

OBJEKTOVÝ PRÍSTUP K PREZENTÁCIÍ KULTÚRNEHO DEDIČSTVA V DIGITÁLNO M PRIESTORE

Vplyv typov písma na výsledok procesu digitalizácie

Mgr. Lenka Plevková, PhD.

Abstrakt

Práca sa zameriava na rozpoznanie typov písma v historických slovacikálnych tlačených dokumentoch od 16. storočia do roku 1914. Cieľom výskumu bolo na základe jednotlivých výstupov zo skenerov navrhnúť optimálne pravidlá digitalizovania historických slovacikálnych dokumentov pre vytvorenie čo najkvalitnejšieho a použiteľného digitálneho obsahu v rámci historických knižníc. Pre prehľadnosť a využiteľnosť daných poznatkov bol vytvorený katalóg, ktorý je súborom typov písiem, ktoré sa používali v našom geografickom priestore viac ako štyri storočia a nachádzajú sa v zbierke Tranovského knižnice. Práca sa venuje analýze typov písma, ktoré sa vyskytovali v časovom období 1477 – 1914 v slovacikálnych tlačených dokumentoch. Popisuje vývoj tlačiarskych aktivít na Slovensku. Zameriava sa na proces digitalizácie ako nevyhnutnej súčasti vlastného výskumu a následne samotný softvér OCR.

Kľúčové slová: digitalizácia, digitálny obsah, typológia písma, tlačiarne a tlačiar, historické slovacikálne dokumenty

Abstract

The thesis focuses on the typology of the font in the historical Slovakia-related printed documents from 16th century until 1914. The objective of research is the basis of outputs from scanners to design optimal rules of the digitalization the old Slovakia-related documents to create the highest quality and usable digital content within historical libraries. For transparency and usability of the knowledge it was created a catalog that sets all the fonts that are used in our geographic area for more than four centuries and they are situated in the collection of Tranovsky library. This thesis is about the types of fonts that appeared in documents in the 1477 – 1914 in Slovakia-related printed documents. It describes the development of printing activities in Slovakia. It focuses on the process of digitization as an essential part of own research and then OCR software as well.

Key words: Digitalization. Digital content. Typology of font. Printer and printers. Historical Slovakia-related documents

Úvod

V súčasnej dobe, keď je pojem digitalizácia taký aktuálny, sa zameriavame na uchovávanie nášho kultúrneho a písomného dedičstva. Nejde len o písomné zbierky, dokumenty a fondy, ale aj o kultúrne pamiatky, fotografie, filmy, dramatické umenie, tradičnú ľudovú hudbu a

kultúru... Všetky tieto diela majú svoje zastúpenie pri vytváraní spoločenských hodnôt spoločnosti.

Obsah knižníc v minulosti i prítomnosti sa rozhodujúcou mierou zúčastňoval a zúčastňuje na výmene informácií a výstavbe duchovných hodnôt spoločnosti. Každá národná kultúra sa usiluje primeraným spôsobom zachovať, interpretovať a využívať písomné kultúrne dedičstvo zhromaždené v knižniciach, archívoch, múzeách alebo u jednotlivcov. Súbor tohto písomného kultúrneho dedičstva nazývame národným knižničným fondom, ktorý je umiestnený v rôznom geografickom a inštitucionálnom priestore. Vytvára archív najdôležitejších informačných a dokumentových hodnôt. Ich výskum a poznanie prináša autentické svedectvo o dejinách národa, hospodárskom, politickom, kultúrnom a spoločenskom živote rôznych vrstiev spoločnosti. Historickú knižnicu nemôžeme považovať len za súbor uložených dokumentov, ale v prvom rade za komunikáciu hodnôt, ich recepciu, výmenu a obohatenie. Historické knižnice disponujú nesmiernym informačným bohatstvom, ktoré má v prípade použitia adekvátnych riešení predpoklad vytvoriť digitálny obsah vysokej intelektuálnej kvality. Tematická rôznorodosť predstavuje špecifický obsah určený špecifickým skupinám používateľov.

Predmetom výskumu v rámci doktorandského štúdia je písomné kultúrne dedičstvo – primárne historické tlačene slovacikálne dokumenty, teda dokumenty, ktoré sa podľa autorstva, jazyka, miesta vydania alebo obsahu týkajú Slovenska alebo Slovákov a boli vydané alebo vytlačené od roku 1544 do roku 1918. Zameriava sa na typológiu (vlastnosti) písma daných historických dokumentov. Typológia písma má zásadný význam pre efektívnosť a kvalitu vytváraného digitálneho obsahu. Vlastnosti spomínaného objektu (písma) sú kľúčové prevažne u starých tlačí. V novších dokumentoch (po roku 1918) by písmo a jazyk nemali predstavovať veľký problém pri vytváraní digitálneho obsahu.

Základným cieľom výskumu bolo na základe jednotlivých výstupov zo skenerov navrhnúť optimálne pravidlá digitalizovania historických tlačených slovacikálnych dokumentov pre vytvorenie čo najkvalitnejšieho a použiteľného digitálneho obsahu v rámci historických knižníc, čo by znamenalo možnosť identifikovať zatiaľ neidentifikovateľné tlačoviny. Prostredníctvom analýzy dostupných tlačí a následnej digitalizácie vybraných dokumentov boli vytvorené vzorky zdigitalizovaných tlačí obsahujúcich kompletné abecedy a následné experimentovanie s možnosťami nastavení jazyka a používateľského vzoru písma programu OCR. Na základe porovnania vzniknutých výstupov boli navrhnuté optimálne pravidlá pre vznik plne rozpoznateľných textov, ktoré je možné aplikovať na optické rozpoznanie znakov historických tlačí v slovenskom jazyku v programe ABBYY FineReader pre potreby prepisu historického písma slovacikálnych dokumentov do elektronickej podoby, ktoré sú skompletizované v katalógu písom. Pre dosiahnutie stanoveného cieľa bolo potrebné určiť parciálne ciele: Vytvorenie modelovej knižnice, ktorá v sebe obsahovala časť zbierky historických tlačených dokumentov z Tranovského knižnice (Tranoscius) od roku 1582 do roku 1914, v čase výskumu uchovávané v depozitári knižnice projektu „Pamäť Slovenska – národné centrum excelentnosti výskumu, ochrany a sprístupňovania kultúrneho a vedeckého dedičstva“ pre potreby digitalizácie historickej knižnice analyzované a následne zaradené podľa stanovenej periodizácie slovenského kníhtlačiarskeho.

Výsledkom výskumu popísaného v dizertačnej práci je vytvorenie katalógu, ako súboru všetkých typov písmen, ktoré sa používali v našom geografickom priestore počas rokov 1582 – 1914 a sú obsiahnuté v spomínaných historických dokumentoch z Tranovského knižnice zaradených v modelovej knižnici vytvorenej pre potreby výskumu, rovnako ako vytvorenie odporúčaných nastavení pre optimálnu tvorbu digitálnych obsahov historických slovacikálnych dokumentov.

V práci sme sa zamerali na analýzu teoretických východísk a na rozbor súčasného stavu danej problematiky. Ohľadom histórie písma na našom území vyšlo niekoľko základných diel. Upriamili sme sa predovšetkým na diela P. Voita – Bibliotheca Strahoviensis, J. Špetka - Dejiny písma a knihy. Knižná kultúra a písmo a A. Húščavu Dejiny a vývoj nášho písma. V teoretickej časti práce je pozornosť upriamená na charakteristiku písma a jeho typov vyskytujúcich sa na našom území počas rokov nášho záujmu. Okrajovo sme sa zamerali aj na analýzu vývoja kníhtlačiarstva na Slovensku.

