

CrossRef – využití DOI pro citation-linking u vědeckých publikací

Jméno studenta, ročník, datum zpracování práce:

Lucie Fabriková, 2010, 7. 12. 2010

Název projektu/programu, jeho nositel, URL:

CrossRef, organizace CrossRef, crossref.org

Stručná charakteristika projektu/programu:

Hlavní funkcionalita CrossRef lze shrnout do dvou bodů: registrace DOI digitálního objektu a příslušných metadat, a služba pro vyhledávání DOI identifikátorů. Základní činností vyhledávačů v CrossRef je přeložit DOI identifikátor na URL adresu, která je zaregistrovaná s tímto DOI. Přístup k digitálnímu objektu plně kontroluje jeho vydavatel – každý vydavatel tedy musí implementovat mechanismy, kterými rozhodne o přístupu uživatele k digitálnímu objektu.

Doba řešení, aktuální stav:

Impulzem k vzniku organizace bylo uvedení projektu DOI-X na Frankfurtském knižním veletrhu v říjnu 1999. Tento projekt představil nový efektivní způsob vyhledávání článků, založený na DOI (digital object identifier) – unikátnímu identifikátoru přidělenému každému digitálnímu objektu, na který se může vztahovat citace.

Samotná organizace CrossRef vznikla v červnu roku 2000 a představuje oficiální registrační úřad identifikátorů DOI pro akademické a odborné publikace. Hlavním záměrem CrossRef je prostřednictvím spolupráce mezi vydavateli zajistit akademické komunitě jednodušší přístup k odborným materiálům, které jsou online dostupné.

Na projektu je soustavně pracováno již od jeho vzniku v roce 2000. Organizace CrossRef v současnosti registruje přes 37 mil. DOI. Dále má projekt širokou členskou základnu – v současné době (v prosinci 2010) je registrováno 1646 knihoven a tisíce akademických a profesionálních vydavatelů.

Cíle projektu:

Současný trend přesouvat veškeré tištěné publikace do elektronické podoby a následně je zobrazovat na Internetu představuje snadnou a rychlou dostupnost informací. Specializované informační servery nabízejí přístup k množství vědeckých článků, akademických publikací a další odborné literatuře. Nalezení příslušného článku nebo kapitoly v knize se postupně stává otázkou několika málo kliknutí, specializované vyhledávače typu Google Scholar umožní efektivní procházení relevantních stránek.

Přes všechny zřejmé výhody elektronického vydavatelství však Internet představuje sporný referenční materiál. Čteme-li si dnes na určité webové adrese zajímavý a přínosný text, neexistuje žádná záruka, že jej nalezneme na témže místě také zítra. Autor odborného článku, který se v seznamu použité literatury nebo seznamu citací opírá o odkazy na různé webové stránky, může být zaskočen – stejně jako čtenáři jeho článku, pokud se náhle změní URL adresy uváděných zdrojů anebo dojde k jejich přesunu.

Chceme-li výše uvedený problém řešit, jsme postaveni minimálně před následující problémy: Jak zajistit trvalou platnost citace? Jak jednoznačně identifikovat práci, kterou chceme citovat?

Odpověď nalezneme ve stanovách organizace CrossRef: „CrossRef si klade za cíl být důvěryhodnou organizací, podporující spolupráci v rámci široké členské základny; směřovatnou a inovativní v oblasti podpory persistentní, udržitelné struktury pro akademickou komunikaci.“¹

Popis projektu a jeho výsledků:

Na tomto místě si jistě zaslouží bližší pozornost DOI – digital object identifier, který představuje stěžejní prvek celého návrhu.

DOI je realizován jako alfanumerický řetězec, který jednoznačně identifikuje digitální objekt. V systému CrossRef je DOI spjat se základními metadaty popisujícími digitální objekt – zejména pak URL odkazujícím na plný text. Hlavními rysy DOI jsou persistence a unikátnost:

- *Persistence* – v případě změny URL odkazujícího na digitální objekt dojde pouze k změně metadat objektu, koncový uživatel žádnou změnu nezaznamená, protože je přesměrován na aktuální URL.
- *Unikátnost* – každý vydavatel má přidělen vlastní unikátní prefix.

