

POKROK V TRANSKRIPCI HISTORICKÝCH RUKOPISŮ

Dušan Katuščák
Lukáš Němec
Vojtěch Říha



**SLEZSKÁ
UNIVERZITA**

FILOZOFICKO-
PŘÍRODOVĚDECKÁ
FAKULTA V OPAVĚ

Úvod

- Historické staré a vzácne tlače, strojopisy a hlavne rukopisy spravidla nie je možné uspokojivo transkribovať pomocou OCR
- Prichádza na pomoc umelá inteligencia
- V snahách sprístupniť historické písomné dedičstvo sa koncentruje pozornosť výskumníkov na *transkripciu* a strojové učenie s použitím *konvolučných neurónových sietí*
- Ide o proces, v ktorom sa nasnímaný *obrázok* mení na *text*.
- V Česku: Vynikajúci nástroj **PERO**, služba **InkCapture**, skvelá práca **Anna Michalcová** (ČAV)

Ciel' prezentácie

1. Vysvetliť metodiku tvorby modelov transkripcie v platforme Transkribus
2. Informovať o výsledkoch projektu Študentskej grantovej súťaže na Slezskej univerzite v Opave (FPF, Oddelenie knihovníctva)
3. Prezentovať výsledky experimentov študentov

Prečo modely?

- Modely slúžia na transkripciu historických textov
- Využíva sa umelá inteligencia
- Na vytvorenie modelu je potrebné stroj naučiť, čo má robiť
- Učenie prebieha tak, že sa manuálne pripraví tréningový set (Train set) a validačný set (Validation set)
- Strany textu je potrebné prepísať čo najpresnejšie do kvality GT (Ground Truth)
- Spustí sa proces trénovania
- Výsledkom trénovania je MODEL
- Na základe čiastkových modelov je možné pripraviť univerzálne supermodely

Platforma Transkribus

Ako to funguje? Všetko je o učení umelej inteligencie!

Dokument 10 000 strán

65

10

10 000 strán

Ground Truth (GT)

Cvičné strany (25-75 strán/5000-15000 slov - transkribovať: model/ručne

Trénovanie modelu: napr. 75 strán

Rozdelíme na: TRAIN SET a VALIDATION SET

TRAIN SET: napr. 65 strán

VALIDATION SET: napr. 10 strán

Spustíme učenie: napr. 250 cyklov (Výsledok je: CER/WER)

Model použijeme na automatickú transkripciu 925 strán

Tvorba modelu automatického rozpoznávání bohemikálního rukopisného textu v platformě Transkribus

Tvorba modelu automatického rozpoznávání bohemikálního rukopisného textu v platformě Transkribus

Podmínky:

- **časové okno 18. – 20. století**
- **žánrová diferenciac**
- **různé rukopisné styly a autoři**
- **regionální sounáležitost**

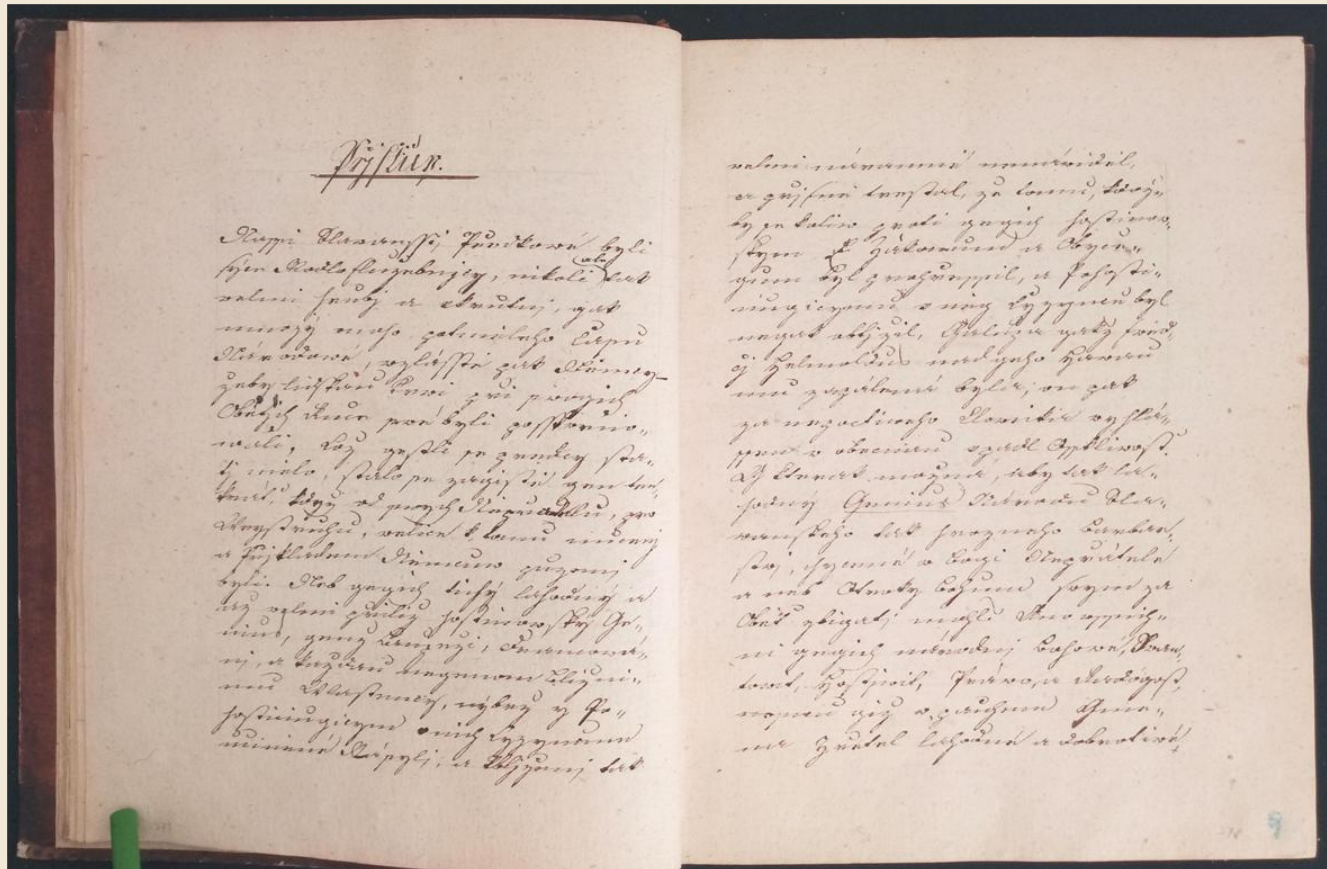
Tvorba modelu automatického rozpoznávání bohemikálního rukopisného textu v platformě Transkribus

**Josef Heřman Agapit Gallaš: 1756 – 1840;
Hranice, Přerov
(přírodní vědy, mytologie, národopis)**

**Otakar Jaroš: 1912 – 1943; Hranice
(vojenství)**

**František Polášek: 1757 – 1824; Příbor
(náboženské texty)**

Mýtické povídky o bozích a bohyních moravských Slovanů



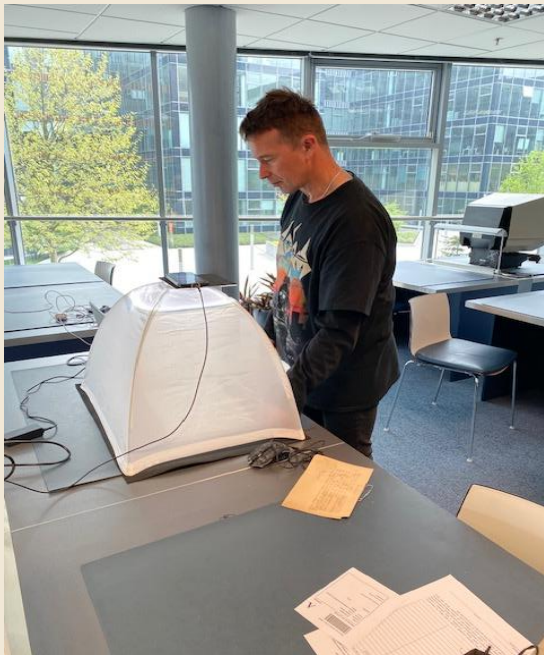
Výzvy:

- Špatná čitelnost původního textu.
- Množství pravopisných chyb.
- Blednutí atramentu.
- Různé písarské styly.

Východisko:

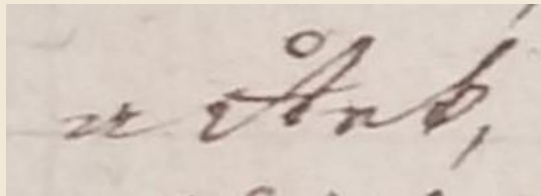
- Dílčí model s chybovostí 8,31 %
- Agregovaný model s chybovostí 6,55 %, který odstranil řadu nejednoznačností.

Mýtické povídky o bozích a bohyních moravských Slovanů

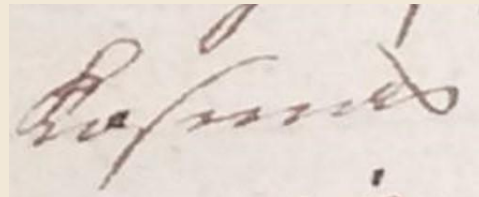


Zachraňujeme historické kulturní dědictví

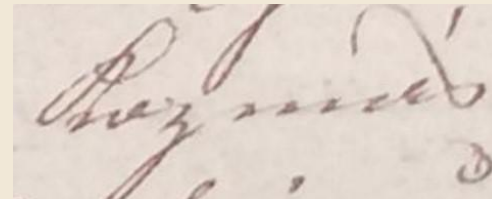
Mýtické povídky o bozích a bohyních moravských Slovanů



