

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Kramerius

PV070

Jan Holman

2. semestr (N-IN PSK)

5. 12. 2014

Název projektu, jeho nositel, URL

Projekt Kramerius: <https://github.com/moravianlibrary/kramerius>

Na projektu se podílejí: Národní knihovna ČR, Akademie věd ČR, společnost Incad

Charakteristika projektu

Kramerius je český open source projekt vyvíjený firmou Incad a financovaný z dotačních programů Akademie věd ČR a Ministerstva kultury ČR.

Systém Kramerius je formou CMS¹ umožňující správu digitálních objektů a jejich zpřístupnění.

Cíle a aktuální stav projektu

Vývoj systému Kramerius iniciovala Národní knihovna po povodních v roce 2002, při kterých došlo k poškození přibližně 800 000 svazků a začaly se vážně řešit problémy spojené s degradací starých tiskovin mj. kvůli nekvalitnímu papíru periodik z 19. století. Bylo proto potřeba řešit otázku digitalizace, archivace, ale i zpřístupňování digitálního obsahu čtenářům.

Systém Kramerius slouží právě ke zpřístupnění digitálních objektů prostřednictvím internetu. Umožňuje řízení dostupnosti dokumentů (dělí je na veřejné a neveřejné), jejich zobrazování a replikaci mezi jednotlivými institucemi, ať už formou exportu a importu, nebo sklizením metadat pomocí protokolu OAI-PMH².

Kramerius je primárně určen pro digitalizované knihovní sbírky, periodika a monografie, je možné ho ale použít ke zpřístupnění řady jiných modelů – například vícesvazkových monografií, map, hudebnin, starých rukopisů, případně článků, kapitol atp.

Kramerius může obsahovat jak zdigitalizované dokumenty, tak tzv. „born digital“ dokumenty, tedy ty, které už vznikly v digitální podobě.

-
1. Content Management System
 2. Open Archives Initiative Protocol for Metadata Harvesting

Popis projektu a jeho výsledků

Systému Kramerius vyvíjela od roku 2003 společnost QBizm³, v roce 2009 ale vývoj převzala firma Incad⁴ a systém prošel řadou změn, z nichž nejvýznamnější je pravděpodobně využití úložiště Fedora Commons⁵. Výsledek je označován jako Kramerius 4 a v současnosti je na něm založeno přibližně 50 digitálních knihoven převážně v ČR, ale už i na Slovensku.

Systém Kramerius ve své současné podobě se dá rozdělit na 3 vrstvy:

První, datová vrstva, je složena z úložiště Fedora Commons a databáze PostgreSQL, ve které se uchovávají vazby mezi objekty ve formě RDF⁶ tripletů. Fedora obsahuje metadata digitálních objektů ve formátu FOXML⁷ a libovolná další data – například textový přepis obsahu oskenované stránky (OCR⁸), popis umístění textu v obrázku ve formátu ALTO⁹ XML, díky kterému je vyhledávaný text v obrázku zvýrazněn, nebo i vlastní digitální obsah, například naskenované strany periodik a monografií ve formátu DjVu, případně celé dokumenty v PDF. V praxi ale bývá digitální obsah uložen na externím úložišti a pouze provázán odkazy.

Druhou, aplikační vrstvu, představuje samotná webová aplikace Kramerius. Ta je napsaná v jazyce Java a může běžet v aplikačním serveru Apache Tomcat¹⁰. Pro účely zobrazování a vyhledávání jsou data uložena v úložišti Fedora Commons indexována pomocí systému Apache Solr¹¹. Import objektu do digitální knihovny proto zahrnuje proces vložení souborů do Fedory a jejich následnou indexaci v Solr.

Kramerius umožňuje nejen zobrazování, ale i správu digitálních dokumentů: jejich import nebo export, replikaci do jiných digitálních knihoven, zpřístup-

3. <http://www.qbizm.cz/>

4. <http://www.incad.cz/>

5. <http://fedorarepository.org/>

6. Resource Description Framework

7. Fedora Object XML

8. Optical Character Recognition

9. Analyzed Layout and Text Object

10. <http://tomcat.apache.org/>

11. <http://lucene.apache.org/solr/>

nění metadat prostřednictvím protokolu OAI-PMH. Umožňuje také spravovat dostupnost jednotlivých digitálních objektů pomocí příznaku public/private. Veřejné (public) objekty jsou přístupné volně přes internet, neveřejné (private) objekty může běžný uživatel pouze vyhledat a případně si zobrazit náhledy obrázků a metadata. Neveřejné dokumenty bývají zpravidla přístupné pouze z vyhrazených počítačů v budově dané instituce, odkud si uživatelé nemohou obsah stahovat a dále ho sdílet.

Kramerius také umožňuje nastavit u dokumentů tzv. „pohyblivou zed“¹². V takovém případě potom aplikace dokumenty zpřístupňuje průběžně v souladu s autorským zákonem. To je velmi praktické zejména v případě, že daná digitální knihovna obsahuje velké množství periodik – není potom potřeba každý rok ručně zpřístupňovat jeden ročník z každého periodika.

Uživatelé si také mohou veřejné dokumenty nechat vygenerovat ve formátu PDF a následně si je stáhnout, nebo si je nechat rovnou vytisknout.

