

IMPLEMENTÁCIA UMELEJ INTELIGENCIE AKO ODPOVEĎ NA NOVÉ VÝZVY INOVATÍVNYCH DIGITÁLNYCH SLUŽIEB KNIŽNÍC

České a slovenské knižnice v priebehu posledných dvoch dekád úspešne zvládli procesy digitalizácie, sprístupňovania a dlhodobej archivácie. V súčasnosti sa ocitáme v ére veľkého množstva digitalizovaných dokumentov roztrúsených medzi inštitúciami naprieč celou Európou. V súvislosti s rozširovaním služieb umelej inteligencie do spoločnosti čelia digitálne knižnice a ich vlastníci novým výzvam, a to najmä v oblasti inteligentných online služieb pracujúcich nad digitalizovaným obsahom. V očiach našich používateľov sa stále viac objavujú očakávania poskytnúť komplexný informačný produkt v reálnom čase, a to napriek existujúcim jazykovým, systémovým a medziinštitucionálnym bariéram, ktoré sú vytvorené nielen rôznymi systémami digitálnych knižníc, ale aj ich metadátovými popismi. V tomto kontexte sa knižnice stanú jednými z inštitúcií, od ktorých ich používatelia budú čoskoro očakávať podobné technologické inovácie. V aktuálnom čísle Vám prinášame rozhovor s námestníkom riaditeľa Moravskej zemskej knižnice Ing. Petrom Žabičkom, v ktorom sa budeme venovať aktuálnym technologickým výzvam v oblasti elektronických služieb knižníc a projektom, ktoré Moravská zemska knižnica v Brne rieši.

<http://doi.org/10.52036/1335793X.2023.SC2.5-12>

Prečo ste ako absolvent Vysokého učení technického nastúpili do Moravskej zemskej knižnice? S akými víziami a cieľmi ste do MZK prichádzali?

Je to jednoduché. Už během studia jsem sice pracoval v soukromé firmě, ale po ukončení studia jsem se musel rozhodnout, jestli nastoupit na rok na vojnu, nebo na rok a půl na civilní služby. Zvolil jsem to druhé, a protože náhoda přeje připraveným, nastoupil jsem na civilku do MZK. Táhlly mě tam počítače, knihovna tehdy připravovala spuštění knihovního systému Aleph v rámci konsorcia CASLIN. Samozřejmě, civilní služba znamenala i to, že jsem strávil řadu dní stěhováním fondu, protože knihovna v té době opouštěla řadu menších depozitářů, které byly vráceny v restitucích.

Protože mě práce s počítači bavila a viděl jsem v knihovně obrovský potenciál rozvoje, zůstal jsem tam i po skončení civilky. Začínal jsem tedy v knihovně v pozici pracovníka oddělení automatizace, který zastával dle potřeby spolu s kolegy role správce serveru, sítě a počítačů, kterých v té době v knihovně bylo méně než dvacet. Byl jsem tak u připojování knihovny k akademické síti, vytvořil jsem první webové stránky knihovny a pomáhal tomu, aby knihovna dokázala poskytovat své služby v online prostředí. Práce v knihovně dala také možnost pracovat s technologiemi, ke kterým bych se jinak snadno nedostal, knihovna byla například připojena od počátku k Internetu rychlostí 10 Mbps, což byl v době modemů a vytáčeného připojení k Internetu rychlostí o několik řádů nižší opravdu zázrak.

Budúci rok to bude 30 rokov, čo pracujete v Moravskej zemskej knižnici, z toho väčšinu na pozícií námestníka riaditeľa. V čase začiatku Vášho pôsobenia bolo aktuálne pripájanie knižníc na internet, webové stránky, online katalógy a poskytovanie prvých elektronických služieb. Dnes sa ocitáme na prahu ďalšej zásadnej technologickej zmeny, a to implementácie umelej inteligencie do služieb knižníc. Ako hodnotíte posledných 30 technologických rokov v kontexte neustáleho vývoja a implementácie nových elektronických služieb knižníc? Považujete príchod umelej inteligencie za podobne prelomový ako príchod internetu a s tým súvisiaci rozvoj elektronických služieb do knižníc?