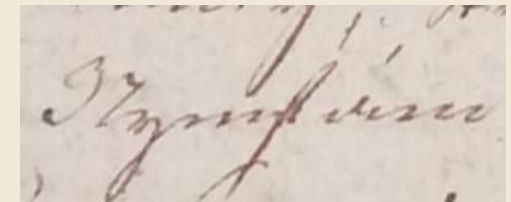
a Řek,



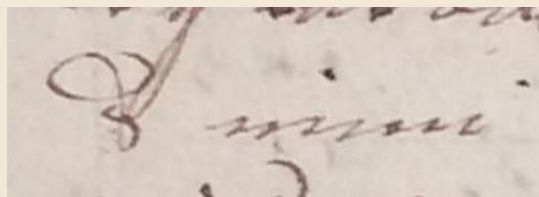
Kosmás



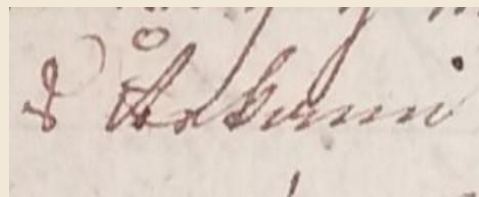
Kozmás



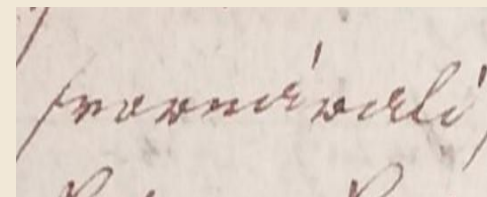
Nymfám



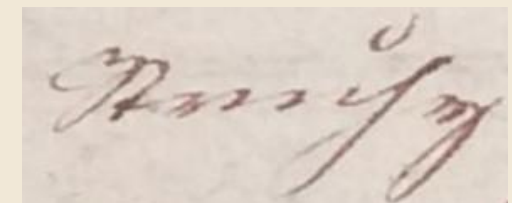
s nimi



s Řekami



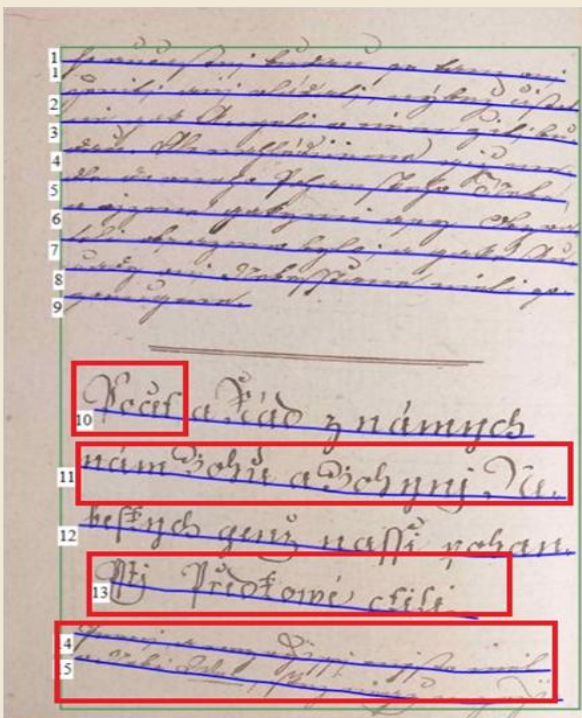
srownáwali ,



Strúhy

Mýtické povídky o bozích a bohyních moravských Slovanů

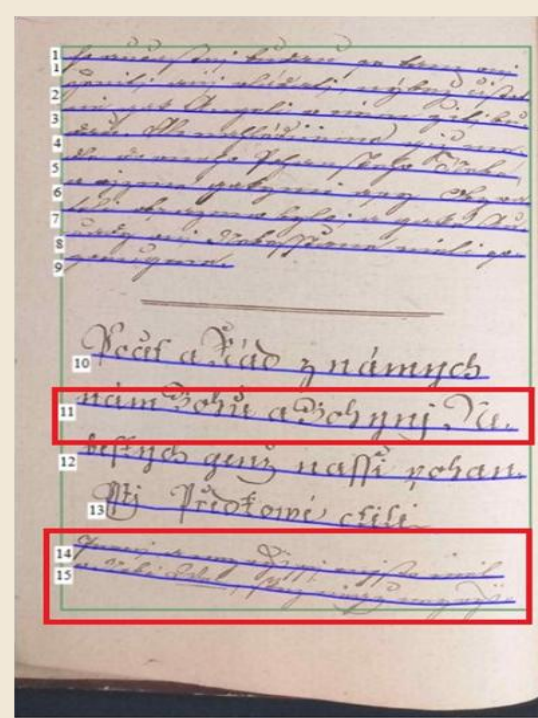
Region 1



1 ho aučaftnj búdaú fe tam onj
 2 ženitj anj vládatj, nýbrž číftot-
 3 ně gak Angeli w něm žitj bú-
 4 daú. Ale nahlédněme giž me-
 5 dle do onoho Pohánfkeho Nebe
 6 a wizme gakými afy Obywa-
 7 teli obfazeno bylo; a gake Au-
 8 řady onj Nebefstané měli, po-
 9 zorúgme.
 10 Počt a Řád známých
 11 nám Bohú a Bohynj, Te-
 12 befkych genž naffi pohan-
 13 f Pedowé cteli.
 14 Prnj a neygwýj mjftommel
 15 Nb bi Wel, frnz něgž eěwýš-

Model Mystic Absolut s chybovostí 8,31 %

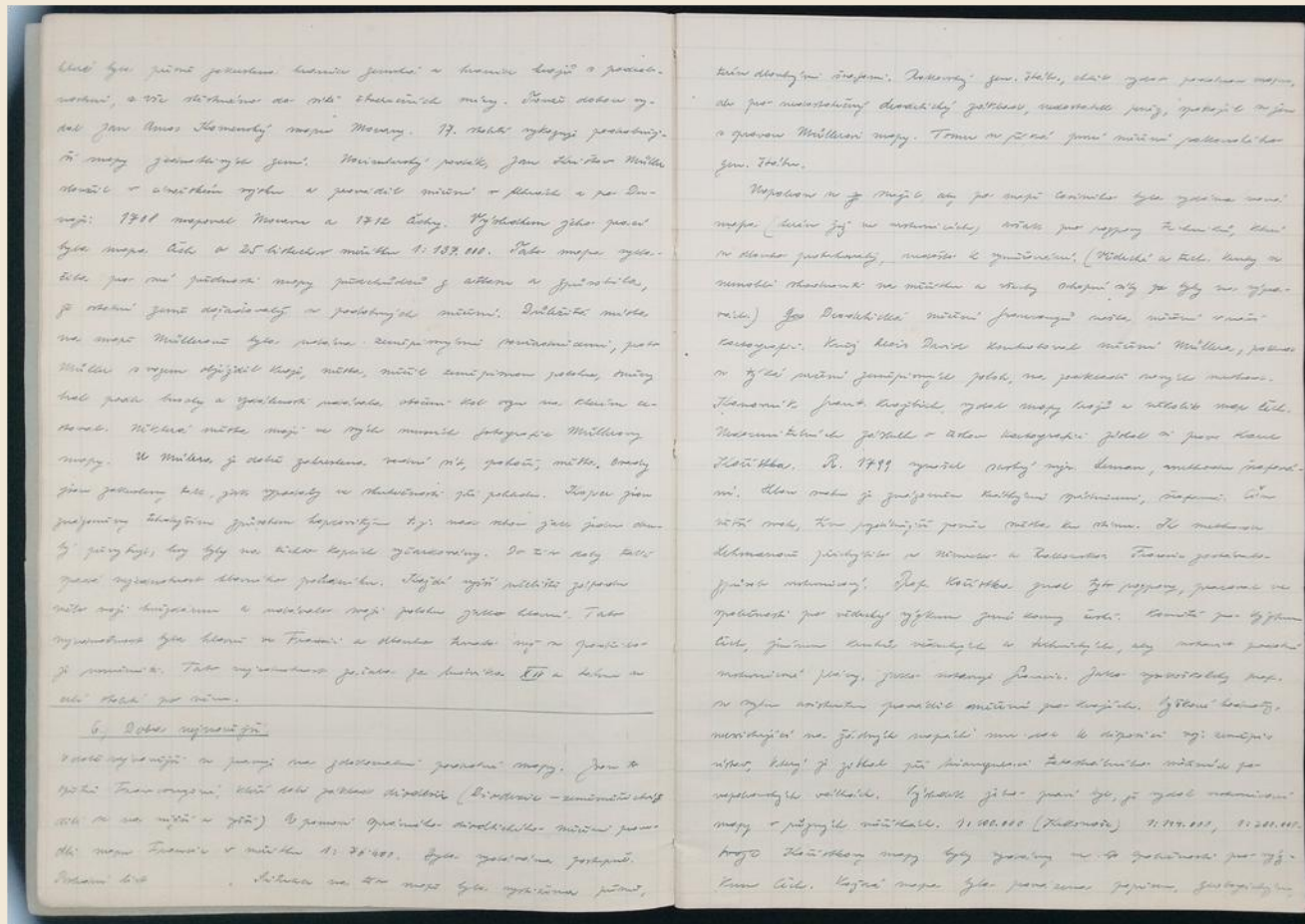
Region 1



1 ho aučaftnj búdaú fe tam anj
 2 ženitj anj vládatj, nýbrž číftot
 3 ně gak Angeli w něm žitj bú-
 4 daú. Ale nahlédněme aiž me-
 5 dle do onoho Pohánfkeho Nebe
 6 a wizme gakými afy Obywa-
 7 teli obfáženo bylo; a gake Au-
 8 řady onj Nebefstané měli, po-
 9 zorúgmne.
 10 Počet a Řád známých
 11 nám Bohú a Bohynj, Ae-
 12 befkych genž naffi pohan-
 13 ftj Předkowé clili.
 14 Prwnj a negwyšřfj mjfto měl
 15 w Nebi Wel frz něgž neywyš-

Agregovaný model s chybovostí 6,55 %

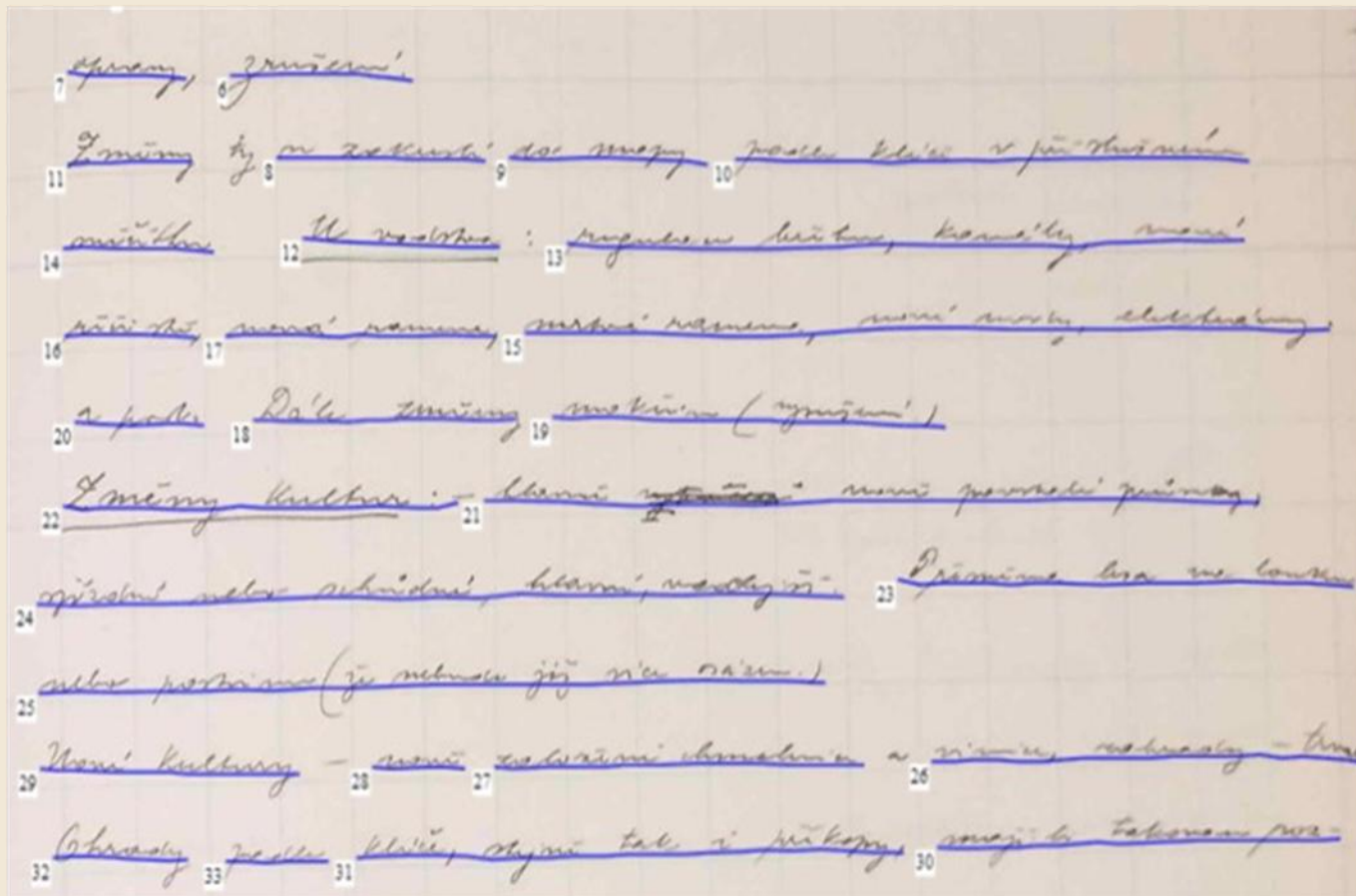
Otakar Jaroš: Nauka o terénu (školní sešit)



Výzvy:

- **Drobné písmo.**
- **Špatně čitelné.**
- **Degradace papíru.**
- **Dílčí vybledlost psaného textu.**
- **ČTVEREČKOVANÝ PAPÍR.**

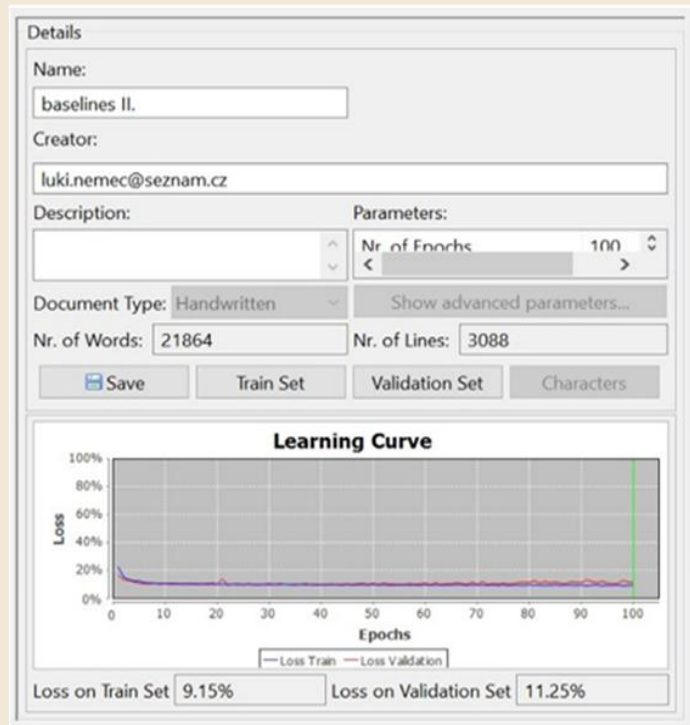
Otakar Jaroš: Nauka o terénu (školní sešit)



Výzvy:

- Drobné písmo.
- Špatně čitelné.
- Degradace papíru.
- Dílčí vybledlost psaného textu.
- Chybná segmentace textu
- ČTVEREČKOVANÝ PAPÍR.

