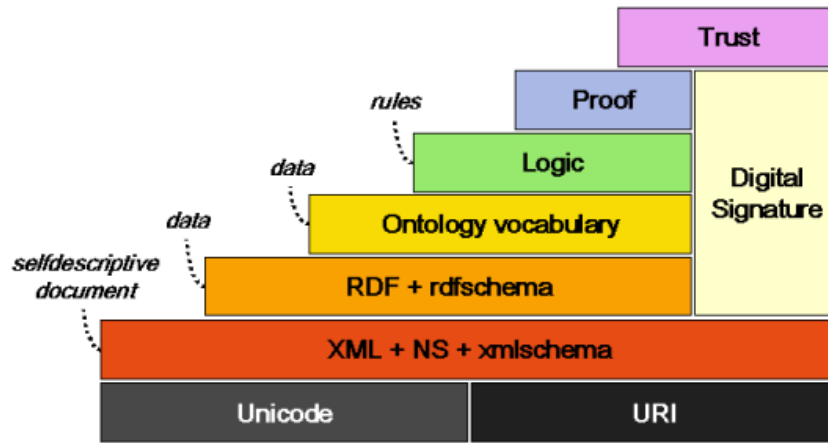
A fenti mechanizmusok elterjedéséhez nem csak új eszközökre, de nagyobb mennyiségű hasznos állításra is szükség van, amelyek már tényleges használati előnyöket tudnak nyújtani. A metaadatok az erre a leginkább alkalmas és kézenfekvő adathalmaz. Itt kapcsolódik a Szemantikus Web a metaadatok világához és cikkünk fő témájához is. De előzőleg még részletesebben ismertetnünk kell a Szemantikus Web felépítését.

A lépcsős modell³

A Szemantikus Web tehát a következtetéseket is lehetővé tevő metaadat infrastruktúra a weben. Ennek legfőbb építőköveit a World Wide Web Consortium (W3C) dolgozza ki és szabványosítja. Az ide vonatkozó legfontosabb ajánlások a Resource Description Framework⁴ (RDF), RDF Schema⁵ (RDFS) és a Web Ontology Language⁶, de a kép ennél jóval összetettebb. A Szemantikus Web összképét az egymásra épülő célok lépcsős modelljével lehet szemléltetni (1. ábra). Ennek rétegei a következők:

1. Az első szinten a Unicode és URI ajánlások biztosítják az egységes nemzetközi karakterformátumot és az egységes címezhetőséget. A Szemantikus Weben belül az URI-k használata jóval túlmutat a megszokott weblap címeken. Egy személyt is azonosíthat mondjuk az e-mail címét tartalmazó URI, vagy egy intézetet a weblapjának a címe. Tulajdonságokat, állításfajtákat is URI-k azonosítanak, így az az állítás mely szerint Zöld Béla az MTA SZTAKI munkatársa, 3 URI segítségével oly módon fogalmazható meg, hogy az más állításokkal az URI-k mentén összekapcsolhatóvá válik.
2. Az XML réteg biztosítja, hogy a Szemantikus Web definíciókat egyéb XML alapú szabványokkal integrálhassuk. Az XML a névtér és séma definíciókkal együtt lehetővé teszi, hogy egy XML dokumentum a létrehozásakor alkalmazott szintaktikai és szemantikus szabályokra hivatkozzon. Ezáltal egy önleíró, validálható dokumentumformátum jön létre.
3. Az RDF és RDFS réteg biztosítja, hogy állításokat tehesünk az objektumokról URI-k segítségével, és olyan szótárakat készítsünk amelyre ezen URI-k hivatkozhatnak. Az első RDF és RDFS specifikáció 1999-ben készült el, de végleges W3C ajánlássá csak 2004 februárjában vált. Fontosságuk miatt ezekkel az ajánlásokkal külön alfejezetben foglalkozunk.
4. Az ontológia réteg lehetőséget ad bonyolultabb szótárak felépítésére. Itt már a különböző fogalmak közötti összefüggések is megadhatók, hasonlóan egy tezauruszhoz.
5. A digitális aláírás réteg a dokumentumok hitelességét hivatott biztosítani.

Ezen rétegek szabványosítása már megtörtént a W3C-nél, a többi réteg még intenzív kutatás alatt áll és csak demonstrációs alkalmazások léteznek. Az elképzelések szerint segítségükkel lehetőség lesz logikai szabályok megadására, következtetések kiértékelésére és az eredmények hitelességének eldöntésére.



1. ábra: A Szemantikus Web lépcsős modellje

RDF, RDFS

Egy RDF leírás metaadat állítások összessége, amely állítások hármassokból állnak⁷: (alany, állítmány, tárgy), ahol a hármass elemei URI-k illetve a tárgy egy szöveges címke is lehet. Ebben az esetben arról további állítások nem tehetők. Az RDF állítások halmaza egy irányított, címkézett gráf, ahol a gráf csomópontjai az alany és tárgy, a gráf élei pedig az állítmányok. Az URI-k alapvető szerepe, hogy egyértelműen azonosítsák az erőforrásokat, így több RDF dokumentum is mondhat állítást ugyanarról az erőforrásról. Ez biztosítja, hogy a metaadatokat egymástól független szereplők definiálhassák, és bármely alkalmazás kombinálhassa, összevonhassa őket.

Az RDF formális leírása is gráfokkal dolgozik és emberi felhasználók számára is sokkal könnyebb gráfokban gondolkodni, de szükség van ezen gráf gépek által feldolgozható ábrázolására is. Az XML vagy n3 leírás⁸ csupán gyakorlati eszközök. Némely RDF szerkesztő is gráfokkal dolgozik, és az XML átírás csak a háttérben, a dokumentum mentésekor történik.

