DL4DH – DIGITAL LIBRARIES FOR DIGITAL HUMANITIES

NOVÝ PROJEKT NA VYTĚŽOVÁNÍ OBSAHU DIGITÁLNÍCH KNIHOVEN

DL4DH – Digital Libraries for Digital Humanities – nový projekt na vytěžování obsahu digitálních knihoven. Knihovna Akademie věd ČR, Národní knihovna ČR a Moravská zemská knihovna v Brně iniciovali projekt, jehož cílem je přispět k vytěžování dat z digitálních knihoven pro vědecké účely zejména v humanitních oborech. Knihovny disponují obrovským množstvím dat v digitální podobě, které jsou zároveň zpravidla popsány kvalitními metadaty umožňujícími jejich vyhledávání a další využití. Tato data jsou uložena v digitálních knihovnách a repozitářích, jejichž design a funkce jsou určeny především pro prohlížení obsahu. Humanitně zaměření vědečtí pracovníci potřebují data obohacená tak, aby vyhledávání v digitálních zdrojích zefektivnilo možnosti jejich využití a vytvořit odpovídající prostředky pro jejich vytěžení, díky čemuž se usnadní důležitá část badatelské práce. Cílem projektu DL4DH je vývoj nových funkcí a nástrojů, které umožní extenzivní využití a vytěžování dat z digitálních knihoven pro potřeby digitálního humanitního výzkumu, a současně příprava aplikovaných vědeckých výstupů využívající tyto nové možnosti. Projekt volně navazuje na evropskou výzkumnou infrastrukturou DARIAH a s výstupy českého výzkumného projektu INDIHU, zaměřeného na vývoj nástrojů pro digital humanities.

Projekt DL4DH – Digital Libraries for Digital Humanities byl schválen Ministerstvem kultury ČR v rámci programu NAKI pod ID DG20P02OVV002 a jeho řešení probíhá v letech 2020 – 2022 . Hlavním cílem projektu DL4DH je vývoj softwarového řešení, které umožní vytěžování dat z digitálních knihoven provozovaných v systému Kramerius.

Přírodní a humanitní vědy již delší dobu sdílejí data napříč obory a systematicky budují datové repozitáře, na které jsou navázány služby jejich základní analýzy. Tyto datové repozitáře, původně budované jako izolované výstupy jednotlivých výzkumných projektů, jsou postupně propojovány napříč pracovišti a obory, což zvyšuje efektivitu jejich využití a umožňuje sdílet výstupy výzkumů v celoevropském kontextu. To prokazatelně vede k rozšíření efektivity i záběru výzkumu a samozřejmě i k otevírání nových otázek. Specifikem humanitních věd je množství odborných i obecných textů, které jsou obsaženy v digitálních knihovnách, jež jsou pro určité obory (například obecnou i matematickou lingvistiku, dějiny literatury, sociologii atd.) klíčovým zdrojem informací a mnohdy základním předmětem výzkumu. Pro další vědní obory jde o podstatný doplňkový zdroj informací a souhrnně jde o nedílnou součást kulturního dědictví. Přestože se jednotlivé obory humanitních věd technologicky rozvíjí, hledání zdrojů pro analytickou práci za využití digitálních nástrojů je stále velmi složité. V tuzemském prostředí existuje jen minimum korpusů, které dokáží poskytovat humanitně-vědní data v podobě vhodné pro strojové zpracování. Realizace projektu tak bude spočívat i v transformaci dat do lépe strukturované podoby, jež po samotném exportu umožní jejich zpracování prostřednictvím různorodých výpočetních operací či skriptů vytvořených samotnými vědci. Projekt nabízí cestu, jak podstatný blok dat o kulturním dědictví otevřít, a to prostřednictvím obohacení existujícího systému Kramerius o nové nástroje pro data mining, které najdou široké uplatnění napříč obory. Za účelem dosažení co nejlepších výsledků vědecké práce budou dostupná data transformována do podoby umožňující jejich využití prostřednictvím automatizovaných procesů. Díky širokému spektru uživatelů stávajícího systému bude dopad projektu velmi vysoký, a dává tak dobrý základ pro skutečný posun v oblasti studia kultury a historie ČR. Projekt zapojuje humanitní vědce z řady oborů (archeologie, knihověda, religionistika, filozofie, jazykověda, …), kteří společně se specialisty na počítačové vědy v první fázi zanalyzují oborové i mezioborové potřeby pro práci s digitálními knihovnami. Výsledky analýzy budou následně využity při vývoji nástrojů v dalších etapách projektu. Dosažení svých cílů bude projekt demonstrovat nejen implementací nových nástrojů, ale též ve formě jejich modelových praktických aplikací pro výzkumné i prezentační účely. Projekt navazuje na dobrou praxi, spočívající ve využití předchozích výsledků aplikovaného výzkumu podporovaného z veřejných zdrojů, které uvádí do praxe, prohlubuje jejich dopad a zvyšuje celkovou udržitelnost a efektivitu.

