Projekt GORAZD – problematika digitalizace staroslověnských slovníků[1]

Štefan Pilát, Slovanský ústav AV ČR, v.v.i., Česká republika, pilatste@slu.cas.cz

The aim of the project *GORAZD: Digital portal of the Old Church Slavonic Language* (2016-2020) is to create an electronic portal, which will make the Old Church Slavonic Dictionaries that have originated in the Institute of Slavonic Studies of the Czech Academy of Sciences accessible on-line. It concerns four volumes of the *Dictionary of the Old Church Slavonic language* (1966-1997), into which the *Addenda to the first Volume of the Dictionary of the Old Church Slavonic language* (2000-2016) will be integrated. Furthermore, the portal will contain also the *The Dictionary of the Most Ancient Old Church Slavonic Literary Monuments*, which follows up the one-volume *Старославянский словарь (по рукописям X-XI веков)* (1994) and also the first volume of the *Greek-Old Church Slavonic Index* (2008-2014). At the same time, the digitalized card index to the Old Church Slavonic Dictionary will be made available. In the article, I describe the aims of the project and I discuss the chosen method of the technical solution of the digitalization of these dictionaries. Moreover, the potential of the scientific use of the electronic Old Church Slavonic dictionaries si considered.

Old Church Slavonic Language, Lexicography, Digitalization

Elektronické zpracování slovníků se od 90. let minulého století stává novým standardem lexikografické práce. Výhody jsou nasnadě – pro uživatele je takový slovník libovolného rozsahu přístupný v zásadě kdekoliv a kdykoliv bez potřeby leckdy fyzicky náročné manipulace s objemnými a těžkými svazky tištěných slovníků, umožňuje prakticky okamžité nalezení patřičného hesla a často i další pokročilé funkce prohledávání slovníkového materiálu. Pro tvůrce elektronických slovníků je zase výhodou snadná aktualizovatelnost obsahu, odpadá tak nutnost náročné, zdlouhavé a rovněž finančně nákladné přípravy nových tištěných vydání slovníků. Elektronické slovníky však lexikografům přinášejí také nové výzvy: technickou přípravu slovníku již nerealizuje editor a nakladatel, ale programátor a další IT pracovníci. To staví moderní lexikografy před nutnost sledovat vývojové trendy i v tak dynamicky se rozvíjejícím odvětví, jako jsou moderní informační technologie.

Vedle četných překladových slovníků živých jazyků se postupně elektronického zpracování dočkávají i historické slovníky slovanských jazyků. Způsobem svého zpracování a

1 Studie čerpá z kapitoly „Projekt GORAZD: Digitální portál staroslověnštiny“ kolektivní monografie *Slovník jazyka staroslověnského – historie, osobnosti a perspektivy*. Praha, 2016, s. 117-125. Text studie byl upraven, rozšířen a aktualizován, aby odpovídal současnému stavu projektu.