Ačkoli se může zdát, že boom umělé inteligence, který přišel v letošním roce, se objevil „z ničeho nic“, opak je pravdou, jen nebyl tolik vidět. V MZK jsme s možnostmi využití systémů strojového učení začali ve spolupráci s kolegy z Vysokého učení technického v Brně experimentovat už v roce 2015 a tato spolupráce trvá dodnes. Kvalitativní skok, ke kterému v letošním roce došlo, s sebou přináší řadu nových možností. Výrazně se rozšířila nabídka různých online služeb na bázi AI, které dnes na vás vyskočí snad už i z každé konzervy. Tohle období bouřlivého rozvoje je příležitostí pro řadu malých firem, které přijdou s nějakým inovativním řešením, dá se ale očekávat, že řada z těchto firem nepřežije a dojde ke konsolidaci trhu, na kterém zůstane několik velkých hráčů. Naštěstí ale během tohoto bouřlivého období vzniká i řada volně dostupných AI aplikací, datasetů a modelů, které výrazným způsobem demokratizují dostupnost AI. Knihovny si tak (podobně jako kdokoli další) budou moci vybrat mezi komerčními službami a mezi volně dostupnými nástroji, které si budou samy provozovat. Není těžké odhadnout, že řada knihovnických specifických AI služeb bude většinou knihoven nabízena prostřednictvím několika velkých knihoven nebo jejich konsorciem. Mimo to samozřejmě budou knihovny využívat obecně dostupné AI služby.



Když se ohlédnu za vývojem automatizace v našich knihovnách za posledních 30 let, vidím (v ČR) roztržitý trh, který od počátku nepřál technologickým monokulturám. Většina knihoven chtěla mít nezávislé technické řešení, byť za cenu řady kompromisů. To sice možná působí značně chaoticky, ale dalo to příležitost vyrůst řadě technologických firem, které ve vzájemném soupeření přinesly do knihovnictví i mnoho technických inovací. Některé z nich úspěšně přežily dodnes a v nejbližších letech uvidíme, jak se jim bude dařit ve světě prostoupeném AI. Přes zdánlivý chaos však se podařilo udržet určitý řád, daný vymáháním jednotných standardů pro bibliografický popis, komunikaci mezi knihovnickými systémy (Z39.50, OAI-PMH a další) nebo pro digitalizaci. Velkou silou proto vidím ve spolupráci knihoven. Díky shodě na společných pravidlech je možná vzájemná spolupráce, je možné dát dodavatelům jasné mantinely, a přesto zůstává prostor pro disentní hlasy, alternativní řešení a hledání nových cest. Přirozeně, v oblastech, které jsou považovány za stabilizované, pak začínají dominovat snahy o optimalizaci nákladů formou konsorciální spolupráce ať už v oblasti sdílené katalogizace, provozování společně využívaných systémů apod. Nástup AI tento trend jen urychlí. AI nástroje budou pravděpodobně nasazovány tam, kde budou mít největší přínos pro uživatele, tedy nad centrálními řešeními, kde jsou vyšší náklady na jejich nasazení a využívání snáze ospravedlnitelné.

Nástup AI však bude postupný. Lidská společnost má svou setrvačnost, danou schopnostmi, preferencemi uživatelů, ekonomickými podmínkami a řadou dalších okolností. Podobné je to i v knihovnách. Myslím si proto, že i změny vyvolané AI se budou v knihovnách projevovat v čase nerovnoměrně i proto, že zejména starší generace neodloží knihu jen proto, že někde na internetu si mohou popovídat s umělou inteligencí – nanejvýš se jí budou ptát, jestli si mají přečíst spíš Stopařova průvodce, nebo Terminátora. Ani předpokládané vyvinutí obecné umělé inteligence pravděpodobně nebude znamenat, že se ze dne na den všechno změní. K tomu bude potřeba ještě dotáhnout vývoj a masovou výrobu humanoidních robotů. I když se mluví o tom, že AI způsobí rychlou radikální změnu a zánik řady profesí, očekávám spíše vznik řady nových nástrojů, které knihovníkům výrazně usnadní prá-

ci a dá jim možnost napřít své úsilí na komunikaci s uživateli knihoven. Na druhou stranu, je možné, že já budu jeden z těch, které AI nahradí už příští rok a toto je můj poslední rozhovor v roli knihovníka.

