

Transkribus ako nástroj pre sprístupnenie dobových archívnych pomôcok na príklade Csákósovho katalógu korešpondencie Koháryovcov

Imrich Nagy

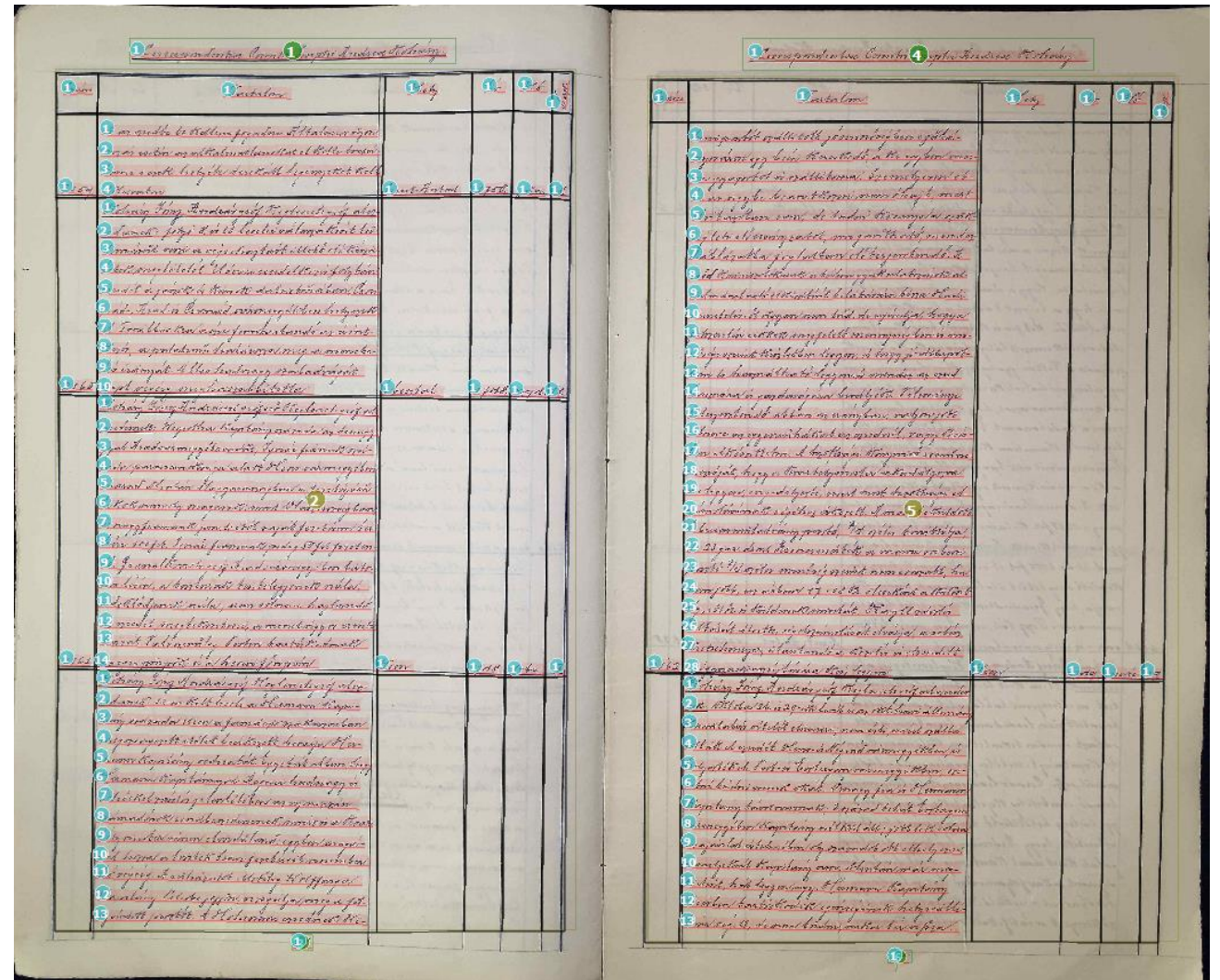
imrich.nagy@umb.sk



Táto prezentácia je výstupom z riešenia projektu APVV-19-0456
SKRIPTOR – Inovatívne sprístupnenie písomného dedičstva
Slovenska prostredníctvom systému automatickej transkripcie
historických rukopisov

