



Umelá inteligencia pomáha sprístupňovať historické písomné dedičstvo

Prof. PhDr. Dušan Katuščák, PhD.

Filozoficko-prírodovedecká fakulta, Ústav bohemistiky a knihovníctva, Slezská univerzita Opava
Štátna vedecká knižnica, Banská Bystrica

dusan.katuscak@fpf.slu.cz



Témy prezentácie

1. Digital humanities
2. Projekt SKRIPTOR APVV (ŠVK a UMB Banská Bystrica)
3. Ceny transkripcie
4. Výsledky transkripcie (modely, sprístupnenie)



Kontext

Digital humanities - Rozpoznávanie textu - Transkripcia

Digital humanities - 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.

PROJEKT READ (HORIZON 2020);

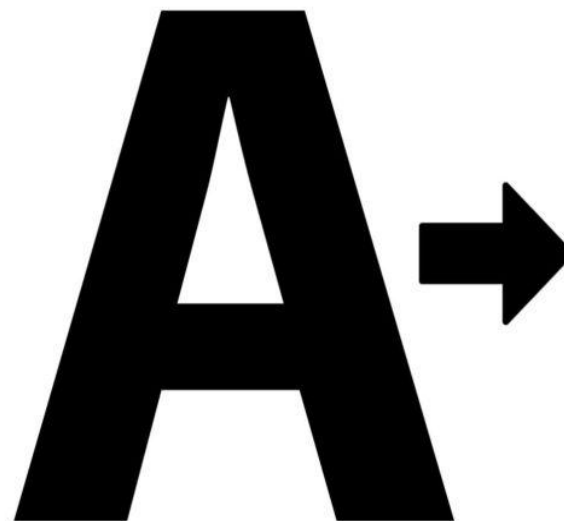
READ-COOP;

TRANSKRIBUS;

PROJEKT SKRIPTOR

OCR (Optical character recognition)

- ▶ Rozpoznávací model je naprogramovaný ručne
- ▶ ABBY FineReader
- ▶ Presnosť transkripcie 87%-92%
- ▶ Až 100% - možno trénovať



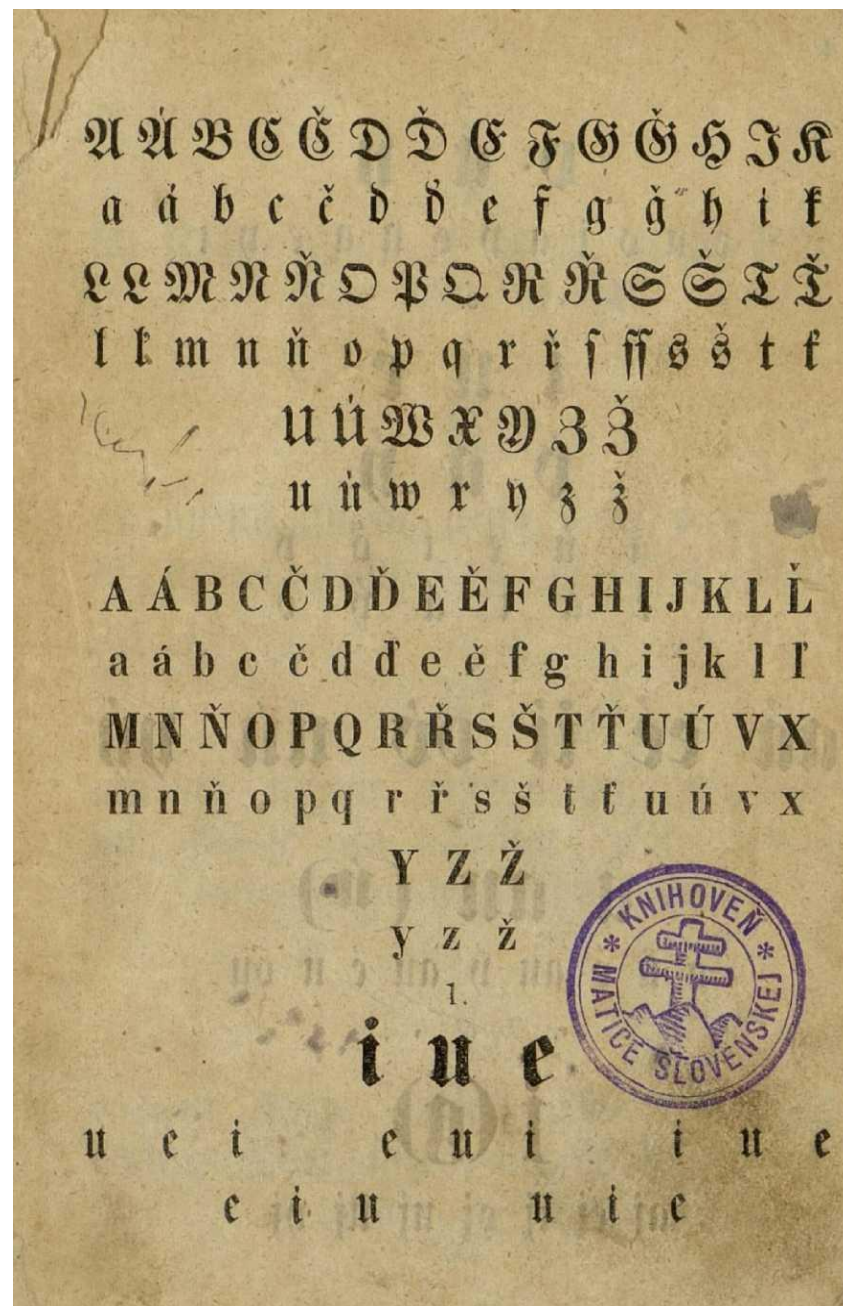
- ▶ **Staré tlačé rukopisy a tlačé - problém!**
- ▶ **OCR = neuspokojivá transkripcia (ale, môžu sa tiež učiť)**

Fraktúra a antikva

Abeceda v šlabikári z roku 1872.