1 Ciele a metódy výskumu v rámci doktorandského štúdia

Predmetom výskumu je písomné kultúrne dedičstvo – primárne historické tlačené slovacikálne dokumenty. Zameriava sa na typológiu (vlastnosti) písma daných historických dokumentov.

„Písomné kultúrne a vedecké dedičstvo je archívna a bibliografická, umelecko-historická a vedecká hnutelná hmotná pamäť ľudského spoločenstva zaznamenaná v písomnej forme, ktorej existencia vyžaduje ochranu pred zánikom, a ktorej účel sa plní v interakcii s človekom – používateľom.“ (Katuščák 2012).

1.1 Ciele výskumu

Prioritným cieľom výskumu v rámci dizertačnej práce bolo na základe jednotlivých výstupov zo skenerov navrhnuť optimálne pravidlá digitalizovania starých tlačných slovacikálnych dokumentov pre vytvorenie čo najkvalitnejšieho a použiteľného digitálneho obsahu v rámci historických knižníc, čo by znamenalo možnosť identifikovať zatiaľ neidentifikovateľné tlačoviny.

Cieľom práce bolo prostredníctvom analýzy dostupných tlačí a následnej digitalizácie vybraných dokumentov vytvoriť vzorky zdigitalizovaných tlačí obsahujúcich kompletne abecedy a následné experimentovanie s možnosťami nastavení jazyka a používateľského vzoru písma programu OCR. Na základe porovnania vzniknutých výstupov bol navrhnutý optimálny postup pre vznik plne rozpoznateľných textov, ktorý je možné aplikovať na optické rozpoznanie znakov historických tlačí v slovenskom jazyku v programe ABBYY FineReader, pre potreby prepisu historického písma slovacikálnych dokumentov do elektronickej podoby.

Čiastkové ciele:

1. Vytvorenie modelovej knižnice, ktorá v sebe obsahuje dostupné knihy z Tranovského knižnice vydané od roku 1582 do roku 1914.
2. Vytvorenie katalógu pre prehľadnosť a využiteľnosť daných poznatkov, ktorý je súborom všetkých typov písiem, ktoré sa používali v našom geografickom priestore počas rokov 1582 - 1914. Súbor typov písma – katalóg typov písma obsahuje všetky typy písiem, s ktorými sme sa počas výskumu stretli a následný popis optimálneho nastavenia skenerov pri tvorbe digitálneho obsahu tlačovín využívajúcich dané typy písiem. Okrem tlačného katalógu písiem bolo súčasťou druhého čiastkového cieľa:
 - vytvorenie súboru obsahujúceho všetky typy písmen, ktoré sú uvedené v tlačnom katalógu,
 - slovník najčastejšie sa vyskytujúcich slov v daných dokumentoch,
 - všetky potrebné nastavenia pre program ABBYY FineReader 12.
3. Vytvorenie kompletnej zdigitalizovanej knihy z modelovej knižnice s vytvorením plne rozpoznateľného digitálneho textu.

1.2 Metódy výskumu a výskumné otázky

Postup riešenia problému dizertačnej práce bol rozdelený do dvoch základných etáp. V každej etape boli stanovené výskumné otázky a následne metodika výskumu a samotný postup riešenia problému.

Prvá etapa výskumu bola zameraná na vytvorenie modelovej knižnice tlačенých slovacikálnych dokumentov od roku 1582 do roku 1914.

Použité vedecké metódy:

- Obsahová analýza a následná syntéza výstupov:
 - Bibliografia tlačí vydaných na Slovensku do roku 1700, 1. – 2. diel (Čaplovič 1984).
 - Bibliografia územne slovacikálnych tlačí 18. storočia, zväzky 1 – 6 (Klimeková et al. 2008a 2008b 2008c 2008d 2008e 2008f).
- Komparácia výstupov na vytvorenie modelovej knižnice podľa periodizácie.
- Zámerný výber výskumného súboru.

Druhá etapa pozostávala z experimentu na konkrétnych exemplároch z modelovej knižnice, vytvorenej v prvej etape.

Použitá vedecká metóda:

- Experiment.

V prvej etape výskumu bolo stanoveným cieľom vytvoriť modelovú knižnicu – zoznam tlačí, ktoré sme následne podrobili výskumu. Jednotlivé kroky experimentu boli nasledovné:

1. Vytvorenie digitálnych vzoriek dokumentov pozostávajúcich z niekoľkých zámerne vybraných strán. Zámerom bolo, aby vzorka obsahovala kompletnú abecedu. Digitálne vzorky boli vytvárané na ručnom skeneri BookEye a automatickom skeneri Treventus v rozlíšení 150, 300 a 600 dpi. Predpokladali sme, že mnohé dokumenty nebude možné skenovať aj prostredníctvom automatického skenera z dôvodu ich defektného stavu a možnosti ďalšieho poškodenia, čo sa nám počas experimentu potvrdilo. V takom prípade boli vytvorené len vzorky dokumentu na ručnom skeneri.
2. Po vytvorení kompletných digitálnych vzoriek sme sa zamerali na analýzu vzniknutých skenov – porovnanie rozdielov a kvality skenovania pri rôznych nastaveniach rozlíšenia (150, 300 a 600 dpi) a rôznych typoch skenovacích zariadení (pozri podkapitola 3.4.2).
3. Pripravené skeny boli následne upravené pre potreby tvorby OCR.
4. Upravené skeny sme následne spracovali prostredníctvom programu OCR, ktorého výstupy nám poslúžili ako vstupy pre ďalšie skúmanie. Pre potreby výskumu sme použili softvér ABBYY FineReader 12.
5. Ďalšia etapa spočívala v analýze výsledkov vzniknutého digitálneho obsahu dokumentu a vymedzenie rozpoznaných a nerozpoznaných znakov.
6. Po analýze výsledkov sme sa zamerali na samotné experimentovanie pre najefektívnejšie nastavenie OCR pri tvorbe digitálneho obsahu.
7. Analýza vzniknutých meraní.

8. Vytvorenie elektronického súboru na podporu softvéru ABBYY FineReader pre rozpoznávanie historických tlačí písaných v slovenskom jazyku, z ktorého bol následne vygenerovaný katalóg písmen a znakov a slovník vyskytujúcich sa slov v nami skúmaných dokumentoch.

2 Výskum v rámci doktorandského štúdia

Kapitola sa zameriava na analýzu problematiky digitalizovania historických tlačených dokumentov a vytvárania plnohodnotných elektronických textov. Popisuje komplexný postup riešenia danej problematiky, a možnosti realizácie ohľadom na súčasné alternatívy v samotnom procese tvorby digitálnych obsahov. Podrobne charakterizuje jednotlivé etapy výskumu s vytýčením čiastkových cieľov a postupmi ich plnenia.

2.1 Prvá etapa výskumu: Vznik modelovej knižnice

Aby bolo možné vytvorenie modelovej knižnice, bolo potrebné stanoviť základné kritériá, ktoré bude nami vytvorená modelová knižnica obsahovať. V prvom kroku bolo nevyhnutné zaoberať sa obsahovou analýzou a následne syntézou výstupov z knižných zdrojov Bibliografia tlačí vydaných na Slovensku do roku 1700, 1. – 2. diel (Čaplovič 1984) a Bibliografia územne slovacikálnych tlačí 18. storočia, zväzky 1 – 6 (Klimeková et al. 2008a 2008b 2008c 2008d 2008e 2008f).

Pre potreby vzniku modelovej knižnice a spôsobu zaradenia kníh do konkrétnych častí knižnice (podľa stanovených parametrov) sme sa zamerali na periodizáciu slovenského kníhtlačiarskeho. Ivona Kollárová (2002) na základe prototypov tlačiarov a vývojových trendov časovo vymedzila 6 období slovenského kníhtlačiarskeho. Za prvý medzník v dejinách slovenského kníhtlačiarskeho považuje rok 1573 ako dokázateľný rok začiatku tlače na slovenskom území, preto etapu pred rokom 1573 nazvala nultou.