Poslední vrstvou je vrstva prezentační. Ta je od samotné webové aplikace oddělena nově od konce roku 2014, kdy vyšel Kramerius ve verzi 5. Nová verze, představená na letošním ročníku konference Archivy, knihovny, muzea v digitálním světě, se od té předchozí neliší nijak zásadně, obsahuje ale nové RESTové rozhraní, díky kterému lze prezentační vrstvu měnit a případně používat jinou, než výchozí. Této možnosti už využila Moravská zemská knihovna v Brně a vytvořila si pro systém Kramerius vlastní mobilní aplikaci. Ta je v současnosti dostupná v beta verzi na webu Google Play¹². MZK zároveň zakoupila několik tabletů, na kterých si budou čtenáři moci prohlížet neveřejné dokumenty, aniž by museli sedět u počítače.

Společnost Incad zároveň připravuje nové rozhraní, které bude od příštího roku součástí aplikace a nahradí to stávající. Toto nové rozhraní je už teď nasazené například na (zatím testovacích) stránkách České digitální knihovny¹³. Tato knihovna je navíc zajímavá tím, že nemá svůj vlastní obsah, ale sklízí data z ostatních digitálních knihoven založených na systému Kramerius po ČR.

12. <https://play.google.com/store/apps/details?id=cz.mzk.kramerius.app>

13. <http://cdk-test.lib.cas.cz/client>

Zhodnocení projektu a jeho přínosu

Systém Kramerius je stále aktivně vyvíjený a vývojáři na stránkách projektu reagují a odpovídají na dotazy uživatelů (knihoven). Systém se mj. díky této aktivní podpoře poměrně snadno instaluje. Moravská zemská knihovna navíc připravila instalační balík pro platformu Docker¹⁴, který instalaci výrazně zjednodušuje.

Kramerius tak představuje velmi vhodný nástroj, pomocí kterého si i menší knihovny zvládnou samy poměrně snadno vytvořit vlastní digitální knihovnu a zpřístupnit tak svůj digitální obsah čtenářům.

Velkou výhodou je také fakt, že vývoj Krameria probíhá pod licencí GNU GPL¹⁵, takže jsou zdrojové kódy volně dostupné a pokud někomu nevyhovuje nebo chybí nějaká funkcionalita, může si aplikaci sám upravit.

Aplikace také velmi dobře škáluje, jak je vidět na příkladu Národní digitální knihovny ČR, která je na systému Kramerius postavená a obsahuje už přes 25 milionů stránek, což znamená více než 30 TB obrázků a 1,6 TB textových metadat.

Mezi nevýhody systému patří například ne příliš uživatelsky přívětivé rozhraní. To by se ovšem mělo v nejbližší době zlepšit s příchodem rozhraní nového. Další nevýhodou, která už je z velké části překonaná je formát DjVu¹⁶, kvůli kterému je třeba mít v prohlížeči speciální plugin. Součástí Krameria je ale i prohlížečka OpenLayers¹⁷, která podporuje propojení se obrázkovým serverem IIPImage¹⁸. Ten dokáže mj. posílat obrázky ve formě dlaždic v tom rozlišení, jaké zrovna uživatel potřebuje. Díky tomu není potřeba žádný zvláštní plugin a aplikace je rychlejší, protože nepotřebuje obrázky nutně načítat v plné kvalitě.

Stávající rozhraní aplikace Kramerius ovšem zatím neobsahuje prohlížečku pro formát PDF. Tyto dokumenty je proto třeba před vlastním prohlížením stáhnout.

Další nevýhoda, kterou ale běžný uživatel nepocítí, vychází z použití úložiště

14. <https://www.docker.com/>

15. GNU General Public License

16. <http://djvu.org/>

17. <http://openlayers.org/>

18. <http://iipimage.sourceforge.net/>

Fedora Commons verze 3.5, která nepodporuje vratné transakce. Kvůli tomu může dojít k problémům ve chvíli, kdy dojde k přerušení procesu importu.

Seznam zdrojů

Github projektu Kramerius: <https://github.com/moravianlibrary/kramerius>

Článek „Kramerius 4 a Kramerius 5“: <http://www.inflow.cz/kramerius-4-kramerius-5>

Článek „Role Systému Kramerius v oblasti tvorby a zpřístupňování digitálních dokumentů“: <http://kzv.kkvysociny.cz/archiv.aspx?id=391>

Metadata v DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Kramerius" />
<meta name="DC.Creator" content="Jan Holman" />
<meta name="DC.Subject" content="Kramerius" />
<meta name="DC.Subject" content="digitální knihovna" />
<meta name="DC.Date" content="2014-12-5" />
<meta name="DC.Description" content=" Esej popisující systém Kramerius,
který slouží k tvorbě digitálních knihoven." />
<meta name="DC.Format"scheme="IMT"content="application/pdf"/>
<meta name="DC.Source"scheme="URL"content="https://github.com/ceskaexpedice/kramerius»
<meta name="DC.Source"scheme="URL"content="http://www.inflow.cz/kramerius-
4-kramerius-5»
<meta name="DC.Source"scheme="URL"content="http://kzv.kkvysociny.cz/archiv.aspx?id=391»
<meta name="DC.Language"scheme="ISO639-1"content="cs"/>
```