

Beszámoló egy, a Katalisten közzétett levelezés alapján

Péter Király írta, 2017. június 19. 15:27
<kirunews@gmail.com>:

Tisztelt Katalist,
egy bő lére eresztett prezentáció ismertetője következik. Herbert van de Sompel, az OAI-PMH, OAI-ORE, SFX link resolver, ResourceSync, Open Annotations, Memento és más jószágok kitalálójátárszerzője ma publikálta Signposting the Scholarly Web: An Overview című előadásának diáit <https://www.slideshare.net/hvdsomp/signposting-overview>. Az általa javasolt technológia a tudományos weboldalak (pl. intézményi vagy országos repozitóriumok, kutatási adat-archívumok, folyóiratok) információinak összekapcsolásában segít.

Ezek az oldalak tipikusan a következő összetevőkből állnak:

- full text szövegek (néha több formátumban, helyben vagy linkelve)
- bibliográfiai leírás(ok) (néha szintén több formátumban, pl. Bibtex, RIS, JSON-LD stb.)
- szerzők oldalai (pl. ORCID, VIAF, authority rekordok oldalai)
- perzisztens URL (pl. URN, PID, DOI, Handle)
- külső bibliográfiai leírások (pl. a Crossref.org-on)
- „landing page” (az a felső szintű oldal, amire pl. egy perzisztens URL mutat, és ami opcionálisan HTML linkeket biztosít az egyes összetevőkre, lásd <http://mek.oszk.hu/00000/00075/>- a MEK-ben található OSZK Tezaurusz főoldala, innen el lehet érni mind a bibliográfiai leírást több változatban, a tartalom különféle formátumú változatait, és sok minden mást).

A Signposting a HTTP fejrészét (Header) használja ezen információk összekapcsolására. Amikor a weben a HTTP protokoll segítségével egy kérést küldünk egy webszervernek, akkor a válasz két részből áll. A válasz eleje az ún. header metaadatokat tartalmaz a szóban forgó erőforrásról, például mi a tartalom típusa (mime content type), mekkora a mérete, az utolsó módosítás dátuma, a gyorstár elévülésének ideje, a

webszerver típusa stb. és csak e fejrész után jön a tulajdonképpeni tartalom például egy HTML oldal, egy PDF fájl, vagy egy TIFF kép. A böngésző (vagy más kliens) a header alapján jeleníti meg számunkra a dokumentumot, vagy ajánlja fel a letöltést. De a header információit sok más eszköz használja: keresőrobotok, tartalombegyűjtő és – elemző robotok, stb. Szinte minden programnyelv tartalmaz eszközkészletet ezen információk olvasására és írására. A metaadatoknak van egy szabványos sémájuk, aminek karbantartását az Internet Engineering Task Force (IETF) nevű testület vállalta. Ez a séma az újabb igények tükrében folytonosan változik, 2010-ben javasolták a „Link” elem felvételét, ami a „Web Linking” szabvány néven híresült el. Ez a szabvány azt teszi lehetővé, hogy linkeket biztosítsunk a tetszőleges URL-ekhez a HTTP headeren belül. Lássunk egy példát:

A <http://dbpedia.org/data/Reykjavik> URL a következő linkeket biztosítja

Link: <<http://creativecommons.org/licenses/by-sa/3.0/>>;

```
rel="license",
<http://dbpedia.org/data/Reykjavik>;
rel="alternate";
type="application/rdf+xml";
title="Structured Descriptor Document (RDF/XML format)",
<http://dbpedia.org/data/Reykjavik>;
rel="alternate";
type="text/n3";
title="Structured Descriptor Document (N3/Turtle format)",
<http://dbpedia.org/data/Reykjavik>;
rel="alternate";
type="application/json";
title="Structured Descriptor Document (RDF/JSON format)",
<http://dbpedia.org/data/Reykjavik>;
rel="alternate";
type="application/atom+xml";
title="OData (Atom+Feed format)",
```

```
<http://dbpedia.org/data/Reykjavik>;
rel="alternate";
type="application/odata+json";
title="OData (JSON format)";
<http://dbpedia.org/page/Reykjavik>;
rel="alternate";
type="text/html";
title="XHTML+RDFa";
<http://mementoarchive.lanl.gov/dbpedia/timegate
e/http://dbpedia.org/data/Reykjavik>;
rel="timegate"
```

