
MŰHELY

KALCSÓ GYULA

Kéziratos források tudományos elektronikus kiadása

Tanulmányom a kéziratos források tudományos elektronikus kiadásáról, elsősorban a kiadás megtervezésének, valamint szabványos kódolásának kérdéseiről szól, nem foglalkozik viszont a kiadás publikálásának a problematikájával. A tudományos elektronikus kiadás fogalmát Maróthy Szilvia munkái alapján használom, a szerző az angol *digital scholarly edition* kifejezés magyar megfelelőjeként javasolja több, a közelmúltban megjelent cikkében.¹ Az ilyen kiadást Patrick Sahle definíciója alapján kritikai reprezentációnak tartja, amely „létező forrásokat, történeti dokumentumokat dolgoz fel, forrást reprezentál és azt metaadatokkal látja el, valamint szövegkritikai munkát végez a forrás reprezentálásakor”.² Sahle nyomán emlékeztet azon elterjedt nézetre, hogy a digitális kiadás paradigmaváltást jelent:

Míg a hagyományos kiadásokban a szöveggondozó írta a szerkesztett szöveget, a digitális kiadásban az a források feldolgozási lépéseiből épül fel, a képi reprezentációtól az átíráson át a kritikai apparátusig. A források különféle reprezentációinak (és n. b. a szöveg különféle forrásainak) „elrejtésére” ezen túl nincs szükség, azok teljes terjedelmükben részei lehetnek a kiadásnak. E változásnak a következménye Sahle szerint a szerkesztés során a szöveg relativizálódása és multiplicitásának (összetettségenek és többszörösségenek) előtérbe helyeződése.³

Debreczeni Attila *Az elektronikus kiadás színe és visszája* című írásában felhívja a figyelmet arra, hogy „a digitális tudományos kiadásnak sem megvalósulásai nincsenek nagy számban, sem konszenzuális teóriája, minden számottevő eredmény ellenére sem”.⁴ Mindezek miatt indokolt lehet olyan teoretikus megfontolások

¹ MARÓTHY Szilvia, „Elektronikus szövegkiadások a könyvtárban”, *Tudományos és Műszaki Tájékoztatás* 64, 6. sz. (2017): 298–309; MARÓTHY Szilvia, „A tudományos elektronikus szövegkiadások feldolgozása”, in *Valóságos könyvtár – könyvtári valóság: Könyvtár- és információtudományi tanulmányok 2018*, szerk. KISZL Péter és CSÍK Tibor (Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar Könyvtár- és Információtudományi Intézet, 2018), 351–356; MARÓTHY Szilvia, „Tudományos szövegkiadások a hálózaton”, *Irodalomtörténeti Közlemények* 122, 5. sz. (2018): 617–633.

² MARÓTHY, „A tudományos elektronikus szövegkiadások feldolgozása”, 351–352.

³ MARÓTHY, „Elektronikus szövegkiadások a könyvtárban”, 299.

⁴ DEBRECZENI Attila, „Az elektronikus kiadás színe és visszája”, in *A humán tudományok és a gépi intelligencia*, szerk. TOLCSVAI NAGY GÁBOR, *A humán tudományok alapkérdései* 3, 48–61 (Budapest: Gondolat Kiadó, 2018), 49.

publikálása, amelyek középkori kéziratos források tudományos elektronikus kiadásának előkészítése, valamint a megvalósítás kezdeti, bátortalan lépései közben körvonalazódtak.⁵

A kéziratok digitális publikálása már hazánkban is elég gyakori, a középkori kódexektől a különféle levéltári forrásokon át az írói levelezésig mindenféle kéziratos műfaj képviselteti magát. Azt azonban aligha mondhatjuk, hogy valamiféle egységes módszertant tükrözne a forrásdigitalizálási és -közlési gyakorlat. Tószegi Zsuzsanna *A szövegdigitalizálás döntési folyamata* című cikkében a digitalizálásnak három szintjét különbözteti meg: a reprodukzív, a reprezentatív és az interpretatív szintet.⁶

A reprodukzív szint a forrásmű formai és tartalmi jegyeit egyaránt tükrözeti (az esetleges hibákkal, eltérésekkel együtt). A digitalizált változat az eredeti művel gyakorlatilag egyenértékű, azzal egyező hatást vált ki. Ebbe a csoportba elsősorban a faksimile állományok (képfájlok) tartoznak