nabízenými funkcemi jsou však tyto projekty značně různorodé. Nejjednodušší typ digitálních slovníků nabízí pouze vyhledávání příslušeného hesla na základě záhlaví, dále je text hesla nestrukturovaný. Tímto způsobem je zpřístupněn na stránkách Jazykovedného ústavu Ľudovíta Štúra Slovenskej akadémie vied pátý díl *Historického slovníku slovenského jazyka (rab – švrkotať* (HSSJ)[2], obdobným způsobem je zpřístupněno i deset doposud vydaných dílů *Slovníku staroruského jazyka (XI.-XIV. st.)* (SDRJ)[3]. Pokročilejší způsob zpracování je patrný u elektronické verze doposud vzniklých dílů *Slovníku ruského jazyka XVIII století* (SRJ18) zpřístupněného v rámci projektu *Фундаментальная электронная библиотека «Русская литература и фольклор»*[4]. Hesla jsou již strukturována a je možné filtrovat výsledky vyhledávání podle vnitřní strukury hesel a podle slovního druhu. Pokročilý způsob zpracování vykazuje také elektronická verze *Slovníku jazyka polského XVII. a XVIII. století* (SJP17)[5] vytvářená Ústavem polského jazyka Polské akademie věd. Hesla jsou v něm přehledně strukturovaná a obsahují i tabulky doložených gramatických tvarů, jsou přítomny prokliknutelné odkazy na synonyma a informace o fázi zpracování hesla; zpětnou vazbu s uživateli nabízí též možnost diskuse u každého hesla. K nejrozsáhlejším a nejambicioznějším projektům tohoto typu náleží jistě portál *Vokabulář webový*[6], který vytváří Oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i. Jeho slovníková část integruje vyhledávání hesel ze všech doposud vydaných slovníků staré češtiny: *Malý staročeský slovník* (MSS), *Slovník staročeský* Jana Gebauera (Gabauer 1913), *Staročeský slovník*, sešit 1–26 (StčS), *Slovníček staré češtiny* Františka Šimka (Šimek 1947), Index *Slovníku staročeských osobních jmen* Jana Svobody a *Heslář lístkového materiálu ke StčS*. V rámci *Elektronického slovníku staré češtiny* (ESSČ) jsou nadále zpracovávána a do databáze zařazována nová hesla. Součástí portálu jsou též pokročilé funkce prohledávání slovníkového materiálu, přístupná je možnost volby prohledávaných databází, dále je možno vyhledávat podle významu, slovního druhu, valence atd.

Zvláštním oddílem slovanských historických slovníků jsou slovníky staroslověnské a církevněslovanské. Nejstarší staroslověnský slovník, slovník Miklošičův (Miklosich 1865), je zpřístupněn v rámci portálu *Monumenta serbica*[7]; je však poněkud udivující, že na těchto stránkách není možné dohledat autory tohoto poměrně rozsáhlého a profesionálně vyspělého

2 [cit. 2018-21-05] <http://www.juls.savba.sk/hssj.html>
3 [cit. 2018-21-05] <http://slovari.ru/default.aspx?s=0&p=2641>
4 [cit. 2018-21-05] <http://feb-web.ru/feb/sl18/slov-abc/>
5 Tiskem vyšlo pouze pět sešitů, od roku 2004 je slovník tvořen jen elektronicky: [cit. cit. 2018-21-05] <http://sxvii.pl>, samostatně je též zpřístupněna kompletní lístková kartotéka k tomuto slovníku: [cit. cit. 2018-21-05] <http://www.rcin.org.pl/publication/20029>
6 [cit. 2018-21-05] <http://vokabular.ujc.cas.cz/>
7 [cit. 2018-21-05] < http://www.monumentaserbica.com/mikl2/main.php>

projektu. Pro zápis textu slovníku jsou užívány moderní UNICODové fonty, slovník nabízí též pokročilé vyhledávací funce, výsledky lze filtrovat podle slovních druhů a hesla lze prohledávat i podle latinských překladů a řeckých ekvivalentů.

Svého elektronického zpracování se v rámci projektu *Cyrillomethodiana*[8] Sofijské univerzity sv. Klimenta Ochridského dočkal též dvoudílný *Starobulharský slovník* (SBR). Tento slovník je okruhem zahrnutých památek a způsobem zpracování blízký jednodílnému *Staroslověnskému slovníku* (SS), narozdíl od něj však zpracovává i četný epigrafický materiál a překlady staroslověnských hesel uvádí jen v současné bulharštině. Jeho elektronická verze prozatím umožňuje pouze přístup k heslům pomocí záhlaví. Cíle projektu jsou však mnohem ambicioznější: odkazy u lokací památek jsou již připraveny na budoucí propojení s korpusem citovaných textů a v přípravě jsou rovněž pokročilé vyhledávací funkce. Digitalizovaný SBR je také vzat jako základ šířeji pojatého historického slovníku bulharštiny, jehož materiál je postupně v elektronické verzi doplňován (nově přidaná hesla jsou barevně odlišena). Užite

ným nástrojem připojeným k historickému slovníku bulharštiny je analyzátor gramatických tvarů ohebných slov, současně je možné u jednotlivých záhlaví vygenerovat celé paradigma včetně pravopisných variant.

