Evropské projekty digitalizace historických slovníků se představily v Praze

Tisková zpráva Slovanského ústavu AV ČR, v. v. i., Praha, 27. dubna 2018

**Na půdě Slovanského ústavu AV ČR, v. v. i. jsou již od roku 2016 v rámci projektu *GORAZD: Digitální portál staroslověnštiny* (MK ČR, program NAKI II, DG16P02H024) digitalizovány staroslověnské slovníky a kartotéka, vytvářené na tomto pracovišti od počátku druhé poloviny 20. století. Zpřístupnění těchto děl široké odborné i laické veřejnosti je plánováno do konce roku 2020. Součástí projektu je i vytvoření a bezplatné zveřejnění softwarových nástrojů, které výrazně zjednoduší tvorbu náročných lexikografických děl, a to jak digitálních, tak tištěných. Projektový tým svolal na 26. dubna mezinárodní workshop „Staroslověnština v digitálním věku“, konaný v Akademickém centru AV ČR v Praze, na němž byly kromě dílčích výsledků projektu GORAZD představeny i další tuzemské a zahraniční projekty, zabývající se digitalizací historických slovníků. Zajímavým osvěžením setkání byla úvodní dramatická scénka, která byla sehrána přímo ve staroslověnštině.**

# Co je projekt GORAZD?

Téměř polovina dne, během něhož se konal workshop **„Staroslověnština v digitálním věku“** byla vyhrazena prezentaci východisek a dílčích výsledků projektu *GORAZD: Digitální portál staroslověnštiny*. Jak uvedl vedoucí projektu PhDr. Štefan Pilátem, Ph.D., hlavními cíli projektu jsou digitalizace a online zpřístupnění zásadních děl české slavistiky vzniklých na půdě Slovanského ústavu: čtyřdílného ***Slovníku jazyka staroslověnského*** (1966-1997) včetně dodatků (vytvářených od roku 2008), který je v současnosti největším slovníkem tohoto jazyka na světě, ukázkové části nového vydání ***Slovníku nejstarších staroslověnských památek***(1. vydání 1994), prvního vydaného svazku ***Řecko-staroslověnského indexu***(2014) a jedinečné staroslověnské kartotéky čítající okolo **milionu kartotéčních lístků**.

Nedílnou součástí projektu jsou také softwarové nástroje na bázi opensourcového dokumentografického systému *INVENIO*, mezi něž patří **Generátor hesel**, který umožňuje automatické generování slovníkových statí na základě vstupu z OCR nebo v textovém formátu, **Editor hesel**, který nabízí uživatelsky velmi přátelské prostředí pro opravu, editaci a psaní hesel a **Exportní modul**, který umožňuje prakticky okamžité vytvoření výstupu z databáze do podoby, z níž lze zhotovit tištěný slovník. Zdrojový kód těchto nástrojů bude uvolněn do konce roku 2018. Vedoucí projektu během workshopu také praktickými ukázkami představil **prezentační systém** GULLIVER (vyvíjený firmou AiP Beroun s. r. o.), pomocí něhož bude možné se slovníky pracovat. Samozřejmostí tohoto systému jsou pokročilé vyhledávací nástroje a prolinkovanost hesel.

Další členové řešitelského týmu se zaměřili na podrobnější představení vzniku a specifik digitalizace jednotlivých slovníků a staroslověnské kartotéky. Tato bude zpřístupněna jednak v samostatném prezentačním systému, kde bude možné pohodlně kartotékou listovat, jednak budou jednotlivé lístky propojeny s příslušnými hesly, čímž se projekt GORAZD zařadí mezi špičku v rámci digitální historické lexikografie. Během jednotlivých příspěvků k projektu byly zdůrazněny i další inovace a přednosti digitalizace zmíněných děl. Pro zvýšení dostupnosti díla pro světovou veřejnost jsou veškeré popisky v digitalizovaném *Slovníku jazyka staroslověnského* překládány do angličtiny, k stávajícím třem současným a třem historickým jazykům přibudou i anglické ekvivalenty ke všem heslům a významům.