Otakar Jaroš: Nauka o terénu (školní sešit)



Layout Analysis Configuration

Selected model:

Baseline detection settings

Minimal baseline length Param-Value:

Baseline accuracy threshold Param-Value:

Use trained separators Param-Value:

Max-dist for merging baselines Param-Value:

Image scaling Param-Value:

Region detection settings

Method

Nr of Text-regions Param-Value:

Východisko:

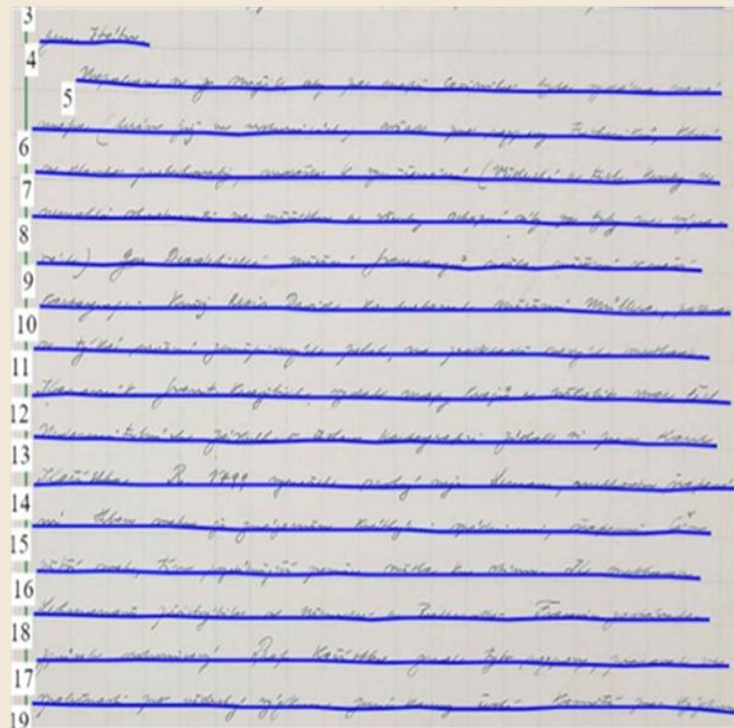
- Experiment s vytvořením modelu na segmentaci textových linek s uvedenými parametry.

Otakar Jaroš: Nauka o terénu (školní sešit)



Odborné vědecké soustředění v Jazernici...víkendový pracovní oddechový pobyt účastníků projektu.

Otakar Jaroš: Nauka o terénu (školní sešit)



4	gen. štábu.	::
5	Napoleon se je snažil, aby po mapě lasiného byla vydána nová	::
6	mapa (terén již ve vrstevnicích, avšak pro rozpory techniků, které	::
7	se dlouho protahovaly, nedošlo k vyměřování (Vědecké a tech. kruhy:	::
8	nemohli shodnouti na měřítku a všechny schopné síly pe byly na	::
9	vách.) Jes Deodetická měření francouzů našla měření v naší	::
10	kartografii. Kněz Alois David kontroloval měření Müllera, pokud	::
11	se týká určení zeměpisných poloh, na podkladě nových method.	::
12	Kanovník frant. krajich vydal mapy krajů a několik map čech.	::
13	Nedocenitelných zásluch o českou kartografii získal si prov koue	::
14	Kořístka. K. 1799 vynošel saský mjr. Lemon, methodu šrafová	::
15	ní. Klon svahu je znázorněn krátkými spádniceni, šrafamá. Čín	::
16	větší svah, tím rozdílnější poměr světla ku stínu. K methode	::
17	způsob vrstevnicový. Pof. Kořístka znal tyto rozpory, pracoval ve	::

- Ukázka experimentálního ověření naší hypotézy s potvrzením správnosti úsudku o nutnosti použití výchozího „pomocného“ modelu pro segmentaci textových linek v případě automatického rozpoznávání rukopisu na „čtverečkovaném“ podkladě.**

Agregovaný model Finale 2.0

Celkové parametry modelu

Details

Name: Language:

Creator:

Description:

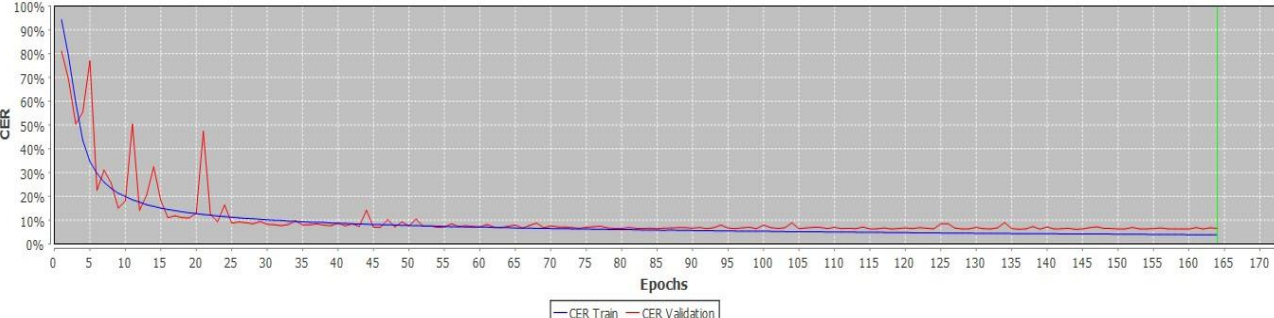
Parameters:

Max epochs	250
Early stopping	20
Epochs trained	164
Learning rate	0.0003
Batch size	24

Document Type:

Nr. of Words: Nr. of Lines:

Learning Curve



CER on Train Set: CER on Validation Set:

Doba trénování: 11 hod. 26 min.

Počet epoch: 164 (bylo nastaveno 250)

Počet stran v kvalitě GT: 514

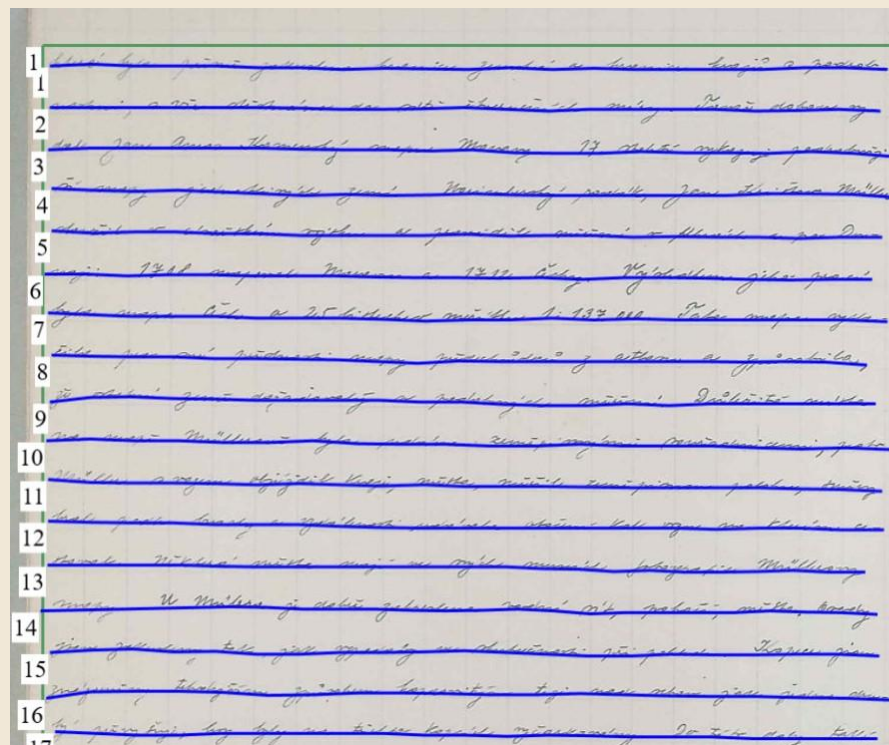
Počet slov trénovací sady / ověřovací sady: 87.945 / 10.808

Počet řádků trénovací sady / ověřovací sady: 14.530 / 1.785

Celková chybovost modelu: 6,56 %

Otakar Jaroš: Nauka o terénu (školní sešit)

Ukázka použití modelu Finale 2.0



Region 1

1 které byla přesně zakreslena hranice zemská a hranice krajů s podrob-
 2 nostmi, s vše stětnáno do sítě čtverečních míry. Tonže dobou vy-
 3 dal Jan Amos Komenský mapu Moravy. 17. stobetí vykazuje podrobněj.
 4 sí mapy jednotlivých zemí. Norimberský rodák, Jan Křištov Müller
 5 sloužil v čísařském vojsku a prováděl měření v Ulhrách a po Dn-
 6 naji 1708 mapoval Moravu a 1712 Čechy. Výsledkem jeho prací
 7 byla mapa lech o 25 listech v měřítku 1: 137.000. Tato mapa vytla-
 8 čila pro své přednosti mapy předchůdců z atlasu a způsobila
 9 že ostatní země dožadovaly se podobných měření. Důležitá místa
 10 na mapě Müllerově byla udána zeměpisnými souřadnicemi, proto
 11 Müller s vozem objížděl kraje, města, měřil zeměpisnou polohu, směry
 12 bral podle busoly a vzdálenosti udávala otočení kol vozu na kterém ce-
 13 stoval Některá města mají ve svých muscích Jstografie Müllerovy
 14 mapy U můlera je dobře zakreslena vodní síť, pohoří, města. osady
 15 jsou zakresleny tak, jak vypodaly ve skutečnosti při pohledu. Kopce jsou

