

Ja tak do práca ľudských možností
vzrastajú malou obetou, ktorú
som na oltár svojho národa obetujem.
Tu sme si poobľubovali – a s ve-
kou na kňazstvo dovozili. Aj tu, ako
druhá sme až do teraz. Len sa nemôže-
me odviazať Jaroslava Brnka, ať



TVORBA MODELU NA AUTOMATICKÉ ROZPOZNÁVANIE RUKOPISU J. M. HURBANA V PLATFORME TRANSKRIBUS

(POSTUPY A SKÚSENOSTI)

Mgr. Alica Kurhajcová, PhD.

Katedra histórie, Filozofická fakulta,
Univerzita Mateja Bela
v Banskej Bystrici

Blahorodny Pane!

Už na slobode súc odpowiedam na Váš ctený dopis zo dňa 22 aprila, s ktorým ste mi i diplom čestného mešťanstwa Vášho mesta poslat' račili. Prijmite odo mňa srdečnú vdaku za toto wyznačenie mňa vtody, keď som odsúdený sedel vo väzení a powedzte mojim pánom spolumešťanom, že wyznačenie toto horkosti môjho wäzenia značne osladilo.

Mimo týchto osobných citow wzbudili ste, páni moji vo mne ale tiež aj iné radostné a blažiace mňa útechy; minulý čas wyoral w srdci mojom hlboké brázdny zármutkow všelijakých, ale tye sú už za mnou, tye pomohol preniest' Pán, a na tye sa ja už neohliadam zpät' ale hladím len k cielu, ktorý wytknul som si za rana mladosti mojej, a k tomuto cielu bežiac widím s potešením velikým jako mi pribýva spolubojownikow.

Činom tím, že ste mňa, tolho už honeneho a prenasledovaného pričislili k mešťanstwu svojmu doklázali ste, že náležite k tým, čo boria sa už od dávna, predávna spolu so mnou za slobodu, rowné právo, rozkvet a blahobyť duchowný i hmotný národa slovenského. Lebo môžu už hovorit' čo chcjú

CIEĽ: Predstaviť hlavné etapy práce a aj postupy pri vytváraní „Modelu J. M. Hurban“

Blahorodny-Blahorodny Pane!

Už na slobode súc odpowiedam na Váš ctený dopis zo dňa 22 aprila, s ktorým ste mi i diplom čestného mešťanstwa Vášho mesta poslat' račili. Prijmite račili. Prijmite odo mňa srdečnú vdaku-~~vdaku~~ za toto wyznačenie mňa vtody, keď som odsúdený sedel vo ~~wäzení-wäzení~~ a powedzte mojim pánom spolumešťanom, že wyznačenie toto horkosti môjho wäzenia značne ~~osladilo-osladilo~~. Mimo týchto osobných citow wzbudili ste, ~~pani-páni~~ moji vo mne ale tiež aj iné radostné a blažiace mňa ~~utechy-utechy~~; minulý čas wyoral w srdci mojom hlboké ~~brázdny-brázdny~~ zármutkow ~~všelijakých-všelijakých~~, ale tye sú už za mnou, tye pomohol preniest' Pán, a na tye sa ja už neohliadam ~~zpät'-zpät'~~ ale hladím len k cielu, ktorý wytknul som si za rana mladosti mojej, a k tomuto cielu bežiac widím s potešením velikým jako mi pribýva spolubojownikow.

Činom tím, že ste mňa, ~~tolho-tolho~~ už ~~honeneho-honeneho~~ a prenasledovaného ~~pričislili-pričislili~~ k mešťanstwu svojmu ~~doklázali-dokázali~~ ste, že ~~náležite-náležite~~ k tým, čo boria sa už od dávna, predávna spolu so mnou za ~~slobodu-rowne-slobodu~~, rowné právo, rozkvet a blahobyť duchowný i hmotný národa slovenského. Lebo môžu už hovorit' čo ~~chcejú-chcejú~~



- Prečo Jozef M. Hurban?

Transkribus ako „kľúč“ k sprístupneniu Hurbanovej nevydanej alebo len čiastočne vydanej rukopisnej pozostalosti (listov, denníkov, kázní a i.); potreba obohatiť platformu o jazykovo slovacikálne texty; osobný motív.