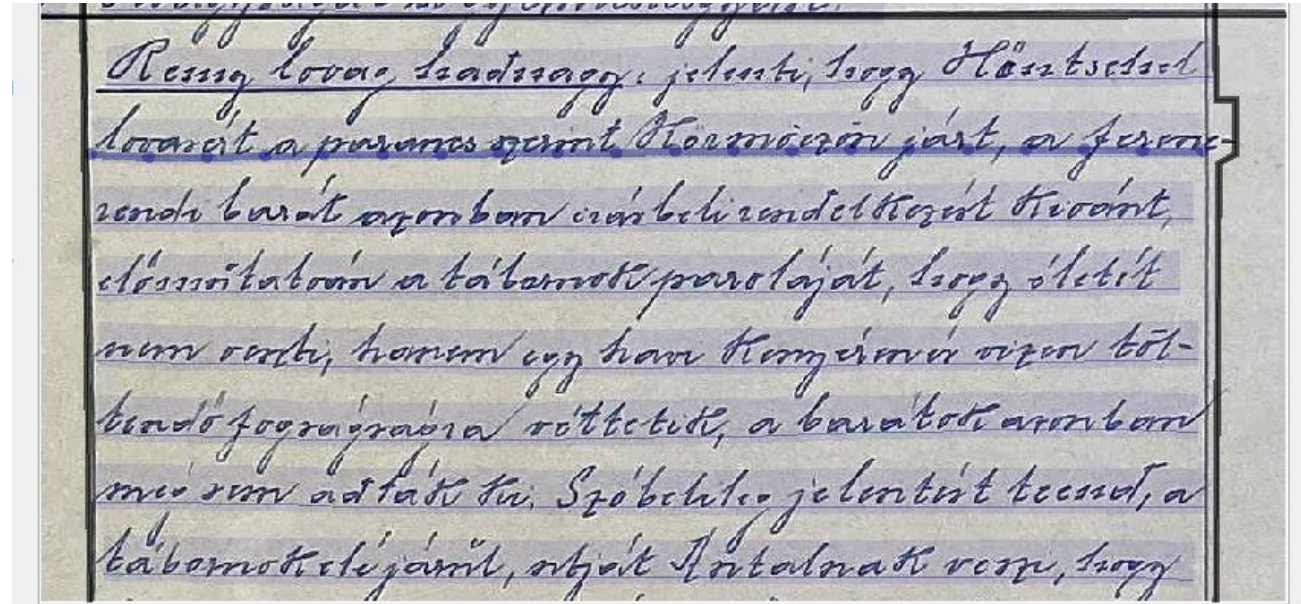
Segmentácia rukopisu (určenie štruktúry a orientácie textu)

- automatická nepoužiteľná
 - manuálna (tabuľky)
- určenie textových rámcov
- určenie hraníc riadkov
- určenie poradia čítania
- časová náročnosť úplnej segmentácie jednej dvojstrany priemerne 10 minút
- spolupráca so študentskou pomocnou vedeckou silou – Matej Smida



Príprava tvorby modelu automatickej transkripcie

- 1. model:** v našom prípade išlo o ručný prepis 29 snímok, čiže digitálnych obrazov predstavujúcich prvých 53 strán zo 4140 stranového rukopisu a obsahujúcich 12 964 slov
- 2. model:** automatická transkripcia a následná korekcia 28 snímok – predstavujúcich ďalších 52 strán (54 – 105) obsahujúcich 14 194 slov
- 3. model:** automatická transkripcia a následná korekcia 29 snímok predstavujúcich ďalších 52 strán (106 – 157) obsahujúcich 16 307 slov



