



# 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](mailto: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**

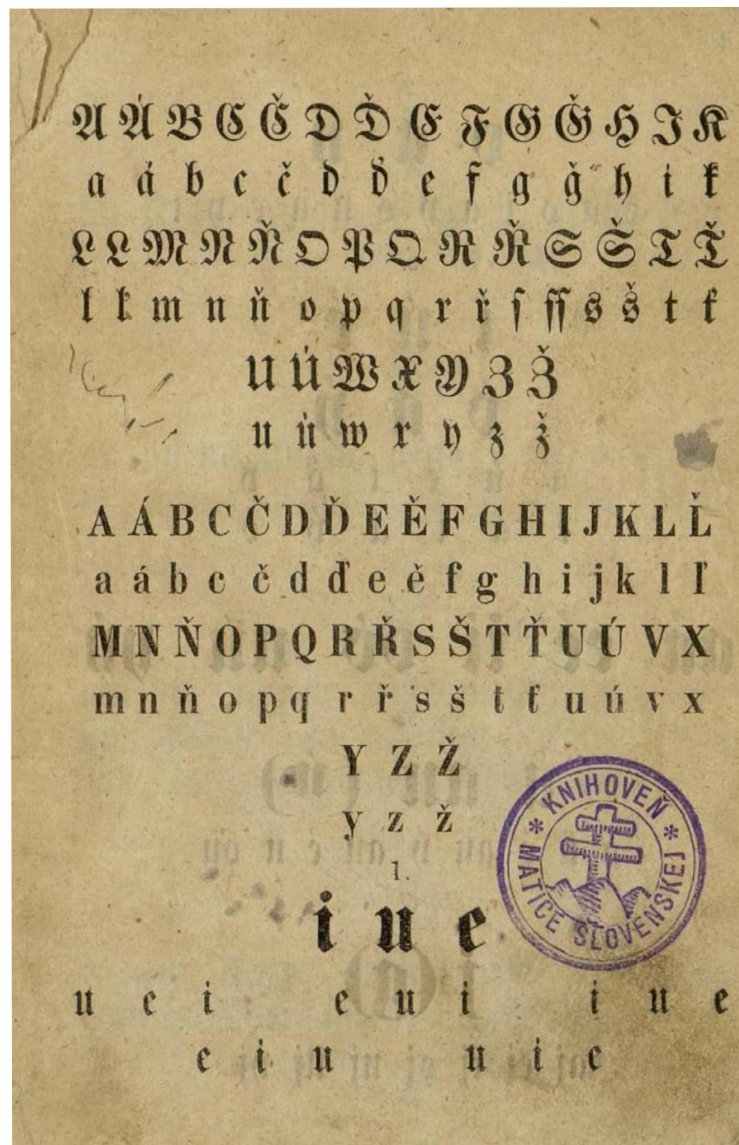


## Fraktúra a antikva

Abeceda v šlabikári z roku 1872.

Fraktúra a antikva

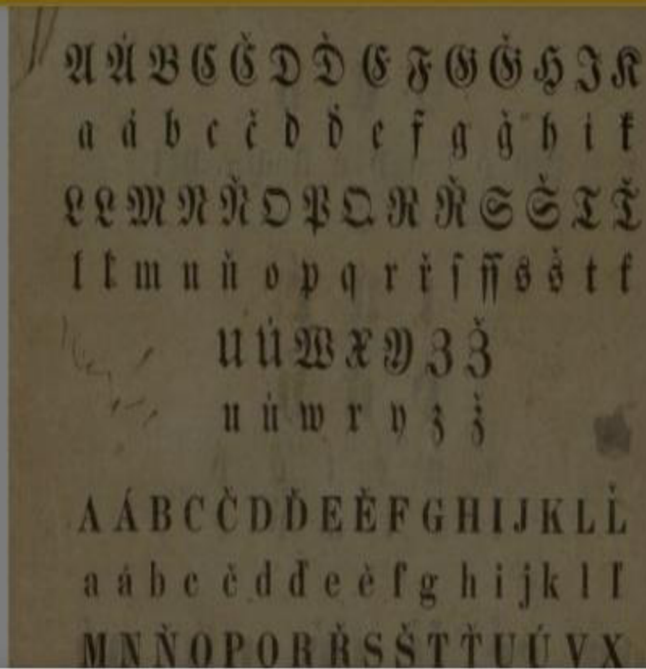
OCR server uspokojivo nerozpoznáva





Hľadať v dokumente ...