PROJEKT DL4DH MÁ TYTO HLAVNÍ CÍLE:

1) DL4DH feeder – rozhraní pro data mining z digitálních knihoven provozovaných v systému Kramerius

Nové softwarové řešení napojené na digitální knihovny provozované v systému Kramerius umožní vyhledání, selekci a následný export všech digitálních dat (obrazových, XML výstupu OCR i textu exportovaného z OCR výstupu, to vše spolu s metadaty) jak v původní podobě, tak i do některého z dalších formátů, umožňujících efektivní (zejména) strojové zpracování (např. CSV, TSV, XML, JSON). DL4DH feeder tak poskytne humanitně orientovaným vědcům možnost vytěžovat data z digitálních knihoven a dále využít výpočetní metody pro výzkum, který je řazen do oblasti digital humanities. Součástí implementace bude řešení evaluace kvality OCR a obohacení plných textů o metadata a základní jazykovou analýzu (tzv. lemmatizaci), což umožní efektivnější vyhledávání v datech a filtraci výsledků, podle zadání a potřeb vědeckých pracovníků z humanitních oborů. Získaná data mohou být následně dále zpracována pomocí dalších nástrojů, jež jsou standardně využívány vědeckou komunitou (a samozřejmě i pro nástroje, jež si jednotliví vědci vyvíjejí pro individuální potřeby), což umožní přípravu jedinečných vědeckých výstupů. Bude možné si zvolit jaká část příslušných metadat bude do exportů zahrnuta. Analyzována bude též podpora a možná implementace standardů pro linked open data, tedy například RDF a SPARQL.
DL4DH feeder umožní dosud nerealizovatelné vytěžování digitálních knihoven provozovaných v systému Kramerius, které bude vzhledem k více než 150 milionům digitálních stran obsažených v těchto knihovnách zcela bezprecedentní příležitostí pro humanitně orientované bádání založené na publikačních výstupech z více než dvou minulých století. Na základě využití nového nástroje DL4DH feeder lze očekávat zcela nové vědecké výstupy v celé oblasti humanitních věd získané díky metodám digital humanities a také nové modely pro analýzu přirozeného jazyka, které budou těmto datům přizpůsobeny a dále podpoří excelentní výzkum v humanitních oborech.

2) Kramerius TEI convertor – konverze a export dat a metadat ze systému Kramerius do formátu TEI
Kramerius TEI convertor umožní konvertovat a exportovat data a metadata ze systému Kramerius ve formátu TEI, který patří mezi hlavní standardy v oblasti digitálních humanitních věd pro detailní popis dokumentů v digitální podobě. Nový softwarový nástroj zajistí kompatibilitu s dalšími projekty a v případě potřeby umožní vzhledem k možnostem TEI obohacení popisu dokumentů z digitálních knihoven v systému Kramerius.

