

Systemy pro tvorbu digitálních knihoven

Vlastimil Krejčíř, krejcir@ics.muni.cz

Ústav výpočetní techniky, Masarykova univerzita, Brno

INFORUM 2006, Praha

Obsah přednášky

- Úvod
- Fedora
- DSpace
- EPrints
- CDSware
- Greenstone
- Lokalizace systémů

Úvod

Systémy pro tvorbu digitálních knihoven:

- realizace „knihovny“ v počítačovém prostředí
- desítky různých systémů
- komerční x **volně dostupné** (opensource) systémy
- obecné x specializované
- přístup k tvorbě obsahu repozitáře

Existuje **svatý grál** mezi systémy?

Představované systémy:

Fedora, DSpace, EPrints, CDSware, Greenstone

Fedora

Flexible Extensible Digital Object and Repository Architecture

- vyvíjen na **Cornell University** a **University of Virginia**
- obecný systém pro uchovávání dat
- staví na teoretických modelech

V současnosti poskytuje pouze robustní jádro – repozitář pro ukládání dat, který poskytuje API pro přístup a správu. Rozhraní pro koncové uživatele si musí instituce vytvořit sama.

Fedora

Digitální objekt

Digitální objekty v systému Fedora:

- metadata a data (dokumenty, fotografie, hudba, ...)
- služby nad daty (poskytování, převody, ...)

Data spojená se službami jsou základním prvkem systému.

Technologie: XML, webové služby – SOAP, WSDL =>
nezávislost na platformě a lokaci.

Fedora

Další vlastnosti

Další vlastnosti systému Fedora:

- export dat ve formátu **METS**
- poskytování metadat – protokol **OAI-PMH**
- metadata **MODS, DC, ...**
- správa verzí
- rozhraní, programové vybavení
- ...

Fedora

Závěr

Uživatel dostává do rukou **motor F#1** – záleží jen na něm, jak s ním naloží. **Je netriviální poskládat výkonný monopol a řídit jej.**

Praktické nasazení? Testy na milionech DOs, avšak zatím ne příliš velké rozšíření. . .

Vhodné pro velké instituce, které to „myslí vážně“.

DSpace

Digital Archive Project

- původně vyvíjen **MIT** a **HP**
- nyní široká a činná celosvětová komunita
- obecný systém pro uchovávání, sdílení a prezentaci dat

Ve srovnání se systémem Fedora je DSpace **kompletní osobní automobil** okamžitě připravený pro svého majitele.

=> Menší flexibilita systému :-(, ale také méně práce :-).

DSpace

Rozhraní

DSpace poskytuje kompletní webové rozhraní:

- uživatelské účty a systém přístupových práv
- procházení obsahem repozitáře, vyhledávání
- vkládání dat uživatelem
- workflow systém
- správa systému
- ...

Vzhled webového rozhraní lze měnit dle potřeb instituce (standard CSS).

DSpace

Další vlastnosti

Podporované standardy:

- export do **METS**
- poskytování metadat přes **OAI-PMH**
- metadata ukládá v **DC**

DSpace

Závěr

DSpace je velmi živý:

- stovky instalací na celém světě (desítky tisíc DOs)
- lokálně modifikované verze pro různé účely
- rychlý vývoj systému
- příspěvek může kdokoli (komunita kolem systému)

Univerzální, praktický a perspektivní systém, ale (zatím) ne úplně dobře modifikovatelný.

EPrints

System **EPrints**:

- vyvíjen na **University of Southampton**
- specializovaný na oblast STI
- kompletní připravený okamžitě k použití
- starší „léty prověřený“

Specializace na oblast STI ovlivňuje celý systém (správa dokumentů, workflow, uživatelské rozhraní, ...).

EPrints

Vlastnosti

Uživatelské **webové rozhraní**:

- vyhledávání, procházení (členění do kategorií dle *LoC*)
- uživatelské účty, jednoduchý systém práv
- propracovaný systém vkládání dokumentů
- správa systému

Webové rozhraní je částečně **statické** – předem generované:

- vložení nového dokumentu se neprojeví ihned
- rychlejší odezva systému při přístupu uživatelů

EPrints

Závěr

System **EPrints** je nasazován v praxi (desítky tisíc dokumentů), zejména v oblasti STI.

Modifikace systému jsou částečně možné – změna vzhledu webového rozhraní, změna postupu při vkládání nového dokumentu, ...

System je vhodný především pro ukládání a prezentaci **textových dokumentů** v případě, kdy uživatelé sami vytvářejí obsah repozitáře.

CDSware

CERN Document Server Software

- vyvíjí CERN ve Švýcarsku
- zaměřený na potřeby CERNU
- uživatelské rozhraní na pomezí Fedory a EPrints
- robustní systém s velkým množstvím funkcí
- možnost placené podpory
- GNU systém

Vysoká modularita systému – využívání pestré kolekce aplikací třetích stran.

CDSware

Vlastnosti

Uživatelské **webové rozhraní** není ve výchozím systému kompletní – nutnost vytvořit si pomocí *maker* a dalších nástrojů základní rozhraní pro různé typy dokumentů.

Z dalších vlastností:

- slovníky klíčových slov a autorit
- možnost modifikovat vyhledávací algoritmus
- propracovaný systém přístupových práv, uživatelské role
- podpora **OAI-PMH**, metadata **MARC 21**
- částečně statický (podobně jako EPrints)

CDSware

Závěr

Hlavní nasazení v CERNu – přes 300 tisíc dig. objektů (dokumenty, fotografie, ...).

„Obrovský“ systém, netriviální instalace (asi 10 dalších aplikací) a správa.

Vhodný pro instituce, které chtějí robustní systém a jsou ochotny si zaplatit podporu ze strany vývojářů.

Greenstone

Greenstone Digital Library Software

- vyvíjí **University of Waikato** (Nový Zéland)
- podpora organizace UNESCO
- GNU systém
- filozofie podobná systému CDSware

System Greenstone byl podrobně testován a popsán v diplomové práci Jakuba Řehana na Fakultě informatiky MU.

Greenstone

Zajímavé vlastností:

- podpora **Z39.50, OAI-PMH, DC**
- import dokumentů převodem do interního formátu
- jádro poskytující služby, modulární systém
- Java aplikace *Librarian* pro správu
- slabší podpora uživatelských účtů a práv

Provést některé úpravy je i přes modularitu systému ne vždy zcela snadné.

Lokalizace systémů

- česky nejlépe Greenstone, částečně CDSware a DSpace
- problémová indexace českých dokumentů
- nefunguje „ceske vyhledavani“

Závěr

Každý systém má své **klady i zápory**. Až při ostrém provozu mnohdy vyjdou najevo nedostatky.

Každý systém je možné **přizpůsobit** s větším či menším úsilím – záleží na konkrétních požadavcích. *Co je snadné v jednom, je obtížné ve druhém a naopak. . .*

Aktuální **funkce a schopnosti** systémů se stále mění a vylepšují.