Az RDF séma az objektumok leírásához osztályokat és osztálytulajdonságokat vezet be. Ezáltal a sémakon alapuló RDF leírások strukturáltabbá válnak, bár bonyolultabb összefüggések leírására csak a következő, ontológiai szinten lehetséges a szintén RDF-en alapuló OWL nyelv segítségével. Ha példaképpen egy objektumot a Munkatárs osztályhoz tartozóként írunk le, akkor a séma szerint az osztályhoz tartozó Neve, Telefonszáma és E-mail címe tulajdonságait adhatjuk meg RDF-ben.

Egy RDF séma megadja azt is, hogy egy osztály mely más osztályok kiterjesztése, valamint hogy milyen kapcsolatban állhatnak ezen osztályok példányai. Az osztályokhoz hasonlóan a kapcsolataikat is egy öröklés szerű hierarchiába szervezhetjük, valamint megadhatjuk az értékészletek és értelmezési tartományukat.

Ellentétben az XML sémákkal, melyek szigorú szintaktikai megkötéseket adnak egy dokumentum felépítésére, az RDF sémák egy névtér felépítését definiálják, szemantikai jelentést rendelve az adott szintaktikához. Ebből következően habár látszólagos rokonság áll fenn az XML és RDF sémák között, egyik sem helyettesítheti a másikat.

A metaadatok világa

Ma már a Dublin Core név mindenki számára ismerősen cseng, és senkinek sem kell elmagyarázni a metaadat fogalom jelentését. Pedig a Dublin Core viszonylag rövid történetre tekinthet vissza. Létrejöttét az 1995-ben Dublinban (Ohio) megtartott műhelykonferenciához kötik, ahol a kb. 50 résztvevő a webes erőforrások leírásának, kategorizálásának szemantikájáról folytatott megbeszéléseket, és ahol nagy nehezen sikerült megegyezniük 13 leíró elemben. Az eredményt a workshop helyszínéről „Dublin Core Metaadatoknak” nevezték el. Azóta is évente tartanak műhelykonferenciát, amely az utóbbi években nemzetközileg elismert és népszerű tudományos konferenciává nőtte ki magát.

A Dublin Core (DC) elemek száma először 15-re majd napjainkban 16-ra bővült. A metaadatszabvány gondozására és fejlesztésére megalakult a Dublin Core Metadata Initiative⁹ (DCMI), melynek célkitűzése a szemantikus együttműködés támogatása. A Dublin Core folyamatosan változik. Az elemkészlet 1.0 verzióját¹⁰ hamarosan felváltotta az 1.1 verzió¹¹, amely mára már szintén túlhaladottá vált. Eközben a Dublin Core egy korábbi verziója ISO szabvány is lett¹². Megjelentek az ún. minősítők¹³ (qualifier-ek), amelyek az egyes elemek jelentését pontosítják (vagy ha úgy tetszik megsokszorozzák, például a dátum elem minősítve lehet a készítés, publikálás vagy módosítás dátuma is). Az elemek tartalmának ábrázolására is számos ajánlást lehet alkalmazni¹⁴. Ezáltal a dátumok, helyszínek, stb. leírása egységesül, illetve egyes elemek esetében szótárak, értékkészletek alakulnak ki (lásd DCMITypes a dokumentumtípusok leírására).

A DCMI sok útmutatót és magyarázatot is kiad a Dublin Core elemkészletekhez, összegyűjti az egyes elemek használati tanácsait, és segít a Dublin Core metaadatok különböző reprezentációinak kidolgozásában. Ide tartozik az is, hogyan írunk le DC metaadatokat XML-ben¹⁵ vagy RDF-ben^{16 17 18}, és hogyan társítsuk ezeket weblapokkal¹⁹. Ezen a ponton tehát találkozunk a Szemantikus Web és a metaadat világ, létrejön az a kapcsolat, amelyen keresztül a két mozgalom előnyeit kölcsönösen élvezni tudjuk.

Nem szabad elfeledkezni arról, hogy a Dublin Core a dokumentumokhoz rendelhető metaadatoknak csak egy fajtája, az ún. leíró metaadat, amely a keresés és böngészés során a könyvtári katalóguscédulához hasonlóan segít abban, hogy egy előzetes elképzelésünk alakuljon ki a dokumentum tartalmáról. Ezen kívül a metaadatoknak több más szerepe is lehet, amelyek csak érintőlegesen vagy egyáltalán nem férnek bele a DC kereteibe. Ide tartoznak például a dokumentum belső struktúráját leíró, karbantartását, megőrzését segítő metaadatok vagy a dokumentumhoz kapcsolódó jogokat és jogosultságokat megadó metaadatok.

A Dublin Core használata rohamosan terjed, és a gyakorlatban is bebizonyosodott, hogy teljesen különböző területeken is egyaránt alkalmazható. Létrejött tehát egy globális konszenzus, amely által a dokumentumok leíró metaadatai globálisan összekapcsolhatók. Ez az összekapcsolás szükségszerűen leegyszerűsítéseket, kompromisszumokat tartalmaz. A Dublin Core felhasználói lokális, kevesebb kompromisszummal és leegyszerűsítéssel járó konszenzusokat kezdtek keresni. Megjelentek az egyes alkalmazási területek saját értelmezései (pl. könyvtárak, oktatás, közigazgatás), melyek nem egyszer új elemeket is felvettek a DC elemei mellé.

