

Filozofická fakulta Masarykovy univerzity

Signposting the Scholarly Web

Esej do předmětu Digitální knihovny

Bernadeta Kurešová
PV070: Digitální knihovny
27. 11. 2017

Úvod

Cílem této práce je čtenáři přiblížit technologii Signposting the Scholarly Web, stručně popsat, jak nástroj funguje a k čemu jej lze využít, případně o něm podat další zajímavé informace. Poté se pokusím zhodnotit naplnění cílů, které si projekt vytyčil a jeho obecný přínos. V závěru práce mohou použité zdroje nasměrovat čtenáře k dalším informacím o metodice Signpostingu. Webové stránky projektu (pouze v angličtině) je možné najít na odkaze: <http://signposting.org/>.

O technologii Signposting

Na vědeckých stránkách najdeme často bibliografické údaje, odkazy na další zdroje, informace o autorovi apod. Lidé tyto údaje snadno detekují, nicméně pro stroje je těžké je rozpoznat, protože weby používají k jejich zápisu různou konvenci.

Signposting je technologie, která pomáhá strojům detekovat a následně získat potřebné informace z vědecky zaměřených webových stránek. Příkladem stroje, který se snaží vytěžit z vědecké stránky informace, může být například indexátor obsahu Google.

Projekt byl spuštěn už před sedmnácti lety a jeho hlavními autory jsou Herbert Van de Sompel a Michael Nelson. Při své práci vycházeli autoři z projektů, které se snažily o něco podobného, pro příklad lze uvést třeba OAI-PMH či Memento. (OAI-PMH je protokol, který slouží ke sklizení metadat, Memento je protokol, který k HTTP přidává časový rozměr.)

Na stránkách projektu však autoři sami uvádějí, že Signposting není a nemá v plánu být nějakým velkým globálním projektem pokoušejícím se o standardizaci vědeckého webu. Cílem jejich snah je spíše shromáždit nápady lidí, kteří strávili spoustu času uvažováním o tom, jak zlepšit vědeckou komunikaci na webu a na základě těchto nápadů pak nabídnout drobné vychytávky (právě již zmiňované šablony) pro weby, které mohou strojům pomoci informace o člancích lépe extrahovat.

Tyto „vychytávky“ pak pojmenovávají jako jednotlivé šablony (patterns), pomocí kterých mohou provozovatelé stránek s vědeckými články zásadní informace sami označit tak, aby je pak bylo pro stroje jednoduché vyhledat a rozpoznat. Provozovatelům webů stačí, aby si do HTML kódu stránky zkopírovali několik řádků kódu vybrané šablony. Kód šablony se skládá z několika standardizovaných odkazů (typed links), pomocí kterých se daří strojům nalézt odpovědi na otázky, jako třeba: „Jaké je DOI této pdf publikace?“, „Kde najít zdroje?“, „Kde najdu ORCID autora této publikace?“¹⁾

Šablony

Šablony lze najít včetně jejich popisu na webových stránkách projektu. V nabídce si uživatel může vybrat z následujících:

- Author – šablona sloužící k tomu, aby na stránkách stroj snadno identifikoval autora (poskytuje standardizovanou formulaci pro záhlaví odkazů HTTP)

- Bibliographic Metadata – šablona napomáhá strojům vyhledávat bibliografická metadata
- Identifier – šablona specifikuje zdrojem použitý URI protokol. (URI, nebo-li Uniform Resource Identifier, česky jednotný identifikátor zdroje, je textový řetězec mající přesně danou strukturu, jenž se používá ke specifikaci zdroje informací především za účelem jejich použití pomocí počítačové sítě.)
- Publication Boundary – šablona, která pomáhá strojům detekovat zdroje, z nichž daný článek či publikace vycházejí
- Resource Type – tato šablona je sémantickým identifikátorem, popisuje téma daného článku či publikace. K tomu využívá sémantické typy z ontologie schema.org.

Fungování a implementace jednotlivých šablon jsou ještě na stránkách projektu přehledně vysvětleny pomocí vizuálních příkladů a podrobněji popsány v odkazovaných odborných článcích. Kromě toho je na webu poměrně přístupně vysvětlena odborná terminologie týkající se tohoto tématu.

Autoři projektu tak usnadňují potenciálnímu uživateli jejich nástrojů se v nich dobře zorientovat, pochopit jejich mechanismus a vybrat si ten, který se mu nejvíce hodí.

My se pro příklad blíže podíváme na šablonu Author.

Příklad šablony Author

Použití šablony typu Author na stránkách autoři ilustrují na této citaci, která může být například uvedena na vědecky orientovaných stránkách:

VAN DE SOMPEL, H., NELSON, M.L. Reminiscing About 15 Years of Interoperability Efforts. *D-Lib Magazine*, Volume 21, Number 11/12 <https://doi.org/10.1045/november2015-vandesompel>.

Pro dohledání autora(ů) na základě URI stránky <https://doi.org/10.1045/november2015-vandesompel> v databázi Orcid slouží HTML hlavička:

```
curl -I "https://doi.org/10.1045/november2015-vandesompel"
```

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

Obdobně lze dohledat autora(y) citované publikace:

ALNOAMANY, Y., WEIGLE, M.C., NELSON, M.L. Characteristics of social media stories. *Int J Digit Libr* (2016), 17:239. <http://link.springer.com/article/10.1007%2Fs00799-016-0185-3>.

