Digital humanities a automatická transkripcia rukopisných textov

Prof. PhDr. Dušan Katuščák, PhD

Abstrakt

Autor uvažuje o koncepte *digital humanities*. Poukazuje na naplnenie tohto konceptu na Slovensku od 70. okov 20. st. Autor považuje pojem *digital humanities* za spoločné pomenovanie a prierezovú metodológiu pre všetky aplikácie informačných a komunikačných technológií (IKT) v spoločenských a humanitných vedách, odboroch a disciplínach a im zodpovedajúcej praxi. Jadro štúdie je zamerané na stručnú charakteristiku európskeho výskumného projektu READ = *Recognition and Enrichment of Archival Documents*[1],[2], ktorého riešenie prebiehalo v rokoch 2016-2019 v rámci programu Horizon 2020. Výskumný projekt podliehal priamo Európskej komisii a bol ročne hodnotené nezávislými hodnotiteľmi[3]. Hlavným výstupom projektu je platforma a nástroj *Transkribus*, ktorý predstavuje zásadnú svetovú inováciu zameranú na transkripciu historických rukopisov a dokumentov. Autor, ako jeden z hodnotiteľov projektu READ popisuje svoje skúsenosti a poznatky získané pri experimentálnej transkripcii rukopisných listov Andreja Kmeťa. Vysvetľuje svoj pohľad na *Digital humanities*, ako na metodologický kontext projektu a stručnú charakteristiku procesu skenovania, nahrávania obrazov, segmentácie a automatickej transkripcie ako aj konkrétne príklady automatickej transkripcie rukopisných listov Andrej Kmeťa a výsledky experimentu.

1.1 Úvod

Digital humanities (Digitálne humanitné vedy DH) považujeme za všeobecné pomenovanie oblasti, ktorá je akousi *strechou* pre rozličné oblasti vedeckej a praktickej činnosti zamerané na využívanie *digitálnych technológií* v spoločenských a humanitných vedách. V podstate ide, podľa nášho názoru, o „staré víno v novej fľaši“, pretože digitálne technológie sa využívajú v spoločenských a humanitných vedách v rôznej miere najmä od 70-tych rokov. *Digital humanities* vnímame ako pojem, ktorý má široký rozsah a nejasný obsah. Pokiaľ ide o *rozsah* pojmu, tento sa nevzťahuje nejakú jednotlivú entitu, ale na to, čo je mnohým jednotlivým entitám *spoločné*. Jednotlivým *entitám* je *spoločné* uplatňovanie toho, čo sa tradične pomenúva termínom *informačné a komunikačné technológie* (IKT), či *digitálne technológie*. *Digitálne technológie*[4] sú *obsahom* pojmu *digital humanities*, pretože sú spoločnou črtou všetkých prvkov množiny entít, ktorých sa pojem týka. *Digital humanities* nepovažujeme ani za vedný odbor ani za vedeckú disciplínu. *Digital humanities* predstavujú *prierezovú metodológiu*, ktorá sa v spoločenských a humanitných vedách aplikuje vo výskume, vývoji, manažmente a praxi.

---

[1] https://read.transkribus.eu

[2] **Mühlberger, Günter**. READ (Recognition and Enrichment of Archival Documents) - 2016-2019. [Projektová štúdia]. Dostupné: https://www.academia.edu/22653102/H2020_Project_READ_Recognition_and_Enrichment_of_Archival_Documents_-_2016-2019

[3] **Christophe DOIN**. Project Officer. European Commission. DG CONNECT C1. EUFO 01/150A. Rue Robert Stumper. L-2350 Luxembourg-Ville. Luxembourgh. Christophe.DOIN@ec.europa.eu
Reinhard Altenhöner. Deputy Director General. Staatsbibliothek zu Berlin - Preußischer Kulturbesitz Zentralabteilung SV/Z. Potsdamer Straße 33, 10785 Berlin. E-Mail: reinhard.altenhoener@sbb.spk-berlin.de
Lorna M. Hughes. Professor of Digital Humanities. Head of Subject. Information Studies. 11 University Gardens. University of Glasgow. Glasgow, G12 8QQ. Scotland. E_Mail: Lorna.Hughes@glasgow.ac.uk
Dušan Katuščák. Professor of Library and Information Science. Silesian University in Opava. Faculty of Philosophy and Science. The Institute of the Czech language and Library Science; State Research Library, Banská Bystrica. Dusan.katuscak@fpf.slu.cz

[4] LIS – Library and Information Science / Studies

Niet pochýb o tom, že Slovensko zachytilo v odbore knižničnej a informačnej vedy a praxe (LIS[5]) prvú vlnu aplikácie digitálnych technológií v spoločenských a humanitných oblastiach koncom sedemdesiatych rokov, a, s istým optimizmom a očakávaním kontinuity a novátorstva sa pozeráme aj do nasledujúcich rokov. S viacerými vedcami, odborníkmi a praktikmi sme mali možnosť podieľať sa na využívaní IKT v LIS. Dajú sa zaznamenať určité rané i pokročilé etapy *či* stupne *digital humanities* v oblasti knihovníctva a informačných systémov, teda v oblasti knižničných a informačných systémov a služieb u nás.

Digitálne humanitné vedy už doteraz významne ovplyvnili LIS, ako aj iné humanitné a spoločenské vedy a v budúcnosti ich počet určite porastie.

1. ***Strojové spracovanie.*** *Prvým stupňom* digital humanities v našom odbore je *strojové spracovanie* národnej bibliografie (SNB) (1968-1975), teda pomerne skoro po objavení integrovaného čipu v roku 1970.
2. ***Mechanizácia a automatizácia****. Druhým stupňom* je *mechanizácia a automatizácia* (SNB) (štátny program P13) 1975-1980.
3. ***Integrácia.*** *Za tretí stupeň* digital humanities možno považovať IKIS a CASLIN (Integrovaný kooperačný informačný systém a Česká a slovenská informačná sieť) (program P18) 1985 a n.
4. ***Informatizácia, kooperácia, integrácia.*** *Štvrtým stupňom* digital humanities je program *informatizácie* spoločnosti zahŕňajúci aj spoločenské a humanitné oblasti a konkrétne v našom odbore ide o projekt KIS3G – portál *Slovenská knižnica* (št. program, sprievodné zavádzanie štandardov, internacionalizácia odboru) 1994-2005. Tento projekt sa v európskom kontexte hodnotí ako zatiaľ najvýznamnejšia praktická služba pre manažmentu znalostí[6].
5. ***Scientizácia.*** *Piaty stupeň* je kvalitatívne novým aspektom *digital humanities*. Ide o reálnu *scientizáciu* nášho odboru, rozsiahlu vedeckú kooperáciu a uplatnenie inžinierskeho prístupu (chemici, biológovia, informatici) na riešenie konkrétnych problémov nášho odboru (kyslý papier). Išlo o výskum konzervovania KNIHA SK (deacidifikácia) (štátny program základného výskumu) 2000-2010 a konkrétne aplikačné riešenia závažného odborného a civilizačného problému zániku papierových nosičov informácií z rokov 1830-1990.
6. ***Digitalizácia.*** *Šiestym stupňom* digital humanities je jedinečný projekt digitalizácie – *Digitálna knižnica a digitálny archív (DIKDA)*, ktorý má národný a európsky rozmer a je koncipovaný ako služba pre humanitné a spoločenské vedy vo veľkoryso financovanom Operačnom programe informatizácie spoločnosti OPIS (OPIS2) (európsky a št. Program) v rokoch 2004-2015.
7. ***Postmoderná vedecká komunikácia.*** *Siedmy stupeň* v našom odbore je poznamenaný sprievodnými fenoménmi *digital humanities*, ako napr. BIG DATA, OPEN DATA, OPEN ACCES, OPEN ARCHIVE, LINKED DATA, umelá inteligencia, vizualizácia dát, využívanie digitálneho obsahu, clouding etc.