3) Metodika přípravy dat z digitálních knihoven pro digital humanities a jejich zprostředkování vědecké komunitě
„Metodika přípravy dat z digitálních knihoven pro digital humanities a jejich zprostředkování vědecké komunitě" bude popisovat, jaká data se v digitálních knihovnách nacházejí, jaký je jejich původ, význam a vzájemné vazby a jakým způsobem lze pomocí v projektu vyvíjených nástrojů tato data těžit a dále využívat. Tomu bude předcházet důkladné zkoumání zvyklostí humanitních vědců při práci s digitálními knihovnami a budou zodpovězeny otázky týkající se jejich dosavadního způsobu získávání analytických dat (elektronické dotazníky, polostrukturované dotazování, pozorování) a budou zjišťovány potřeby humanitně orientovaných badatelů týkající se automatického těžení dat. Metodika představí dostupné možnosti exportu dat a na praktických příkladech uvede způsoby jejich konkrétního využití vědci.

PRŮBĚH ŘEŠENÍ V PRVNÍM ROCE PROJEKTU

Řešení projektu bylo zahájeno v březnu 2020 a prvním úkolem projektového týmu byla detailnější specifikace funkčnosti jednotlivých softwarových výstupů projektu. Za účelem ukládání metadat a dat, která budou rozšiřovat a obohacovat obsah uložený v systému Kramerius, byla jako součást řešení definovaná pomocná databáze Kramerius+.
Výsledkem práce řešitelského týmu v úvodních měsících byla technická specifikace celého řešení.
Základem vyvíjeného řešení je software DL4DH Feeder provozovaný jako webová služba určená pro přímé využití vědeckými pracovníky. Pro ukládání metadat a obohacených textových dat, které nelze získat ze systému Kramerius, bude vytvořen pomocný systém Kramerius plus. Součástí vyvíjeného řešení bude i možnost propojení webové služby DL4DH Feeder a systému Kramerius plus s externími databázemi a nástroji a jejich využití při obohacování metadat a textových dat nebo při vizualizaci dat vytěžených ze systému Kramerius. Nový nástroj TEI Converter umožní export dat ze systému Kramerius a databáze Kramerius plus ve formátu TEI (viz https://tei-c.org/guidelines/p5/).

MODUL KRAMERIUS PLUS (KRAMERIUS+)

Rozšiřující modul Kramerius+ bude sloužit pro uchování obohacených textových dat a metadat, které jsou uloženy v systému Kramerius, jedná se na např. o lemmatizaci a automaticky rozpoznané entity v plných textech, záznamy z knihovních systémů, odkazy na autority a bibliografické záznamy, apod.). Tyto údaje se využijí pro kvalitnější vyhledávání digitálních dokumentů, k filtraci jejich obsahu a umožní obsah Krameria vědecky zpracovávat, případně exportovat pro další použití.
Kramerius+ by měl fungovat jako externí modul systému Kramerius, měl by mít vlastní API a měl by komuni-