– írja Tószegi.⁷ Az ilyen közlések a forrás különböző minőségű képét publikálják többnyire valamilyen szabványos formátumban, esetenként a képet tartalmazó PDF-ben. Ilyen például az Országos Széchényi Könyvtár nyelvemlékes honlapja (például a *Szabács viadala* PDF-változata)⁸ vagy a Magyar Elektronikus Könyvtár (MEK) digitalizált nyelvemlékei (például a *Régi magyar nyelvemlékek* 33. darabjaként megjelent Apor-kódex DVD-mellékletének böngészhető változata).⁹ Ezek a „digitális faksimilek” jobbára az autopszia helyettesítésére valók, nem teszik azonban lehetővé a keresést a szövegekben.

A reprezentatív szint a forrásmű tartalmát helyezi előtérbe, de alapvetően nem változtatja meg a szöveg lineáris olvasatát. Ezen a szinten az analóg szövegből digitalizált szöveget állítunk elő, amelynek információtartalma a számítógép nyújtotta szokásos eszközökkel könnyebben kereshető.¹⁰

⁵ A munkálatok a 2018 őszén, az Eszterházy Károly Egyetemen megalakult Digitális Bölcsészeti Kutatócsoportban (<https://uni-eszterhazy.hu/hu/bmk/bmk/a-bmk-hirei/c/digitalis-bolcseszeti-kutato-csoport>) indultak meg, kezdeti célként a csoport három ómagyar kéziratos forrás, a Tihanyi alapítólevél, a Halotti beszéd és könyörgés, valamint az Ómagyar Mária-siralom tudományos elektronikus kiadásának az elkészítését tűzte ki. A tanulmány a koncepció ez idáig körvonalazódott elemeit adja elő.

⁶ TÓSZEGI Zsuzsanna, „A szövegdigitalizálás döntési folyamata”, *Könyvtári Figyelő* 52, 2. sz. (2006): 245–260, 245.

⁷ Uo.

⁸ A *Szabács viadala* faksimiléje, hozzáférés: 2020.02.29, http://nyelvemlekek.oszk.hu/sites/nyelvemlekek.oszk.hu/files/mny2_b.pdf.

⁹ Az Apor-kódex faksimiléje, közléteszi HAADER Lea, KOCIS Réka, KOROMPAY Klára és SZENTGYÖRGYI Rudolf, hozzáférés: 2020.02.29, <http://mek.oszk.hu/16300/16329/html>.

¹⁰ TÓSZEGI, „A szövegdigitalizálás döntési folyamata”, 245.

A szövegdigitalizálás „hőskorából” származó szövegeken túl¹¹ – amelyek gyakran mutattak fel kezdetleges, mondhatni „barbár” megoldásokat¹² – mára a Tószegi cikkének megírása idején még nem létező kétrétegű PDF-ek is ilyen jellegű közléseknek számítanak, igaz, a reprodukív és a reprezentatív szint sajátos keverékének tekinthetők. Ezekben a forrás képe „mögött” megtalálható a – többnyire optikai karakterfelismeréssel (angol betűszavával OCR-rel) előállított – szöveg, amelyben ily módon kereshetünk is. Az OCR minősége – amely a történeti szövegek esetében általában nem jó –, illetőleg az utómunkára, főként a kézi korrektúrára fordított idő meghatározza a használhatóságot. Ilyen közlés például a Cornides-kódex MEK-beli változata, amely a *Codices Hungarici* sorozat nyomtatásban megjelent VI. kötetének OCR-ezett PDF-je.¹³

A digitalizálás legmagasabb szintjét a Tószegi által „interpretatív szintű” szövegek közléseknek nevezett kiadások jelentik:

Az interpretatív szinten az eredeti forrás tartalmához hozzáadódik a feldolgozást végző szakemberek tudása és tapasztalata, melynek eredményeként új minőség jön létre. Az eredeti művet kiegészítő elemek (amelyek lehetnek magyarázatok, mutatók, hipertext hivatkozások, vagy a szövegtől eltérő műfajú elemek: hang- és videofájlok stb.) megbontják az eredeti szöveg lineáris egységét.¹⁴

