Štátna vedecká knižnica v Banskej Bystrici

Pavol Maliniak – Imrich Nagy

Digital humanities

nástroje sprístupňovania historického dedičstva

Zborník abstraktov

Banská Bystrica 2022

ISBN 978-80-8227-012-2

Zborník abstraktov z vedeckej konferencie s medzinárodnou účasťou *Digital humanities – nástroje sprístupňovania historického dedičstva* konanej v Štátnej vedeckej knižnici v Banskej Bystrici v dňoch 12. a 13. októbra 2022.

Compilation of abstracts from the scientific conference with international participation *Digital humanities – tools for making historical heritage accessible*, held at the State Scientific Library in Banská Bystrica on 12-13 October 2022.

AGENTÚRA
NA PODPORU
VÝSKUMU A VÝVOJA

Tento text je výstupom z riešenia projektu APVV-19-0456 SKRIPTOR – Inovatívne sprístupnenie písomného dedičstva Slovenska prostredníctvom systému automatickej transkripcie historických rukopisov.

This text is an output from the project APVV-19-0456 SKRIPTOR – Innovative access to the written heritage of Slovakia through a system of automatic transcription of historical manuscripts.

Odborný garant konferencie: prof. PhDr. Dušan Katuščák, PhD.

Recenzenti: prof. Ing. Milan Konvit, PhD.
Jan Odstrčilík, Ph.D.

ISBN 978-80-8227-012-2

Štátna vedecká knižnica v Banskej Bystrici

Digital humanities

Nástroje sprístupňovania historického dedičstva

Zborník abstraktov

Ed. Pavol Maliniak – Imrich Nagy

Banská Bystrica 2022

OBSAH

CONTENTS

Predhovor

Zborník abstraktov je sprievodným výstupom vedeckej konferencie *Digital humanities – nástroje sprístupňovania historického dedičstva* konanej v Banskej Bystrici v dňoch 12. a 13. októbra 2022. Konferencia a zborník sú plánovanými výstupmi projektu SKRIPTOR. Ide o projekt APVV-19-0456 (2020 – 2024) s názvom *Inovatívne sprístupnenie písomného dedičstva Slovenska prostredníctvom systému automatickej transkripcie historických rukopisov [Innovative disclosure of written heritage of Slovakia through the automatic transcription of historical manuscripts].* Riešiteľské organizácie projektu sú: Univerzita Mateja Bela v Banskej Bystrici (zodpovedný riešiteľ doc. Mgr. Imrich Nagy, PhD.); Štátna vedecká knižnica v Banskej Bystrici – partner (garant prof. PhDr. Dušan Katuščák, PhD.).

Všeobecný kontextový rámec projektu SKRIPTOR a výskumu transkripcie historických dokumentov s použitím umelej inteligencie tvorí oblasť vedeckej a praktickej činnosti nazývanej *Digital humanities*. *Digital humanities* (digitálnu humanistiku) možno považovať, podľa nášho názoru, za spoločné pomenovanie a prierezovú metodológiu pre všetky aplikácie informačných a komunikačných technológií (IKT) v spoločenských a humanitných vedách, odboroch a disciplínach a v im zodpovedajúcej praxi. Táto metodológia sa komplexne uplatnila v projekte *READ*, ktorý sa realizoval v rámci programu *Horizon 2020*.

Priame podnety na zameranie slovenského projektu SKRIPTOR nám poskytli práve poznatky a nástroje európskeho projektu výskumu READ *Recognition and Enrichment of Archival Documents*, ktorého riešenie prebiehalo v rokoch 2016 – 2019. Autorom a koordinátorom projektu bol prof. Günter Mühlberger z Univerzity v Innsbrucku. Udržateľnosť projektu READ je zabezpečená v združení READ-COOP (*A European Cooperative Society*). Združenie má v roku 2022 vyše 100 členov z 24 krajín. Jedinou členskou krajinou zo strednej a východnej Európy je zatiaľ Slovensko, hoci v oblasti transkripcie sú, samozrejme, aktívne aj ďalšie krajiny.

Kľúčový inovatívny nástroj na transkripciu historických rukopisných dokumentov je *Transkribus*. Je to komplexná platforma na digitalizáciu, rozpoznávanie textu podporované umelou inteligenciou, ako aj na prepis a vyhľadávanie historických dokumentov – z akéhokoľvek miesta, kedykoľvek a v akomkoľvek jazyku. S *Transkribus Lite* je možné použiť *Transkribus* v prehliadači osobných počítačov a smartfónov. Mnohé z funkcií klienta *Transkribus Expert* môžu byť použité aj v *Transkribus Lite*. Platforma *Transkribus* integruje nástroje vyvinuté výskumnými skupinami v celej Európe vrátane *Skupiny pre rozpoznávanie vzorov a technológie ľudského jazyka* Technickej univerzity vo

Valencii a skupiny *CITlab University* v Rostocku. V súčasnosti s platformou pracujú tisíce historikov, archivárov, knihovníkov... Platforma bola vytvorená v kontexte dvoch predchádzajúcich projektov EÚ *tranScriptorium* (2013 – 2015) a READ (2016 – 2019).

Na Slovensku sme začali pracovať s platformou Transkribus v roku 2017 a informovali sme verejnosť o našich prvých modeloch transkripcie rukopisov. Spočiatku išlo o individuálnu iniciatívu, ktorá vďaka osvietenej ústretovosti ľudí z Univerzity Mateja Bela prerástla do inštitucionálneho výskumu v projekte SKRIPTOR.

V platforme Transkribus používame stroj umelej inteligencie *HTR+* (Handwritten Text Recognition) a *PyLaia*. Tieto stroje zatiaľ nemôžu okamžite automaticky transkribovať rôzne historické rukopisy. Najprv musí byť stroj vyškolený na konkrétny typ písma a rukopisu. Hlavným cieľom praktických experimentov v projekte SKRIPTOR v súčasnosti je tvorba modelov transkripcie.

Spoločnou víziou vedcov, expertov a iných používateľov z oblasti písomného dedičstva je, aby sa verejne dostupné modely transkripcie postupne stali užitočným spoločným nástrojom pre automatickú transkripciu historických dokumentov. Je potrebné dosiahnuť takú úroveň, aby už nebolo potrebné tvoriť pre každú zbierku rukopisov a tlačí samostatné modely. Pre používateľov by malo ísť o akúsi „čiernu skrinku“ (black box), v ktorej umelá inteligencia sama vyberie z integrovaných modelov najvhodnejší model transkripcie historických tlačí, rukopisov, strojopisov a iných dokumentov, ktoré používateľ chce študovať alebo sprístupniť. K tomuto cieľu však vedie dlhá cesta a je nevyhnutné vytvoriť množstvo parciálnych modelov.

Zmyslom projektu SKRIPTOR je, aby súčasťou spoločného medzinárodného úsilia boli aj naši odborníci a aby budúca „čierna skrinka“ bola pripravená poskytnúť pomoc všetkým pri transkripcii slovacikálnych historických zbierok a dokumentov (slovenčina, čeština, latinčina, maďarčina, poľština a i.). V súčasnej fáze vývoja je dôležité zamerať pozornosť na tvorbu modelov transkripcie na základe väčších zbierok, ktoré obsahujú stovky a tisíce strán.

Výskumníci zapojení do projektu SKRIPTOR sa po počiatočnej nedôvere k možnostiam transkripcie postupne stávajú expertmi. Príspevky na našej konferencii sú dokladom toho, že si osvojujú nástroj Transkribus, že zvládli základné pracovné postupy a že čoraz dôkladnejšie spoznávajú funkcionalitu platformy Transkribus. Darí sa im vytvárať veľmi dobré až excelentné modely transkripcie archívnych dokumentov a starých tlačí.

Predstavitelia *Digital humanities* na Slovensku majú k tejto iniciatíve ako k podozrivej novote rozličné postoje. Od nadšených prejavov súhlasu a obdivu po veľmi rezervované až odmietavé postoje (typu „to nie je nič pre nás“, „máme iné starosti“, „umelá inteligencia nenahradí nás expertov“). Často ide o

reakcie, ktoré na jednej strane síce verbálne deklarujú záujem o „digitalizáciu“ a „umelú inteligenciu“, no na druhej strane svedčia o nedostatočných vedomostiach o problematike a možnostiach digitalizácie a využitia umelej inteligencie. Problémom je zrejme aj fakt, že Transkribus nie je hotový nástroj, „policový softvér“ hotový na „klikanie“, ale nástroj, ktorý sa kolektívne stále tvorí a zdokonaľuje. Postoje niektorých svedčia skôr o uprednostnení tradičných paradigiem práce a výskumu ako o reálnej snahe hľadať inovatívne nástroje sprístupnenia a interpretácie nášho obrovského historického písomného dedičstva ako súčasti európskeho kultúrneho dedičstva.

Výskumníci projektu SKRIPTOR sa *prednostne* venujú platforme Transkribus a transkripcii rukopisných zbierok a okrajovo aj transkripcii tlačí. Existuje celý rad iných nástrojov transkripcie: *OCR4all*, ktorý bol vyvinutý na digitalizáciu starých tlačí. Aplikácia *eScript* slúži na transkripciu rukopisov a tlačí. Nástroj *Rescribe* je určený pre stolné počítače na vykonávanie rozpoznávania OCR na obrazových súboroch, súboroch PDF a Knihách Google. Jedným z použiteľných nástrojov transkripcie je aj *Pero.cz*. Systém *ABBYY Cloud OCR SDK* je veľmi kvalitná aplikácia v cloude prostredníctvom webového rozhrania API. Aj ku *ABBYY Cloud OCR SDK* existuje viac ako 10 alternatív. Najlepšou alternatívou je *Online OCR*, ktoré je zadarmo. Ďalšie skvelé stránky a aplikácie podobné *ABBYY Cloud OCR SDK* sú okrem *Transkribusu* aj *Kofax Omnipage, Geekersoft OCR Word Recognition* a *i2OCR*. Pred výskumníkmi v budúcnosti stojí úloha vypracovať kritériá hodnotenia funkcionality a kvality nástrojov, aplikácií a platforiem transkripcie.

Našou úlohou je tvoriť modely, ktoré umožnia transkripciu písomného dedičstva z našej kultúrnej a jazykovej oblasti, pre ktorú sú charakteristické určité druhy písma, jazyky, znaky, štýly, diakritika a pod.

Pre náš projekt SKRIPTOR sú veľmi cenné pracovné kontakty s výskumníkmi z akademických pracovísk, knižníc a archívov. Ide o príspevky v *bloku Digital humanities*. Globálny pohľad na digitalizáciu a modernizáciu elektronických služieb štátnych archívov v európskom kontexte poskytuje príspevok autorov M. Pékovej a D. Ryšku. O masovej digitalizácii v Slovenskej národnej knižnici v Martine informuje J. Kováčik. O skúsenostiach a výsledkoch digitalizácie zbierok v ÚK SAV prináša správu I. Kollárová. O digitalizácii fondu komorskogrófskeho úradu v Banskej Štiavnici informuje J. Michelík. J. Šedivý poskytuje poznatky a skúsenosti o digitalizácii prameňov k dejinám Bratislavy. Veľmi cenná je tiež iniciatíva a projekt P. Labanca zameraný na Digital humanities v historických vedách v Trnave.

V *bloku Transkribus* našu konferenciu významne obohacuje Z. Szatucsek z Maďarského národného archívu, ktorý dokladá, že transkripciou sa môžu a majú zaoberať hlavné národné pamäťové inštitúcie a súčasne prezentuje skúsenosti s platformou Transkribus.

Jednotliví výskumníci projektu SKRIPTOR informujú o stave výskumu a tvorbe modelov pre zvolené archívne zbierky. D. Katuščák popisuje transkripciu niekoľkých historických tlačí a rukopisov. P. Maliniak vysvetľuje postup a skúsenosti s transkripciou rukopisných kázní Izáka Abrahamidesa. Študentka K. Kováčová opisuje možnosti transkripcie nemeckej rukopisnej kuchárskej knihy z roku 1667. P. Kunec a absolvent štúdia histórie M. Katreniak prinášajú poznatky o transkripcii kanonických vizitácií. M. Mikušková a L. Nižníková opisujú prístup k transkripcii historickej tlače. O. Tomeček opisuje prístup a výsledky transkripcie novolatinského rukopisu reambulačného protokolu. M. Bôbová uvažuje o možnostiach využitia digitalizácie vo výskume dejín knižnej kultúry. A. Kurhajcová vysvetľuje postupy a skúsenosti s transkripciou rukopisu J. M. Hurbana. I. Nagy sa venuje postupu a výsledkom transkripcie Csákósovho katalógu korešpondencie Koháryovcov.

Naše dosiahnuté výsledky, know-how a skúsenosti ašpirujú zaviesť revolučnú a inovatívnu platformu Transkribus na Slovensku a podnietiť výskum aj v Čechách. Usilujeme sa rozvinúť medzinárodné kontakty a zaviesť poznatky jednak do systému vzdelávania a jednak do praxe pamäťových a fondových inštitúcií prostredníctvom projektov výskumu a vývoja.

Prof. PhDr. Dušan Katuščák, PhD.

Preface

The collection of abstracts is an accompanying output of the scientific conference *Digital humanities – tools for making historical heritage accessible*, held in Banská Bystrica on October 12-13, 2022. The conference and the collection are planned outputs of the SCRIPTOR project. This is the project APVV-19-0456 (2020 – 2024) with the title *Innovative access to the written heritage of Slovakia through the system of automatic transcription of historical manuscripts*. The research organizations of the project are: Matej Bel University in Banská Bystrica (responsible researcher Doc. Mgr. Imrich Nagy, PhD.); State Scientific Library in Banská Bystrica – partner (guarantor Prof. PhDr. Dušan Katuščák, PhD.). The general contextual framework of the SKRIPTOR project and research into the transcription of historical documents with the use of artificial intelligence is the field of scientific and practical activity called *digital humanities*. Digital humanities can be considered, in our opinion, as a common name and a cross-cutting methodology for all applications of information and communication technologies (ICT) in social and human sciences, fields and disciplines and in their corresponding practice. This methodology was comprehensively applied in the *READ* project, which was implemented as part of the *Horizon 2020* program.

The knowledge and tools of the European research project READ *Recognition and Enrichment of Archival Documents*, which took place between 2016 and 2019, gave us the direct stimulus for the focus of the Slovak SKRIPTOR project. The author and coordinator of the project was prof. Günter Mühlberger from the University of Innsbruck. The sustainability of the READ project is ensured in the READ-COOP (*A European Cooperative Society*) association. In 2022, the association has more than 100 members from 24 countries. Slovakia is currently the only member country from Central and Eastern Europe, although, of course, other countries are also active in the field of transcription.

A key innovative tool for the transcription of historical manuscript documents is *Transkribus*. It is a comprehensive platform for digitization, text recognition supported by artificial intelligence, as well as for the transcription and search of historical documents – from any place, at any time and in any language. With *Transkribus Lite*, it is possible to use *Transkribus* in the browser of personal computers and smartphones. Many of the functions of the *Transkribus Expert* client can also be used in *Transkribus Lite*. The *Transkribus* platform integrates tools developed by research groups across Europe, including the *Pattern Recognition and Human Language Technology*

Group of the Technical University of Valencia and the *CITlab* University Group in Rostock. Currently, thousands of historians, archivists, librarians work with the platform... The platform was created in the context of two previous EU projects *tranScriptorium* (2013 – 2015) and READ (2016 – 2019).

In Slovakia, we started working with the Transkribus platform in 2017 and informed the public about our first manuscript transcription models. Initially, it was an individual initiative, which, thanks to the enlightened hospitality of people from the Matej Bel University, grew into an institutional research in the SKRIPTOR project.

In the Transkribus platform, we use the artificial intelligence engine *HTR+* (Handwritten Text Recognition) and *PyLaia*. These machines cannot yet automatically transcribe various historical manuscripts instantly. First, the machine must be trained for a specific type of font and handwriting. The main goal of practical experiments in the SCRIPTOR project is currently the creation of transcription models.