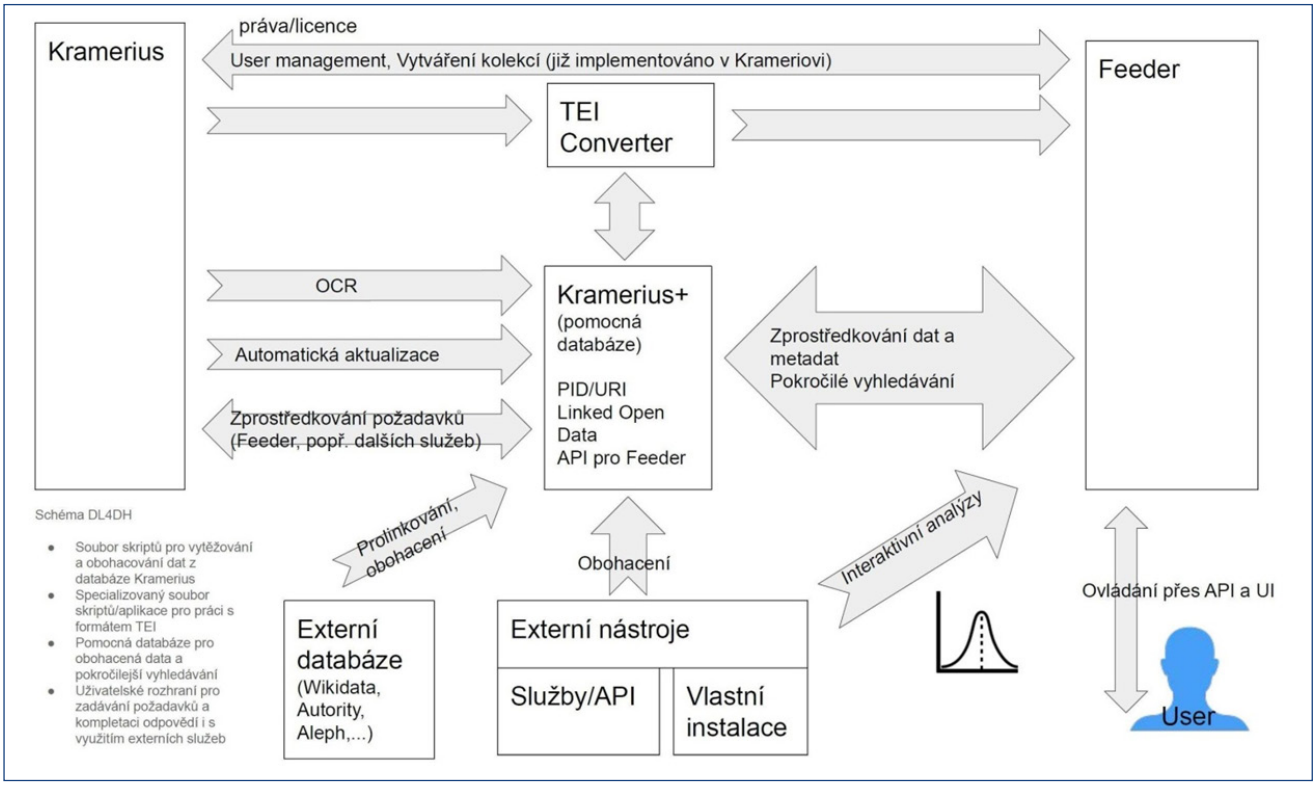

Schéma řešení.

kovat s jádrem Krameria i s DL4DH Feederem. Krameriusplus bude využívat existující plné texty publikací, které jsou uložené v systémech Kramerius a které obvykle vznikly pomocí OCR (ve formátu prostého textu nebo ALTO, viz https://www.loc.gov/standards/alto/).

SPECIFIKACE FUNKCÍ MODULU KRAMERIUS+

- Jednotlivé záznamy v Krameriu+ budou svázány s dílčími objekty v Krameriu na úrovni stránky publikace. Pokud to výhledově půjde, lze uvažovat o navázání přímo na dílčí objekty v ALTO (např. odstavce), které však není v současnosti dostupné pro všechny dokumenty v Krameriu. Řešení bude v tomto ohledu univerzální.
- V Krameriu+ budou uložena metadata rozšiřující dosavadní popis dokumentů v systému Kramerius o informace podstatné pro pokročilé dotazy a obsahové vyhledávání v dokumentech. Zvolená technologie musí umožnit efektivní prohledávání těchto obohacených dat (řádově miliony metadatových položek a desítky milionů stran textu).
- Kramerius+ se bude starat o obohacování metadat a textových dat voláním interních, popř. externích služeb s požadovanými parametry. Součástí popisu služeb bude i jejich účel (např. lemmatizace, vyhledání záznamů v národních autoritách ap.). Zajistí také uložení metadat o volané službě (název služby, verze, nastavené parametry, datum volání a vrácení výsledku).
- Výstupy z externích služeb spolu s aktuální podobou textových dat ve formátu XML TEI předá Kramerius+ nástroji TEI Converter, který zajistí konverzi a integraci těchto výstupů do formátu TEI. Výsledný dokument se uloží v Krameriu+.
- Validní TEI export pro API komunikující s DL4DH Feederem bude umět složit Kramerius+ s využitím nástroje TEI Converter.
- Při zpracování výstupů pro DL4DH Feeder se v maximální míře budou využívat metadata ze systému Kramerius. Existující údaje se tak nebudou zbytečně duplikovat v pomocné databázi, pokud pro to nebudou existovat jiné závažné důvody (zejména efektivita a rychlost zpracování dotazů).
- Obohacené záznamy se budou za účelem vývoje, testování a poloprovozu vytvářet pro cíleně zvolené skupiny dat z Krameria. Úzké hrdlo při velkých objemech je zejména dostupný prostor na úložištích. Pro vlastní provoz tak bude důležitá možnost vytvářet fronty uživatelských požadavků pro zpracování publikací do obohaceného formátu. Při do-