S příchodem umelej inteligencie je spojených veľa príležitostí na inováciu súčasných online služieb knižníc. Ktoré súčasné služby podľa Vás budú obohatené či postupne nahradené umelou inteligenciou? Ako sa podľa Vás zmenia knižnice s príchodom umelej inteligencie v nasledujúcich 5 rokoch?

Stál jsem u vzniku několika projektů, jejichž cílem je právě transformace online služeb knihoven v nejbližších letech. Přesto, že tyto projekty vznikly v loňském roce, kdy se ještě o AI v mainstreamu tolik nemluvilo, bylo jejich společným jmenovatelem úsilí o automatizaci manuálních činností knihovníků při tvorbě metadat a zvýšení relevance online služeb uživatelům. Moje predikce tedy přinejmenším zčásti jsou založené na znalosti konkrétních cílů běžících projektů nebo existujících AI služeb.

Je jednoznačné, že dojde k proměně digitálních knihoven. Jejich součástí bude možnost překladu textu do libovolného jazyka, vytváření souhrnů ve zvoleném jazyce nebo hlasové čtení textu ve zvoleném jazyce. To jsou ty nejjednodušší služby, s jejichž nasazením již experimentujeme a největší výzvou při jejich nasazení bude vhodné řešení jejich uživatelského rozhraní a cena takové služby. Podobně jen otázkou času je možnost diskuse s AI nad určitým dokumentem, v budoucnu snad i s uživatelem vybranou množinou dokumentů nebo s celým obsahem digitální knihovny. Vůbec by se hlasová komunikace s digitální knihovnou nebo katalogem knihovny, resp. chatbotem reprezentujícím knihovnu mohla stát standardní variantou komunikačního rozhraní. Největší výzvou zde možná bude udržet si důvěru uživatele, že to, co mu AI píše nebo říká, je pravda. Tam mají současné systémy značné rezervy, ale je možné, že právě tam dojde už třeba během několika měsíců k radikálnímu zlepšení.

Změn dozná i samotné vyhledávání. Bude možné vyhledávat nejen v metadatech a textu, ale stejným dotazem i v obrázcích. Samozřejmostí bude vyhledávání relevantních dokumentů nebo informací podle významu, odpadne tedy nezbytnost přesného zadání dotazu. Obdobně bude možné prohledávat i zvukové nebo videodokumenty, největší brzdou tak bude to, že většina dokumentů v knihovnách stále čeká na digitalizaci a často (třeba i u nás v MZK) i na prostou katalogizaci, která by klasickými metodami mohla trvat ještě několik desetiletí. Zde vidím také velký prostor pro změnu. AI nástroje pro tvorbu metadat určitě výrazně urychlí procesy katalogizace i digitalizace, byť možná za cenu určité nedokonalosti ve srovnání s dnešními standardy. Zde je ale nutné si uvědomit, že současné katalogizační pravidla mají svůj původ v době, kdy se data ukládala na magnetické pásky a že ve světě fulltextového vyhledávání, AI technologií a digitalizace se mění i požadavky na podobu dat.

Změny se zcela jistě projeví i ve fyzických službách knihoven – AI asistenti si svou niku určitě najdou, ale domnívám se, že komunikaci s živým člověkem bude řada lidí stále preferovat. K nahrazení knihovníků ve službách v krátkém horizontu pravděpodobně nedojde, vážně to začne být teprve, až se na trhu objeví finančně dostupní humanoidní roboti.