- 0. Etapa: 1477 – 1573
- 1. Etapa: 1574 – 1578
- 2. Etapa: 1579 – 1720
- 3. Etapa: 1721 – 1789
- 4. Etapa: 1790 – 1870
- 5. Etapa: po roku 1871

2.1.1 Tvorba výskumnej vzorky

Do vzoriek boli zaradené všetky dostupné dokumenty z Tranovského knižnice nachádzajúce sa v depozitári knižnice projektu Pamäť Slovenska. Celkovo tak bolo vytvorených 118 vzoriek zaradených do modelovej knižnice.

Tabuľka 1 Modelová knižnica vytvorená pre potreby výskumu dizertačnej práce

Etapy periodizácie	Celkový súčet	Jazyk				
		D	HU	LAT	PL	SK
1477 - 1573	0					
1574 - 1578	0					
1579 - 1720	21			2		20
1721 - 1789	37			2		35
1790 - 1870	26	5	1			20

po roku 1871	16				1	15
nezaradené	18			1		17
Celkový súčet	118	5	1	5	1	107

2.1.2 Tvorba výskumného formulára

Výsledky vykonaných experimentov boli zaznamenávané do formulára, pričom ku každej vzorke bol vytvorený samostatný formulár. Pre potreby výskumu sme rovnako vytvorili jeden spoločný formulár, ktorý obsahoval všetky potrebné údaje pre kompletnú modelovú knižnicu. Údaje zaznamenané v jednotlivých formulároch boli zhodné s údajmi v spoločnom formulári, ktorý následne slúžil pre potreby analýz jednotlivých etáp výskumu.

Štatistické vyhodnotenia jednotlivých výstupov, ich analýza a optimálne nastavenie sa stali základom pre vytvorenie vzorkovníka typov písiem a databázy s potrebnými údajmi pre zefektívnenie digitalizácie historických slovacikálnych dokumentov.

2.2 Druhá etapa výskumu: Experiment na výskumnej vzorke

Počas druhej etapy výskumu sme sa zamerali na zodpovedanie základnej otázky:

Do akej miery má typológia typu písma historických slovacikálnych tlačí vplyv na tvorbu plnohodnotného digitálneho obsahu?

2.2.1 Tvorba digitálnych vzoriek do modelovej knižnice

Pri vytváraní digitálnych vzoriek do modelovej knižnice bolo nevyhnuté zamerať sa v prvom rade na kvalitu a stav poškodenia papiera a väzby zdrojových dokumentov. Ich stav poškodenia bol základným ukazovateľom pre stanovenie spôsobu digitalizácie. Viac ako 90% pôvodných dokumentov bolo fyzicky poškodených, preto sme sa rozhodli využiť pre tvorbu digitálnych vzoriek len ručný skener a predísť tak neúmyselnému porušeniu papiera alebo väzby.

Z každého dokumentu bolo skenovaných priemerne 5 dvojstrán. Počet skenovaných strán sa líšil v závislosti od kvality a množstva písmen na strane. V modelovej knižnici sa nachádza 1180 strán vzoriek.

Čas skenovania každej vzorky sa podrobne zaznamenával do záznamového formulára. Zaznamenaný čas sa samozrejme môže odlišovať od nami nameraných hodnôt aj v závislosti od typu a kvality skenovacieho stroja.

Čas trvania digitálnych výstupov jednotlivých dokumentov je priamoúmerný výške zvolenej hodnoty dpi. Kvalitnejšie výstupy z ručného skenera si vyžadujú dlhšiu dobu skenovania. Pri hodnote 150 dpi trvalo vytvorenie digitálnej vzorky jednej strany priemerne 23 sekúnd, pričom pri hodnote 600 dpi sa čas posunul až na hranicu 1 minúta 37 sekúnd.

2.2.2 Analýza vzniknutých vzoriek

Po vytvorení kompletných digitálnych vzoriek sme sa zamerali na analýzu vzniknutých skenov. Porovnávali sme kvalitu skenovania pri všetkých skúmaných nastaveniach hodnoty dpi.

Analýzou vytvorených vzoriek sme zistili, že nízka hodnota dpi pri tvorbe digitálneho obrazu nemusí vždy pri následnom rozpoznávaní znakov predstavovať veľký problém. Pri kvalitnej tlačenej predlohe nie je problematické optické rozpoznávanie znakov ani pri hodnote 150 dpi. Táto hodnota je však kvalitatívne nízka, preto by sme ju pre digitalizáciu historických dokumentov neodporúčali. Hodnoty 300 dpi a 600 dpi a kvalita vytvoreného obrazu je v mnohých prípadoch veľmi podobná a voľným okom málo rozpoznateľná. Preto pokiaľ je pôvodný dokument v dobrom zchovalom stave, nie je potrebné pre vytváranie digitálnej podoby dokumentu nastavenie hodnoty na 600 dpi.

Takýmto krokom je možné ušetrenie až približne polovice času potrebného na skenovanie jednej strany, čo pri 500 stranovej publikácii môže predstavovať až 3 hodiny 45 minút ušetreného času. Pri hodnote 300 dpi by bol čas potrebný na vytvorenie kompletného digitálneho obrazu približne 3 hodiny 7 minút, pričom na tvorbu digitálneho obrazu v hodnotách 600 dpi by bolo potrebných približne 6 hodín 52 minút.

Po rozbere získaných vzoriek a na základe výsledkov z vytvorených záznamov skúmania sme vytvorili podklad pre základné odporúčania tvorby digitálnych obrazov historických tlačí pre potreby vzniku plnohodnotných digitálnych textov.

2.2.3 Úprava vzoriek pred procesom OCR

Pre docielenie kvalitného a súčasne aj vizuálne dobre spracovaného výsledného produktu kompletného procesu digitalizácie, je potrebné zamerať sa aj na úpravu digitálnych obrazov pred samotným procesom optického rozpoznávania znakov. Proces úpravy digitálnych vzoriek sme vykonávali prostredníctvom programu Scan Tailor.

1. Oprava orientácie a ozertanie strán
2. Natočenie strany
3. Výber obsahu a tvorba okrajov
4. Vytvorenie výstupu

Počas experimentu na vzorkách modelovej knižnice sme sa zamerali aj na porovnanie čierneho výstupu a farebného výstupu po vykonaných úpravách. Čierno-biely mód umožňuje následné čistenie nevyžiadaných častí dokumentu, text je častokrát zreteľnejší, avšak dodatočné úpravy jasu a kontrastu v programe ABBYY FineReader nie sú podporované. Možné úpravy v programe sú využiteľné pre farebný výstup, kde nevzniká problém s využitím úprav jasu, kontrastu... Výsledný dokument vytvorený vo farbe po uskutočnení procesu optického rozpoznávania znakov je vizuálne totožný s pôvodným tlačeným vydaním.