statečných kapacitách může být obohaceno vše v Krameriovi.
- Technologie pro pomocnou databázi Kramerius+ bude zvolena v průběhu implementační analýzy.
- Data Krameria+ by mělo být možné zpřístupnit jako Linked Open Data.
- Kramerius+ musí podporovat správu přístupů v souladu s nastavením licenčních podmínek v základním Krameriu.
- Bude probíhat automatická aktualizace metadat na základě zdrojového záznamu z knihovního systému a báze národních autorit (přebírání MARC záznamů).
- Měla by probíhat automatická aktualizace metadat/dat mezi Krameriem a Krameriem+
- Součástí řešení bude interní analytika využívání systému kvůli vylepšování funkcí i textových dat a metadat.
- K obohacení metadat budou využita metadata z archivních balíčků a zároveň bude možné se přes perzistentní identifikátory provázat na zdrojový obrázek v plném rozlišení.

SPECIFIKACE OBOHACOVÁNÍ DAT

Cílem procesu obohacování je doplnit metadata a plný text (vzniklý OCR) ze systému Kramerius o strukturní značky, jazykovou analýzu, rozpoznané entity a další údaje. Obohacování bude zpravidla plně automatizovaný proces, který využije zejména externí služby a nástroje, schopné zpracovat data z Krameria v požadované formě a rozsahu. V případech, kdy je to možné, mohou být využity i externí výpočetní kapacity (např. infrastruktura LINDAT/CLARIAH-CZ).

Vlastní nástroje pro zpracování dat budou vyvinuty pouze v případech, které jsou kritické pro chod systému a nemají existující funkční alternativu.

Integrace systému s externími i interními nástroji musí být modulární: proces zpracování bude možné administrátorskými zásahy upravovat a např. připojovat nové nástroje a upravovat parametry obohacování (nastavení vstupního a výstupního formátu apod.).

Součástí systému budou také údaje o průběhu obohacení textových dat a metadat: označení externí, popř. interní aplikace, webová/lokální adresa, verze aplikace, nastavení parametrů pro zpracování.

Mezi základní požadavky pro obohacení patří:
- Integrace funkcí pro jazykové zpracování (využití volně dostupné infrastruktury LINDAT/CLARIAH -CZ):
- Tokenizace (rozdělení na slova a věty), lemmatizace (určení základního tvaru slova) a určení slovního druhu (např. nástrojem UDPipe)
- pro texty v češtině, němčině, latině, příp. dalších jazycích.
- Rozpoznání a určení typu pojmenovaných entit (např. pomocí nástroje NameTag).
- Přinejmenším pro češtinu, další jazyky zatím nejisté.
- Převod do TEI – bude zajištěn nově vyvinutým nástrojem TEI Converter, jehož specifikace je součástí této zadávací dokumentace.