- Na účely tvorby modelu: listy rodine (1846 — 1887); písané latinkou (slovenčina, slovakizovaná čeština, ojedinele latinské, nemecké či maďarské slová); Literárny archív SNK v Martine.

TRANSKRIBUS & HURBANOVE LISTY

✓ Po prevzatí digitálnych kópií príprava malých vzoriek vlastnoručne a „bezchybne“ prepísaných strán listov („Ground Truth“),

✓ Spustenie tréningu, vyvíjanie prvých menších verzií modelu a jeho vylepšovanie (Model JMH 28 s presnosťou znakov 7,98 %).

1. ETAPA: Príprava a trénovanie modelu (december 2020 – október 2021)

„Malé“ verzie modelu na transkripciu rukopisu J. M. Hurbana							
Názov modelu	Cvičný súbor / Training set		Overovací súbor / Validation set	Chybovosť na úrovni znakov CER (%)		Segmentácia	Cykly / Epochs
	strany / pages	slová / words		cvičný (TS)	validačný (VS)		
Model JMH 1	26	7 673	7	0,79	15,43	M	50
Model JMH 2	31	8 583	2	0,96	27,64	M	50
Model JMH 3	28	7 936	3	0,86	15,40	M	50
Model JMH 4	26	2 959	3	0,62	18,90	A	50
Model JMH 5	28	3 189	1	0,67	18,02	A	50
Model JMH 6	27	3 059	2	0,59	18,77	A	50
Model JMH 7	56	11 127	6	1,78	15,75	M+A	50
Model JMH 8	46	8 191	14	0,58	10,98	M+A	100
Model JMH 28	104	18 099	16	0,81	7,98	M+A	250

M – manuálna (dôkladná) segmentácia; A – automatická (strojová) segmentácia

✓ Metódou priebežného rozširovania cvičného súboru o automaticky transkribované strany (t. j. na základe aktuálnej verzie modelu); ich následná korektúra a pripojenie k predošlým „bezchybne“ pripraveným stranám (Ground Truth).

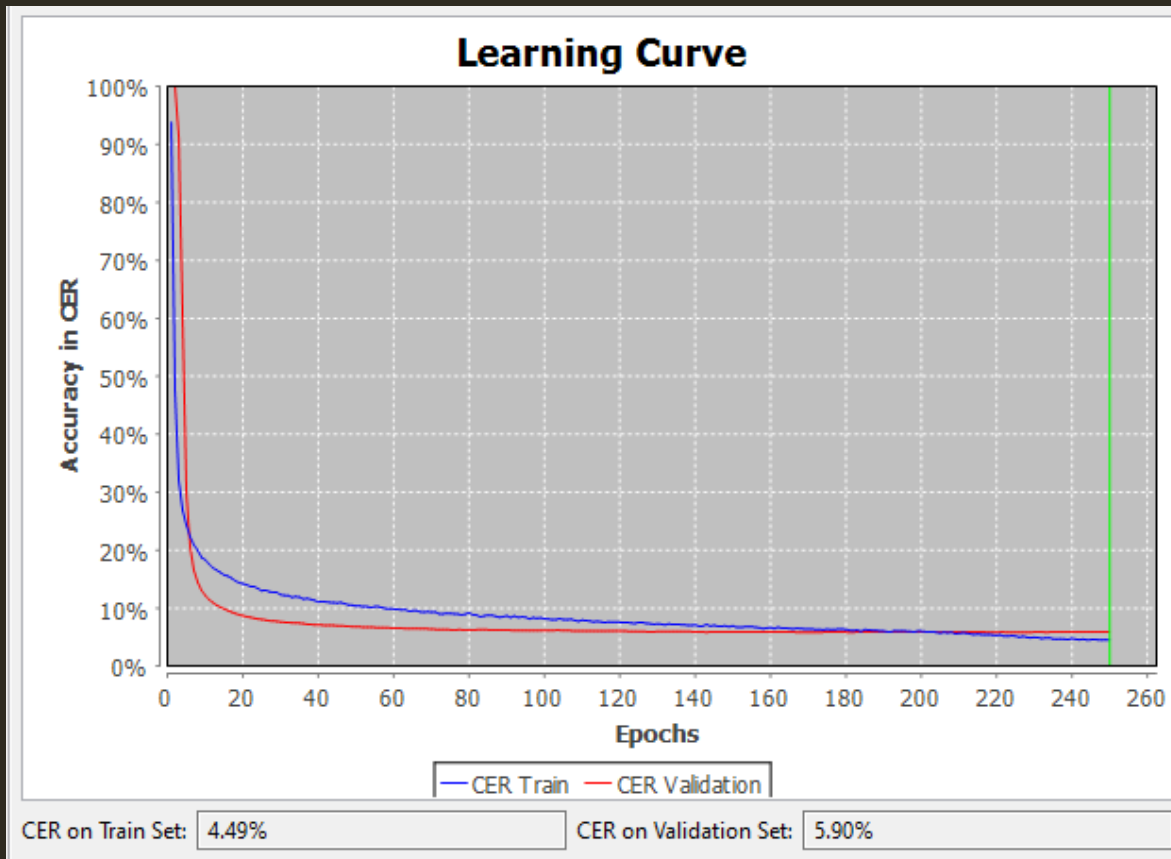
✓ Metódou použitia „base modelu“: t. j. „zuniverzálnenie“ špecifického modelu pomocou modelu na transkripciu rukopisu Józsefa Kissa (Transkribus Hungarian handwriting 19th-20th Century).

