

VALÓS TÉRBEN – AZ ONLINE TÉRÉRT

Networkshop 31: országos konferencia

2022. április 20–22.
Debreceni Egyetem

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2022



A kötet megjelenését támogatta az
Energiaügyi Minisztérium

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Workshop

2022. április 20–22. Debreceni Egyetem, konferencia előadásainak közleményei

ISBN 978-615-82243-0-7

DOI: [10.31915/NWS.2022](https://doi.org/10.31915/NWS.2022)

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével
Budapest
2022

Borítókép: [freepik.com](https://www.freepik.com)

TARTALOMJEGYZÉK

Előszó	5
Lencsés Ákos: A nyílt tudomány pénzügyi vonatkozásai	7
Farkas Katalin: Centenáriumi média-adattár és virtuális kiállítás létrehozásának tanulságai az SZTE Klebelsberg Könyvtárban	13
Bódog András: A nyílt archívumi információs rendszer (OAIS) szabványának honosítása.....	20
Perlaki Attila: Oktatást segítő gamifikációs alkalmazások, mint szakdolgozati témák	27
Csapó Noémi – Dani Erzsébet: APPropó fejlődés – A Bács-Kiskun Megyei Katona József Könyvtár mobilapplikációja.....	32
Simon András: Integrált könyvtári rendszerek tranzakciós rekordjainak vizsgálata, a könyvtári állomány digitalizálásának tervezésekor.....	41
Németh Márton: Az OSZK Webarchívum nemzetközi kapcsolatai.....	58
Antal Péter: A mesterséges intelligencia kihívásai a XXI. század társadalmára	70
Hajdu Csaba – Szilágyi Zoltán: Modern robotikai technológiai ismeretek oktatása „Teljes spektrumú” oktatási módszerrel	77
T. Nagy László – Boda István Károly – Tóth Erzsébet: E-tananyagfejlesztés virtuális 3D környezetben.....	84
Palencsárné Kasza Marianna: Digitális átállás – Minőség – lehetőségek az EQAVET terén.....	92
Nagy Gyula: Nemzetközi kitekintés a felsőoktatási könyvtárak világára: a EUGLOH könyvtári workshopja	99
Babocsay Gergely: Az európai természettudományi gyűjtemények digitális integrációja: határ a csillagos ég.....	108
Somorjai Noémi: Egyenlőtlenségek a tudományos kutatás területén. Az amatőr kutatók szerepe	114
Molnár Dániel – Dani Erzsébet: Robotok a könyvtárban: Hogyan válhat a robotika a könyvtári mindennapok részévé?	122
Horváthné Felföldi Helga: Digitalizáció a szakképzésben. A Szakmajegyzékben szereplő szakmák digitáliskompetencia jártassági szintjeinek felülvizsgálata	130
Kalcsó Gyula: Ne csak útra csomagoljunk! Miért fontos a csomagolás a digitális megőrzésben?	138
Karsa Zoltán István – Szeberényi Imre: A CIRCLE felhő elmúlt évtizede	146
Bobák Barbara – Kasza Péter: Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus <i>Rerum Ungaricarum</i> libri kéziratának digitális kiadása (esettanulmány)	154
Egyed-Gergely Júlia – Vajda Róza, Gárdos Judit – Horváth Anna – Meiszterics Enikő – Micsik András – Martin Dániel – Marx Attila – Pataki Balázs – Siket Melinda: Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok	161
Szemes Botond – Bajzát Tímea – Fellegi Zsófia – Kundráth Péter – Horváth Péter – Indig Balázs – Dióssy Anna – Hegedüs Fanni – Pantyelejev Natali – Sziráki Sarolta – Vida Bence – Kalmár Balázs – Palkó Gábor: Az ELTE Drámakorpuszának létrehozása és lehetőségei.....	170



