

Masarykova univerzita

Fakulta informatiky



# **Kramerius a MZK**

PV070

Maroš Grman

7. semester

25.11.2016

## Názov projektu, jeho nositeľ, URL

Názov projektu: Kramerius a MZK. Jeho nositeľ: Petr Žabička, Moravská zemská knihovna Brno. URL: <http://www.digitalniknihovna.cz/mzk/>

## Charakteristika projektu

Moravská zemská knihovna, sídliaca v Brne, je knižnicou s univerzálnym knižničným fondom. Plní funkciu krajskej knižnice Juhomoravského kraja a od roku 2009 je aj výskumnou organizáciou. MZK využíva systém Kramerius, čo je open source projekt vyvíjaný firmou Incad. Tento systém je primárne určený pre sprístupňovanie digitálnych dokumentov.

## Aktuálny stav

Digitálna knižnica MZK využíva pre ukladanie a prezentáciu digitálneho obsahu systém Kramerius. Tento systém je používaný viacerými českými knižnicami. Neustály vývoj systému je pre Moravskou zemskou knihovnu zásadný. Kľúčové bolo pridanie API<sup>1</sup> vo verzii 5. Toto umožnilo vytváranie nových webových a mobilných aplikácií, ktoré využívajú dáta a funkcie Krameria prostredníctvom API. Aplikácie majú prístup iba k verejným dokumentom a autorsky chránené diela sú neprístupné. V súčasnej dobe MZK využíva nové webové rozhranie digitálnej knižnice<sup>2</sup> vyvinuté Janem Rychtařem. Tento nový klient nahradil klienta vyvinutého firmou Incad, ktorý je dostupný iba zo zamestnaneckej siete MZK. Najväčšou výhodou tohto webového rozhrania je prístup k viacerým digitálnym knižniciam v Českej republike.

Čo sa týka mobilných aplikácií pre Kramerius, ich vývoja sa ujala MZK v roku 2014. Ako prvá bola vyvinutá aplikácia pre operačný systém Android a v nasledujúcom roku začala vznikať aplikácia pre iOS. Obsahom digitálnej knižnice sú:

- monografie
- noviny a časopisy (periodiká)
- rukopisy a staré výtlačky
- mapy
- archíválie<sup>3</sup>
- hudobniny

---

1 Application programming interface

2 <http://www.digitalniknihovna.cz/mzk/>

3 dokument uchovávaný v archíve

- obsahy CD z fondu MZK
- zvukové dokumenty (nahrávky z gramofónových platní)

## Ciele projektu

Cieľom je dlhodobé uchovávanie dokumentov a ich sprístupňovanie v digitálnej podobe on-line z dôvodov

- zánik, rozpad a opotrebovanie dokumentov
- vysoká požičiavanosť
- záujem užívateľov alebo inštitúcií

## Popis projektu a jeho výsledkov

V súčasnosti MZK využíva Kramerius verziu 5, ktorá je vyvíjaná a publikovaná od roku 2013. Svojou funkčnosťou nadväzuje na predchádzajúcu verziu systému. Ako jadro systému je použitý open source repozitár Fedora<sup>4</sup>. Na ukladanie obrázkov je použitý systém imageserver<sup>5</sup>. Pri vývoji sú využívané ďalšie voľne dostupné technológie tretích strán - Apache, Apache Tomcat, Apache Solr, Postgres SQL. Systém je založený na technológii Java a je možné ho prevádzkovať ako samostatnú webovú aplikáciu v ľubovoľnom J2EE kontajneri (napr. Apache Tomcat). Vo verzii 4 a 5 je užívateľské rozhranie orientované na fulltextové vyhľadávanie.

Po najnovšej aktualizácii je systém Kramerius spúšťaný v aplikácii Docker<sup>6</sup>. Jednotlivé komponenty sú zapnuté v samostatných kontajneroch, ktoré sú na sebe závislé a spúšťajú sa pomocou aplikácie docker-compose<sup>7</sup>. V kontajneroch bežia všetky komponenty, okrem databázy Postgres. Jednotlivé dockerovské repozitáre sú dostupné na: <https://hub.docker.com/r/moravianlibrary/>.

