

# Automatikus átírás szlovákiai provenienciájú történeti dokumentumok esetében – tapasztalatok és kutatási perspektívák a Bél Mátyás Egyetemen

Nagy Imrich

[imrich.nagy@umb.sk](mailto:imrich.nagy@umb.sk)



AGENTÚRA  
NA PODPORU  
VÝSKUMU A VÝVOJA

Ezt a munkát a Szlovák Kutatási és Fejlesztési Ügynökség támogatta az APVV-24-0049 szerződés alapján: „DIGINARCH-AI: Történeti dokumentumok digitalizálása és intelligens tartalomelemzése a szakmai levéltári tevékenység és a történettudomány számára.”

# Mesterséges intelligencia a történettudományokban

A történeti kutatás alapja a **történeti forrás**, és erre épül a leendő történészek szakmai képzése is:

- **Források azonosítása és feltárása (forráskutatás)**
  - ✓ Bevezetés a történeti tanulmányokba, kiegészítő történeti segédtudományok
- **Források átírása**
  - ✓ Paleográfia
- **Források fordítása**
  - ✓ A források nyelvei: latin, német, magyar
- **Források értelmezése**
  - ✓ Diplomamunka szeminárium

**Kiválthatók-e ezek a tevékenységek mesterséges intelligenciával?**

Ha az MI mindezt „megtanulja”, marad-e még tér a történész szakmai ítéletének, kreativitásának és felelősségének?

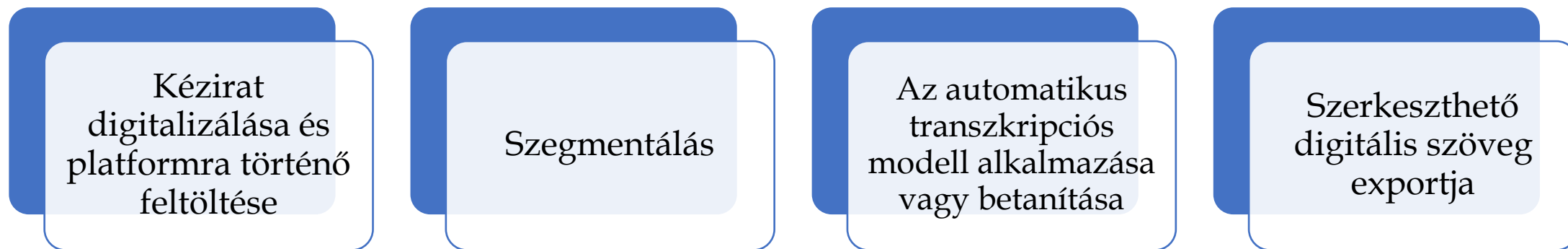
# Eddigi tapasztalataink a mesterséges intelligencia eszközeinek alkalmazásával a történettudomány területén

❖ APVV-19-0456 SKRIPTOR projekt: Szlovákia írásos kulturális örökségének innovatív hozzáférhetővé tétele történeti kéziratok automatikus átírásával

- Projekt időtartama: 2020. 07. 01. – 2024. 06. 30.
- Módszer: kézírásfelismerés (HTR+) a Transkribus gépi tanulási platformon
- Főbb eredmények:
  - Slovak Supermodel M1 (SSM1) – <https://zenodo.org/records/11109087>
  - Slovák Supermodel P&T1 (SSPT1) – <https://zenodo.org/records/11218527>
  - TOMEČEK, Oto – NAGY, Imrich: *A Besztercebányai város 1820. évi reambulációs jegyzőkönyve* (forráskiadás történeti dokumentumok automatikus transzkripciójával)
    - Belianum, Besztercebánya, 2024. – 295 p. – ISBN 978-80-557-2189-7
    - DOI: <https://doi.org/10.24040/2024.9788055721903>
  - A projekt valamennyi eredménye online elérhető:  
<https://repo.umb.sk/communities/3c590b46-9620-4e79-9a19-acb728520bd1>
- Új egyetemi kurzusok bevezetése (MA szint):
  - *Történeti források digitalizálása és hozzáférhetővé tétele*
  - *Történeti kéziratok automatikus transzkripciója*



# A történeti dokumentum automatikus átírásának fázisai



- A modell betanításához szükséges adatok (Ground Truth készlet) előkészítése idő- és szakértelem-igényes folyamat
- A Transkribus legalább ~50, ideálisan ~100 oldal kéziratos szöveg átírását igényli a sikeres tanuláshoz
- Ezért a Transkribus különösen alkalmas nagy terjedelmű, kézírás szempontjából egységes dokumentumok transzkripciójára