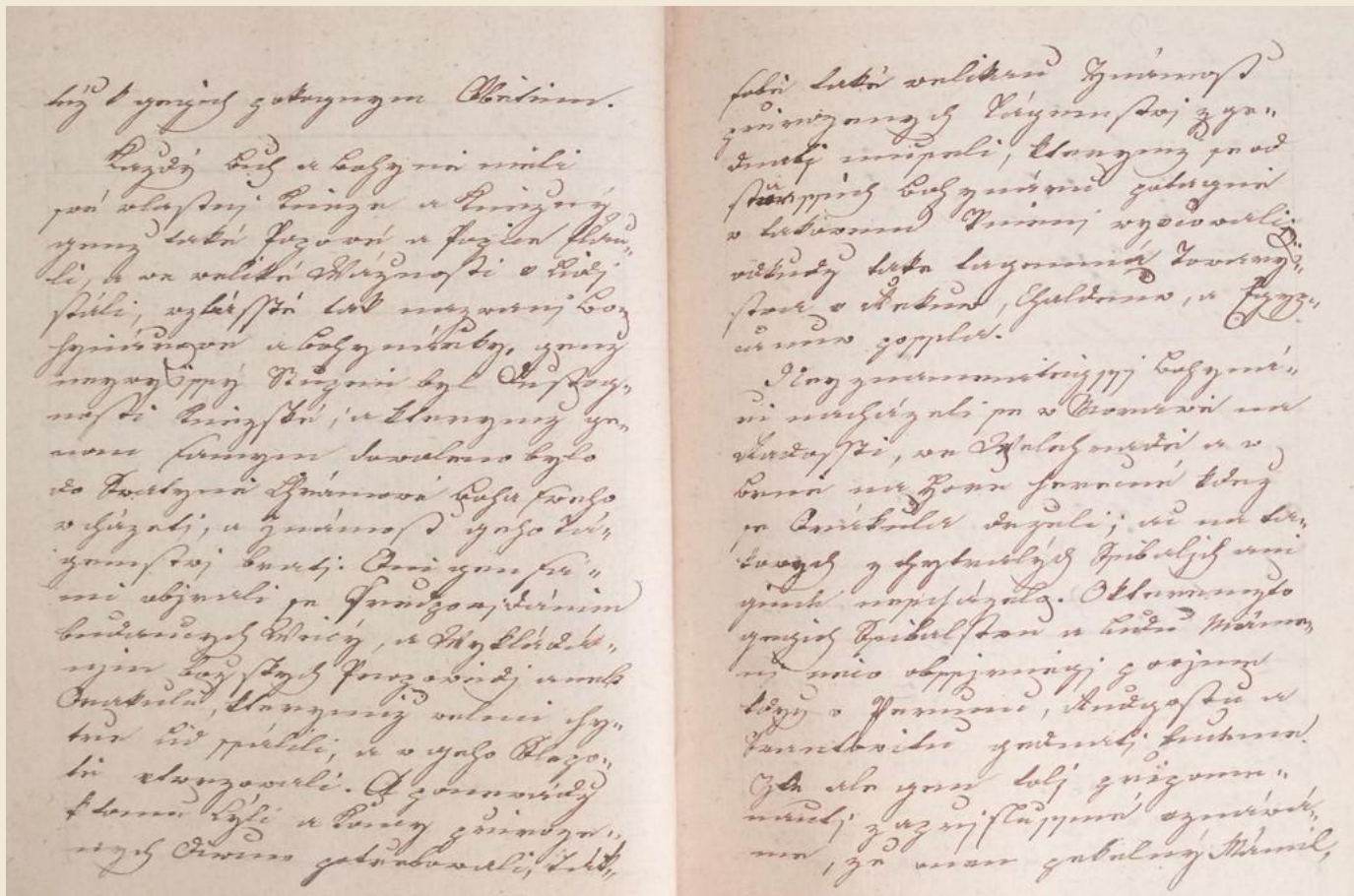
Josef Heřman Agapit Gallaš : *Mystické povídky o bozích a bohyních*

Ukázka použití modelu Finale 2.0

	Region 1
1	1 a neb Pop pohodlně seděj a wffe což se
2	2 w Chrámě bliže nj djlo widěj a flyffe-
3	3 tj mohlj neb Uffi gegj byly tak mjftr-
4	4 ně Sformowané že y flabo proně-
5	5 ffená Sliwa zřetedlně od něg flyfa-
6	6 ná bytj mohla. Proftředkem onoho
7	7 Túlipanta, a neb Zwonečka Kwjtk
8	8 genž se mezy geho Rtami wynachá-
9	9 zel, mohl sedjcy wněm bohynář ne
10	10 gen wffecho widěj nýbrž y také
11	11 aúftná Orakúla wydawatj čímž
12	12 powěrečnj Pohané, neznage Popúw
13	13 fwych keykljřkych Proftředkúw a
14	14 Sfalby w Slepotě fwogj welmi upew-
15	15 nen byl. Zdaliž wfifichni Slawani
16	

Agregovaný model Finale 2.0

Ukázky rukopisných vzorků:



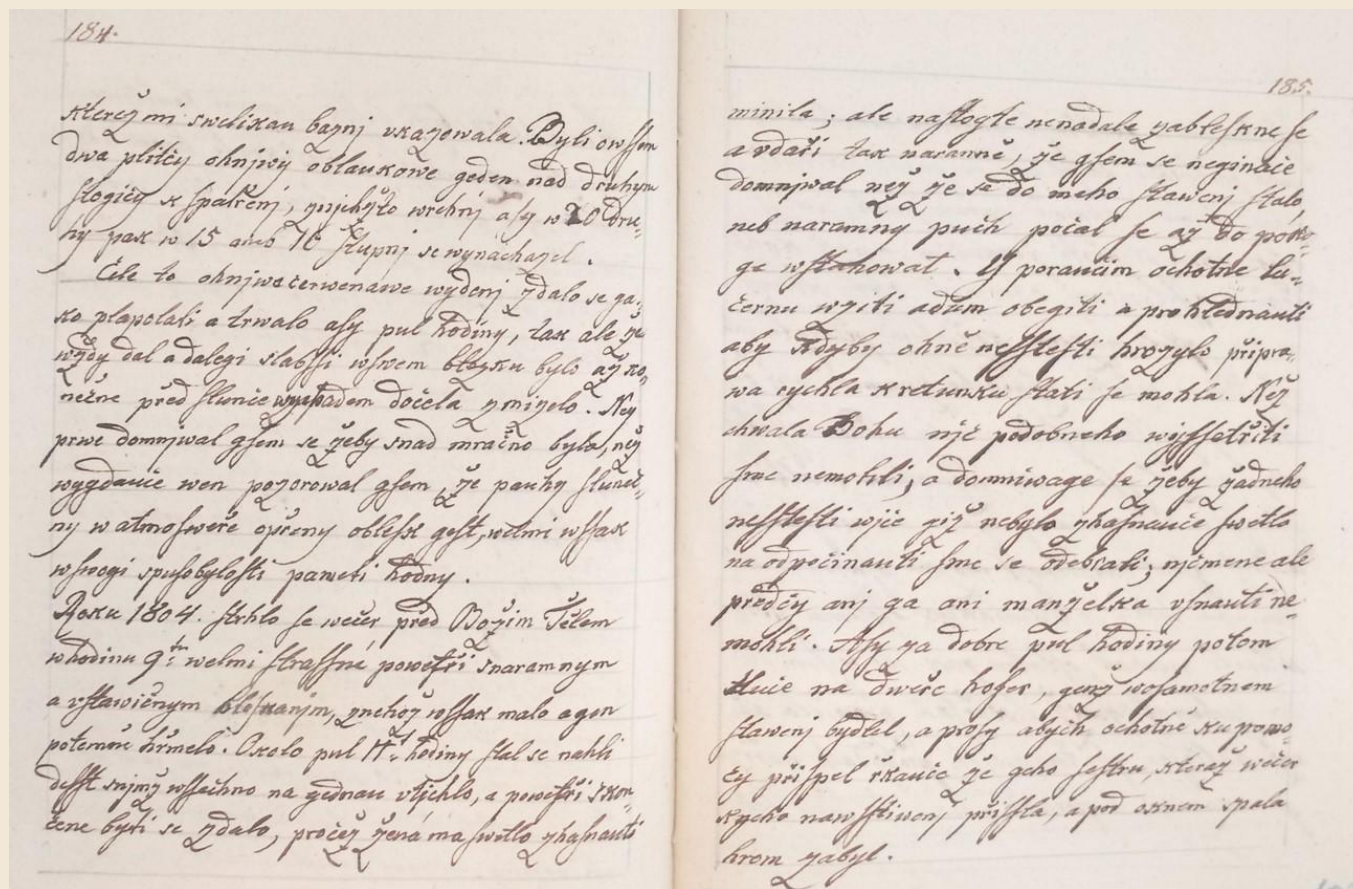
Josef Heřman Agapit Gallaš Mythické povídky o bozích a bohyních moravských Slovanů

Uloženo v:

**MZA Brno, G 11, sign. 838, čeština,
papír, rukopisná kniha, originál,
vázáno v tvrdých, polokožených
deskách, šířka 195 mm, výška 250
mm, stopy po pův. pag., starší fol.
125; stará sign.: Schr. 224, pův. 287,
červ.**

Agregovaný model Finale 2.0

Ukázky rukopisných vzorků:



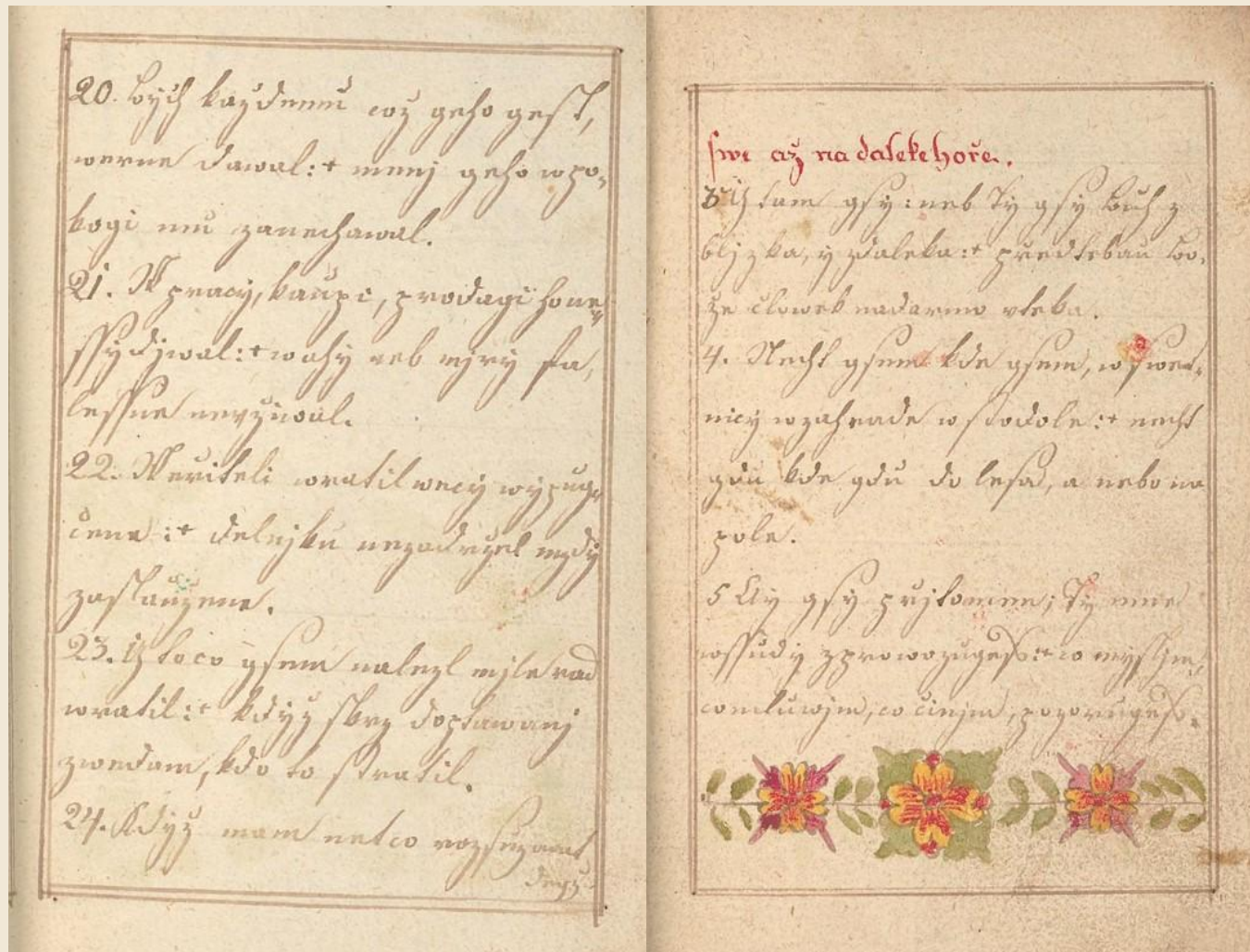
Josef Heřman Agapit Gallaš Fyzické památky města Hranice a okolí.