The shared vision of scholars, experts and other users in the field of written heritage is that publicly available transcription models will gradually become a useful common tool for automatic transcription of historical documents. It is necessary to reach such a level that it is no longer necessary to create separate models for each collection of manuscripts and prints. For users, it should be a kind of “black box” in which artificial intelligence itself selects from the integrated models the most suitable transcription model of historical prints, manuscripts, typescripts and other documents that the user wants to study or make available. However, this goal has a long way to go and it is necessary to create a number of partial models.

The purpose of the SKRIPTOR project is for our experts to be part of the joint international effort and for the future “black box” to be ready to provide assistance to everyone in the transcription of Slovak historical collections and documents (Slovak, Czech, Latin, Hungarian, Polish, etc.). At the current stage of development, it is important to focus attention on the creation of transcription models based on larger collections that contain hundreds and thousands of pages.

Researchers involved in the SKRIPTOR project, after initially distrusting the possibilities of transcription, gradually become experts. The contributions at our conference are proof that they are adopting the Transkribus tool, that they have mastered the basic workflows, and that they are getting to know the functionality of the Transkribus platform more and more thoroughly. They manage to create very good to excellent transcription models of archival documents and old prints.

Representatives of *digital humanities* in Slovakia have different attitudes towards this initiative as a suspicious novelty. From enthusiastic expressions of approval and admiration to very reserved or dismissive attitudes (such as "it's nothing for us", "we have other concerns", "artificial intelligence will not replace us experts"). These are often reactions that, on the one hand, verbally declare an interest in "digitalization" and "artificial intelligence", but on the other hand, testify to insufficient knowledge about the issues and possibilities of digitization and the use of artificial intelligence. The problem is probably also the fact that Transkribus is not a finished tool "off the shelf software" ready to be "clicked", but a tool that is still being created and improved collectively. The attitudes of some indicate a preference for traditional work and research paradigms rather than a real effort to find innovative tools for making accessible and interpreting our huge historical written heritage as part of the European cultural heritage.

The researchers of the SKRIPTOR project primarily focus on the Transkribus platform and the transcription of manuscript collections, and marginally also print transcription. There are a number of other transcription tools: *OCR4all*, which was developed to digitize old prints. The *eScript* application is used to transcribe manuscripts and prints. *Rescribe* is a desktop tool for performing OCR on image files, PDF files, and Google Books. One of the useful transcription tools is *Pero.cz*. The *ABBYY Cloud OCR SDK* system is a high-quality application in the cloud through a web API. There are also more than 10 alternatives to *ABBYY Cloud OCR SDK*. The best alternative is *Online OCR*, which is free. Other great websites and applications similar to the *ABBYY Cloud OCR SDK* are, in addition to *Transkribus, Kofax Omnipage*, *Geekersoft OCR Word Recognition*, and *i2OCR*. In the future, researchers face the task of developing criteria for evaluating the functionality and quality of transcription tools, applications and platforms.

Our task is to create models that will enable the transcription of written heritage from our cultural and linguistic area, which is characterized by certain types of writing, languages, signs, styles, diacritics, etc.

Working contacts with researchers from academic workplaces, libraries and archives are very valuable for our SKRIPTOR project. These are posts in *the Digital humanities block*. A global view of the digitization and modernization of electronic services of state archives in the European context is provided by the contribution of the authors M. Péková and D. Ryška. J. Kováčik informs about mass digitization in the national library in Martin. I. Kollárová reports on the results, experiences and results of the digitization of collections at the Central Library of the Slovak Academy of Science. J. Michelík informs about the digitization of the fund of the Chamber of Commerce and Industry in

Banská Štiavnica. J. Šedivý provides knowledge and experience about the digitization of sources for the history of Bratislava. The initiative and project of P. Labanc focused on "digital humanities" in historical sciences in Trnava is also very valuable.

In the *Transkribus block*, our conference is significantly enriched by Z. Szatucsek from the Hungarian National Archive, who proves that the main national memory institutions can and should deal with transcription and at the same time presents experiences with the Transkribus platform. Individual researchers of the SCRIPTOR project inform about the state of research and the creation of models for selected archival collections. P. Maliniak explains the process and experiences with the transcription of the manuscript sermons of Isaac Abrahamides. Student K. Kováčová describes the possibilities of transcription of a German manuscript cookbook from 1667. P. Kunec and history graduate M. Katreniak bring knowledge about the transcription of canonical visitations. M. Mikušková and L. Nižníková describe the approach to the transcription of historical press. O. Tomeček describes the approach and results of the transcription of the Neo-Latin manuscript of the reambulation protocol. M. Bôbová considers the possibilities of using digitization in the research of the history of book culture. A. Kurhajcová explains the procedures and experiences with the transcription of J. M. Hurban's manuscript. I. Nagy deals with the process and results of the transcription of Csákós's catalog of the Koháry correspondence.

The achieved results, know-how and experience led us to the effort to introduce the revolutionary and innovative Transkribus platform in Slovakia and stimulate research in the Czech Republic as well. We strive to develop international contacts and introduce knowledge both in the education system and in the practice of memory and fund institutions through research and development projects.

Prof. PhDr. Dušan Katuščák, PhD.

Digitalizácia a modernizácia elektronických služieb štátnych archívov v európskom kontexte

Monika Péková – Dan Ryška

Ministerstvo vnútra Slovenskej republiky, odbor archívov a registratúr, Križkova 7, 811 04 Bratislava, monika.pekova@minv.sk
ryska.mail@gmail.com

Najväčšia a najdôležitejšia časť archívneho dedičstva Slovenskej republiky je uložená v štátnych archívoch, pričom najstaršie archívne dokumenty pochádzajú už zo začiatku 12. storočia. Štátne archívy svojím postavením a úlohami majú nenahraditeľný význam v živote každej demokratickej spoločnosti, predstavujú pamäť národa a sú základom identity spoločnosti, východiskom zachovania informačnej kontinuity vo verejnom sektore, tvoria integrálnu súčasť informačných štruktúr a zdrojov spoločnosti.

Ministerstvo vnútra SR vypracovalo Národnú koncepciu rozvoja štátnych archívov s výhľadom do roku 2028, ktorú vláda SR schválila uznesením vlády SR č. 456 z 13. júla 2022. Jedným z troch primárnych cieľov vedúcich k ochrane archívneho dedičstva je digitalizácia štátnych archívov.
Primárnou požiadavkou na dosiahnutie stanoveného zámeru je nevyhnutná modernizácia štátnych archívov, digitalizácia obsahu a procesov s cieľom zvyšovať kvalitu a rozsah elektronických služieb, ako aj ochranu archívnych dokumentov, zmazanie aktuálneho archívneho dlhu popri priebežnej digitalizácii a nasadenie aktuálnych moderných informačných technológií. Na naplnenie týchto cieľov je potrebné zrealizovať tieto úlohy:

1. Vytvoriť podmienky na komplexnú digitalizáciu novoprichádzajúceho archívneho obsahu a pilotnú digitalizáciu historického obsahu.
2. Dobudovať funkcionality a kapacity Elektronického archívu Slovenska a zabezpečiť jeho podporu.
3. Vybudovať špecializované digitalizačné pracoviská vybavené knižnými, prieťahovými a veľkoformátovými skenermi a skenermi na mikrofilmy vrátane digitalizačného workflow softvéru na spracovanie digitálnych kópií.
4. Zaviesť ArchivLog na evidenciu životného cyklu archívnych dokumentov.
5. Sprístupniť archívne dokumenty z pilotnej prevádzky digitalizačných pracovísk v Elektronickom archíve Slovenska.

6. Implementovať systém automatizácie popisovania metadát s využitím technológie strojového učenia.
7. Implementovať systém monitorovania procesu digitalizácie a sledovania produktivity práce digitalizačných pracovísk.
8. Zabezpečiť migráciu všetkých zdigitalizovaných archívnych dokumentov z lokálnych úložísk na centrálne dátové úložisko, zabezpečiť pravidelné zálohovanie archívu s dôrazom na kritické dáta.
9. Zabezpečiť dostatočnú kapacitu na uloženie digitálnych kópií archívnych dokumentov vyhotovených systematickou digitalizáciou.
10. Dovybaviť bádateľne o príručné skenery a počítače s cieľom znížiť zaťaženie pracovníkov bádateľne.
11. Oboznámiť verejnosť s prínosmi a fungovaním elektronického archívu so zameraním na perspektívnych užívateľov.

Ministerstvo vnútra SR plánuje zriadiť desať digitalizačných pracovísk a vybaviť ich nevyhnutnou technikou, infraštruktúrou a personálom s cieľom začať výkonnú digitalizáciu archívneho obsahu. Taktiež plánuje dobudovať nové funkcionality informačného systému Elektronický archív Slovenska s cieľom sprístupniť archívny obsah občanom, zvýšiť efektivitu a kvalitu práce archívov a znížiť náklady spojené so spracovávaním archívnych dokumentov. Ide o vybudovanie digitalizačno-informačného systému opierajúceho sa technologicky už o existujúci systém vrátane využitia dátových centier Ministerstva vnútra SR s cieľom dobudovať potrebnú serverovú a úložnú kapacitu. Výsledkom navrhnutých postupov a opatrení bude skvalitnenie služieb občanovi, šetrenie finančných prostriedkov a ochrana životného prostredia. Zverejnením digitálnych kópií sa podstatne skráti čas na vyhľadanie a získanie relevantných a jedinečných informácií na účely uplatnenia práv občanov, pre potreby riešenia ich životných situácií, dôkazového materiálu v súdnych konaniach, historického výskumu a zvyšovania národného povedomia.

Digitization and modernization of electronic services of state archives in the European context

Monika Péková – Dan Ryška

Ministry of the Interior of the Slovak Republic – department of archives and registries, Križkova 7, 811 04 Bratislava, monika.pekova@minv.sk ryska.mail@gmail.com

The largest and most important part of the archival heritage of the Slovak Republic is stored in the state archives, while the oldest archival documents date back to the beginning of the 12th century. Due to their position and tasks, state archives are of irreplaceable importance in the life of every democratic society, they represent the nation's memory and are the basis of the society's identity, the starting point for preserving information continuity in the public sector, they form an integral part of the information structures and resources of society. The Ministry of the Interior of the Slovak Republic developed the National Concept for the Development of State Archives with a view to 2028, which was approved by the Government of the Slovak Republic by Resolution of the Government of the Slovak Republic no. 456 of July 13, 2022. One of the three primary goals leading to the protection of archival heritage is the digitization of state archives.

The primary requirement for achieving the set goal is the necessary modernization of state archives, digitalization of content and processes for the purpose of increasing the quality and scope of electronic services, as well as increasing the protection of archival documents, erasing the current archival debt along with continuous digitization and deploying current modern information technologies. To fulfill these goals, it is necessary to complete the following tasks:

1. develop conditions for the comprehensive digitization of newly arriving archival content and the pilot digitization of historical content,
2. to complete the functionalities and capacities of the Electronic Archive of Slovakia and ensure its support,
3. build specialized digitization workplaces equipped with book, draft and large-format scanners and microfilm scanners, including digitization workflow software for processing digital copies,
4. introduce ArchivLog for recording the life cycle of archival documents,

5. to make available archival documents from the pilot operation of digitization workplaces in the Electronic Archive of Slovakia,
6. implement a metadata description automation system using machine learning technology,
7. implement a system for monitoring the digitization process and monitoring the work productivity of digitization workplaces,
8. ensure the migration of all digitized archive documents from local storage to the central data storage, ensure regular backup of the archive with an emphasis on critical data,
9. ensure sufficient capacity for storing digital copies of archival documents made by systematic digitization,
10. to equip the research rooms with handheld scanners and computers in order to reduce the workload of the research room workers,
11. to acquaint the public with the benefits and functioning of the electronic archive, focusing on prospective users.

The Ministry of the Interior of the Slovak Republic plans to establish 10 digitization workplaces and equip them with the necessary technology, infrastructure and personnel in order to start the efficient digitization of archival content. It also plans to complete the development of new functionalities of the Electronic Archive of Slovakia information system with the aim of making archive content accessible to citizens, increasing the efficiency and quality of archive work, and reducing the costs associated with processing archive documents. It is about the construction of a digitization and information system technologically based on an already existing system, including the use of data centers of the Ministry of the Interior of the Slovak Republic for the purpose of completing the necessary server and storage capacity. The result of the proposed procedures and measures will be the improvement of services to the citizen, the saving of financial resources and the protection of the environment. The publication of digital copies will significantly reduce the time to search and obtain relevant and unique information for the purposes of exercising the rights of citizens, for the needs of solving their life situations, evidential material in court proceedings, historical research and raising national awareness.

Masová digitalizácia v Slovenskej národnej knižnici a jej výsledky

Ján Kováčik

Slovenská národná knižnica, Nám. J. C. Hronského 1, 036 01 Martin, jan.kovacik@snk.sk

Masová digitalizácia fondov Slovenskej národnej knižnice (SNK) sa začala v roku 2012 a bola výsledkom implementácie národného projektu *Digitálna knižnica a digitálny archív* realizovaného v rámci Prioritnej osi 2 Operačného programu Informatizácia spoločnosti spolufinancovaného zo Štrukturálnych fondov EÚ. V ostatných desiatich rokoch sa v novovybudovanom Konzervačnom a digitalizačnom centre SNK vo Vrútkach zdigitalizovalo (a v prípade potreby očistilo, sterilizovalo a zreštaurovalo) viac než 70 miliónov strán dokumentov, čo predstavuje viac než 2,8 milióna objektov (monografií, čísel seriálov, článkov, špeciálnych a archívnych dokumentov), ktoré boli zároveň sprístupnené odbornej a laickej verejnosti – všetky digitalizované objekty bez ohľadu na ich autorskoprávny status sú voľne dostupné v Digitálnej knižnici SNK na vyhradených koncových zariadeniach v priestoroch SNK, pričom desiatky tisíc voľných diel, na ktoré sa už nevzťahuje autorskoprávna ochrana (vrátane tisícov archívnych dokumentov z Literárneho archívu SNK a starých a vzácnych tlačí zo zbierok SNK), je voľne dostupných online z ktoréhokoľvek miesta na svete. V roku 2023 bude dosiahnutý ďalší významný míľnik, ktorým je online a bezodplatné sprístupnenie viac než jedného milióna tzv. obchodne nedostupných diel pre používateľov, ktorí k nim budú chcieť pristúpiť z krajín Európskeho hospodárskeho priestoru, v dôsledku čoho sa rozhodujúca časť slovacikálnej vydavateľskej produkcie 20. storočia a prvej dekády 21. storočia stane v digitalizovanej forme široko dostupná verejnosti, a to nielen pre výskum a vzdelávanie. Uvedený postup umožnilo uzavretie rozšírenej hromadnej licenčnej zmluvy s autorskou spoločnosťou LITA v roku 2022, a to na základe transpozície smernice EÚ o autorskom práve a právach súvisiacich s autorským právom na digitálnom jednotnom trhu z roku 2019 do slovenského právneho poriadku.

Príspevok poskytuje stručný opis procesov digitalizácie a jej výsledkov so zameraním na sprístupňovanie digitalizovaného obsahu verejnosti prostredníctvom Digitálnej knižnice SNK a jeho dosahy na budúcnosť poskytovania knižnično-informačných služieb nielen zo strany SNK.