3 z 118 stránok



Ewanjelický Šlabikář wypracovaný z naložení wel. Seniorátu gemerského

Autor

Ormis, Samuel

1864-1909

Vydavateľská údaje

V Rožňavě: Tiskem Michala Kovács, 1872

Typ dokumentu

> 21\$ © dd \$ s > © acibcc b j c f g g f ) t t í m n ň o q r í f f ō é t f U ti 28 £ 3 ) 3 3 u t t i D r n \$ | A Á B C Č D Ď E Ě F G H I J K L Ľ a á b c ě d ě e ě f g h i j k l | M N Ň O P Q R A S Š T Ť U Ú V X m n ň o p q r í s š 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 u e

ORMIS, Samuel a 1864-1909. Ewanjelický Šlabikář wypracovaný z naložení 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>

Wysledná kvalita nasnímanej slovacikálnej produkcie zodpovedá kvalite fyzickej predlohy snímaných dokumentov. Defektv ako orezaný text, znížená kvalita tlače, či poškodenie papierového nosiča boli v rámci digitalizačnej linky redukované na minimum snímaním stránok

# 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 rámečku je rozpoznaná celkom neprijateľne a rovnako sú chyby v rozpoznávaní antikvy – červené značky.
- ▶ Zdroj: Slovenská národná knižnica: Katalóg digitálnej knižnice. Dostupné: Ewanjelický Šlabikář wypracowany z nałożeni 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 dokonalo.  
Již svůj denný na oblak běh,  
Vlhou denný se umenšovalo  
A chladnější, byl zepřevr dech;  
Oblehy pak krásou šarlatovan  
Tri kápady byly zbarvené,  
Když sem střebel v lese sílu novou  
Odpočívám v haňšti zelené.

2.  
Do kichanbá sem se samotnosti  
Oddalil od krestu dediny  
Ukrily mne svou vollosti  
Smolau se potici sovčiny.  
Ticho jestit v celém temném háji,  
Vetve sovčím jen se časem hmar,  
Zvuky píšťalky mé slyšaji  
Jed se — ale lisko ukichvan.



Reč Štefana Moyzesa  
pred Forgáčom

1861


Váša Excellencia,  
Najosviätejší paní grof a uhorsko-kráľovský kancelár,  
pane najmilostivejší!

Dvojkrásobná je pohľadka, ktorú nám, Vyšľanstvu a Skutku Slovinskéj, svätej náctvrdá povinnosť, aby sme pred vaším obličím  
Váš Excellencie Vásij pokorne predstavili.

Co sme totiž dľa vysokého úradníckeho postavenia, jako i dľa  
velikodostojného vaju Excellencie Vásij ľahko ľahil mohli, to i  
zo skutočnej zkusenosti našej už zväna, že totiž Excellencie Vasie  
najpoučenejšiu prosbu našu, o potvrdení Skutku Slovinskéj a padron  
živ najjasnejšieho cisárskorostického praveho najvyššieho vysokým  
tvojmú prostredkovaním milostivo potverdovú vás ľah

**Dobšinský**  
Dejiny literatúry

23  
Dejiny literatúry československej.  
Podvrh dejin literatúry československej.  
Deje literatúry československej delíme na 4 doby:  
1, Od najstarších dob až po Husa po r. 1410, alebo vek mladosti;  
2, Od Husa až po bitku na Bielyj Hore do 1620, alebo vek mužskosti;  
3, Od belohorskej bitky až po uvedenie nemčiny do škôl a úradov, alebo vek úpadku do 1774;  
4, Od uvedenia nemčiny až na časy naše, alebo vek znovuzrodenia literatúry českej, ako aj slovenskej, proči, nič zvlášte od Bernoláka.  
Reč československá a jej nárečia.  
a. Česká.  
Českú reč delíme na starú od 9. st. a novú od 11. st.