Cíle projektu GORAZD

Cílem projektu *GORAZD: Digitální portál staroslověnštiny (2016-2020)*[9] je vytvořit elektronický portál, který bude on-line formou zpřístupňovat staroslověnské slovníky, které vznikly nebo vznikají na půdě Slovanského ústavu AV ČR v Praze, a taktéž zpřístupní unikátní staroslověnskou lístkovou kartotéku, na níž jsou tato lexikografická díla založena. Díky digitalnímu zpracování budou moci být tyto slovníky nejen uchovány a zpřístupněny mnohem širšímu okruhu uživatalů, než jsou v současnosti jejich často obtížně dostupné tištěné verze, ale taktéž budou moci být nadále inovovány, doplňovány a vylepšovány, což umožní jejich udržování vždy na úrovni aktuálního vědeckého poznání.

Jádrem projektu je připravovaná digitalizace ***Slovníku jazyka staroslověnského*** (dále SJS). Tento čtyřsvazkový slovník zahranuje nejen úplný výčet slovní zásoby památek kanonické staroslověnštiny, ale též mladších církevněslovanských opisů památek vzniklých v cyrilometodějském období nebo majících s ním těsný vztah. Do fondu tohoto slovníku jsou taktéž zařazeny památky českocírkevněslovanské a jejich mladší opisy z jiných slovanských

8 [cit. 2018-21-05] <http://histdict.uni-sofia.bg/>

9 Projekt je financován z grantu Ministerstva kultury ČR, program NAKI II, č. grantu DG16P02H024, příjemcem je Slovanský ústav AV ČR, v.v.i.

prostředí. V letech 2010-2016 vyšly po osmi sešitech *Dodatky k I. dílu SJS* (DSJS), jež jsou koncipovány jako jeho V. díl. V rámci projektu *Gorazd* budou tyto dodatky a opravy přímo zapracovány do hesel SJS, čímž bude naplněn cíl udržování jeho hesel v aktuálním stavu. Výhledově je počítáno s tím, že po dokončení digitalizace SJS budou podle metodiky *Dodatků* již přímo v digitální podobě zpracována hesla II.-IV. dílu, aby slovník zůstal co do způsobu svého zpracování homogenní ve všech svých částech. Výraznou proměnou digitální verze SJS oproti jeho tištěnému originálu bude též změna jazyka lexikologických, gramatických a faktografických výkladů, jímž byla doposud latina. Pro zpřístupnění obsahu SJS co nejširšímu okruhu uživatelů v mezinárodním měřítku budou nově v digitální verzi SJS tyto výkladové části slovníku poskytovány v angličtině. Doplněny budou rovněž anglické překlady významů hesel, které jsou doposud uváděny v češtině, ruštině a němčině.

Hesla v elektronické verzi SJS budou strukturována tak, aby byly umožněny pokročilé vyhledávací funkce a různé možnosti filtrování výsledků. Omezit výběr prohledávání bude možné podle gramatických charakteristik hesel (u sloves např. podle prézentní třídy a vidu, u substantiv např. podle rodu a kmene), dále též podle výskytu v jednotlivých památkách nebo podle skupin památek na základě jejich obsahu (evagelia, žaltáře atd.) a původu (kanonické památky, postkanonické památky, památky českocírkevněslovanské). Bude možné též obrácené vyhledávání podle českých, anglických, ruských a německých překladových ekvivalentů a také podle ekvivalentů latinských, řeckých a starohornoněmeckých. V plném rozsahu hesel bude možné fulltextové vyhledávání. Tyto pokročilé funkce též poskytnou cenné a nové statistické údaje o celkovém počtu hesel spadajících do té či oné kategorie. Bude tak např. možné s úplnou přesností stanovit poměry jednotlivých deklinačních či konjugačních typů nejen v rozsahu celého SJS, ale taktéž samostatně pro každou památku v něm zařazenou. Postupně budou s jednotlivými hesly SJS propojovány lístky naskenované staroslověnské kartotéky, díky čemuž bude uživatel moci dohledat všechny výskyty daného lexému i u hesel, která nejsou exhaustována. Celá lístková kartotéka bude paralelně dostupná ve formě samostatné databáze, čímž bude zpřístupněn tento unikátní archivní materiál všem badatelům.