---

[6] **European Commission**. The factsheets present an overview of the state and progress of eGovernment in European countries. Joinup is a joint initiative by the Directorate General for Informatics (DG DIGIT) and the Directorate General for Communications Networks, Content & Technology (DG CONNECT). Production/Publishing: ISA Editorial Team, Wavestone Luxembourg S.A. May 2018. – Dostupné: https://joinup.ec.europa.eu/sites/default/files/inline-files/eGovernment_in_Slovakia_2018_0.pdf

Zaujímavý metodologický mikrosystém *„digitálnej vedy“* rozvíja J. Steinerová vo svojom koncepte v kontexte *informačnej vedy*[7]. V podstate ide o inováciu prekonaného modelu vedeckej komunikácie zo 60-tych rokov s dôrazom na OPEN SCIENCE, OPEN DATA, OPEN ACCESS. Podľa nej:

> *„digitálna veda znamená nielen návrh zložitých socioekonomických systémov a nástrojov vedeckej komunikácie, ale aj premenu vedy od hierarchickej organizácie poznania smerom k horizontálnym interdisciplinárnym prepojeniam. V nich sa vynárajú nové disciplíny a metódy prostredníctvom otvorenosti zdrojov a nástrojov. Objavujú sa aj nové formy kolaborácie a vedeckej metriky s novými publikačnými kanálmi pri verifikovaní a viacnásobnom využívaní dát a výsledkov. Pritom sa predpokladá otvorenie vedy smerom k participácii budúcich výskumníkov vo vedeckých komunitách.“*(Steinerová, 2014).

Vďaka prierezovej metodológii *digital humanties* doznieva a pomaly sa prekonáva stará paradigma založená v odbore LIS na *kumulácii* a kladie sa dôraz na *využívanie* nahromadených záznamov a najmä poznatkov a dát. Záznamy o dokumentoch a dokumenty v digitálnej forme sa tvoria, uskladňujú v databázach a repozitoch už desiatky rokov. Avšak *využívanie* digitálnych záznamov je nedostatočné (pre vedu, výskum, vzdelávanie, zábavu, priemysel, hospodárstvu, podnikateľov, verejný a privátny sektor).

1.2 Atribúty digital humanities

V *digital hmanities* sa uplatňujú nové spôsoby výskumu a využitia *digitálnych technológií*[8]. Pre výskum v DH je charakteristická:

1. kooperácia bádateľov vo výskumných projektoch,
2. scientizácia v spoločenských a humanitných vied,
3. interdisciplinarita - informatika, chémia, história, ekonómia, medicína, sociológia, pedagogika, psychológia...
4. tímovosť (medziinštitučná, medzištátna, univerzity, knižnice, archívy, galérie, múzeá),
5. výrazné zapojenie IKT vo výskume, vzdelávaní a v sprístupňovaní poznatkov,
6. umelá inteligencia (*Hidden Markov Model (HMM*) - rozpoznávanie reči, rukou písaného písma, gest, bioinformatika.

Pojem *digital humanities* považujeme za spoločné pomenovanie pre všetky aplikácie informačných a komunikačných technológií (IKT) v spoločenských a humanitných vedách, odboroch a disciplínach a im zodpovedajúcej praxi.

V spoločenských a humanitných vedách a praxi sa *využívajú poznatky a nástroje z odborov a disciplín IKT*[9]. Pritom tok poznatkov nie je len jednostranný od IKT

---

[7] **Steinerová, Jela**. 2014. Digitálna veda – východiská, problémy a princípy. In *ITLib*, 2014, č. 1. Dostupné: https://itlib.cvtisr.sk/archiv/2014/1/digitalna-veda-vychodiska-problemy-a-principy.html?page_id=2626

[8] **Digitálne technológie** – a) odvetvie vedeckých alebo inžinierskych poznatkov, ktoré sa zaoberajú tvorbou a praktickým využívaním digitálnych alebo počítačových zariadení, metód, systémov atď. b) pokroky v digitálnej technológii, digitálne zariadenie, metóda, systém atď. vytvorené pomocou týchto znalostí; vynález internetu a ďalších digitálnych technológií, c) uplatňovanie týchto poznatkov na praktické účely, napríklad v oblasti digitálnej komunikácie a sociálnych médií. (Podľa: https://www.dictionary.com/browse/digital-technology)

[9] *Teoretická informatika (aj pre prírodné vedy); Aplikovaná informatika; Softvérové inžinierstvo (aj pre prírodné vedy); Hospodárska informatika; Telekomunikácie; Vojenské komunikačné a informačné systémy; Telekomunikačná technika; Telekomunikačné systémy; Počítačové inžinierstvo; Umelá inteligencia; Informačné systémy; Teória informácie; Riadenie*

k odborom a praxi spoločenských a humanitných vied, pretože aplikácia poznatkov, metód a nástrojov IKT v spoločenských a humanitných vedách vyvoláva spätne požiadavky voči IKT. Príkladom takejto interakcie v odboroch LIS môžu slúžiť požiadavky *na integrované knižnično-informačné systémy, infraštruktúru a workflow digitalizácie, optické rozlišovanie znakov, textové analýzy, nástroje vyhľadávania informácií, dlhodobé archivovanie digitálneho obsahu, formáty dát, databázy* a pod. Ak sa poznatky, metódy a nástroje disciplín IKT využívajú v spoločenskovedných a humanitných odboroch a praxi, možno ich považovať za odbory patriace pod spoločnú strechu či dáždnik odborov *digital hmanities.*

1.3 *Digital humanities* a projekt READ

Projekt READ má všetky atribúty metodológie *digital humanities*. Projekt bežal v rámci programu Horizon 2020. Je to výskumný a inovačný program, číslo zmluvy No 674943. Projekt skončil 30. júna 2019. Záverečné hodnotenie projektu bolo 12.09.2019 v Luxemburgu. Autorom a koordinátorom projektu je prof. Günter Mühlberger (University of Innsbruck, Digitisation and Digital Preservation Group).

Univerzita v Innsbrucku od roku 2016 skúma základné technológie rozpoznávania rukopisu, analyzuje rozloženia a vyhľadávanie kľúčových slov pre historické dokumenty v spolupráci s 13 ďalšími partnermi z Európy. Na všetkých troch oblastiach sa podieľajú výskumné tímy univerzít vo Valencii, Rostocku, Technickej univerzity vo Viedni a ďalšie výskumné inštitúcie zastúpené v projekte READ.

Projekt READ bol financovaný Európskou úniou sumou približne 8,2 milióna EUR. Financovanie sa končí 30. júna 2019. Formujú sa však rôzne nadväzujúce projekty, v ktorých bude pokračovať základný aj aplikovaný výskum. Autor tejto štúdie sa usiluje o zapojenie slovenských a českých inštitúcií do tohto výnimočného vedeckého inovačného úsilia spadajúceho do konceptu *digital humanities*. Technologická a vedecká inovácia projektu READ je založená na využívaní *umelej inteligencie* ako jednej z perspektívnych disciplín *informatiky*.

1.4 Význam platformy Transkribus[10]

V platforme *Transkribus* sa implementujú výsledky základného výskumu. Vytvorenie výskumnej platformy Transkribus bolo okrem základného výskumu jedným z hlavných cieľov projektu READ. Približne 2,5 milióna EUR z vyššie uvedených 8,2 milióna EUR sa investovalo do rozvoja tejto výskumnej infraštruktúry, ktorá postavila digitalizáciu, rozpoznávanie, prepis a vyhľadávanie v historických dokumentoch na technologicky úplne nový základ. Technológia, ktorá je založená na metódach strojového učenia, má mimoriadny význam, pretože:

- archívy, knižnice a múzeá, ktoré chcú zlepšiť prístup k svojim zbierkam,
- vedci humanitných vied, ktorým je umožnené budovať výskum na úplne novom základe („Digitálne humanitné vedy“),
- široká verejnosť, ktorá ťaží z drasticky zlepšeného prístupu k "rodinným údajom" v archívoch, a

procesov; Robotika (aj pre strojárstvo); Kybernetika; Technická kybernetika; Ostatné príbuzné odbory informačných a komunikačných technológií;

[10] **Mühlberger, Günter**. READ. D3.4. READ Platform Business Implementation. Report for Period 3. [*Confidential*]. 05.08.2019. H2020 Project 674943.

- počítačoví vedci a poskytovatelia technológií, ktorí dostávajú veľmi významné súbory údajov pre svoj výskum, a teda im umožňuje vyvíjať vylepšené algoritmy a metódy.

Transkribus má transformačnú silu pre celý proces tvorby hodnoty pri digitalizácii historických dokumentov. Podľa štatistiky NUMERIC (2010) sa v európskych archívoch nachádza 26,98 miliárd strán. Predpokladá sa, že z tohto objemu sa postupne bude digitalizovať asi 10,45 miliárd strán. V slovenských archívoch je odhadom 170 km archiválií. V Českej republike sa už viac ako dvadsať rokov kooperatívnym spôsobom buduje digitálna knižnica rukopisov Manuskriptorium, v ktorej sa nachádza vyše 46 000 plne digitalizovaných dokumentov a asi 400 000 popisných záznamov[11]. V archívoch na jeden meter pripadá asi 7 000 strán. Bolo by ideálne, keby súčasťou digitalizácie mohla byť aj automatická konverzia vybratých archívnych rukopisných, strojopisných a iných materiálov. Preto Transkribus!