Uloženo v:

**MZA Brno, G 11 Sbirka rukopisů
Františkova muzea Brno, sign. 658,
čeština, latina, papír, rukopisná
kniha, originál, vázáno v tvrdých
deskách, šířka 215 mm, výška 270
mm, pův. pag. 236, nová fol. 128;
stará sign.: Schr. 223, pův. 288, červ.**

Agregovaný model Finale 2.0

Ukázky rukopisných vzorků:



František Polášek
Pravé poznání boha...

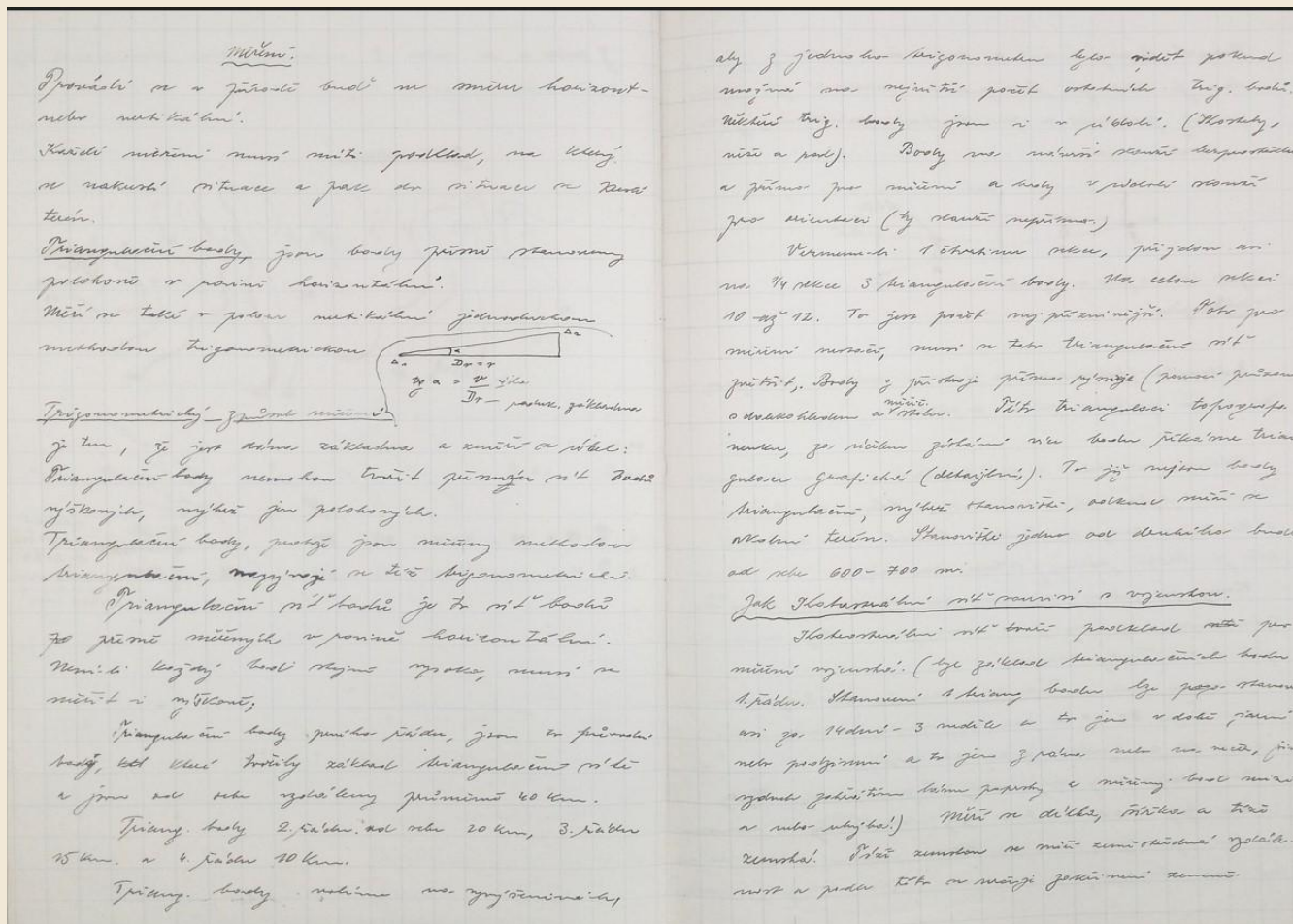
Uloženo v:

**Sbírková knihovna Vlastivědného
muzea v Olomouci, Sign. K-24073,
dostupný z:**

[https://new.manuscriptorium.com/
hub/catalog/default/detail/single/
manuscriptorium%7CVMO-VMOK
24073_0U6ABL2-cs?lang=cs](https://new.manuscriptorium.com/hub/catalog/default/detail/single/manuscriptorium%7CVMO-VMOK_24073_0U6ABL2-cs?lang=cs)

Agregovaný model Finale 2.0

Ukázky rukopisných vzorků:



Otakar Jaroš
Nauka o terénu, školní sešit,
čtverečkovaný papír

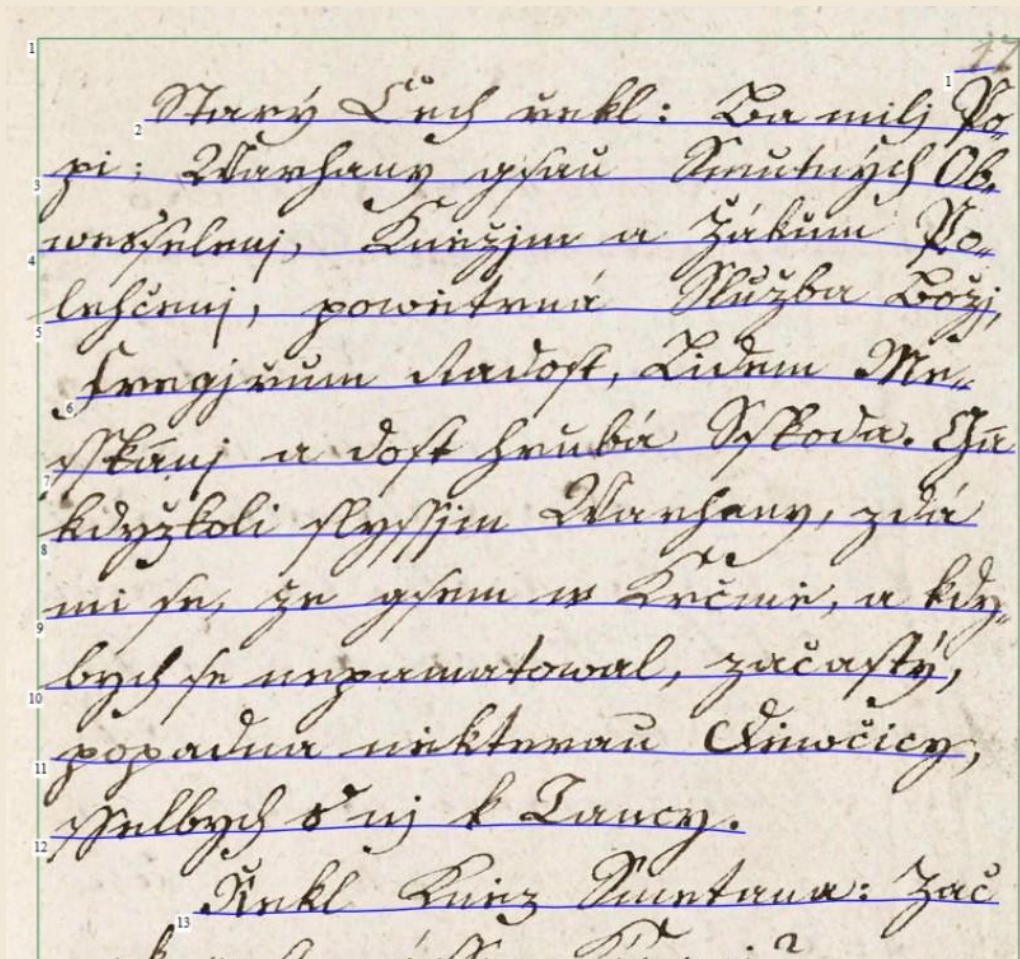
Uloženo v:

Historická expozice 71. mech.
praporu „Sibiřského“ v Hranicích.
Zapůjčeno z pozůstalostní sbírky
rodiny.

Digitalizát vytvořen s laskavým
svolením kurátora muzea nrtn.
Radima Cába.

Agregovaný model Finale 2.0

Ukázka použití modelu na textu, který nebyl trénován



Region 1

- 1 17
- 2 Stary Čech řekl: Ba milj Po-
- 3 pi: Wárhany gfaú Gmútných Ob-
- 4 webfelegj, Kněžjm a Žákúm Po-
- 5 lehčenj, powětrú Slúžba Božj,
- 6 Přepjřúm Radoft, Lidem Me-
- 7 řkánj a doft hrúbá Sřkoda. Gu-
- 8 kdyžloli flyřřjm Wárheny, zdá
- 9 mi fe že gfm w Krčině, a kdy-
- 10 bych fe nepamatowal, začařky
- 11 popadna některou děwčicy
- 12 řelbych s nj k Čaúcy.
- 13 Řekl Kněz búletanx : Zač
- 14 pak geft náffe Kázágj?
- 15 Stary Čech řekl: Rázánj wá-

Zdroj:

**Rozličné písně
starožitné**

Uloženo v:

**Moravská zemská
knihovna v Brně
RKP-0048.022; Rkp 59**

AUTOMATICKÁ
TRANSKRIPCE
POMOCÍ PLATFORMY
TRANSKRIBUS

VOJTĚCH ŘÍHA



Pozehnana brádrů o křesťově,
Křesťů a nerozdilna křesťice
Božka, Duch Otce, Duch Dín
Duch Duch Dvati Ami v wž

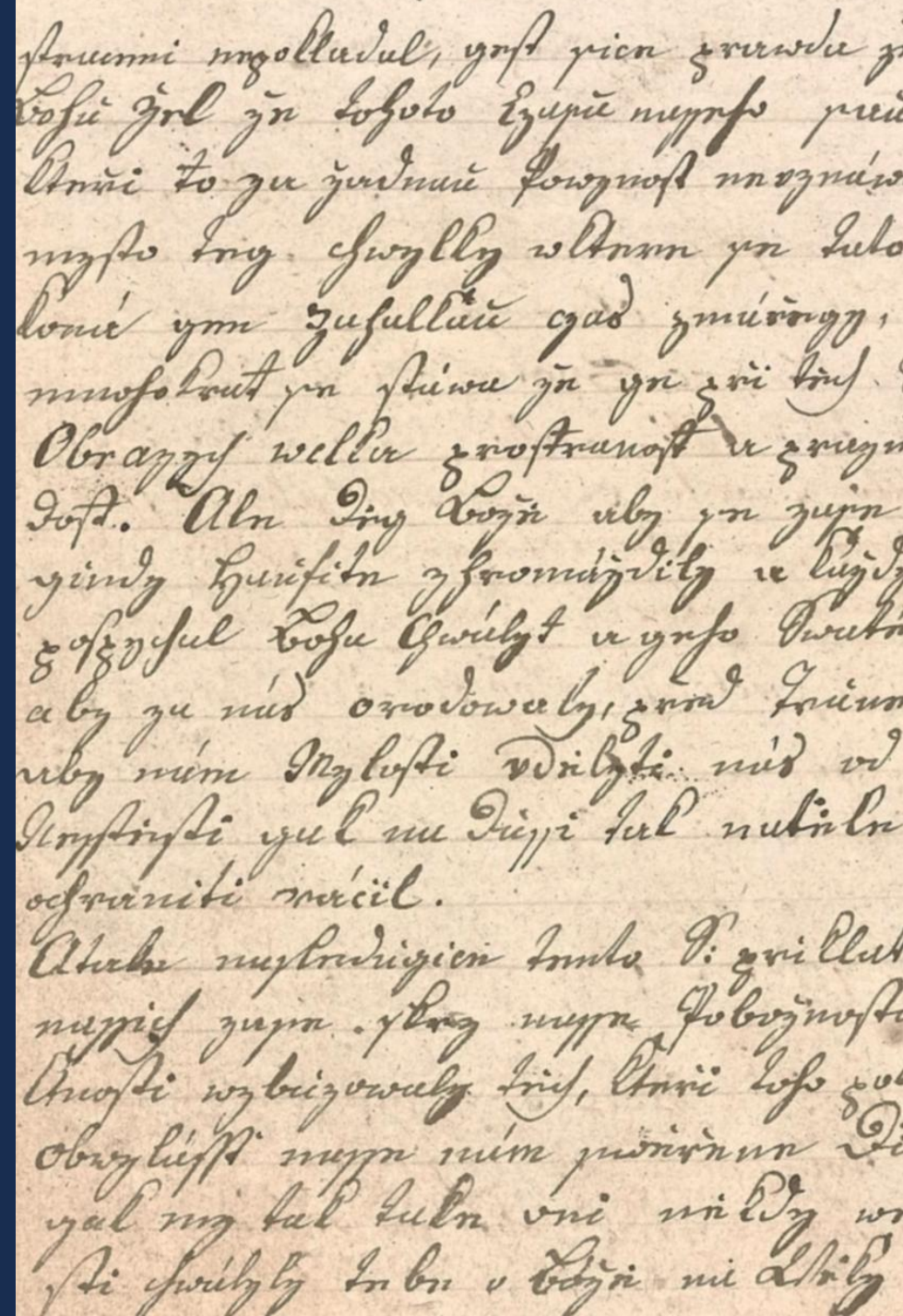
sti, a nešodnosti se nekoril, a neponizoval.
Ach Bože Troj jediný! tobě patří čest,
sláva a dílo činění. Ale ty Bože! v sa,
me pochvále oblybeni nemáš, Ty od nás
žádáš dětinství větění, ty od nás žádáš
lásku, podanost, a důvěrnost. *fošle Bože*
my ti dnes lylo Avěti měni a me vědaū