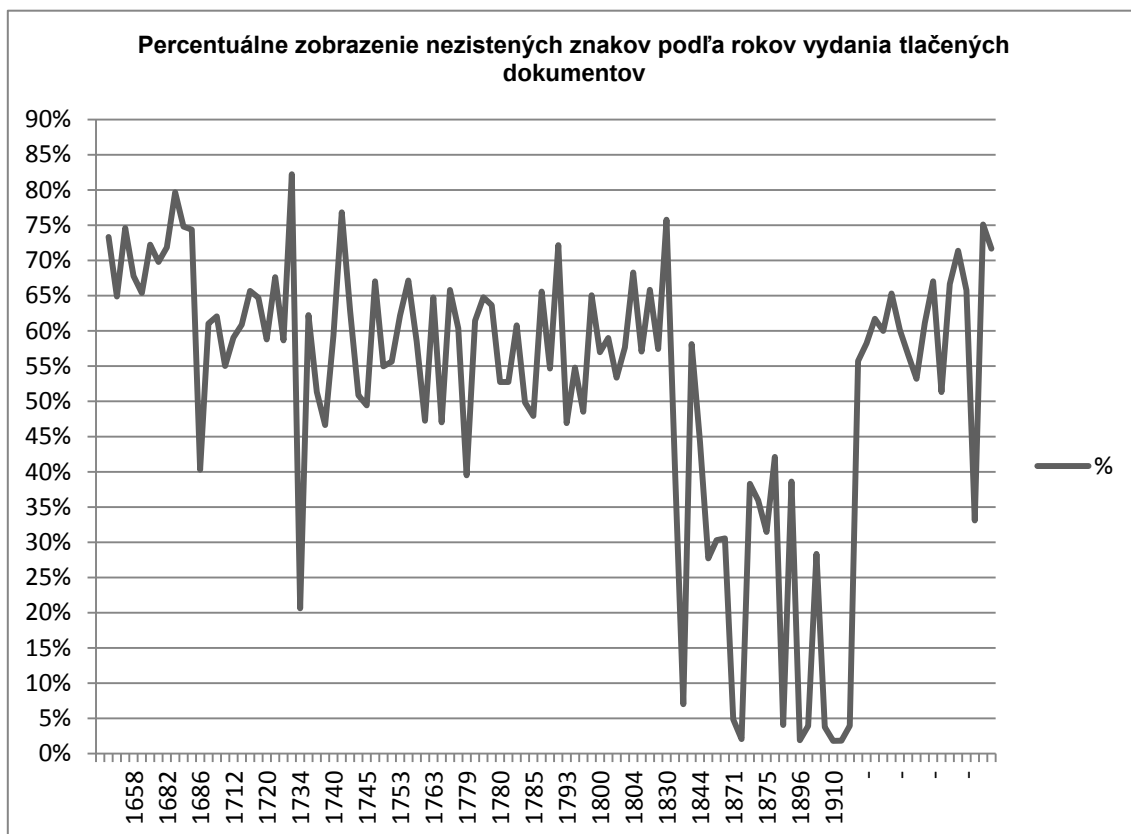
2.2.4 Optické rozpoznávanie znakov (OCR) v základných nastaveniach

Zamerali sme sa výlučne na historické dokumenty písané v slovenskom jazyku. Z dostupných 118 dokumentov bolo následne vybraných 107 dokumentov obsahujúcich nami zvolený jazyk. Proces optického rozpoznávania znakov bol vykonávaný v programe ABBYY FineReader 12.

Pre všetky vzorky bol zvolený slovenský a český jazyk rozpoznávania, ako jediný úkon vykonaný v nastaveniach, pretože jazyk v skúmaných dokumentoch je najpodobnejší vyššie

spomínaným jazykom. Program automaticky rozpoznal oblasti textu na zvolených stranách a následne strany podrobil procesu „čítania“.

Pred analýzou nameraných výsledkov bolo potrebné stanoviť konkrétne percento neúspešnosti, ktoré sme v následných analýzach považovali za primerané množstvo nezistených znakov daných dokumentov. Primeraná percentuálna hodnota nezistených znakov bola stanovená na maximálne 5% z celkového počtu znakov v dokumente. Všetky vzorky, z ktorých sme v prvom procese OCR získali požadovaných maximálne 5% neúspešnosti sme nezaradili do druhého kola procesu OCR.



Graf 1 Percentuálne zobrazenie nezistených znakov podľa rokov vydania tlačéných dokumentov

Po zatriedení nameraných hodnôt neúspešnosti do etáp slovenského kníhtlačiarstva sme analyzovali priemernú a rovnako aj prostrednú percentuálnu hodnotu neúspešnosti procesu OCR pre možnosť efektívnejšieho porovnania nameraných hodnôt.

Tabuľka 2 Percentuálne znázornenie nerozpoznaných znakov podľa periodizácie

Etapy periodizácie		Priemerná percentuálna neúspešnosť	Stredová percentuálna neúspešnosť
0.	1477 – 1573		
1.	1574 – 1578		
2.	1579 - 1720	61.52%	61.05%
3.	1721 - 1789	57.86%	59.34%
4.	1790 - 1870	50.26%	55.88%
5.	do roku 1871	16.20%	4.07%

	nezaradené	60.83%	61.14%
--	------------	--------	--------

Zo získaných výsledkov je zrejmé, že najproblematickejším obdobím výskumu je druhá etapa periodizácie. Predpokladáme, že nezaradené dokumenty pochádzajú väčšinou z obdobia do roku 1800, čoho vyplýva rovnako veľmi vysoká percentuálna neúspešnosť.

Procesom optického rozpoznávania znakov bez akýchkoľvek doplňujúcich nastavení program vyhodnotil celkovo 332 446 znakov z toho 169 109 nezistených, čo predstavuje priemernú percentuálnu hodnotu 52,56% a prostrednú percentuálnu hodnotu 58,50% neúspešne identifikovaných znakov. Výsledná hodnota je veľmi ovplyvnená 5. etapou periodizácie v období od roku 1871, kde zo 16tich skúmaných vzoriek splnilo hranicu neúspešnosti pod 5% nerozpoznaných znakov až 9 vzoriek. Pre potreby výskumu boli následne vzorky z 5. etapy periodizácie rozdelené podľa typu písma a výsledné percentuálne hodnoty neúspešnosti vyrátané samostatne pre vzorky obsahujúce gotický typ písma a vzorky obsahujúce latinku. Percentuálna neúspešnosť rozpoznávania znakov vo vzorkách písaných gotickým typom písma bola 35,29%, pričom vzorky písané latinkou získali hodnotu percentuálnej neúspešnosti 3,73%.

Po vykonaní všetkých meraní a následnej analýze získaných hodnôt OCR v základných nastaveniach sme pristúpili k procesu optického rozpoznávania znakov s použitými nastaveniami.

2.2.5 Optické rozpoznávanie znakov (OCR) s použitím nastavení

Na základe zistení z predchádzajúcich meraní OCR v programe ABBYY FineReader bez nastavení sme sa prvotne zamerali na tvorbu vzorov písmen z dôvodu neúspešnosti programu rozpoznať dané znaky.

Prvým krokom bolo nastavenie jazyka dokumentu.

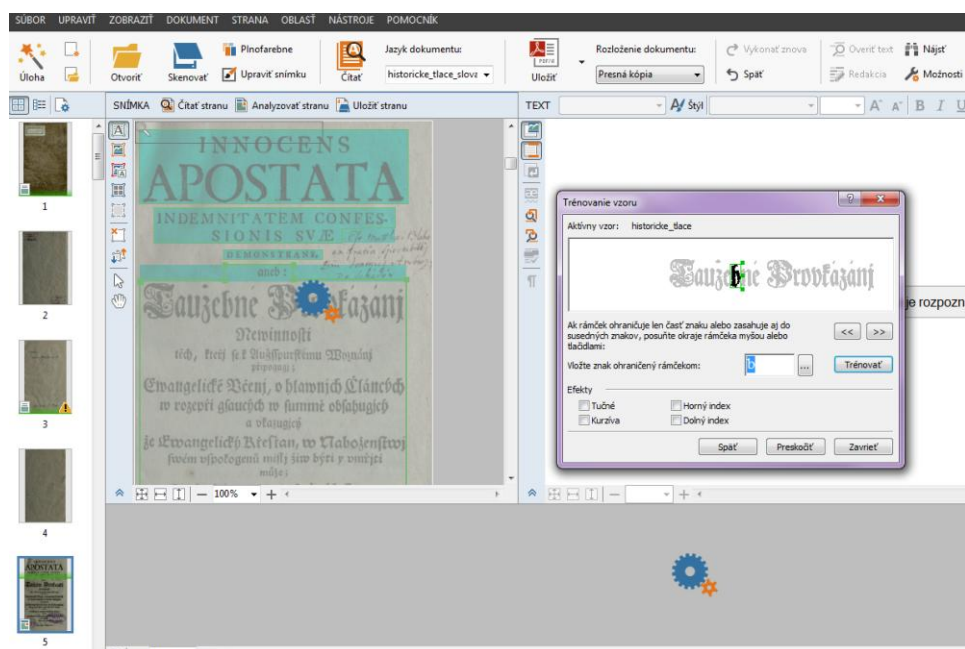
V druhom kroku sme vytvorili používateľský vzor písmen „historicke_tlace“.