1.5 Archívne dedičstvo Slovenska[12]

Archívne dedičstvo Slovenska je v správe 47 štátnych archívov. V roku 2009 predstavoval 27 000 archívnych fondov a archívnych zbierok. Mal celkový rozsah 185 000 bežných metrov (bm), teda 185 kilometrov archívnych dokumentov, 1 480 000 archívnych škatúľ, cca 740 000 000 kusov archívnych dokumentov. Prírastky archívnych dokumentov predstavujú približne 3000 bm/rok.

Povzbudzujúce je, že k archívnemu dedičstvu existuje na určitej úrovni prístup cez archívne pomôcky v elektronickej forme. Podľa stavu pred 10 rokmi bolo asi 4000 archívnych pomôcok, ktoré však boli len vo forme obrázkov (bez možnosti vyhľadávania). V SNK sme na požiadanie Slovenského národného archívu skenovali všetky archívne pomôcky a odovzdali sme ich archívnej správe aj s vykonaným OCR. Predpokladá sa, že budú (už sú?) všetky dostupné na internete.

Z celkového počtu 4000 archívnych pomôcok bolo ca 2800 inventárov vyhotovených prostredníctvom písacieho stroja, 200 inventárov vyhotovených rotaprintom, 650 inventárov vyhotovených v MS Word, 350 inventárov vyhotovených v aplikácii Bach – Inventáre. Z cca 4000 archívnych pomôcok bolo v roku 2010 prístupných 275 (= 7 %).

V rámci výskumu a implementácie platformy Transkribus na Slovensku by bolo vhodné preskúmať, ako táto platforma môže pomôcť sprístupniť všetky archívne pomôcky v digitálnej forme širokej verejnosti. Archívne pomôcky sú *de facto* len indexy ku fondom a zbierkam, podobne ako sú katalógy knižníc pomôckami v prístupe ku knižničným zbierkam a fondom. Bežne sú archívne pomôcky všeobecne dostupné v režime *Creative commons* CC0.

Ďalší výskum by mohol podporiť úsilie archívov o sprístupnenie archívnych dokumentov do roku 1526, sprístupnenie archívnych pomôcok a sprístupnenie matrík, katastrálnych záznamov a pod, nakoľko v súčasnosti obsahuje Elektronický

[11] PSOHLAVEC, Tomáš. Digitální knihovna Manuscriptorium. In: *Libraries V4 in the Decoy of Digital Age. Proceedings of 6th Colloquium of Library and Information Experts of the V4+ Countries held from 31st May – 1st June 2016 in Brno*. – Brno : Moravská zemská knihovna v Brně, 2016. S.(cze) 367-374. – ISBN 978-80-7051-216-6 (brož.)
[12] PÉKOVA', Monika – HANUS, Jozef. 2010. Digitalizácia a sprístupnenie obsahu v štátnych archívoch SR. In: *Konferencia Digitálna knižnica, Jasná pod Chopkom, 2010.*

archív Slovenska len minimálny počet verejne dostupných digitálnych historických dokumentov.

1.6 Unikátne vlastnosti Transkribus

Transkribus je jedinou platformou na svete, ktorá umožňuje aj netechnickým používateľom trénovať špecifické neurónové siete a modely, ktoré sú potom schopné rozoznávať rukopisy a tlače v akomkoľvek jazyku a písme s dobrými alebo veľmi dobrými výsledkami.

Na konci roka riešenia projektu READ bolo v systéme Transkribus 409 344 jedinečných obrázkov strán, ktoré obsahovali asi 40 mil. slov, ktoré vytvorili používatelia systému ako školiace, tréningové údaje. Až do konca projektu bolo užívateľmi vyškolených takmer 3 000 modelov. Doteraz boli modely automatického rozpoznávania vytvorené pre tieto jazyky: *nemčina, fínčina, angličtina, arabčina, švédčina, perzština, holandčina, sýrčina, latinčina, španielčina, macedónčina, ruština, jidiš, francúzština, hebrejčina, francúzština, dánčina, pravoslávna cirkevná ruština, slovanská a srbská cyrilika, bengálčina, taliančina, osmanská turečtina, portugalčina, poľština, nórčina, stará taliančina, gréčtina, stará nórčina, stará španielčina, stredoveká nemčina, stredoveká holandčina, stredoveká francúzština, stredoveká latinčina a slovenčina*.

Pokiaľ ide o *slovenčinu*, tá sa ocitla v zozname v záverečnej správe o projekte READ len vďaka samostatnej a osobnej iniciatívnej práce prof. Dušana Katuščáka a vďaka experimentu popísanom v tejto štúdii. Zo spomínaných 3000 modelov transkripcia bol na Slovensku vytvorený len *jeden* model. Získal ako jeden z 500 klientov povolenie pracovať so systémom Transkribu. Išlo o prácu, ktorej autor venoval asi 1000 hodín a ktorá bola financovaná len z vlastných zdrojov autora. Dosiahnuté výsledky, know-how a skúsenosti nás vedú k úsiliu o to, aby sa revolučný a inovatívny nástroj systému Transkribus zaviedol a na Slovensku do systému vzdelávania a do praxe pamäťových a fondových inštitúcií prostredníctvom projektu výskumu a vývoja.

Zhromažďovanie údajov (teda BIG DATA) je najväčšou hodnotou uchovávania a sprístupňovania písomného dedičstva s pridanou hodnotou, ktorú predstavujú starostlivo prepisované historické dokumenty v štandardných formátoch čo dovoľuje priamo opakovane použiť tieto zbierky pre ďalšie procesy strojového učenia.

Trhové ceny historických skriptov sa pohybujú od 10 EUR až do 30 EUR alebo viac za jednoduchú angličtinu a nemčinu za konkrétny rukopis. Ak predpokladáme 15 EUR za stranu ako priemerné náklady, tak v projekte READ operátori vygenerovali peňažnú hodnotu 4 - 6 miliónov EUR. Je zrejmé, že tieto údaje sú jedným z najdôležitejších kapitálových zásob novozaloženej READ-COOP SCE a pôsobivým potvrdením základnej koncepcie výskumu smerujúcej k novým poznatkom a súčasne komerčnému využitiu nástrojov, ktoré sú výsledkom aplikácie poznatkov.

1.7 Otvorenosť platformy Transkribus

Platforma Transkribus je „otvorená“ pre ľudí i pre stroje:

- Môže ju používať každý, kto si na platforme vytvorí *účet*.
- Kto má vytvorený účet, môže si zadarmo stiahnuť *expertného klienta*, cez ktorého používa platformu

- Všetky *služby* na platforme sú bezplatné.
- Na pripojenie počítačov klientov k platforme je k dispozícii rozhranie API.
- Väčšina softvérových nástrojov sú otvorené zdroje a je možné ich získať alebo stiahnuť prostredníctvom GitHubu.

Obsah, obrázky, súbory, zbierky, vytvorené modely, transkripcie nahrané na platformu sú v predvolenom nastavení *súkromné*, ale to nevyhnutne nie je v rozpore s konceptom „otvorenosti“.

1.8 Potvrdená efektívnosť automatickej transkripcie

Tlačené publikácie zo 16. až 19. storočia sa dajú rozpoznať s mierou chybovosti výrazne nižšou ako jedno percento, jednotlivé rukopisy s 2 až 5% a kolektívne rukopisy s 6 až 10%. Pred niekoľkými rokmi by tieto čísla boli úplne nemysliteľné.

Automatický prepis s plaformou Transkribus poskytuje často takmer bezchybný text. To je však možné iba školením, trénovaním systému a trpezlivým vytvorením modelu pre špecifický rukopis alebo zbierku. Je to tiež jeden z najsilnejších argumentov na používanie platformy, pretože umožňuje každému jednotlivému používateľovi trénovať zodpovedajúce modely presne podľa jeho požiadaviek. V praxi to znamená, že ak máme jednou rukou písaný text vyše 10 000 strán (napr. Laučekova zbierka), *vytrénujeme* model na 50-70 stranách. Potom už ostatné strany dokáže Transkribus automaticky transkribovať so slušnou presnosťou a prinajmenšom podstatne uľahčí editovanie textu, jeho úpravy, preklad, plnotextové vyhľadávanie atd.

1.9 Experiment

O automatickej transkripcii rukopisných textov už desiatky rokov snívajú historici, lingvisti, archivári, knihovníci, dokumentaristi a všetci, ďalší, ktorí prichádzajú do styku s rukopisnými textami[13]. Postupne sa automatický prepis rukopisov stáva skutočnosťou. Je za tým mohutný medzinárodný základný výskum v oblasti umelej inteligencie a tisíce hodín práce.