- 33-1 Remy·lovag·hadnagya·jelenti·hogy·Köntschel↵
- 33-2 lovasért·a·parancs·szerint·Körmöczön·járt·a·ferenc↵
- 33-3 rendi·barát·azonban·írásbeli·rendelkezést·kivánt,↵
- 33-4 előmutatóan·a·tábornok·paroláját·hogy·életét↵
- 33-5 nem·veszi·hanem·egy·havi·kenyéren·és·vizen·töl-↵
- 33-6 tendő·fogságságra·véttetik·a·barátok·azonban↵
- 33-7 mégsem·adták·ki·Szóbelileg·jelentést·teend·a↵
- 33-8 tábornok·elé·járul·utját·Antalnak·veszi·hogy↵

Hodnotenie modelu automatickej transkripcie

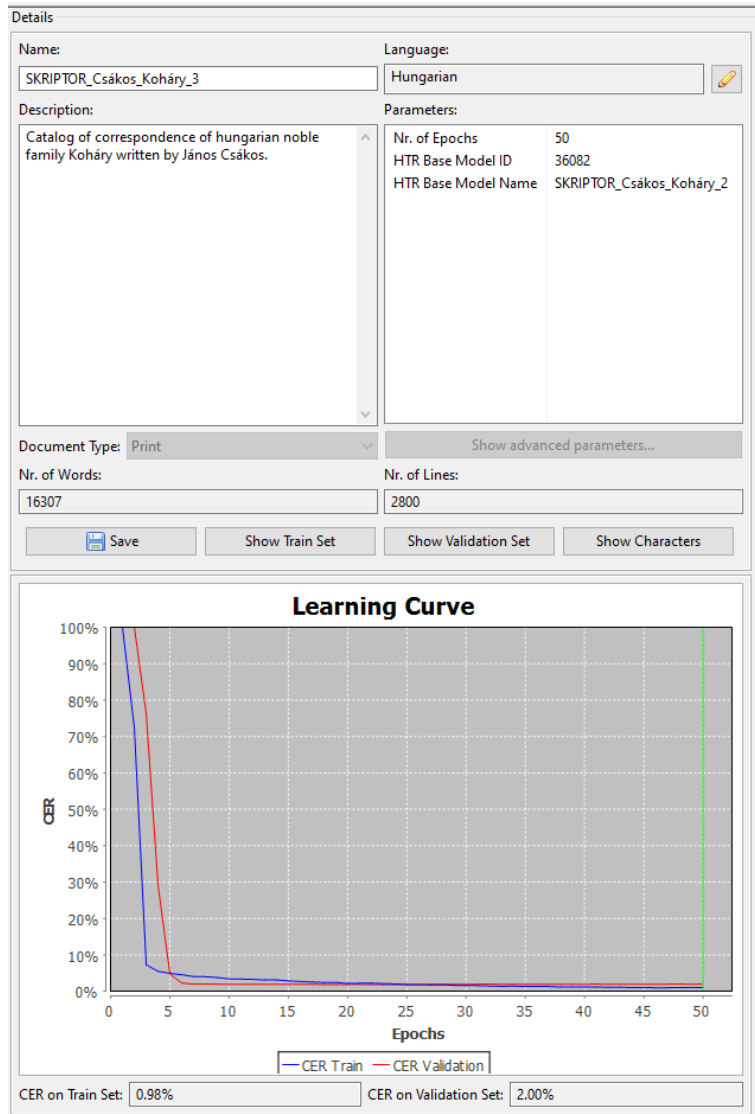
- Hodnotí sa na chybovosti automatického prepisu (CER - character error rate)

Model 1: 4,11% CER

Model 2: 3,08 % CER

Model 3: 2,00 % CER

Model X: 2,67 % CER
(36 772 slov)



Porovnanie modelov 3 a X

Model 3

CER: 1,56 %

WER: 8,37 %

Model X

CER: 1,65 %

WER: 8,25 %

Correspondentiæ Comitum Stephani II. ac Josephi Andreae Koháry.

Szám

Tartalom

Hely

Év

Hó

Nap

2340.

vel Blaskovits írja, írja, de maga is jól tudja, hogy nem érí be avval a pénzzel, melyet Forgách letett részére a szécse-nyvi jászágból-joszágból. Bár csak addig segítse valamivel, míg Castellinéval általeshetne által eshetne az alföldi jászág iránt, hogy ott hitelezőt ott contentálhatná. András öcsésétől is akart kölcsönt kérni, de talán neki is szükség ge léssen a pénzre, és így nem tudott kihez fordulni. Ó benne van minden reménysege, föl akarta keresni, miután úgy értesült, hogy az ünnepek után écsben-Bécsben lesz.