Pred procesom čítania bolo potrebné zamerať sa na výber textu. Program ABBYY FineReader ponúka možnosti vymedzenia tabuliek, textu, obrázkov... Pre OCR vzorky sme sa zamerali na výber a označenie textových častí dokumentu.

Čítanie s tréningom

Podrobné učenie sa programu, tj. „čítanie s tréningom“ je zdĺhavý proces, prostredníctvom ktorého prebieha podrobné učenie sa konkrétnych znakov vyskytujúcich sa vo zvolenom dokumente. Okrem samotných písmen sa v ňom môžu nachádzať rôzne iné znaky využívané v texte v jednotlivých obdobiach tlače.

Prostredníctvom vymedzenia textu vlastnoručne, dokáže používateľ zabrániť nežiaducemu označeniu znakov, ktoré nie je potrebné rozpoznať prostredníctvom OCR.



Obrázok 1 Ukážka "čítania s tréningom"

V procese „čítania s tréningom“ sme postupne prichádzali na mnohé spôsoby učenia konkrétnych písmen. Program v procese čítania veľmi často zamieňal podobné typy písmen, preto bolo preň ich čítanie veľmi zložitá a výsledok ich rozpoznania nevyhovujúci. Pre niektoré po sebe nasledujúce písmená sme preto vytvorili jeden konkrétny znak pre efektívnejšie rozpoznanie spomínaných znakov.

Počas výskumu sme tak vytvorili spojenia najčastejšie sa vyskytujúcich po sebe idúcich písmen: „st“, „sk“, „ss“, „sl“, „ch“, „ij“.

Najčastejšie problémy pri využití OCR

Po prvých experimentoch v oblasti rozpoznávania znakov s ich učením sme analyzovali niekoľko závažných problémov:

1. Označenie nezisteného znaku aj v prípade jeho správneho rozpoznania.

Počas OCR vzoriek v modelovej knižnici sme postupovali dvoma spôsobmi. Pri prvej polovici vzoriek sme vykonávali „čítanie s tréningom“ na každej strane dokumentov. Týmto spôsobom sme do používateľského vzoru zaradili podstatnú časť znakov – písmen. V druhej polovici vzoriek sme vykonávali experiment s rozpoznávaním znakov bez procesu „čítania s tréningom“. Po vyznačení textových častí dokumentu sme spustili optické rozpoznávanie znakov automaticky bez použitia „tréningu“. Analýza získaných výsledkov poukazuje na nesprávne označenie nezistených znakov aj v prípade jeho správneho rozpoznania. Program správne rozpoznal daný znak, ale jeho označenie v celkovom hodnotení úspešnosti bolo negatívne. Veľmi intenzívne daný problém ovplyvňovala skutočnosť, že program nedisponoval slovami vyskytujúcimi sa v texte skúmaných dokumentov.

Problém sa nám podarilo čiastočne vyriešiť vytvorením slovníku slov nachádzajúcich sa v skúmaných dokumentoch. Vytvorený slovník s názvom „historicke_tlace_slovacika_slova“ sme založili na základe českého jazyka, pretože jeho podobnosť so skúmaným textom je

najbližšia. Po „čítaní s tréningom“ a následnej úprave rozpoznaných strán zvoleného dokumentu boli všetky neznáme vyskytujúce sa slová v daných dokumentoch pridávané do slovníka. Výsledkom je vytvorenie slovníka slov, najčastejšie sa vyskytujúcich v nami skúmaných slovacikálnych tlačiarach.

2. Neustála záměna písmen.

Problém neustálej záměny niektorých písmen sa nám nepodarilo vyriešiť do úplnej spokojnosti. Počet navzájom zamenených písmen sa postupným učením znakov, a tak vytváraním obsiahlejšej databázy znakov podarilo eliminovať, ale úplne odstránenie problému sa nám počas výskumu nepodarilo. Predpokladáme, že dôvodom záměny písmen je ich veľká podobnosť, ktorú program aj napriek veľkému množstvu používateľských vzorov daného znaku počas nášho výskumu nedokázal úspešne rozpoznať vo všetkých výskytoch..

Tabuľka 3 Najčastejšie sa navzájom zamieňajúce znaky

Pôvodný znak	Zamieňaný znak
m	w
n	u
y	h
k	t
s	f
u	v
l	t
l	k
e	c
é	ě

Pôvodný znak	Zamieňaný znak
č	ě
l	I
i	j
d	b
B	V
l	1
r	t
C	E
č	é
D	O

Pôvodný znak	Zamieňaný znak
i	í
j	í
I	1
M	W
ř	ť
ú	ň
ú	ů
ů	ň

3 Výstupy dizertačnej práce

Hlavným cieľom výskumu dizertačnej práce bolo navrhnutie optimálnych pravidiel a postupu digitalizovania historických tlačených slovacikálnych dokumentov pre vytvorenie čo najkvalitnejšieho a použiteľného digitálneho obsahu v rámci historických knižníc.

Pre splnenie cieľa boli vytvorené základné výstupy dizertačnej práce:

- Odporúčané optimálne pravidlá digitalizácie historických slovacikálnych tlačí
- Katalóg písmen a znakov
- Slovník vyskytujúcich sa slov
- Elektronická verzia historickej slovacikálnej knihy s plne rozpoznaným textom
- Súbor „historicke_tlace“ pre OCR historických slovacikálnych dokumentov na podporu programu AbbyyFinereader

3.1 Odporúčané optimálne pravidlá digitalizácie historických slovacikálnych tlačí

Na základe výstupov z výskumu dizertačnej práce a následných analýz získaných výsledkov sme navrhli odporúčané optimálne pravidlá digitalizácie tlačených historických slovacikálnych dokumentov:

- *Minimálna odporúčaná hodnota 300 dpi, pri nekvalitnom, poškodenom stave dokumentu 600dpi.*
- *Mimoriadnu pozornosť preba venovať úprave vytvorených skenov hlavne pri väčších nastaveniach dpi z dôvodu zviditeľnenia všetkých nedostatkov tlače, papiera.*
- *Úprava vytvorených skenov vo farebnom prevedení / odtieňoch sivej.*
- *V prípade potreby úprava jasu a kontrastu pre lepšiu viditeľnosť skúmaného textu (nie je možná pri čierno-bielej snímke).*
- *Pri procese OCR využitie používateľského vzoru „historicke_tlace“, ktorý je možné importovať do programu ABBYY FineReader pred spustením procesu čítania.*
- *V programe ABBYY FineReader nastavenie jazyka dokumentu na „historicke_tlace_slovacika_slova“, ktorý je súčasťou používateľského vzoru „historicke_tlace“.*

3.2 Katalóg písmen a znakov

Pre efektívne a používateľsky optimálne využitie vytvoreného katalógu písmen sme počas tvorby používateľského vzoru vytvárali jeden kompletný zoznam všetkých vyskytujúcich sa znakov v skúmaných dokumentoch. Program ABBYY FineReader pracoval s jedným základným používateľským vzorom, z ktorého následne dokázal generovať opticky rozpoznaný text. Tvorbu centralizovaného zoznamu všetkých písmen do jedného používateľského vzoru sme navrhli pre efektívnejšie využitie nami vytvoreného súboru, ktorý obsahuje kompletné nastavenia programu, všetky naučené používateľské vzory znakov a slovník vyskytujúcich sa slov. Pri tvorbe digitálnych textov z historických slovacikálnych dokumentov tak nie je potrebné zamerať sa výlučne na jeden typ písma, ale v procese OCR je možné súbežne vytvoriť digitálne knihy obsahujúce viaceré typy písma.