Digitální verze SJS do budoucna otevře možnost snadného doplňování dalších památek a revize zařazení památek stávajících. V první řadě čeká na doplnění část staroslověnských památek kanonických, které byly nalezeny roku 1975 v klášteře sv. Kateřiny na Sinaji. Jedná se o *Dimitrijův žaltář* před nedávnem fototypicky vydaný společně se *Sinajskými medicinskými lístky* (Miklas 2012) a též o doposud nevydaný tzv. *Sinajský misál* (Tarnanidis 1988, 5/N, str. 103-108). Již nyní je možné uvažovat i o doplnění některých

dalších památek českocírkevněslovanského původu – těmi jsou např. *Legenda o sv. Anastázii* (Čajka 2011) nebo *Modlitba sv. Řehoře* či *Modlitba vyznání hříchů* (Vepřek 2013). Budoucí revizi si jistě zaslouží i stávající zařazení textů v SJS. Tak například namísto jen částečně excerpovaného a do SJS jen výběrově zařazeného *Nikoľského evangelia* (Daničić 1864), bude lepší v úplnosti zařadit některé z dalších čtveroevangelií archaického typu, v úvahu připadá zejména *Typografské* nebo *Haličské evangelium* (Aleksejev 1998, 36-37). Namísto zastaralého Kałužniackého vydání *Christinopolského apoštolu* (Kałužniacki 1846), které představuje mladší redakci textu, bude vhodnější apoštol citovat podle některého z přepisů archaické redakce, snad zcela nově podle *Crkolezského apoštolu* (Bogdanović 1986), možná však bude efektivnější použít *Hilferdingův apoštol č. 14*[10], který je již z Kałužniackého edice částečně do SJS zařazený. Podle nejnovějších poznatků bude též třeba prověřit i zařazení stávajících starozákonních textů do SJS a zvážit i případné doplnění některými dalšími památkami. Tyto a mnohé další úkoly bude schopna otevřít budoucím zpracovatelům právě úspěšná digitalizace SJS.

V rámci projektu *Gorazd* se převedení do digitální podoby dočká též ***Slovník nejstarších staroslověnských památek*** (dále SNSP*)*, který v roce 1994 vyšel pod názvem *Старославянский словарь (по рукописям X-XI веков)* (SS). Oproti SJS tento jednosvazkový slovník zpracovává výlučně jen slovní zásobu staroslověnských kanonických památek, struktura hesel je zjednodušená, což odpovídá jeho primárnímu určení zájemcům z řad studentů a veřejnosti. Digitalizovaná verze tohoto slovníku však bude navazovat na připravované druhé doplněné a opravené vydání tohoto slovníku, které bude zahrnovat nejen doplnění hesel podle nově nalezených památek, ale nově též anglické ekvivalenty významů hesel vedle současných českých a ruských. Probíhá převod výkladového aparátu tohoto slovníku z ruštiny do češtiny, což učiní tento slovník snadno přístupný českým studentům a dalším zájemcům z řad veřejnosti, a tak bude zachován a rozšířen jeho původní účel. Údajem, který má SNSP oproti SJS navíc, je uvádění počtu výskytů u každého hesla. Do počtu 100 je uváděno přesně, u větších počtů zaokrouhleně na stovky. Díky tomu bude možné filtrovat výsledky vyhledávání v SNSP i podle frekvenčního indexu. Plánovaná možnost paralelního zobrazení hesel ze SJS a SNSP pak umožní snadné porovnávání způsobu jejich zpracování v obou lexikografických dílech.