Fraktúra a antikva

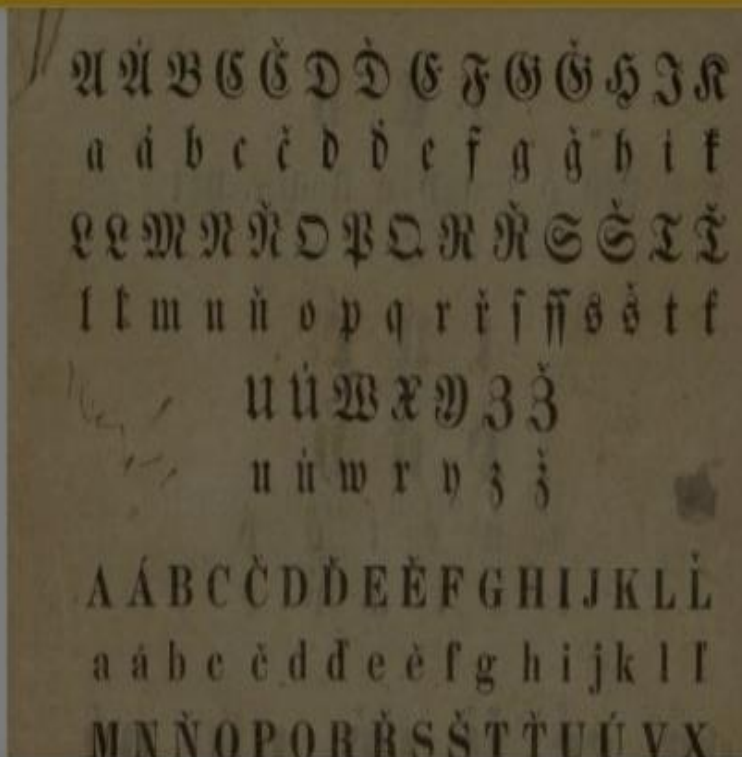
OCR server uspokojivo nerozpoznáva





Hľadať v dokumente ...

3 z 118 stránok



Ewanjelický Šlabikář wypracowaný z nałożení wel. Seniorátu gemerského

Autor

Ormis, Samuel

1864-1909



Vydavateľské údaje

V Rožňavě: Tiskem Michala Kovács, 1872.

Typ dokumentu

>' 21 \$ @ d d \$ s > @ a c i b c c b i j c f g g f) t t i m n ñ o q r í f f f ô é t f U t i 28 £ 3) 3 3 u f f i d r n \$ I A Á B C Ć D Ď E Ě F G H I J K L Ľ a á b c ĉ d ě e ě f g h i j k l (1 M N Ń O P Q R A S Š Ť Ť U Ú V X m n ñ o p q r í s š t t f u ú v x Y Z Z y z ž i . t u e i t c i c u t i u e c i u i e

ORMIS, Samuel a 1864-1909. Ewanjelický Šlabikář wypracowaný z nałożení wel. Seniorátu gemerského. V Rožňavě: Tiskem Michala Kovács, 1872. s. [3]. Dostupné tiež z: <https://onk.snk.sk/view/c9e7ce09-2c3b-4c72-85e9-304947366228?page=0ce8d83e-678a-4a10-a0e8-2b03438fa662>

Nedostatočná OCR transkripcia

- ▶ Automatické OCR abecedy v Ormisovom šlabikári z roku 1872
- ▶ Poznámka: V riadku pod obrázkom je zobrazený výsledok automatického OCR. Fraktúra v červenom rámčeku je rozpoznaná celkom neprijateľne a rovnako sú chyby v rozpoznávaní antikvy – červené značky.
- ▶ Zdroj: Slovenská národná knižnica: Katalóg digitálnej knižnice. Dostupné: Ewanjelický Šlabikář wypracowaný z na'ložení wel. Seniorátu gemerského | Slovenská národná knižnica | Digitálna knižnica (snk.sk).
- ▶ Nedostatočná alebo nemožná transkripcia rukopisov - len obrázky
- ▶ Ukážky nasledujú - obrázky bez transkripcie

Belopotocký

Zlaté slunce

4) 1)
Zlaté slunce bylo dokonale.
Již svůj denný na obloze běh,
Vlisko denní se umenšovalo
A chladnější, byl zepřívár dech;
Oblehy pak krásou šarlátovan
Tri kápade byly sbervené,
Když sem střebel v lese sílu novou
Odpočívám v hanošti zelené.

2.
Do tichanbé sem se samotnosti
Oddalil od kresku dediny
Uhrily mne svou rozložitostí
Smolan se posici sročiny.—
Ticho jestit v celém temném háji,
Vetve sročiny jen se časem hnau,
Zvuky píšťalky mé obyvají
Tež se — ale ležko utichnav.

