

EPrints - open repository software, community and services

Jméno: Bc. Petr Šimkovič

UČO: 73204

Ročník: 2.

Datum zpracování: 18. 1. 2008

URL projektu: <http://www.eprints.org>

Stručná charakteristika projektu

Systém EPrints je určen zejména pro správu vědeckých a technických informací a pro sdílení výsledků a novinek vědeckého výzkumu. Příkladem takových informací jsou diplomové práce, vědecké články, technické zprávy, příspěvky z konferencí apod. Možné je však i ukládání multimediálních a digitálních záznamů či souborů. EPrints je kompletní systém připravený k okamžitému nasazení. Je vyvíjen na University of Southampton ve Velké Británii. Šíření je pod hlavičkou organizace GNU. Rozvoj programu je finančně podporován britskou Společnou komisí pro informační systémy JISC (The Joint Information Systems Committee, <http://www.jisc.ac.uk/>) a americkou Národní vědeckou nadací NSF (The National Science Foundation, <http://www.nsf.gov/>).

Historie

- 1999 říjen – zrozena myšlenka EPrints
- 2000 listopad – vydána verze 1.0, podpora OAI 0.2
- 2001 leden – vydána verze 1.1, podpora OAI 1.0
- 2002 duben – vydána verze 2.0
- 2002 červenec – EPrints pod hlavičkou GNU Project, podpora OAI 2.0
- 2007 leden – vydána verze GNU EPrints 3.0

Současný stav

Na celém světě je nyní známo 242 EPrints archivů. Celkový počet záznamů v těchto archivech činí 405826. Tímto se tedy patrně řadí mezi nejrozšířenější eprintové řešení. Komunikace s externím prostředím je zajištěna implementací OAI-PMH.

Cíle projektu

Projekt je hlavně vytvářen pro budování volně přístupných (open access) repozitářů. Proto je tedy používán hlavně v akademickém a vědeckém prostředí, což je jeho primárním

záměrem. Díky svému open source charakteru smí koncoví provozovatelé tento systém modifikovat či dokonce přidávat nové funkce. Pro větší rozšíření byla vytvořena multijazyková podpora. Tu zajišťují speciální XML soubory s příslušnými národními frázemi. Pro neustálé zlepšování, práci s uživateli a řešení jejich potřeb byla vytvořena tzv. EPrint community.

Popis projektu

Mezi hlavní rysy systému patří jeho modulární návrh. Tímto je tak umožněna snadnější a flexibilnější údržba, stejně tak zapojení produktů třetích stran a propojení s externími službami. Např. s PubMed (medicínská databáze), CrossRef (referenční systém umožňující propojení mezi citacemi a texty článků elektronických časopisů napříč mezi vydavateli), Google Maps nebo ParaCite (experimentální služba vyvíjená samotnými řešiteli tohoto projektu umožňující hledat v citovaných položkách). Při nainstalování potřebných dodatečných aplikací je možné vytvářet fulltext a umožnit tak vyhledávání nejen v metadatech, ale i v samotných datech (např. PDF, HTML nebo RTF souborech).

Specialitou systému EPrints je podpora archivů – což jsou v podstatě zcela samostatné digitální knihovny s vlastní konfigurací, které běží nad jednou instalací systému EPrints. Je tedy možné, aby jeden server obsluhoval samostatné a navenek zcela odlišné archivy dvou i více institucí. Archivy jsou obvykle řešeny pomocí virtuální serverů webového serveru Apache, který systém užívá.

EPrints je tedy webová aplikace, která pro svůj běh a administraci vyžaduje použití i příkazového řádku. Její provoz je umožněn pod operačními systémy Unix, Solaris, Mac OS X či Windows. Zdrojové kódy jsou napsány ve skriptovacím jazyce Perl. Pro uložení a vyhledávání metadat vyžaduje databázi MySQL. Jako webový server je zvolen Apache webserver. Samotná konfigurace je pak zajišťována pomocí XML souborů.

Jak již bylo zmíněno, metadata jsou uložena v databázi MySQL. Ty jsou dvojího druhu:

- systémová (vlastní ID, datum vložení)
- archivní (lze customizovat, např. rok, název, autor, ...)

Podobně tak systém rozlišuje mezi třemi uživatelskými rolami:

- administrátor
- autor (vkládání dokumentů)
- editor (ověření, schvalování a smazání dokumentů)

S tímto souvisí i patřičné workflow práce s jednotlivými dokumenty. Postup vkládání položek je velmi propracovaný a uživatelé (předpokládá se, že publikují vědci) mohou vytvořit velmi podrobná metadata. Obyčejný uživatel smí svoji vlastní položku, která již byla finálně schválena pro vložení do repozitáře editovat. Takto editovaná položka však musí opět projít celým schvalovacím procesem. Zde je dobré zmínit, že systém nepoužívá termín digitální objekt, ale užívá termín položka (Item). Položka je ale v principu zase jen digitálním objektem – zapouzdřuje metadata i data. Vlastní dokument se pak tedy skládá z metadat a dat, které mohou být tvořeny různými formáty (HTML, PDF, ...) a tedy nejen jedním souborem. Pro výměnu metadat se využívá standardů Dublin Core, METS nebo MODS.