Cílem mimo jiné je automaticky analyzovat metadata již dostupná v ALTO (nechceme tato metadata nově vytvářet, pouze efektivně využít). Objem takto zpracovatelných dat poroste tak, jak se bude rozšiřovat obsah digitální knihoven. Zpracování jednotlivých funkcí bude závislé na vstupní analýze a nemusí být všechny realizovány, pokud se k tomu nepodaří nalézt technické prostředky a jejich vývoj nebude v rámci projektu proveditelný. Jde např. o tyto funkce:
- Identifikace obrázků na stránkách, jejich polohy a popisků.
- Rozlišení veršovaného a prozaického textu.
- Rozpoznání struktury textu (sloupce atd.) a podílu prázdného místa na stránce (v %).
- Rozpoznání a správná interpretace tabulek, např. pomocí nástroje Tabula (https://tabula.technology/).
- Analýza typografie.

DALŠÍ MOŽNOSTI PRO OBOHACENÍ

Cílem je pokusit se najít služby, které by dokázaly naplnit následující požadavky, případně vyhodnotit, nakolik je možné v rámci projektu jejich úpravou dosáhnout požadovaného cíle a poskytnou takový výsledek pro Feeder. Bude implementováno podle časových možností a podrobnější vstupní analýzy v průběhu řešení projektu. Výsledkem může být i test s vyhodnocením nasaditelnosti v budoucnu.

PŘÍKLADEM JSOU TYTO FUNKCE:
- Klasifikace obrázků podle základního typu obsahu – fotografie/mapa/kresba (atp.). Lze vycházet též z návěstí (popisků) obrázků.
- Identifikace rukopisných přípisků.
- Rozpoznávání bibliografických odkazů, resp. obecně vylepšení práce s referencemi v rámci seznamu literatury, např. pomocí nástroje Grobid (https://github.com/kermitt2/grobid).
- Segmentace slovníkových a encyklopedických textů, např. pomocí nástroje Grobid-Dictionaries (https://github.com/MedKhem/grobid-dictionaries).
- Obohacování o článková metadata za využití obo-

rové nebo článkové bibliografie (např. Česká literární bibliografie, Bibliografie dějin Českých zemí, Bibliografie české lingvistiky), vč. prolinkování na konkrétní článek a zpět.

SPECIFIKACE FUNKCIONALITY A ROZHRANÍ NÁSTROJE „TEI CONVERTER“ PRO OBOHACOVÁNÍ DAT Z DIGITÁLNÍCH KNIHOVEN PROVOZOVANÝCH V SYSTÉMU KRAMERIUS

TEI Converter je softwarové řešení, které komunikuje s modulem Kramerius+, aby rozpoznaný text digitalizátu obohatil o relevantní data a metadata a aby uživateli poskytl sadu dat v požadované kvalitě (metadata, úroveň obohacení textu) ve formátu XML TEI. Jeho funkce budou využívat správci modulu Kramerius+ a prostřednictvím nástroje DL4DH Feeder, jej budou využívat vědečtí pracovníci v oblasti digitálních humanitních věd. TEI Converter nebude mít vlastní grafické uživatelské rozhraní.

TEI Converter musí poskytovat kvalitně vytvořené a zdokumentované API (s využitím Swaggerhubu, Apiary nebo podobných nástrojů), které umožní modifikaci jednotlivých kroků v procesu obohacování a poskytování dat.

POPIS POŽADOVANÉ FUNKČNOSTI

1. Při vytváření nebo aktualizaci záznamů v systému Kramerius+ vytvoří TEI Converter výchozí podobu dokumentu TEI (tj. elementu <text>) na základě výstupů OCR (ve formátu prostého textu nebo ALTO).
2. Převod výstupů z volaných služeb na odpovídající prvky a atributy TEI, případně sloučení s dokumentem TEI z předchozích kroků.
3. Sestavení kompletního dokumentu, resp. sady dokumentů ve formátu TEI na základě metadat ze systému Kramerius a dat, popř. metadat z modulu Kramerius+. TEI Converter zajistí např. vytvoření hlavičky (<teiHeader>) s metadaty o digitalizátu a jeho zpracování a obohacování nebo spojení samostatných stran publikace do jednoho dokumentu (v elementu <tex>). Dokumenty budou obsahovat metadata, text a jeho obohacení v takové úrovni, kterou si nastavil uživatel prostřednictvím nástroje DL4DH Feeder.