Szent-János

1724.

ápr.

13.

2341.

Koháry Borbála Koháry II. Istvánnak: Köszöni, azt 100 rhénes frt-ot, melyet Blaskovits által adatott. Ugyan még nem vette kézhez, mivel még csak tegnap vette kézhez levelét. Mivel feltett szándékát máius havában véghez nem vihette azért, mert egyik unokája megbetegedett, és nem volt kire bízni a nyavalyást. Most már javulóban van állapota amiak-annak, akinek a kinc-a lába volt görbe. Nagyon sok jót tett vele.

Szent-János

1730.

jún.

20.

Koháry Ignác főkapitány atyjának Koháry András Józsefnek: Február 11.-iki levelét vette. A katonai szöki-szöke-vények ügyében atyja és a vármegye intézkedését köszöni, kéri őket fölküldetni. Mit határozott a szökött katonák katonák ügyében, még nem tudja. De az alispánnak alispánnak Weivel-Weigel hadnagynak Weigel hadnagynak írott 14.-én kelt leveléből kitetszik, Gosztonyi nehezen érkezett a megyegyűlésre, megyegyűlésre, s talán most is függőben vannak a sereg conservatióit célzó s a vármegye által eldöntendő állapotok-állapotok. Az elszökött katonák katonák megküldött lajstromban lajstromban megnevezett Krausz és Oravec-Teschinben Oravec Teschinben fogattak el. A tábornagy hadi törvényt látatott fölöttük, Krausz-Krausz Loszlauban fölakasztatott.-ezt fölakasztatott.-ezt saliat szemivel lát-

Correspondentiæ Comitum Stephani II. ac Josephi Andreae Koháry.

Szám

Tartalom

Hely

Év

Hó

Nap

2340.

vel Blaskovits írja, írja, de maga is jól tudja, hogy nem érí be avval a pénzzel, melyet Forgách Forgách letett részére a szécse-nyvi jászágból-joszágból. Bár csak addig segítse valamivel, míg bastellinéval általeshetne Castellinéval által eshetne az alföldi jászág iránt, hogy ott hitelezőt ott contentálhatná. András öcsésétől is akart kölcsönt kérni, de talán neki is szükség ge léssen a pénzre, és így nem tudott kihez fordulni. Ó benne van minden reménysege, föl akarta keresni, miután úgy értesült, hogy az ünnepek után écsben-Bécsben lesz.

Szent-János

1724.

ápr.

13.

2341.

Koháry Borbála Koháry II. Istvánnak: Köszöni, azt 100 rhénes frt-ot, melyet Blaskovits által adatott. Ugyan még nem vette kézhez, mivel még csak tegnap vette kézhez levelét. Mivel feltett szándékát máius havában véghez nem vihette azért, mert egyik unokája megbetegedett, és nem volt kire bízni a nyavalyást. Most már javulóban van állapota amiak-annak, akinek a kinc-a lába volt görbe. Nagyon sok jót tett vele.

Szent-János

1730.

jún.

20.

Koháry Ignác főkapitány atyjának Koháry András Józsefnek: Február 11.-iki levelét vette. A katonai szöki-szöke-vények ügyében atyja és a vármegye intézkedését köszöni, kéri őket fölküldetni. Mit határozott a szökött katonák ügyében, még nem tudja. De az alispánnak alispánnak Weivel-had-nagynak Weigel hadnagynak írott 14.-én kelt leveléből kitetszik, Gosztonyi nehezen érkezett a megyegyűlésre, megyegyűlésre, s talán most is függőben vannak a sereg conservatióit célzó s a vármegye által eldöntendő állapotok-állapotok. Az elszökött katonák katonák megküldött lajstromban megnevezett Krausz és Oravec-Teschinben Oravec Teschinben fogattak el. A tábornagy hadi törvényt látatott fölöttük, Krausz-Krausz Loszlauban fölakasztatott.-fölkasztatott.-ezt saliat szemivel lát-