Dáta, ktoré prúdia z Krameria sú využívané viacerými klientmi, ktorí sú tiež umiestnení v kontajneroch aplikácie Docker.

1. Ostrá verzia, ktorá je dostupná pre verejnosť
2. Starý klient, ktorý je dostupný iba zo zamestnaneckej siete
3. Testovacia verzia
4. Špeciálne modifikovaný klient, ktorý je dostupný iba z počítačov umiestnených v študovniach MZK.

Klient č. 4 sa používa na prístup k neverejným dokumentom. Otvára sa pomocou prehliadača IceWeasel<sup>8</sup>, ktorý je modifikovaný takým spôsobom, aby v ňom nebolo

4 <http://www.fedora-commons.org>

5 hierarchicky rozdelené úložisko obrázkov

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

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

8 Odnož webového prehliadača Mozilla Firefox

možné kopírovanie, sťahovanie, printscreen a iná funkcionálnosť, ktorá by umožňovala stiahnutie alebo iné odcudzenie nedostupných dokumentov. Okrem neverejných hudobní, je možné všetky dokumenty v knižnici tlačiť.

Momentálny stav zdigitalizovaných dokumentov je taký, že všetkých dokopy (verejných a neverejných) je 152 432. Z toho najviac je monografií, až 145 207. Novín a časopisov je iba 1 941, čo sa môže zdať málo, ale jednotlivé noviny majú až desiatky ročníkov a v každom z ročníkov môžu byť stovky vydaní, pokiaľ časopis vychádzal denne. Žiaľ 90 % týchto dokumentov je neverejných, presne 138 246. Takže z domu si čitateľ prečíta iba 10 % zo všetkých zdigitalizovaných dokumentov.

Do budúcnosti má v pláne MZK pokračovať v digitalizácii primárne týchto dokumentov:

1. dokumenty 19. a 20. storočia, ktoré sú najviac ohrozené degradáciou kyslého papiera
2. najznámejšie a najžiadanejšie publikácie od roku 1801
3. najznámejšie historické dokumenty od roku 1801
4. dokumenty voľne prístupné širokej verejnosti
5. ostatné typy dokumentov podľa kritérií významu, unikátnosti, fyzického stavu a podobne.

Všetky tieto dokumenty sú uložené takým spôsobom, že rodič má informácie o svojich potomkoch. Najlepším príkladom tohto spôsobu ukladania sú periodiká. Samotné periodikum má informácie o ročníkoch, ročníky o jednotlivých vydaniach a tieto vydania o samotných stranách. Monografie obsahujú väzbu priamo na stránku. V prípade viacväzbových monografií majú najskôr väzbu na zväzok a až potom na stránky. Každý jeden z dokumentov v digitálnej knižnici má svoje metadáta<sup>9</sup>. Tie sú definované v špecifických formátoch a dajú sa vyhľadávať v katalógoch. Všetky metadátové záznamy majú svoj identifikátor. Medzi tieto identifikátory patrí:

- **sysno (system number)**, napríklad: 001090970
- **signatúra**, napríklad: 4-1235.668. Ide o lokálny identifikátor dokumentu, v podstate číslo knihy
- **čiarový kód**, napríklad: 2610406249. Identifikátor exemplára
- **uuid**, napríklad: uuid:9ebcb206-24b7-4dc7-b367-3d9ad7179c23. 32-miestny identifikátor digitálneho dokumentu pre akúkoľvek úroveň dokumentu (celé periodikum a strana). Generuje sa náhodne. Pravdepodobnosť, že vzniknú dve rovnaké uuid je veľmi nízka.
- **urn:nbn**, napríklad: urn:nbn:cz:mzk-004ztc. Trvalý identifikátor nezávislý na umiestnení zdroja, je určený pre dokumenty národných bibliografií (nbn - national bibliografic number). V identifikátore je vidieť, v akej zemi a inštitúcii sa dokument digitalizoval. V Česku má na starosti pridelovanie identifikátorov urn:nbn Národní knihovna ČR (systém CZIDLO na adrese [resolver.nkp.cz](http://resolver.nkp.cz)).