Počas tvorby katalógu sme zaznamenávali údaje rozoznatelnosti jednotlivých znakov uvedené v časti poznámky. Získané údaje sme zaradili do stupnice od 1 do 5, podľa kvality čitateľnosti daného znaku. Hodnotenie bolo rovnaké ako v škole, kde 1 predstavuje bezproblémové rozpoznanie znaku a 5 veľmi problematické rozpoznávanie znaku. Pri najproblematickejších znakoch, v stupnici označené číslom 4 a 5, sme uvádzali najčastejšie zameniteľné znaky.

Tabuľka 4 Čitateľnosť písmen a znakov uvedených v katalógu

Zna	Poznámka
a	1
á	2
A	2 (A – U)
b	5 (b – d)
B	5 (B – V)
c	5 (c – e)
C	4 (C – E)
č	5 (č – ě – é)
d	5 (d – b)
D	4 (D – O)
d'	3
e	5 (e – c)
E	4 (E – C)
é	5 (é – ě – č)
ě	5 (ě – é – č)
f	5 (f – s)
F	3
g	2
G	3 (G – S)
h	4 (h – v)
H	2 (H – Y)
i	5 (i – í – j)
í	5 (í – i – j)

Zna	Poznámka
I	4 (I – l – 1)
i	5 (i – í – j)
J	2
k	5 (k – t – l)
K	3 (K – R)
l	5 (l – k – t)
L	1
m	4 (m – w)
M	4 (M – W)
n	5 (n – u)
N	2
ň	5 (ň – ů – ú)
o	3
O	4 (O – D)
ó	3 (ó – d)
p	2
P	2
r	4 (r – t)
R	3 (R – K)
ř	4 (ř – t')
s	4 (s – f)
S	3 (S – G)
t	5 (t – k – l) 4 (t – r)

Zna	Poznámka
T	1
t'	4 (t' – ř)
u	5 (u – n) 4 (u – v)
U	2 (U – A)
ů	5 (ů – ň – ú)
ú	5 (ú – ň – ů)
v	4 (v – u)
V	5 (V – B)
w	4 (w – m)
W	4 (W – M)
X	1
y	4 (y – h)
Y	2 (Y – H)
ý	3 (ý – h)
z	2
Z	2
ž	3 (ž – j)
Ž	2
!	4 (nezistený znak)
:	3
;	3
(	3 (- /)

3.3 Slovník vyskytujících sa slov

Počas experimentu zameraného na optimálne nastavenia programu ABBYY FineReader na OCR pre vygenerovanie čo najpresnejšieho digitálneho prepisu bolo potrebné vytvorenie slovníka obsahujúceho čo najväčší počet slov nachádzajúcich sa v skúmaných vzorkách. Počas výskumu boli postupne do slovníka vkladané všetky slová, ktoré neboli programom na spracovanie OCR rozpoznané. Slovník „historicke_tlace_slovacika_slova“ bol vytvorený na základoch českého jazyka, preto niektoré slová nebolo do slovníka potrebné ukladať.

3.4 Elektronická verzia historickej slovacikálnej knihy s plne rozpoznávaným textom

Ako vzorový dokument bola zvolená kniha s Identifikačným číslom - XIX.G.259 s názvom: INNOCENS APOSTATA INDEMNITATEM CONFESSIONIS SVAE DEMONSTRANS, aneb: Taužebné Provkázánj Newinnosti těch, ktoř se k Augsspurskému Wyznánj připogují; 1750.

Záverom výskumu v dizertačnej práci bolo na základe vykonaného výskumu a následného vyšpecifikovania používateľského vzoru a jeho doplnenia o potrebné znaky a slová vytvorenie elektronickej verzie historickej knihy s plne rozpoznaným textom.

3.5 Súbor „historicke_tlace“ pre OCR historických slovacikálnych dokumentov na podporu programu AbbyyFinereader

Na základe porovnania a analýz vytvorených výstupov z prvej a druhej etapy výskumu boli navrhnuté optimálne pravidlá pre vznik plne rozpoznateľných textov, ktoré je možné aplikovať na optické rozpoznanie znakov historických tlačí v slovenskom jazyku v programe ABBYY FineReader pre potreby prepisu historického písma slovacikálnych dokumentov do elektronickej podoby.

Jedným z prioritných cieľov bolo vytvorenie používateľského vzoru, ktorý v sebe obsahuje:

- všetky možnosti nastavenia potrebné pre efektívnejšie rozpoznanie historických slovacikálnych dokumentov,
- naučené znaky historických písmen,
- slovník obsahujúci slová vyskytujúce sa vo vzorkách dokumentoch v modelovej knižnici vytvorenej pre potreby dizertačnej práce.

Vytvorený súbor bol pomenovaný „historicke_tlace“ a je možné ho importovať do programu ABBYY FineReader pre potreby OCR.

Prínosy dizertačnej práce

Prioritným prínosom predkladanej dizertačnej práce je vytvorenie katalógu písmen nachádzajúcich sa v tlačených slovacikálnych dokumentoch uložených v zbierke Tranovského knižnice.

Okrem tlačenej verzie katalógu bol vytvorený súbor na podporu programu ABBYY FineReader pre potreby digitalizácie starých slovacikálnych tlačí, ktorý okrem znakov obsahuje slovník vyskytujúcich sa slov.

Možnosť využitia spomínaného katalógu na úrovni odboru mediamatiky a kultúrneho dedičstva je pre potreby vytvorenia plne rozpoznateľných digitálnych textov zapožičanej časti zbierky Tranovského knižnice, v rámci ktorej bol celý výskum riešený. Vzhľadom na už vytvorený zoznam nachádzajúcich sa slov v skúmaných vzorkách je tvorba plne rozpoznateľných textov kvalitnejšia a súčasne časovo menej náročná. Výskumníci sa nemusia zaoberať výučbou základných písmen a znakov, pričom sa môžu plne zamerať na kontrolu najproblematickejších a najčastejšie zameniteľných písmen, čo sa nám počas celého výskumu nepodarilo plne dosiahnuť. Na základe pozorovaní počas výskumu predpokladáme, že zvyšujúcim sa počtom vytvorených kníh sa percento problematicky rozpoznateľných a zamieňaných znakov bude znižovať.

Katalóg, či už v tlačenej alebo elektronickej podobe, je rovnako možné využiť vo vedeckej sfére, v oblasti digitalizovania historických tlačí, uchovávaní písomného kultúrneho dedičstva, ako jednu oblasť vstupných dát pri realizácii analýz typov písmen a možnosti ich zámeny a rozpoznania v obdobných typoch dokumentov.

Optické rozpoznávanie historických tlačí sa na území Slovenska rieši individuálne vo výskumných pracoviskách, ktoré si svoje poznatky chránia. Momentálne nie je voľne prístupný podporný materiál pre tvorbu OCR zameraný na historické slovaciká, preto zhotovený elektronický súbor písmen a slov (používateľský vzor programu) môže slúžiť ako základ pre vytvorenie uceleného podporného programu určeného pre optické rozpoznávanie slovacikálnych dokumentov.