Az elem szintaxisa egyszerű: a linkeket kacsacsőrben kell megadni, egymástól vesszővel elválasztva. A linkekhez opcionálisan metaadatok kapcsolhatók, ezeket kettősponttal különítjük el. A fenti példában háromféle metaadattal találkozunk: A „rel” a link típusát jelenti. A „license” jelentése, hogy a link a tartalom licencinformációjára mutat. A „alternate” egy alternatív dokumentumtípusra mutat, a „timegate” pedig a Memento link (ez egy sajátos időbélyeg, szintén Van de Sompel találmánya, lásd <http://mementoweb.org/about/>). A „type” jelöli a MIME (vagy dokumentum-) típust, a „title” pedig egy emberi fogyasztásra szánt címke.

A Web Linking szintaxisa és általános szemantikája tehát 2010 óta adott, a Signposting ezt a keretet egészíti ki a tudományos weboldalak elemei összekapcsolásának szemantikájával, úgy hogy a "rel" attribútum értékészletét további lehetőségekkel gazdagítja a következő típusokkal:

author – szerző
collection – a cikket alkotó fájlok összessége (cikk szövege, ábrák, kiegészítő anyagok stb.)
describedby – a dolgot leíró bibliográfiai rekord
describes – annak a dolognak az azonosítója amiről a bibliográfiai leírás szól
item – a gyűjtemény egy darabja
identifier – az azonosító
type – típus (schema.org vagy más névtérből vett jelölő)

Mivel a signposting létező szabványokra épít a tartalomtípus tekintetében a MIME szabvány az irányadó, amiben szerepel néhány bibliográfiai formátum:

BibTeX – „application/x-bibtex”
CiteProc JSON – „application/vnd.citationstyles.csl+json”
RIS – „application/x-research-info-systems”

Azonban számos bibliográfiai szabvány formátuma az XML („application/xml”), amiket nem lehet a MIME-on belül megkülönböztetni, ezért a szerzők egy további attribútum, a „profile” használatát javasolják ilyen esetekben (a type kiegészítéseként és nem helyett):

MARC XML – „<http://www.loc.gov/MARC21/slim>”
MODS – „application/xml <http://www.loc.gov/mods/>”
Simple Dublin Core – „application/xml <http://purl.org/dc/elements/1.1/>”
Qualified Dublin Core – application/xml <http://dublincore.org/documents/dcmi-terms/>”

Mivel mindez így meglehetősen absztrakt, lássunk néhány típuspéldát. A példákban a nyíl jelentése: A erőforrásról szeretnénk linket biztosítani B erőforrásra, vagyis Link-kel kezdődő sorok az A erőforrás HTTP headerjében szerepelnek.

Szerző:

Viaf azonosító linkelése:

```
Link: <http://viaf.org/viaf/31835361/>
rel="author"
```

Több szerző ORCID oldalának linkelése:

```
Link: <http://orcid.org/0000-0002-0715-6126>
rel="author",
<http://orcid.org/0000-0003-3749-8116>
rel="author"
```

Bibliográfiai metaadatok linkelése:

Perzisztens azonosító [UTF-8?]^â

Bibliográfiai leírás

```
Link: <https://doi.org/10.1007/s10958-016-2844-8>
rel="describedby"
type="application/vnd.citationstyles.csl+json"
```

Bibliográfiai leírás [UTF-8?]^â

Perzisztens azonosító

```
Link: <https://doi.org/10.1007/s10958-016-2844-8>
rel="describes"
```

Landing page [UTF-8?]^â

Bibliográfiai leírás

```
Link: <http://citation-needed.services.springer.com/v2/references/10.1007/s10958-016-2844-8?format=refman&flavour=citation>
rel="describedby"
type="application/x-research-info-systems",
```

<<http://citation-needed.services.springer.com/v2/references/10.1007/s10958-016-2844-8?format=bibtex&flavour=citation>>
rel="describedby"
type="application/x-bibtex",
<<https://doi.org/10.1007/s10958-016-2844-8>>
rel="describedby"
type="application/vnd.citationstyles.csl+json"

Teljes szöveg [UTF-8?]