# A Koháry-család levelezésének katalógusa - Csákós János feldolgozása (20. század első fele)

UMB\_lmrich.nagy... / Koháry\_corresp... / 148.jpeg Done ↻

Process with AI 0 unsaved changes Save

1 2356.	<p>1 Grospsitsch h. ágens mindezeket írásba foglalhatná.</p> <p>2 Kéri Ő Mélt. értesitené erről sógorát, anélkül, hogy</p> <p>3 öt megnevezné, ajánlja a terv elfogadását, mert</p> <p>4 az ezredes már idősebb ember és meglehetősen testi</p> <p>5 fogyatékosai vannak, ez a teher csak ne-</p> <p>6 hány évre szólna. Levélíró igen örülne, ha a Ko-</p> <p>7 háry-házzal kapcsolatban lehetne is oly parancs</p> <p>8 nokság alá kerülni, mely iránt hódolattal visel-</p> <p>9 tetne.</p>	1 Neutitschein	1 1767.	1 jun.	1 17.
1 2357.	<p>1 Agnes A. József hadnagy, seregbíró Koháry József An-</p> <p>2 drás grófnak: Névnapi köszöntés. Köszöni, hogy ezre-</p> <p>3 desénél beajánlotta, mégis kéri abban a kegyben ré-</p> <p>4 szesíteni, hogy századparancsnokként hozza ja-</p> <p>5 vaslatba a Gyulafehérvárt állomásozó Hoche</p> <p>6 alezredes századánál. Az ezrednél jelenleg szám-</p> <p>7 féletti tisztségek nincsenek.</p>	1 Neutitschein	1 1767.	1 nov.	1 30.
1 2358.	<p>1 Aigner Gaspár kvartélymester Koháry József An-</p> <p>2 drás grófnak: az ezrednél</p>	1 Tordas	1 1738.	1 okt.	1 30.

# Történeti szövegek automatikus átírása a mesterséges intelligencia korszakában

- ❖ A nagy nyelvi modellek (LLM) megjelenésével a gépi tanulás fogalma átkereteződött, és „mesterséges intelligenciaként” kezdtek rá hivatkozni
- ❖ A mindentudó, azonnali és erőfeszítés nélküli megoldást kínáló eszköz illúziója
- ❖ Valóság?
  - Kiváló teljesítmény a szegmentálás terén (az MI sok esetben jobban felismeri a szöveg szerkezetét, mint a Transkribus)
  - A Transkribusnál magasabb felbontású digitalizátumokat igényel
  - Figyelmeztetés nélkül is képes olyan átírást adni, amely nem felel meg a valóságnak, csak imitálja azt
- ❖ Jövő?
  - Az automatikus transzkripció eszközök és a nagy nyelvi modellek integrációja
  - Az MI nem váltja ki a történészt – inkább új típusú szakmai felelősséget és ellenőrzési kompetenciát igényel

**Sorsz. Tárgy**

Sorsz.	Tárgy	Hely	Kelet	Fo.
2261	Gróf Koháry úr arra nevezetett uszodát foglalában, s hogy abbul ő ki ne toloncoltassék, s odább is építhesse részei es mind az lerombolásának s mind az üresen maradt fenntartásának, az udvari népnek nagyobb alkalmatossága végett, azon kívül, hogy mindez iránt ismeretes gondoskodással akarna, az is megígéri, hogy az esztendőnként azoknak, kik az uszodában laknak, ...	Csallóköz	1763. febr. 25.	26
2262	Tekintetes Uram, Szeretett Atyám, méltóztassék kegyesen megengedni, hogy az én szolgám, ki most ottan szolgál, megmaradhasson, s engedtessek meg számára egy kis lakás fenntartása...	Csallóköz	1763. márc. 12.	27
2263	Nagyságos Gróf úrnak méltóságos és kegyes levele, hogy ... az árendásnak meghagyassék, hogy a vásár után való dézsmát megfizesse, melyre nézve a parancsolat irattatott, s azt leróni nem akarván, a ...	Csallóköz	1763. máj. 4.	28
2264	Kérem Nagyságodat szolgálatomat ajánlva, hogy kegyesen méltóztassék ... azon kívánságomat is megengedni, hogy a kért irat elintéztessék és hozzám küldessék	Tardos	1763. aug. 18.	29
2265	Gróf Koháry úrnak szolgálatomat ajánlván, méltóságod kegyességét kérem, hogy az ... elintéztetvén nekem küldessék ...	Veszprém	1763. okt. 12.	30
2266	Méltóságos Gróf úrnak alázatos levele, melyben instál és esedezik, hogy méltóztassék azon házát megtartani...	—	—	31