Signálnu informáciu o práci s platformou Transkribus som zverejnil v jednom blogu a v statuse na Fecebooku. Bol som prekvapený veľkým záujmom o túto prácu. Je to pochopiteľné, pretože mnohí historici, jazykovedci, knihovníci, pedagógovia a i. sú čoraz vzdelanejší v používaní nových technológií vo svojej práci a chápu, že inovácie, ktoré im prácu uľahčia sú veľmi dôležité.

Transkribus, pochopiteľne, nenahrádza odbornú a vedeckú erudíciu historikov a archivárov. Automatická transkripcia je len jedným z krokov vedeckej práce historikov. Ďalej nasleduje historický výskum textu a kontextu transkribovaných textov a informácií, editovanie textov získaných transkripciou, identifikácia entít, kľúčových slov, ktoré sú v texte objavené (dátumy, mená osôb, názvy geografických jednotiek, korporácií a pod.).

Zmyslom rozsiahlejšej transkripcie s použitím špičkovej platformy Transkribus je sprístupnenie unikátnych zbierok, dokumentov, archívnych jednotiek, ktoré sa

[13] V roku 1991 som sa v spolupráci s ing. Jánom Mišíkom pokúšal použiť systém na rozpoznávanie znakov na automatický prepis rukou písaných katalogizačných záznamov zo starého katalógu Slovenskej národnej knižnice (Matice slovenskej). Výsledkom bola účinnosť IRIS OCR transkripcie ca 35/40% a transkripcia bola nepoužiteľná.

nachádzajú v archívoch spravidla len v jednom exemplári. V tom je rozdiel medzi výskytom jednotiek v knižniciach a archívoch. V archívoch sú *jedinečné*, autentické originálne dokumenty, zbierky, archívne jednotky, kým v knižniciach sú *tituly* dokumentov, ktoré majú často stovky až tisíce *exemplárov*.

Po transkripcii historických textov a rukopisov je možné digitálny obsah editovať, interpretovať, použiť a *sprístupniť* na *využitie* v širšom meradle aj vo verejných informačných systémoch a službách. Navyše, transkribovaný originálny text, napríklad v latinčine, maďarčine, nemčine, alebo v inom jazyku je možné aspoň približne ďalej automaticky *preložiť* do iného jazyka. Tým sa dosť podstatne mení charakter práce archivárov a historikov.

Prinášam pre záujemcov výsledky mojej práce[14].

1.10 Čo bolo predmetom experimentu?

Na experiment som vybral zbierku rukopisnej prevažne slovenskej korešpondencie Andreja Kmeťa, uloženej v Knižnici Slovenského národného múzea v Martine, a to po predchádzajúcom láskavom súhlas riaditeľky múzea *dr. Márie Halmovej*.[15] Listy Andreja Kmeťa (SNM, Martin) z rokov 1841-1908. Osobnosťou Andreja Kmeťa, vrátane spracovania častí jeho korešpondencie sa zaoberá systematicky Karol Hollý[16], [17] a uvádza aj ďalšie zdroje, ktoré sa týkajú Kmeťovej rukopisnej pozostalosti.

Pre ďalšie experimenty som skenoval materiály z Archívu rodu ZAY, Bučiansky archív SNA, Laučekova zbierka, 1500-1800) (ca 5000 strán). V budúcnosti máme v pláne skenovať, virtuálne skompletizovať a sprístupniť celú zbierku nevydaných rukopisov Martina Laučeka. Collectaneu využívali historici tak, že citovali alebo prekladali priamo niektoré jej časti. Našim cieľom je naskenovať, transkribovať a umožniť preklad celej zbierky alebo aspoň jej vybratých častí. Celkove má ísť o 22 zväzkov a odhadovaný rozsah je viac ako 10 000 strán. Ide totiž o mimoriadne cennú zbierku najmä pre dejiny evanjelickej cirkvi, ale tiež pre dejiny nášho novoveku. Mimoriadne záslužnú prácu v spracovaní, skenovaní, preklade a vydávaní prameňov k dejinám evanjelickej cirkvi na Slovensku robí už od roku 2004 *Združenie evanjelikov augsburského vyznania Považského seniorátu (ZEAVPS) v Dolnom Srní* a jeho mimoriadne aktívny avšak skromný zberateľ a organizátor aktivít predseda združenia Mgr. Pavel Černaj. P. Černaj zhromaždil aj informácie o Martinovi Laučekovi[18] a jeho rukopisnej zbierke *Collectanea*, opierajúc sa najmä o základnú monografiu Jána Ďuroviča[19].

---

[14] Podrobné inštrukcie pre prácu s platformou Transkribus obsahuje dobrá a dostupná dokumentácia. V tejto štúdii uvádzam len základné informácie a poznatky z konkrétneho experimentu, na ktorý som potreboval ca 1000 hodín, nakoľko som celý systém potreboval naštudovať, zoznámiť sa s architektúrou, dokumentáciou. Nadobudol som skúsenosti, know-how a expertízu, ktorú popisujem len všeobecne.

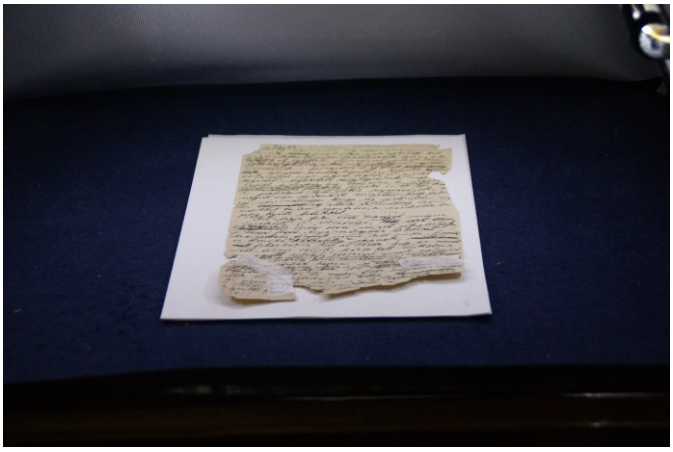
[15] Moja vďaka za to, že ma v priestoroch knižnice strpeli a poskytli mi všestrannú pomoc patrí archivárke Mgr. Viere Varínskej a knihovníčke PhDr. Anne Peťovej. Za pomoc pri zisťovaní informácií o okolnostiach a podmienkach pôsobenia Andreja Kmeťa v Prenčove ďakujem pani Oľge Kuchtovej z Banskej Štiavnice. Za možnosť skenovať v Slovenskom národnom archíve v Bratislave zbierku Martina Laučeka (Collectanea) ďakujem Ústrednej archívnej správe Ministerstva vnútra a za odbornú pomoc PhDr. Eve Kowalskej, DrSc. Z Historického ústavu SAV v Bratislave.

[16] HOLLÝ, Karol: *Andrej Kmeť a slovenské národné hnutie : Sondy do života a kreovanie historickej pamäti do roku 1914.* Bratislava : Veda – Historický ústav SAV, 2015. 279 s. ISBN 978-80-224-1480-7

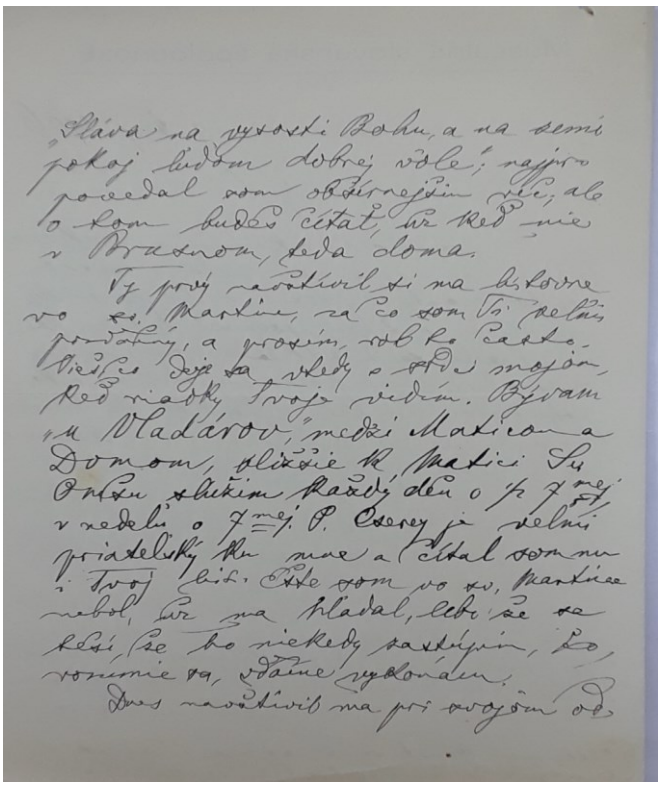
[17] HOLLÝ, Karol: *Veda a slovenské národné hnutie : snahy o organizovanie a inštitucionalizovanie vedy v slovenskom národnom hnutí v dokumentoch 1863-1898.* Bratislava : Historický ústav SAV v Typoset Print s.r.o., 2013.