Mass digitization in the Slovak National Library and its results

Ján Kováčik

Slovak National Library, Nám. J. C. Hronského 1, 036 01 Martin, jan.kovacik@snk.sk

The mass digitization of the holdings of the Slovak National Library (SNL) began in 2012 and was the result of the implementation of the national project Digital Library and Digital Archive, implemented as part of Priority Axis 2 of the Operational Program Informatization of Society, co-financed by EU Structural Funds. In the past ten years, more than 70 million pages of documents have been digitized (and, if necessary, cleaned, sterilized and restored) in the newly built Conservation and Digitization Center of SNL in Vrútky, which represents more than 2.8 million objects (monographs, serial numbers, articles, special and archival documents), which were also made available to the professional and lay public – all digitized objects, regardless of their copyright status, are freely available in the SNL Digital Library on dedicated terminal devices in the SNL premises, while tens of thousands of free works that are no longer subject to copyright protection (including thousands of archival documents from the SNL Literary Archive and old and rare prints from the SNL collections) is freely available online from anywhere in the world. In 2023, another important milestone will be reached, which is the availability of more than one million so-called commercially unavailable works online and free of charge for users who want to access them from the countries of the European Economic Area, as a result of which a decisive part of the Slovak publishing production of the 20th century and the first decade of the 21st century will become widely available to the public in digitized form, and not only for the needs of research and education. The mentioned procedure made it possible to conclude an extended collective license agreement with the LITA copyright company in 2022, based on the transposition of the EU directive on copyright and copyright-related rights in the digital single market from 2019 into the Slovak legal order.

The paper provides a brief description of digitization processes and its results with a focus on making digitized content accessible to the public through the SNL Digital Library and its impact on the future of the provision of library and information services not only by the SNL.

Digitalizácia historických zbierok v Ústrednej knižnici SAV, v. v. i.

Ivona Kollárová

Ústredná knižnica SAV, v. v. i., Klemensova 19, 814 99 Bratislava, ivona.kollarova@savba.sk

Prvé úvahy o digitalizácii historických zbierok sa vynárali po roku 2000 a ich výsledkom bolo sformovanie jednoduchej stratégie výberu dokumentov založenej na poznaní zbierok a jej unikátnych častí. Historické fondy, obsahujúce predovšetkým tlačené knihy, prinášajú ako objekt digitalizácie veľkú mieru nežiaducej duplicity a multiplicity digitalizovaných objektov. Základným kritériom je preto jedinečnosť, a teda v istom zmysle neprístupnosť pamiatky. Ak je spojená s informačnou atraktívnosťou, tak je sprístupnenie v digitálnej podobe zároveň prevenciou rizík vyplývajúcich z klasického sprístupňovania. Digitalizačnú stratégiu postupne formovali tieto princípy a zároveň podnety bádateľov, ich potreba mať k dispozícii konkrétny dokument ako prameň výskumu. V roku 2005 začalo v Univerzitnej knižnici v Bratislave pôsobiť digitalizačné pracovisko a dohoda o spolupráci umožnila dostať do virtuálneho prostredia prvé historické dokumenty. V tom čase sme tak mali k dispozícii nielen najvyšší dostupný štandard digitalizácie, ale aj spracovania digitálnych objektov a tvorby metadát vo formáte XML v intenciách projektu Manuscriptorium – najväčšej digitálnej knižnice rukopisov.

Míľnikom formovania hybridnej knižnice bolo zriadenie vlastného digitalizačného minipracoviska v roku 2008 v rámci projektu *NISPEZ – Národný informačný systém podpory výskumu a vývoja na Slovensku – prístup k elektronickým informačným zdrojom*. Technologické osamostatnenie prinieslo väčšiu dynamiku a dodnes bolo vytvorených viac ako 300 digitálnych objektov. V tomto období sme pristúpili aj k digitalizácii zbierky mikrofilmov – bohatého fondu reprodukcií slovacikálnych tlačí a rukopisov zachovaných v mnohých inštitúciách na území Slovenska, ale aj mimo neho. Uplynulých 12 rokov nebolo len obdobím monotónneho skenovania dokumentov, ale umožnilo zapojiť sa do nadinštitucionálnych kooperácií. Zbierka pamätníkov alebo tzv. štambuchov zo 16. – 19. storočia sa stala základom spolupráce s platformou *Inscriptiones alborum amicorum*, online prostredím na analýzu, sprístupňovanie a vyhľadávanie pamätníkov v európskych pamäťových inštitúciách a zápisov v nich. Zbierka obrazových dokumentov bola sprístupnená v spolupráci s platformou *Pammap – Pamäť mesta Bratislava*.

V roku 2020 sa začala spolupráca s Centrom vedecko-technických informácií SR. Vznikla tzv. virtuálna študovňa – webové prostredie pre digitálne objekty rukopisov a historických tlačí. V roku 2019 sa Ústredná knižnica SAV stala pilotným partnerom spoločnosti Piql Slovakia, vyvíjajúcou inovatívnu bezmigračnú technológiu digitálneho filmu na dlhodobé uchovávanie digitálneho obsahu. V roku 2021 poskytla Európska únia podporu projektu *eArchiving of Engineering and Science Library*, ktorý vznikol na základe spolupráce nórskej spoločnosti Piql, španielskej spoločnosti Airbus a Ústrednej knižnice SAV. Je to projekt vývoja systémov digitálneho uchovávania dát v súlade so špecifikáciami eArchivingu. Jeho dôležitou súčasťou je vývoj nových modulov systému piqlConnect poskytujúceho užívateľsky prívetivý a jednoduchý spôsob uchovávania informácií. V intenciách tejto spolupráce sa do virtuálneho prostredia dostane významná časť rukopisnej zbierky Evanjelického lýcea v Bratislave, tzv. školský archív obsahujúci „matriky" školy, záznamy dokumentujúce pôsobenie školy v rokoch 1714 – 1900. Výsledok by mal spĺňať predovšetkým kritérium pohodlného vyhľadávania biografických dát študentov lýcea a bude zároveň novou skúsenosťou sprístupnenia obsiahleho a z hľadiska tvorby metadát špecifického historického prameňa.

Príspevok je retrospektívou zachytávajúcou zámery, míľniky a úskalia mikrodejín sprístupňovania kultúrneho dedičstva v jednej inštitúcii. Takýto pohľad na viac ako dve desaťročia formovania a uskutočňovania digitalizačných zámerov ukazuje nesúlad medzi víziami inštitúcie, očakávaniami jej klientov na jednej strane a jej možnosťami a obmedzeniami na druhej strane. Pohľad ponad úskalia a problémy však umožňuje vnímať, ako Ústredná knižnica SAV počas dvoch desaťročí postupne prostredníctvom niekoľkých projektov prerámcovala digitálne sprístupňovanie historických dokumentov od digitalizácie vybraných dokumentov ku sprístupneniu kompaktných zbierok.

Digitization of historical collections in the Central Library of Slovak Academy of Sciences

Ivona Kollárová

Central Library of Slovak Academy of Sciences, Klemensova 19, 814 99 Bratislava, ivona.kollarova@savba.sk

First thoughts about the digitization of historical collections emerged after the year 2000, and their result was the formation of a simple strategy for selecting documents, based on knowledge of collections and their unique components. Historical funds, primarily containing printed books, bring a large degree of unwanted duplication and multiplicity of digitized objects as an object of digitization. Therefore, the basic criterion is the uniqueness and thus in some sense the inaccessibility of the monument. If it is associated with informational attractiveness, then making it available in digital form is also a prevention of risks arising from the standard way the monument was being made available. The digitization strategy was gradually shaped by these principles and at the same time by the researchers' suggestions and their need to have a specific document available as a source of research. In 2005, a digitization workplace started to operate at the University Library in Bratislava, and a cooperation agreement made it possible to get the first historical documents into a virtual environment. At that time, we had at our disposal not only the highest available standard of digitization, but also of the processing of digital objects and the creation of metadata in XML format in the intentions of the Manuscriptorium project – the largest digital library of manuscripts.

A milestone in the formation of the hybrid library was the establishment of its own digitization mini-workplace in 2008 as part of the *NISPEZ project – National Information System for Research and Development Support in Slovakia – access to electronic information resources*. Technological independence brought greater dynamism and more than 300 digital objects have been created to date. During this period, we also started digitizing the microfilm collection – a rich fund of reproductions of Slovak prints and manuscripts preserved in many institutions in Slovakia, but also outside of it. The past 12 years were not only a period of monotonous scanning of documents, but it enabled involvement in supra-institutional cooperation. Collection of monuments or so-called "Štambuchy" from the 16th – 19th centuries has become the basis of cooperation with the *Inscriptiones alborum amicorum* platform, an online environment for the analysis, access and

research of monuments and inscriptions in European memory institutions. The collection of visual documents was made available in cooperation with the platform *Pammap – Memory of the City of Bratislava*.

In 2020, began the cooperation with the Center for Scientific and Technical Information of the Slovak Republic. The so-called virtual study room – web environment for digital objects of manuscripts and historical prints – was created. In 2019, the Central Library of Slovak Academy of Sciences became a pilot partner of the company Piql Slovakia, which is developing innovative migration-free digital film technology for long-term preservation of digital content. In 2021, the European Union provided support for the *eArchiving of Engineering and Science Library* project, which was created based on the cooperation between the Norwegian company Piql, the Spanish company Airbus and the Central Library of Slovak Academy of Sciences. It is a project which develops digital data storage systems in accordance with eArchiving specifications. Its important part is the development of new modules of the piqlConnect system, providing a user-friendly and simple way of storing information. In the intentions of this collaboration, a significant part of the manuscript collection of the Evangelical Lyceum in Bratislava will be stored in this virtual environment, the so-called the school archive; containing school "registers" which record documentation of the school's activities in the years 1714 – 1900. The result should primarily meet the criterion of convenient search of biographical data of lyceum students and will also be a new experience of declassifying a comprehensive historical source specific from the point of view of creating metadata.

The contribution is a retrospective capturing the intentions, milestones and pitfalls of the microhistory of making cultural heritage accessible in one institution. Such a view of more than two decades of forming and implementing digitization intentions shows a discrepancy between the visions of the institution, the expectations of its clients on the one hand, and its possibilities and limitations on the other. Looking beyond the pitfalls and problems, however, allows us to see how the Central Library of Slovak Academy of Sciences gradually over two decades, through several projects, reframed the digital access of historical documents from the digitization of selected documents to the accessibility of compact collections.

Fond Hlavného komorskogrófskeho úradu v Banskej Štiavnici a jeho digitalizácia

Juraj Michelík

Slovenský národný archív, špecializované pracovisko Slovenský banský archív, Radničné námestie 16, 969 01 Banská Štiavnica, juraj.michelik@minv.sk

V roku 1571 rakúsky cisár vydal tzv. Maximiliánov banský poriadok, ktorým bol ustanovený hlavný komorský gróf. Jeho úlohou bolo centralizovať banskú správu v Hornom Uhorsku a hájiť záujmy panovníka. Úrad bol zriadený v Banskej Štiavnici ako najvýznamnejšom banskom meste v stredoslovenskej oblasti. Od 17. storočia až do 19. storočia Hlavný komorskogrófsky úrad (ďalej len HKG) sústreďoval dôležité banské dokumenty z celého Horného Uhorska. Dnes je archívny fond HKG najrozsiahlejším a najvýznamnejším fondom Slovenského banského archívu v Banskej Štiavnici. Obsahuje vzácne banské mapy, najstaršiu z roku 1642, plány banských strojov a zariadení, berggerichty a ordináriá. Fondové oddelenie Banské mapy a plány HKG bolo v roku 2007 zapísané do medzinárodného registra UNESCO Pamäť sveta. Zápisom do registra sa Slovenská republika zaviazala zvýšiť ochranu archívnych dokumentov ich digitalizáciou. Touto úlohou bolo vládou SR poverené Ministerstvo vnútra SR, ktoré zriadilo špecializované digitalizačné pracovisko s veľkoformátovým skenerom CRUSE vo vtedajšom Štátnom ústrednom banskom archíve v Banskej Štiavnici, kde sú banské mapy a plány HKG uložené. Za desať rokov systematickej digitalizácie sa podarilo celý rozsiahly archívny fond zdigitalizovať a sprístupniť bádateľom. Časť digitalizátov je sprístupnená širokej verejnosti v Elektronickom archíve Slovenska.

Fund of the Main Chamber Count Office in Banská Štiavnica and its digitization

Juraj Michelík

Slovak Mining Archives (Slovak National Archives), Radničné námestie 16, 969 01 Banská Štiavnica, juraj.michelik@minv.sk

In 1571, the Austrian emperor issued the so-called Maximilian's mining order, by which the head chamber count was appointed. His task was to centralize the mining administration in Upper Hungary and defend the interests of the monarch. The office was established in Banská Štiavnica as the most important mining town in the Central Slovak region. From the 17th century until the 19th century, the Main Chamber Count Office (MCCO) gathered important mining documents from all over Upper Hungary. Today, the MCCO archive fund is the largest and most important fund of the Slovak Mining Archive in Banská Štiavnica. It contains rare mining maps, the oldest from 1642, plans of mining machines and equipment, berggerichts and ordinaries. In 2007, the fund department of MCCO called the Mining Maps and Plans was entered into the UNESCO World Heritage International Register. By entering the register, the Slovak Republic has committed itself to increasing the protection of archival documents by digitizing them. The Ministry of Interior of the Slovak Republic was entrusted with this task by the Government of the Slovak Republic, which established a specialized digitization workplace with a large-format CRUSE scanner in the then State Central Mining Archive in Banská Štiavnica, where mining maps and MCCO plans are stored. During the ten years of systematic digitization, the entire extensive archive fund was digitized and made available to researchers. Part of the digitized files is made available to the general public in the Electronic Archive of Slovakia.

Digitalizácia prameňov k dejinám Bratislavy

Juraj Šedivý

Katedra archívnictva a muzeológie, Filozofická fakulta, Univerzita Komenského v Bratislave, Gondova 2, 811 02 Bratislava 1, juraj.sedivy@uniba.sk

Už približne jednu generáciu je historiografia na Slovensku konfrontovaná s digitálnym svetom. Kým samotnú interpretáciu prameňov (napr. využitím umelej inteligencie alebo Digital humanities vo všeobecnosti) takmer neovplyvnil, tak prvotnú heuristickú fázu spracúvania, sprístupňovania a filtrovania historických dát ovplyvnil už oveľa výraznejšie. Digitalizáciu prameňov na Slovensku možno rozdeliť do štyroch základných fáz (z ktorých niektoré prebiehali či ešte aj existujú paralelne): 1. digitalizovanie cimélií na e-nosičoch (CD, neskôr DVD-ROM), 2. individuálna digitalizácia ad hoc podľa záujmu bádateľov, 3. snaha o koordinované digitalizovanie nadkritických más prameňov (napr. Slovakiana), 4. dlhodobé iniciatívy menších zoskupení smerujúce k online sprístupňovaniu doteraz menej prístupných prameňov (napr. PamMap.sk). Z hľadiska záberu prameňov možno rozlišovať:

a) **inštitucionálnu** digitalizáciu prameňov (primárny cieľ je obvykle vytváranie digitálnej kópie na účely ochrany prameňov – napr. v MMB digitalizácia zbierky fotografií; obvykle sú digitalizáty prístupné len vnútri inštitúcie a dopĺňanie metadát je individuálne);

b) **rezortnú** digitalizáciu smerujúcu k prehľadu o zbierkach naprieč inštitúciami v rámci jedného sektoru pamäťových inštitúcií (napr. SNG iniciovala a viaceré galérie (ako GMB) sa pripojili k portálu Webumenia; pre múzeá vznikol intrarezortný portál CEMUZ a pre archívy sa pripravuje Elektronický archív Slovenska, na ktorom sú však zatiaľ prístupné len zoznamy fondov a zbierok, ako aj niektoré archívne pomôcky, postupne by mali pribúdať aj digitalizáty archívnych dokumentov ako takých; medzinárodným typom vnútrorezortného portálu sú dva medzinárodné projekty Monasterium.net, napr. databáza stredovekých listín – z bratislavských archívov SNA a AMB a rakúsko-slovenský projekt CRARC (digitalizáty úradných kníh z AMB a ŠABA – hoc aj len s minimom metadát);

c) **tematické** portály majú obvykle len parciálny záber (konkrétne k Bratislave napr. DiFMOE zameriavajúci sa na nemecké (spočiatku najmä tlačené) a v nemčine písané pramene – len k dejinám Bratislavy ponúka vyše 5000 jednotiek, z nich najcennejšie sú noviny s OCR sprístupneným obsahom);

d) **komplexné** pamäťové portály, ako je Slovakiana (necelých 70 000 digitalizátov zo SNA, SNM, MMB a ďalších) alebo PamMap.sk (nielen cca.