# Stav a výzvy současné digitální historické lexikografie

Prvním z hostů workshopu byl Mgr. Vít Boček, Ph. D., z etymologického oddělení Ústavu pro jazyk český AV ČR, v. v. i., který podrobil analýze různé možnosti digitalizace a online zpřístupnění právě dokončovaného ***Etymologického slovníku jazyka staroslověnského***, který materiálově čerpá ze *Slovníku jazyka staroslověnského*. Připomenul, že i během historie tištěného vydání (od r. 1989) tohoto obsahově i koncepčně ojedinělého etymologického slovníku byla vystřídána řada textových editorů, takže by bylo nutno pro různé části díla užít jiné digitalizační metody.

Obsahově se *Slovníku jazyka staroslověnského* nejvíce blíží projekt ***HISTDICT***Sofijské univerzity sv. Klimenta Ochridského (<http://histdict.uni-sofia.bg/>), který představila Dr. Gergana Ganeva. Ambiciózním cílem tohoto projektu je rozšíření původního staroslověnského slovníku o lexémy z mladšího, církevněslovanského období a jejich propojení s korpusem původních textů. Dr. Ganeva upozornila na problémy se sháněním odborných pracovních sil, které jsou pro takové projekty zcela nezbytné.

Prof. Viktor Baranov z Iževské státní technické univerzity (<http://manuscripts.ru/>) představil mezinárodní projekt ***Manuskript***, který zahrnuje rozsáhlý korpus církevněslovanských textů vzniklých především ve východoslovanském prostředí. Prof. Baranov ukázal pokročilé možnosti vytváření badatelského virtuálního prostředí, které umožňuje automatickou korpusovou analýzu velkého množství rukopisů. Zmínil také problematiku autorských práv na přepisy rukopisů a apeloval na důležitost technické aktualizace databází za účelem maximálního prodloužení životnosti projektu.

Zkušenostmi ze zpracování ***Elektronického slovníku staré češtiny***, přístupného prostřednictvím webového portálu Vokabulář webový (<http://vokabular.ujc.cas.cz/>), se podělila PhDr. Miloslava Vajdlová z oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v.v. i., která mj. seznámila přítomné se současným stavem vědeckého hodnocení digitálních výsledků.

Projekt digitalizace ***Slovníku středověké latiny v českých zemích*** (<http://lb.ics.cas.cz/>) představili zástupci Kabinetu pro klasická studia Filosofického ústavu AV ČR, v. v. i. Mgr. Pavel Nývlt, Ph.D., a Mgr. Jan Ctibor upozornili na problematiku převodu správné segmentace heslových statí do elektronické podoby a přítomné nadchli precizně zpracovanou digitální kartotékou, čímž poukázali na důležitost předdigitalizační kvalitativní úpravy publikovaného materiálu.

Posledním přednášejícím byl dr. Krzysztof Nowak z Ústavu polského jazyka Polské akademie věd v Krakově (Polsko), jehož důkladný příspěvek ukázal výhody využívání mezinárodního standardního formátu TEI pro tvorbu ***Elektronického slovníku středověké latiny v Polsku*** (<http://scriptores.pl/elexicon/>).

Účastníci se shodli na důležitosti a přínosnosti podobných, úzce zaměřených workshopů, které napomáhají mezinárodnímu sdílení zkušeností s digitálními lexikografickými projekty. Workshop totiž jasně ukázal podobnost problémů, s nimiž se jednotlivé řešitelské týmy potýkají, a zároveň ilustroval různé metody jejich překonávání.

# Odkaz pro doplnění informací k projektu GORAZD

[Čajka, F., Chromá, M., Knoll, V., Pilát, Š., Spurná, K. *GORAZD: Digitální portál staroslověnštiny.* Edice Věda kolem nás. Co to je... Praha: Academia 2017.](http://www.vedakolemnas.cz/miranda2/m2/sys/galerie-download/VKN_74WEB.pdf)

# Kontakt pro média

Mgr. Vladislav Knoll, Ph.D., tel: 224 800 259, e-mail: [knoll@slu.cas.cz](mailto:knoll@slu.cas.cz)