Dohnány - práce

1842

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

Nám P. Rodicové dobré chęgje, do Brětis-  
lavi nás poslali, a by sme se mi tu na  
učili náš tekauý řewot na dobré winakla  
dati, tedy M. Br. pjalno mosyme všickny sobě  
rozmjšlet, naše powinnosti, co od nás P. R.  
poraduj, a usilowat se máme to vše učiniti,  
aby M. R. na nás se darmo nęstřawati.

Ján Hollý  
rukopis

č. 2. 97. ~~Na Den Mena~~ Na Den Mena. S. N. Derčikovi  
Sak sa teša' vlánci na Sar, keď nastane Teplota,  
A spaňilimi nabé zaodiva' listami Hage:  
Tak sa mnobi' radugu', keď nam Den tvešo dochádza  
Mena pre víborni Muži; však watsšo neciti'  
Kihdo we swem ze wsecb ti Pragicič drdci plesana  
Než ga sam: ponewac' žaden celoročne milegš',  
Žaden aniž mne wzacnegš' nad tento nebiva'  
Vlawni' Den, co ti zaswateni', Dobrodince pripada  
Kecb teda dneš' Dášnici, Kerich obgasunge wissu'  
Duch' Milu' a tubla' amita' pod Hlami Horu'caš'

## Sľowanstwj.

7. Podie sem kle prijklad sobe brati! ,  
Prjlo za krajného učera. Měje se trpytel skwěclami  
na welebných nebesjeh. Ja spēchal iněoblika myšl' do bytj  
prátelj, Mlicemi abychom se wzagemně někde wyrazili  
w rozmluwě. Nejwáhodnějšj farowě mjesto se nam wě  
la některa' ~~hospoda~~ hospoda; neboť sme <sup>ohřij' bjtj' dnes lysesj'</sup> ~~nechtěli~~ <sup>nechtěli</sup> ~~zjstla~~  
kbañneka. Hospoda tedy do nř sme wstoupili. Byl' osāen,  
na; lampy se leskly po klenutě' zjmi oswěcujj twarě  
ctwěraciwnjeh. Postulu, genž to byli pohromade. An sme  
bljže wstoupili; spatřili sme znaměho nam - garozlyj  
genjz - gak se titworj, gra opjly' před Postuluy hospod,  
skěho. Ei, ei pomyslili sme sobě, garowj to ryhřj pried,  
wj! myslit by člověk, kdyby geg nespatrił, že se sworn  
cnostj kleran se na každem smětis ku wychlamba' chel  
za žura do nebe na ohnivem wozě wstoupiti. Poslu,

J. M. Hurban

1838

Prívet

K slov. mládeži a priateľom jej srdca  
v Lipt. Sv. Mikuláši 1/III 1914.

Nie knieža som, nie a vládca tohto sveta,  
čo na porúcki majú prostriedky,  
v tíniach života, tŕnkach, potrebách  
spomáhať veľkodušne, zástiť  
pred doktorom pekelnicia; tu klesajúcu  
starobu podprieť silným ramenom,  
tu a chladnej tóne neustanno vyprst  
zastúhu na výslu — ach! nie som z nich;  
a preba ľah ste, mládež naša drahá,  
tých svojich prestupoch, nepodliqnuť  
srdce Masovarovi nica si zvoliti  
za protektora! — Nie to omyl, sľeba?