Záver

Tvorba digitálnych historických knižníc a s ňou spojené optické rozpoznávanie znakov historických dokumentov je jeden zo základných atribútov uchovávaní a sprístupňovania písomného kultúrneho dedičstva ďalším generáciám. V priebehu 423 rokov od roku 1477 bolo na území dnešného Slovenska vytlačených približne 30 000 územne slovacikálnych dokumentov. Mnohé z nich sa zachovali dodnes a predstavujú neodmysliteľnú súčasť nášho kultúrneho a spoločenského života.

Hlavným cieľom predkladanej dizertačnej práce bolo vytvorenie podporného systému pre tvorbu digitálnych obsahov práve spomínaných tlačí a zameranie sa tak na historické a kultúrne pramene odzrkadľujúce vtedajší život na našom území.

Počas výskumu sme mali k dispozícii 118 dokumentov z obdobia od roku 1582 do roku 1914 zo zbierky Tranovského knižnice. Na základe jednotlivých výstupov zo skenerov sme navrhli optimálne možnosti digitalizovania historických tlačených územne slovacikálnych dokumentov pre vytvorenie čo najkvalitnejšieho a použiteľného digitálneho obsahu v rámci historických knižníc, čo znamená možnosť identifikovať zatiaľ neidentifikovateľné tlačoviny.

V priebehu práce sme narazili na rôzne problémy, ktoré ovplyvnili smerovanie nášho výskumu. Zamerali sme sa na zaradenie získaných vzoriek do periodizácie slovenského kníhtlačiarstva, analýzu jazykového zastúpenia a na vytvorenie zoznamu názvov historických kníh, ktorý vzhľadom na poškodené stavy jednotlivých exemplárov nebolo možné skompletizovať.

Prioritnou časťou výskumu bolo zameranie sa na optické rozpoznávanie znakov vyskytujúcich sa v historických slovacikálnych dokumentoch, na základe ktorých bol následne vytvorený katalóg písmen a znakov. Jedným z hlavných problémov, ktorý významne ovplyvnil výskumnú časť práce bola chýbajúca zložka, obsahujúca korpus jazyka vyskytujúceho sa v skúmanom časovom období na našom území. Vzniknutá situácia bola čiastočne vyriešená vytvorením slovníka používaných slov, ktorým sa nám podarilo aspoň sčasti daný problém potlačiť.

Ciele dizertačnej práce boli naplnené, o čom svedčí vytvorenie tlačeného katalógu písmen a znakov nachádzajúcich sa v dokumentoch tlačených na našom území v obdobiach od roku 1582 do roku 1914, z ktorých sme mali k dispozícii výskumné vzorky, ako aj vytvorenie elektronického súboru pod názvom „historicke_tlace“ ako podporného používateľského vzoru programu ABBYY FineReader obsahujúceho spomínané písmená a znaky a súčasne aj slovník vyskytujúcich sa slov.

Vytvorený používateľský vzor môže slúžiť ako základ pre tvorbu uceleného podporného programu softvéru ABBYY FineReader pre OCR historických slovacikálnych tlačí. Alternatívu vidíme v kooperácii skupiny riešiteľov z viacerých oblastí výskumu,

zaoberajúcich sa danou problematikou. Základnou požiadavkou vyplývajúcou z prevedeného skúmania je zameranie sa na analýzu korpusu vtedy používaného jazyka na území Slovenska, pretože v súčasnej dobe dokáže program pracovať len s vloženými slovami v uloženom tvare.

Prostredníctvom prvej analýzy vykonanej v retrospektívnych bibliografiách zameranej na tlač do roku 1700, evidované v Slovenskej národnej knižnici boli vybrané tri vzorky dokumentov vyprodukované od jedného tlačiaru. Výber sa orientoval na tri etapy jeho tvorby, a to na začiatok, stred a koniec pôsobenia tlačiaru. Bola tak vytvorená tabuľka, ktorá obsahuje eventuálnu výskumnú vzorku zoznamu tlačiarov a tlačiarňí na území Slovenska s pridelenými signatúrami jednotlivých vzoriek pre zjednodušenie následnej manipulácie s dokumentmi (pozri Príloha D). Vytvorením vzorky vidíme možnosť ďalšieho výskumu na obohatenie používateľského vzoru na podporu OCR a pomôže tak obsiahnuť širšie pole písmen a znakov vyskytujúcich sa v historických územne slovacikálnych tlači.

Zoznam použitej literatúry

ABBY. Čo je OCR. [online]. ABBY, ©2014 [cit. 2013-11-10]. Dostupné na: http://www.sk.abby.com/products/personal/finereader/whatis_ocr.

ANDERSON, Niall. IMPACT Briefing Paper: Optical Character Recognition. [online]. IMPACT, ©2010 [cit. 2013-12-14]. Dostupné na: <http://www.impact-project.eu/uploads/media/IMPACT-ocr-bp-pilot-1b.pdf>

AUGUSTÍNOVÁ, Eva. 2002. Dejiny tlačiarňí na Slovensku 16. – 19. storočie. Rukopis.

AUGUSTÍNOVÁ, Eva. 2013. Základný výskum historických knižných dokumentov a fondov: Výskumný projekt Pamäť Slovenska – národné centrum excelentnosti výskumu, ochrany a sprístupňovania kultúrneho a vedeckého dedičstva. Žilina: Poradca, s.r.o. ISBN 978-80-8162-000-3.

BREZA, Vojtech. 1997. Tlačiarne na Slovensku 1477-1996. Martin: Matica slovenská. ISBN 80-85170-25-6.

ČAPLOVIČ, Ján. 1972. Bibliografia tlači vydaných na Slovensku do roku 1700. 1. diel. Martin: Matica slovenská.

ČAPLOVIČ, Ján. 1984. Bibliografia tlači vydaných na Slovensku do roku 1700. 2. diel. Martin: Matica slovenská.

FÖLDES-PAPP, Károly cit. podľa KRUPA, Viktor a GENZOR, Jozef. 1989. Písma sveta. Bratislava: Obzor. ISBN 80-215-0011-5.

HINCOVÁ, Katarína a HÚSKOVÁ, Alexandra. 2007. Slovenský jazyk pre 1. – 4. ročník stredných škôl. Prvé vydanie. Bratislava: Slovenské pedagogické nakladateľstvo. ISBN 978-80-10-00690-8.

HÚŠČAVA, Alexander. 1951. Dejiny a vývoj nášho písma. Bratislava: Slovenská akadémia vied a umení.

IMPACT Project: Improving Access to Text. [online]. IMPACT, ©2012 [cit. 2013-12-20]. Dostupné na: <http://www.impact-project.eu>.

InfoLib. Digitalizácia. [online]. ©2012 [cit. 2013-12-10]. Dostupné na: <http://old.infolib.sk/index/podstranka.php?id=1299>.

JENSEN, Hans cit. podľa KRUPA, Viktor a GENZOR, Jozef. 1989. Písma sveta. Bratislava: Obzor. ISBN 80-215-0011-5.

JEŘÁBEK, Vlastimil. 2011. Rekonstrukce historického písma na základě podrobné analýzy: bakalárska práca. Brno: Masarykova univerzita.

KATUŠČÁK, Dušan. Stratégia digitalizácie kultúrneho, vedeckého a intelektuálneho dedičstva SR. [online]. [cit. 2013-03-23]. Dostupné na: http://www.p3.sk/domain/flox/files/download/Stratgia_digitalizcie_kultrneho,_vedeckho_a_in_telektulneho_dediTstva.ppt.