Digitální objekty, kterým lze přiřadit DOI:

- časopisy – DOI lze přiřadit na úrovni názvu, ročníku, čísla a článku;
- kniha;
- sborníky z konferencí;
- disertační a diplomové práce;
- technické zprávy;
- standard;
- záznam v databázi;
- jednotlivé součásti určitého díla – odstavec článku, obrázek, tabulka, soubor s daty, graf.

DOI je přiděleno pouze definitivní práci (Definitive Works), nikoli pre-printu, post-printu nebo osobní verzi (Personal Version)².

CrossRef neshromažďuje plné texty, ale metadata o digitálních objektech – zejména URL, na kterých jsou plné texty dosažitelné. Metadata zadává vydavatel do systému CrossRef ve formátu XML, mezi hlavní položky patří:

- DOI;
- zdroj digitálního objektu – URL;
- název článku;
- jména všech autorů;
- přesný název časopisu, do kterého článek patří; dále jeho ročník, číslo a příslušná čísla stran;
- reference.

1 “CrossRef’s goal is to be a trusted collaborative organization with broad community connections; authoritative and innovative in support of a persistent, sustainable infrastructure for scholarly communication.”[4]

2 Veřejně přístupná verze práce, která byla přijata vydavatelem, ale nebyla formálně zpracována a publikována.

Základní scénář práce s CrossRef obsahuje následující kroky:

1. Autor vytvoří digitální objekt, připraví související metadata.
2. Vydavatel vytvoří na URL zadaném v metadatach webovou stránku „response page“, která přísluší vytvořenému digitálnímu objektu. Tato stránka obsahuje plnou bibliografickou citaci a mechanismus, kterým může uživatel získat přístup k plnému textu. Uživatelé autentizovaní na vydavatelově serveru mají zpravidla automatický přístup k plnému textu, ostatní by měli mít možnost tzv. „pay-per-view“, tedy zaplacení za zhlédnutí práce. Na stránce „response page“ by měl být přítomný abstrakt práce.
3. V co nejkratším časovém sledu po zveřejnění článku by měl do systému CrossRef vydavatel zadat DOI, které přidělil objektu spolu s jeho metadaty (zejména URL odkazující na „response page“).

Uživatel, který chce zhlédnout publikaci s určitým DOI, zadá tento identifikátor do vyhledávací služby na webové stránce <http://dx.doi.org/>. Poté je z centrálního registru DOI vybrána položka metadat, příslušících zadanému DOI, která obsahuje aktuální URL uvedené vydavatelem. Následně je prohlížeč přesměrován na tuto stránku vydavatele.

Struktura DOI

Příklad DOI: 10.1016/S0140-6736(87)90593-9

Řetězec DOI se skládá ze dvou základních částí: prefixu a sufixu, které jsou odděleny dopředným lomítkem „/“. Všechny prefixy začínají dvojcíslím „10“ následovaným „.“, dále je alespoň čtyřciferné číslo, přiřazené vydavateli organizací CrossRef. Vydavatel může mít zaregistrováno více prefixů.

Na rozdíl od pevně daného formátu prefixu, strukturu sufixu si může volit každý vydavatel sám. Jediná omezení se týkají znaků, použitelných pro zápis sufixu. Dovolené jsou následující symboly:

„a–z“, „A–Z“, „0–9“ a „-._;()/"

Velké a malé znaky nejsou při vyhodnocování DOI identifikátoru rozlišovány.

Existuje řada doporučení pro vytváření sufixů – mezi základní patří stručnost (z důvodu snížení rizika chyb vzniklých při přepisu), použití ISBN v případě knih (doporučení AAP³), strukturování sufixu do více úrovní (oddělených např. tečkou), nebo možnosti snadného rozšíření sufixu (např. v případě nutnosti zavedení hlubší struktury – přidání obrázku apod.)

Z prefixu nelze jednoznačně odvodit, kdo je vydavatelem daného díla – DOI zůstává zachováno i v případě, že dojde ke změně vlastnických práv vztahujících se na něj.