Sebestyén Ádám: Az ELTEdata szemantikus adatbázis legújabb fejlesztései.....	179
Szlamka Erzsébet: Új trendek a tanulási eredmények tanúsításában	185
Tóth Máté – Héjja Balázs: Webshop indítása közkönyvtári környezetben.....	192
Etlinger Mihály – Hernády Judit: A kiadás hagyatéka / a hagyatéka kiadása: A Régi Magyar Költők Tárának hálózati kiadásáról.....	199
Varga Emese – Makkai T. Csilla: „Ki a fenének kell collstok?” A digitális szöveg rejtett mértékegységei	204
Dobás Kata – Fazekas Júlia: ITIdata – Egy irodalmi adatbázis fejlesztése Wikibase alapon és ennek hasznosítása Kosztolányi Dezső forrásjegyzékénél	211
Sörény Edina: Kézai Simon Program – digitális családi fotóarchívum.....	219
Fülöp Tiffany – Molnár Tamás – Hoczopán Szabolcs: Open Monograph Press e-könyvplatform a Szegedi Tudományegyetemen	227
Palkó Gábor: Mesterséges intelligencia, digitális bölcsészet, kulturális örökség: trendek és eredmények.....	235
Pergéné Szabó Enikő – Bátfai Mária Erika: A tudományos publikálás támogatása a Debreceni Egyetemi és Nemzeti Könyvtárban	241
Csirmazné Rezi Éva: Nemzetközi kiadványazonosítók és kötelezpéldányok kezelése az OSZK OKP (Országos Könyvtári Platform) rendszerében	250
Alföldi István – Dióssy Anna Laura: Digitálisan született kutatási anyagok megőrzése: a relációs adatbázis mint born-digital objektum	262
Fekete Norbert: HTR-modellépítés és kézírásfelismerés nagyméretű, többszerzős szövegtörzsen. A Transkribus alkalmazása az Arany János hivatali iratokon.....	271
Horváth Péter – Kundráth Péter – Palkó Gábor: ELTE Népdalkorpusz – magyar népdalok gépileg annotált adatbázisa	276
Nagy György: IKT eszközök alkalmazása az alsó tagozatos környezetismeret órákon.....	284
Köpösdí Zsuzsa – Molnár Tamás: Multimédiás, interaktív és adaptív tananyagok létrehozásának lehetőségei H5P keretrendszerrel	289
Jankó Tamás: Munka 4.0 – Ipar 4.0 – Szakképzés 4.0 – : A digitális kompetencia jövőbeni fejlesztési útjai	296
Békésiné Bognár Noémi Erika – Nagy Andor: Megújuló könyvtári statisztika: az egységes adatstruktúra és a korszerű megjelenítés kialakításának útján	304
Bolya Máttyás: Kézírtos dallamlejegyzések feldolgozása MI-vel támogatott digitális környezetben	310
Maróthy Szilvia – Seláf Levente – Vigyikán Villó: Régi magyar verskorpusz összeállítása stilometriai és számítógépes metrikai kutatásokhoz	324
Szűcs Kata Ágnes: Kézírtos források transzformációinak lehetőségei a közgyűjteményekben.....	330
Fellegi Zsófia: A digitális filológia infrastruktúrái. A DigiPhil megújulásáról.	338
Mihály Eszter: Mi az a dHUpla? A Digitális Bölcsészeti Platform bemutatása.....	345
Nemeskey Dávid Márk – Palkó Gábor: Szemantikus névelém-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása)	359

Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus
Rerum Ungaricarum libri kéziratának digitális kiadása¹
(esettanulmány)

Bobák Barbara
Magyar Tudományos Akadémia Bölcsészettudományi Kutatóközpont
Irodalomtudományi Intézet, DigiPhil
bobak.barbara@abtk.hu