Specifikace funkcionality a rozhraní nástroje „DL4DH Feeder” pro vytěžování dat z digitálních knihoven provozovaných v systému Kramerius

DL4DH Feeder je softwarové řešení napojené na systém Kramerius a na modul Kramerius+, které zajišťuje selekci, vizualizaci a následný export dat do některého z požadovaných formátů (CSV, TSV, XML, JSON, ALTO, PAGE), umožňující další (zejména) strojové zpracování. Software by měly mít možnost nasadit knihovny využívající systém Kramerius (viz registrkrameriu.mzk.cz/). Uživatelé, kteří budou využívat jeho funkce, budou vědečtí pracovníci v oblasti digitálních humanitních věd. DL4DH Feeder bude mít vlastní uživatelské rozhraní umožňující plné využití jeho funkcí.

DL4DH Feeder musí poskytovat kvalitně vytvořené a zdokumentované API (s využitím Swaggerhubu, Apiary nebo podobných nástrojů), které umožní vytěžování dat ze systému Kramerius a pomocné databáze Kramerius plus.

ZÁKLADNÍ POŽADOVANÁ FUNKČNOST:

DL4DH Feeder

- bude obsahovat správu uživatelů a autentifikaci přes Shibboleth
- umožní extrakci, exporty a vizualizace tagovaných údajů a jejich vazeb
- podá informaci o kvalitě dat v Krameriu a kvalitě obohacení, případně o kvalitě jednotlivých verzí dat
- bude provázán na uživatelské kolekce v Krameriovi (bude jen jedna správa uživatelů pro Krameria i Feeder)
- bude podporovat formáty exportu: TEI, T/CSV, JSON, plain text, PAGE, ALTO (pokud je k dispozici v systému Kramerius) + RAW export (data v originální podobě)
- umožní vyhledávání publikací na základě geografických údajů zmíněných v textu publikace výběrem z mapy
- umožní vyhledávání na základě dotazů zohledňujících tagované prvky v textu publikace (zejména chronologie a místa)

V rámci dostupných zdrojů bude možné rozšiřovat funkčnost Feederu o další prvky. Patřit mezi ně mohou např.:

- Napojení na více digitálních knihoven Kramerius najednou (je potřeba řešit duplicity).
- Lepší vytěžitelnost digitalizovaných dokumentů, nebo částí dokumentů, které mají encyklopedický nebo slovníkový charakter (např. rejstříky a soupisy).
- (Polo)automatická extrakce do podoby databáze/linked data.
- Identifikace knižního dekoru, prázdného místa na stránce, rukopisných přípisků.
- Rozpoznání struktury textu (sloupce), typografie apod.
- Evidence historie dotazů a porovnání, zda se ve výstupu Feederu něco změnilo.
- Možnost realizovat reprodukovatelný export – (nástroj AVOBMAT to např. řeší JSON souborem zaznamenávajícím všechna vyhledávací nastavení, tentýž

soubor lze pak nahrát k zopakovaní vyhledávání).