Paní Bohu služba wděčná
A křesťicím Radost věčná.
Křesťicím modlitbě, aby ona
tjm spisegi a bezpečnegi do nebe k Bo-
hū se man ozi, maha dno s vidi a vřidati
zdravugi tebe mradosny den
narozeny tveho zaviteg knom
oneg světe gšny panenko potěše-
ny a radosti srdce meho tebe

MODEL AGREG-8

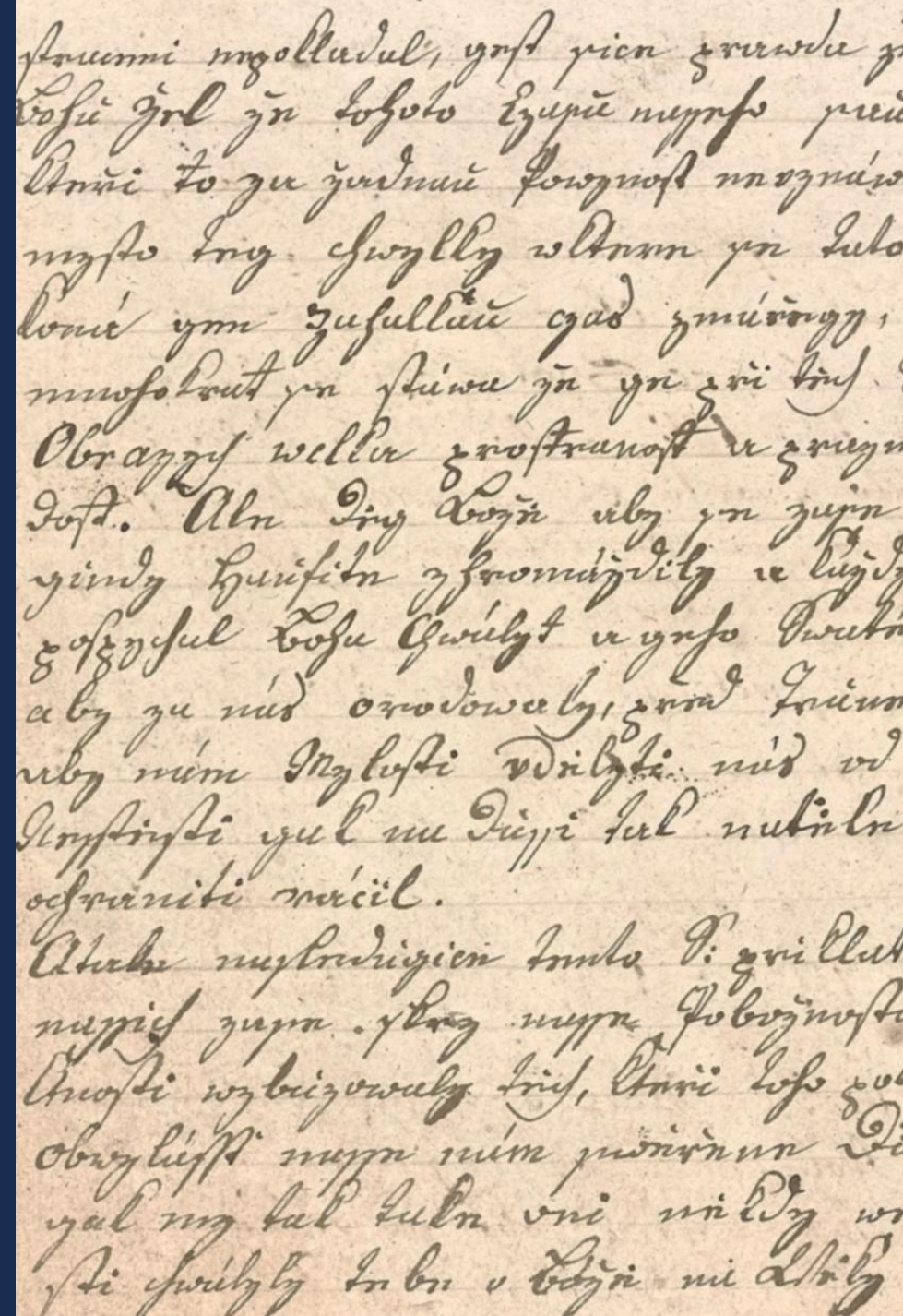
Model Agreg-8

- Společné rysy dokumentů
 - Století 18.-19. století
 - Tematika modliteb a katolických písní
 - Podobnost písma – cílem je postupně rozšiřovat data pro dosažení větší rozmanitosti znaků
 - Licence Creative Commons (BY-NC-SA)



Model Agreg-8

- HTR model kombinující 5 rukopisných dokumentů
 - Česká modlitební kniha
 - Cesta Svatocellenská
 - Radostná cesta
 - Modlitby, písně a litanie
 - Modlitební knížka



1 . Česká modlitební kniha

- Datace: 1733-1766
- Instituce: Muzeum Jindřichohradecka
- Signatura: RK 071
- Rozměry: 14 cm x 10,5 cm
- **Problematika zasahování do susedících řádků**

chem брвату на рлне на була а
у захаку даки ржохосту
була се леми доми брасени се ст
оржже welmi тосне роздхавени
де зазни талеру ссич mich асп
бтале спрасени ме ду ссе тиску сжат
х садене жад те члел роздхави
кжа ден аву с тимне дедене жатр
жати а асила патен те чду талел
онудентрехо брватохо патстир
итебе ластаме роздхави ги отан
задо стир а семте бне дн едл стуба се

1. Če

- Data
- Instit
- Signa
- Rozn
- Prob
- sous

malu a odem zdalu a vnutne zad:
nu vice gemu odlauiditi neto zl
o^{hu} z rodice ney sveteg sluzetena g^o g^o d^o n:
e sluzodne zobetowanij manje ney
met sluzmi mylostij obdazila ruzj:

chem svalym naplnena byla a

anirxojostij
brasenij se to
rozdnawenij n
tich mich a spr
dusse ti sluzekat
tel rozdnawij
ie dedenekat pr
en tech d^o tal
ateho nam sluzim
namidi o tam

zadosim g^o semte b^o neg^o sveteg sluzbase

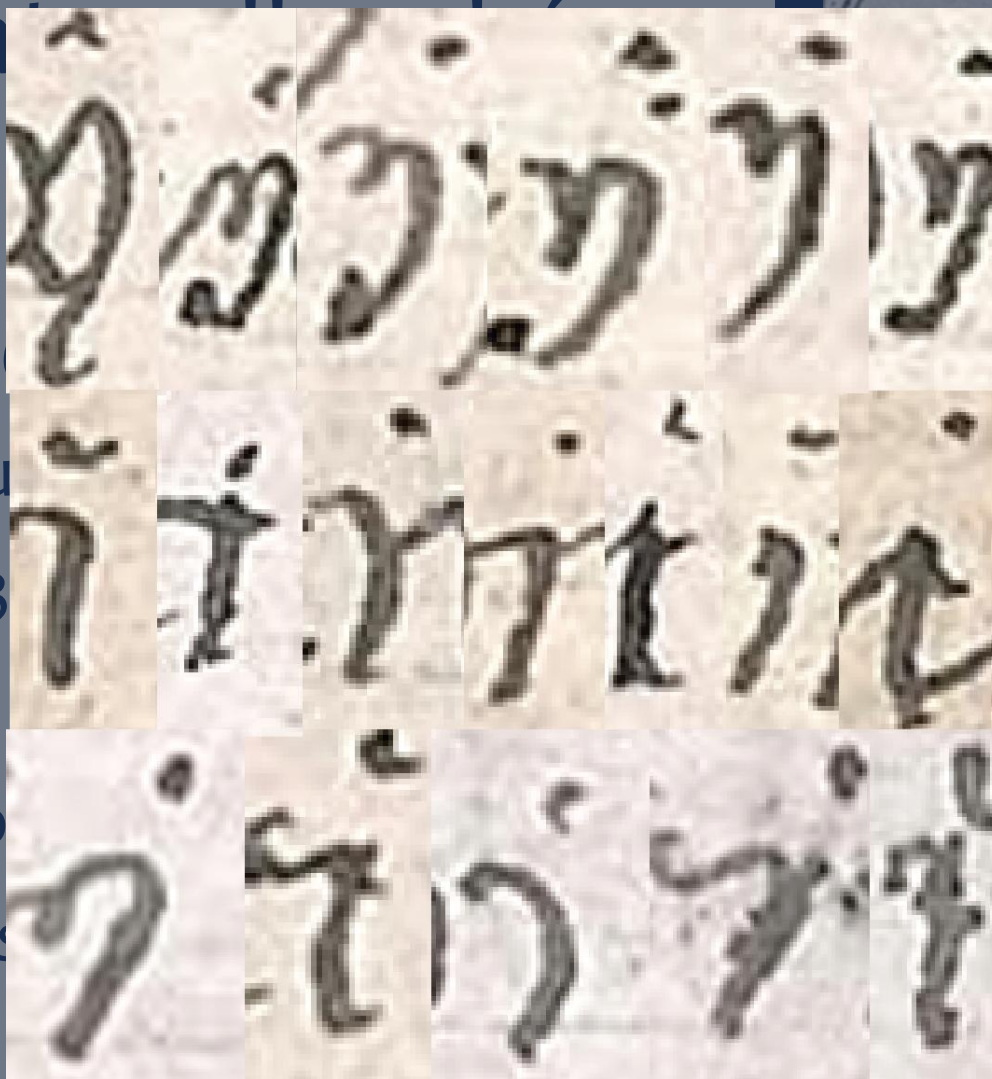
2 . Cesta Svatocellenská

- Datace: 1733-1766
- Instituce: Muzeum Jindřichohradecka
- Signatura: RK 037
- Rozměry : 18 cm x 11,5 cm
- Problematika podobnosti i/y
- Používání obrovských iniciál (segm.)