30 000 digitalizátov historických prameňov, ale aj encyklopédia a mapová aplikácia prepojená s portálom towns.sk, kde je historický atlas mesta Bratislavy – na rozdiel od Slovakiany aj s množstvom privátnych prameňov). Cieľom príspevku je porovnať plusy a mínusy portálov a pozrieť sa na ne kritickým okom užívateľa.

Digitization of sources for the history of Bratislava

Juraj Šedivý

Department of archive studies and museology, Faculty of Arts, Comenius University in Bratislava, Gondova 2, 811 02 Bratislava 1,
juraj.sedivy@uniba.sk

For approximately one generation now, historiography in Slovakia has been confronted with the digital world. While it hardly influenced the interpretation of sources (e.g. by using artificial intelligence or digital humanities in general), it influenced the initial heuristic phase of processing, making available and filtering historical data much more significantly. The digitization of sources in Slovakia can be divided into 4 basic phases (some of which took place or still exist in parallel): 1) digitization of heirlooms on e-carriers (CD, later DVD-rom), 2) individual ad-hoc digitization according to the interest of researchers, 3) the effort for coordinated digitization of supercritical masses of sources (e.g. Slovakiana), 4) long-term initiatives of smaller groups aimed at online access to previously less accessible sources (e.g. PamMap.sk). From the point of view of capturing sources, we can talk about

a) **institutional** digitization of sources (the primary goal is usually the creation of a digital copy for the purpose of protecting sources – e.g. in the Bratislava City Museum it was the digitization of a collection of photographs; usually the digitized images are accessible only inside the institution and the addition of metadata is individual);

b) **departmental** digitization leads to an overview of collections across institutions within one sector of memory institutions (e.g. Slovak National Gallery initiated and several galleries (such as Bratislava City Gallery) joined the portal Webumenie; the CEMUZ intra-departmental portal was created for museums, and the Electronic Archive of Slovakia is being prepared for archives, on which, however, only lists of funds and collections as well as some archival aids can be accessed for now, digitized archival documents as such should gradually be added; the international version of intra-ministerial portal are the two international projects Monasterium.net (database of medieval documents – from the Bratislava archives of Slovak National Archives and Bratislava City Archive) and the Austrian-Slovak project CRARC (digitalizations of official books from Bratislava City Archive and State Archiv in Bratislava – albeit with only a minimum of metadata)

c) **thematic** portals usually only have a partial scope (specifically, for Bratislava, e.g. DiFMOE focusing on German (initially mainly printed) sources

and sources written in German – it offers more than 5,000 units on the history of Bratislava alone, of which the most valuable are newspapers with OCR accessible content);

d) **complex** memory portals such as Slovakiana (almost 70,000 digitized items from Slovak National Archives, Slovak National Museum, Bratislava City Museum and others) or PamMap.sk (not only approx. 30,000 digitized historical sources, but also an encyclopaedia and map application linked to the towns.sk portal, where historical city atlas of Bratislava can be found – unlike Slovakiana, also with many private sources). The goal of the contribution is to compare the pros and cons of these portals and look at them with a critical eye of the user.

Možnosti, výhody, prínosy, prekážky a nástrahy vyučovania Digital humanities v historických vedách (niekoľko postrehov z prípravy vysokoškolského vyučovacieho predmetu)

Peter Labanc

Katedra histórie, Filozofická fakulta, Trnavská univerzita v Trnave, Hornopotočná 23, 918 43 Trnava, peter.labanc@truni.sk

Dne je už klišé tvrdiť, že digitálne technológie zasahujú do každej oblasti ľudskej činnosti. Je to fakt, ktorý už nepodlieha diskusii. Tá by sa však mala zamerať na to, akým spôsobom vieme a môžeme čo najlepšie využiť spomenuté technológie v humanitných vedách, v našom prípade v historickom výskume. Súčasťou týchto úvah musí byť bezpodmienečne aj otázka prípravy budúcich historikov – terajších študentov na využitie informačných technológií nielen pri ich štúdiu, ale tiež potenciálne v ich profesionálnom živote budúceho historika, resp. zamestnanca pamäťových inštitúcií.

Príspevkom k tejto diskusii je vytvorenie nového dvojsemestrálneho univerzitného predmetu Digital humanities v historických vedách a archeológii, začleneného do odporúčaného študijného programu na Katedre histórie a Katedre klasickej archeológie Filozofickej fakulty Trnavskej univerzity v Trnave so začatím výučby v akademickom roku 2022/2023. Predložený príspevok v sebe sumarizuje skúsenosti z formovania a vytvárania tohto predmetu. Tie sa dajú rozdeliť do niekoľkých skupín v chronologickej postupnosti ich riešenia.

Vzhľadom na nepreberné množstvo spôsobov uplatnenia digitálnych nástrojov v historických vedách bolo v prvej fáze našich úvah najprv potrebné špecifikovať najvhodnejšie nástroje využiteľné študentmi nielen pri ich ďalšom štúdiu, ale tiež aj v ďalšom uplatnení. Táto selekcia sa nezaobišla bez porovnávania výhod a prínosov jednotlivých možností v aktuálnych trendoch historického výskumu. Na zreteli bolo potrebné mať zároveň aj strednodobé a dlhodobé hľadisko ich využitia, udržateľnosti a prínosu v ďalšom profesionálnom živote študentov historických vied. A to bez ohľadu na to, či budú pracovať priamo vo vedeckovýskumnej oblasti, alebo v profesiách s históriou previazaných len okrajovo alebo nepriamo.

Do formovania predmetu musela prehovoriť aj materiálna stránka a potenciál udržateľnosti náročnejších IT nástrojov a infraštruktúry v kontexte dlhodobého podfinancovania humanitných vied na Slovensku. Z tejto núdze sa podarilo urobiť cnosť v podobe vystavania predmetu na cenovo udržateľných softvérových platformách.

Possibilities, advantages, benefits, obstacles and pitfalls of teaching Digital Humanities in historical sciences (a few observations from the preparation of a university course)

Peter Labanc

Department of History, Faculty of Arts, Trnava University in Trnava, Hornopotočná 23, 918 43 Trnava, peter.labanc@truni.sk

Today, it is already a cliché to say that digital technologies affect every area of human activity. It is a fact that is no longer subject to debate. However, the debate should focus on how we know and can best use the mentioned technologies in the humanities, in our case, in historical research. These discussions must also include the question of preparation of future historians – current students for the use of information technology not only during their studies, but also potentially in their professional life as a future historian, or employee of memory institutions.

The contribution to this discussion is the creation of a new two-semester university subject called Digital Humanities in historical sciences and archeology included in the recommended study program at the Department of History and the Department of Classical Archeology of the Faculty of Arts of the University of Trnava in Trnava, starting in the academic year 2022/2023. The submitted contribution summarizes the experiences of forming and creating this university subject. These experiences can be divided into several groups in the chronological order of their solution.

Due to the endless number of ways of applying digital tools in historical sciences, in the first phase of our considerations, it was first necessary to specify the most suitable tools to be used by students not only in their further study, but also in further application. This selection did not go without comparing the advantages and benefits of individual options in the current trends of historical research. At the same time, it was necessary to take into account the medium and long-term point of view of their use, sustainability and contribution in the further professional life of students of historical sciences. And that regardless of whether they will work directly in the scientific and research field, or in professions connected to history only marginally or indirectly.

The material side and the potential of sustainability of more demanding IT tools and infrastructure in the context of long-term underfunding of the humanities in Slovakia had to speak into the formation of the subject. This necessity has been turned into a virtue in the form of building and basing the subject on cost-effective software platforms.

Miért vágtunk bele saját fejlesztésbe? A Magyar Nemzeti Levéltár tapasztalatai a Transkribusszal és a szolgáltató alapú megközelítésekkel

Szatucsek Zoltán

Magyar Nemzeti Levéltár, 1014, Budapest Bécsi kapu tér 2-4, Magyarország, szatucsek.zoltan@mnl.gov.hu

Napjainkban talán egyetlen olyan újítás sem ígér a levéltárak számára olyan látványos áttörést, mint amit a kézírásfelismerés tud nyújtani. A mára már tekintélyes számban digitalizált gyűjtemények jelentős része ugyan online elérhető, de a tartalom gépi feldolgozásának, strukturált kinyerésének és szemantikus kereshetőségének legnagyobb akadálya a géppel olvasható szövegek hiánya. Az Európai Bizottság által két fejlesztési cikluson keresztül támogatott Transkribus platfhorm gyors elterjedése hozzájárult ahhoz, hogy a korábban csak kódok szintjén elérhető algoritmusok a hétköznapi felhasználók számára is kipróbálhatóvá, elérhetővé tegyék a technológiát. A gyorsan bővülő felhasználói közösség tapasztalatai egyrészt felhívták a figyelmet a Transkribus korlátaira, a szolgáltatások nagy volumenben való igénybevételének finanszírozhatóságára, a történeti kéziratok nagyfokú diverzitására, különösen az olyan speciális gyűjteményekre, mint a táblázatok vagy térképek. A Magyar Nemzeti Levéltár felismerve a lehetőséget, amit az őrizetében lévő levéltári anyag teljes szövegű kereshetőségének ígérete jelent, több projektben kezdett ismerkedni azokkal a folyamatokkal, amelynek az eredményeként szolgáltatásaiba tudja ágyazni az íly módon feldolgozott tartalmakat. Ennek a munkának az eddigi legnagyobb eredménye az 1828. évi összeírás teljes névanyagának, mintegy 170 ezer oldalnak és 2 millió egykorú személynévnek a publikálása. Jelen előadás a Transkribusszal folytatott próbálkozásokat bemutatva ismerteti az European Digital Treasures projekt során öt ország nemzeti levéltárával együtt megvalósított feldolgozást és azokat a tapasztalatokat, amelynek eredményeként a Transkribus alternatívájaként a levéltár saját munkafolyamat fejlesztésébe kezdett.

Why do we need our own processes? Experiences with Transkribus and vendor based solutions in the National Archives of Hungary

Zoltán Szatucsek

National Archives of Hungary, 1014, Budapest Bécsi kapu tér 2-4, Hungary, szatucsek.zoltan@mnl.gov.hu

Today, probably no other innovation promises for the archives such a remarkable breakthrough as handwriting recognition. While a vast number of collections digitised in the past years and decades and are now available online, the biggest obstacle to automatic processing, structured retrieval and semantic searchability is the lack of machine-readable texts. The rapid adoption of the Transkribus platform, supported by the European Commission over two development cycles, has helped to bring the algorithms, previously available only in code, to the everyday user. On the other hand the experiences of a rapidly expanding user community have drawn attention to the limitations of Transkribus, the financial feasibility of large-scale use of its services, the diverse needs of the sector and the great diversity of historical manuscripts, especially for special collections such as spreadsheets and maps. Recognising its potential for the archival material in its custody, the Hungarian National Archives has started to explore HTR technics in order to be able to embed new content into its services. The most spectacular result of this work so far is the publication of the complete names of the 1828 census, some 170,000 pages and 2 million personal names. This presentation will give an insight into the experiments inside and outside of Transkribus, the work carried out with five national archives in the framework of the European Digital Treasures project, and the experience gained from it, which has led us to start developing our own workflow as an alternative to the existing platforms.

Transkripcia historických dokumentov v kontexte Digital humanities

Dušan Katuščák

Štátna vedecká knižnica v Banskej Bystrici, Lazovná 9, 975 58 Banská Bystrica / Ústav bohemistiky a knihovnictví, Filozoficko-přírodovědní fakulta, Slezská univerzita v Opavě
Masarykova třída 343/37, 746 01 Opava, Česká republika,
dusankatuscak@gmail.com

V polytematickom príspevku sa autor venuje téme transkripcie (OCR a HTR) a niektorým výsledkom transkripcie historických dokumentov v projekte SKRIPTOR. Vysvetľuje metodologický kontext transkripcie v Digital humanities a poukazuje na globálne trendy v oblasti dokumentácie v pamäťových a fondových inštitúciách a systémoch (robotizácia, umelá inteligencia). Autor poukazuje na objektívne existujúce napätie medzi prudkým technologickým rozvojom, najmä technológie *digitalizácie, umelej inteligencie, virtuálnej reality, vizualizácie, rozšírenej reality,* ktoré nevyhnutne potrebujú „potravu" v podobe *dokumentov, vecných a osobných dát a poznatkov*, aby mohli knižnice, archívy a iné informačné inštitúcie poskytovať čo najkvalitnejšie digitálne informačné služby. Na jednej strane považuje za dôležité *vytvárať*, *zhromažďovať* digitálne dáta a poznatky a na druhej strane *využívať* tieto dáta a poznatky. Zdôrazňuje, že dnes už nie je hlavným cieľom a problémom samotná *tvorba* mohutných digitálnych archívov a repozitárov dát. Ťažisko sa prenáša na potrebu *využívania* dát a poznatkov v nových technologických projektoch a nástrojoch, ako je napr. rozšírená (obohatená) realita. Na základe vlastných experimentov opisuje konkrétne nevyhnutné kroky a postupy transkripcie historických dokumentov. Stručne opisuje proces tvorby modelov na základe vlastných nasnímaných a verejne dostupných digitálnych rukopisných a tlačených zdrojov. Zmieňuje sa o nákladoch na transkripciu. Z digitálnych zbierok Štátnej vedeckej knižnice v Banskej Bystrici v rámci projektu SKRIPTOR vykonal experimenty s týmito dokumentmi: časopis *Jitřenka* (Schwabacher) JPG formát, 1840, 128 s.; publikácia *Jánošík* (Schwabacher) JPG formát, 1862, 32 s.; noviny *Obzor* (Antikva a Schwabacher), PDF formát, 1866, 293 s.; monografia *Církev Ewanjelicko-Lutheránská* (Schwabacher), JPG formát, 1861, 375 s. Ďalej tvoril modely na základe českých zdrojov: noviny *Moravské noviny* (Antikva a Schwabacher), PNG formát, 1849, 20 s.; noviny *Opavský Besedník* (Antikva a Schwabacher) vo formáte PNG, 1861, 9 s. Vo vlastných modeloch transkripcie pre fraktúru a antikvu dosiahol parametre tréningového súboru

CER 0,03 % a validačného súboru CER 1,78 %. V transkripcii unikátnej horno- a dolnolužickej srbčiny podľa časopisu *Lužica* (1909) dosiahol presnosť transkripcie 99,2 % pri danom písme. V rámci projektu SKRIPTOR pracuje autor so študentmi na transkripcii historickej cca 500-stranovej rukopisnej nemeckej kuchárskej knihy z roku 1667 z archívu v Jeseníku. Na transkripciu boli nasnímané v archíve v Rimavskej Sobote s použitím zariadenia ScanTent a softvéru DocScan dokumenty: Mestečko Tisovec. *Kurentálny protokol* (Derivovaný súbor PDF, 500 MB, cca 500 obr., 175 dvojstrán. Rukopis, slovenčina, ca 1780 a n.) a *Školská zápisnica*, rukopis, slovenčina, 1836, cca 110 MB. Potrebný je ďalej dôkladný opis dokumentov, tréning modelu, transkripcia, sprístupnenie. Uvádza, ktoré dokumenty sú verejne dostupné cez platformu Read&Search. Opisuje stav rozpracovanosti svojho projektu virtuálnej zbierky *Collectanea* Martina Laučeka, v rámci ktorého sa usiluje jednak sústrediť všetky nasnímané zväzky Collectaney z rôznych zdrojov (SNA, SNK, OSzK, UKB), vytvoriť model transkripcie a sprístupniť virtuálnu zbierku Collectanea. Napokon navrhuje inštitucionalizovať transkripciu historických dokumentov na Slovensku a načrtáva zameranie takejto novej entity.