---

<sup>9</sup> štruktúrované dáta nesúce informácie o primárnych dátach

Dáta a Metadáta, ktoré tvoria jednotlivé dokumenty sa do Krameria MZK dostávajú pomocou digitalizačných liniek. Primárnou digitalizačnou linkou je NDK, spoločný projekt medzi MZK a NK ČR (Národní knihovna České republiky). Tento projekt je zameraný na knižnú produkciu vydanú po roku 1800. Plán na rok pre MZK je naskenovať cca 3 mil. knižných strán. V MZK sa digitalizačná linka nachádza v -1. poschodí a je to kľúčový zdroj dát.

Druhou digitalizačnou linkou je Meditor. Ten slúži na menšie digitalizácie, na konverziu obrázkov, vytvorenie metadátového popisu a import dát do Krameria. Meditor je v dnešnej dobe už zastaraný, do budúcnosti je plánovaný prechod na Proarc<sup>10</sup>. Cez meditor sú do Krameria nahrávané aj zvukové dokumenty. Meditor je tiež schopný editovať už existujúce objekty a to týchto metadátových štandardoch: DublinCore, Mods, FOXML alebo METS.

## Zhodnotenie projektu a jeho prínos

Systém Kramerius používaný Moravskou zemskou knihovnou je veľkým prínosom pre knihovníkov, z dôvodu archivácie starých dokumentov a taktiež aj pre čitateľov, ktorým umožňuje jednoduchú prácu s dokumentmi. Vďaka platforme Docker je správa a inštalácia systému zjednodušená. Taktiež je v prípade zlyhania niektorého z kontajnerov jeho opätovné spustenie triviálnou záležitosťou. V momentálnom stave obsahuje digitálna knižnica MZK až 40 miliónov naskenovaných stránok. Najnovší vzhľad digitálnej knižnice prináša jednoduchší a zrozumiteľnejší spôsob práce s digitálnymi dokumentmi. Za veľké plus považujem možnosť stiahnutia aplikácie Kramerius do mobilného zariadenia. Všetky zdrojové kódy, okrem najnovšieho klienta, sú dostupné online na GitHubu. Jednou z nevýhod tohto systému je zatiaľ nepodporovanie transakcií, čo môže vytvárať problémy pri zlyhaní nedokončeného procesu. Na záver by som odporučil každému navštíviť digitálnu knižnicu MZK a vyskúšať si prezrieť nejakú dostupnú knihu.

## Zoznam zdrojov

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

GitHub projektu Kramerius Incad: <https://github.com/ceskaexpedice>

Stránka Moravskej zemskej knihovny: <http://www.mzk.cz/>

Digitálna knižnica Moravskej zemskej knihovny:  
<http://www.digitalniknihovna.cz/mzk/>

Zkušenosti s digitalizací v MZK Brno, digitální knihovna Kramerius: Využití digitálních technologií v památkové péči

<sup>10</sup> <https://github.com/proarc/proarc/wiki>

Příspěvek do Infoboxu: Digitální knihovna MZK: její vývoj, mobilní aplikace a uživatelé

Lenka Damborská, Uživatelské testování nového webového rozhraní pro Digitální knihovnu MZK - zhodnocení

Interview zo zamestnancami Moravskej zemskej knižnice

Zamestnanecká dokumentácia

## Metadata

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
```

```
<meta name="DC.Title" content="Kramerius a MZK" />
```

```
<meta name="DC.Creator" content="Maroš Grman" />
```

```
<meta name="DC.Subject" content=" Kramerius a MZK" />
```

```
<meta name="DC.Date" content="2016-11-25" />
```

```
<meta name="DC.Description" content="Esej ktorá popisuje využitie systému Kramerius MZK" />
```

```
<meta name="DC.Format" scheme="IMT" content="application/pdf" >
```

```
<meta
```

```
name="DC.Source" scheme="URL" content="https://github.com/moravianlibrary/kramerius»
```

```
< meta
```

```
name="DC.Source" scheme="URL" content="https://github.com/ceskaexpedice/kramerius»
```

```
< meta name="DC.Source" scheme="URL" content=" http://www.mzk.cz »
```

```
< meta
```

```
name="DC.Source" scheme="URL" content="http://www.digitalniknihovna.cz/mzk/»
```

```
<meta name="DC.Language" scheme="ISO639-1" content="sk"/>
```