Dotazování a vyhledávání podle DOI

Rozlišujeme dva hlavní formáty dotazů; prvním je „legacy pipe delimited format“ – tento dotaz může obsahovat 10 (resp. 12) polí oddělených znakem „|“, kde pořadí jednotlivých polí je pevně dáno. Příklady polí, které mohou být uvedeny: ISBN, TITLE, VOLUME, PAGE a další.

Druhý typ dotazu je podstatně komplexnější a je v současné době více prosazován – dotaz XML. Kromě elementů analogických polím v prvním typu dotazů může uživatel navíc zadat například atribut *match* s hodnotami „fuzzy, optional, exact“ upřesňujícími požadovanou míru shody určitého elementu.

Cited-by linking

Každý člen organizace CrossRef může díky službě „Cited-by Linking“ získat seznam DOI identifikujících publikace, které citují jeho obsah.

Scénář popsaného procesu:

1. Vydavatel A zaregistruje svůj článek A1 s určitým DOI a metadaty.
2. Následně vydavatel B vloží do systému svůj článek B1, ve kterém cituje článek A1 vydavatele A. Tedy ve svých referencích (v elementu <citation_list>) uvádí autora článku A1, název časopisu, ze kterého článek pochází, rok vydání a další specifické položky.
3. V systému CrossRef vzniká mezi články A1 a B1 vztah „cited-by“.
4. Pokud vydavatel A položí dotaz, kdo cituje článek A1, obdrží DOI a metadata článku B1.

Seznam publikací, které citují článek, může získat pouze osoba s vlastnickými právy k tomuto článku.

Primární způsob získávání cited-by odkazů je zadání Cited-by Linking dotazu. Formát dotazu je určen XML schématem CrossRef XML query schema, odpověď na takový dotaz je opět XML – ve schématu CrossRef XML query result schema.

Poplatky

Vědci, kteří vytvářejí digitální objekty, ani knihovny nemusí za přístup k systému CrossRef nic platit. Naproti tomu vydavatelé platí roční poplatek, jehož výše závisí na ročním obratu tržeb vydavatele. Například v roce 2010 platili vydavatelé s ročním obratem do \$1 milionu poplatek \$275, tedy 5 160Kč; vydavatelé s ročním obratem přes \$500 milionů platí \$50,000 – téměř 1 milion Kč.

Kromě výše zmíněných poplatků platí vydavatelé za každý vložený digitální objekt; za vložení jedné kapitoly knihy s více než 250 kapitolami zaplatí \$0.15, tedy zhruba 3Kč. Vložení standardu nebo diplomové práce stojí \$1. Další oblast platby představují pokuty ve výši \$2 za vytvoření tzv. konfliktu – tedy za pokus o registraci již zaregistrovaného digitálního objektu (digitální objekty jsou považovány za nerozlišitelné v případě, že mají shodná metadata).

Vlastní zhodnocení projektu a jeho přínosu:

Velmi mě zaujala idea celého projektu – možnost jednoznačně identifikovat každý článek i další digitální objekty. Organizace CrossRef si vytyčila nelehký úkol, potýká se například s problémy výskytů duplicitních objektů – zejména v případě pre-printů. Přesto se domnívám, že tento projekt znamená velký přínos pro akademickou obec, knihovny i vydavatele.

Seznam literatury/zdrojů, URL:

[1] Oficiální webové stránky organizace CrossRef: <<http://www.crossref.org/help/>>

[2] Brand, Amy. 2001. CrossRef Turns One. Volume 7 Number 5. ISSN 1082-9873. D-Lib Magazine. <<http://www.dlib.org/dlib/may01/brand/05brand.html>>

[3] Oficiální stránky organizace Scitation: <<http://scitation.aip.org/jhtml/crossref.jsp>>

[4] Stránka history/mission na oficiálních webových stránkách organizace CrossRef: <<http://www.crossref.org/01company/02history.html>>

Metadata v DC:

<i>dc:title</i>	CrossRef - využití DOI pro citation-linking u vědeckých publikací
<i>dc:creator</i>	Fabriková, Lucie
<i>dc:subject</i>	CrossRef, DOI, citation-linking
<i>dc:date</i>	2010-12-7
<i>dc:description</i>	Describes main features of project CrossRef and its usability in citation-linking
<i>dc:type</i>	text
<i>dc:language</i>	"cs"