[18] *Martin Lauček, služobník Slova Božieho Cirkvi augsburského vyznania v Skalici*. Centuria diplomatum et epistolarum Thurzonianarum. Sto Turzovských listov. Diel 1. Ed. Pavel Černaj. Dolné Srnie : ZEAVPS, 2016. 78 s. – ISBN 978-80-89486-13-7

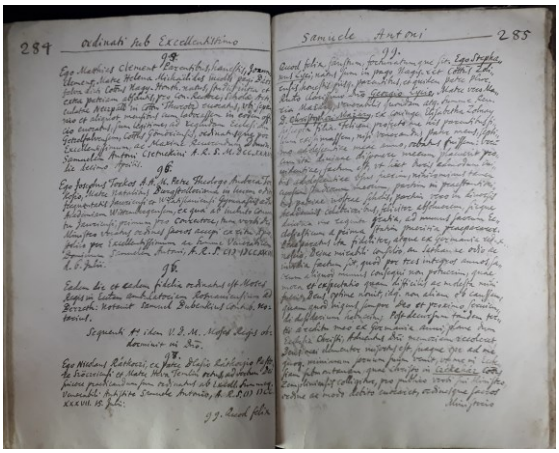
[19] Ďurovič, Ján: Martin Lauček, tolerančný kňaz – spisovateľ. Myjava 1933.



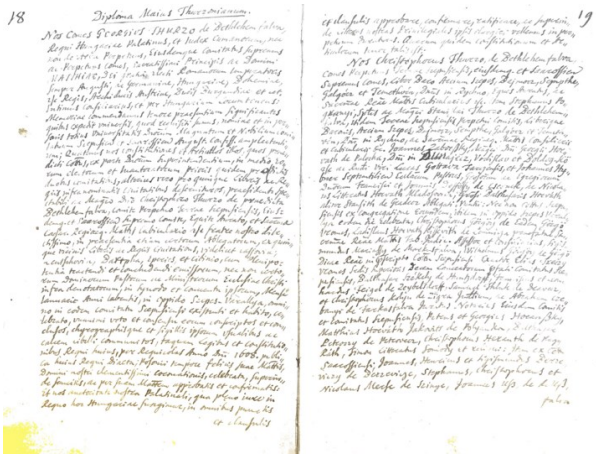
Obrázok 1 Starší rukopisný list Andreja Kmeťa



Obrázok 2 Rukopisný list Andreja Kmeťa



Obrázok 3 Ukážka latinského rukopisu Martina Laučeka. Collectanea zv. 18.



Obrázok 4 Ukážka Laučekovho rukopisu sa týka Juraja Thurzu

1.1.1.1 SKENOVANIE

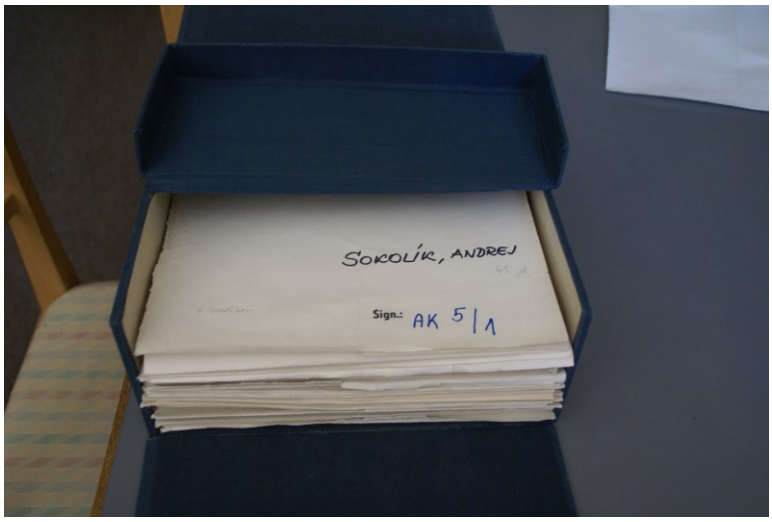
Skenovanie prebehlo 23.-30. 05.2018 v Knižnici SNM. Na skenovanie som použil zariadenie ScanTent (skenovací stan) a aplikáciu DocScan. Toto zariadenie som použil zámerne, aby som overil celý worflow Transkribus, vrátane ponúkaného zariadenia ScanTent a DocScan. Je známe, že mnohé archívy už majú časti zbierok viac-menej kvalitne skenované. Mnou zvolené zariadenia majú význam v prípadoch, ak zbierky ešte nie sú skenované. Takisto je známe, že z bádateľní archívov bežní vedci a používatelia nesmú vynášať archiválie a amatérske fotenie strán mobilmi alebo fotoaparátmi je problematické, ak ide o väčšie súbory (tisíce strán). Preto je ScanTent a DocScan dobrou a dostupnou voľbou, ktorá je s určitými praktickými výhradami (formát, zaostrovanie, kvalita) prijateľná. Treba si však uvedomiť, že v tomto prípade ide o *fotografovanie* a nie o *skenovanie* v pravom technologickom zmysle slova.

Obrázok 5 Skenovací stan ScanTent

Skenoval som kompletný obsah piatich krabíc. Niektoré listy boli na viacerých stranách, tiež neúplné strany, vakáty a pod. Jeden obraz mohol obsahovať aj viac strán rukopisu. Vo fáze skenovania sa vytvárajú obrazy a nie strany, pokiaľ sa strany neskenujú osve. Vhodnejšie je listy skenovať podľa strán, jednotlivo, pretože ak sa skenuje list ako dvojstrana, musí sa prácne usporadúvať poradie strán v postprocesingu.

Čas skenovania bol spolu ca 15-20 hodín. Skenovanie bolo v režime „single“ podľa jednotlivých listov, nie „series“, (s automatickým snímaním po obrátení strany), nakoľko rukopisný materiál je na samostatných listoch rôzneho formátu. Časť materiálu tvoria originály listov, časť fotokópie. Najmä originály listov sú často na krehkom papieri, ktorý by si vyžadoval konzervačné zásahy. Vizitky a podobné menšie formáty papiera – DocScan žiadal „move closer“ asi „priblížiť“, riešil som podložením čistej stránky formátu A4 pod chýbajúce časti listu. Niektoré listy boli poškodené (chýbal roh, poškodené strany listu. Systém v takom prípade hlásil „no page found“. Riešil som to tak, že som podložil bielu stranu ako podložku pod list aj pod chýbajúce časti, potom DocScan zaostril.

Niektoré zložky z 1. krabice som musel skenovať znovu, nakoľko som nevenoval spočiatku potrebnú pozornosť zaostrovaniu. Pri ďalších krabiciach som zaostrovaniu venoval viac pozornosti. DocScan zaostruje na plochu listu na niekoľkých miestach, červené a zelené značky. Keď je zaostrenie uspokojivé, zobrazí „OK“, potom možno stlačiť spúšť. Na skenovanie bol použitý mobilný telefón Samsung Galaxy 6 s operačným programom Android. Nejasný bol pre mňa proces prenosu dát zo Samsungu (Android) do MacBook Air (operačný systém iOS). Napokon som použil počítač s Windows a stiahol som obrázky z Pictures zo Samsungu do iného počítača.



Obrázok 6 Zložky listov Andreja Kmeťa v v archívnej krabici

Systém DocScan je možné pri skenovaní napojiť priamo na server a platformu Transkribus (v Insbrucku či Rostocku) a skenovať priamo do platformy Transkribusu, ktorý zabezpečí experimentálnu transkripciu rukou písaného textu do tlačenej latinky alebo iného písma. Túto možnosť som nevyužil. Niektoré operácie s Transkribus si vyžadovali použitie *Preview, Adobe Acrobat, File Zilla* a i.