Moravská zemská knižnice je jedným z technologických lídrov nielen v množstve, ale i náročnosti národných a medzinárodných projektov. Prečo sú projekty pre Moravskú zemskú knižnicu tak dôležité a čo Vám priňajú? Aký riešený IT projekt je pre Vás najvýznamnejší?

Pokud se knihovna nemá stát jen skanzenem dob minulých, musí držet krok s technologickým vývojem. Pokud nemáme být pouze pasivními příjemci toho, co přichází zvenčí, musíme se snažit využít nové technologie pro rozvoj knihovny a jejich služeb. Zároveň platí, že podmínkou pro to, aby se nějaká technologie ujala, je její rozšíření mezi uživatele. Proto se snažíme nejen něco nového vytvořit, ale rozšířit takto vzniklá řešení mezi další knihovny. K tomu může vést řada cest, tou asi nejlepší je otevřenost. Tedy vytvářet volně dostupný software s otevřeným kódem a svobodnou licencí, poskytovat online služby využitelné více institucemi apod. K tomu všemu se ideálně hodí projekty, které na jedné straně pomohou zaplatit náklady na vývoj nových technologií a na straně druhé umožní vytvořit jádro uživatelské základny.

Díky tomu, že má MZK již řadu let statut výzkumné organizace, máme výrazně lepší přístup k projektovému financování než běžná knihovna. To má však i opačný důsledek. Aby si MZK svůj statut výzkumné organizace udr-

žela, musí se výzkumu věnovat, což v podmínkách ČR znamená tvořit dostatečné množství dostatečně kvalitních výsledků, a tedy i projektů. Snažíme se však přitom projekty připravovat tak, aby posouvaly knihovnu žádoucím směrem, tedy směrem rozšíření nabídky služeb, zvýšení jejich kvality, zmenšení časové náročnosti činností apod.

Samozřejmě, některé projekty se zaměřují na „velké“ věci, jako jsou například Obálky knih, portál Knihovny.cz nebo Česká digitální knihovna, jiné projekty jsou zaměřené na úzkou, specifickou oblast, která byla doposud opomíjena. Tak jsme se v minulosti zabývali problematikou starých map nebo v současnosti problematikou optického rozpoznávání hudby (projekt OmniOMR). Nemohu však vybrat jeden nejvýznamnější projekt, už proto ne, že většina řešených projektů má společného jmenovatele: rozvoj online služeb s důrazem na rozvoj digitální knihovny. Vážím si zejména toho, že můžeme spolupracovat s řadou partnerských institucí u nás i v zahraničí. Kdybych měl jmenovat ty nejvýznamnější, musím zmínit Národní knihovnu ČR, se kterou sdílíme i digitalizační linku, Knihovnu Akademie věd ČR nebo Vysoké učení technické v Brně, které je naším hlavním technologickým partnerem v oblasti AI technologií. Významná pro nás je i spolupráce s Univerzitou Karlovou, díky níž jsme zapojeni i do velké výzkumné infrastruktury pro digitální humanitní vědy LINDAT/CLARIAH-CZ. Slibně se začíná vyvíjet i spolupráce s CVTI SR, jejímž dokladem je i toto speciální číslo časopisu ITlib.

Vo Vašich projektoch a aktivitách kontinuálne rezonuje digitalizácia, sprístupnenie digitalizovaného obsahu a dlhodobá archivácia. Aké sú Vaše aktuálne plány v oblasti ďalšieho rozvoja digitalizačného pracoviska a jeho prezentačnej platformy (systém Kramerius), včítane dlhodoberj archivácie? Rátate v nich aj s budúcim uplatnením umelej inteligencie?