Krista a Jani Swata: Kteraz
žadneho neopaustiti a zadnim
nepohrdass w zlydni o Janne
Dobrotiwa Oczima Ktilosti
a wiprod nam utweho mileh
dina wssch hrstehuro odpustit
ni, abichom pobožnau Ktilo
twe Swate Bozeti sobie pr
pominali, Potom wiecznetjo
Blahostawenstwi dosahli skrz
stiedrost Jana nasseho Bezisse
Krista czechoz ysi ti Janne Kreci
porodila Ktemiz v Dohem otcem
j v duchem Swatim ziw gest ak
aluce w trojici dokonali. Zú

2. Cesta Svatých

- Datace: 1733-176
- Instituce: Muzeu
- Signatura: RK 03
- Rozměry : 18 cm
- Problematika po
- Používání obrov



Fixilla a Jani Sweta: Kteraz
žadného neopauštil a žadním
ordass w zřetledni o Janmo
otowa Oczima Ktilosti
proa nam utweho mlety
wofech hrstehuro odpustit
bichom pobožnau Ktil
Swate Bozeti sobie p
inani, Potom wiecznetjo
ostawentwoi dosahli hrz
kost Jana nasseho Gezisse
a czehoz ysi ti Janmo Kreci
ila Ktemiz i Dohem otcem
i duchem Swatim ziw gest ak
alucze w trogici dokonali. Zuh

2. Ce

■ Datac

■ Institu

■ Signat

■ Rozm

■ Proble

■ Použí

ti vobrygi wylkiffet
We Gmenú
Trogyce,
tebe stuzze Ustia 20

Fixilla a Jani dwieta: Kteraz
Kadmo...
...stiff a zadnim
...edni o Janno
...sima Ktilosti
...utweho mlety
...stchuro odpustit
...oboznau Ktilo
...ocseti sobie p
...om wiecnetjo
...oi dosahli stuz
...nasseho Gezisse
...ti Janno Kreci
...Dohem otcem
...m ziro gest ak
...abuge w trogyce dokonali Zuh

3. Radostná cesta

- Datace: 1829-1884
- Instituce: Moravské zemské muzeum
- Signatura: ST 2272
- Rozměry: 19,5 cm x 16,5 cm
- Různost při psaní diakritiky

la opūstena, pro bolest do Mado
padnauē, skoro miltwa sy žustala
O nejdobrotiwější Jmeno Ma
Prye tvo Owatē opūstēni, neopau
mne, w mých zarmutených, a prot
stwi, ne neč mne padnāiti, i
yalau: málo myslu, neb za
odwrāt ~~od mne~~ nemoc, Inor,
nebezpečné wod rozlyti, študly
Ojē, Laurky, Dromoly, a lru p
aw rsech žli nestěti.
O nejdobrotiwější, Opūstena
Bozi! Prye tvo O: opūstēnj, ne
Prye mne w mōm časem žiro
Orozlastni ale neopauštāij m
při loně žirota mezo, neopa
mniš. Dž na širletné hostely 16
bada, od cledo Owa opūst
plijy bolest, ležeti bada, neop
mne Dž šračeny, gazy ^um
nejswetější Jmeno wyec ne

3. Rado

Büch Büch Büch

Pa ritz

utunula

Lūdiz

nüle

Beni

assemohaucy

Putomahi

zhuxi

- D
- Instituce: Moravské zemské muzeum
- Signatura: S
- Rozměry: 19 m
- Různost při psaní diakritiky

4 . Modlitby, písně a litanie

- Datace: okolo roku 1826
- Instituce: Moravské zemské muzeum
- Signatura: ST 2193
- Rozměry: 22 cm x 18 cm
- Různé druhy písma v jednom dokumentu (unclear tag)

twym, we wyznani prawe
stawu Trogyce wieczne po
a w mocy Wlechnosti, Ge
se skaneti, prosyme tebe, a
wyry stalosti, proti wssim
wenswym opatreni byly, st
Gezisse Krysta pana naszego
s teba u ziw gest, a bral u
gednoti Duchu swateho r
Duh na weky weka Am

3. Otcinnym a z z druwat Mar
Dlawa Oty z Pym z Dufu
z: Galozj byla nu gwelka nje
wzlicy wy nu Libly Wnili

Sidzj na wile Kuznart l O
Janina: drosim w kofu:
Dlawa Oty z ul Otcinnym. O
z z druwat z druwat Merga:

5 . Modlitební knížka

- Datace: 1700-1750
- Instituce: Muzeum Jindřichohradecka
- Signatura: RK 087
- Rozměry: 22 cm x 16 cm
- Průsvitnost textu protilehlých stran

Starší posvěcení tvaú štoci
Kozšcau pěstovaní, a k giste
naději Spasení našeho přite
men byl, deq snilost abychom
na tebe, kterého zde pod Spúš
baú Chleba vidjme, w Nebesyc
wěčně, patřili, a tam se s tebáú
wěčně, radovali, Amen.

Kněz Kalich pozdvihúge

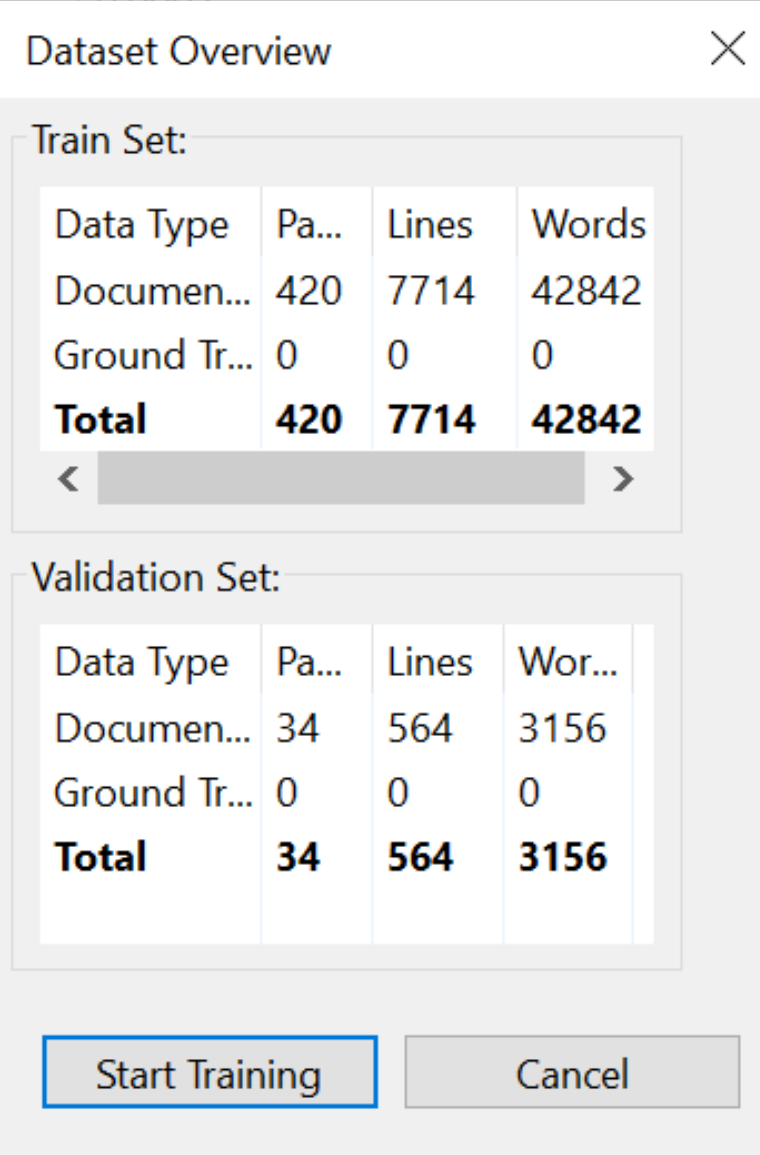
3 Jan Krysta Pána nevíthná kren
teče, a gíe a gíe

Modlitba

Dřevotá kren, která
gíe s Nienšvétégšých Kán
Spasitele meho k obmitj wšech
mých hříchů w ogně, tekla, ob
meg wšechnú šecístotú hříchů
mých, a mně očistěnaú, k wěčné
mu životú zachoweg. Dřevotá
Gežíšy Kryste! genz gíe chěl

Model Agreg-8

- Počet trénovacích cyklů: 250
- Délka trénování: 21h 37m
- Celkový počet slov: 45998
 - *Trénovací sada*: 42842
 - *Validační sada*: 3156
- Počet stran GT: 454
- Výška řádku: 140 px



Dataset Overview

Train Set:

Data Type	Pa...	Lines	Words
Documen...	420	7714	42842
Ground Tr...	0	0	0
Total	420	7714	42842

Validation Set:

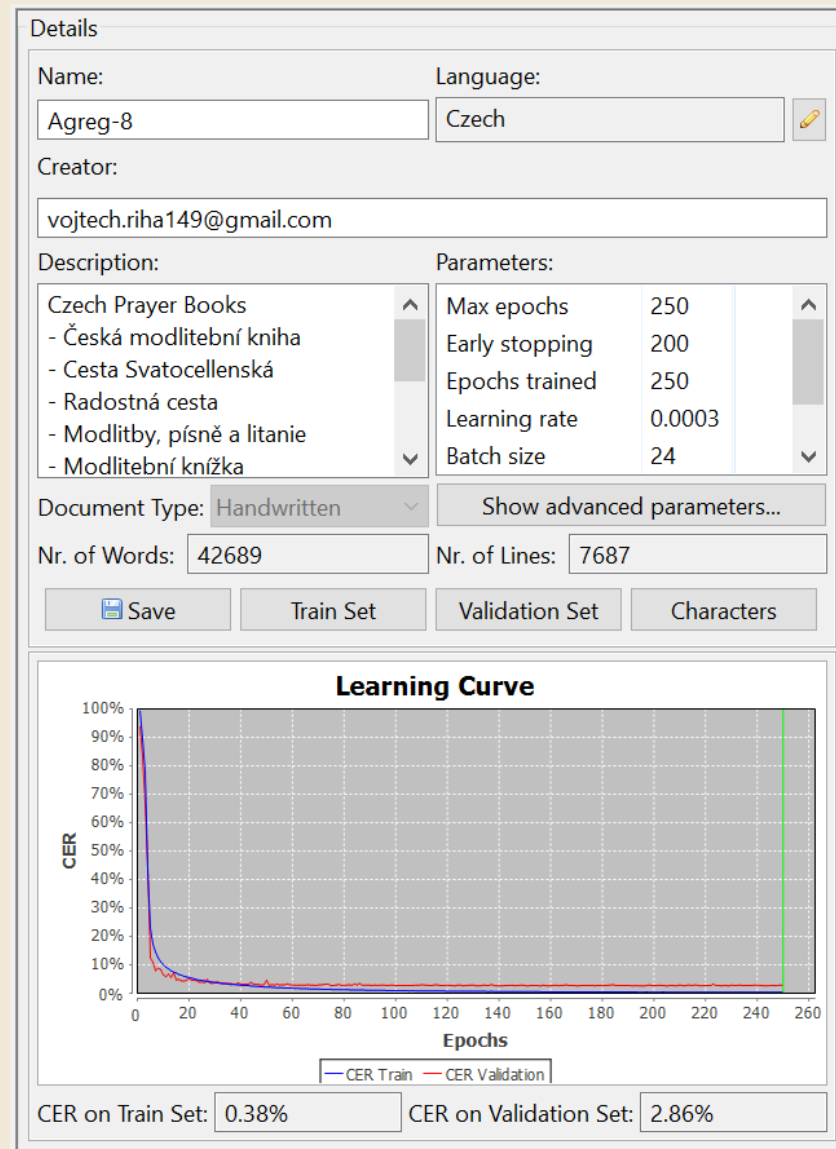
Data Type	Pa...	Lines	Wor...
Documen...	34	564	3156
Ground Tr...	0	0	0
Total	34	564	3156

Start Training Cancel

Model Agreg-8

■ Výsledek

- CER na trénovací sadě: **0,38%**
- CER na validační sadě: **2,86%**
- nejlepší epocha CER: **2,60%**
- nejlepší epocha WER: **16,57%**



Model Agreg-8

Model 'Agreg-8'

Name: Agreg-8 Language: Czech

Creator: vojtech.riha149@gmail.com

Description: Czech Prayer Books
- Česká modlitební kniha
- Cesta Svatocelestenská
- Radostná cesta
- Modlitby, písně a litanie
- Modlitební knížka
Doplnění verze 7 (20 stran 5. dokumentu)
Korekce

Parameters:

Max epochs	250
Early stopping	200
Epochs trained	250
Learning rate	0.0003
Batch size	24
Normalized height	140
Omitted Tags	unclear

Show advanced parameters...