KATUŠČÁK, Martin. 2012. Komplexné spracovanie písomného kultúrneho a vedeckého dedičstva SR: dizertačná práca. Žilina: Žilinská univerzita v Žiline.

KLIMEKOVÁ, Agáta a AUGUSTÍNOVÁ, Eva a ONDROUŠKOVÁ, Janka. 2008a. Bibliografia územne slovacikálnych tlačí 18. Storočia. I zväzok. Martin: Slovenská národná knižnica. ISBN 978-80-89301-17-1.

KLIMEKOVÁ, Agáta a AUGUSTÍNOVÁ, Eva a ONDROUŠKOVÁ, Janka. 2008b. Bibliografia územne slovacikálnych tlačí 18. Storočia. II zväzok. Martin: Slovenská národná knižnica. ISBN 978-80-89301-18-8.

KLIMEKOVÁ, Agáta a AUGUSTÍNOVÁ, Eva a ONDROUŠKOVÁ, Janka. 2008c. Bibliografia územne slovacikálnych tlačí 18. Storočia. III zväzok. Martin: Slovenská národná knižnica. ISBN 978-80-89301-19-5.

KLIMEKOVÁ, Agáta a AUGUSTÍNOVÁ, Eva a ONDROUŠKOVÁ, Janka. 2008d. Bibliografia územne slovacikálnych tlačí 18. Storočia. IV zväzok. Martin: Slovenská národná knižnica. ISBN 978-80-89301-20-1.

KLIMEKOVÁ, Agáta a AUGUSTÍNOVÁ, Eva a ONDROUŠKOVÁ, Janka. 2008e. Bibliografia územne slovacikálnych tlačí 18. Storočia. V zväzok. Martin: Slovenská národná knižnica. ISBN 978-80-89301-29-4.

KLIMEKOVÁ, Agáta a AUGUSTÍNOVÁ, Eva a ONDROUŠKOVÁ, Janka. 2008f. Bibliografia územne slovacikálnych tlačí 18. Storočia. Súbor. Martin: Slovenská národná knižnica. ISBN 978-80-89301-16-4.

KOLLÁROVÁ, Ivona. 2002. Pokus o logickú periodizáciu dejín slovenského kníhtlačiarsťva. In: Kniha 2001 – 2002. Martin: Slovenská národná knižnica. s. 38-47. ISBN 80-89023-19-3.

- KRAJČOVIČ, Rudolf. 1980. Svedectvo dejín o slovenčine. Martin: Matica slovenská.
- KRUPA, Viktor a GENZOR, Jozef. 1989. Písma sveta. Bratislava: Obzor. ISBN 80-215-0011-5.
- LIBA, Peter. 2006. Bibliografía v procese. Martin: Slovenská národná knižnica. ISBN 80-89023-85-1.
- MÁLEK, Matěj. 2014. Čitelnost písma, návrh písma s vysokou čitelností a rasterizace fontů: diplomová práce. Brno: Masarykova univerzita.
- Metodický manuál pre zabezpečenie konverzie textových objektov. [online]. Spolok slovenských knihovníkov, ©2009 [cit. 2013-09-08]. Dostupné na: http://www.opis.culture.gov.sk/uploads/a5/80/a580a462e150a07746df2adc977619c5/05_metodicky_manual_pre-zabezpecenie-konverzie-textovych-objektov.pdf
- MORAVČÍK, Fridrich. 1992. Zákonitosti tvorby písma. Pezinok: Milan Moravčík. ISBN 80-901394-0-X.
- MUZIKA, František. 2005. Krásne písmo ve vývoji latinky. I. a II. díl. Praha: Paseka. ISBN 80-7185-740-8.
- MUZIKA, František. Význam a história písma. [online]. [cit. 2014-01-10]. Dostupné na: <http://ssuszv.edupage.org/files/pismo.pdf>.
- NĚMEC, Václav. Pomocné vědy historické [online]. ©2014 [cit. 2014-11-10]. Dostupné na: <http://www.dejepis.com/ucebnice/pomocne-vedy-historicke>.
- NUPSESO GROU. ABBYY FineReader XIX. [online]. ABBYY, ©2012 [cit. 2013-11-10] Dostupné na: <http://nupseso.eu/softver/konverzia-ocr/abbyy-finereader-xix>.
- Predpis č. 183/2000 Z. z. Zákon o knižniciach, o doplnení zákona Slovenskej národnej rady č. 27/1987 Zb. o štátnej pamiatkovej starostlivosti a o zmene a doplnení zákona č. 68/1997 Z. z. o Matici slovenskej.
- RAITT, David. Digital Library Initiatives Across Europe. [online]. Computers Libraries, ©2010 [cit. 2014-01-20] Dostupné na: <http://www.infoday.com/cilmag/nov00/raitt.htm>.
- REITZ, Joan M. Online Dictionary for Library and Information Science. [online]. ABC-CLIO, ©2014 [cit. 2014-01-20]. Dostupné na: http://www.abc-clio.com/ODLIS/odlis_A.aspx.
- Slovník [online]. Bratislava: Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied. 2014 [cit. 2014-01-25]. Dostupné na: <http://slovník.juls.savba.sk/>.
- SNK. Spýtajte sa knižnice. [online]. ©2010 [cit. 2013-12-15]. Dostupné na: http://www.snk.sk/swift_data/source/OS/Spytajte%20sa%20kniznice/2007-05.pdf.

SNYDER, Henry L. a HUTCHINSON, Heidi L.. 1994. Cataloging of the Hand Press: A Comparative and Analytical Study of Cataloging Rules and Formats Employed in Europe. Hague : IFLA/SAUR. ISBN 3-598-23400-7.

ŠPETKO, Jozef. 1963. Dejiny písma a knihy. Bratislava: Slovenské pedagogické nakladateľstvo.

ŠPETKO, Jozef. 1969. Knižná kultúra a písmo. Martin: Matica slovenská.

ŠTĚRBOVÁ, Eva. 2006. Rekonstrukce fraktury ze 16. století a vytvoření fontu daného písma: bakalárska práca. Brno: Masarykova univerzita.

VALACH, Július. 1987. Staré tlače a tlačiari na Slovensku. Martin: Matica slovenská.

V. ECSÉDY, Judit. 2010. A régi magyarországi nyomdák betűi és díszei XVII. század. Budapest: Balassi kiadó-országos széchényi könyvtár. Hungaria typographica II. 1. kötet: nyugat- és észak-magyarországi nyomdák. ISBN 978-963-506-791-6.

VOIT, Petr. 2008. Encyklopedie Knihy: Starší knihtisk a příbuzné obory mezi polovinou 15. a počátkem 19. století. I a II. díl. Praha: Nakladatelství Libri. ISBN 978-80-7277-390-9.

VOIT, Petr. 2011. Tiskové písmo Čech a Moravy první poloviny 16. století. In: Bibliotheca Strahoviensis. roč. 2011, č. 10, s. 105-202. ISSN 1213-6514.

VOIT, Petr. 2013. Český knihtisk mezi pozdní gotikou a renesancí: I severinsko-kosořská dynastie 1488-1557. Praha: KLP. ISBN 978-80-86791-98-2.

WAGNER, Bettina a REED, Marcia. 2010. Early Printed Books as Material Objects: Proceedings of the conference organized by the IFLA Rare Books and Manuscripts Section, Munich, 19-21 August 2009. Hague: IFLA. ISBN 978-3-11-025324-5.

WHITNEY, Lance. IBM helping Europe scan historical documents. [online]. Cnte, ©2010 [cit. 2014-01-20]. Dostupné na: http://news.cnet.com/8301-11386_3-20014774-76.html.

Recenzenti: prof. PhDr. Dušan Katuščák, PhD., doc. PaedDr. Zdena Kráľová, PhD.