Szám	Tárgy	Hely	Év	Hó	Nap
2356.	Grospitsch h. ágens mindezeket írásba foglalhatná. Kéri Ő Mélt. értesítené erről sógorát, anélkül, hogy őt megnevezné, ajánlja a terv elfogadását, mert az ezredes már idősebb ember és meglehetősen testi fogyatékoságai vannak, ez a teher csak néhány évre szólna. Levélíró igen örülne, ha a Koháry-házzal kapcsolatban lehetne is oly parancs nokság alá kerülni, mely iránt hódolattal viseltetne.	Neutitschein	1767.	jun.	17.
2357.	Agnes A. József hadnagy, seregbíró Koháry József András grófnak: Névnap köszöntés. Köszöni, hogy ezredesénél beajánlotta, mégis kéri abban a kegyben részesíteni, hogy századparancsnokként hozza javaslatba a Gyulafehérvárt állomásozó Hoche alezredes századánál. Az ezrednél jelenleg számféletti tisztségek nincsenek.	Neutitschein	1767.	nov.	30.
2358.	Aigner Gaspár kvartélymester Koháry József András grófnak: az ezrednél megüresedett kvartélymesteri tisztséget személyével kéri betölteni és kéri egy a hadnagyi rangnak adományozását már azért is, mert a profuntmester zászlóssá neveztetett ki. Hátíratként a válasz, a melyben tudatik vele, hogy amúgy is tb. zászlósnak neveztetett ki.	Tordas	1738.	okt.	30.
2359.	Aigner Gáspár zászlós, kvartélymester Koháry József Anhas grófnak: Mult hó 28.-án kelt rendeletére Mosanner hadn.nak a csatolt specificatio szerinti járandóságait a Medgyesen utalványozott, de visszaszolgáltatni rendelt pénzek helyett a bécsi bankhivatálnál utalványozott 10,000 frt-ot illetőleg utalványt előterjeszti. Beküld egyben egy kimutatást; meny nyi portio jár az ezrednek és miként rendeztettek ez illetékek, hogy Wunderlich zászlós az elszámolás kapcsán a hátralékokat megállapíthassa.	Varaglia	1739.	febr.	12.

# Az MI felhasználási lehetőségei a történettudományokban

- Szövegfeldolgozás
- Szövegfordítás
- Szövegösszefoglalás / sumarizáció
- Tartalomelemzés



A szöveg tényleges hozzáférhetővé tétele

- ✓ Az eredményesség és relevancia alapvetően függ az adatt mennyiségtől, vagyis a digitalizált szövegtől, amely az elemzés tárgya
- Ezeket a hipotéziseket a gyakorlatban kívánjuk tesztelni és igazolni az APVV-24-0049 DIGINARCH-AI projekt keretében „Történeti dokumentumok digitalizálása és intelligens tartalomelemzése a szakmai levéltári tevékenység és a történettudomány számára”
- A projekt időtartama: 2025. 09. 01. – 2029. 08. 31.
- **Konzorcium tagjai:**
  - Bél Mátyás Egyetem, Besztercebánya
  - Kassai Állami Levéltár
  - Szlovák Nemzeti Könyvtár, Turócszentmárton
  - Fórum Kisebbségkutató Intézet, Somorja
  - Állami Tudományos Könyvtár, Besztercebánya
- A projekt témája: a lakosság kényszerű áttelepítése Szlovákia és Magyarország között (1945 – 1948)

# Az előttünk álló kutatási kihívások

## ✓ digitalizálás és levéltári dokumentumok konzerválása

- a Telepítési Hivatal szlovákiai kirendeltségeinek (Kassa és Rimaszombat) teljes levéltári anyaga, összesen körülbelül 58 folyóméternyi dokumentum
  - különféle kritériumok szerint összeállított, kényszertranszportokba besorolt személyek névjegyzékei
  - az áttelepítésre, illetve kivételek megadására vonatkozó döntések indokai
  - az áttelepítettek utáni vagyonok birtokba vétele
  - az áttelepítésre vonatkozó politikai döntések
  - személyes kérelmek és panaszok
  - az áttelepítéssel összefüggő helyi hivatalok ügyrendje

## ✓ automatikus transzkripció modellek kifejlesztése az AI-alapú Transkribus eszköz számára a dokumentumok automatikus átírásához

- egy nagyméretű digitális szöveges adatbázis létrehozása

# Az előttünk álló kutatási kihívások

- ✓ intelligens modell megtervezése és megvalósítása történeti szövegek kulcsrészleteinek összefoglalására (sumarizációjára), valamint azok bemutatása speciálisan tervezett alkalmazáson keresztül
  - Az alkalmazás két fő modulból fog állni:
    - egy kommunikációs modulból, amely biztosítani fogja a levéltári alkalmazott és a létrehozott rendszer közötti kommunikációt természetes nyelven, az alkalmazás felhasználói felületén és az NLP (angolul: natural language processing) modellen keresztül
    - egy összegző (szummarizációs) modulból, amely biztosítani fogja magát a létrehozott adatbázisban tárolt történelmi dokumentumok szövegének összegzését
- ✓ a szóban forgó levéltári fondok tartalomelemzésének elvégzése a háború utáni Szlovákia történeti kontextusában, a digitalizált dokumentumokból nyert összefoglaló MI-eredmények felhasználásával

## Végezetül a bevezetőben feltett kérdésekre adott válasz

1. MI a történeti tudományokban a közeljövőben nem lesz képes teljes mértékben helyettesíteni az emberi szakmai tevékenységet
2. MI jelentősen felgyorsíthatja a kutatás egyes fázisait a szakember – az erudált történész – felülvizsgálata alatt
3. Kiterjedt adatbázisok esetében a tartalomelemzési képességének köszönhetően értékes adatokat kínálhat fel különféle statisztikai korrelációk formájában, amelyeket az erudált történész értelmezhet
4. Következésképpen a jövő történészeinek szakmai képzésében elengedhetetlen, hogy megfelelő teret teremtsünk a mesterséges intelligencia eszközeinek alkalmazására a történész munkájának egyes fázisaiban

Köszönöm a figyelmet!