Bibliográfiai leírás

Link:

<<http://journals.plos.org/plosone/article/citation/ris?id=10.1371%2Fjournal.pone.0115253>>
rel="describedby"
type="application/x-research-info-systems",
<<http://journals.plos.org/plosone/article/citation/bibtex?id=10.1371%2Fjournal.pone.0115253>>
rel="describedby"
type="application/x-bibtex",
<<https://doi.org/10.1371/journal.pone.0115253>>
rel="describedby"
type="application/vnd.citationstyles.csl+json"

Landing page/cikk [UTF-8?]

Perzisztens azonosító

Link: <<https://doi.org/10.1017/jns.2015.28>>
rel="identifier"

Szerzői oldal (pl. egyetemi honlapon) [UTF-8?]

Szerzői azonosító

Link: <<https://orcid.org/0000-0003-1419-2405>>
rel="identifier"

Landing page [UTF-8?]

Link:

<<http://nar.oxfordjournals.org/content/44/9/4005.full>>
rel="item"
type="text/html",
<<http://nar.oxfordjournals.org/content/44/9/4005.full.pdf>>
rel="item"
type="application/pdf",
<<http://nar.oxfordjournals.org/content/suppl/2016/04/12/gkw229.DC1/nar-00308-f-2016-File002.pdf>>
rel="item"
type="application/pdf"

Cikk különféle formátumokban illetve kiegészítő anyagok [UTF-8?]

Link:

<<http://nar.oxfordjournals.org/content/early/2016/06/07/nar.gkw501>>
rel="collection"

Perzisztens azonosító típusának megadása

Cikk:

Link: <<http://schema.org/ScholarlyArticle>>
rel="type"

Adattár:

Link: <<http://schema.org/Dataset>>
rel="type"

Szerző:

Link: <<http://schema.org/Person>>
rel="type"

A javaslat banalitása ellenére eme szétszórt információk összegyűjtése nem triviális feladat. Saját tapasztalatból tudom, hogy a vonatkozó információk összeszedése egyetemi méretben igencsak nehéz és manuálisan szinte lehetetlen feladat, nem beszélve az elérhető információk minőségéről. Még a nagyon drága fizetős adatbázisokban (Web of Science, Scopus) is találkozunk nagyon komoly bibliográfiai hiányosságokkal, rossz minőségű adatokkal. A Signposting egy komoly infrastrukturális előrelépés, ami segíti az adatszolgáltatókat, hogy a felhasználók könnyebben elérhessék az adataikat. A MEK esetében úgy látom, hogy adatszinten minden adott az implementációhoz, ugyanakkor pont az ő esetükben, tekintve hogy ezeket az adatokat számos egyéb csatornán is eljuttatják a felhasználóhoz, kevesebb a gyakorlati haszna, mint teszem azt az MTMT vagy más tudományos repozitórium esetében.

Nem tudom, hogy van-e magyar szolgáltatás, ami használja a technológiát, ha van, akkor örömmel látnám a készítőik tapasztalatát.

Bővebb információ:

<http://signposting.org/>

Irodalom:

* Van de Sompel, H., and Nelson, M.L. (2015) Reminiscing About 15 Years of Interoperability Efforts. D-Lib Magazine, 21(11/12). DOI:10.1045/november2015-vandesompel, <https://doi.org/10.1045/november2015-vandesompe>

* Van de Sompel, H., Rosenthal, D., and Nelson, M.L. (2016) Web Infrastructure to Support e-Journal Preservation (and More). Arxiv preprint. arXiv:1605.06154 ;
<http://arxiv.org/abs/1605.06154>

* Van de Sompel, H., and Nelson, M.L. (2015) Achieving Meaningful Interoperability for Web-Based Scholarship.
<http://www.slideshare.net/hvdsomp/interoperability-for-web-based-scholarship>

A legjobbakat!
 Király Péter

--

Péter Király
 software developer

GWGD, Göttingen – Europeana – eXtensible
 Catalog – The Code4Lib Journal
<http://linkedin.com/in/peterkiraly>

☺☺☺☺☺

Horváth Ádám hozzászólása

Adam HORVATH <horvadam@gmail.com>

Címzett: Katalist katalist@listserv.niif.hu

Tárgy: Re: [KATALIST] Signposting: tudományos weboldalak HTTP alapú szemantikus linkelése