V roce 2014 byl vydáním osmého sešitu dokončen I. díl *Řecko-staroslověnského indexu* (ŘSI). Tento první díl (v rozsahu hesel α΄ – γωνιᾶιος) bude rovněž v rámci projektu *Gorazd* digitalizován a zpřístupněn. S dalším postupem prací na tomto lexikografickém díle je

10 Российская национальная библиотека (Санкт-Петербург), Гильф. 14.

počítáno i s postupným rozšiřováním jeho elektronické verze. Důležitou funkcí bude postupné obousměrné projování hesel ŘSI s hesly v SJS, díky čemuž budou oba slovníky navzájem prostupné. V elektronické verzi ŘSI budou hesla strukturována tak, aby byly poskytnuty pokročilé vyhledávací funkce nejen podle řeckých záhlaví, ale též řeckých podzáhlaví, příslušnosti řeckých hesel k slovnímu druhu a podle záhlaví staroslověnských odstavců. Fulltextově bude též možné vyhledávat jednotlivé rukopisné lokace a biblická místa, pokud jsou v textu hesel citována.

Technické řešení

První návrh digitalizace SJS podalo již v roce 1992 nakladatelství LEDA s. r. o., vypracoval jej RNDr. Rudolf Červenka. V návrhu se počítalo nejen s elektronickým zpracováním k tisku tehdy ještě chybějících posledních sešitů IV. dílu, ale také s digitalizací již vydaných částí a následnou distribucí celého SJS na CD-ROM. Již tehdejší návrh počítá s využitím technologie OCR pro převedení obsahu slovníku do elektronické podoby (bez bližší specifikace technických podrobností), taktéž dr. Červenka navrhuje obohacení obsahu slovníku doplněním anglických překladů a také využitím dalších pramenů (bez upřesnění). Pro ilustraci dobových technických poměrů můžeme uvést, že nakladatelství LEDA doporučuje autorské pracoviště vybavit čtyřmi počítači typu IBM PC 286 s pevným diskem 40 MB a monochromatickou obrazovkou. V porovnání se současnými technickými možnostmi byl tehdejší záměr digitalizace tak komplikovaného díla, jako je SJS, vskutku mimořádně odvážný, nicméně k jeho realizaci nedošlo ani v podobě přípravných či testovacích prací.[11]

Digitalizace slovníků historických jazyků přinášela a stále ještě přináší četná úskalí, protože mnohé staré jazyky užívají zvláštní abecedy nebo znaky, které donedávna nebyly nebo ještě dodnes nejsou plně zahrnuty do mezinárodně přijatých standardů. Tento problém se ještě v nedávné době týkal i staroslověnštiny. Pro její počítačový zápis byla užívána řada různých vzájemně nekompatibilních fontů, bez jejichž instalace nebyl staroslověnský text obvykle vůbec čitelný. Absence obecně přijatého standardu značně komplikovala i vývoj pokročilejších softwarových řešení schopných se staroslověnskými texty pracovat, i ty totiž byly závislé na rozložení znaků v konkrétním fontu. Situaci neusnadňoval ani fakt, že staroslověnština užívá abecedy dvě: hlaholskou a cyrilskou. Až mezinárodní standard

11 V propagačním materiálu k *Slavnostní prezentaci díla Slovníku jazyka staroslověnského / Lexicon linguae palaeoslovenicae*, která se konala u příležitosti dokončení tohoto slovníku 29.9.1997 v budově Kanceláře AV ČR, je ještě zmiňováno uvažované „doplněné kompletní vydání SJS, ať už v podobě knižní či v případném zpracování počítačovém (elektronickém), pokud se najde finanční řešení.“

UNICODE 8.0 z června roku 2015 zahrnuje většinu cyrilských a hlaholských znaků potřebných pro zápis staroslověnských textů. Staroslověnštinu lze již díky tomuto standardu zapisovat nezávisle na použitém fontu, čímž se otevřely dveře pro systematické a mezinárodně standardizované digitální zpracování staroslověnských materiálů. V projektu *Gorazd* byl použit unicodový font *CyrillicaBulgarian10U*, užívaný též v příbuzném projektu *Cyrillomethodiana*, který řešitelskému týmu laskavě poskytla prof. Anna-Marija Totomanova z Univerzity sv. Klimenta Ochridského v Sofii.