Naskenovaný digitálny *obsah* (obrazy) bol:

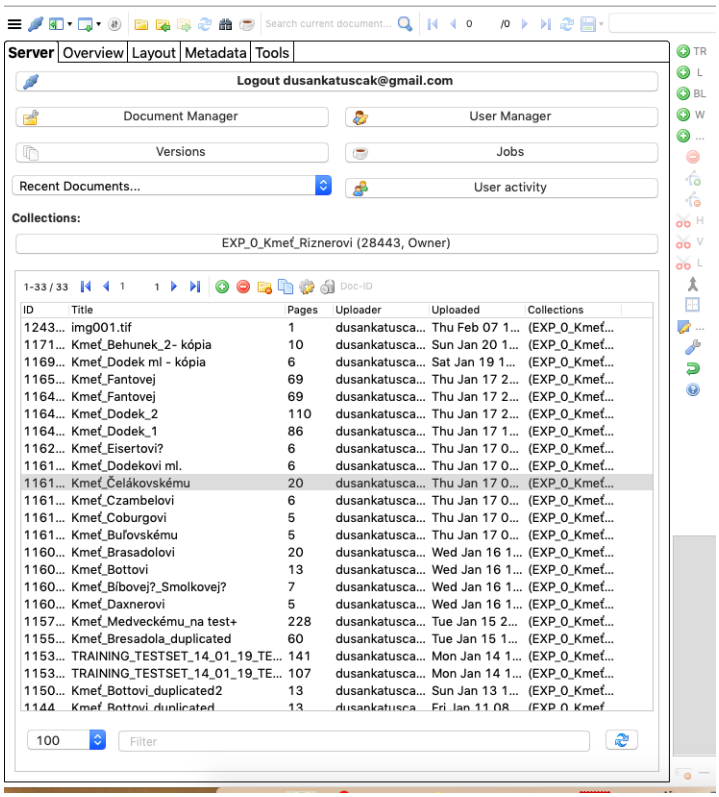
1. pripravený na ďalšie spracovanie v softvéri DocScan (identifikácia obsahu, metadáta)
2. nahratý bez úprav na CD ROM na použitie v SNM podľa uváženia vedenia SNM a Archívu.

3. Obrazy boli pripravené na nahratie do platformy Transkribus a na ďalšie spracovanie v softvéri Transkribus. Nasledovalo *nahrávanie, segmentácia a transkripcia* rukopisného textu.

Digitálny obsah som rozdelil tak, ako sa nachádza v archívnych krabiciach. Napálil som teda 5 kompaktných diskov (CD), ktoré som protokolárne odovzdal riaditeľke SNM EM v Martine dr. Márii Halmovej. Správcovia zbierky teraz môžu použiť digitálny obsah a celý ho zverejniť. Ďalej môžu vložiť do každej krabice jedno CD. Potom môžu rozhodovať tom, komu umožnia prístup na CD alebo opäť umožnia prácu s pomerne krehkými papierovými originálnymi archívnymi listami.

1.1.1.2 NAHRÁVANIE DIGITÁLNYCH OBRAZOV PO SKENOVANÍ

Skenované obrazy je možné spracovať buď lokálne, alebo ich upravovať po *importe* na vzdialený server Transkribus. Pred *importom* na server a pred používaním platformy Transkribus je potrebné zaregistrovať sa, stiahnuť si platformu a vytvoriť si svoju vlastnú privátnu zbierku, ktorá je dostupná výlučne tomu, kto ju vytvoril, ak sa nerozhodne inak. Je možné, aby *transkriber* umožnil prístup k niektorým operáciám napríklad študentom, operátorom, *kooperantom*. Môže umožniť prístup k vlastnej zbierke na prípravu tréningovej vzorky, editáciu po transkripcii a pod. Automatická transkripcia sa vykonáva výlučne na vzdialenom serveri s použitím infraštruktúry Transkribus. Lokálne je možné s vlastnými dokumentami a zbierkami pracovať podľa potreby.



Obrázok 7 Importované súbory (strany, vlastník, dátum, zbierka)

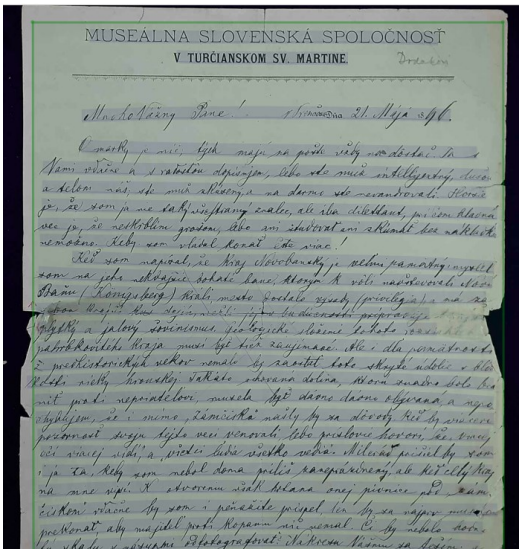
Pred importom je potrebné vytvoriť si vlastnú zbierku (collection), zložku (folder). Nahrávanie, import obrazov jednorazovo je možný do veľkosti 500 MB. Ak je objem importovaných obrazov väčší, obrazy je možné rozdeliť do viacerých súborov

a importovať ich postupne. Väčšie súbory obrazov je možné nahrať, importovať aj s použitím FTP klienta, cez URL alebo DFG Viewer METS. Obrazy sa môžu importovať ako PDF i JPG,TIFF a i. Zbierka importovaných obrazov, vytvorených skenovaním listov Andreja Kmeťa má 11,7 GB v rozlíšení 300 dpi. Nehodnotil som efektívnosť rozlíšenia pri skenovaní vo vzťahu k presnosti automatickej konverzie v Transkribus, hoci, hypoteticky, môže byť tento vzťah významný.

Moje skúsenosti ukazujú, že pred importom je vhodné skontrolovať digitálne obrazy, ich kvalitu, ostrosť, úplnosť, orientáciu strán a pod. Po určitých skúsenostiach som importoval súbory vo formáte PDF.

1.1.1.3 SEGMENTÁCIA

Po importe súborov na server sa musí vykonať na serveri automatická *segmentácia*. Pri segmentácii textu a obrazov musí byť klient pripojený na aplikáciu na serveri. *Segmentácia* znamená, že sa obraz rukopisného textu dokumentu, ktorý je zatiaľ na serveri ako obraz, rozdelí automaticky na bloky, oblasti, riadky textu. Ak je to potrebné, môžu sa urobiť manuálne korekcie. Ide pritom napríklad o spájane a rozdeľovanie blokov, rozširovanie ohraničenia segmentu a pod.



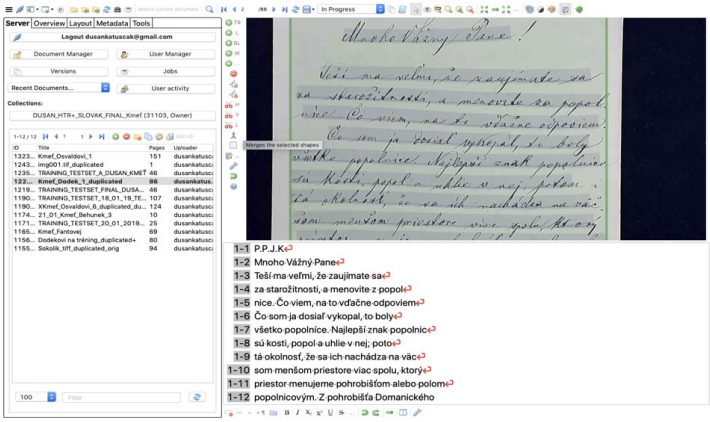
Obrázok 8 Ukážka segmentovaného textu zo zbierky listov Andreja Kmeťa. Označený je blok textu strany a riadky.

1.1.1.4 TRÉNING STROJA HTR[20]

Z importovanej zbierky sa podľa určitého algoritmu vyberie vzorka strán (*dataset*) ktorá slúži na tréning a vytvorenie *modelu* pre určitý typ rukopisu. Na to je potrebné ukázať stroju *správne príklady* textu. Stroj sa v podľa tréningovej sady naučí vzory písma a slov. Ak je zbierka textov od viacerých rúk, je potrebné vybrať primeranú veľkosť testovacej vzorky. Výber strán je možné urobiť aj automaticky tak, aby bola vzorka pripravená podľa určitých ca 20 000 slov. Tréningový dataset sa tvorí priamo v editore klienta Transkribus jednak lokálne, ako aj na serveri. V podstate je potrebné pozorne a veľmi presne prepísať rukopis v editore podľa riadkov, nič neopravovať.

[20] HTR = Historical Text Recognition. Ide o rozpoznávanie textov historických listov, pohľadníc, rukopisov a stredovekých dokumentov. Stroj **HTR engine** z Computational Intelligence Technology Lab (CITlab).