Transcription of historical documents in the context of "digital humanities"

Dušan Katuščák

State Scientific Library in Banská Bystrica, Lazovná 9, 975 58 Banská Bystrica / Institute of the Czech Language and Library Science, Faculty of Philosophy and Science, Silesian University in Opava, Masarykova třída 343/37, 746 01 Opava, Czech Republic, dusankatuscak@gmail.com

In this polythematic contribution, the author addresses the topic of transcription (OCR and HTR) and some results of the transcription of historical documents in the SKRIPTOR project. He explains the methodological context of transcription in "digital humanities" and points to global trends in the field of documentation in memory and fund institutions and systems (robotization, artificial intelligence). The author points to the objectively existing tension between the rapid technological development, especially the technologies of *digitization, artificial intelligence, virtual reality, visualization, augmented reality*, which inevitably need "sustenance" in the form of *documents, material and personal data and knowledge*, so that libraries, archives and other information institutions can provide the highest quality digital information services possible. He considers it important, on the one hand, *to create and collect* digital data and knowledge and, on the other hand, *to use* this data and knowledge. He emphasizes that today *the creation* of massive digital archives and data repositories is no longer the main goal and problem. The focus is transferred to the need *to use* data and knowledge in new technological projects and tools, such as augmented (enriched) reality. Based on his own experiments, he describes the specific necessary steps and procedures needed for the transcription of historical documents. He briefly describes the process of creating models based on his own scanned and publicly available digital manuscript and printed sources. He mentions the cost of transcription. He conducted experiments with the following documents from the Slovak Scientific Library in Banská Bystrica´s digital collections as part of the SKRIPTOR project: magazine *Jitřenka* (Schwabacher) JPG format, 1840, 128 p.; publication *Jánošík* (Schwabacher) JPG format, 1862, 32 p.; newspaper *Obzor* (Antikva and Schwabacher), PDF format, 1866, 293 p.; monograph *Církev Ewanjelicko-Lutheranská* (Schwabacher), JPG format, 1861, 375 p. He also created models based on Czech sources: the newspaper *Moravské noviny* (Antikva and Schwabacher), PNG format, 1849, 20 p., the newspaper *Opavský Besedník* (Antikva and Schwabacher) in PNG

format, 1861, 9 p. In his own transcription models for fracture and antiquity, he achieved the parameters of the training set CER 0.03% and the validation set CER 1.78%. In the transcription of the unique Upper and Lower Lusatian Serbian according to the magazine *Lužica* (1909), he achieved a transcription accuracy of 99.2% for the given script. As part of the SKRIPTOR project, the author is working with students on the transcription of an approx. 500-pages long, historical, handwritten German cookbook from the year 1667, which is from the archive in Jeseník. The following documents were scanned for transcription in the archive in Rimavská Sobota using the ScanTent device and the DocScan software: The town of Tisovec. *Curental protocol* (Derived PDF file, 500 MB, approx. 500 images, 175 double pages. Manuscript, Slovak, approx. 1780 and later), and *School memorandum*, manuscript, Slovak, 1836, approx. 110 MB. Furthermore, a thorough description of the documents, model training, transcription, and of access are needed. He states which documents are publicly available via the Read&Search platform. He describes the state of development of his project regarding the Martin Lauček's *Collectanea* virtual collection, in which he is trying to gather together all scanned Collectanea volumes from various sources (SNA, SNK, OSzK, UKB), create a transcription model and make the Collectanea virtual collection accessible. Finally, he proposes to institutionalize the transcription of historical documents in Slovakia and outlines the focus of such a new entity.

Automatická transliterácia rukopisnej kazateľskej tvorby Izáka Abrahamidesa (prvé modely a predpoklady ďalšieho postupu)

Pavol Maliniak

Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 974 01 Banská Bystrica, pavol.maliniak@umb.sk

Zachované rukopisné dielo Izáka Abrahamidesa Hrochotského (1557 – 1621), evanjelického kazateľa vo Zvolene, tvorí významný a doposiaľ málo využívaný prameň k náboženským, kultúrnym, sociálnym a jazykovým pomerom. Zbierka kázní doteraz nebola bádateľsky sprístupnená v edícii, poskytuje preto bohaté možnosti na prácu v platforme Transkribus. Prednosti automatického rozpoznávania rukopisu sa prejavujú zvlášť v tomto prípade, keď ide o text písaný rovnakou rukou v krátkom časovom úseku (písmo sa preto zásadne nemení) a rukopis dosahuje relatívne veľký rozsah.

Rukopisná postila predstavuje Abrahamidesov originál z prelomu 16. a 17. storočia, písaný súbežne novogotickým aj humanistickým písmom v rozsahu 722 strán. Niekoľko strán je vakantných, pričom porušené s čiastočne chýbajúcim textom sú iba štyri strany. Postila pozostáva z 30 kázní a troch prednášaných rečí. Venujú pozornosť moru, drahote a hladu, odpustkom, strigám, pominuteľnosti ľudského života, úžere a osmanskému ohrozeniu. Jednotlivé kázne pôvodne tvorili osobitné zošity. Neskôr boli v pozmenenom poradí zviazané v novej väzbe. Jazykom postily je slovakizovaná čeština, dokument preto tvorí významnú jazykovú pamiatku. Rukopis je uložený v Literárnom archíve Slovenskej národnej knižnice v Martine, kde ho pracovníci archívu zdigitalizovali s rozlíšením 200 DPI. Po počiatočných obavách sa takéto rozlíšenie ukázalo ako dostačujúce na prácu v prostredí Transkribus. Prameň navonok vystihuje čeština s rôznym podielom slovakizmov (ale i hungarizmov a germanizmov), avšak tento dominantný jazyk dopĺňajú početné pasáže v latinčine, sporadické pasáže v gréčtine, veľmi zriedkavo aj v hebrejčine. Hoci je text vyhotovený jednou pisárskou rukou, vzhľadom na používanie rôznych jazykov sa mení paleografická charakteristika. Z uvedených dôvodov do manuálneho prepisu (ortograficky vernej transliterácie) neboli zahrnuté grécke a hebrejské pasáže, ktoré nedosahujú dostatočný počet slov potrebných na trénovanie modelu. V rukopise za určujúci možno považovať český a latinský text, ktoré sa navzájom líšia štýlom aj druhom písma. Grafémy v obidvoch typoch textu sa síce odlišujú, zároveň však majú aj spoločné znaky alebo sa prelínajú.

Po segmentácii jednotlivých riadkov v rukopise sme pristúpili k vytváraniu modelov v softvéri Transkribus. Prvý model tvorila cvičná vzorka s počtom 10 119 slov. Chybovosť znakov (CER) bola v cvičnom súbore 0,26 % a v overovacom súbore 11,44 %, čo možno považovať za stále zrozumiteľný výsledok. Druhý model, ktorý pracoval s menšou cvičnou vzorkou (2 652 slov), vykazoval chybovosť v cvičnom súbore 0,27 % (veľmi podobná hodnote predchádzajúceho modelu) a v overovacom súbore 9,68 % (mierny pokles). Tretí model pracoval so zlúčením väčšej časti dovtedy využívaných cvičných vzoriek (11 434 slov). Chybovosť v tomto modeli bola v cvičnom súbore 1,55 %, avšak v dôležitejšom overovacom súbore klesla na 7,99 %, čo je použiteľný výsledok. Na základe modelov sa ukazuje, že ich efektivita sa zvyšuje s počtom slov zadávaných do cvičnej vzorky. Bude preto potrebné pracovať s počtom až 15 000 slov, ktoré odporúča Transkribus. Dôvodom je paleografická náročnosť rukopisu. Zámerom pri tvorbe ďalších modelov je znížiť chybovosť približne na 5 %, v ideálnom prípade aj nižšiu.

Automatic transliteration of the manuscript preaching work of Isaac Abrahamides (first models and assumptions for further progress)

Pavol Maliniak

Department of History, Faculty of Arts, Matej Bel University in Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, pavol.maliniak@umb.sk

The preserved manuscript work of Isaac Abrahamides Hrochotius (1557 – 1621), an evangelical preacher in Zvolen, constitutes an important and so far little-used source for religious, cultural, social and linguistic matters. The collection of sermons has not yet been made available for research in an edition; that is why it provides wide possibilities for work in the Transkribus platform. The merits of the automatic handwriting recognition become evident especially in this case, when it comes to the text written by the same hand in a short period of time (the script therefore does not fundamentally change) and the handwriting reaches a relatively large range.

The manuscript Postil represents Abrahamides' original from the turn of the 16th and 17th centuries, written simultaneously in neo-Gothic and Humanistic script, in the extent of 722 pages. Several pages are vacant, while only four pages are disrupted with partially missing text. The Postil consists of 30 sermons and three presented speeches. They pay attention to plague, costliness and hunger, indulgences, witches, the transience of human life, usury and the Ottoman Threat. Individual sermons originally formed separate notebooks. Later, they were bound in a new book binding in a changed order. The language of the Postil is Slovakized Czech; therefore the document forms a significant linguistic monument. The manuscript is stored in the Literary Archive of the Slovak National Library in Martin, where the archive staff digitized it with a resolution of 200 DPI. After initial concerns, such resolution proved to be sufficient for working in the Transkribus environment. The source outwardly epitomizes Czech with a varying portion of Slovakisms (but also Hungarianisms and Germanisms), however this dominant language is complemented by numerous passages in Latin, sporadic passages in Greek, and very rarely also in Hebrew. Although the text is made by one scribal hand, the palaeographic characteristics change due to the use of different languages. For the stated reasons, the Greek and Hebrew passages, which do not reach a sufficient number of words needed for training the model, were not included in the manual transcription (orthographically faithful transliteration). The

Czech and Latin texts, which differ from each other in style and type of the script, can be considered to be decisive in the manuscript. Even though the graphemes in both types of text are different, they have common features or they overlap.

After the segmentation of individual lines in the manuscript, we proceeded to create models in the Transkribus software. The first model consisted of a training sample with 10,119 words. The character error rate (CER) was 0.26% in the training set and 11.44% in the testing set, which can be considered a still understandable result. The second model, which worked with a smaller training sample (2652 words), showed an error rate of 0.27% in the training set (very similar to the value of the previous model) and 9.68% in the testing set (a slight decrease). The third model worked with the unification of a larger part of the training samples used until then (11,434 words). The error rate in this model was 1.55% in the training set, but dropped to 7.99% in the more important testing set, which is a usable result. On the basis of the models, it is shown that their effectiveness increases with the number of words entered into the training sample. It will therefore be necessary to work with the number of up to 15,000 words recommended by Transkribus. The reason is the palaeographic complexity of the manuscript. The intention when creating other models is to reduce the error rate to approximately 5%, ideally even lower.

Ukázka práce transkripce v platformě Transkribus na příkladu vzácné kuchařské knihy z roku 1667

Klára Kováčová

Ústav bohemistiky a knihovnictví, Filozoficko-přírodovědní fakulta, Slezská univerzita v Opavě, Masarykova třída 343/37, 746 01 Opava, Česká republika, F190283@fpf.slu.cz

Rukopisné archiválie jsou již po tisíce let nesmírnou pokladnicí informací. Bohužel, i přes veškerou snahu, kterou člověk vynakládá na zachování těchto neocenitelných pokladů, se jednoho dne fyzicky rozpadnou. Díky posunu v technice a moderním zařízením se nám daří je zanechat budoucím generacím, například formou digitalizace apod. Klasické transkripce textů, které pomáhají tyto skvosty udržovat, jsou již ve světě celkem známé. Kdežto transkripce historických rukopisů jsou zatím stále brány jako celkem nové odvětví, které se nacházejí ve stádiu výzkumů a sbírání co největšího množství materiálů k práci s nimi.

Hlavním cílem referátu proto bude poukázat na možnosti, jak uplatnit nástroje umělé inteligence (v našem případě platformy Transkribus = komplexní platforma na digitalizaci, rozpoznávání textu pomocí umělé inteligence, dále přepisu a vyhledávání historických dokumentů v jakémkoliv jazyku v rámci transkripce). Dokázat, že tato celkem stále nová metoda, zajišťuje zlepšení dostupnosti a celkového zpřístupnění historických rukopisů. Poukážeme nejen na nutnost vytvoření vlastního modelu transkripce, ale představíme vlastní model, na kterém jsme práci s umělou inteligencí použili. Model byl hledán dle přesně daných a stanovených kritérií. Jedná se o starou kuchařskou knihu z roku 1667, nalezenou náhodou na půdě Vodní tvrze v Jeseníku, dnešním sídle Vlastivědného muzea Jesenicka v České republice. Dnes je již sbírka uložena ve Státním okresním archivu Jeseník. Obsahuje soupis kuchařských receptů, především panských jídel. Celý originální název v německém jazyce zní: *Koch und Einmachbuch von Allerley Eingemachten Sachen von Zucker, Hönig und aller Früchten, auch und erschiedlicher gueten Speisen*, tedy volně přeloženo jako *Kuchařka potravin a zavařenin. Konzervované všechny věci z cukru, medu a všeho ovoce, a také z různých dobrých potravin*. Tento velmi objemný rukopis, který skýtá přes 800 stran, je psán v typu písma kurent, tedy německou verzí běžné novogotické kurzívy používané od 16. století. Autor této kuchařské knihy nebyl zjištěn. Právě díky tomuto vlastnímu modelu jsme schopni ověřit použitelnost jak zařízení ScanTent, potřebného pro prováděnou transkripci, tak i samotného softwaru DocScan, který se používá

při snímání rukopisů. Budeme tedy schopni získat poznatky a data pro závěr a ověření funkcionality platformy Transkribus a představit tím inovativní přístup ke zpřístupnění rukopisného písemného dědictví. Součástí referátu budou i přiložené fotografie přímo z prováděného výzkumu na vybraných rukopisech a také úspěšné ukázky modelů dané transkripce z platformy Transkribus. Celý tento výzkum byl i součástí bakalářské práce na dané téma, které spadalo do oblasti Digital humanities, předmětem zájmu ale stále zůstala oblast archivnictví.

An example of transcription work in the Transkribus platform using the example of a valuable cookbook from 1667

Klára Kováčová

Institute of the Czech Language and Library Science, Faculty of Philosophy and Science, Silesian University in Opava, Masarykova třída 343/37, 746 01 Opava, Czech Republic, F190283@fpf.slu.cz

Manuscript archives have been an immense treasury of information for thousands of years. Unfortunately, despite all the effort that man puts into preserving these priceless treasures, one day they will physically disintegrate. Thanks to the advances in technology and modern equipment, we manage to leave them to future generations, for example, through digitization, etc. Classical transcriptions of texts that help preserve these masterpieces are already quite well-known in the world. Whereas the transcriptions of historical manuscripts are still considered as a completely new industry, which is in the stage of research and collecting as much material as possible to work with them.

The main goal of the paper will therefore be to point out the possibilities of how to apply artificial intelligence tools (in our case, the Transkribus platform = a comprehensive platform for digitization, text recognition using artificial intelligence, as well as transcription and search of historical documents in any language within the transcription). To prove that this still quite new method ensures the improvement of availability and overall accessibility of historical manuscripts. We will not only point out the necessity of creating our own transcription model, but we will present our own model on which we used the work with artificial intelligence. The model was searched according to precisely given and established criteria. This is an old cookbook from 1667, found by chance on the grounds of the Water Fortress in Jeseník, today the seat of the History and Geography Museum of Jeseníky Region in the Czech Republic. Today, the collection is stored in the Regional State Archive Jeseník. It contains a list of cooking recipes, mainly for meals of the nobility. The full original title in German reads: *Koch und Einmachbuch von Allerley Eingemachten Sachen von Zucker, Hönig und aller Früchten, auch und erschiedlicher gueten Speisen*, thus loosely translated as *Cookbook of foods and preserves. Canned all things of sugar, honey, and all fruits, and also of various good foods*. This very voluminous manuscript, which contains over

800 pages, is written in the Kurrent font, a German version of the common Neo-gothic cursive used since the 16th century. The author of this cookbook has not been identified. Thanks to this own model, we are able to verify the usability of both the ScanTent device, which is needed for the transcription being carried out, and the DocScan software itself, which is used when scanning manuscripts. We will therefore be able to obtain knowledge and data to conclude and verify the functionality of the Transkribus platform and that way present an innovative approach to making manuscript written heritage available. The paper will also include attached photographs directly from the research conducted on selected manuscripts, as well as successful examples of models of the given transcription from the Transkribus platform. This entire research was also part of the Bachelor's Thesis on the given topic, which fell into the area of "digital humanities", but the area of archival science still remained as a subject of interest.