Porovnanie chybovosti modelov

Model: SKRIPTOR_Csákós_Koháry_X2

48 623 slov

CER: 2,99 %

Strany originálu	CER	WER
158-159	2,41	12,34
160-161	1,93	9,55
162-163	2,07	10,34
164-165	2,70	11,60
166-167	1,97	9,90
168-169	5,21	14,59
170-171	3,12	13,06
172-173	2,62	12,98
174-175	1,91	10,74
176-177	4,07	17,23
178-179	6,18	19,63
180-181	1,70	8,53
182-183	1,59	8,05
184-185	1,94	10,43
186-187	3,52	13,18
188-189	2,10	9,86
190-191	1,83	9,44

Model: SKRIPTOR_Csákós_Koháry_4

Base Model: SKRIPTOR_Csákós_Koháry_3

8 466 slov

CER: 2,34 %

Strany originálu	CER	WER
164-165	2,46	11,22
172-173	2,49	11,86
180-181	1,72	9,62
188-189	2,12	10,12

**Porovnanie modelov – zmena chybovosti v modeli 4
oproti modelu X**

Strany originálu	CER	WER
164-165	- 0,24	- 0,38
172-173	- 0,13	- 1,12
180-181	+ 0,02	+ 1,09
188-189	+ 0,02	+ 0,26

Výstupy z Transkribusu

digitálny text, ktorý možno exportovať v rôznych formátoch (pdf, docx, txt, xlsx) so všetkými výhodami ďalšieho spracovania textu v príslušnom formáte (napr. vyhľadávanie konkrétneho reťazca znakov či celých slov a výrazov)

Szám	Tartalom	Hely	Év	Hó	Nap.
2159.	it az ezredbe be kellene fogadni. Altalános ujonczozás esetén az alkalmatlanokat el kelle bocsátani s ezek helyébe derekább legényeket kell fölavatni	Szent-Antal	1750.	jun.	8.
2160	<u>Koháry József András gróf Keglevich gróf alezredesnek</u> : folyó 8. és 10. levelei válaszként tudomásul veszi a végrehajtást illető előkészületek megtételét. Udvari rendelkezés folytán ezredét a jászok és kúnok districtusában, Csongrád-, Arad-és Csanád vármegyékben helyezik el. Továbbiakra nézve fenntartandó az érintkezés: a palatinus határozza meg a menetelés irányát. Allio hadnagy szabadságát szept. végeig meghosszabbította.	Ebental	1748.	szept.	18.

Prezentácia výsledkov transkripcie

Pre sprístupnenie ponúka vhodný nástroj samotný Transkribus. Ide o webové rozhranie „read&search“, ktorý ponúka prezeranie digitalizovaného originálu paralelne s jeho prepisom s graficky zvýraznenou konkordanciou na úrovni riadkov.

<https://readcoop.eu/readsearch/>

The screenshot displays the Transkribus web interface. At the top, the logo 'Transkribus lite' is visible, along with navigation links for 'Collection', 'Search', 'Jobs', 'Credits', 'Features', and 'Info'. The user is logged in as 'imrich.nagy@umb.sk'. The current document is 'Koháry_corresp_54_105', page 22. The interface shows a side-by-side view of the original handwritten text and its digital transcription. The transcription is color-coded to show word-level alignment between the two versions. A search bar at the top right contains the text 'GT | imrich.nagy@umb.sk'. Below the document view, a list of search results is shown, including the entry '2239. Esterházy Antal gróf Koháry István Zólyom 1724. ápr. 25.' and a snippet of the transcription: 'II. grófnak: A hűsvé- ti ünnepek alkalmából köszönti; köszöni mélt. gráciáját'. A 'Download' button is located at the bottom right of the interface.