Egy ilyen hierarchikus konszenzus-rendszert egyik európai K+F projektünkben éltünk meg, amelyben közösségi rádióknak metaadatsémára volt szükségük. Több szervezet szövetkezve hozta létre a SOMA²⁰ metaadatsémát, amely a European Broadcasting Union rádiók számára készült EBU Core²¹ sémát finomította tovább. Az EBU Core pedig természetesen a Dublin Core-ra épül, annak elemeit magyarázza, értelmezi. Például a Date.Issued elem kitöltendő tartalmát az első sugárzás dátumaként határozza meg, ajánlást tesz a személynevek rögzítési formájára, és egy új minősítőt vezet be Role néven, amellyel a személyek szerepkörét (riporter, rendező, interjúalany, stb.) egy adott listából kiválasztva adhatók meg. Ebben a konszenzus-hierarchiában az egyre mélyülő szinteken egyre speciálisabb, specializáltabb közösítések lehetségesek, míg a lényeg, a globális összekapcsolhatóság a Dublin Core szintjén sértetlen marad.

A fent leírt helyzetet és a kialakult használati módot figyelembe véve szükségessé vált a dinamikusan változó metaadatsémákat valamilyen rendszerbe foglalni. Az első ilyen kísérlet a Warwick framework²² volt, ami arra ad lehetőséget, hogy több séma szerinti metaadatot egy metaadatként ábrázolhassunk oly módon, hogy az egyes sémákhoz tartozó elemeket külön csomagokban helyezzük el. Az a kifinomultabb rendszer, amelyet a Dublin Core nyelvtanának is lehet nevezni, rögzített alapsémákat (névtereket) és ún. alkalmazási profilokat^{23 24} különböztet meg. Az előbbi egy szabványosító szervezet által kiadott eredeti metaadatséma, míg az utóbbi egy speciális terület vagy alkalmazás számára létrehozott módosított, finomított séma. Az alkalmazási profil definíció szerint egy vagy több névtérből gyűjt össze elemeket, de nem vezethet be új elemeket. A kiválasztott elemeket az alkalmazás profil az alábbi szabályok szerint módosíthatja:

- finomíthatja az értelmezését, de csak az eredeti értelmezés szűkítésével. Például a formátum elem finomítása rádióműsorok esetén a műsor hosszának és a hanganyag minőségi paramétereinek (sztereómonó, mintavételsűrűség, stb.) megadása lehet.
- módosíthatja az elem előfordulási módjait: kötelező, ismételhető, stb.
- megszabhatja az elem értékészletét egy adott szókészlettel (pl. riport, magazin, hírműsor, zene...) vagy egy ún. kódolási sémával (pl. dátum formátuma)

Ennek a modellnek a továbbfejlesztésén alapul az itt ismertetendő egységes metaadatséma megadási mód, melyet a következő fejezetben részletezünk.

Fontos kérdés, hogyan társíthatunk Dublin Core metaadatokat egy dokumentumhoz. Az egyik lehetőség magában a dokumentumban elhelyezni azokat. A HTML nyelv esetén ezt megtehetjük a fejlécben elhelyezett META tagok segítségével, vagy egy különálló metaadat-rekord hivatkozásával (ez lehet XML vagy RDF/XML formátumú). Mivel az RDF leírás címzési módszere az URI, az RDF leírás tetszőleges helyen lehet, az URI egyértelműen azonosítani fogja, hogy mire vonatkoznak az adott metaadat állítások. A keresők dolgát megkönnyítendő a HTML oldalunkon hivatkozhatunk erre az RDF dokumentumra, de ez nem szükséges. Az URI definíciójából adódóan olyan dolgokhoz is megadhatunk metaadatokat amelyekhez az RDF állomány nem kapcsolható, illetve nem is szükséges hogy ezen objektumok a weben ténylegesen elérhetők legyenek (pl. városok, emberek).

A metaadatok hozzáféréseinek egyik leghatékonyabb serkentője az Open Archives Initiative²⁵, amely OAI-PMH protokollja révén nagy archívumok metaadataihoz nyújt egységes hozzáférést. Az OAI azonban még az 1.0 verziójú Dublin Core-t veszi alapul.

Mindezen eredmények magyar alkalmazására vonatkozóan komoly lemaradás tapasztalható. Hiányoznak a megfelelő ajánlások magyar fordításai, alkalmazásuk speciális magyar problémáinak feloldásai. A DC elemkészletnek nincs még hivatalos magyar fordítása (pedig több mint 20 nyelven már elérhető), bár a NAVA projekt és a Magyar Elektronikus Könyvtár készített már javaslatokat a fordításra. Hazai metaadatsémák definiálására is viszonylag kevesen vállalkoztak még, ezek közül megemlítenénk néhányat, ahol figyelmet szenteltek a DC és más nemzetközi ajánlásoknak való megfelelésre. Ilyen volt a kormányzati KIKERES²⁶ projekt, a NAVA (Nemzeti AudioVizuális Archívum), és ilyen a jelenleg induló NDA (Nemzeti Digitális Adattár)²⁷. Az MTA SZTAKI és az Országos Széchényi Könyvtár közös HEKTÁR²⁸ nevű projektje az OAI és a Dublin Core használatának terjesztését segíti a könyvtárak és kulturális archívumok között. Az MTA SZTAKI Elosztott Rendszerek Osztálya által üzemeltetett nyílt rádiós archívum²⁹ a StreamOnTheFly nevű európai K+F projektünkben közösségi és regionális rádiók számára kidolgozott DC alapú metaadatsémát és adatcsereformátumot használja.