Kasza Péter
SZTE BTK Klasszika-Filológia és Neolatin Tanszék
kasza.peter@oszk.hu

2020 pandémia sújtotta nyarán Petneházi Gábor kollégával rendkívüli felfedezést tettünk az olaszországi Trento érseki könyvtárának (*Vigilianum*) kéziratai között: megtaláltuk Báthory István erdélyi fejedelem (1571–1586) és lengyel király (1576–1586) udvari történetírójának, az olasz Gian Michele Brutónak (latinos alakban használt nevén: Brutus) egy addig ismeretlen, magyar történeti tárgyú kéziratát. A felfedezés nemcsak szűkebb szakmai berkekben keltett komoly visszhangot, de a hír a nyomtatott és elektronikus sajtó ingerküszöbét is átlépte, és ennek nyomán szinte az összes komolyabb országos és vidéki lap beszámolt a Brutus-kézirat előkerüléséről.

Az alábbi közleményben egyrészt arról számolunk be, miért volt olyan jelentős ez a felfedezés, illetve milyen kihívások elé állította a szöveg feldolgozására vállalkozó filológusokat, másrészt arról, hogy a modern digitális technológia hogyan segített megfelelni ezeknek a kihívásoknak.

Egy európai rangú humanista és műve

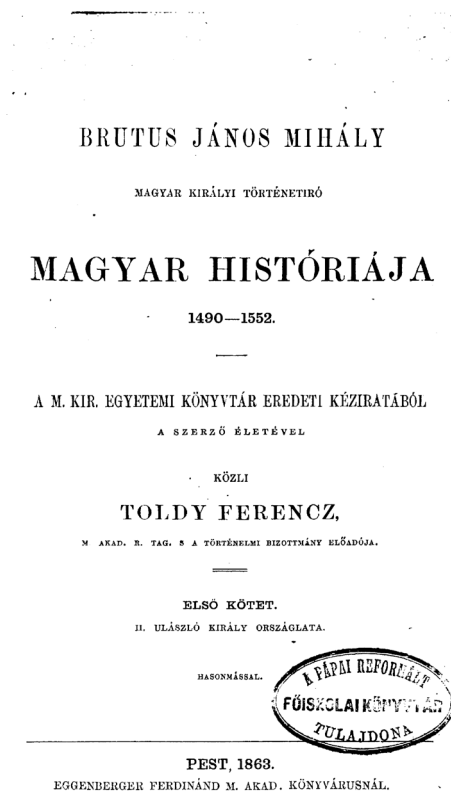
Gian Michele Bruto 1517-ben született Velencében. Ágostonos szerzetes volt, aki utóbb, a protestáns eszmék hatására kilépett a rendből. Képzett humanista volt, aki tudását igyekezett különböző fejedelmi udvarokban értékesíteni és így megélhetéshez jutni. Ennek folyamánként szinte egész Európát beutazta és jelentős hírnévre tett szert, mint jó tollú és szakavatott történetíró.

Magyarországon a 16. század a történetírás virágkora. Ennek hátterében a politikai helyzet keresendő. A Mohács után (1526) előbb két, majd Buda elestével (1541) három részre szakadó ország humanistái igyekeztek választ találni a kínzó kérdésre, mi okozhatta a 15. század végén még közép-európai nagyhatalomnak számító Magyar Királyság váratlan és szédítő gyorsaságú összeomlását. Az okok és felelősök keresése vezetett 1526 után a kortörténetírás felvirágzásához. Erősítette a tendenciát, hogy az országban két hatalmi központ jött létre: a Bécsből irányított Habsburg-oroszág rész állt szemben a Gyulafehérvár centrummal előbb a Szapolyai, majd a Báthory család uralma alatt fokozatosan formát öltő Erdélyi Fejedelemséggel. A két versengő államalakulat nemcsak a csatamezőn, de a propaganda mezején is háborúban állt egymással, így mindkét udvar humanisták sorát alkalmazta, hogy megírják a számukra kedvező, az ő legitimitációjukat erősítő kortörténeti narratívát. Brutus meghívása ebben a kontextusban merült föl még 1571 körül, majd a nagy tekintélyű történetíró ténylegesen 1574-ben érkezett meg Erdélybe és állt Báthory István fejedelem szolgálatába. Feladata a magyar történelem Mátyás halála és Báthory trónra lépte (1490–1571) közti évtizedeinek megírása volt. Korabeli levelezésekből származó adatok szerint a *Rerum Ungaricarum Libri*