Reč Štefana Moyzesa
pred Forgáčom

1861

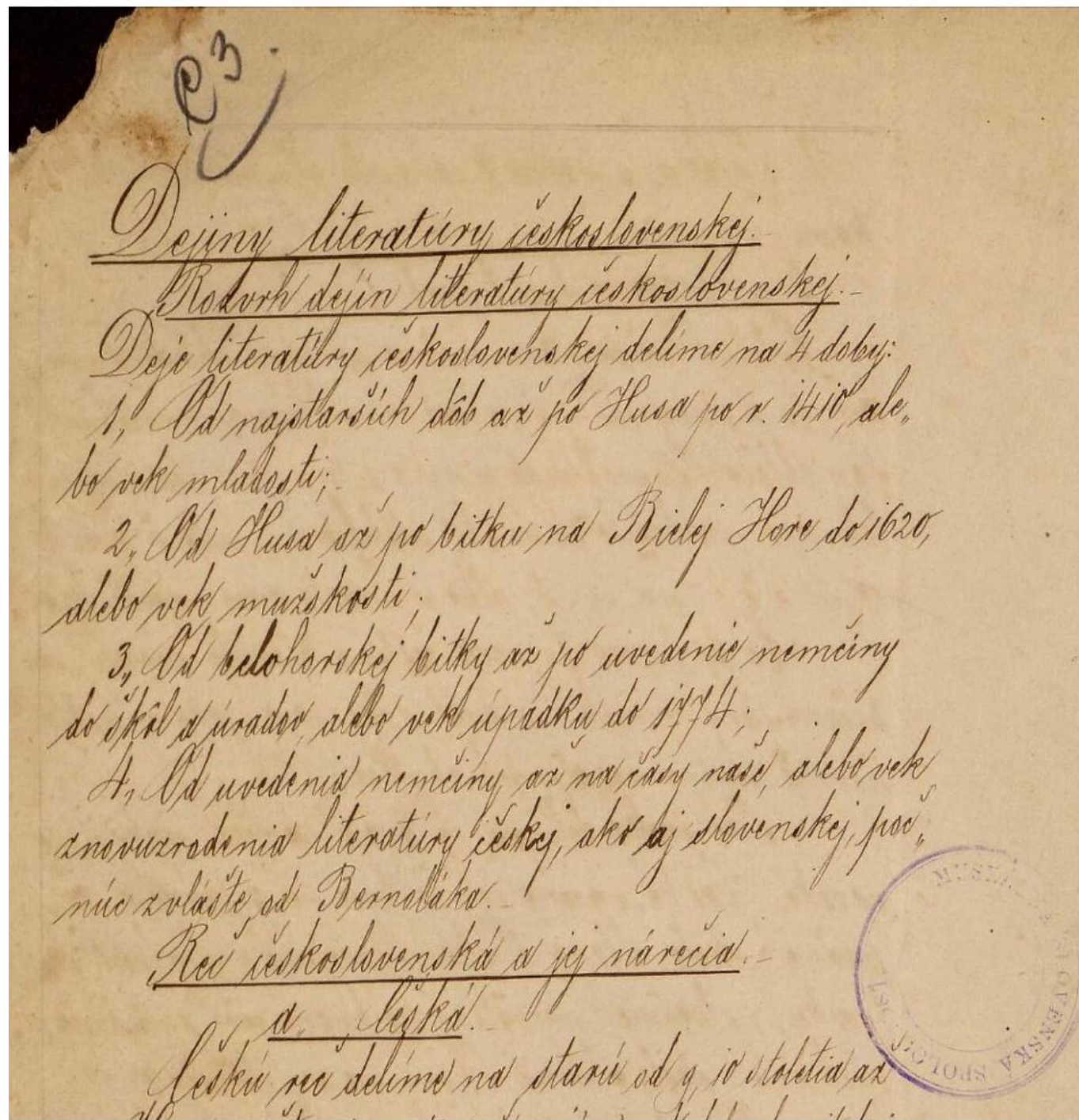
Vasae Excellentiae,
Najosviadenejši pán grof a uhorsko-kráľovský kancelár,
pane najmilostivejši!

Dvojkráľovná je pohromka, ktorá nám, vylstenstvu a štátu slovenskému, svätú náležať povinnosť, aby sme pred milostivú obličiu Vašej Excellentie Vásij pokorne predstavili.

Co sme totiž dľa vysokého úradníckeho postavenia, jako i dľa veľkodušného vzrušenia Excellentie Vásij ľahko ľuďom mohli, to i zo skutočnej zvesnosti našej už vieme, že totiž Excellentie Vasae najpoživenejšiu prosbu našu, o povznesenie štátu slovenského a podpora živu najjasnejšieho cisárskoráľovského praprotu najvyššieho službenia vysokým svojím prostredkovaním milostivo podporovať násť.

Dobšinský

Dejiny literatúry



Dohnány - práce

1842

Proč sme přišli do Brětislavi?
Pracuj každý s chutí usilownau
Na národů roli dedičné
Česty mohau býti rozličné,
Genom vůli všickni měgme rovnau.
Jan Kollár.

Nám P. Rodičové dobré chtejíce, do Brětis-
lavi nás poslali, a by sme se mi tu na
učili náš tekaučý řěwot na dobré winakla
dati, tedy M. Br. přijao mosíme všickny sobě
rozmýšlet, naše powinnosti, co od nás P. R.
poradují, a usilowat se máme to vše učiniti,
aby M. R. na nás se darmo nečtěwili.

Ján Hollý
rukopis

č. 2. 97. ~~Na Den Mena~~ Na Den Mena. S. N. Dercikovi

Tak sa teša dlavici na Sar, keď nastane Tepló,
A spaňilimi nabe' zaodiva' listami Flage:
Tak sa mnohi' radugu, keď nam Den' tvošo docbádza
Mena prewiborni' Mužu; wšak watsšo neciti
Kikdo we swem ze wšech ti Pragicich drdci plesana
Než ga sam: ponewac' zaden celoročne milegš, i
Zaden aniz mne wzacnegš nad tento nebiva'
Slawni' Den, co ti zaswateni, Dobrodvice pripada
Kecb teda dneš Dašnici, Kterich obgasunge wiffu
Duch Miln' a tubla amita' nad Glni' Haru' cast

Słowanstwoj.