Možnosti vyhľadávania v read&search

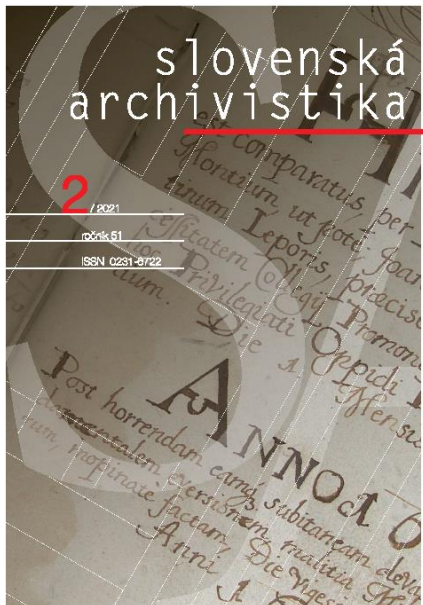
Objektom vyhľadávania sú samostatné slová, ktoré možno upraviť zástupnými znakmi pre rozšírenie vyhľadávania aj na alternatívne tvary slova:

Napr. pri zadaní výrazu „Esterh?zy“ sa zobrazí výskyt mena s diakritikou (Esterházy) aj bez nej (Esterhazy). Podobne pri zadaní výrazu „Antal*“ sa zobrazí výskyt slova Antal a zároveň aj výskyt všetkých tvarov slova s príponami (napr. Antalra, Antalban a pod.).

The screenshot displays the Transkribus Lite search interface. At the top, the logo 'Transkribus lite' is visible on the left, and navigation links for 'Collection', 'Search', 'Jobs', 'Credits', 'Features', and 'Info' are in the center. On the right, the user 'imrich.nagy@umb.sk' is logged in, with a 'Log out' button and a language dropdown set to 'EN'. Below the navigation is a search bar containing the text 'Esterházy' and a 'Search' button. Underneath the search bar, there are five filter buttons: 'Author', 'Uploader', 'Title', 'Collection Name', and 'Script type'. A results summary bar shows '18 Results' and a dropdown menu set to '10'. The first search result is titled '/ Koháry_corresp_1_50 / Page 28' and contains a snippet of text in Hungarian: '... Correspondentiæ Comitum Pauli et Alexandri Esterházy ac Stephani II. et Emer. Koháry Szám Tartalom ... rendezésére. Fülel 1680. jan. 18. 3. Esterházy Pál gróf Koháry István bárónak: ér- tesítii megérkezéséről, s... igazodhatott. Sempte 1678. jul. 22. 4. Esterházy Sándor gróf Koháry István II. grófnak Új évi üdv kívánat...'. Below the text are three horizontal strips of handwritten manuscript with red bounding boxes highlighting the search term 'Esterházy'. The second result is titled '/ Koháry_corresp_1_50 / Page 17' and contains a snippet: '... Földregés és rázkodásokról érkeznek hírek, Bécsűjhelyen is volt ilyen. Pest 1756. jan. 30. Esterházy Borbála ...'. Below this text is another strip of handwritten manuscript with red bounding boxes highlighting the search term 'Esterházy'.

Namiesto záveru – má zmysel transkripcia Csákósovho katalógu?

- segmentácia si bude vyžadovať 690 hodín práce (pri predpokadanom čase 10 minút na spracovanie 1 snímky – dvojstrany z katalógu)
- automatická transkripcia cez nástroj Transkribus: od 373,- € do 684,- € (pri cene od 0,18 € do 0,33 € za 1 prepísanú snímku)
- výsledok digitálny text umožňujúci fulltextové vyhľadávanie, vkladanie metadát, usporiadanie a triedenie v rámci tabuľkového formátu etc.
- otvára možnosť vytvoriť moderné archívne pomôcky pre samotný fond Koháry – Coburg (1241) 1321 – 1945, časť: korešpondencia a nasledujúci archívny/historický výskum



Ďakujem za pozornosť.

Možnosti aplikácie metódy digitálnej transkripcie historických rukopisných textov pri sprístupňovaní archívnych fondov = The Possibilities of application the method of digital transcription of historical manuscript texts in the process of accessing the archival fonds / Imrich Nagy.

In: Slovenská archivistika. - Bratislava : Ministerstvo vnútra Slovenskej republiky, 2021. - ISSN 0231-6722. -

Roč. 51, č. 2 (2021), s. 53-67. Dostupné na:

https://www.minv.sk/swift_data/source/verejna_sprava/odbor_archivov_a_registratur/archivnictvo/slovenska_archivistika/SA%202-2021,%20roc.%2051.pdf