1 Jelen tanulmány a Digitális Örökség Nemzeti Laboratórium és az OSZK-SZTE Kulcsár Péter Historiográfiai Kutatócsoport támogatásával készült.

című, monumentális latin nyelvű munka húsz könyvre tagolva az 1582–83 évekre már készen állt, de Báthory haláláig (1586) nem jelent meg. Brutus ekkor előbb váratlan húzással Habsburg-szolgálatba lépett, és megpróbálta némi átdolgozás után Bécsben kiadatni élete fő művét, de nem járt sikerrel. A kutatás jelenlegi állása szerint 1591-ben azért tért vissza Erdélybe a *Rerum Ungaricarum Libri* legutolsó változatával aládájában, hogy ott próbálja kiadni a művet, de halála (1592) meggátolta ebben.

A kiadásra csak évszázadokkal később, a 19. század második felében kerülhetett sor. De mire a Magyar Tudományos Akadémia megbízásából Toldy Ferenc hozzáfogott, hogy sajtó alá rendezze Brutus-művét, már csak két, erősen csonkult kézirat állt a rendelkezésére. Noha Brutus levelezése alapján tudni lehetett, hogy a *Rerum Ungaricarum Libri* húsz könyvet tartalmazott, ezek tetemes része a 19. századra elveszett. A bécsi *Nationalbibliothek*ban levő példány csak az 1–3, illetve 6–8. könyveket őrizte meg, tehát összesen hat könyvet, az eredeti mennyiség alig harmadát. A pesti Egyetemi Könyvtár két kötetes példánya ennél jóval bővebb szöveganyagot tartalmaz: hézagokkal ugyan, de megvan benne a *Rerum Ungaricarum Libri* 1–13. könyve, illetve a 16. egy kisebb részlete, vagyis hozzávetőlegesen a teljes mű kétharmada. Árnyalja ugyanakkor a képet, hogy míg a pesti példány 1–8. könyve gondos és megbízható másolat, addig a maradék könyvek nem csupán töredékesek, de számos kéztől származó, hevenyészett, sokszor értelmezhetetlenül romlott szöveget őriztek csak meg. Toldynak tehát ebből kellett gazdálkodnia és ezekre alapozta a három kötetben megjelent szövegkiadást.



1. kép: Toldy Ferenc által sajtó alá rendezett Brutus-mű főcímlapja

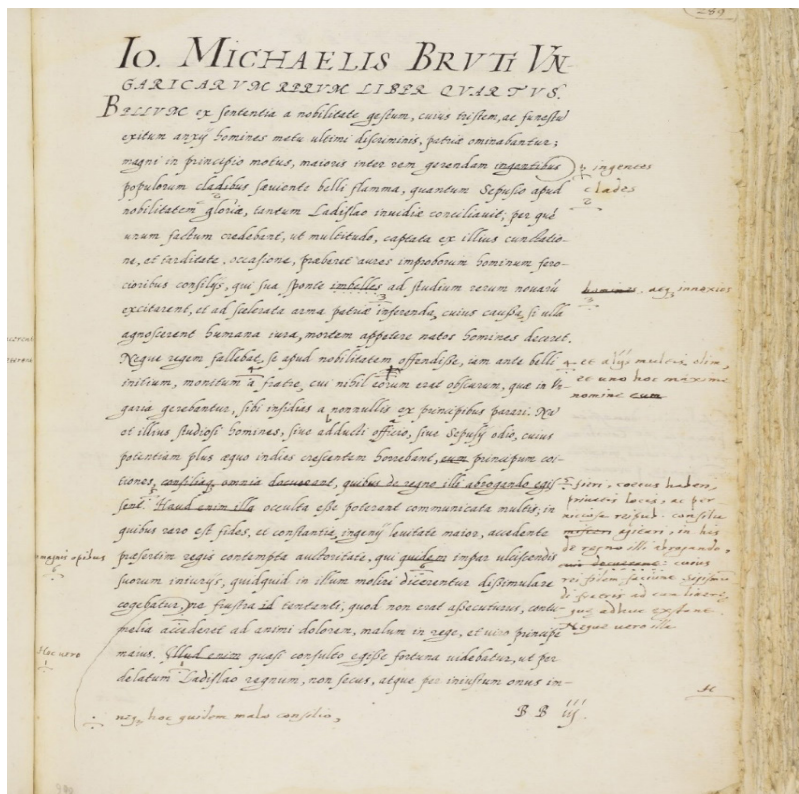
A trentói kézirat jelentőségét viszont az adja, hogy lényegében teljesnek tekinthető. Brutus autográf javításait tartalmazza, így hitelessége megkérdőjelezhetetlen. Két kötetben, összesen 2200 oldalon hozza a *Rerum Ungaricarum Libri* teljes húsz könyvét. Az utolsó könyv ugyan