7. Podie sem hle prjklad sobě brati! ,
Plylo za krásného večera. Měsje se ^{Kyillar.} trpytel shwedlami
na welebných nebesjeh. Ja spechal iněblika ^{dobrych} přátel
přátelj, Mlicemi abychom se wzajemně někde vyrazili
w rozmluwě. Nejwáhodnějšj farowé místo se nam jda
la některá ~~hospoda~~ hospoda; neboť sme ^{chtěl byti dnes kyselj} ~~nechtěl~~ ^{mystic}
krásného. Hospoda tedy do nás sme wstoupili byli ozaře,
na; lampy se leskly po klenutě sji osweicij twarě
ctweraicijeh. Posluba, genz to byli pohromade. An sme
bljže wstoupili, spatřili sme znaměho nam - garo zly
genjs - gak se titworj, gra opily před Posluby hospod,
ského. Ei, ei pomysliti sme sobě, garowj to rytyr pried,
wij! myslit by člověk, kdyby geg nespatrił, že se swaru
cnostj kleran se na každem smětis ku wyhlamba, chue
za žura do nebe na ohnivem wozě wstoupiti. Poslu,

J. M. Hurban

1838

Hviezdoslav

1914

Prívet

K slov. mládeži a príležitosti jej zjazdu
v Lipt. Sv. Mikuláši 11/II 1914.

Nie knieža som, nie a vládcom tohto sveta,
čo na poručíci majú prostriedky,
v tíniach rútia, tvrdosti, potrebách
spomáhať veľhodusne, zástiť
pred dočkou pekelnaja; tu klesajúcu
starobu podprieť silným rúmekom,
tu a chladnej tóne neuvrúania vypríeť
zásluhu na výslu — ach! nie som a nič:
a preba ľeá ste, mládež naša dráha,
tých svojich prestorekúch, nepodlegujúch
súdc Masoviarum mňa si zvolili
za protektora! — Nie to omýľ, sáľba?



Sládkovič - kázeň

1852

A. Sládkovič
Kázeň w ned 11 prosna
v Uhrovi 1852.

Pane, nezranítej twých wěřicích
Ans mi w slabostech klesáme,
Nybrž wrhtedni na nás bojujících
Když w utrpení padáme.
O jin wíru Twého uieni,
Upewniz nawždy w dušech našich
Kterau, čas, wet, let+ nermeni
Tu ráis dás w srdce slabých,
By ona jak welka sak nála
Byla w srdcích Twých křestani
Nam tak jednauc wydobila
Křiš nebeských občanů ames.

Věti v Boha snadno se nám zda křestane,
Ja zato mám že není snadno. Wšak mnozí před
ce wíru w Boha tak považují, jako wěc jenž ne-
škodí když wěříme, a není sak těžka jako sobě

Palkovič

Latinský rkp., testament

1.
Nos Capitulum Ecclesiae
Metropolitana Strigoniensis,
memoria commendamus, tenore presentium
significantes, quibus expedit Universis: quod
post mortem Reverendissimi olim Domini
Georgii Palkovici Propositi Sancti Stepha-
ni de Castro Strigoniensi, Fratris, Socii,
et Canonici nostri, die 27^a Januarii a.c.
isthic Strigonii inventam, duplici or-
dine Litteralia Instrumenta, inter posthu-
ma Eiusdem Scripta, reperta: primum
illorum, Minuta Testamenti, per extensum
propria Defuncti manu, Anno 1823. concepta,
subinque diversis temporibus, uti id ex co-
lore atramenti, et ipso scriptura tractu pa-
set, pariter propria manu modificata, et
aucta; - alterum autem in puro, perinde
autographum Defuncti, quantum quidem ex

Šoltésová

rkp

Prispevok k židovskej otázke.

Hlas so Slovenska.

3 Ste niet národa na celom našom pozem-
4 dom svete, ktorý by tam bol po celom ňom rozle-
5 gnutý, ako je národ židovský - a tiež niet národa
1 a tom svete, ktorý by práve voju národnost
2 tak zhmýslnu, vypočítave zapieral a podľa príhod-
3 nosti pripovoľal si ktorúkoľvek inú, národnost,
4 radne národnost byť u2dzi ktorými a z ktorých
5 je - nakoľko je ktorá práve životnou inou. V prav.

Slovenská filharmónia 1949

Strojopis, programový bulletin

1 5. XI. 1949 - 10 hod.

2 Konc. sieň Pov. vnútra. dir. Dr. Ľud. Rajter.

3 Koncert pre študentov.

4 Program 5 1. / J. S. Bach : "Sarabanda a Gavotta." ✓

6 . / J. Haydn : "Ochsen-Menuet." ✓

7 . / W. A. Mozart : "2 nemecké tance." ✓

8 . / M. Moyzes : "Mazurka." ✓

9 . / A. Dvořák : "III. časť zo symf. IV. G-dur." ✓

10 . / -"-" : "Slovanský tanec čís. 8." ✓

11 . / B. Smetana : "Z českých luhov a hájov." ✓

12 symf. báseň .

13 . / A. Moyzes : "4 veta zo Symfonie č. V." ✓

HTR+ = Handwritten Text Recognition

- ▶ **Umelé neurónové siete (ANN)**
- ▶ **Stroj - Rozpoznávací model sa sám učí na príkladoch!**
- ▶ **Množstvo dát, trénovanie, učenie stroja**
- ▶ **CER (Character Error Rates)** je miera chybovosti znakov (porovnáva pre danú stranu celkový počet znakov (n) vrátane medzier s minimálnym počtom vložení (i), nahradenia (s) a vymazania (d) znakov, ktoré sú potrebné. získať výsledok *Ground Truth*. Ide teda o chyby v porovnaní s presným textom.
- ▶ **Vzorec na výpočet CER** je nasledujúci: $CER = [(i + s + d) / n] * 100$. Každá malá chyba v prepise je štatisticky plnohodnotná chyba. To znamená, že každá chýbajúca čiarka, „u“ namiesto „v“, dodatočná medzera alebo dokonca veľké písmeno namiesto malého písmena sú zahrnuté v CER ako chyba.
- ▶ **PyLaia** – ako **HTR+** s možnosťou nastavovať parametre