OCR zpracování SJS

I po téměř 25 letech po vzniku prvního návrhu na digitalizaci SJS zůstával základní technický problém tentýž: SJS existoval pouze v tištěné verzi, přičemž se jedná o mimořádně náročný text obsahující šest hlavních jazyků (staroslověnština, stará řečtina, latina, čeština, ruština, němčina; vedle toho místy i stará horní němčina, hebrejština a aramejština), čtyři hlavní abecedy (latinka, řecká alfabeta, ruská cyrilice, staroslověnská cyrilice; vedle toho vzácně i hlaholice a hebrejská abeceda) a složitě strukturovaná hesla. Možností byl buď jeho ruční přepis do textového editoru nebo použití metody automatického rozpoznávání textu (OCR). První varianta by při rozsahu a složitosti SJS znamenala takovou míru nasazení lidských zdrojů a tím i finanční a časovou náročnost, že byla již v prvních úvahách o projektu vyloučena. Jít cestou OCR zpracování však představovalo výzvu inovativního řešení nových technických problémů: automatické rozpoznávání staroslověnštiny není současnými OCR aplikacemi podporováno, taktéž rozpoznávání staré řečtiny má svá omezení. Jako součást projektu tedy vznikl cíl vytvoření specializovaných nástrojů, které rozpoznávání staroslověnštiny na běžně dostupném OCR produktu FineReader 12 (nebo novější) společnosti ABBYY umožní. Tyto nástroje bude možné dále přizpůsobit pro automatické rozpoznávání jakéhokoliv tištěného staroslověnského textu, např. edicí staroslověnských a církevněslovanských památek. Součástí těchto nástrojů budou definice staroslověnštiny a staré řečtiny jako uživatelských jazyků s kompletním výčtem všech potřebných znaků podle tabulek nejnovějšího standardu UNICODE (v tuto chvíli /22.5.2018/ verze 10.0, zveřejněna je již betaverze 11.0)[12]. Pro zkvalitnění výsledků rozpoznávání bude též sestaven specializovaný staroslověnský slovník pro OCR, jehož materiál bude vycházet z citovaných dokladů chystaného druhého vydání SNSP, dále z kanonických staroslověnských památek, jež máme k disposici v UNICODE textové podobě, jako je *Zografské* a *Mariánské čtveroevangelium* a *Sborník Supraslský*, z památek církevněslovanských to bude *Čtyřicet homilií Řehoře Velikého*

12 [cit. 2018-05-22] <http://unicode.org>

na evangelia. Úspěšnost rozpoznávání textu OCR taktéž výrazně navýší funkce učení uživatelských vzorů, která umožní nastavit rozpoznávání přesně podle typů užívaných v tištěném SJS. Nicméně i tak probíhají rozsáhlé korektury takto získaného textu, kvůli složitosti struktury SJS nemohou být výsledky nikdy zcela bezchybné. Avšak i takto využití OCR technologie znamená výraznou úsporu jak časovou, tak finanční. Výstupním formátem automaticky rozpoznaného textu SJS jsou soubory formátu ALTO (XML)[13], který byl vyvinut americkou *Knihovnou Kongresu* (Library of Congress) speciálně pro digitalizaci knižních publikací. Výhodou tohoto formátu je zachování přesné informace o umístění textu na stránce, což významně usnadněje korektorské práce. Pro vytvoření výstupu v tomto formátu je využíván produkt ABBYY Recognition Server 4.0.

Vývoj nových aplikací pro digitaliazaci slovníků

V rámci projektu *Gorazd* budou nově vyvinuty tři specializované nástroje využitelné pro automatizaci digitalizace vícejazyčných slovníků výkladového či encyklopedického charakteru. Tyto nástroje budou volně dostupné na základě licence GNU GPL 3 a budou uzpůsobené pro práci s dokumentografickým systémem INVENIO[14], což je open sourcový framework pro správu digitálních knihoven, vyvíjený CERNem (Evropská organizace pro jaderný výzkum). V tomto dokumentografickém systému, který je přizpůsoben speciálním potřebám historických slovníků, jsou provozovány též slovníkové databáze vzniklé v rámci projektu *Gorazd.* Obsah těchto databází bude koncovým uživatelům zpřístupněn prostřednictvím prezentačního systému Gulliver[15], vyvíjeným firmou AiP Beroun s. r. o.