Systém EPrints nabízí kompletní uživatelské rozhraní. Vzhled a částečně i samotné funkce rozhraní jsou dobře konfigurovatelné. K dispozici jsou standardní služby – vyhledávání, odkazy na jednotlivé části systému, procházení obsahem repozitáře, výpis nejnovějších přidaných položek, dále odkazy na vstup do uživatelské oblasti, registrace

nového uživatele a další. Správci mohou využívat další funkce, které umožňují celý systém konfigurovat a obsluhovat. Při procházení a vkládání dokumentů do repozitáře lze využít i stromové struktury, která uživatelům usnadňuje samotnou orientaci. Při vyhledávání je možné pak nastavit i samotný výpis dat a umožnit tak i jejich export (ASCII, HTML, BibTeX, Dublin Core, MODE, METS, XML nebo OpenURL).

Samotný projekt nabízí i mnoho služeb, volně dostupných i placených. Sem zejména patří tréninkové školící programy a řešení importu dat ze starších archivů. Pro méně programátorsky schopné provozovatele je nabízena placená customizace systému dle vlastních specifických požadavků, rovněž tak i dopsání a zavedení různých pluginů. Do placených služeb patří i nabízený hosting a následná údržba repozitáře.

V systému EPrints se předpokládá, že vkládání dokumentů nebude příliš časté. Proto je část rozhraní pro přístup k systému tvořena statickými HTML stránkami, které se jednou za čas (obvykle v době nejmenší zátěže, například v noci) generují. Tak se značně urychluje přístup k datům. Pokud je nějaká položka vložena do repozitáře, pak se ve veřejné části webu objeví až po té, co je spuštěno nové generování HTML stránek. Tento přístup je volen zejména v případech, kdy se předpokládá skladování obrovského množství dat.

Vlastní zhodnocení

Pokud se budeme dívat na EPrints jako na software určený zejména pro vědecké a akademické prostředí, můžeme v klidu říci, že je pro toto použití zcela vhodný. Jistě o tomto svědčí i to, že mu byla udělena cena UKUUG (UK Unix and Open Systems User Group) Award v roce 2005. Pokud bychom ale chtěli systém např. pro uchovávání fotek, měli bychom se raději podívat po jiném systému. Obecně tento systém není vhodný pro hromadné vkládání velkého množství dokumentů. K hlavním kladům řadím jeho poměrně velké konfigurační možnosti. Naopak k záporům to, že pro jeho provoz je potřeba nainstalovat poměrně velké množství perlowských modulů, i když je pak jeho vlastní instalace velmi rychlá a systém je možné dokonce ihned užívat. Nutno také vyzdvihnout jeho výraznou podporu ze stran jeho autorů, nabízených služeb a zejména tzv. technical mailing listu vhodného především pro administrátory systému.

Literatura

<http://www.eprints.org/>

<http://en.wikipedia.org/wiki/Eprints>

<http://www.oucs.ox.ac.uk/rts/eprints/>

Metadata v DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title"
      content="EPrints - open repository software, community and services" />
<meta name="DC.Subject" content="EPrint" />
<meta name="DC.Subject" content="open access" />
<meta name="DC.Subject" content="repository" />
<meta name="DC.Creator" content="Petr Šimkovič" />
<meta name="DC.Description"
      content="Systém EPrints je určen zejména pro správu vědeckých a technických
informací a pro sdílení výsledků a novinek vědeckého výzkumu. Příkladem takových
informací jsou diplomové práce, vědecké články, technické zprávy, příspěvky z konferencí
apod. Možné je však i ukládání multimediálních a digitálních záznamů či souborů. EPrints je
kompletní systém připravený k okamžitému nasazení. Je vyvíjen na University of
Southampton ve Velké Británii. Šířen je pod hlavičkou organizace GNU." />
<meta name="DC.Date" content="18.1.2008" />
<meta name="DC.Type" content="Text" />
<meta name="DC.Format" content="application/pdf" />
<meta name="DC.Format.medium" content="computerFile" />
<meta name="DC.Identifier"
      content="http://www.fi.muni.cz/~xsimkov/files/eprints_esej.pdf" />
<meta name="DC.Source" content="http://www.eprints.org/" />
<meta name="DC.Source" content="http://en.wikipedia.org/wiki/Eprints" />
<meta name="DC.Source" content="http://www.oucs.ox.ac.uk/rts/eprints/" />
<meta name="DC.Language" content="cze" />
```