vstup
48x48

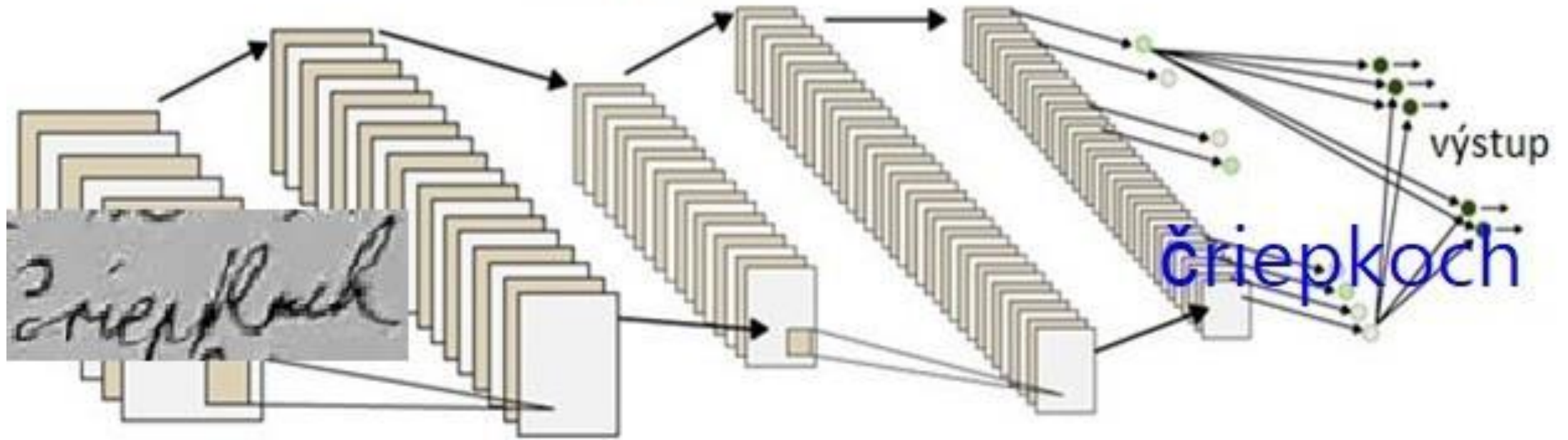
6@44x44

6@22x22

12@18x18

12@9x9

mapy funkcií



konvolúcia max-pooling

konvolúcia max-pooling

mapy funkcií

klasifikácia

Platforma Transkribus



- ▶ *Transkribus expert klient* – [Transkribus | AI powered Handwritten Text Recognition \(readcoop.eu\)](https://readcoop.eu)
- ▶ kľúčový inovatívny nástroj pre transkripciu historických rukopisných a tlačných dokumentov
 - ▶ komplexná platforma na digitalizáciu,
 - ▶ rozpoznávanie textu podporované umelou inteligenciou
 - ▶ prepis a vyhľadávanie historických dokumentov –
 - ▶ z akéhokoľvek miesta, kedykoľvek a v akomkoľvek jazyku

Ekosystém Platformy Transkribus



- **READ-COOP SCE** (Societas Cooperativa Europeae – SCE), od 1. júla 2019 [About us - READ-COOP \(readcoop.eu\)](#).
 - Cieľ udržať a ďalej rozvíjať platformu *Transkribus*
 - 100 000 používateľov
 - 40 mil obrazov
 - 20 mil rozpoznaných strán
- **Transkribus Lite** – jednoduchá transkripcia v prehliadači [Transkribus](#)
- **Read&Search** – portál na sprístupnenie zbierok a dokumentov [Read&search - READ-COOP \(readcoop.eu\)](#)
- **ScanTent a DocScan** – podpora snímania [ScanTent - Professional scanning with your Smartphone \(readcoop.eu\)](#)

Ceny a efektívnosť



Manuálna transkripcia - 10-15 eur/strana

Automatická transkripcia - Transkribus ca 0,12 € - 0,14 €/strana s DPH

Transkribus? [Transkribus Credits & Pricing - READ-COOP \(readcoop.eu\)](https://readcoop.eu)

Cena snímania:

ScanTent a DocScan – 1000 strán ? 50 € (0,05 €/strana)

Profesionálne skenovanie 742-60 obr. (670 obr) = ca 1000 € s DPH (bez postproc.). Teda ca 1,50 €/strana

Hospodáriť s kreditmi!



500 CREDITS FREE

Sign up for free

- ✓ Get 500 Credits free on signup
- ✓ Only applicable to one "collection"
- ✓ Can be used with any "engine"

❓ How many pages can I process with this? ▾

Engine	Handwritten Pages	Printed Pages
PyLaia	500	3000
HTR+	400	2500

Pages can be mixed and matched

3 000 CREDITS 720€ ONE-TIME

Add to cart

- ✓ Can be bought **as often as you like**
- ✓ Shareable with **any** "collection"
- ✓ Can be used with any "engine"

❓ How many pages can I process with this? ▾

Engine	Handwritten Pages	Printed Pages
PyLaia	3 000	18 000
HTR+	2 400	15 000

Pages can be mixed and matched

3 000 CREDITS 648€ ONE-TIME

Become a Member

- ✓ Can be bought **as often as you like**
- ✓ Shareable with **any** "collection"
- ✓ Can be used with any "engine"

❓ How many pages can I process with this? ▾

Engine	Handwritten Pages	Printed Pages
PyLaia	3 000	18 000
HTR+	2 400	15 000

Pages can be mixed and matched



Suma / kredity	Stroj PyLaia Počet strán rukopis/cena €	Stroj PyLaia Počet strán tlač/cena €	Stroj HTR+ Počet strán rukopis/cena €	Stroj HTR+ Počet strán rukopis/cena €
648 €/3000	3 000/0,216€	18 000/0,036€	2 400/0,27€	15 000/0,043€
1944 €/10000	10 000/0,194€	60 000/0,0324€	8 000/0,24€	50 000/0,038€