Mivel a metaadatok, metaadatsémák használata még nem terjedt el az országban, nem csoda, hogy fel sem merült az igény ezeknek a sémáknak a nyilvántartására, rendszerezésére, kezelésére. Reméljük, hogy eljön az idő, amikor a következőkben leírt módszereket Magyarországon is alkalmazni fogják.

Metaadatséma-nyilvántartások

A metaadatsémák konkrét és precíz leírásai általában a szabványosítási szervezetek ajánlásaiban találhatóak meg. Ezek a statikus dokumentumok azonban nem tudják megfelelően tükrözni a változásokat, és nem adnak megfelelő áttekintést a séma belső és külső kapcsolatrendszeréről. Ezt felismerve több szervezet kísérletezik interaktív metaadat nyilvántartások (metadata registry) használatával. Ilyen például a DCMI által üzemeltetett nyilvántartás³⁰, melyben a Dublin Core elemeinek aktuális definícióját, történetét és kapcsolatait lehet könnyen megkeresni. Hátránya, hogy csak a DC saját sémáit tartalmazza, és nem terjed ki a DC számos kapcsolódó sémájára.

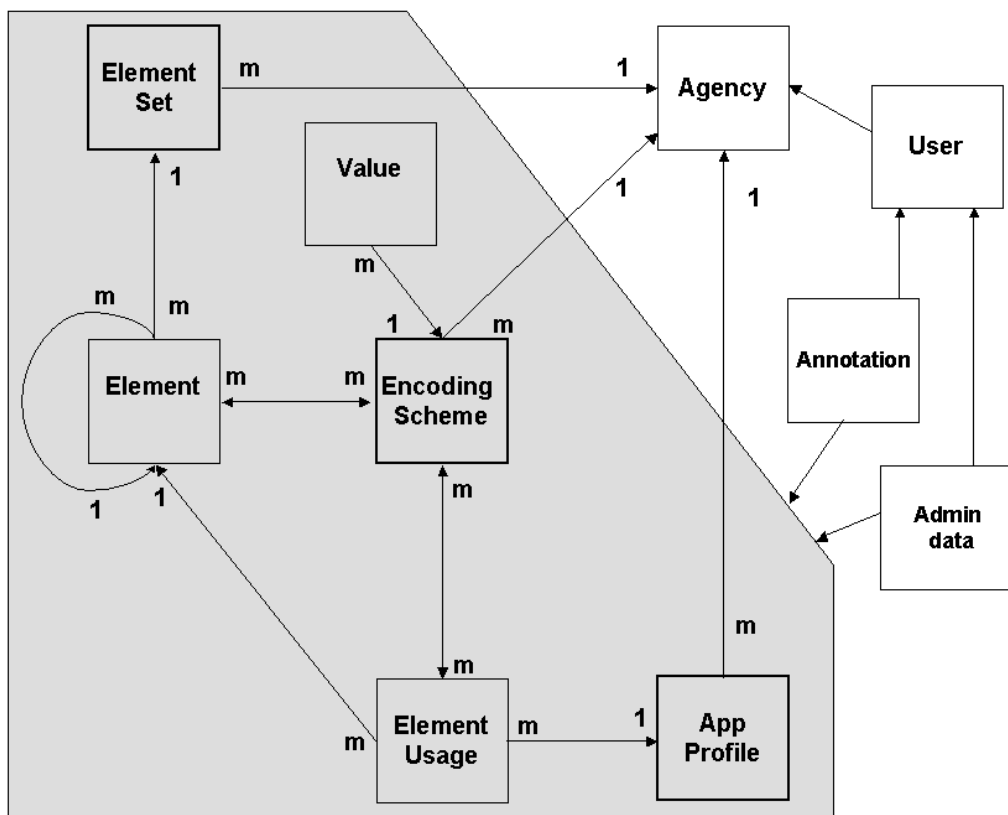
A CORES nyilvántartás³¹

A CORES projekt keretén belül egy olyan átfogó metaadatséma-nyilvántartás készült el, amely egy átgondolt modell alapján rendszerezi és összefogja a különböző metaadatsémákat, és ezen felül támogatja új

sémák készítését is³². Sokszor adódhat, hogy a meglévő metaadat sémák nem megfelelőek egy adott alkalmazáshoz annak különleges igényei miatt vagy egyszerűen az adott területen még nem történt metaadat szabványosítás. Ekkor rosszabb esetben a fejlesztők saját sémákat készítenek ezzel teljesen inkompatibilissá téve alkalmazásukat más metaadat sémákkal vagy jobb esetben már meglévő sémák részeit használják fel. Ezt felismerve a projektben résztvevő angol partnerünk (UKOLN) már korábbi projektjeiben is egy módszert körvonalazott, amelynek segítségével nagymértékben megkönnyíthető az alkalmazás profilok (application profile) elkészítése már meglévő sémák alapján, illetve a már meglévő profilok újrafelhasználása. A CORES projekt ezen korábbi projektek munkájára építve próbálja meg tökéletesíteni és népszerűsíteni ezt a megoldást.

A módszer lényege, hogy az alkalmazás specifikus profilokat már létező, széles körben elterjedt és elfogadott sémák elemeinek felhasználásával, esetleg azok jelentésének finomításával, értékészletükre tett megszorításokkal, tudják elkészíteni az alkalmazás fejlesztők. Így ha alkalmazásuk egy már ismert séma elemeire építik, akkor azt a sémát ismerő többi alkalmazás képes lesz ezen elemek értelmezésére, feldolgozására, míg a többi, újonnan definiált vagy máshonnan származó, számukra ismeretlen elemet figyelmen kívül hagyhatják.