Jak už jsem naznačil, projekty v oblasti AI jsou v MZK navázané zejména na rozvoj digitalizace a digitalizační linky. Problematiku dlouhodobé archivace řešíme v současnosti už jen v omezeném rozsahu pro speciální druhy dokumentů, jako jsou například datová CD (systém CDArcha), protože pro 95 % naší digitální produkce využíváme dlouhodobé úložiště provozované Národní knihovnou ČR. Největší praktický dopad si slibuji od nasazení AI nástrojů do procesu digitalizace, kde v současnosti dominují manuální činnosti při úpravách skenů a tvorbě metadata. Věřím, že tyto činnosti se i díky projektům podaří v nejbližších letech automatizovat tak, aby se pracovníci mohli věnovat především těm činnostem, na které zatím stroje nestačí, tedy zejména skenování dokumentů. Co se týče rozvoje digitální knihovny, resp. systému Kramerius, na kterém je naše digitální knihovna postavena, tam budou hrát AI nástroje, jak už jsem uvedl v předchozích odpovědích, klíčovou roli.

Nejde však jen o ně. MZK nedávno přešla na novou verzi systému Kramerius, Kramerius 7. Ten je založen na moderních principech, jeho jádro si však s sebou stále nese určitou historickou zátěž, technologický dluh vycházející z toho, že některé jeho části vznikly už před mnoha lety a je potřeba je přepracovat, i když uživatel na první pohled žádnou změnu nezaznamená. Tyto úpravy také usnadní další rozvoj systému i směrem k využití AI technologií.

Prvním krokem bude využití existujících AI služeb – již zmíněné překlady, souhrny a čtení, případně i zadávání požadavků hlasem. K tomu později přistoupí indexace obrázků, vyhledávání částí textu nebo obrázků pomocí vektorů a věřím, že se podaří zprovoznit i chatbota pro komunikaci s dokumenty. Tohle všechno však bude vyžadovat i netriviální úsilí při přípravě dat. Bude nutné u značného množství dokumentů udělat nové OCR (pomocí AI nástroje PERO OCR, který vznikl také v rámci kooperačního projektu pod vedením VUT v Brně). V digitální knihovně MZK je totiž značné množství dokumentů, jejichž OCR, dělané klasickými technologiemi, je velmi nedokonalé, nebo úplně chybí. Jedná se zejména o stará periodika, dokumenty tištěné frakturou, rukopisy apod.

S pribúdajúcimi technologickými možnosťami je stále viac atraktívnejšie poskytovať nadstavbové služby nad digitalizovaným obsahom. V tomto kontexte neustále rastie význam, potreba a záujem o digitálne humanity. Aké sú Vaše súčasné aktivity a vízie v tejto oblasti?

Hlavním posláním knihovny je uchovávat a zpřístupňovat informace. V kontextu digitálních humanitních věd je proto primárním úkolem knihovny poskytovat vědcům data potřebná pro jejich výzkum. Přitom jsme samozřejmě limitováni, ale i podporováni autorským zákonem. Naši primární snahou je proto co nejotevřenější zpří-

stupnění těch dokumentů, na které se autorský zákon nevztahuje. K tomu slouží jednak samotné API Krameria a pak řada nadstavbových aplikací. Organizačně je pak značnou pomocí účast knihovny v už zmíněné velké infrastruktuře LINDAT/CLARIAH-CZ, kde MZK plní především roli poskytovatele dat nejen svých, ale skrze portál Česká digitální knihovna, jehož provozování MZK před rokem převzala, i dat z dalších českých digitálních knihoven.



Z konkrétních aktivit musím upozornit na to, že díky MZK vznikla brána mezi API Krameria a standardy IIF, díky které mohou vědci z celého světa přistupovat k dokumentům uloženým v kterémkoli systému Kramerius prostřednictvím mezinárodně standardizovaných komunikačních rozhraní. Za zmínku stojí i nástroje vytvořené v projektu DL4DH (Digital Libraries for Digital Humanities), na kterém jsme se podíleli a které umožňují data exportovaná z Krameria obohatit pomocí externích nástrojů a zpřístupnit je už obohacené vědcům.