mondat közben megszakad, de minden jel arra mutat, hogy nem a kézirat csonkult, hanem maga Brutus nem írta tovább a szöveget. A felfedezés így lehetőséget kínál arra, hogy teljességében megismerjük a 16. század egyik legjelentősebb történetírójának, mindeddig éppen a hibás kéziratokra alapozott, nehezen érthető és csonka kiadás miatt elhanyagolt fő művét.

Ilyen terjedelmű és jelentőségű kézirat lényegében évszázados távlatban nem került elő. Így jogos elvárás, hogy legalább a latin szöveget mihamarabb hozzáférhetővé tegyék a szakmai közösség számára, még ha ez nem is jelent rögtön a kéziratok összevetésén alapuló, tudományos igényű kritikai kiadást, pláne nem a nagyközönség számára is hozzáférést biztosító magyar fordítást. De a mielőbbi közzététel kívánalmain túl a szöveget feldolgozó kutatók számára is létkérdés, hogy a hatalmas kéziratot és annak óriási szöveganyagát valahogy kezelhetővé, kereshetővé és így megközelíthetővé, feldolgozhatóvá tegyék. Ez praktikus azt jelenti, hogy a kéziratot számítógépre kell vinni. Nem elég a digitális fotókról beszélni, hiszen azok csak azt biztosítják, hogy a fizikailag Trentóban levő kéziratot megbízható kópia nyomán bárhol tanulmányozni lehessen. A *Vigilianum* jóvoltából és a római Magyar Nagykövetség közbenjárásának köszönhetően a kézirat teljes digitalizált változata már a birtokunkban van, de ez még nem oldja meg a szöveg kezelhetővé tételének problémáját.

Még egy évtizede sem nagyon lett volna erre más megoldás, mint hogy a kéziratot sorról sorra begépeljük. Csakhogy 2200 oldalnyi sűrű szövegről van szó. Ennek hatékony átgépelése, még paleográfiai gyakorlattal rendelkező kutató számára is komoly terhelést jelent. Egy nap átlagosan mindössze 4-5 oldal átírásának elkészítése lehetséges, így a 2200 oldal átírása egy ember közel másfél-két évi munkáját jelentené. Nem beszélve arról, hogy a szövegben óhatatlanul számos hiba, elütés marad. Ha az ellenőrzést is hozzávesszük, lényegében a pusztán és valamennyire megbízható közlés közel három évet vesz igénybe.

Számunkra azonban kivételes lehetőséget biztosított az a tény, hogy bár hatalmas mennyiségű szövegről van szó, a 2200 oldalból 2000 teljesen egységes kézírással íródott.