Projekt Skriptor



- ▶ Na Slovensku sme začali pracovať s platformou *Transkribus* v roku 2017
- ▶ Projekt APVV SKRIPTOR – SVK BB a UMB BB (APVV-19-NEWPROJECT-17816 (2020-2024). *Inovatívne sprístupnenie písomného dedičstva Slovenska prostredníctvom systému automatickej transkripcie historických rukopisov*)
- ▶ *Transkribus* - používame stroj umelej inteligencie *HTR+* (Handwritten Text Recognition) a *PyLaia*
- ▶ Tieto stroje zatiaľ nemôžu okamžite automaticky transkribovať rôzne historické rukopisy
- ▶ Učiť stroj na konkrétny typ písma a rukopisu.
- ▶ Hlavným cieľom praktických experimentov v projekte SKRIPTOR je v súčasnosti tvorba *modelov* transkripcie.
- ▶ Transkripcia geografických slovacikálnych historických zbierok a dokumentov (slovenčina, tiež čeština, latinčina, maďarčina, poľština ai.)
- ▶ Tvorbu modelov transkripcie na základe väčších zbierok, ktoré obsahujú stovky a tisíce strán
- ▶ Tvoríme veľmi dobré až excelentné modely transkripcie archívnych dokumentov a starých tlačí

Výskumníci SKRIPTOR



- ▶ P. Maliniak - transkripcia rukopisných kázní Izáka Abrahamidesa
- ▶ P. Kunec a absolvent štúdia histórie M. Katreniak - transkripcii kanonických vizitácií.
- ▶ M. Mikušková a L. Nižníková - transkripcii komplikovanej historickej tlače
- ▶ O. Tomeček transkripcie novolatinského rukopisu reambulačného protokolu.
- ▶ M. Bôbová - digitalizácie vo výskume dejín knižnej kultúry
- ▶ A. Kurhajcová - transkripcioa rukopisu J. M. Hurbana.
- ▶ I. Nagy - transkripcie Csákósovho katalógu korešpondencie Koháryovcov
- ▶ D. Katuščák – transkripcia historických rukopisov a tlačí

Výstupy projektu APVV Skriptor (2022)

- ▶ **KATUŠČÁK Dušan a Imrich NAGY, eds.**
 - ▶ Automatická transkripčia slovacikálnych historických dokumentov [elektronická editovaná kniha] / zost. Dušan Katuščák, Imrich Nagy ; rec. Milan Konvit, Jan Odstrčilík. - 1. vyd. - Banská Bystrica : Belianum. Vydavateľstvo Univerzity Mateja Bela v Banskej Bystrici, 2022. - 200 s. - ISBN 978-80-557-2020-3. – Dostupné na: DOI: <https://doi.org/10.24040/2022.9788055720203>
 - ▶ URI: <https://repo.umb.sk/xmlui/handle/123456789/372>
- ▶ KATUŠČÁK, Dušan: Digital Humanities a automatická transkripčia rukopisných textov. In: *ITLib : informačné technológie a knižnice*, roč. 24, č. 1, 2020, s. 6 – 16. Dostupné na: [3.pdf \(cvtisr.sk\)](#)
- ▶ KATUŠČÁK, Dušan: Umelá inteligencia pomáha sprístupňovať písomné dedičstvo. In: *Knihovna : knihovnícká revue* [online], roč. 33, č. 2, 2022, s. 27 [cit. 2022-11-20]. Dostupné na: <https://knihovnarevue.nkp.cz/archiv/2022-2/recenzovane-prispevky/umela-inteligencia-pomaha-spristupnovat-pisomne-dedicstvo>

BŔBOVÁ, Mária. Aplikácia automatickej transkripcie na príklade knižničného katalógu zo začiatku 19. storočia. In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 7, s. 147-167. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

KATUŠČÁK, Dušan. Metodológia a metodika transkripcie historických textov. In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 1, s. 19-47. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

KATUŠČÁK, Dušan. Výkladový slovník pojmov a termínov. : KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 9, s. 193-204. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

KUNEC, Patrik. Automatická transkripcia protokolov z kanonických vizitácií farností Zvolenského arcidiakonátu z polovice 18. storočia v platforme Transkribus. In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 5, s. 84-101. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

KURHAJCOVÁ, Alica. Keď sa stroj učí čítať Hurbanove listy. In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 6, s. 125-146. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

MALINIÁK, Pavol. Postila Izáka Abrahamidesa Hrochotského a automatické rozpoznávanie jeho rukopisu. In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 2, s. 48-64. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

NAGY, Imrich. Sprístupnenie Csákósovho katalógu korešpondencie Koháryovcov pomocou automatickej transkripcie. In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 3, s. 66-84. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

NIŽNÍKOVÁ, Lucia a Michaela MIKUŠKOVÁ. Modely automatickej transkripcie štvorjazyčného diela J.A.

Komenského Orbis pictus (1798). In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap. 8, s. 168-191. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

TOMEČEK, Oto. Automatická transkripcia reambulačného protokolu Banskej Bystrice z roku 1820. . In: KATUŠČÁK, Dušan a Imrich NAGY (eds), Automatická transkripcia slovacikálnych historických dokumentov, Kap.