Plánujeme tyto nástroje zprovoznit nad indexem České digitální knihovny, kam logicky patří, a v rámci možností je ve spolupráci s partnery dále rozvíjet. Jednou z prvních vlašťovek – projektů, které využívají data z digitální knihovny – je projekt DIGEOCAT & Lib, jehož jsme partnerem a jehož hlavním řešitelem je Univerzita Palackého v Olomouci.

Zrovna oblast digitálních humanitních věd je zároveň oblastí, se kterou může hodně zamíchat nástup obecné umělé inteligence. Pokud vznikne entita, které dokáže hledat nové vztahy a vyvozovat nové a zároveň správné nové závěry z dostupných dat, nebo si dokonce klást nové relevantní výzkumné otázky, dostanou vědci z oblasti digital humanities do ruky na jedné straně velmi silný nástroj, na druhé pak konkurenta, který je připraví o práci. Tolik vize.

Vo svojej práci ste dosiahli významné výsledky v oblasti projektov zameraných na sprístupňovanie digitalizovaných máp s georeferenciami. Pripravujete v tejto oblasti nejaké novinky, prípadne nadviazanie na výsledky v už zmienenom projekte DIGEOCAT & Lib, ktorý sprostredkuje službu digitálny storytelling? Predpokladáte v tejto službe aj využitie umelej inteligencie?

Problematika zpřístupnění starých map byla hlavním impulsem k tomu, abych prosazoval použití imageserveru jako standardní součásti systému Kramerius, což vedlo následně mj. i k volbě formátu jpeg2000 jako standardu pro zpřístupnění digitalizovaných dokumentů v ČR. V oblasti starých map se nyní v MZK zaměřujeme zejména na přenos samostatně existujících funkcionalit přímo do systému Kramerius. Díky tomu je dnes možné v Krameriu dokumenty (nejen mapy, ale i grafiky, případně jiné objekty obsahující v metadatech geografické souřadnice) vyhledávat, nalezené georeferencované mapy zobrazit v překryvu nad současnou mapou, nebo prezentovat v Krameriu jednotlivé listy mapových děl (mapy rozdělené na více samostatně vydaných mapových listů) ve formě kladu listů, tedy v kontextu celého díla. Málo viditelná, ale důležitá je i aplikace mapseries.mzk.cz, prostřednictvím které dáváme katalogizátorům starých mapových děl k dispozici sadu katalogizačních šablon pro zpracování jednotlivých listů těchto děl. V rámci jednoho z aktuálně řešených projektů bychom rádi získali také sadu nástrojů na podporu kvalitní digitalizace velmi rozměrných map.

Projekt DIGEOCAT & Lib, na který se ptáte, je trochu specifický v tom, že v něm MZK hraje především roli poskytovatele dat, aplikaci samotnou vyvíjí hlavní řešitel, Univerzita Palackého, prostřednictvím externí firmy.

V době přípravy projektu však ještě neexistovaly vhodné hotové AI nástroje pro analýzu starých cestopisů, proto v tomto projektu nedošlo k nasazení AI. Na druhou stranu vznikne v rámci tohoto projektu zajímavá datová sada, která může posloužit buď jako trénovací data pro budoucí AI nástroj, nebo dokonce jen vzorová data, na základě nichž dostane obecný jazykový model zadání, jak má analyzovat další cestopisy. Musím zdůraznit, že se v tomto případě jedná o netriviální úlohu, protože odborníci, kteří se na projektu podílejí, musí v mnohdy více než 100 let starých cestopisech identifikovat místní názvy a správně je umístit na mapu. Zatímco ostatní části analýzy cestopisů by pravděpodobně mohlo AI brzy převzít, samotná geolokace starých místních názvů z cizích zemí, převedených do dobové češtiny, je úloha, které bude pravděpodobně ještě nějakou dobu doménou lidí.

V současnosti ste riešiteľom projektov, ktoré majú veľký medzinárodný a národný potenciál z hľadiska poskytovania komplexných inovatívnych a centralizovaných elektronických služieb naprieč inštitúciami a ich digitalizovanému obsahu. Vedeli by ste nám viac priblížiť projekty Smart digilinka, Orbis Pictus, semANT a PERO? Aký je najväčší rozdiel medzi projektmi PERO a Transcribus?