Automatická transkripcia historických prameňov obsahujúcich viac rukopisov na príklade kanonických vizitácií

Martin Katreniak – Patrik Kunec

Novohradské múzeum a galéria, Námestie Kubínyiho 38/3, 984 01 Lučenec, martin.katreniak222@gmail.com
Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 974 01 Banská Bystrica, patrik.kunec@umb.sk

Automatická transkripcia historických prameňov by mohla nájsť uplatnenie aj pri prepise špecifického prameňa k cirkevným dejinám, ktorými sú kanonické vizitácie. V slovenských archívoch (či už štátnych alebo cirkevných) sú totiž prístupné v zdigitalizovanej podobe. Na našom území bolo konanie vizitácií jednotlivých farností nariadené spolu zo závermi Tridentského koncilu, ale správy o výsledkoch vizitácií v písomnej podobe sa nám dochovali len od začiatku 17. storočia. Texty vizitačných protokolov poskytujú historikom zaujímavé informácie nielen o podobe náboženského života v konkrétnej farnosti, ale často obsahujú informácie aj o fungovaní školy, o cirkevných stavbách a hmotnom majetku či o finančných príjmoch kňaza.

Na automatickú transkripciu v rámci projektu TRANSKRIBUS boli vybrané vizitačné protokoly z fondov Banskobystrického diecézneho archívu, konkrétne išlo o vizitačné protokoly z rokov 1756 a 1783. Cieľom výskumu bolo zistiť možnosti aplikovania metódy automatickej transkripcie na protokoly kanonických vizitácií písané po latinsky, ktoré sa často skladajú z viacerých rukopisov. Hlavným cieľom bolo zistiť, akým spôsobom je najideálnejšie pristupovať k automatickej transkripcii takéhoto typu rukopisného prameňa. Čiastkovým cieľom bolo tiež preskúmať, ako ovplyvňuje pridávanie ďalších rukopisov presnosť trénovaného modelu. Sekundárnym cieľom bolo zistiť, kedy sa dá považovať využitie metódy automatickej transkripcie za efektívne a aké sú minimálne podmienky na vytvorenie funkčného malého modelu skladajúceho z jedného alebo aj viacerých rukopisov. Ciele výskumu spočívali v troch parciálnych úlohách, prípadne fázach výskumu. V rámci prvej fázy boli vytvorené tri modely, pričom každý bol založený na jednom rukopise konkrétneho vizitátora. Cieľom bolo získať čo najnižšiu mieru chybovosti pri automatickom prepise v programe TRANSKRIBUS a preskúmať minimálne požiadavky na vytvorenie funkčného malého modelu. Druhá fáza výskumu bola zameraná na vytvorenie modelu skladajúceho sa z viacerých rukopisov.

Cieľom bolo vytvoriť jednotný model, ktorý by dokázal dostatočne presne transkribovať všetky rukopisy jedného protokolu. Tretia fáza výskumu bola zameraná na preskúmanie možností využívania už vytvoreného, tzv. základného modelu (base model) pri vytváraní nových modelov založených na jednom alebo aj viacerých rukopisoch. Vo všetkých troch rovinách bol výskum realizovaný len pri vizitačnom protokole z roku 1783, pri práci s vizitačným protokolom z roku 1756 bola realizovaná len prvá fáza (zvyšné dve budú realizované dodatočne, referované o nich bude počas plánovanej konferencie).

V rámci prvej úlohy sme zistili, že na vytvorenie funkčného modelu stačí aj značne menší počet tréningového materiálu (čiže prepísaných strán z jednotlivých vizitačných protokolov), ako je odporúčaných 15 000 slov. Modely, ktoré sme vytvorili pri transkripcii vybraných častí vizitácie z roku 1783, dosiahli hodnotu chybovosti (CER = Character Error Rate) v overovacom súbore v rozmedzí 8,75 % až 2,42 %. Najpresnejší z týchto modelov bol pritom vytrénovaný na cvičnej vzorke obsahujúcej len 2 991 slov. Modely vytvorené pri transkripcii vybraných častí vizitácie z roku 1756 vykazovali chybovosť okolo 7 % (hodnoty CER pri troch rôznych rukopisoch v overovacích súboroch: 6,99 %, 7,05 %, 6,97 %). Aj pri modeloch vytváraných z tohto prameňa siahal počet slov v trénovanej vzorke v rozmedzí od 3 285 po 5 482 slov. Pri sledovaní príčin miery chybovosti sa ukázalo, že presnosť malých modelov výrazne závisí nielen od presnosti prepisu vybraných pasáží, ale tiež od detailnej segmentácie parciálnych riadkov skenovaného textu prameňa a aj od konzistentnosti rukopisu naprieč celým dokumentom.

Pri kanonickej vizitácii z roku 1783 sa riešenie úloh posunulo aj do druhej a tretej fázy. V druhej fáze bol vytvorený jeden tréningový model založený na odlišných rukopisoch, ktoré si však boli paleograficky podobné. Finálny model dosiahol hodnotu chybovosti 3,70 %. Išlo o plne funkčný model, ktorý dokázal „prečítať“ všetky tri vybrané rukopisy s dostatočnou presnosťou. Porovnaním čiastkových výsledkov s referenčnými modelmi sme tiež zistili, že spoločné trénovanie malo jednoznačne pozitívny dosah nielen na celkovú presnosť finálneho modelu, ale aj na zlepšenie presnosti malých modelov pri jednotlivých rukopisoch.

V tretej fáze bolo vyskúšané použitie tzv. base modelu (vybraný bol model s názvom *NeoLatin_Ravenstein_1643-1772*). Výsledky pokusu ukázali, že aplikovanie base modelu malo pozitívny vplyv na presnosť nových modelov založených na jednom, ako aj na viacerých rukopisoch. Po aplikovaní base modelu na referenčný model, založený na jednom rukopise, sme dosiahli zlepšenie chybovosti z 8,75 % na 5,50 %. Po aplikovaní base modelu na kombinovaný model, skladajúci sa z troch rukopisov, bolo dosiahnuté zlepšenie

chybovosti z 3,70 % na 3,25 %. Využitie base modelu sa teda preukázalo ako veľmi vhodný postup na zlepšenie presnosti prepisu pri malých modeloch. Efektívnosť tohto postupu ale závisí od voľby takého base modelu, ktorý vznikol prepisom podobného typu prameňa napísaného v rovnakom jazyku (latinčine) a v približne rovnakom čase.

Dosiahnuté výsledky poukazujú na efektivitu využitia aj podstatne menších cvičebných súborov (training set). Spoločné trénovanie viacerých rukopisov sa javí v mnohých smeroch ako efektívnejšia alternatíva oproti ich samostatnému trénovaniu. Využitie base modelu sa preukázalo ako jednoznačne výhodné pri vytváraní nových malých modelov založených na jednom alebo aj viacerých rukopisoch. Tieto postupy budú ďalej rozvíjané pri tvorbe modelov v rámci automatickej transkripcie vybraných vizitačných protokolov z územia Slovenska.

Automatic transcription of historical sources containing several manuscripts on the example of canonical visitations

Martin Katreniak – Patrik Kunec

Novohrad Museum and Gallery, Námestie Kubínyiho 38/3, 984 01 Lučenec, martin.katreniak222@gmail.com
Department of History, Faculty of Arts, Matej Bel University in Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, patrik.kunec@umb.sk

Automatic transcription of historical sources could also find use in the transcription of a specific source for church history, which are canonical visitations. They are accessible in digitized form in Slovak archives (either state or church). In our territory, the conduct of visitations of individual parishes was ordered together with the conclusions of the Council of Trent, but reports on the results of visitations in written form have only survived from the beginning of the 17th century. The texts of the visitation protocols provide historians with interesting information not only about the form of religious life in a specific parish, but also often contain information about the functioning of the school, about church buildings and material property, or about the priest's financial income.

For automatic transcription within the TRANSKRIBUS project, visitation protocols from the holdings of the Banská Bystrica Diocesan Archive were selected, namely visitation protocols from the years 1756 and 1783. The aim of the research was to find out the possibilities of applying the automatic transcription methods to protocols of canonical visitations written in Latin, which often consist of several manuscripts. The main goal was to find out the most ideal way to approach the automatic transcription of this type of manuscript source. A sub-goal was also to examine how adding more manuscripts affects the accuracy of the trained model. The secondary goal was to find out when the use of the automatic transcription method can be considered effective and what are the minimum conditions for creating a functional small model, which consists of one or more manuscripts. The objectives of the research consisted of three partial tasks, or phases of the research. Within the first phase, three models were created, each based on one manuscript of a specific visitor. The aim was to obtain the lowest possible error rate during automatic transcription in the TRANSKRIBUS program and to examine the minimum requirements for creating a functional small model.

The second phase of the research was focused on creating a model consisting of several manuscripts. The aim was to create a unified model that could accurately enough transcribe all the manuscripts of one protocol. The third phase of the research focused on examining the possibilities of using the already created, so-called base model in the creation of new models, based on one or more manuscripts. In all three levels, the research was realized only for the visitation protocol from 1783, when working with the visitation protocol from 1756, only the first phase was realized (the remaining two will be realized subsequently, they will be reported on during the planned conference).

As part of the first task, we came to the conclusion that for the creation of a functional model is sufficient a much smaller amount of training material (i.e. transcribed pages from individual visitation protocols) than the recommended 15,000 words. The models we created while transcribing selected parts of the visitation from 1783 achieved an error rate (CER = Character Error Rate) in the testing set in the range of 8.75% to 2.42%. The most accurate of these models was trained on a training sample containing only 2991 words. The models created during the transcription of selected parts of the visitation from 1756 showed an error rate of around 7% (CER values for three different manuscripts in the testing sets: 6.99%, 7.05%, 6.97%). Even with models created from this source, the number of words in the trained sample reached in the range of 3285 to 5482 words. When observing the causes of the error rate, it became clear that the accuracy of small models depends not only on the accuracy of the transcription of selected passages, but also on the detailed segmentation of partial lines of the scanned text of the source and also on the consistency of the manuscript across the entire document.

During the canonical visitation of 1783, the solution of the tasks also moved into the second and third phase. In the second phase, one training model was created based on different manuscripts, which were, however, paleographically similar. The final model achieved an error rate of 3.70%. It was a fully functional model that could “read” all three selected manuscripts with sufficient accuracy. By comparing partial results with reference models, we also found that joint training had a clear positive impact not only on the overall accuracy of the final model, but also on improving the accuracy of small models for individual manuscripts.

In the third phase, the use of the so-called base model was tested (the model named *NeoLatin_Ravenstein_1643-1772* was selected). The results of the experiment showed that the application of the base model had a positive effect on the accuracy of new models based on one as well as on several manuscripts. After applying the base model to the reference model, based on

one manuscript, we achieved an improvement in the error rate from 8.75% to 5.50%. After applying the base model to the combined model, consisting of three manuscripts, an improvement in the error rate was achieved from 3.70% to 3.25%. Thus, the use of a base model has proven to be a very suitable procedure for improving the accuracy of the transcription for small models. However, the effectiveness of this procedure depends on choosing such a base model, which was created by transcribing a similar type of source, written in the same language (Latin) and at approximately the same time.

The achieved results point to the effectiveness of using even significantly smaller training sets. In many ways, the joint training of several manuscripts appears to be a more effective alternative compared to their separate training. The use of the base model has proven to be clearly advantageous when creating new small models based on one or more manuscripts. These procedures will be further developed during the creation of models within the automatic transcription of selected visitation protocols from the territory of Slovakia.

Využitie softvéru Transkribus na automatickú transliteráciu štyroch typov fontu štvorjazyčnej historickej tlače

Michaela Mikušková – Lucia Nižníková

Univerzitná knižnica Univerzity Mateja Bela v Banskej Bystrici, Tajovského 40, 974 01 Banská Bystrica, michaela.mikuskova@umb.sk, lucia.niznikova@umb.sk

Platforma Transkribus, ktorá využíva technológiu strojového učenia, bola vyvinutá na rozpoznávanie historických rukopisných textov a ich automatickú transkripciu. Analogicky by mohla fungovať aj v prípade historických tlačí. Rozhodli sme sa túto hypotézu overiť na diele J. A. Komenského *Orbis Pictus*. Vydanie z roku 1798 vyšlo z dielne prešporského tlačiara Š. P. Webera, je napísané v štyroch jazykoch a pri tlači boli použité štyri typy písma. Na prepis sme uprednostnili metódu transliterácie pred transkripciou s cieľom zachovať všetky pôvodné grafémy. Keďže transliterácia umožňuje spätne rekonštruovať pôvodnú podobu slov, výsledný prepis tlače môže byť atraktívny nielen pre historikov, ale aj pre jazykovedcov či grafológov. Pri trénovaní tlače nie je nevyhnutné dodržať počet slov odporúčaný pre rukopisné texty, preto sme do cvičného súboru zaradili osem snímok (jedna snímka = dvojstrana originálneho dokumentu) s počtom 2 047 a 1 878 slov, do overovacieho súboru dve snímky s počtom 605 a 773 slov. Už prvé dva modely, v ktorých sme trénovali štyri fonty súčasne, priniesli relatívne dobré výsledky, ktoré sa zvyknú dosahovať pri historických tlačiach. Aby sme modely vylepšili, použili sme tri metódy: oprava chýb z manuálnej transkripcie cvičného súboru, trénovanie každého fontu samostatne, pridanie ďalších strán do cvičného súboru. Oprava chýb a vytrénovanie nových modelov priniesli mierne zlepšenie. Výsledky samostatného trénovania jednotlivých fontov neboli také uspokojivé, ako sme očakávali – naopak model fontu švabach s CER 4,78 % na overovacom súbore vykázal najhoršie výsledky, porovnateľné s rukopisnými dokumentmi. Pridaním ďalších strán do cvičného súboru sa potvrdila téza, že väčší rozsah textu, a teda vyšší výskyt jednotlivých grafém, pomáhajú softvéru pri rozpoznávaní a učení sa. V cvičnom súbore bolo 4 107 a 3 938 slov, v overovacom súbore 603 a 772 slov. Vytrénovaný model s CER 0,36 % na cvičnom a CER 1,02 % na overovacom súbore sa dá považovať za úspešný. V ďalšej fáze môžeme zisťovať jeho využiteľnosť na historických tlačiach toho istého kníhtlačiara, príp. iných tlačiach z toho istého obdobia.