Text prepisovať podľa súdobého jazykového úzu a gramatiky, aj s chybami a podľa ďalších inštrukcií a návodov, ktoré sú k tejto operácii k dispozícii. Poradie častí textu, tagovanie, výber a redakciu kľúčových slov, deskriptívne metadáta a pod. určuje autor transkripcie a tvorca modelu transkripcie. Výsledok transkripcie je potom viditeľný a zhodnotený na *testovacom* datasete. Ak je výsledok uspokojivý, možno automaticky transkribovať ďalšie súbory alebo celú zbierku.

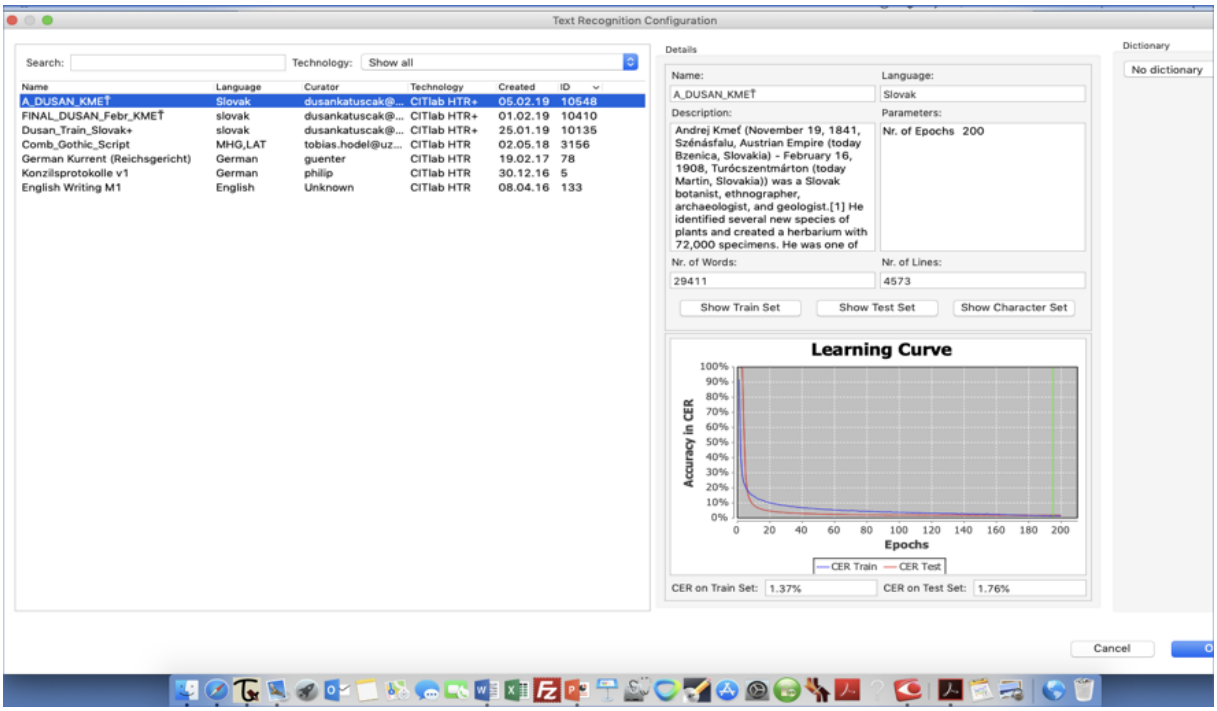


Obrázok 9 Príklad z editovania strany pre tréningový set

1.1.1.5 AUTOMATICKÁ TRANSKRIPCIA

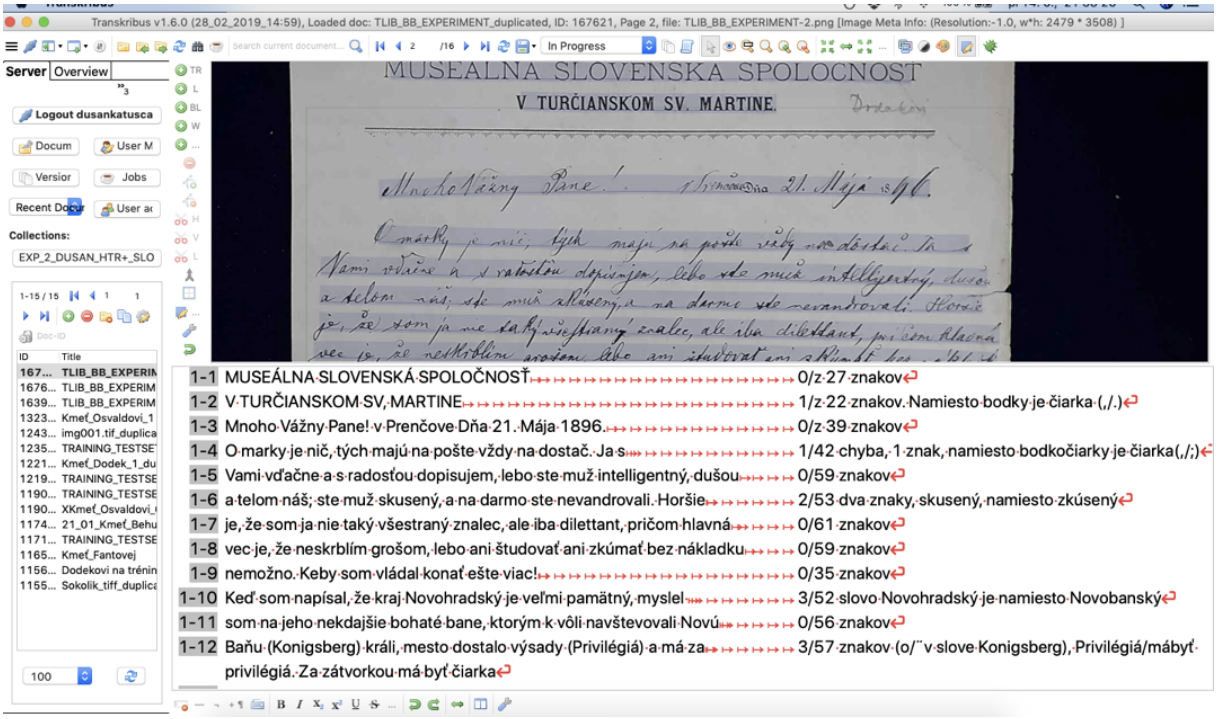
Automatická transkripcia slúži ako základ pre vedecké editovanie, v ktorom je možné text korigovať, explicitne pridávať ďalšie dáta, kontextové informácie, dešifrovanie dát, tagovať, dávať poznámky, metadáta, anotácie, opravy diakritiky, skratky, mále a veľké písmená, paleografické spracovanie, ligatúry a pod.

Automatickú transkripciu som urobil po spustení tréningu a testovania. Použil som vlastný model transkripcie a spustil som transkripciu s použitím HTR+.



Obrázok 10 Obrazovka s údajmi po automatickej konverzii s použitím vlastného modelu A_DUSAN_KMET.

Výsledkom učenia v automatickej transkripcii textu rukopisu Andreja Kmeťa je 1,37% v tréningovom datasete a 1,76% v testovacom datasete (CER – Character Error Rates). Tréningový set obsahoval 29 411 slov a 4 573 riadkov. Model možno nasadiť na celú zbierku.



Obrázok 11 Ukážka výsledku „surovej“ automatickej transkripcie. Pri jednotlivých riadkoch je uvedená chybovosť, ktorú som pridal sám.

Date of training	Transkribus Training method	Training set size		Test set size		CER	CER	Model
RRRRMMDD	HTR/HTR+	pages	words	pages	words	training	test	Name in Transkribus
20181228	CITlabHTR	40	7685	29	5119	6,73%	22,81%	EXP_Dusan_Kmet
20181231	CITlabHTR	74	11458	63	11260	7,95%	17,02%	X_TEST_Slovak_Kmeť
20190110	CITlabHTR	101	16700	87	15171	9,28%	12,25%	10_01_19_XXTEST_Slovak
20190111	CITlabHT+	108	18259	80	15419	1,71%	7,50%	11_01_19_KMET_Slovak_HTR+
20190115	CITlabHT+	193	30377	99	15495	1,5%	3,9%	15_01_192TESTXX
20190118	CITlabHT+		21386			1,04%	2,92%	18_01_TEST_AAA
20190120	CITlabHT+		20254			1,01%	2,78%	20_01_2019_TEST_Január
20190205	CITlabHT+		29411			1,37%	1,76%	A_DUSAN_KMET

Obrázok 12 Prehľad mojich pokusov a omylov z práce s platformou Transkribus (od chybovosti 22,81% ku chybovosti 1,76%). Efektívnosť transkripcie sa výrazne zlepšila, keď mi prof. Muehlberger HTR+.