- Citace – viz níže
  - Extrakce citací pro linkování (pokus o identifikaci citací v textech a jejich prolinkování na příslušný dokument/článek).
  - Integrace článkových citací, příp. umožnit uživateli, aby si mohl vytvořit citaci.
  - Možnost vyhledávání podle citace.
  - Možnost vyhledání všech citací (např. u konkrétního autora apod.).
- Extrakce dat z tabulek a grafů (identifikace stránky s tabulkou).
- Vyhledání publikací potenciálně souvisejících s vybranou pasáží textu (na základě automatické obsahové analýzy/identifikace klíčových slov). Teoreticky lze řešit i vícejazyčně - přeložit a prohledat zahraniční biblio katalogy.
- Generování výstupu a automatické předání pro API existujících externích nástrojů pro textové analýzy (např. Voyant Tools). Účelem by mělo být, aby správce systému mohl nastavit UI (knoflíky) pro uživatelský export dat do externích nástrojů. Primárně by to mělo být určeno pro uživatele, kteří nejsou schopni sami používat REST API, nebo prostě pro rychlé a snadné jednorázové užití.
- Automatické překlady zobrazovaného textu (okno s automatickým překladem).
- Síťové vizualizace pro entity nebo záznamy.
- Recommender systém (např. podle žánru, předmětového hesla, podobné ortografie)
- Vizualizace TEI XML s entitami (a další anotací – viz příklad): http://lindat.mff.cuni.cz/services/teitok/parczech-ps7-1.0/index.php?action=file&cid=xml-files/2013-011/2013-011-01-000.xml
- Možnost využití nástroje Grobid (https://github.com/kermitt2/grobid)
- Automatická detekce a extrakce poznámkového aparátu (užitečné zejména pro publikace historiků, kde jsou často poznámky až na konci textu).
- Wrappery pro API DH Feederu v Pythonu/R.
- Možnost otáčení stran dokumentu (viz současný klient Krameria).

Ve výběrovém řízení na dodavatele programátorských prací byla jako nejvýhodnější vyhodnocena nabídka firmy InQool, se kterou byla v říjnu 2020 uzavřena smlouva. Do konce roku 2020 budou probíhat analytické práce na jednotlivých částech řešení a definice struktury pomocné databáze Kramerius+. Na základě těchto příprav, které probíhají v interakci s řešitelským týmem, budou v roce 2021 zahájeny programátorské práce a následně bude zahájeno testování dílčích částí.

ZÁVĚR

Projekt DL4DH – Digital Libraries for Digital Humanities je dalším rozšiřováním spolupráce mezi knihovnami a humanitně orientovanými vědeckými pracovišti v oblasti digital humanities. Volně navazuje například na infrastrukturu DARIAH (v ČR společně s CLARIN pod názvem LINDAT/CLARIAH-CZ) nebo projekt INDIHU jehož cílem byl vývoj nástrojů pro digital humanities. Přímé zapojení humanitně orientovaných pracovníků do přípravy a řešení projektu dává dobré předpoklady, že výstupy projektu budou odpovídat jejich požadavkům a umožní širší využití obsahu digitálních knihoven pro výzkum pomocí metod digital humanities.

Zdroje

BURDICK, Anne, Johanna DRUCKER, Peter LUNENFELD, Todd Samuel PRESNER a Jeffrey T. SCHNAPP. Digital_humanities. Přeložil David VICHNAR. Praha: Academia, 2019. XXI. století. ISBN 978-80-200-2865-5.

MILLSON – MARTULA, Christopher a Kevin B. GUNN. The digital humanities: implications for librarians, libraries, and librarianship. London: Routledge, Taylor & Francis Group, 2019. ISBN 978-1-138-39316-5.

LEVENBERG, Lewis, Tai NEILSON a David RHEAMS. Research methods for the digital humanities. Cham: Palgrave Macmillan, 2018. ISBN 978-3-319-96712-7.

HLADÍK, Radim. Zdrojový kód jako literární technologie?: O vytváření faktů v digitálním věku. Teorie vědy: problémy vědy, techniky a společnosti. 2017, 39(1), 31-56. ISSN 1210-0250.

Systém Kramerius. Homepage [online]. [cit. 2020-11-15]. Dostupné na: https://system-kramerius.cz/

TEI. 2020. Homepage [online]. [cit. 2020-11-15]. Dostupné na: https://tei-c.org/

DARIAH-EU. 2020. Homepage [online]. [cit. 2020-11-15]. Dostupné na: https://www.dariah.eu/

LINDAT/CLARIAH-CZ. 2020. Homepage [online]. [cit. 2020-11-15]. Dostupné na: https://lindat.cz/

INDIHU. 2020. Homepage [online]. [cit. 2020-11-15]. Dostupné na: https://indihu.cz/

Ing. Martin Lhoták
Knižnica akadémie vied ČR
lhotak@knav.cz