2. ETAPA: vylepšovanie modelu (október 2021 – august 2022)

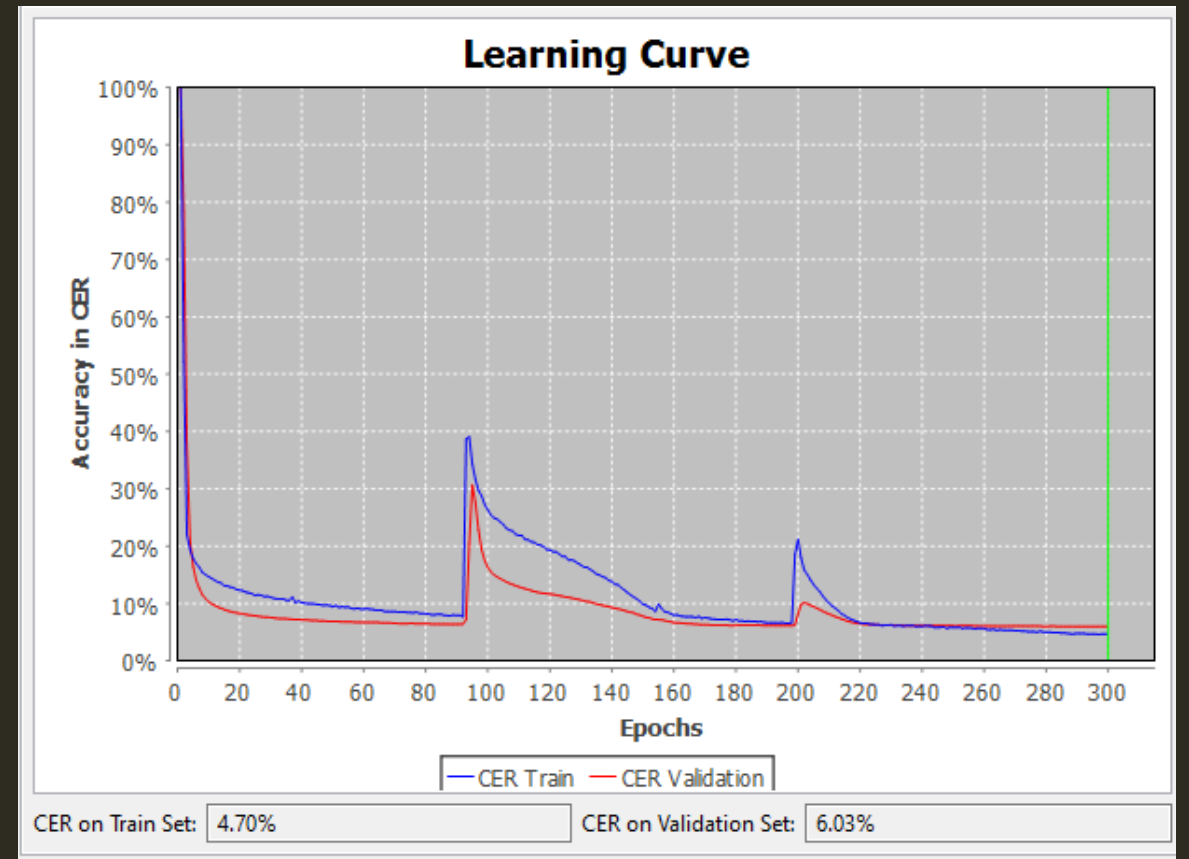
Názov modelu	Trénovanie stredne veľkých verzií modelu						Cykly / Ep.
	Cvičný súbor / Training set		Overovací súbor / Validation set		Chybovosť na úrovni znakov (CER) %		
	strany	slová	strany	slová	cvičný TS %	validačný VS %	
Model JMH 30	123	24 513	16	3 341	0,79	7,01	250
Model JMH 35	327	62 597	33	6 168	5,35	6,86	100
Model JMH 36	369	72 888	46	9 555	3,85	6,32	200
Model JMH 40	424	83 583	34	6 698	4,26	6,28	200
Model JMH 45	423	83 593	34	6 682	4,18	6,16	200
Model JMH 50	560	101 241	62	13 092	4,49	5,90	250
Model JMH SK+HU	530	98 700	60	12 016	4,70	6,03	300