4, s. 101-124. Banská Bystrica: Belianum, 2022. ISBN 978-80-557-2020-3. Dostupné: <https://doi.org/10.24040/2022.9788055720203>

Prehľad najlepších modelov v projekte SKRIPTOR (2020 – 2022)

Vlastník modelu (owner)	Názov dokumentu v TRB	Rok(y) vzniku/ pokrytie	Počet str. dok	Jazyk/ jazyky	Typ		Názov modelu	ID modelu	Počet slov TRAIN	Počet riadko v TRAIN	CER TRAIN	CER	
					písma RKP/PRT	Ruky						VALID	Stroj
MALINIÁK	Postila Izáka Abrahamidesa	1600/1601	722	slo/cze; lat	novogotické humanistické písmo (RKP)	1	Abraham model 3	1120455	11434	1817	1,56%	7,99%	HTR+
NAGY	Csákósov katalóg korešpondencie Koháryovcov	1944/1945	4140	lat/hun / ger	moderná kurzíva	1	SKRIPTOR_Csákos_Koháry_3	45232	16307	2800	0.98%	2.00%	HTR+
TOMEČEK	Reambulačný protokol	1820	245	lat	novolatinská kurzíva (RKP)	1	Metales Model 4	1129960	7690	1773	1,54%	4,25%	HTR+
KURHAJCOVÁ	J. M. Hurban	1838/1887	902	slo	kurzíva (RKP)	1	Model J. M. Hurban 50 Model SK+HU	47833 47892	103966 101721	13318 12765	5,10% 5,00%	8,00% 8,70%	PyLaia
KUNEC	Kanonické vizitácie	1754/1756	350	lat	novolatinská kurzíva (RKP)	3+	BBBDA - Kanonická vizitácia - 5th	45220	7134	992	0,32%	6,07%	HTR+
BÔBOVÁ	Elenchus librorum	1802/1852	512	lat; ger	novogotická kurzíva (RKP)		Latinský+ Nemecký	45004+ 30919	7596	2316	1,5%	3,08%	HTR+
NIŽNÍKOVÁ MIKUŠKOVÁ	Orbis pictus	1798	94	lat; hun; ger, cze	antikva; kurzíva; fraktúra; švabach (PRT)	tlač	Model 5	44220	4107	1318	0,36%	1,02%	HTR+
KATUŠČÁK ²⁰⁴	Andrej Kmet'	1860/1908	3000	slo	kurzíva, kurent (RKP)	1	Andrej Kmet Kmet_Py	36009 49016	28672 26749	4703 4395	1,87% 2,10%	5,79% 5,30%	PyLaia

MODELY TRANSKRIPCIE RUKOPISNEJ KOREŠPONDENCIE ANDREJA KMEŤA

MODELS OF TRANSCRIPTION OF ANDREJ KMEŤ'S HANDWRITTEN DOCUMENTS

DATE	Method	Model	Training set		Validation set		CER accuracy set		Number of cycles (epochs)	CER/WER	
			pages	lines	pages	lines	training	validation		characters	words
20190125	CITlabHT+	10135	125	22549	26	3497	1.15%	3.37%	200	5.97%	21.60%
20190201	CITlabHT+	10410	152	29905	46	4499	1.27%	2.97%	200	6.19%	22.13%
20190205	CITlabHT+	10548	166	29411	46	4573	1.37%	1.84%	200	5.91%	21.87%
20201012	CITlabHT+	26809	111	18071	98	2921	0.44%	7.25%	500	6.08%	21.87%
20210410	CITlabHT+	31888	119	19291	13	3126	1.15%	5.16%	200	3.77%	12.27%
20210821	CITlabHT+	36009	185	28672	26%	4703	1.8%	5.79%	200	2.48%	7.73%