Hviezdoslav

1914

Sládkovič - kázeň

1852

A. Praxelova  
v Uhrovi 1852.

Věrazen w ned 11 postau

Pane. nezamítej tvých věřících  
Ans mi w slabostech klesáme,  
Nybož wshledni na nás bojujících  
Když w utrpení padáme.  
O jiu wíru Twého učení,  
Upewniz nawždy w dušech našich  
Kterau, čas, met, lett nezmeni  
Tu ráčs dáti w srdce slabých,  
By ona jak welka' sak měla  
Byla w srdcích Tvých Křestani  
Kam sak jednauc wydobila  
Křisť nebeských občanů ames.

Věriti w Boha snadno se Vám zda' Křestane',  
Ja' zato mám že není snadno. Wšak mnozí před  
ce wíru w Boha tak považují, jako wěc jenž ne-  
škodí když věříme, a není sak těžka jako sobě

Palkovič

Latinský rkp., testament

1.  
Nos Capitulum Ecclesiae  
Metropolitanae 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 Concanonici nostri, die 27<sup>a</sup> Januarii a.c.  
isthic Strigonii interventam, duplici or-  
dine litteralia Instrumenta, inter posthu-  
ma Ejusdem 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 scripturae tractu pa-  
set, pariter propria manu modificata, et  
aucta; - alterum autem in puro, perinde  
autographum Defuncti, quantum quidem ex



Šoltésová

rkp

1 Príspevok k židovskej otázke.

2 Hlas so Slovcami.  
1

3 Miete niet národa na celom našom pozem-  
4 dom svete, ktorý by tak bol po celom ňom rozle-  
5 gnutý, ako je národ židovský - a tiež niet národa  
1 a tom svete, ktorý by práve svoju národnosť  
2 tak zhmýslné, vypočítave zapieral a podľa príhod-  
3 osti privojoväl si ktorúkoľvek inú, národnosť,  
4 radne národnosť byď u 2 či ktorúsi a z ktorých  
5 je - uakolko je ktorá práve pozitívnu mieri. V prav.

## Slovenská filharmónia 1949

Strojopis, programový bulletin

5

1 5. XI. 1949 - 10 hod.

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

3 "Koncert pre študentov."