Kedves Péter,

a *Magyar Nemzeti Múzeum Központi Könyvtára* az online olvasói katalógusában a BIBFRAME leírással is rendelkező rekordok esetén, mint például ennél a rekordnál:
<http://link.hnm.hu/lib/bib/MNMKVT19563>

az alábbi link tageket helyezi el. Ezt a megoldást egyébként még mint "link discovery service" ismerem meg. A cél az lenne, hogy a robotokat átírányítsuk a géppel jobban olvasható formátumok felé. Nem tudom, hogy a robotok (pl. Google) használja-e? Erre nincs egyértelmű bizonyíték. Valószínűleg igen. A link tageket a Mongúz Kft. helyezte el az oldalainkon. És akkor itt vannak az általunk használt link tagek, melyek mind az adott MARC rekord különböző formátumú BIBFRAME-es verzióhoz vezetnek el:

```
<link rel="alternate" type="application/rdf+xml"
title="Structured Descriptor Document (RDF/XML
format)"
```

```
href="http://data.hnm.hu/sparql?query=define%20s
ql%3Adescribe-
mode%20%22LOD%22%20%20DESCRIBE%20%
3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%
2Flib%2Fbib%2FInstance%2FMNMKVT19563%3
E&output=application%2Frdf%2Bxml" />
```

```
<link rel="alternate" type="text/n3" title="Structured
Descriptor Document (N3/Turtle format)"
href="http://data.hnm.hu/sparql?query=define%20s
ql%3Adescribe-
mode%20%22LOD%22%20%20DESCRIBE%20%
3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%
2Flib%2Fbib%2FInstance%2FMNMKVT19563%3
E&output=text%2Fn3"/>
```

```
<link rel="alternate" type="application/rdf+json"
title="Structured Descriptor Document (RDF/JSON
format)"
href="http://data.hnm.hu/sparql?query=define%20s
ql%3Adescribe-
mode%20%22LOD%22%20%20DESCRIBE%20%
3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%
2Flib%2Fbib%2FInstance%2FMNMKVT19563%3
E&output=application%2Frdf%2Bjson"/>
```

```
<link rel="alternate" type="application/atom+xml"
title="Structured Descriptor Document (OData/
Atom format)"
href="http://data.hnm.hu/sparql?query=define%20s
ql%3Adescribe-
mode%20%22LOD%22%20%20DESCRIBE%20%
3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%
2Flib%2Fbib%2FInstance%2FMNMKVT19563%3
E&output=application%2Fatom%2Bxml"/>
```

```
<link rel="alternate" type="application/odata+json"
title="Structured Descriptor Document (OData/
JSON format)"
href="http://data.hnm.hu/sparql?query=define%20s
ql%3Adescribe-
mode%20%22LOD%22%20%20DESCRIBE%20%
3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%
2Flib%2Fbib%2FInstance%2FMNMKVT19563%3
E&output=application%2Fodata%2Bjson"/>
```

```
<link rel="alternate" type="text/csv" title=
"Structured Descriptor Document (CSV format)"
href="http://data.hnm.hu/sparql?query=define%20s
ql%3Adescribe-
mode%20%22LOD%22%20%20DESCRIBE%20%
3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%
2Flib%2Fbib%2FInstance%2FMNMKVT19563%3
E&output=text%2Fcsv"/>
```

```
<link rel="alternate" type="application/microdata+json" title="Structured Descriptor Document (Microdata/JSON format)" href="http://data.hnm.hu/sparql?query=define%20%3A%3Adescribe-mode%20%22LOD%22%20%20DESCRIBE%20%3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%2Flib%2Fbib%2FInstance%2FMNMKVT19563%3E&amp;output=application%2Fmicrodata%2Bjson" />
```

```
<link rel="alternate" type="text/html" title="Structured Descriptor Document (HTML+Microdata format)" href="http://data.hnm.hu/sparql?query=define%20%3A%3Adescribe-mode%20%22LOD%22%20%20DESCRIBE%20%3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%2Flib%2Fbib%2FInstance%2FMNMKVT19563%3E&amp;output=text%2Fhtml"/>
```

```
<link rel="alternate" type="application/ld+json" title="Structured Descriptor Document (JSON-LD format)" href="http://data.hnm.hu/sparql?query=define%20%3A%3Adescribe-mode%20%22LOD%22%20%20DESCRIBE%20%3Chttp%3A%2F%2Fdata.hnm.hu%2Fid%2Fcolls%2Flib%2Fbib%2FInstance%2FMNMKVT19563%3E&amp;output=application%2Fld%2Bjson"/>
```