Document Type: Handwritten

Nr. of Words: 42689 Nr. of Lines: 7687

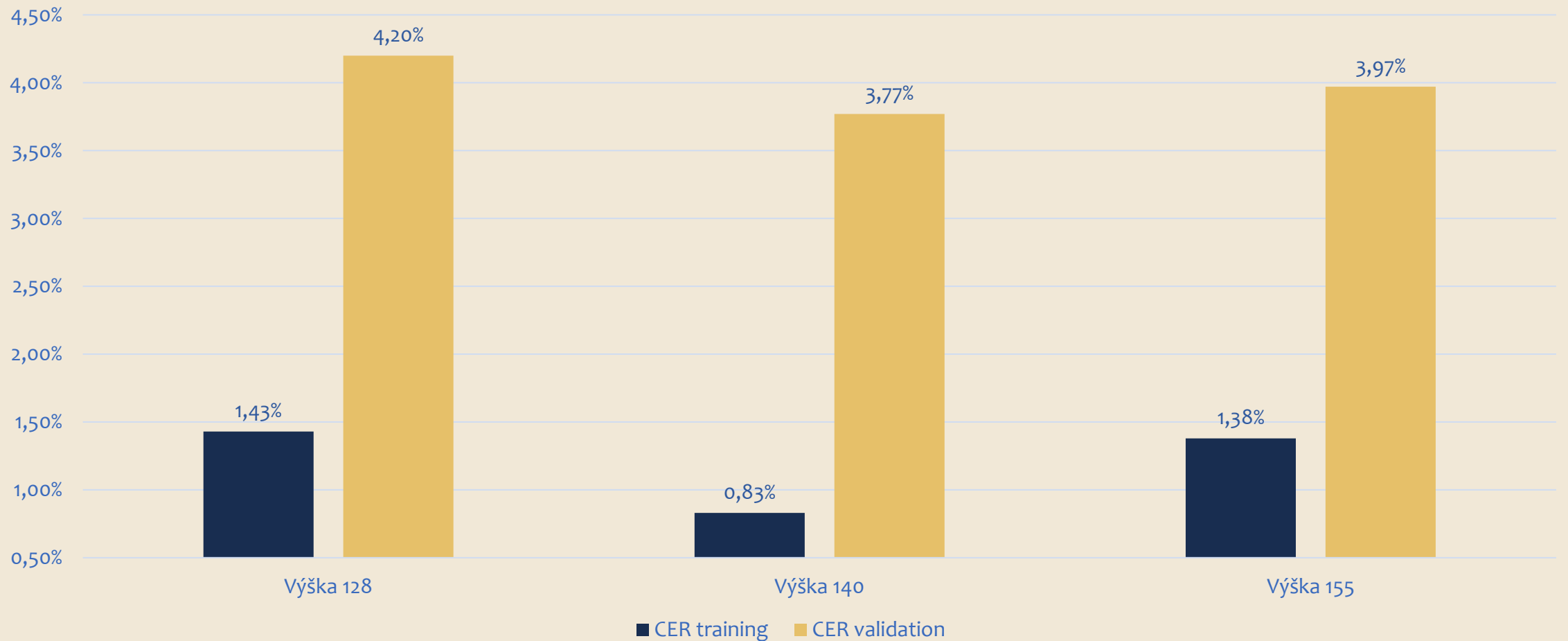
Save Train Set Validation Set Characters

Learning Curve

CER on Train Set: 0.38% CER on Validation Set: 2.86%

Korekce výšky řádku u modelu ČMK-70

Úspěšnost dílčích modelů

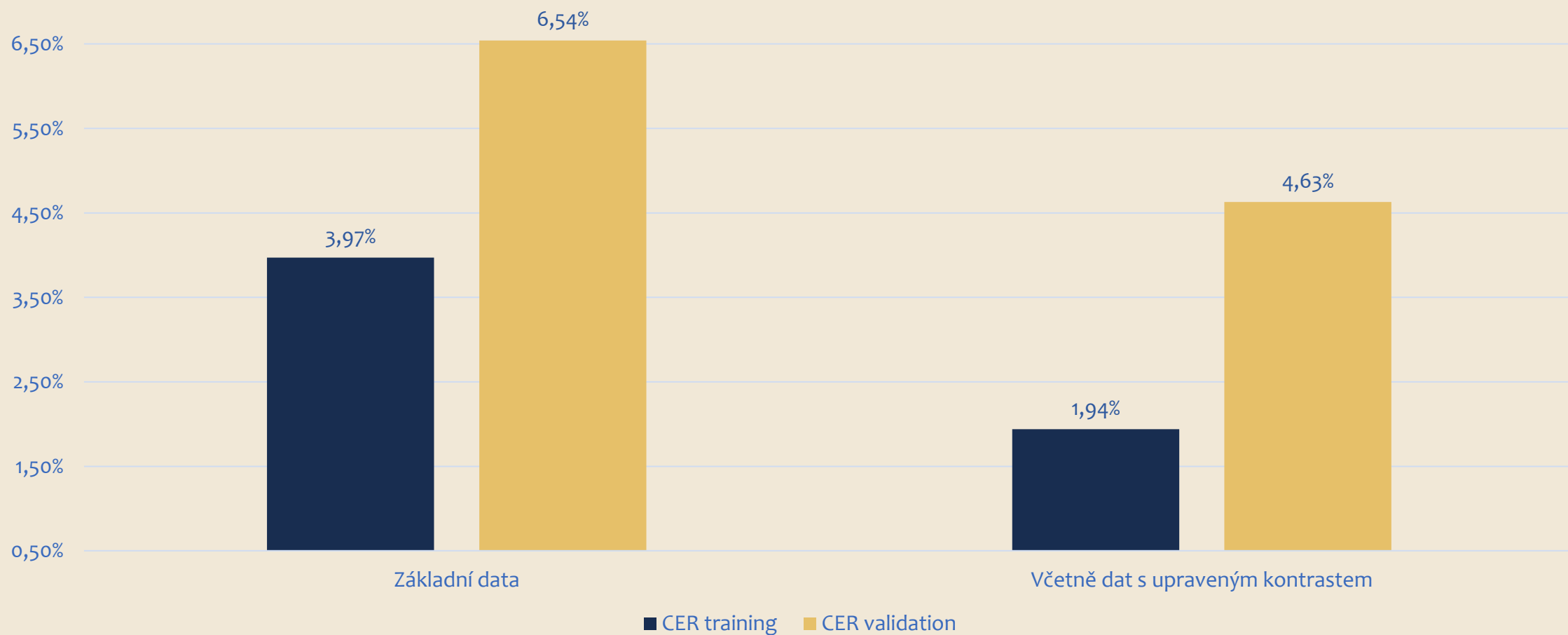


Augmentace obrazových dat

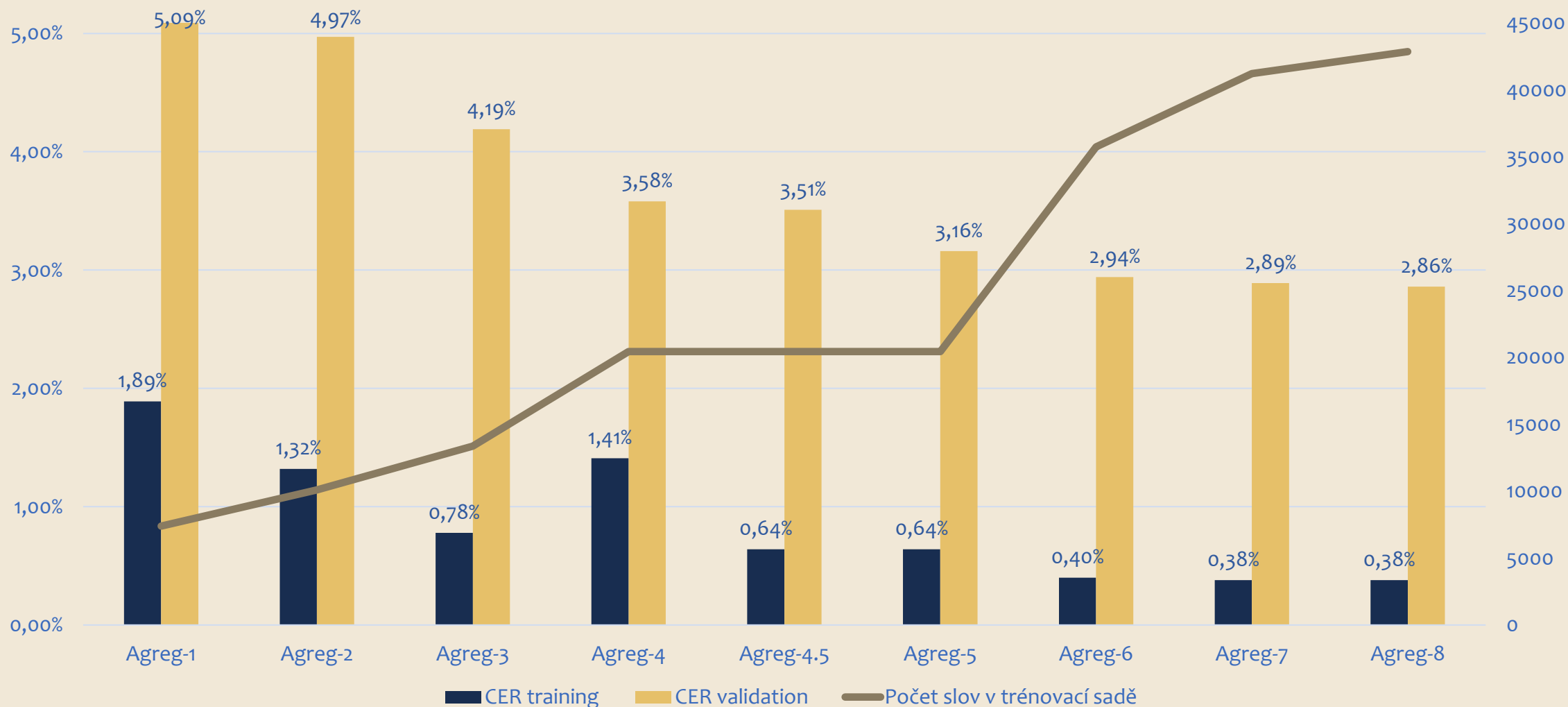
1. Geometrické transformace
 - *Rotace, škálování*
2. Fotometrické úpravy
 - **Změna kontrastu (+40, -20), jasu (+10, -10)**
 - *ostrosti, šumu apod.*
3. Deformace
 - **Změna šířky či výšky obrazových dat (+- 10, 20%)**
 - *Práce s oříznutím obrázku*
 - *Přidávání stínů a rozmazávání*

Augmentace obrazových dat modelu u ČMK-20

Úspěšnost dílčích modelů



Vývoj úspěšnosti a počtu slov modelu AGREG



Závěr

- Na našich dvou modelech jsme vytvořili nový **CZECH supermodel_SGS**
- Základ tvoří 10 různých dokumentů (různé rukopisné styly)
- 947 stran Train set GT (Chybovost CER 2.98%)
- 93 stran Validation set GT (Chybovost CER 5.83%)

Výsledek:

- S naším modelem lze transkribovat podobné rukopisy s přesností 94,17%

prof. PhDr. Dušan Katuščák, PhD.
dusankatuscak@gmail.com

Lukáš Němec
luki.nemec@seznam.com

Vojtěch Říha
vojtech.riha149@gmail.com



**SLEZSKÁ
UNIVERZITA**

FILOZOFICKO-
PŘÍRODOVĚDECKÁ
FAKULTA V OPAVĚ