Server Overview Layout Metadata Tools

Logout dusankatuscak@gmail.com

Document... Find

Document Mana User Manager

Versions Jobs

Recent Documents... User activity

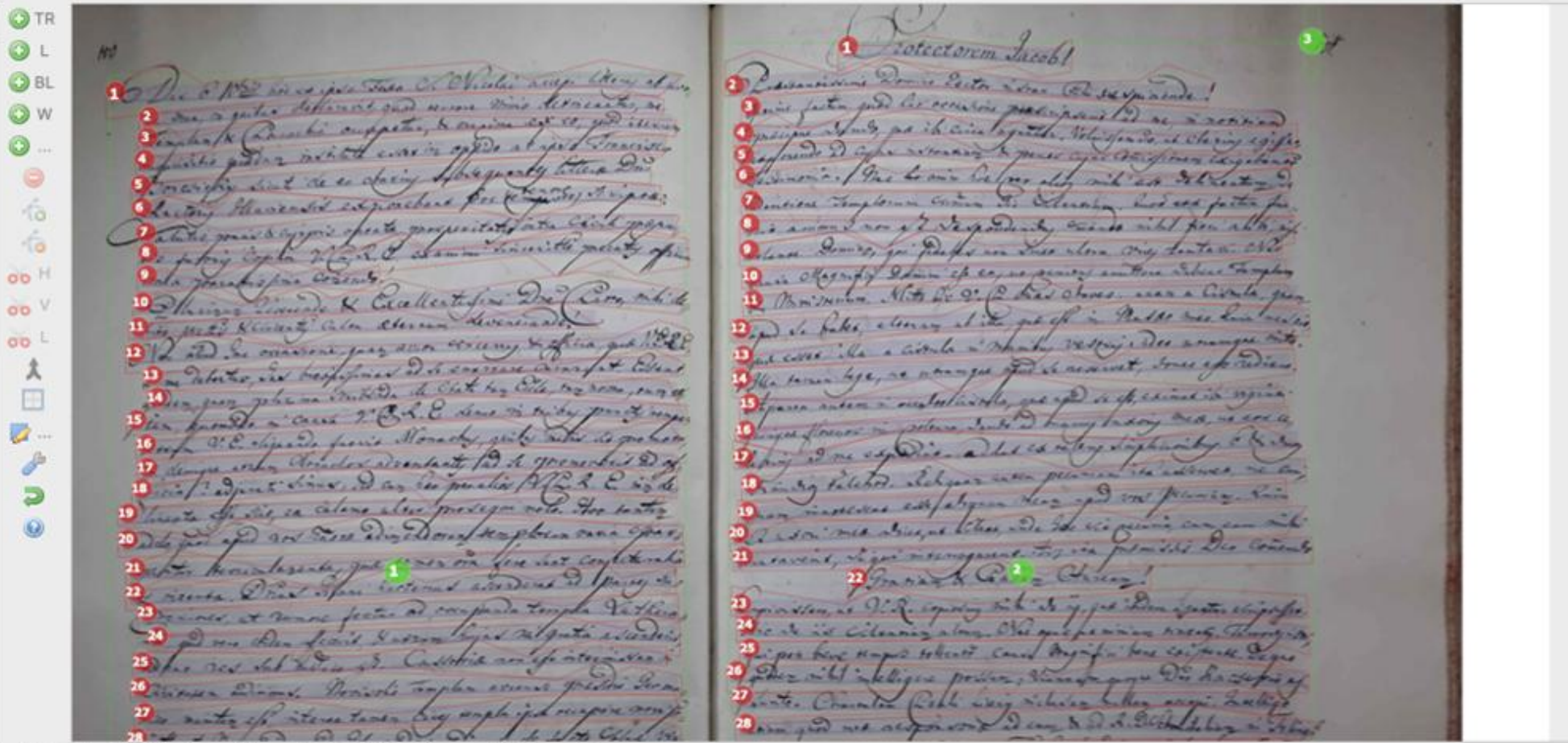
Collections: LAUČEK_SKRIPTOR Col-ID

Documents Model Data

1-44 / 44 1 1

Doc-ID

ID	Title
416681	LAUČEK_MARTIN_SNA_ZV_13_4
416679	LAUČEK_MARTIN_SNA_ZV_13_3
416675	LAUČEK_MARTIN_SNA_ZV_13_2
416674	LAUČEK_MARTIN_SNA_ZV_13_1
416672	LAUČEK_MARTIN_SNA_ZV_10
416303	LAUČEK_MARTIN_SNA_ZV_9
416291	LAUČEK_MARTIN_SNA_ZV_8
416285	LAUČEK_MARTIN_SNA_ZV_7
415335	Laucek_SNK_1069_8
415297	Laucek_SNK_1069_8
415281	Laucek_SNK_1069_7a
415270	Laucek_SNK_1069_7
415219	Laucek_SNK_1069_6
415209	Laucek_SNK_1069_5
415180	Laucek_SNK_1069_4
415171	Laucek_SNK_1069_3_2
415164	Laucek_SNK_1069_3a
415159	Laucek_SNK-1069_3



- 1-1 Die 6 10bris hoc ab ipso Fato S. Nicolai accepi Litteras ab u
- 1-2 ge mea, in quibus declaravit, quod terrore Nimio detineantur, ne
- 1-3 Templum & Parochia occupatur & maxima ex eo, quod iterum
- 1-4 inquisitio quædam instituta esses in oppido ab ipso Francisco
- 1-5 Borcsicky sicuti de eo clarius subsequentes litteræ Dñi
- 1-6 Rectoris Illaviensis exponebant hoc temppestes scriptæ
- 1-7 Salutis, pravus & cuiusvis opitatæ prosperitatis intra Seculi præsen
- 1-8 tis fatoris copiden V. C. R. E. examini sinceritte precatuſ officia
- 1-9 mea paratissima Comendo

HP LaserJet Pro MFP M26nw

Server Overview Layout Metadata Tools

Logout dusankatuscak@gmail.com

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: JESENIK_kovacova (140241, Owner) Col-ID

Documents Model Data

1-15 / 15 Doc-ID

ID	Title	Pages	Uploader	Uploaded	Co
985...	JESENIK_PDF_z_JPG_OREZ	336	dusankatus...	Tue May 03...	(JE
931...	Jesenik_ok 20	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 19	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 18	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 17	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 16	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 15	1	dusankatu...	Sun Feb 2...	(JE
931...	Jesenik_ok 14	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 13	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 12	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 11	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 10	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 9	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 8	1	dusankatus...	Sun Feb 27 ...	(JE
931...	Jesenik_ok 7	1	dusankatus...	Sun Feb 27 ...	(JE

100 Filter

1-1 14.

1-2 zuckhert, und von einen warmen tag in Torten auf

1-3 finger hoch aufgesetzt, von Zeugen darein gefult

1-4 aber darunter die schnitl von getron gelegt.

1-5 rechter hir gebachen, so laufft es schon auf

1-6 Die gar hoch Mande Torten

1-7 Man bereits. 1. lb. Mandt aufs Kleinest, darunter

1-8 5. Kreutter stetzl Putter, vnd 2. ganze frische Bier-

1-9 unnd von 6. Ayern die Cler, und gar wol zu khert



Rozpracované dokumenty a zbierky

- **ANDREJ KMEŤ - CA 3000 S. - RUKOPIS - PREVAŽNE SLOVENČINA**
- **MARTIN LAUČEK - MIN 20 000 S. - RUKOPIS - PREVAŽNE LATINČINA**
- **KUCHÁRSKA KNIHA Z ROKU 1667 - ARCHÍV JESENÍK – 800 S., RUKOPIS, NEMECKÝ KURENT**
- ***MESTEČKO TISOVEC*. KURENTÁLNY PROTOKOL - RUKOPIS, SLOVENČINA, CA 500 S. 1780**
- **ŠKOLSKÁ ZÁPISNICA, RUKOPIS, SLOVENČINA, 1836, CA 60 S. 1836 – RUKOPIS, SLOVENČINA, MAĎARČINA**
- **ČASOPIS LUŽICA 1909 – TLAČ, VLASTNÝ MODEL TRANSKRIPCIA – PRESNOŠŤ 99,2% (BEŽNÉ OCR – 87%-92%)**

Ciele - moje a v projekte Skriptor

1. Od obrázkov k textu a prekladu!
2. Agregované modely

Name

0_SKRIPTOR_WEB_read&search

1_SKRIPTOR_SLOVAK kurent

2_SKRIPTOR_NEOLATIN

3_SKRIPTOR_CZE_GER kurent

4_SKRIPTOR_PRINT_West slavonic

5_SKRIPTOR_STROJOPIS

6. ŠTUDENTI_DIGI

7. Andrej Kmet' writen letters (Slov...

8. SKRIPTOR_LAUČEK Martin