Üdvözlettel:
Horváth Ádám

☺☺☺☺☺☺

Ungváry Rudolf hozzászólása

Dátum: Mon, 19 Jun 2017 18:15:06 +0200
Küldő: [Ungváry Rudolf <ungvaryr@gmail.com>](mailto:ungvaryr@gmail.com)
Címzett: "Adam HORVATH" <horvadam@gmail.com>, "Katalist" <katalist@data.hnm.hu>
Tárgy: Re: [KATALIST] Signposting: tudományos weboldalak HTTP alapú szemantikus linkelése

Kedves Péter,
(megint) elismerés a minőségi tálalásért.

Attól tartok, hogy ami a szemantikus web és kapcsolódó részeinek műszaki megvalósítását illeti, mintha lassan a megoldások bőségének zavarával kellene majd szembenézni.

Ami persze nem rossz.

Ungváry Rudolf

☺☺☺☺☺☺

Király Péter válasza

Dátum: Mon, 19 Jun 2017 22:31:51 +0200
Küldő: [Péter Király <kirunews@gmail.com>](mailto:kirunews@gmail.com)
Tárgy: Re: [KATALIST] Signposting: tudományos weboldalak HTTP alapú szemantikus linkelése

Kedves Ádám,

Nagyon köszönöm a választ és egyúttal gratulálok a linkek implementálásához is! Bevallom, erről a fejlesztésről eddig nem tudtam.

A HTML fejrészébe ágyazott link régebb óta használatos technológia mint a HTTP linkek, és a szintaxis meg is egyezik. Tudomásom szerint a keresőrobotok használják – egyéb technológiák mellett – ezt is. A Google esetében a Google Analytics segítségével meg lehet nézni, hogy mit is indexelnek, és ha adott prezentációtípus csak ilyen linkeken keresztül érhető el, akkor az bizonyíték a használatra.

Annak, hogy a tartalom különféle reprezentációira is lehessen hivatkozni többféle neve és megvalósítási formája van. A szemantikus webes megoldásokban leginkább a „content negotiation” kifejezéssel szoktam találkozni, amire az általad is népszerűsített Cool URI is épül. Ennek lényege, hogy a formai és nyelvi változatoknak ugyanaz az URI-jük és a HTTP kérés fejrészének kísérőinformációin keresztül mondhatjuk meg, hogy mely formai vagy nyelvi változatot szeretnénk elérni.

Például a következő kérésben:

```
GET /people/alice HTTP/1.1
Host: www.example.com
Accept: text/html, application/xhtml+xml
Accept-Language: en,
```

de az Accept sorral állítjuk be a formai, az Accept-Language sorral a nyelvi preferenciáinkat. Ámde mivel ezt a böngészőben nem könnyű szabályozni, a webszervereken be szoktak állítani valamilyen URL sablont, aminek a segítségével a felhasználók az URL-en belül tudják szabályozni ezeket a változatokat. (A Cool URI dokumentum leír néhány technológiát a probléma kezelésére: <https://www.w3.org/TR/cooluris/>).

A Signpostban két újdonság van:

- 1) nem a HTML, hanem a HTTP rétegben teszi elérhetővé a linkek felsorolását, így képek vagy más bináris állományok esetében is lehetőség van a kontextus megértését segítő linkeket megadására.
- 2) A linkek szemantikáját kifejezetten könyvtári terminusokkal bővíti.

A Signpost továbbá a példák segítségével propagálja (de nem írja elő) a cool URI-k használatát, ami például a Nemzeti Múzeumi dokumentumok esetében azt jelentené, hogy a link nem hordozná

a jelenlegi szolgáltatási technológia sajátosságait, pl. nem egy SPARQL (vagy egyéb) keresőkérdés lenne, hanem egy azonosító.

A legjobbakat!
Király Péter

--

Péter Király
software developer

GWGD, Göttingen – Europeana – eXtensible
Catalog – The Code4Lib Journal
<http://linkedin.com/in/peterkiraly>