2. kép: A Brutus-kézirat egy oldala

Nem Brutus-autográf ugyan a szöveg, de Brutus írnokának levelekből ismert kézírását tartalmazza. Az egységes kézírás pedig lehetőséget adott arra, hogy az MI-alapú Transkribus-szolgáltatást² alkalmazva hidaljuk át a problémát. Petneházi Gábor kollégával ezért felvettük a kapcsolatot a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetében működő DigiPhil munkacsoporttal, és velük összefogva kezdtük meg a Brutus-szöveg feldolgozását.

Előbb az SZTE Klasszika-Filológia és Neolatin Tanszékén működő Neolatin Doktori programnak a latinban és a kéziratolvasásban egyaránt jártas doktorandusz hallgatóinak (Csapó Fanni, Kintli Dóra, Majoros Máté) bevonásával átírtunk 200 kéziratoldali szöveget, majd Bobák Barbara bevonásával megkezdődött a projekt Transkribus-szakasza.

Automatikus kézírásfelismerés

A DigiPhil munkacsoport 2018-ban kezdett el a Transkribus átíróprogrammal dolgozni, Arany János akadémiai hivatali levelein.³ Ez a projekt volt az első, amely magyar nyelvű kézirat átíratását célozta meg a Transkribus segítségével és ebből a munkából született meg az első – és azóta többször fejlesztett – magyar nyelvű modell is, amely a kézírásfelismeréshez elengedhetetlen.⁴ Az évek során a Transkribus jelentős népszerűsége tette szert – dacára annak, hogy ma már nem ingyenes a szolgáltatás⁵ –, amit mi sem mutat jobban, mint hogy egyre többféle nyelvre egyre több modell áll rendelkezésre. Népszerűsége többek között annak köszönhető, hogy a kézírásfelismerés teljes folyamatát magába foglalja, amit egy felhasználóbarát grafikus felülettel tesz elérhetővé. A Transkribus szolgáltatásban a kéziratok átíratása csupán néhány egyszerűbb lépésből áll. A beszkenelt kéziratokról készült képfájlok feltöltését követően a képfájlok szegmentálást végzünk, amely folyamatot layout analízis segít, illetve ennek során a program beazonosítja azokat a területeket, ahol szöveg található. A teljes szöveget és szövegegységeket úgynevezett dobozokba teszi és sorokra, valamint szavakra szegmentálja azt. A szegmentálást követően nyelvi modellt kell választanunk, amely az adott kézirat nyelvének és szerzőjének karakterkészletét sajátosságait foglalja magába, azok felismertetésére szolgál. A Transkribusban számos nyilvános modell⁷ közül választhatunk, így a latin nyelvű Brutus kéziratot a kutatócsoport először tesztelte a meglévő latin nyelvű modelleket. Mivel egyik sem hozott eredményt, a Brutus kéziratához saját modell létrehozására volt szükség, melynek tanítóanyagaként az említett kézzel átírt 200 oldal szolgált. Az átíró program dokumentációja⁸ a tanítóanyag méretét 5 000 és 15 000 szó között határozza meg, amely mennyiség az adott nyelv komplexitásától, illetve a kézirat sokszínűségétől függően változhat. A 200 oldal annak érdekében határoztuk meg, hogy hatékonyabb legyen a nyelvi modell, amely így egészen pontosan 54 344 szón tanulhatott. Bár a modell egyelőre csak a kutatócsoport számára érhető el, a közeljövőben bekerül a nyilvános modellek közé.

2 A Transkribus online szolgáltatás, melynek segítségével nyomtatott szövegek és kéziratok felismertetését és átíratását végezhetjük. Az alábbi weboldalról tölthető le ingyenesen, regisztrációt követően: <https://readcoop.eu/transkribus/download/>

3 Bobák, Barbara és Gábori Kovács, József (2019) Kézírásfelismerés Arany János levelein. In: Networkshop 2019. HUNGARNET Egyesület, Budapest, pp. 38 – 44.

4 A magyar nyelvre létrehozott modellek köre azóta bővült egy 19. századi, Kiss József levelezését feldolgozó projektnek köszönhetően: Mihály, Eszter (2021) A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben. In: Online térben az online térért, Budapest, pp. 91–109.