Tyto projekty majú spoločného predovšetím kľúčového řešitele, Michala Hradiše z Vysokého učení technického v Brně. Ten stojí se svým vývojářským týmem za všemi kľúčovými výsledky těchto projektů, tedy za nástroji využívanými AI technologie. Zatímco PERO je projekt, který byl v minulém roce ukončen, ostatní tři projekty jsou v prvním roce řešení z pěti let. Pokud bych se měl zaměřit na projekt PERO, jeho výhodou je velmi dobrá podpora češtiny, tedy českých znaků jak v rukopisných, tak v tištěných textech. To je důležité například pro frakturu (švabach), kde zahraniční nástroje včetně Transcribusu selhávají, protože nebyly trénovány na českých textech. PERO se přitom umísťuje na prvních místech při soutěžích srovnávajících kvalitu různých OCR systémů. Systém PERO je volně dostupný, kolegové na VUT se jej nyní snaží dále vylepšit a najít ekonomický model, který by jim umožnil jej dále rozvíjet.

Další tři projekty na projekt PERO jistým způsobem navazují v tom smyslu, že je jejich cílem vytvořit AI nástroje pro zrychlení procesu digitalizace odstraněním lidské práce při tvorbě metadat (Smart digilinka), nebo AI nástroje pro lepší vytěžování informací z digitalizovaných dokumentů (Orbis Pictus a semANT).

Hlavními cíli projektu Orbis Pictus je identifikace a kategorizace obrázků nebo jiných objektů na stránkách, identifikace osob na obrázcích a – pokud to bude možné – i jejich pojmenování. Jak ukazují první výsledky, bude možné nejen získat díky kategorizaci obrázků zajímavé datasety různých typů objektů, ale otevřou se tak digitální knihovně i možnosti vyhledávání obrázků podle podobnosti, vyhledávání obrázků podle textového popisu nebo automatické vytváření textových popisů obrázků, které ocení například nevidomí. Co se identifikace osob týče, jedná se o samostatnou úlohu, v rámci které jsou hledána vyobrazení stejných osob (aniž bychom věděli, o koho jde) na různých obrázcích, a následně propojení takto identifikovaných skupin obrázků s databází národních autorit.

Projekt semANT je z hlediska knihovny v základních rysech podobný: jeho cílem je rozčlenit texty digitalizovaných dokumentů na logické celky (např. články, hesla v encyklopedii apod.), takto vzniklé entity tematicky charakterizovat a umožnit jejich vyhledávání podle podobnosti. Součástí projektu je pak i analýza témat z hlediska jejich vývoje v čase.

Mimo výše uvedených projektů řeší MZK jako partner Univerzity Karlovy i projekt OmniOMR, jehož cílem je identifikovat hudební notaci v dokumentech, její převod do počítačově srozumitelné podoby (odtud zkratka OMR – Optical Music Recognition). Projekt se zaměřuje především na zpracování rukopisných hudebnin, jeho hlavním přínosem bude možnost vyhledávat hudební skladby jen podle podobnosti melodie.

Ak by sme začali skladať všetky Vaše projekty ako puzzle, zistili by sme, že koncepčne vytvárajú jeden komplex do seba zapadajúcich aktivít, ktoré sa na seba systémovo nadväzujú. Prezradíte nám, akú technologickú platformu alebo službu uvidí používateľ po zložení všetkých jej dielikov? Aké máte ďalšie dlhodobé vízie na digitálne služby?