4 Program

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

6 2. / J. Haydn : "Ochsen-Menuet." ✓

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

8 4. / M. Moyzes : "Mazurka." ✓

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

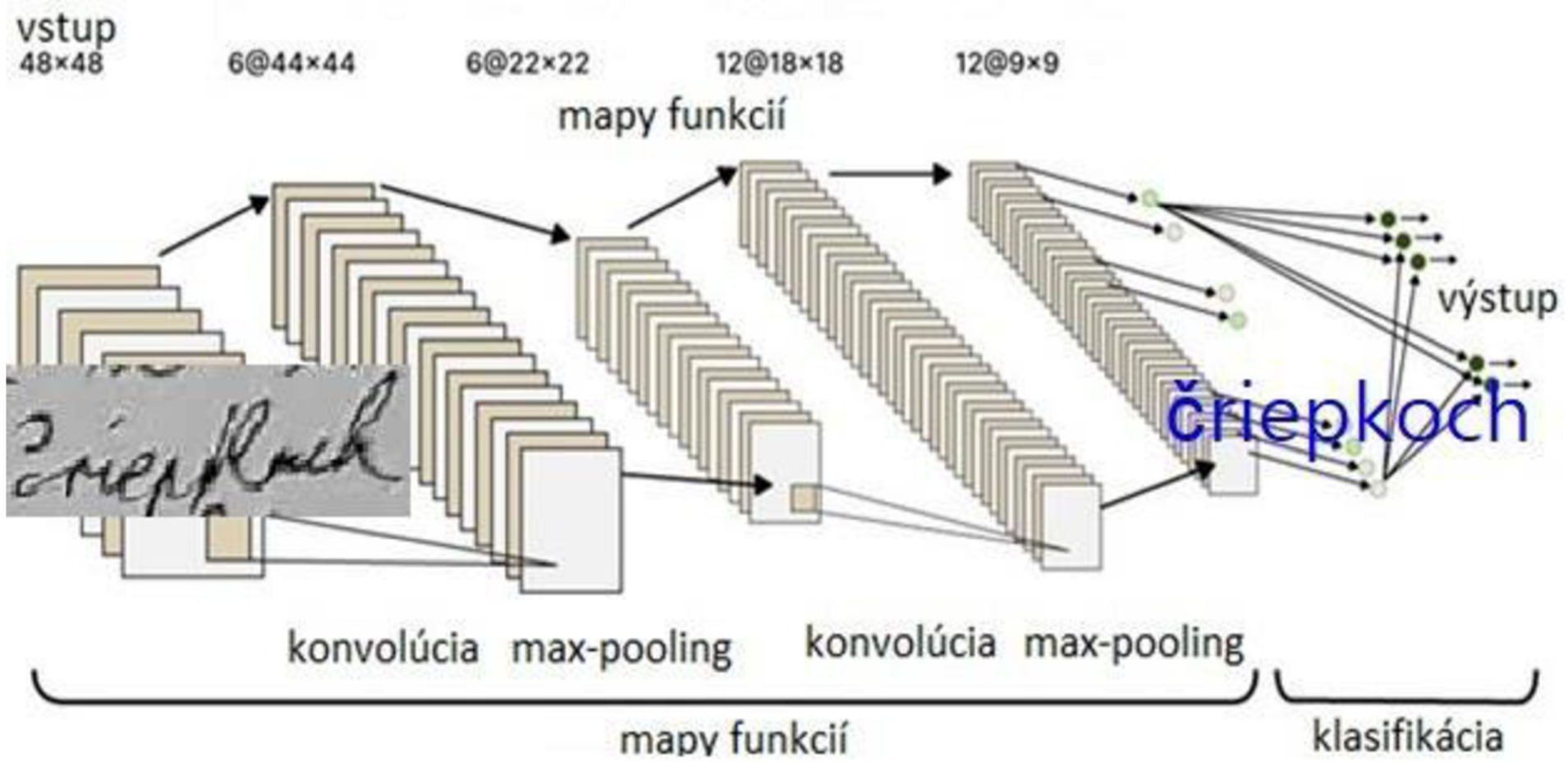
10 6. / " " : "Slovanský tanec čís. 8." ✓

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

12 8. / 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, tréovanie, 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



3  
číepkoch

číepkoch

# Platforma Transkribus



- ▶ *Transkribus expert klient* – [Transkribus | AI powered Handwritten Text Recognition \(readcoop.eu\)](https://readcoop.eu)
- ▶ klúčový inovatívny nástroj pre transkripciu historických rukopisných a tlačených 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

All prices shown on this page include Austrian VAT of 20%.





<b>Suma / kredity</b>	<b>Stroj PyLaia Počet strán rukopis/cena €</b>	<b>Stroj PyLaia Počet strán tlač/cena €</b>	<b>Stroj HTR+ Počet strán rukopis/cena €</b>	<b>Stroj HTR+ Počet strán rukopis/cena €</b>
<b>648 €/3000</b>	3 000/0,216€	18 000/0,036€	2 400/0,27€	15 000/0,043€
<b>1944 €/10000</b>	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>

BÔBOVÁ, 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 <sup>204</sup>	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%