Ezekben a kiadásokban a szerkesztők a digitális szövegben különböző hozzáadott információkat kódolhatnak, amelyek az adatbázisba szervezéshez vagy az összetettebb keresések végrehajtásához szolgálhatnak alapul. Az annotáció tartalma általában attól függ, hogy milyen célból publikálják az elektronikus kiadást. A magyar gyakorlatban van példa papíralapú kritikai kiadásokhoz hasonló apparátus használatára, de nyelvészeti annotációt tartalmazó digitális korpuszokat is fejlesztettek már. Az előbbire példák a Sermones Compilati Kutatócsoport elektronikus szövegek közlései, többek között az Érdy-kódex digitális változata,¹⁵ amelyben hipertextes lábjegyzetek formájában szerepelnek a szövegkritikai annotációk, az utóbbira pedig az MTA Nyelvtudományi Intézetében fejlesztett ómagyar korpusz¹⁶ vagy a *Történeti magánéleti korpusz*.¹⁷

¹¹ Ilyenek pl. a MEK magyar irodalmi klasszikusokat népszerű, olcsó kiadásokból közreadó digitalizált szövegei.

¹² Vö. *Filológia és digitális barbárság: A BIÖP tanulmányi napja*. ELTE BTK, 2004. március 4, hozzáférés: 2019.12.19, <http://magyar-irodalom.elte.hu/biop/barbar/>.

¹³ A Cornides-kódex kétrétegű PDF-ben, hozzáférés: 2020.02.29, <http://mek.oszk.hu/08900/08981/pdf>.

¹⁴ TÓSZEGI, „A szövegdigitalizálás döntési folyamata”, 245.

¹⁵ Az Érdy-kódex elektronikus kritikai kiadása, hozzáférés: 2019.12.19, <http://sermones.elte.hu/erdy>.

¹⁶ A Magyar Generatív Történeti Szintaxis projekt ómagyar korpusza, hozzáférés: 2019.12.19, <http://omagyarkorpusz.nyttud.hu/>.

¹⁷ Az MTA NYTI Történeti magánéleti korpusza, hozzáférés: 2019.12.19, <http://tmk.nyttud.hu/>.

Az „interpretatív szint” definíciója azonban kiegészítésre szorul. Tószegi elfeledkezett arról, hogy interpretálni csak már valamilyen formában létrejött digitális objektumot lehet, vagyis az interpretatív szint szükségképpen összekapcsolódik a reprodukívval, a reprezentatívval vagy mindkettővel. Az annotált képek (reproduktív + interpretatív) és az annotált szövegek (reprezentatív + interpretatív) hazánkban sem példa nélküliek. Azonban a minden kétséget kizáróan legfejlettebb elektronikus kiadásnak számító, a képet és az annotált szöveget összekapcsoló, rugalmasan kereshető, az adatokkal műveletek végzését lehetővé tevő, szabványos és időtálló formátumban létrehozott közléseket hiába keresünk.¹⁸

Mi lehet az a szabványos és időtálló formátum, illetőleg módszer a digitális filológiában, amely a magas minőségű elektronikus tudományos kiadások esetében a bölcsek követ jelentheti? A kérdés napjainkig nyitott. Aktualitását mi sem jelzi jobban, mint a Text Encoding Initiative 2019 őszén, Grazban tartott nemzetközi konferenciájának a címe: *What is text, really? TEI and beyond (Mi a szöveg valójában? A TEI és azon túl)*. Ez a cím is megmutatja, hogy a digitális filológia, valamint az elektronikus kiadások új textológiaelméleti problémákat hoztak felszínre, amelyek – mint arra Debreczeni Attila is utal fenti cikkében – továbbra is megoldásra várnak. Hans Walter Gabler idézi saját fordításában:

Az előttünk álló feladat ezért a tudományos kiadást mint digitális tudományos kiadást megvalósítani. Ez megköveteli a médium sajátosságainak a maga teljességében való feltárását. Jelenleg a digitális közegben megvalósult kiadások még nagymértékben a nyomtatott kiadások imitációi maradnak. Amit azonban a mediális váltás megkövetel, az a teljes újrakonceptualizálás és következésképpen újramodellezés.¹⁹

Fontos kérdés továbbá az is, hogy a már kidolgozott módszerek mennyire nyugszanak elméletileg jól felépített alapokon.

