

Digitální knihovny

Praktická práce s digitální knihovnou

Jindřich Marek

Praktická práce s digitální knihovnou

Staneme se badateli a tvůrci/kurátory digitální knihovny.

1. Prozkoumáme skryté funkce Krameria (NDK).
2. Vytvoříme vlastní malou digitální sbírku v systému Omeka.net.

Jak funguje digitální knihovna (technicky)?

- frontend (to, co vidí uživatel)
 - vyhledávací pole, filtry, čtečka obrázků
- data (to, co systém uchovává)
 - skeny, fotografie (TIFF, JPEG2000)
 - texty z obrázků (ALTO XML, TXT)
- backend a metadata (to, s čím pracuje knihovník)
 - rozhraní pro popis objektů (Dublin Core, případně složitější formáty)
 - systém pro správu oprávnění a sbírek

1. Pokročilý badatel

- analýza záznamů v různých digitálních knihovnách
- práce s API systému Kramerius

Analýza záznamů

- různé digitální knihovny
 - Kramerius
 - Manuscriptorium
 - oborové repozitáře apod.

Checklist

- jaký dokument jste vybrali? (název, URL, příp. PID/URN)
- metadata: jaká pole vidíte v uživatelském rozhraní? odhadněte použitý standard (Dublin Core, MODS...)
- authority: je autor prokliknutelný? je v záznamu uvedeno číslo authority?
- zpřístupnění: je dostupný obraz (scan), fulltext nebo obojí? lze stáhnout?
- práva: jaká licence / právní věta je uvedena? (public domain, CC?)
- dlouhodobé uchovávání: najdete informaci o použitém formátu (JP2, PDF/A), kontrolních součtech (checksum) či archivní verzi?
- hodnocení UI: uveďte 1 silnou a 1 slabou stránku pro běžného uživatele

Práce s API systému Kramerius

- sken knihy je pro počítač jen „slepý“ obrázek.
- abychom mohli vyhledávat, potřebujeme OCR
- úkol: na <https://kramerius5.nkp.cz> najděte knihu nebo periodikum z 19. století a přepněte zobrazení na text

Všimněte si případných chyb v textu (špatně přečtené znaky).

Všimněte si, jak je prezentovaná struktura textu.

Jaký to má vliv na vyhledávání?

Kniha a její metadata (Kramérius API)

- UUID vyhledané knihy
 - <https://ndk.cz/view/uuid:bfc1db80-79c4-11e2-b930-005056827e51>
- popis v DC
 - <https://ndk.cz/search/api/v5.0/item/uuid:bfc1db80-79c4-11e2-b930-005056827e51/streams/DC>
- IIIF manifest (odkaz nefunguje)
- dokumentace k API:
<https://github.com/ceskaexpedice/kramerus/wiki/ClientAPIDEV>
- UUID obrázku, který vás zajímá, např.
 - <https://ndk.cz/view/uuid:bfc1db80-79c4-11e2-b930-005056827e51?page=uuid:4cd23a83ec650dd3d888985ca0fca0c5>
- obrázek v nejvyšším rozlišení
 - <https://ndk.cz/search/api/v5.0/item/uuid:4cd23a83ec650dd3d888985ca0fca0c5/full>
- plný text k obrázku
 - https://ndk.cz/search/api/v5.0/item/uuid:4cd23a83ec650dd3d888985ca0fca0c5/streams/TEXT_OCR

Citace a PID (Kramérius)

- digitální knihovna není jen web, linky se nesmějí měnit
- URN:NBN (Uniform Resource Name: National Bibliography Number) - trvalý a neměnný odkaz na digitální dokument
- úkol
 - najděte u dokumentu v Kramériovi tlačítko pro generování citací
 - na co odkazuje a jaké jsou možnosti?