Use of Transkribus software for automatic transliteration of four font types of four-language historical print

Michaela Mikušková – Lucia Nižníková

University Library of Matej Bel University in Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, michaela.mikuskova@umb.sk, lucia.niznikova@umb.sk

The Transkribus platform, which uses machine learning technology, was developed to recognize historical manuscript texts and automatically transcribe them. It could also work analogously in the case of historical prints. We decided to verify this hypothesis on J. A. Komenský's work *Orbis Pictus*. The edition from 1798 was published by the workshop of the printer Š. P. Weber from Prešporok, it is written in four languages and four fonts were used for printing. For transcription, we preferred the method of transliteration over transcription with the aim to preserve all the original graphemes. Since the transliteration makes it possible to reconstruct the original form of words, the resulting print transcription can be attractive not only to historians, but also to linguists and graphologists. When training print, it is not necessary to comply with the number of words recommended for manuscript texts, that is why we included 8 images into the training set (one image = two pages of the original document) with the number of 2,047 and 1,878 words, 2 images with the number of 605 and 773 words into the testing set. Already the first two models, in which we trained four fonts at the same time, brought relatively good results, which are usually achieved with historical prints. To improve the models, we used three methods: correcting errors from manual transcription of the training set, training of each font separately, adding more pages to the training set. Correcting errors and training new models brought a slight improvement. The results of training individual fonts separately were not as satisfactory as we expected – on the contrary, the Schwabach font model with a CER of 4.78% on the testing set showed the worst results, comparable to manuscript documents. By adding more pages to the training set, the thesis was confirmed that a larger range of text, and thus a higher presence of individual graphemes, helps the software in recognition and learning. There were 4,107 and 3,938 words in the training set, 603 and 772 words in the testing set. A trained model with a CER of 0.36% on the training set and a CER of 1.02% on the testing set can be considered successful. In the next phase, we can determine its usability on historical prints of the same printer, or other prints from the same period.

Vytváranie modelu na automatickú transkripciu novolatinského rukopisu reambulačného protokolu Banskej Bystrice v prostredí platformy Transkribus

Oto Tomeček

Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 974 01 Banská Bystrica, oto.tomecek@umb.sk

Digitalizácia spočívajúca v automatickom prepise historického rukopisného textu sa v súčasnosti stáva vítanou pomôckou historického výskumu, ako aj prípravy textových edícií historických dokumentov. Bádateľovi umožňuje rýchle vyhľadávanie a orientáciu v dokumente. Obrovským benefitom je tiež jeho eventuálne sprístupnenie odbornej a širšej laickej verejnosti v zrozumiteľnej a dobre čitateľnej podobe. Na automatický prepis historického rukopisného textu je možné využiť na tento cieľ utvorenú platformu Transkribus.

Samotnému procesu automatickej transkripcie rukopisného textu predchádza niekoľko nevyhnutných krokov. Prvým z nich je nasnímanie historického dokumentu prostredníctvom fotoaparátu mobilného telefónu pomocou ScanTentu a jeho následné importovanie do systému. Potom nasleduje segmentácia dokumentu, počas ktorej je potrebné vymedziť bloky textového poľa, označiť jednotlivé riadky a určiť poradie ich čítania. Po segmentácii dokumentu nasleduje manuálny prepis časti dokumentu, ktorý je nevyhnutný na vytvorenie a vytrénovanie modelu, prostredníctvom ktorého je možné naučiť program identifikovať jednotlivé alfanumerické znaky.

Na overenie možností automatickej transkripcie sme si zvolili rukopisný dokument zaznamenávajúci reambuláciu chotára Banskej Bystrice z roku 1820, ktorý je označený jednoslovným názvom *Metales*. Dokument, uložený v Slovenskom banskom archíve v Banskej Štiavnici, bol napísaný v latinskom jazyku jednou pisárskou rukou v priebehu rokov 1820 – 1823. Celkový rozsah dokumentu zviazaného v samostatnej knihe je 254 strán.

Celý dokument sme nafotili dvakrát prostredníctvom fotoaparátov na dvoch rôznych mobilných zariadeniach (Huawei P40 Pro a Google Pixel 4). Jednotlivé snímky zachytávajúce dvojstrany dokumentu sme roztriedili podľa kvality ich nasnímania. Najlepšiu verziu každej snímky sme následne importovali vo formáte JPEG do systému Transkribus. Po automatickej segmentácii dokumentu, ktorá sa neosvedčila, sme pristúpili k manuálnej segmentácii. Počas nej bolo potrebné vymazať všetky chybne označené polia, riadky a znaky. Nasledovalo spresnenie plochy textových polí, označenie poradia čítania jednotlivých riadkov a kontrola označenia dĺžky jednotlivých

riadkov. Práve manuálna segmentácia dokumentu je, spolu s manuálnym prepisom textu, najzdĺhavejšou časťou prípravných prác predchádzajúcich automatickej transkripcii. Na manuálny prepis textu, ktorý poslúži ako vzorka na vytrénovanie modelu, sme vybrali prvých 26 snímok dokumentu obsahujúcich 49 samostatných strán textu a 6 910 slov (necelá polovica z počtu 15 000 slov odporúčaných autormi Transkribusu).

Pri vytváraní modelu sme túto vzorku rozdelili približne v pomere 5 : 1 (autori Transkribusu odporúčajú pomer 10 : 1) na tréningový set (5 702 slov) a validačný, teda overovací set (1 208 slov). Na tréningovom súbore (Training set), na ktorom sa Transkribus učí identifikovať jednotlivé znaky, sa podarilo dosiahnuť chybovosť znakov (character error rate) na úrovni 1,20 %. Dôležitejší overovací súbor (Validation set), ktorý slúži na praktické odskúšanie modelu, ukazuje, ako dokáže program prečítať text podľa toho, čo sa naučil v tréningovom súbore. Pri prvom modeli sa v tomto prípade podarilo dosiahnuť chybovosť na úrovni 5,35 %.

Rovnakú vzorku textu sme sa následne pokúsili vylepšiť využitím takzvaného Base modelu, teda modelu voľne prístupného na platforme Transkribus. Pre naše potreby najlepšie vyhovoval Base model označený ako Neolatin Ravenstein 1643 – 1772. Tento model sa spomedzi všetkých dostupných modelov najviac približoval nášmu textu časovo aj typom písma. Po vylepšení využitím Base modelu sme vytvorili vlastný model číslo 2, pri ktorom sa chybovosť v tréningovom sete nepatrne zvýšila na 1,33 %, avšak v dôležitejšom validačnom sete sa znížila na hodnotu 4,74 %. Výsledok s chybovosťou znakov na úrovni pod 5 % sa pri rukopisoch považuje za veľmi dobrý výsledok. Aj pri vzorke textu s výrazne menším počtom slov, v porovnaní s odporúčaným počtom slov, sa podarilo dosiahnuť veľmi dobré výsledky. Hlavným výsledkom tohto snaženia je vytvorenie modelu, ktorý je možné použiť na automatickú transkripciu konkrétneho vybraného rukopisného historického dokumentu.

Creating a model for automatic transcription of the Neo--Latin manuscript of the reambulatory protocol of Banská Bystrica in the environment of the Transkribus platform

Oto Tomeček

Department of History, Faculty of Arts, Matej Bel University in Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, oto.tomecek@umb.sk

Digitization, consisting of automatic historical manuscript text transcription, is currently becoming a welcome aid to historical research, as well as to the preparation of text editions of historical documents. It enables the researcher to quickly search and orientate themselves in the document to the lay public in an understandable and easily readable form. The Transkribus platform, created for this purpose, can be used for the automatic transcription of historical manuscript text.

The process of automatic transcription of manuscript text itself is preceded by several necessary steps. The first of them is to capture a historical document using a mobile phone camera with the help of ScanTent and then to import the document into the system. This is followed by segmentation of the document, during which it is necessary to define blocks of the text field, mark individual lines and determine their reading order. Document segmentation is followed by manual transcription of a part of the document, which is necessary for creating and training a model, through which the program can be taught to identify individual alphanumeric characters.

To verify the possibilities of automatic transcription, we chose a handwritten document recording the reambulation of Banská Bystrica from 1820, which is marked with the single word *Metales*. The document, stored in the Slovak mining archive in Banská Štiavnica, was written in Latin by a single scribal hand during the years 1820 – 1823. The total scope of the document, bound in a separate book, is 254 pages.

We photographed the entire document twice using the cameras of two different mobile devices (Huawei P40 Pro and Google Pixel 4). We have sorted the individual images capturing both sides of the document according to the quality of their capture. We then imported the best version of each image into the Transkribus system in JPEG format. After automatic document segmentation, which did not prove successful, we proceeded to manual segmentation. During the process of manual segmentation, it was necessary to delete all wrongly marked fields, lines and characters. This was followed

by the refinement of the area of the text fields, the marking of the reading order of individual lines and the checking of the marking of the length of individual lines. The manual segmentation of the document is, together with the manual transcription of the text, the most time-consuming part of the preparatory work prior to the automatic transcription. To manually transcribe the text, which then serves as a sample for training the model, we selected the first 26 images of the document containing 49 separate text pages and 6910 words which is less than half of the 15,000 words recommended by the authors of Transkribus. When creating the model, we divided this sample approximately in a ratio of 5:1 (the authors of Transkribus recommend a ratio of 10:1) into a training set (5,702 words) and a validation, i.e. verification, set (1,208 words). It was possible to reach a character error rate of 1.20% on the training set, which Transkribus uses to learn to identify individual characters. However, the verification file is more important as it is used for practical testing of the model because it shows how the program can read text based on what it learned in the training file. In this case, the first model managed to achieve an error rate of 5.35%. We subsequently tried to improve the same text sample by using the so-called Base model, i.e. a model freely accessible on the Transkribus platform. For our needs, the Base model marked Neolatin Ravenstein 1643 – 1772 was the best fit. Among all the available models, this model was the closest to our text both in terms of time and font. After improving the transcription using the Base model, we created our own model number 2, in which the error rate slightly increased to 1.33% in the training set, but decreased to 4.74% in the more important validation set. A result with a character error rate below 5% is considered a very good result for manuscripts. Even with a text sample with significantly fewer words than the recommended number of words, very good results were achieved. The main result of this effort is the creation of a model that can be used for the needs of automatic transcription of a particular selected handwritten historical document.

Nové možnosti digitálneho prostredia vo výskume dejín knižnej kultúry

Mária Bôbová

Štátna vedecká knižnica v Banskej Bystrici, Lazovná 9, 975 58 Banská Bystrica, maria.bobova@svkbb.eu

Referát predstaví proces tvorby modelu rukopisu, využívajúc platformu Transkribus, ktorá je súčasťou projektu APVV Skriptor. Model je prioritne určený pre historický rukopisný katalóg cirkevnej knižnice Kňazského seminára sv. Karola Boromejského v Banskej Bystrici. Katalóg vznikol začiatkom 19. storočia a používal sa do polovice 19. storočia. Je súčasťou rozsiahlejšieho rukopisného dokumentu *Elenchus librorum*, ktorý je v súčasnosti vo vlastníctve Štátnej vedeckej knižnice v Banskej Bystrici a ktorý okrem dvoch knižničných katalógov zahŕňa aj prírastkové zoznamy a výpožičný katalóg Seminárnej knižnice.

Rukopisný katalóg predstavuje dôležitý prameň pre dejiny knižnice Kňazského seminára sv. Karola Boromejského (1807 – 1950), pretože približuje jej prvé obdobie pôsobenia. Na stotridsiatich siedmich listoch je rozpísaných 3 607 záznamov o knihách. Súpis kníh je rozčlenený do desiatich tematických celkov (Biblia, patrológia, dogmatika, scholastika, katechetika, polemická teológia, morálna teológia, asketika, homiletika, liturgika, cirkevné a svetské dejiny, právo, filozofia, pedagogika a iné), ktoré majú aj ďalšie vnútorné členenie. Jednotlivé strany v katalógu sú rozdelené stredovou čiarou na dve časti. Jedna časť slúžila na zápis záznamov o knihách v tvare názov, autor, miesto vydania, rok a formát. Druhá časť bola vymedzená značeniu poznámok o zmenách, či už išlo a nové prírastky do knižnice, predaj kníh, alebo preraďovanie kníh v katalógu. Doplnky pochádzali z rokov 1813 – 1816, 1848, 1850 – 1852.

Vnútorné i vonkajšie členenie katalógu a hlavne jazyk záznamov a typ písem (prevažná väčšina kníh bola písaná v latinčine, druhým najpoužívanejším jazykom bola nemčina) sa stali základnými bodmi na segmentáciu a charakteristiku modelu, ktorý v sebe nesie isté špecifiká. Jeho podstatou bolo hľadanie konsensu medzi antikvou a fraktúrou. Uplatnenie vytvoreného modelu vidíme v prepise ďalších častí dokumentu *Elenchus librorum*, ako i iných rukopisných knižničných historických katalógov.

New possibilities of the digital environment for researching the history of book culture

Mária Bôbová

State Scientific Library in Banská Bystrica, Lazovná 9, 975 58 Banská Bystrica, maria.bobova@svkbb.eu

The paper will present the creation process of a manuscript model using the Transkribus platform, which is part of the Skriptor project of the Slovak Research and Development Agency (SRDA). The model is primarily intended for the historical manuscript catalogue of the church library of the Clerical Seminary of St. Karol Boromejský in Banská Bystrica. The catalog was created at the beginning of the 19th century and was used until the mid-19th century. It is part of the larger manuscript document *Elenchus librorum*, which is currently in the possession of the State Scientific Library in Banská Bystrica and which, in addition to two library catalogues, also includes accession lists and the catalogue of the Seminary Library. The handwritten catalog is an important source for the history of the library of the Clerical Seminary of St. Karol Boromejský (1807 – 1950), because it approximates the first period of its work. There are 3,607 book records listed on thirty-seven sheets. The list of books is divided into ten thematic units (Bible, patrology, dogmatic theology, scholastics, catechetics, polemical theology, moral theology, asceticism, homiletics, liturgy, church and secular history, law, philosophy, pedagogy and others), which also have other internal division. Individual pages in the catalogue are divided into two parts by a central line. One part was used to write records about books, namely the title, author, and place of publication, year and format. The second part was limited to noting changes, whether it was new accessions in the library, sale of some books or to note the books' rearranging in the catalogue. The accessions came from the years 1813 – 1816, 1848, 1850 – 1852.

The internal and external division of the catalogue and especially the language of the records and type of writing (the vast majority of books were written in Latin, the second most used language was German) became the basic points for the segmentation and characterization of the model, which carries certain specifics. Its essence was the search for a consensus between the Latin script and the Fracture. We can see the application of the created model in the transcription of other parts of the *Elenchus librorum* document as well as in other manuscript library historical catalogues.

Tvorba modelu na automatické rozpoznávanie rukopisu J. M. Hurbana v platforme Transkribus (postupy a skúsenosti)

Alica Kurhajcová

Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 974 01 Banská Bystrica, alica.kurhajcová@umb.sk

Výskum kultúrnych dejín 19. storočia a v rámci nich každodenného života a osobných vzťahov sa prakticky nezaobíde bez štúdia ego-dokumentov, medzi ktorými zaujíma osobitné miesto osobná korešpondencia. Predpokladom rýchlej a efektívnej práce s rukopisnými súkromnými dokumentmi (a nielen s nimi) je ich „približný" prepis a možnosť vyhľadávať v ňom kľúčové slová, v ideálnom prípade verejný prístup k presnej transkripcii dokumentov, ktoré môžu byť v budúcnosti upravené do podoby komentovanej pramennej edície. Komplexné riešenie v tomto smere ponúka platforma Transkribus s technológiou na automatické rozpoznanie rukopisných textov. Kým časť korešpondencie slovenských osobností 19. storočia sa pripravila na vydanie už v druhej polovici 20. storočia, tá zvyšná, nevydaná alebo len čiastočne vydaná, medzi ktorou nachádzame aj listy Jozefa Miloslava Hurbana, by mohla byť aj vďaka tomuto digitálnemu nástroju sprístupňovaná a editovaná v najbližších rokoch či desaťročiach. Technológiu strojového čítania, na ktorom sa Transkribus zakladá, sme sa z uvedených dôvodov rozhodli odskúšať a vyhodnotiť na vybranej zbierke listov J. M. Hurbana, evanjelického farára, spisovateľa, literárneho kritika a popredného protagonistu slovenského kultúrneho a národnopolitického života 19. storočia.