VÝSLEDOK: VYTVORENIE DVOCH VARIANTOV ŠPECIFICKÉHO MODELU NA AUTOMATICKÚ TRANSKRIPCIU RUKOPISU J. M. HURBANA

Model J. M. Hurban: chybovosť na úrovni znakov v overovacom súbore: 5,90 % (CER)



Model J. M. Hurban SK+HU: chybovosť na úrovni znakov v overovacom súbore: 6,03 % (CER)



Overovanie funkčnosti oboch verzií modelu na nových, dosiaľ netranskribovaných stranách:

- od excelentného prepisu (2,10 %, resp. 2,49 % chybné prečítaných znakov) po stále použiteľný prepis (8,70 %), v jednom prípade stále zrozumiteľný (12,70 %*).

* Opraviť v abstrakte (pôvodne 12,38 %).

3. ETAPA: využiteľnosť modelu v praxi (august – september 2022)

CHYBOVOSŤ NA ÚROVNI ZNAKOV (CER) V %		
Názov modelu	Model JMH	Model JMH SK+HU
STRANA 1	4,40	4,75
STRANA 2	7,50	7,72
STRANA 3	7,42	8,70
STRANA 4	8,13	9,28
STRANA 5	5,48	5,55
STRANA 6	5,91	5,91
STRANA 7	6,64	5,63
STRANA 8	8,70	7,73
STRANA 10	5,53	5,30
STRANA 11	7,50	8,16
STRANA 12	6,42	5,50
STRANA 13	2,69	3,65
STRANA 14	3,61	5,09
STRANA 15	12,70	10,35
STRANA 17	3,36	4,01
STRANA 19	2,49	3,54
STRANA 20	2,77	2,10
STRANA 21	2,86	3,13
STRANA 22	2,80	3,28
STRANA 23	3,19	4,79
Priemer (CER)	4,96	5,33

KATEGÓRIE CHÝB PRI „ČÍTANÍ“ LISTU:

- **irelevantné:**
 - interpunkčné (výkričník, bodka, čiarka),
 - diakritické: Mikulaši → Mikuláši; wazenia → wäzenia; Väžení → Väzení; môjho → mojho,
 - delenie slov: DrHurban → Dr_Hurban.
- **závažnejšie:**
 - podobné, resp. nejednoznačné znaky: vite → rite; Prikladam → Prikładám; príležitostisti → príležitost,
 - odklon od zabehaných ťahov písmen: vikekkkost → Príležitost,
 - nesprávny zápis skratiek a číslic: H / Pos → NB, 1876 → 1870.

vite rite týchto národow!!! Sláwa
Slovenskému Národu!
Váš, Bratia moji,
tu w 66ty den fotografowany



DrHurban

Dané vo Väzení
dňa 13 febr. (w 70ty den
wäzenia mojho) 1870.