A metaadatsémák modellezése



2. ábra: A CORES modell

A rendszer a 2. ábrán látható RDFS modellt alkalmazza a metaadat sémák és alkalmazás profilok megadására. A szürke részben található entitások a metaadat sémák leírásáért felelősek. Az elemkészlet (ElementSet) és elem (Element) entitások a hagyományos metaadat sémák leírására szolgálnak, a kódolási séma (Encoding Scheme) és érték (Value) entitások pedig az azokhoz rendelhető megszorítások, minősítők megadására. Az újdonságot az alkalmazási profil (Application Profile) és elem felhasználás (Element Usage) entitások jelentik, amelyek az alkalmazás profilok leírására szolgálnak. A profil elemeit az elem felhasználások adják, amelyek egyértelműen kapcsolódnak egy már létező elemkészlet egy eleméhez. Azért, hogy ne csak már létező elemkészletek elemeit lehessen felhasználni, az alkalmazások készítőinek

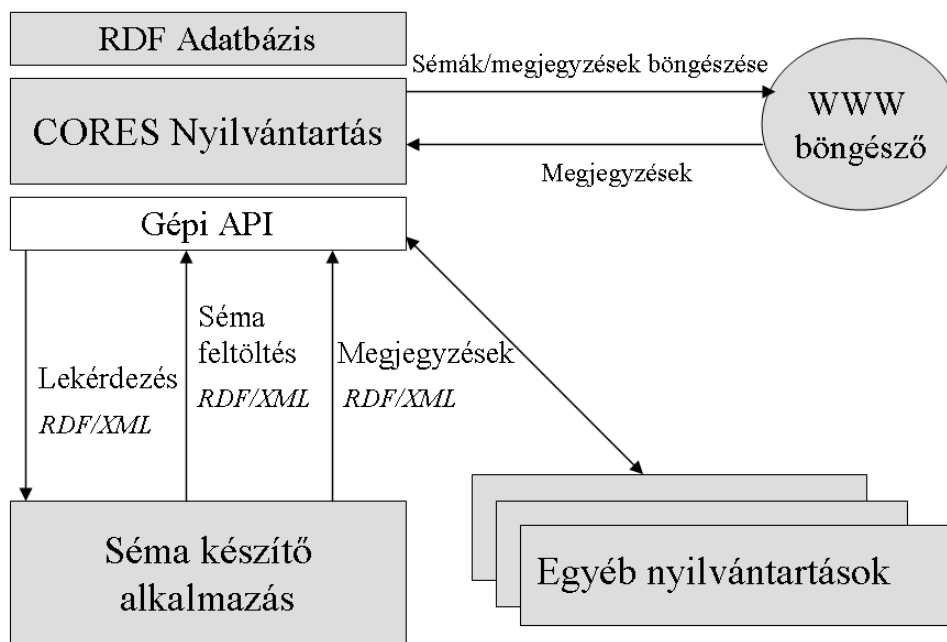
lehetőségük van új sémák létrehozására is, amelyben eddig nem létező elemeket is leírhatnak, majd felhasználhatnak a profiljukban.

A szürke részből kilógó entitások a felhasználói azonosításért és a jogosultságok kezeléséért felelősek, ezzel biztosítva az alapjait egy olyan metaadat nyilvántartásnak amelyet széles közönség használhat. A szervezetek sémáinak és profiljainak a karbantartását az adott szervezethez tartozó felhasználók végezhetik.

A modell másik kiegészítése az egyes entitásokhoz fűzhető megjegyzések, kérdések, amelyekben a felhasználók leírhatják tapasztalataikat, ötleteiket az adott sémáról, sémaelemről, és az azt használó közösség megvitathatja a felmerülő problémákat. A megjegyzéseken túl az egyes sémák fejlődése az adminisztratív metaadatok segítségével is nyomon követhetők, amelyekből kiderül, hogy ki és mikor módosított a sémán vagy profilon.

A rendszer felépítése

Hogy mindez működhessen, szükség van egy olyan nyilvántartásra, amelyben minél több ismert és elterjedt szabványos séma megtalálható, hogy a profilok készítői kedvükre válogathassanak. Szükség van egy olyan séma és profil szerkesztő eszközre, amely ezen nyilvántartásban képes keresni és lehetőséget biztosít az ott talált elemek újrafelhasználására. A profil elkészülte után pedig lehetőséget ad a szerkesztőnek, hogy a saját profilját is feltöltse. Ez a mechanizmus biztosítja, hogy mások is rátalálhassanak az adott profilra és azt ők is felhasználhassák: ezzel szorosabb együttműködést biztosíthatnak az alkalmazásaik között, maguknak pedig munkát takaríthatnak meg, hogy nem kell egy új profilt átgondolni és megcsinálni, hanem egy esetleg már jól bejáratottat használhatnak.



3. ábra: a CORES nyilvántartás architektúrája

Az elkészült rendszer architektúráis felépítése a 3. ábrán látható. A metaadat nyilvántartás egy RDF adatbázisra épül, ennek feladata az adatbázisában szereplő adatok és azok közötti kapcsolatok megjelenítése, lehetővé téve azok könnyű felderítését valamint az adatokon végezhető keresést. Eléréséhez csupán egy web böngésző szükséges. A nyilvántartás másik feladata egy programok számára elérhető felület (API) biztosítása, amelyen keresztül távoli alkalmazások számára lekérdezhető és kereshető a tartalom. Ez az ún. gépi API szolgál az új sémák feltöltésére is.

Az új sémák/profilok előállítására egy Java nyelven írt, platform független kliens program használható, amely képes a nyilvántartás API-val kommunikálni, ezzel lehetővé téve, hogy az új profilokat egy folyamatosan bővülő adatbázis alapján készíthessük el.