Aplikace nově vytvořené v rámci projektu Gorazd:

- **Gorazd Generátor hesel** je pythonová aplikace umožňující generovat hesla podle definovaných formálních pravidel. Za tímto účelem inovativně využívá formální gramatiku zapsanou v jazyce knihovny ANTLR. Díky tomu je možné automatizovaně vytvářet hesla různé hloubky a struktury z výchozího nestrukturovaného textu, který je získán např. zpracováním tištěného materiálu pomocí technologie OCR nebo přímým importem z textového souboru. To umožňuje provádět digitalizace slovníků historických jazyků bez nutnosti přepisu jednotlivých hesel lidskou silou, což výrazně zefektivní a urychlí převod takových publikací do strukturované digitální podoby. Výsledné XML soubory jsou primárně

13 [cit. 2018-05-22] <https://www.loc.gov/standards/alto>
14 [cit. 2018-05-22] <http://invenio-software.org>
15 [cit. 2018-05-22] <http://www.aipberoun.cz/?q=node/15>

určené pro dokumentografický systém INVENIO, mohou být však využitelné i jiným způsobem.

- **Gorazd Editor hesel** je webová aplikace integrovaná do dokumentografického systému INVENIO, která umožňuje efektivní prohlížení a opravu výsledků zpracovaných Generátorem hesel včetně zobrazení obrazových originálů (pokud text vznikl pomocí technologie OCR) a také tvorbu nových hesel, tudíž je tento software použitelný i nezávisle na Generátoru hesel. Při využití formátu ALTO (XML) Editor hesel označuje v obrazové předloze editovaný text, což usnadňuje korektury a redakční úpravy. Z Editoru hesel je možné k heslům uložených v databázích znovu přistupovat, editovat jejich obsah a měnit jejich strukturu. Pro vkládání neobvyklých znaků obsahuje nástroj virtuální klávesnice. Při ukládání editovaného hesla vytváří v rámci systému INVENIO vždy nový XML soubor, proto je možné se vždy vrátit k dříve uloženým verzím a sledovat editační historii, což má svůj význam zejména při redakčních úpravách hesel. Tato webová aplikace je přizpůsobena pro rozšířený webový prohlížeč Google Chrome, díky tomu mohou pracovníci s Editorem hesel pracovat na jakémkoliv běžném počítači připojeném k internetu bez nutnosti instalace speciálního softwaru.
- **Gorazd Export hesel** je pythonová apliakce, jejímž učelem je exportovat obsah slovníkových databází uložených v dokumentografickém systému INVENIO za účelem tisku nebo webové prezentace hesel. Hesla je možné exportovat buď jako celek nebo jen určitý výběr. Pomocí XSL transformací je text hesel možné požadovaným způsobem naformátovat. Data jsou exportována ve formátu HTML. V případě předtiskové přípravy je vytvořen jeden HTML soubor, který je možné otevřít v běžných textových editorech, např. Microsoft Word, a přeuložit jej do standardních formátů dokumentů, např. doc(x) nebo rtf. Tímto způsobem je možné získat snadno a efektivně podklady pro tištěné verze slovníků historických jazyků uložených v dokumentografickém systému INVENIO.

♦

Projekt Gorazd představuje ambiciózní plán na interdisciplinární propojení historické lexikografie a nejmodernějších výdobytků informačních technologií. Současně umožní zpřístupnění uvedených staroslověnských slovníků širokému okruhu veřejnosti a také poskytne nové možnosti využití jejich obsahu, který bude do budoucna snadno aktualizovatelný a vždy jej bude možné poskytovat na úrovni současného vědeckého poznání. Softwarová platforma, na které bude systém vystavěn, bude open sourcová a otevřená pro

další rozšiřování. Z technologického hlediska proto nic nebrání budoucímu přidávání dalších staroslověnských a církevněslovanských slovníků postavených na koncepci blízké SJS, což otvírá cestu k případnému budoucímu vytvoření integrovaného tezauru staroslověnštiny a církevní slovanštiny všech redakcí na základě mezinárodní spolupráce vědeckých institucí.