Logout dusankatuscak@gmail.com

Document... Find

Document Mana User Manager

Versions Jobs

Recent Documents User activity

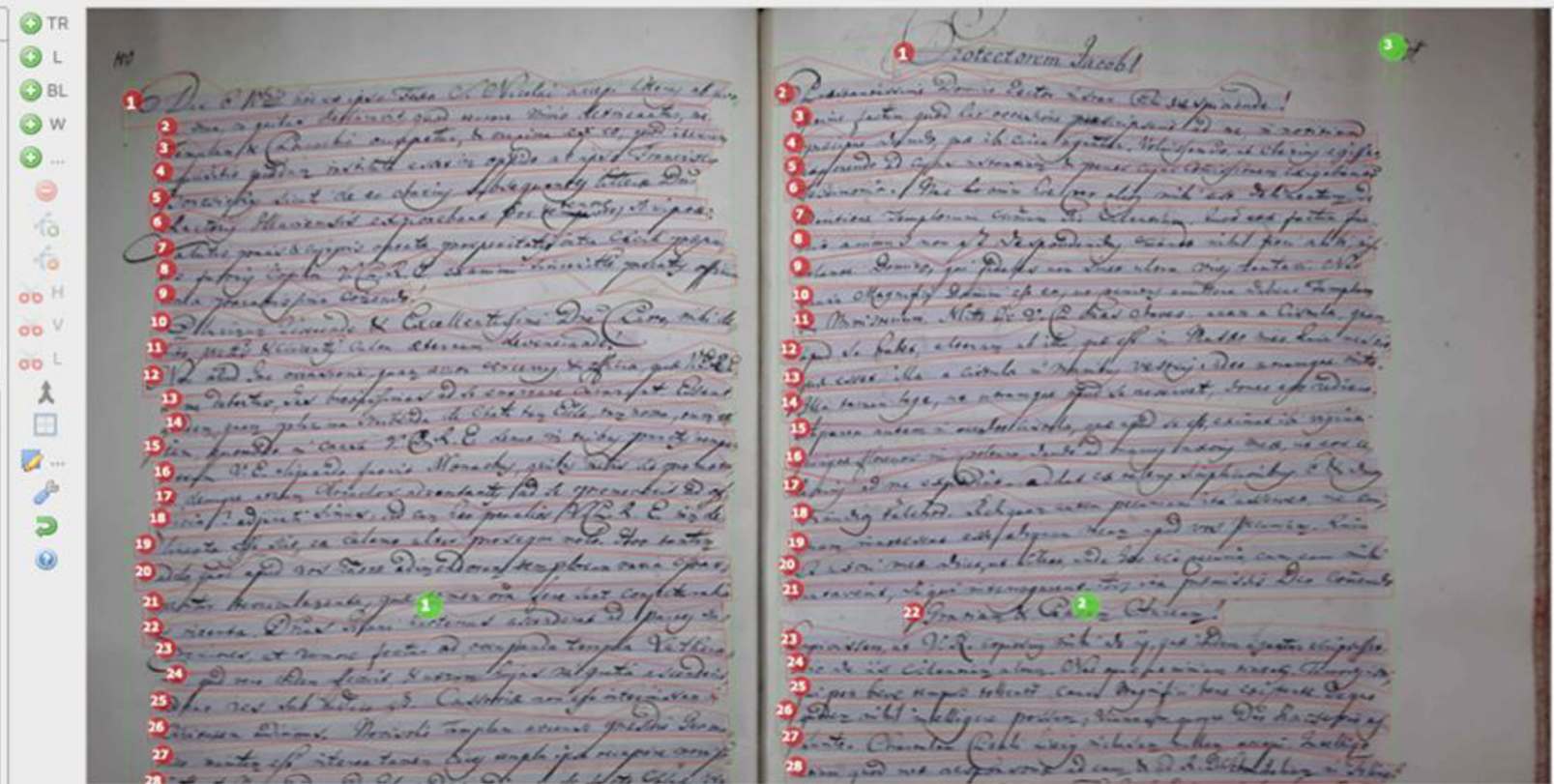
Collections: LAUČEK\_SKRIPTOR Col-ID

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 accepti Litteras ab u
- 1-2 ge mea, in quibus declaravit, quod terrore Nimio detineatur, 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 opitatae 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



Search current document... 1 / 1 In Progress

Server Overview Layout Metadata Tools

Logout dusankatuscak@gmail.com

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: JESENİK\_kovacova (140241, Owner) Col-ID

Documents Model Data

1-15 / 15 1 1 Doc-ID

ID	Title	Pages	Uploader	Uploaded	Co
985...	JESENİK_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...	<b>Jesenik_ok 15</b>	<b>1</b>	<b>dusankatu...</b>	<b>Sun Feb 2...</b>	<b>(JE)</b>
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 zuckert, 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