5 A munkafolyamatok zavartalan működése érdekében a BTK Irodalomtudományi Intézet krediteket vásárolt a Transkribus szolgáltatáshoz.

6 <https://readcoop.eu/transkribus/credits/>

7 <https://readcoop.eu/transkribus/public-models/>

8 <https://readcoop.eu/transkribus/resources/how-to-guides/>

Parameters:	
Max epochs	250
Early stopping	20
Epochs trained	84
Learning rate	0.0003
Batch size	24
Normalized height	64

3. kép: a Brutus nyelvi modell metaadatai

A Transkribus szolgáltatás számos előnye közé sorolható, hogy felhasználói felülete – amelynek megléte sem triviális az átíró szolgáltatások között⁹ – könnyen kezelhető, kifejezetten felhasználóbarát. Az átírás minden szakasza, maga a modellépítés is néhány kattintással elvégezhető, kedvezve mindazon felhasználóknak, akik kevésbé járatosak a különböző online szolgáltatások, programok világában. Az átíratás folyamatának következő lépése, hogy a létrejött Brutus modell segítségével lefuttatjuk a HTR¹⁰ modult, amely magát a kézírásfelismerést végzi a modellben rögzült nyelvi és karakterkészletű sajátosságok alapján. A Transkribusban lehetőség van arra, hogy az átírás egyes lépéseinek végeredményét kézzel korrigáljuk. A szöveg szegmentálását, illetve az átíratott szöveget is tudjuk javítani, kiegészíteni, törölni vagy pótolni.

A Brutus nyelvi modell és hatékonysága

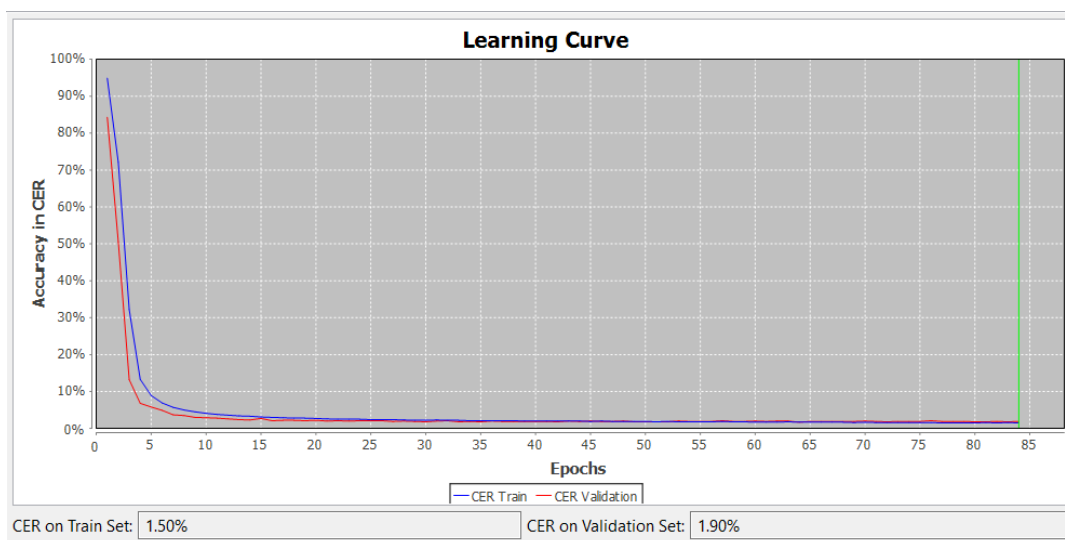
A Brutus-kézirat gépi átíratása rendkívül eredményesnek tekinthető. A Transkribus dokumentációja kitér arra, hogy a nyelvi modellek hatékonyságát mutató karakterhiba-arány¹¹ hogyan értelmezhető.¹² Az automatizált átírásnál a 10% alatti hibaarányt eredményező nyelvi modell magas hatékonyságúnak tekinthető. Esetünkben a Brutus modell karakterhiba-aránya 1,90%-ot ért el, azaz az átírt szöveg kevesebb, mint 2%-át rontotta csak el. Az alábbi koordináta-rendszer a Brutus nyelvi modell tanulási görbét (Learning Curve) mutatja:

⁹ A Google Tesseract például kizárólag parancssoros módon használható.