A nyilvántartás használata

Regisztráció

A nyilvántartásban való kereséshez, böngészéshez nem szükséges regisztráció. Mivel a nyilvántartás nem egy konkrét szervezet metaadat sémáinak tárolását célozza meg, hanem szervezetek minél szélesebb körét átfogó nyilvántartás létrehozása a célja, ezért nem megoldható, hogy egy központi személy/csoport foglalkozzon az adatok karbantartásával. E helyett a nyilvántartás az adatok karbantartását azok gazdáira bízta. Érthető okokból ezen adatok feltöltése és módosítása regisztrációhoz és megfelelő jogosultságokhoz kötött. A nyilvántartásban egyelőre bárki regisztrálhat új szervezetet, ebben az esetben ő lesz az adott szervezet adminisztrátora, és ő dönthet arról, hogy másoknak engedélyezi-e, hogy a szervezetéhez csatlakozzanak, ezzel jogot adva nekik a szervezet adatainak módosítására. Természetesen később lehetősége van tagok kizárására is. A szerveren létrehozott azonosítók felhasználása a kliensben is történhet, például amikor a kliensből sémákat töltenek fel a nyilvántartásba.

Böngészés

A nyilvántartás weboldalán a főmenüben választhatjuk ki, hogy melyik entitásoknál szeretnénk kezdeni a böngészést. Ezekről a rendszer felépítésénél már esett szó: ügynökségek, elem készletek, elemek, kódolási sémák, alkalmazási profilok vagy elem felhasználások. A későbbi könnyebb eligazodást biztosítandó ezek a linkek minden oldal láblécében is megtalálhatóak egyéb hasznos linkek társaságában. Ezek közül tetszőlegesen kiválasztva az azonos típusú entítások listázáshoz jutunk, ahol azok neve és a legfontosabb kapcsolódó adatok láthatók. A sor végén található a részletes megjelenítéshez vezető „Detail” link.

Element Set:: The Dublin Core Element Set v1.1	
ID	http://www.cores-eu.net/registry/reg/elementSet/dc
Name	The Dublin Core Element Set v1.1
Version	1.1
Creation date	1999-07-02
Status	DCMI recommendation
Description	The Dublin Core metadata vocabulary is a simple vocabulary intended to facilitate discovery of resources.
Classification	
Responsible Agency	The Dublin Core Metadata Initiative
Namespace Name/URI	
Namespace Prefix	
Specification	http://dublincore.org/usage/terms/terms-latest.html
Annotations	There are no annotations for this resource. Add new annotation.
Administrative metadata	List administrative metadata for this resource. (0)

4. ábra: Egy elemkészlet megjelenítése a nyilvántartásban

A részletes megjelenítésnél (4. ábra) az adott entáshoz kapcsolható összes metaadatot láthatjuk: link a részletes specifikációra, verziószám, dátum, az adott elemért/sémáért felelős ügynökség, stb. Emellett találhatunk az entítások közötti navigálást lehetővé tévő linkeket, táblázatos formában összefoglalva: sémák, profilok elemei, kapcsolódó kódolási sémák, finomítások, elem felhasználások, stb.

Keresés

Ha nem tudjuk az adott elem pontos helyét, és nem akarunk a kapcsolatok között elveszni, természetesen lehetőség van az entítások keresésére is. A kereső oldal a főmenüből érhető el, és az entítások minden tulajdonsága között keres. Az eredmények az entítások listázásánál használt táblázattal azonos formában jelennek meg, ahonnan a „Detail” linkkel léphetünk tovább a részletes leíráshoz.

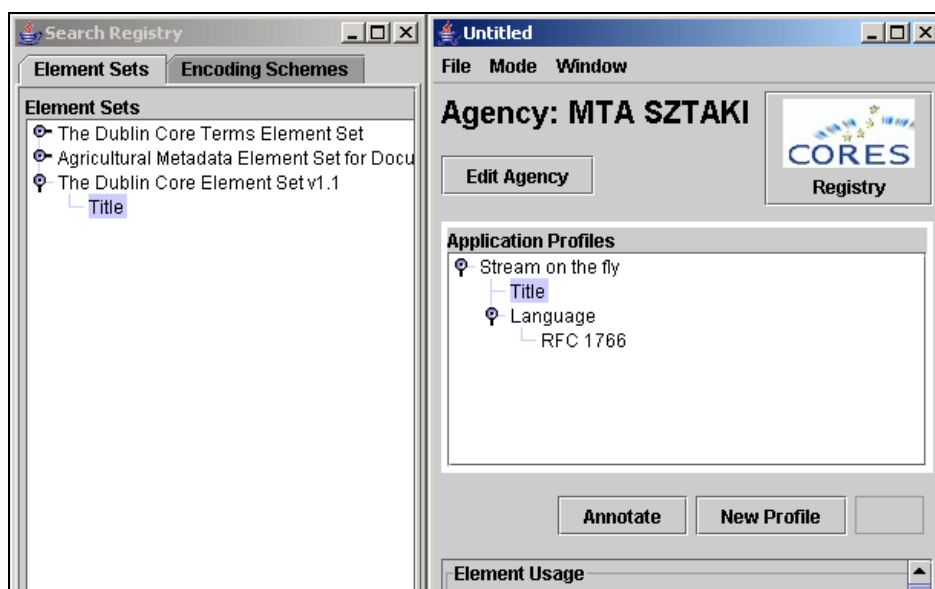
Megjegyzések

Az adott sémák szakértői, felhasználói számára lehetőség van az egyes elemek, profilok használata során felmerült ötletek, szakmai kérdések megvitatására, kommentek hozzáfűzésére. Ez a művelet is regisztrációhoz kötött. A megjegyzések az adott entitás részletes oldaláról érhetők el.