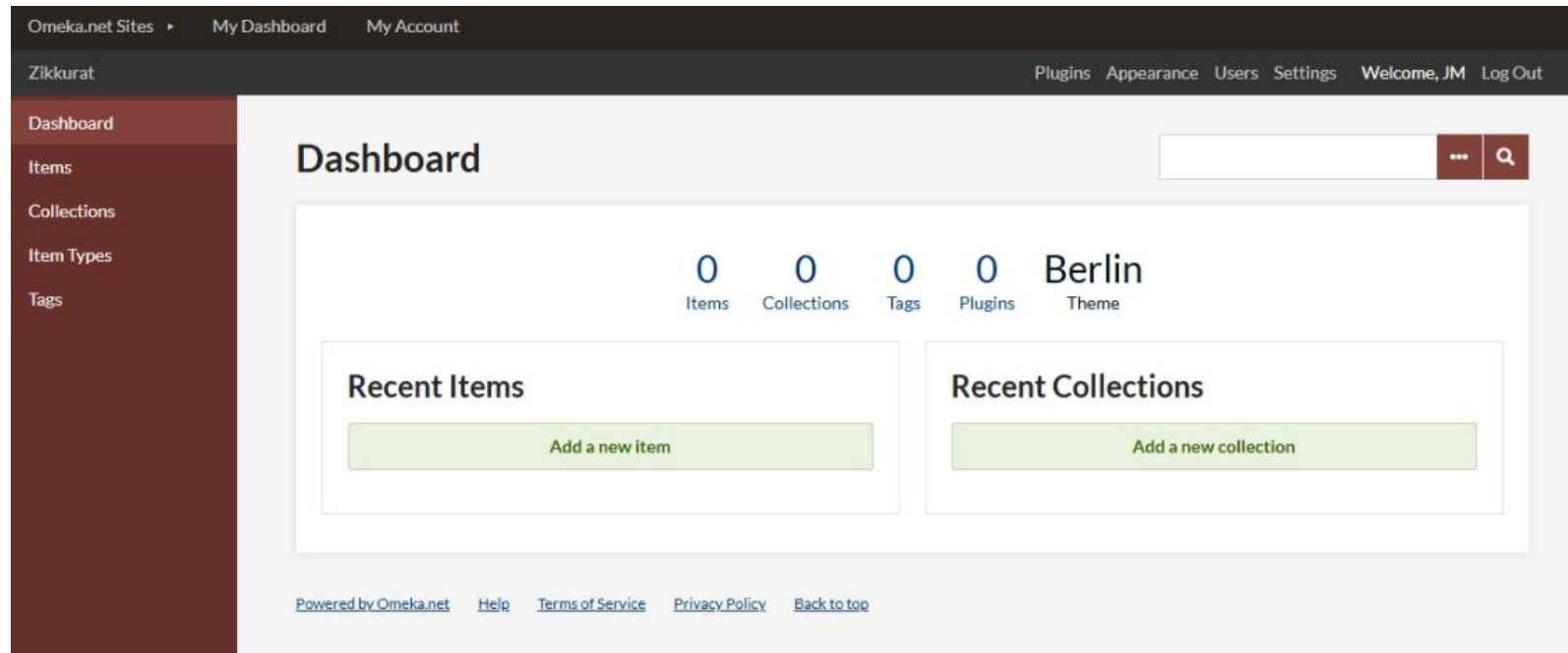
Přínos IIF

International Image Interoperability Framework

- standard, který umožňuje sdílet obrázky napříč různými institucemi bez jejich stahování
- sdílejí se jen strukturovaná data (tzv. Manifest)
- úkol
 - najděte v Manuscriptoriu odkaz na IIF Manifest a přidejte jej jako zdroj na <https://projectmirador.org>
 - najděte jiný IIF Manifest (např. na <https://www.e-codices.unifr.ch>) a udělejte s ním totéž
 - porovnejte oba rukopisy a podívejte se také, jak vypadá IIF Manifest

2. Role kurátora

- tvoříme vlastní digitální knihovnu / online výstavu



Předpoklady

- soubor obrázků (minimálně 3-5), ze kterých budeme vytvářet digitální knihovnu
- účet na Omeka.net (<https://www.omeka.net/signup>), tlačítko „Try Omeka for free“ = zkušební verze na měsíc, je potřeba jen e-mailová adresa
- je dobré znát prezentaci k metadatům v Moodle (<https://dl1.cuni.cz/course/view.php?id=5216>), zejména část zaměřenou na popisná metadata (Dublin Core)

Co je Omeka (S)?

- <https://omeka.org>
- open-source platforma pro publikování na webu zaměřená na kulturní dědictví
 - vytvořena v USA speciálně pro menší knihovny, galerie, muzea a univerzity
- starší a novější verze (Omeka Classic, Omeka S)
- budeme pracovat s cloudovou verzí: Omeka.net
 - založená na Omeka Classic

Úkol: prezentace obsahu v Omeka.net

- krok 1: stránka (Site)
 - založte stránku (bude přístupná na <https://mojestranka.omeka.net>)
- krok 2: sbírka (Collection)
 - založte sbírku
 - nastavte ji jako veřejnou
 - vyplňte metadata podle standardu Dublin Core
 - proč jsou kvalitní metadata klíčová pro dlouhodobé uchování a vyhledávání?
- krok 3: jednotky (Items)
 - nahrajte obrázky
 - přiřadte je ke sbírce
 - nastavte je jako veřejné
 - vyplňte metadata podle standardu Dublin Core

Výsledek: stránka v Omeka.net

- prohlédněte si výsledek a odpovězte na otázky

Co bylo na nastavování knihovny nejsložitější? (Nahrání?
Vyplňování metadat?)

Jak by se tento proces dal urychlit/škálovat pro reálnou knihovnu?

Reflexe a diskuse

Co ze surových dat a metadat vytváří příběh přístupný koncovému uživateli na webu?

Jaký byste zvolili archivní formát pro data a metadata? Jak byste zajistili migraci dat?

4. Závěr

- digitální knihovna je komplexní systém propojující objekty, metadata, vyhledávací indexy a prezentační vrstvu

Jak dnes získané zkušenosti využijete pro závěrečné prezentace?