¹⁰ A HTR feloldása Handwritten text recognition, azaz kézírásfelismerés, képfájlok átalakítása feldolgozásra alkalmas szövegfájl formátummá.

¹¹ Character error rate (CER): a hibásan átírt karakterek aránya százalékos formában.

¹² <https://readcoop.eu/transkribus/howto/how-to-train-a-handwritten-text-recognition-model-in-transkribus/>



4. kép: a Brutus nyelvi modell tanulási görbéje

Az ábrán látható koordináta-rendszer y-tengelyén (Accuracy in CER) a karakterhiba-arány pontossága látható, százalékos formában. Ezen a tengelyen a görbe mindig 100%-tól indul, hiszen a nyelvi modell a tanulási folyamat elején nem rendelkezik információkkal az adott kézírásról. A tanulás során a modell fejlődik, melynek mértékét a görbe esése mutatja.

Az x-tengelyen (Epochs) azon szakaszok arányos leosztása látható, ahol a tanulási folyamat megáll, és a Transkribus értékeli a tanulás eredményességét. A nyelvi modell létrehozásakor megadható ezen szakaszok mennyisége – esetünkben 85 szakasz –, mellyel tehát azt határozhatjuk meg, hányszor kerüljön kiértékelésre a betanítási adat. Az ábra alján látható még a konkrét tanítóanyag¹³ (1,50%), illetve a tesztoldalakon¹⁴ (1,90%) ejtett karakterhibák százalékos aránya. A modellépítésnél a tanítóanyag 5–10%-át tesztoldalként különíti el a program, de kézzel is válogatható vagy bővíthető a tesztanyag.

A Transkribusban tehát a nyelvi modelleknek nemcsak a metaadatait láthatjuk, hanem a tanulási görbéjét is, amely megmutatja a modell tanulási folyamatát (kék vonal), illetve eredményességét (piros vonal). Az átíró program működésének minőségén túl a gyorsasága is számottevő. Bár a modellépítéshez szükséges tanítóanyag átírása elkerülhetetlen, ez általában a teljes anyag maximum 20%-át jelenti. A Brutus kézirat több mint 80%-nak átírása a program segítségével egy napba sem telt, az igen alacsony hibaaránynak köszönhetően pedig az átírat ellenőrzése sem túl időigényes. A megspórolt munkaidőn és energián túl mind a feldolgozandó kézirat mennyisége, mind az egységessége – amely a Brutus nyelvi modell magas hatékonyságát is eredményezte – alátámasztotta a Transkribus melletti döntést.

¹³ Train Set

¹⁴ Validation Set



Bibliográfia

Bobák, Barbara és Gábori Kovács, József (2019) Kézírásfelismerés Arany János levelein. In: Networkshop 2019. HUNGARNET Egyesület, Budapest, pp. 38 – 44.

<http://ocs.mtak.hu/index.php/nws/2019/paper/view/23/40>

Mihály, Eszter (2021) A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben. In: Online térben az online térért, Budapest, pp. 91–109.

<http://ocs.mtak.hu/index.php/nws/2021/paper/view/69/94>

Transkribus – Public models. Hozzáférés: 2022.06.16.

<https://readcoop.eu/transkribus/public-models/>

Transkribus – How to guides. Hozzáférés: 2022.06.16.

<https://readcoop.eu/transkribus/resources/how-to-guides/>

Transkribus – Model training. Hozzáférés: 2022.06.16.

<https://readcoop.eu/transkribus/howto/how-to-train-a-handwritten-text-recognition-model-in-transkribus/>