Všechny zmíněné aktivity se točí kolem digitální knihovny. Samozřejmě, digitální knihovna obsahuje jen asi 10 % fondů knihovny, proto i snaha proces digitalizace zefektivnit a zrychlit. Pokud bych tedy měl mluvit o vizi ideálního stavu, je to stav, kdy jsou digitalizované fondy všech knihoven, jsou online přístupné na jednom místě a pokud uživatel hledá nějaký dokument nebo informaci, zeptá se na ně přívětivého a pravdomluvného AI kurátora. K tomuto cíli povede ještě dlouhá cesta, ale věřím, že se tato technologická platforma bude stále jmenovat Kramerius, i když to možná bude verze, která se bude značně lišit od té dnešní.

Práce na technologických projektech je nejen personální, ale aj administrativnou výzvou pre štátne a verejné inštitúcie, zvlášť pre knižnice. Akým spôsobom sa snažíte motivovať a stabilizovať kľúčových zamestnancov Vašej knižnice? Aké pracovné pozície majú u Vás vždy otvorené dvere?

Vzhľadom k charakteru našich projektů, které hodně zabíhají do IT oblasti, jsou pro nás klíčoví programátoři nebo lidé s programátorskou zkušeností. Knihovna samozřejmě nemůže nabídnout svým zaměstnancům tak vysoké platy, jaké by mohli dostat v soukromých firmách. Proto se knihovna snaží nabízet svým zaměstnancům takové pracovní podmínky, které by v soukromých firmách neměli. Konkrétně v případě studentských programátorských pozic je naší výhodou fyzická blízkost Fakulty informatiky MU nebo o něco vzdálenější Fakulty informačních technologií VUT, která studentům umožňuje rozdělit efektivně svůj čas mezi práci a studium bez velkých časových ztrát způsobených přesuny.

Na Slovensku, ale i v Českej republike momentálne rezonuje téma centrálnej implementácie, interoperability a dlhodobej udržateľnosti knižničných systémov. Ako vnímate súčasnú situáciu v ČR a SR v kontexte novej implementácie otvorených knižničných systémov ako napríklad Folio? Bude Folio vhodné pre všetky typy knižnic a považujete ho za dostatočne interoperabilné pre prácu s dátovými repozitármi a European Open Science Cloud (EOSC)?

Je pravda, že v systému Folio vidím slibné řešení pro budoucnost. Je to open source systém, za kterým stojí řada firem i knihoven v čele s firmou EBSCO. Zejména fakt, že si tento systém vybrala Kongresová knihovna a investuje do něj na naše poměry obrovské prostředky, je zárukou, že má tento systém před sebou dobrou perspektivu a nestane se slepou uličkou.

Fakt, že Folio má otevřený zdrojový kód, umožňuje jej v případě potřeby rozšiřovat a modifikovat podle aktuálních potřeb, i když i zde je potřeba určité opatrnosti, protože jakákoli rozšíření se buď musejí stát podporovanou součástí Folia, nebo zůstane jejich podpora a aktualizace trvale zátěží pro ty knihovny, které takové rozšíření využívají. Řadu integrací je však možné realizovat pomocí vhodných API, které takovou zátěž snižují.



Aké predpokladáte zmeny vo vedeckej komunikácii a hodnotení vedy v kontexte príchodu umelej inteligencie a otvorenej vedy?

Umělá inteligence může do hodnocení vědy přinést značné změny. První vlaštovky už jsou tu. Například hodnocení citovanosti s přihlédnutím k tomu, z jakého důvodu byla daná práce citována, je velmi významným krokem správným směrem, který se, věřím, povede zavést do běžné praxe hodnocení vědeckých výstupů. V dnešní době již také existuje řada nástrojů na bázi velkých jazykových modelů, které usnadňují vědeckou práci. Jejich použití je však zatím spojeno s určitým rizikem – že například strojem vytvořený souhrn odborného článku nevystihne jeho klíčové závěry nebo dokonce bude obsahovat závěry, které se nebudou shodovat se skutečností. Věřím ale, že se tyto porodní bolesti podaří překonat, a pak vědci dostanou do rukou velmi efektivní nástroje pro svou práci nebo možná i velmi pracovitě kolegy do svého týmu.

Zhováral sa Mgr. Tomáš Fiala.