1.11 Budúcnosť Transkribus

Projekt končí 30. júna 2019. Odborníci a inštitúcie majú záujem o pokračovanie a vývoj služby Transkribus. V súčasnosti (2019) je viac ako 20 000 používateľov Transkribus. Pokračovanie: výskum a implementácie výsledkov sa predpokladajú v rámci projektu EU NewsEye (https://www.newseye.eu/project/about/). Vzniká READ-

COOP (**S**ocietas **C**ooperativa **E**uropeae - SCE). Dňa 1. júla 2019 sa projekt READ mení na **Európsku družstevnú spoločnosť** (SCE). Družstvo READ-COOP bude slúžiť na udržanie a ďalší rozvoj platformy Transkribus a súvisiacich služieb a nástrojov[21].

1.12 Možné ciele pokračujúceho výskumu

V ďalšom výskume by bolo vhodné zamerať pozornosť na tieto oblasti:

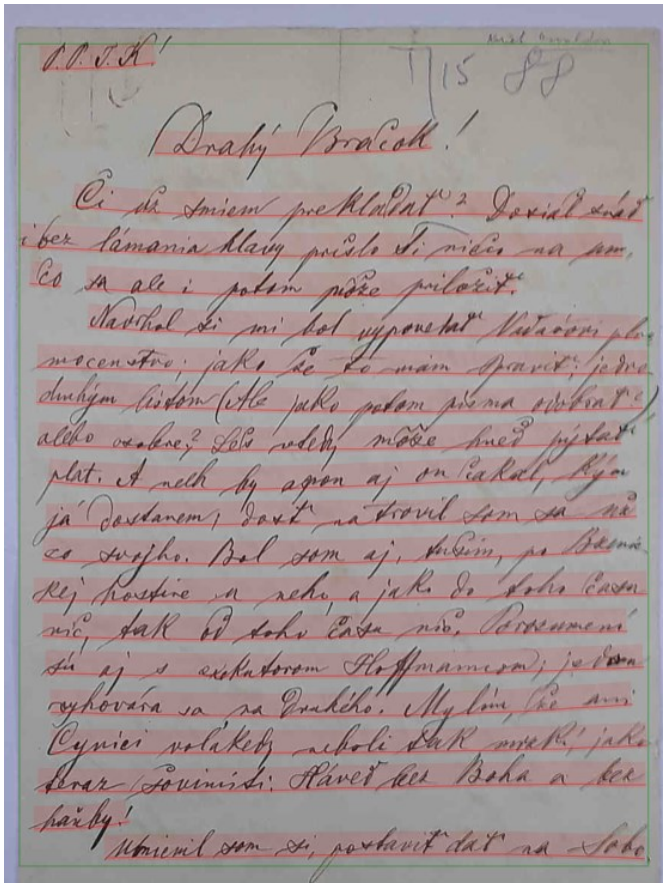
Hlavný cieľ ďalšieho výskumu:

Implementovať na Slovensku najnovšie poznatky z výskumu automatického rozpoznávania textov historických dokumentov

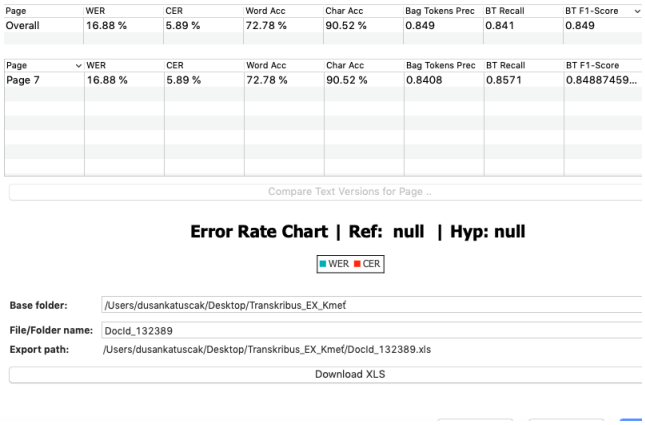
Procesy, ktoré budú viesť k dosiahnutiu hlavného cieľa:

a) výber a štandardný popis rozsiahlejších rukopisných zbierok európskeho a národného významu
b) digitalizácia vybratých historických dokumentov podľa plánu experimentov s cieľom potvrdiť alebo zlepšiť doteraz známe postupy a hodnoty vzhľadom na nasledujúci proces segmentácie textu a automatickú transkripciu (korelácia medzi rôznymi podmienkami a kvalitou skenovania a transkripciou)
c) zdieľanie digitálnych dokumentov s archívmi a inými inštitúciami, ktoré ich budú môcť používať podľa vlastnej úvahy ako náhradu papierových dokumentov
d) tvorba modelov, tréning a analýza modelov automatickej transkripcie podľa novovekých a moderných zbierok a jazykov (najmä slovenčina, čeština, maďarčina, latinčina, nemčina, poľština),
e) overenie a zhodnotenie použiteľnosti dostupných modelov transkripcie z výskumu v projekte READ
f) zoznámenie sa s najlepšou praxou automatického rozpoznávania textov historických dokumentov v Európe, najmä v Nemecku, Rakúsku, Španielsku, Maďarsku, Veľkej Británii, Fínsku, Holandsku, Srbsku, využitie informácií a skúseností na Slovensku
g) automatická transkripcia podstatnej časti rukopisnej Laučekovej zbierky a jej virtualizácie, teda virtuálna jedna digitálna prezentácie zväzkov, ktoré sa nachádzajú na geograficky rozličných miestach (SNA, SNK, UK, Maďarsko)
h) výskum možností zvýšenie efektívnosti rozpoznávania rukopisných textov a textov historických dokumentov prostredníctvom systému Transkribus a súvisiacich nástrojov,
i) sprístupnenie transkribovaných a interpretovaných zbierok cez digitálny repozitár širokej verejnosti,
j) tvorba dokumentácie, ktorá bude slúžiť pre archívy, knižnice, akademické pracoviská ako aj fyzické osoby na automatickú transkripciu textov

[21] Predpokladáme, že vďaka ústretovosti a porozumeniu Univerzity Mateja Bela v Banskej Bystrici a doc. Imricha Nagya, PhD z Katedry histórie, budeme môcť pokračovať v tejto zaujímavej výskumnej inovatívnej práci v rámci nejakého projektu.



Obrázok 13 Experimentálna automatická transkripcia tejto strany



Page	WER	CER	Word Acc	Char Acc	Bag Tokens Prec	BT Recall	BT F1-Score
Overall	16.88 %	5.89 %	72.78 %	90.52 %	0.849	0.841	0.849

Page	WER	CER	Word Acc	Char Acc	Bag Tokens Prec	BT Recall	BT F1-Score
Page 7	16.88 %	5.89 %	72.78 %	90.52 %	0.8408	0.8571	0.84887459...

Obrázok 14 Výpis efektívnosti a chybovosti automatickej transkripcie

1.13 Záver. Efektívnosť platformy Transkribus

Naše skúsenosti overené experimentom potvrdzujú, že jednotlivé rukopisy možno automaticky transkribovať, pričom chybovosť môže byť 2 až 5%, kolektívne rukopisy(zbierky) majú 6 až 10%. Výsledky transkripcie sú čitateľné, použiteľné a možno ich exportovať (DOC, TXT, PDF, TEI, METS atd), editovať, redigovať, korigovať. V experimente sme dosiahli chybovosť (CER) 1,76%.

Z hľadiska vnímania, porozumenia a použitia transkribovaného textu vo všeobecnosti podľa autorov Transkribus platí, že: a) ak sa striktne počíta chybovosť “slov” a ak chybovosť slov je 30%, tak text je pre človeka ešte pochopiteľný a použiteľný, b) ak

sa striktne počíta chybovosť “znakov” a ak chybovosť znakov je 15%, tak text je ešte pre človeka pochopiteľný a použiteľný.

V experimente som „dosiahol“ chybovosť *slov* 16,88% (z 30% prijateľných).
V experimente som „dosiahol“ chybovosť *znakov* 5,89% (z prijateľných 15%).
Presnosť transkripcie slov na hodnotenej strane bola 72,78%.
Presnosť znakov na tejto strane bola 90,52%.

Platforma Transkribus je skvelou pomôckou pre svedomitých a trpezlivých bádateľov, ktorým podstatne uľahčí *doladenie* transkripcie. Platforma nie je, a sotva niekedy bude, určená pre „klikavcov“, teda používateľov, ktorí sú zvyknutí viac „klikať“ ako inovovať.