Zde bude HTML hlavička vypadat takto:

```
curl -I "http://link.springer.com/article/10.1007%2Fs00799-016-0185-3"
```

```
Link: <http://orcid.org/0000-0003-4018-5153>  
; rel="author",  
<http://orcid.org/0000-0002-2787-7166>  
; rel="author",  
<http://orcid.org/0000-0003-3749-8116>  
; rel="author"
```

(Zdroj obrázků a příkladů je citován pod číslem 1.)

Cíle projektu

Hlavním cílem projektu je zjednodušit strojům vyhledávání informací o vědeckých člancích a zefektivnit tak způsob sdílení informací z vědeckých výzkumů. Realizace projektu spočívá implementaci jednotlivých šablon, čím více šablon je na data použito, tím více z nich pak stroje mohou při vyhledávání získat informací.

Pro úspěch projektu je klíčové, aby byly šablony využívány co nejširším spektrem uživatelů. V současné době vybrané šablony využívají weby DataCite Search, Pangaea, UCD Digital Library a DANS (Data Archiving and Networked Services).

Zhodnocení projektu

Abych dokázala projekt smysluplně zhodnotit, zajímala jsem se o to: – jak podle odborníků z daného oboru projekt naplňuje cíle, které si vytyčil, – jak široké je spektrum uživatelů této technologie (a o jak významné uživatele se jedná) – jak dobře se projekt sám prezentuje a jak je tím pádem přístupný potenciálním zájemcům. V závěru hodnocení se také dostanu k osobnímu pohledu na tuto technologii.

V prvé řadě jsem se zabývala vědeckými články a recenzemi. Narazila jsem například na velmi zajímavou recenzi Dr. Davida Rosenthala (ta mi mimo jiné pomohla blíže pochopit, jak technologie pracuje), který navštívil přednášku samotných autorů projektu. Podle něj projekt splňuje cíle, které si předsevzal. Dr. Rosenthal hodnotí Signposting the Scholarly Web velmi pozitivně, byť zmínil některá úskalí, která se musí při implementaci standardizovaných šablon v HTML nutně objevit.

Další podobně laicky přístupné recenze jsem nenašla, nicméně množství vědeckých článků zabývajících se projektem (navíc uvedených v prestižních vědeckých publikacích) ukazuje na to, že projektu se daří. Projekt byl také prezentován na významných vědeckých konferencích (v CERNu, FORCE11).

Jak už jsem zmínila, projekt v současnosti využívají DataCite Search, Pangaea, UCD Digital Library a DANS (Data Archiving and Networked Services). Jedná se o významné vyhledávače a shromaždiště vědeckých publikací, z čehož usuzuji, že projekt našel poměrně široké praktické využití.

Možná by bylo využití Signposting ještě širší, kdyby bylo více energie věnováno propagaci projektu. Například pokud jde o webovou prezentaci technologie, má podle mého názoru určité rezervy. Především chybí popis projektu v jiných jazycích než angličtině, šedý a velmi jednoduchý

design stránek s jedinou černobílou ilustrací není komerčně příliš přitažlivý. Nicméně je možné, že pro velké klienty z oblasti správy vědecký dat stránky v této podobě více, než dostačují.

Pokud jde o mé osobní hodnocení projektu, vidím ho jako velice přínosný, protože může pomoci propojit vědecké práce, sdružit ty, které se zabývají podobným tématem. Snadnější vyhledávání vědeckých informací může vést k razantnějšímu vědeckému pokroku, může ušetřit vědeckým pracovníkům spoustu práce a dát do souvislosti myšlenky z článků napsaných na jiném místě a v jiné době, které by jinak nikdo nespojoval.

Zdroje

Při zpracování eseje jsem čerpala z následujících zdrojů. Doporučuji je také k dalšímu vzdělávání v dané problematice.

- 1) Signposting the Scholarly Web [online]. 2017 [cit. 2017-11-26]. Dostupné z: <http://signposting.org/>
- 2) VAN DE SOMPEL, Herbert a Michael. L. NELSON. About 15 Years of Interoperability Efforts [online]. D-Lib Magazine, 2015 [cit. 2017-11-26]. Dostupné z: https://digitalcommons.odu.edu/cgi/viewcontent.cgi?referer=http://scholar.google.cz/&httpsr edir=1&article=1005&context=computerscience_fac_pubs
- 3) ROSENTHAL, David. Signposting the Scholarly Web. DSHR's Blog [online]. 2015 [cit. 2017-11-26]. Dostupné z: <http://blog.dshr.org/2015/12/signposting-scholarly-web.html>
- 4) DANS - Data Archiving and Networking Services. DANS - Data Archiving and Networking Services [online]. Netherlands, 2017 [cit. 2017-11-26]. Dostupné z: <https://dans.knaw.nl/en/about>
- 5) Příspěvatelé Wikipedie, Uniform Resource Identifier [online], Wikipedie: Otevřená encyklopedie, c2017, Datum poslední revize 4. 10. 2017, 12:40 UTC, [cit. 2017-11-26], Dostupné z: https://cs.wikipedia.org/w/index.php?title=Uniform_Resource_Identifier&oldid=15402344