A sémakészítő kliens használata

Amennyiben a nyilvántartást böngészve mégsem találunk számunkra megfelelő sémát, abban az esetben szükséges lehet, hogy ezt magunk készítsük el. Erre egy kliens alkalmazás szolgál, amely képes kommunikálni a nyilvántartással az interneten keresztül, így lehetőség van a tartalmának újrafelhasználására és az elkészített sémák visszatöltésére (5. ábra).

Az elkészíthető sémák három csoportra oszthatók: alkalmazási profilok, elem készletek és kódolási sémák. Egy átlagos felhasználónak valószínűleg csak az alkalmazási profilok készítésére lesz szüksége ezért a kliens egyszerű módban indul, azaz csak profilok készítésére van lehetőség. Elem készletek és kódolási sémák készítése már a szakértők feladata, mivel az jóval több egyeztetést és kompromisszumot igényel az adott téma felhasználói részéről. Ezek szerkesztésére a szakértő módban van lehetőség.



5. ábra: A sémakészítő kliens kereső és szerkesztő ablaka

Mint arról már feljebb is esett szó az alkalmazási profilok nem definiálnak új elemeket, csak már meglévő sémák elemeit használják fel elem felhasználások képében, esetleg azokhoz kódolási sémát rendelnek. Ennek érdekében a felhasználónak lehetősége van a nyilvántartásban keresni. A kliens a nyilvántartás API-ján keresztül letölti a találatok definícióját, majd a felhasználó a megfelelő találatot drag-and-drop módszerrel a profiljába húzhatja. Ezzel létrehoz egy elem felhasználást, amely az adott elemre mutat. A kliensben a drag-and-drop műveletek természetesen nem csak a keresés eredményeire alkalmazhatók, szakértő módban lehetőség van a dokumentum részei közötti használatára is.

Sémák létrehozása

Új elem készletek, kódolási sémák létrehozásához már nem elég a keresések eredményeinek átmásolása, ilyenkor teljesen ki kell tölteni a sémát definiáló táblázatot. Új elemek létrehozásakor lehetséges már létező elemek finomítása. Ez a kapcsolat szintén drag-and-drop módszerrel adható meg az adott elem részletes adatainál.

Munkánk során nagy segítség lehet a már mások által elkészített sémák újrafelhasználása. A profilunk már bejárattott, széles körben elterjedt sémára építésével sok felesleges munkától kímélhetjük meg magunkat,

egyben biztosíthatjuk az alkalmazásunk bizonyos szintű együttműködésének lehetőségét az adott sémára építő többi alkalmazással.

Az újrafelhasználás módja, hogy a keresés funkcióval megkeressük a nyilvántartásban a szükséges elemet, kódolási sémát és azt drag-and-drop módszerrel a saját sémánk megfelelő helyére húzzuk: elem felhasználásnak, finomításnak vagy kódolási sémának.

```
<rdf:Description rdf:about='http://www.renardus.org/#elementUsage-10'>
  <rdf:type rdf:resource='http://www.cores-eu.net/registry/reg/ElementUsage' />
  <NS0:uses rdf:resource='http://purl.org/dc/elements/1.1/language' />
  <rdfs:label>Language | ISO639-2</rdfs:label>
  <NS0:useComment>The language code is the ISO 639-2, three letter code. SUB will provide a mapping
  between the two letter and three letter language code but this will also be found on the LoC site - ISO 639-
  2: http://lcweb.loc.gov/standards/iso639-2/englangn.html. Best Practice: Use codes rather than text.
  Recommended means if there is any spoken or written text, supply.</NS0:useComment>
  <NS0:obligation>Recommended</NS0:obligation>
  <NS0:maximumOccurrence>Unbounded</NS0:maximumOccurrence>
  <NS0:isUsageIn rdf:resource='http://www.renardus.org/#applicationProfile-1' />
  <NS0:associatedEncodingScheme rdf:resource = 'http://purl.org/dc/terms/ISO639-2' />
  <rdfs:isDefinedBy rdf:resource = "http://www.cores-eu.net/registry/schema/renardus/renap.xml" />
</rdf:Description>
<rdf:Description rdf:about='http://purl.org/dc/elements/1.1/language'>
  <rdf:type rdf:resource='http://www.w3.org/1999/02/22-rdf-syntax-ns#Property' />
  <rdfs:label>Language</rdfs:label>
  <NS0:isElementOf rdf:resource='http://www.cores-eu.net/registry/reg/elementSet/dc' />
</rdf:Description>
<rdf:Description rdf:about='http://purl.org/dc/terms/ISO639-2'>
  <rdf:type rdf:resource='http://www.cores-eu.net/registry/reg/EncodingScheme' />
  <rdfs:label>ISO 639-2</rdfs:label>
</rdf:Description>
```

6. ábra: Példa a kliens által készített RDF sémára (egy elemfelhasználás)

Mentés és feltöltés

Az elkészült sémákat mindenképpen érdemes a helyi gépre is elmenteni, mert egyelőre nincs mód a sémák visszatöltésére a nyilvántartásból, így azok későbbi módosítása csak az elmentett helyi példány alapján lehetséges.