Seznam použité literatury:

DSJS: *Slovník jazyka staroslověnského.* Díl V. Addenda et corrigenda. Seš. 53-58. Praha: Slovanský ústav AV ČR, v.v.i., 2010-2015.

ESSČ: *Elektronický slovník staré češtiny.* Vajdlová a kol. [online]. [cit. 2018-21-05] <http://vokabular.ujc.cas.cz/informace.aspx?t=ESSC&o=slovniky>

HSSJ: *Historický slovník slovenského jazyka*, Díly 1-7. Bratislava: Veda 1991-2008.

MSS: *Malý staročeský slovník*. Bělič, J. – Kamiš, A., – Kučera, K. Praha: Státní pedagogické nakladatelství 1978.

ŘSI: *Řecko-staroslověnský index.* Seš. 1-8. Praha: Slovanský ústav AV ČR, v.v.i., 2008-2014.

SBR: *Старобългарски речник.* Иванова-Мирчева et al. T. I (А–Н). T. II (О–У). София: Валентин Траянов 2009.

SDRJ: *Словарь древнерусского языка (XI—XIV вв.).* Вып. 1-10. Москва: Институт русского языка им. В. В. Виноградова РАН 1988-2013.

SJP17: *Słownik języka polskiego XVII i 1. połowy XVIII wieku.* Zesz. 1-5, Kraków: Instytut Języka Polskiego PAN 1996-2004.

SJS: *Slovník jazyka staroslověnského.* Díly I.-IV. Praha: Academia 1966-1997.

SNSP: *Slovník nejstarších staroslověnských pomátek.* Lexikální databáze připravovaná v rámci projektu Gorazd.

SRJ18: *Словарь русского языка XVIII века.* Вып. 1-19. Санкт-Петербург: Наука 1984-2011.

SS: *Старославянский словарь (по рукописям X-XI веков).* Bláhová a kol. Москва: Русский язык 1994.

StčS: *Staročeský slovník*, sešit 1–26 (N–při). Praha: Academia 1968–2008.

Aleksejev 1998: Алексеев, А. А. et al. *Евангелие от Иоанна в славянской традиции.* Санкт-Петербург: Российское библейское общество.

Bogdanović 1986: Богдановић, Д. – Велчева, Б. – Наумов, А. *Болгарский апостол XIII века: рукопись Дечани-Црколез 2.* София.

Čajka 2011: Čajka, F. *Církevněslovanská legenda o svaté Anastázii*. Praha: Slovanská ústav AV ČR, v.v.i., 2011.

Daničić 1864: Даничић, Ђ. *Никољско јеванђеље.* Београд: Државна штампарија 1864.

Gabauer 1913: Gabauer, J. *Slovník staročeský.* Díl I (A–J), díl II (K–N). Praha: 1900-1913, reprint: Praha: Academia 1970.

Kałużniacki 1846: Kałużniacki, Ae. *Actus epistlaeque apostolorum palaeoslovenice ad fidem codicis Christinopolitani saeculo XIIo scripti*. Vindobonae.

Miklas 2012: Miklas, H. *Psalterium Demetrii Sinaitici (monasterii sanctae Catherianae codex slav. 3/N) adiectis foliis medicinalibus.* Wien: Holzhausen 2012.

Miklosich 1865: Miklosich, F. *Lexicon Palaeoslovenico–Graeco-Latinum.* Vindobonae: 1862-1865, reprint: Aalen: Scientia Verlag, 1977.

Šimek 1947: Šimek, F. *Slovníček staré češtiny.* Praha: Matice česká, Orbis 1947.

Tarnanidis 1988: Tarnanidis, I. *The Slavonic Manuscripts discovered in 1975 at St. Catherine's Monastery on Mount Sinai.* Thessaloniki.

Vepřek 2013: Vepřek, M. *Modlitba sv. Řehoře a Modlitba vyznání hříchů v církevněslovanské a latinské tradici.* Olomouc: Univerzita Palackého v Olomouci.