Vytypované listy, ktoré v rokoch 1846 – 1887 adresoval J. M. Hurban najužšej rodine, pochádzajú z osobných fondov deponovaných v Literárnom archíve SNK v Martine. Vypovedajú o dobových skutočnostiach v najširšom zmysle slova, pričom precizujú Hurbanovo vnímanie a prežívanie aktuálnej reality, poodhaľujú jeho vzťah s manželkou, deťmi a svokrom a vôbec každodennosť rodiny Hurbanovcov. Jazyk a písmo sa v listoch menia pod vplyvom rôznych okolností (napr. v závislosti od miesta pôsobenia príjemcu, aktuálnych rečových pomerov, pisateľovho zdravotného stavu, veku). Najčastejšie sú písané latinkou, a to raz v dobovej slovenčine, inokedy v (biblickej, resp. slovakizovanej) češtine, z času na čas obohatené o latinské (menej maďarské) výroky či nárečové slová z okolia Myjavy. Do latinkou písaných textov, ktoré sú objektom nášho skúmania, ojedinele vstupuje nemčina písaná kurentom, prípadne ruština a srbčina značené azbukou (s ktorými sme prakticky nepracovali).

V príspevku sa prioritne nevenujeme osobnosti J. M. Hurbana. Na tomto mieste si kladieme za cieľ predstaviť hlavné etapy práce a v rámci nich jednotlivé postupy, ktoré sme uplatnili pri tvorbe špecifického modelu na automatickú transkripciu Hurbanovho rukopisu (*Modelu J. M. Hurban*). Prvú etapu charakterizovala príprava neveľkých vzoriek vlastnoručne prepísaných listov a vyvíjanie prvých menších verzií modelu (trénovaných na 26 až 56 stranách). Po počiatočných neúspešných pokusoch sa nám na konci tejto fázy podarilo znížiť mieru chybovosti prepisu na necelých 8 %. Pri vylepšovaní skúšobného modelu sa nám najviac osvedčila metóda priebežného rozširovania cvičného súboru o automaticky (t. j. na základe aktuálnej verzie modelu) transkribované strany. V druhej etape sme týmto spôsobom vytvorili šesť rôzne veľkých verzií modelu, pričom pri poslednom z nich (trénovanom na 560 stranách, 101 241 slovách) sme dosiahli zníženie miery chybovosti o 2,08 percentuálneho bodu (zo 7,98 % na 5,90 %). Funkčnosť tohto modelu sme overovali v záverečnej etape, a to na nových, dosiaľ netranskribovaných listoch. Chybovosť znakov na 20-stranovej vzorke sa prejavila v rôznej miere – od 2,49 % (vynikajúci prepis) až po 8,70 % (použiteľný), len v jednom prípade 12,38 % (stále zrozumiteľný). Výsledky nateraz vyznievajú optimisticky.

Creating a model for automatic recognition of J.M. Hurban's handwriting in the Transkribus platform (procedures and experiences)

Alica Kurhajcová

Department of History, Faculty of Arts, Matej Bel University in Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, alica.kurhajcová@umb.sk

Research into the cultural history of the 19th century and, within it, everyday life and personal relationships is practically impossible without the study of ego-documents, among which personal correspondence occupies a special place. A prerequisite for fast and efficient work with handwritten private documents (and not only with them) is their "approximate" transcription and the possibility to search for key words in it, ideally public access to the exact transcription of documents that can be edited in the future in the form of an annotated source edition. A comprehensive solution in this regard is offered by the Transkribus platform with technology for automatic recognition of handwritten texts. While part of the correspondence of Slovak personalities of the 19th century was prepared for publication already in the second half of the 20th century, the remaining, unpublished or only partially published texts, among which we also find the letters of Jozef Miloslav Hurban, could also be available and edited thanks to this digital tool in the coming years or decades. For the stated reasons, we decided to test and evaluate the machine reading technology on which Transkribus is based on a selected collection of letters by J. M. Hurban, an evangelical priest, writer, literary critic and a leading protagonist of Slovak cultural and national political life in the 19th century.

Selected letters addressed to Hurban's closest family between 1846 and 1887, come from his personal funds deposited in the Literary Archive of the Slovak National Library in Martin. They help us understand realities of the time in the broadest sense, specifying Hurban's perception and experience of the current reality, revealing his relationship with his wife, children and father-in-law and the everyday life of the Hurban family in general. The language and writing in letters change under the influence of various circumstances (depending on the recipient's place of work, current speech conditions, the writer's health, age, etc.). They are most often written in Latin cursive, sometimes in contemporary Slovak, sometimes in (biblical or Slovakized) Czech, from time to time enriched with Latin (less Hungarian)

sayings or dialect words from the Myjava area. The texts written in Latin script, the object of our investigation, rarely include German written in Kurent, also known as German cursive or Russian and Serbian written in Cyrillic script (which we practically did not work with). In the article, we do not primarily focus on the personality of J.M. Hurban. At this point, we aim to present the main stages of our work and within the stages also the individual procedures we applied in the creation of a specific model for the automatic transcription of Hurban's manuscript *(the J.M. Hurban Model)*. The first stage was characterized by the preparation of small samples of handwritten letters and the development of the first smaller versions of the model (exercised on 26 to 56 pages). After initial unsuccessful attempts, at the end of this phase we managed to reduce the transcription error rate to just fewer than 8%. When improving the test model, the method of continuous expansion of the training file with automatically transcribed pages (i.e. based on the current version of the model) proved to be the most effective for us. In the second stage, we created six different sized versions of the model in this way, while in the last one (exercised on 560 pages, 101,241 words) we achieved 2.08 percentage points reduction in the error rate (from 7.98% to 5.90%)). We verified the functionality of this model in the final stage, namely on new, so far not transcribed sheets. The character error rate on the 20-page sample varied from 2.49% (excellent transcription) to 8.70% (usable transcription), only in one case the error rate was 12.38% (however the transcription was still intelligible). So far, the results sound optimistic.

Transkribus ako nástroj na sprístupnenie dobových archívnych pomôcok na príklade Csákósovho katalógu korešpondencie Kohároyovcov

Imrich Nagy

Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 974 01 Banská Bystrica, imrich.nagy@umb.sk

Metóda automatického rozpoznania rukopisných historických textov (HTR+) pomocou nástroja Transkribus má veľký potenciál pri digitálnom sprístupňovaní dobových archívnych pomôcok, ako sú inventáre, katalógy, registre, súpisy a pod. Túto triviálnu tézu budem analyzovať a dokazovať na príklade číselného katalógu korešpondencie z archívneho fondu Koháry – Coburg, ktorý v rokoch 1944 – 1945 vypracoval bratislavský archivár János József Csákos. Katalóg tvorí obsiahly súbor regestov 6 632 listov na 4 140 stranách formátu A3 v tabuľkovej podobe spísaných moderným kurzívnym písmom. Pre potreby automatickej transkripcie sa použili vyhotovené digitalizáty s rozlíšením 192 dpi. Proces segmentácie (prípravy digitalizátov na automatické rozpoznanie textu) rukopisu je časovo najnáročnejšou fázou automatickej transkripcie. Katalóg síce má jednotnú tabuľkovú úpravu a nástroj Transkribus umožňuje aplikáciu užívateľom preddefinovaných rámcov tabuľky na celý dokument, v praxi je to však realizovateľné iba pri dokumentoch s predtlačenými tabuľkami. Manuálna segmentácia jednej dvojstrany si vyžadovala priemerne 10 minút – pre celý súbor je teda potrebné počítať so 690 hodinami práce. V Transkribuse sa na základe manuálne prepísanej vzorky 29 digitalizátov obsahujúcich 53 strán rukopisu vytrénoval základný model s úrovňou chybovosti v znakoch (CER) 4,11 %. Na základe neho sa s ďalšou vzorkou 28 digitalizátov vytvoril vylepšený model s mierou chybovosti 3,08 %. Ak je dostatočne rozsiahly súbor digitalizátov daného rukopisu, možno dosiahnuť vylepšenie základného modelu pridaním ďalších vzoriek manuálne korigovaných strán. Už pri chybovosti okolo 3 % aplikovaného modelu možno hovoriť o preukázanej úspešnosti a praktickej využiteľnosti Transkribusu a metódy HTR+ na sprístupnenie dobovej archívnej pomôcky: väčšina chýb zaznamenaných pri automatickom prepise sa koncentruje do oblasti interpunkcie a diakritiky, čo nijako významne neovplyvňuje zrozumiteľnosť a praktickú použiteľnosť získaného výstupu digitálneho editovateľného dokumentu v praxi. Samotná aplikácia modelu

pre konkrétny dokument a následné získanie editovateľného textového výstupu (prepisu pôvodného rukopisného dokumentu) nie je užívateľsky ani časovo náročnou činnosťou. Je však potrebné počítať s finančnými nákladmi na automatický prepis rukopisu, ktoré sa pri súčasnom biznis modeli Transkribusu pohybujú približne v rozmedzí 0,18 € – 0,33 € (v závislosti od množstva a spôsobu zakúpenia kreditov pre transkripciu) za spracovanú snímku digitalizátu. V prípade kompletného prepisu Csákósovho katalógu sa teda pohybujeme v cenovom rámci približne 373 až 684 €. Myslím si, že sú tu na mieste konštatácie o hodnote za peniaze.

Transkribus as a tool for making periodical archival aids available using the example of Csákós's catalogue of the Koháry correspondence

Imrich Nagy

Department of History, Faculty of Arts, Matej Bel University in Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, imrich.nagy@umb.sk

The method of automatic recognition of handwritten historical texts (HTR+) using the Transkribus tool has great potential in the digital access of period archival aids such as inventories, catalogues, registers, lists, etc. We will analyse and prove this trivial thesis using the example of the numerical catalogue of correspondence from the Koháry-Coburg archival fund, which was prepared by the Bratislava archivist János József Csákos in 1944 – 1945. The catalogue consists of a comprehensive set of records of 6,632 sheets on 4,140 pages of A3 format in tabular form, written in modern cursive script. For the needs of automatic transcription, digitized images with a resolution of 192 dpi were used. The segmentation process (preparation of digitized data for automatic text recognition) of the manuscript is the most time-consuming phase of automatic transcription. Although the catalogue has a uniform tabular arrangement and the Transkribus tool allows users to apply predefined table frames to the entire document, in practice this is only feasible for documents with pre-printed tables. Manual segmentation of one double page required an average of 10 minutes – so for the entire file, it is necessary to count with 690 hours of work. In Transkribus, a basic model with a character error rate (CER) of 4.11% was trained based on a manually transcribed sample of 29 digitized copies containing 53 manuscript pages. Based on this model, an improved model with an error rate of 3.08% was created with an additional sample of 28 digitized images. If the digitized set of the given manuscript is large enough, the basic model can be improved by adding additional samples of manually corrected pages. Even with an error rate of around 3% of the applied model, we can talk about the proven success and practical usability of Transkribus and the HTR+ method for making historical archival aids available: most of the errors recorded during automatic transcription were in punctuation and diacritics, which do not significantly affect the comprehensibility and practical usability of the obtained output, a digital editable document. The very application of the model for a specific document and the subsequent obtaining of an editable text output

(transcription of the original handwritten document) is neither a user- nor time-consuming activity. However, it is necessary to take into account the financial costs for the automatic transcription of the manuscript, which with the current business model of Transkribus are approximately in the range of € 0.18 – € 0.33 (depending on the amount and method of purchase of transcription credits) per processed digitized image. In the case of a complete copy of Csákós's catalogue, the price ranges from approximately € 373 to € 684. I think the discussion about the value for money would be appropriate.

Zoznam autorov

Mgr. Mária Bôbová, PhD.
Štátna vedecká knižnica v Banskej Bystrici
Lazovná 9, 975 58 Banská Bystrica
maria.bobova@svkbb.eu

Mgr. Martin Katreniak
Novohradské múzeum a galéria
Námestie Kubínyiho 38/3, 984 01 Lučenec
martin.katreniak222@gmail.com

Prof. PhDr. Dušan Katuščák, PhD.
Štátna vedecká knižnica v Banskej Bystrici
Lazovná 9, 975 58 Banská Bystrica
Ústav bohemistiky a knihovnictví, Filozoficko-přírodovědní fakulta,
Slezská univerzita v Opavě
Masarykova třída 343/37, 746 01 Opava, Česká republika
dusankatuscak@gmail.com

Doc. Mgr. Ivona Kollárová, PhD.
Ústredná knižnica SAV, v. v. i.
Klemensova 19, 814 99 Bratislava
ivona.kollarova@savba.sk

Mgr. Ján Kováčik
Slovenská národná knižnica
Nám. J. C. Hronského 1, 036 01 Martin
jan.kovacik@snk.sk

Bc. Klára Kováčová
Ústav bohemistiky a knihovnictví, Filozoficko-přírodovědní fakulta,
Slezská univerzita v Opavě
Masarykova třída 343/37, 746 01 Opava, Česká republika
F190283@fpf.slu.cz

Mgr. Patrik Kunec, PhD.
Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
patrik.kunec@umb.sk

Mgr. Alica Kurhajcová, PhD.
Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
alica.kurhajcová@umb.sk

Doc. Mgr. Peter Labanc, PhD.
Katedra histórie, Filozofická fakulta, Trnavská univerzita v Trnave
Hornopotočná 23, 918 43 Trnava
peter.labanc@truni.sk

Doc. PhDr. Pavol Maliniak, PhD.
Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
pavol.maliniak@umb.sk

Ing. Mgr. Juraj Michelík
Slovenský národný archív, špecializované pracovisko Slovenský banský archív
Radničné námestie 16, 969 01 Banská Štiavnica
juraj.michelik@minv.sk

Mgr. Michaela Mikušková
Univerzitná knižnica Univerzity Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
michaela.mikuskova@umb.sk

Doc. Mgr. Imrich Nagy, PhD.
Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
imrich.nagy@umb.sk

Mgr. Lucia Nižníková
Univerzitná knižnica Univerzity Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
lucia.niznikova@umb.sk

Mgr. Monika Péková
Ministerstvo vnútra Slovenskej republiky, odbor archívov a registratúr
Križkova 7, 811 04 Bratislava
monika.pekova@minv.sk

Dan Ryška
ryska.mail@gmail.com

Zoltán Szatucsek
Magyar Nemzeti Levéltár (National Archives of Hungary)
1014, Budapest Bécsi kapu tér 2-4, Hungary
szatucsek.zoltan@mnl.gov.hu

Prof. PhDr. Juraj Šedivý, MAS, PhD.
Katedra archívnictva a muzeológie, Filozofická fakulta, Univerzita Komenského v Bratislave
Gondova 2, 811 02 Bratislava 1
juraj.sedivy@uniba.sk

PhDr. Oto Tomeček, PhD.
Katedra histórie, Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici
Tajovského 40, 974 01 Banská Bystrica
oto.tomecek@umb.sk

Digital humanities – nástroje sprístupňovania historického dedičstva
Zborník abstraktov

Zborník abstraktov z vedeckej konferencie s medzinárodnou účasťou *Digital humanities – nástroje sprístupňovania historického dedičstva* konanej v Štátnej vedeckej knižnici v Banskej Bystrici v dňoch 12. a 13. októbra 2022.

Odborný garant konferencie:	prof. PhDr. Dušan Katuščák, PhD.
Zostavovatelia:	doc. PhDr. Pavol Maliniak, PhD., doc. Mgr. Imrich Nagy, PhD.
Recenzenti:	prof. Ing. Milan Konvit, PhD., Jan Odstrčilík, Ph.D.
Jazyková korektúra:	PaedDr. Ivan Očenáš, PhD.
Anglický preklad textov:	Mgr. Róbert Címer, Mgr. Zuzana Hušlová, Bc. Katarína Němcová, Bc. Mária Onderufová
Spolupráca:	Mgr. Mária Bôbová, PhD., Ing. Ivana Poláková, PhD.
Grafická úprava:	PaedDr. Dušan Jarina
Vydavateľ:	Štátna vedecká knižnica v Banskej Bystrici, v spolupráci s Filozofickou fakultou Univerzity Mateja Bela v Banskej Bystrici
Tlač:	DALI-BB, s.r.o.
Rok vydania:	2022
Číslo publikácie:	ŠVK BB 738/2022
ISBN	978-80-8227-012-2