Természetesen az egész rendszer fő célja, hogy az új sémákat ne csak helyi használat céljából mentjük el, hanem azokat feltöltés által tegyük közzé a nyilvántartásban, ezzel segítve mások munkáját és a saját sémánk fejlődését is! A rendszer hitelességét biztosítandó a feltöltés a megfelelő jogosultsághoz kötött. A regisztrálás a nyilvántartás honlapján tehető meg, ahol a felhasználó kiválaszthatja, hogy mely szervezethez tartozik, vagy hozhat létre újat is. Feltöltéskor az azonosítót megadva a rendszer meghatározza, hogy mely szervezethez tartozik az adott felhasználó és ezt egyezteteti a feltöltendő adatokat karbantartó szervezettel. Amennyiben a kettő megegyezik, akkor engedélyezi a feltöltést.

Összegzés

A metaadatsémák egyre szélesebb körű alkalmazása előbb vagy utóbb kikerülhetetlenné teszi egy általános módszertan létrehozását, amely megoldja a sémák egységes modell szerinti reprezentációját, a sémák kapcsolatainak és összefüggéseinek kezelését, a sémák világának áttekinthetőségét. Egy ilyen módszertan

számos előnnyel kecsegtet. Többek között könnyebbé válik új sémák létrehozása az újrafelhasználás révén. Az újrafelhasználásnak és a szigorú modellnek köszönhetően egységesebbé, együttműködőbbé válnak a használatban lévő sémák, amely viszont konkrét gazdasági előnyökhöz, megtakarításokhoz is vezethet akár országos szinten is. Végül, a modell alapján működő nyilvántartások megteremtik a sémák áttekintésének, a sémák közti eligazodásnak a lehetőségét.

Az e cikkben ismertetett megoldás egy lehetséges módja a fenti célok elérésének. A modell előnye, hogy a Szemantikus Web eszközeit használja fel a sémák modellezésére, ezzel mintegy összekapcsolva a Szemantikus Web és a metaadatok világát. Ez a kapcsolat a Szemantikus Web hasznára is válhat, mivel ott is szükség van minél nagyobb mennyiségű és egységes metaadatra, amelyet az újfajta szolgáltatások bemeneti adatként használhatnak.

A CORES metaadat nyilvántartás folyamatosan üzemel a <http://cores.dsd.sztaki.hu> címen, ahol böngészhető a jelenlegi nyilvántartás, valamint rendelkezésre áll egy teszt nyilvántartás is, melyben a séma szerkesztés és feltöltés is kipróbálható.

Köszönet

A szerzők ezúton köszönik a CORES projekt tagjainak segítségét (PricewaterhouseCoopers, Fraunhofer-Gesellschaft, UKOLN, MTA SZTAKI Elosztott Rendszerek Osztály). A CORES projektet az EU támogatta.

¹ CORES projekt honlapja: <http://www.cores-eu.net/>

² Szemantikus Web: <http://www.w3c.org/2001/sw/>

³ Semantic Web Layers: <http://www.w3c.org/2001/12/semweb-fin/w3csw>, 3. pont

⁴ RDF: <http://www.w3.org/RDF/>

⁵ RDFS: <http://www.w3.org/TR/rdf-schema/>

⁶ OWL: <http://www.w3.org/TR/owl-ref/>

⁷ RDF: Concepts and Abstract Syntax: <http://www.w3.org/TR/rdf-concepts/>

⁸ RDF/XML Syntax Specification: <http://www.w3.org/TR/rdf-syntax-grammar/>

⁹ Dublin Core Metadata Initiative: <http://www.dublincore.org/>

¹⁰ DC 1.0: <http://dublincore.org/documents/1998/09/dces/>

¹¹ DC 1.1: <http://dublincore.org/documents/dces/>

¹² ISO 15836:2003: Information and documentation - The Dublin Core metadata element set

¹³ Dublin Core Qualifiers: <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>

¹⁴ DCMI Metadata Terms: <http://dublincore.org/documents/dcmi-terms/>

¹⁵ Guidelines for implementing Dublin Core in XML: <http://dublincore.org/documents/dc-xml-guidelines/>

¹⁶ Expressing Simple Dublin Core in RDF/XML: <http://dublincore.org/documents/dcmes-xml/>

¹⁷ Expressing Qualified Dublin Core in RDF/XML: <http://dublincore.org/documents/dcq-rdf-xml/>

¹⁸ DCMI Schemas: <http://www.dublincore.org/schemas/>

¹⁹ Expressing Dublin Core in HTML/XHTML meta and link elements: <http://dublincore.org/documents/dcq-html/>

²⁰ Shared Online Media Archive: SOMA Metadata Element Set: http://soma-dev.sourceforge.net/SOMA_Metadata_1.htm

²¹ EBU Core Metadata Set for Radio Archives: http://www.ebu.ch/trev_t3293.pdf

²² The Warwick Framework: <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>

²³ Application profiles: mixing and matching metadata schemas: <http://www.ariadne.ac.uk/issue25/app-profiles/>

²⁴ DCMI Usage Board Review of Application Profiles: <http://dublincore.org/usage/documents/profiles/>

²⁵ Open Archives Initiative: <http://www.openarchives.org/>

²⁶ KIKERES: <http://www.kikeres.hu/>

²⁷ NDA/NAVA: http://www.ihm.hu/programok/nda_nava/

²⁸ HEKTÁR: <http://hektar.sztaki.hu/>

²⁹ <http://radio.sztaki.hu>

³⁰ DCMI Registry: <http://dublincore.org/dcregistry/>

³¹ CORES metaadat nyilvántartás: <http://cores.dsd.sztaki.hu>

³² Rachel Heery, Pete Johnston, Csaba Fülöp, András Micsik: *Metadata schema registries in the partially Semantic Web: the CORES experience*, http://www.siderean.com/dc2003/102_Paper29.pdf