~~Prikładám jednu fotografiiu~~
pre brata professora w Mikulaši.

~~Prikładám ešte jednu fotografiiu~~
moju pre milých SwätoMartinských
mojich priatelow a priateľkyne k f-ru-

Vam kám Presidentky Žiweny, čím
to preto pri tejto príležitosti, že už sme
pusili skúsili tu vo fognari, fognazi, ako moje do
Turca posielky sú lapané. Prosim odoslat
to ta - keď s-bude k-vikekkkost-Príležitost.

vite rite týchto národow!!! Sláwa národow!!! Sláwa
Slovenskému Národu!
Váš-Váš, Bratia moji-moji,
tu w 66ty den fotografowany-fotografowany
DrHurban-Dr Hurban
Jane-Dané vo Väzení-Väzení
dňa 13 febr.-febr. (w 70ho len-70ty den
wazenia-mojho)-1876-wäzenia mojho) 1870.
H-Prikładam-NB Prikładám jednu fotografiiu
pre brata professora w Mikulaši-Mikuláši.
Pos-Prikładam-NB Prikładám ešte jednu fotografiiu
moju pre milých SwätoMartinských-SwätoMartinských
mojich priatelow a priateľkyne k f-ru-
Vam kám Presidentky Žiweny, čím Žiweny, čím
to preto pri tejto príležitosti, príležitosti, že už sme
pusili skúsili tu vo fognari, fognazi, ako moje do
Turca posielky sú lapané. Prosim odoslat-odoslat
to ta - keď s-bude k-vikekkkost-Príležitost.

Stále zrozumiteľný prepis (CER 12,70 %)

„CIT“ NA ČÍTANIE J. M. HURBANA

OTÁZKY:

- otázka hranice možnosti vylepšovania modelu
- otázka možnosti vytvorenia univerzálneho modelu na rozpoznanie rukopisov slovenských osobností 19. storočia

LIMITY:

- nepoužitelnosť modelu na cudzí rukopis
- (ne)použitelnosť modelu na Hurbanov rukopis z 30. rokov

MOŽNOSTI:

- použiť verejne dostupné modely na prepis kurentu (German Kurrent XIX M2) či azbuky (Russian generic handwriting 1)
- vyhľadávať kľúčové slová
- pripraviť komentovanú edíciu vybraných listov (či modernú databázu Hurbanovej siete kontaktov?)

Soetuskó maj drahi!

Sak mi ta luto bolo se si take
lutostivi a rozbaveni list pi
sal, a tak sme ti chceli hned
odpovedat ale pre prace sme
nemohli, vet tyho stebu do
bre misli, a jeho list si sle
rozumel, a nebuo tak roztra
teni, pisalsi se detom neco
posles citame a nie, se censur
posilas, ani tej nebolo, muj Soe
tuskó mili! Sak mi ta velmi
radi mame len sa nam pekne
spravuj a dobre sa uc sak ti
to B. Buch a Ymistus seko
pomoz. Mi sme Soatki odbavo
vati zdraui, opakaji, aj deti,
su zdrave a mali veliku ra
dost se stromku, krajju sa
a raduju, a hoster sa sankuje.

*Ket pisess pekerise pis noco
lechoat preicitane*

Rukopis Anny Hurbanovej (CER 19,35 %)
LA SNK, 138M1, 1865

Ja tak do práca ľudského možem sa
vziasť svojou malou obetou, ktorú
som na oltar svojho národa obžijem.
Tu sme si poobľubovali – a s ve-
kou na kľúčom dobroty. Aj tu, ako
druhá sme až do teraz. Len sa nemôže-
me odtrhnúť Jaroslava Brnka, ať



Ďakujem za pozornosť.

Mgr. Alica Kurhajcová, PhD.

alica.kurhajcova@umb.sk