Mindazonáltal a nemzetközi digitális filológiában több évtized alatt kikristályosodtak bizonyos konszenzusok, és ha még nem is álltak össze egy elméletileg is egységes, viszonylag stabil módszertanná, bizonyos elemeik megkérdőjelezhe-

¹⁸ A szerző egyetlen ilyen hazai közlést ismer, amely azonban az időtállóság követelményének sajnos nem tudott eleget tenni. A Debreceni Egyetemen készült el a Magyar Nyelv Éve alkalmából egy multimédiás összeállítás a Tihanyi alapítólevél és a Tihanyi összeírás nyelvemlékekről (<http://mny-tud.arts.klte.hu/kiallitas/intro.html>). Ez tartalmaz digitalizált betűhív szöveget, fordítást, a szöveggel összekapcsolt képet, valamint az egyes szövrányrészletek részletes leírását bibliográfiával, sőt térkép-melléleteket, valamint hangfájlokat is. A közlés koncepciója közel áll az ideális elektronikus tudományos faksimilékhez (lásd alább), szinte csak a rugalmas keresés lehetőségét hiányolhatjuk. Nagy hátránya viszont a formátuma: a Flash böngészőbővítmény, mely a megjelenítésért felel, napjainkra elavult, nem támogatott, a böngészőkben csak speciális beállításokkal jeleníthető meg, és várhatóan a nem túl távoli jövőben előfordulhat akár olyan felhasználói konfiguráció is, amelyen nem fog futni.

¹⁹ Hans Walter GABLER, „Foreword”, in *Digital Scholarly Editing: Theories and Practices*, eds. Matthew James DRISCOLL and Elena PIERAZZO, XIII–XIV (Cambridge: Open Book Publishers, 2016), XIV. Idézi: DEBRECZENI, „Az elektronikus kiadás színe és visszája”, 49.

tetlenek. Ilyen például az annotáció jelölőnyelvekkel történő megvalósítása. A jelölőnyelvek evolúciója mára egyértelműen kijelölte a leginkább platformfüggetlen és szabványos megoldást, amely a robusztus SGML-től az XML-ig vezetett:

a lassan 20 éve [az idézett cikk megszületésekor, K. Gy.] ISO-szabványként elfogadott SGML (Standard Generalized Markup Language) – szabványos jelölőnyelv dokumentumok belső szerkezetének leírására, beleértve az egyes elemeket jelölő címkék (tagok) definiálásának módját is – segítségével elvben bármilyen dokumentum leírható, függetlenül az azt tároló és megjelenítő számítógépes környezettől, nagyfokú bonyolultsága miatt azonban nem terjedt el sem Európában, sem világszerte a várt mértékben. A technológia viszont a maga nemében nagyszerű és egyedülálló, kár lett volna kiaknázatlanul hagyni. Éppen ezért a webes technológiák fejlesztésével foglalkozó World Wide Web Konzorcium 1998-ban egy új ajánlást – de facto szabványt – jelentett meg, az XML-t

– írta Bíró Szabolcs 2004-ben.²⁰ Az XML tehát az 1986-ban megjelent ISO 8879-es SGML-szabványnak a célszerű szűkítése. Más szabványokkal együtt (például az ISO/IEC 10646-tal, amely a Unicode konzorcium karakterkódolási ajánlásának a szabványa, egy érvényes XML-dokumentum csak ezzel lehet kódolva) képes biztosítani a teljes platformfüggetlenséget. Bíró cikkének megírásakor már készült a legnagyobb nemzetközi szövegekódolási konzorcium, a Text Encoding Initiative újabb ajánlása, a P5 (Proposal 5), amely már XML-alapon tett javaslatokat szövegekódolási eljárásokra. Jelenleg ennek 3.6.0 verziószámú változata van érvényben, amelyet 2019. július 16-án adtak ki. Részben igazat kell adnunk Palkó Gábornak, amikor azt írja:

A digitális filológiában évtizedek alatt egyeduralkodóvá vált az XML dokumentumleíró szabvány és az arra épülő Text Encoding Initiative (TEI) ajánlás, amelyet kifejezetten bölcsészettudományi használatra, textológus szakemberek közössége hozott létre [...]. A kilencvenes években még komoly viták folytak arról, hogy milyen hátrányai lehetnek a dokumentumok leírására kifejlesztett hierarchikus jelölőnyelveknek, de a TEI végül alternatíva nélkül maradt...²¹

A viták azonban ma is zajlanak. Fennállásának évtizedei alatt a TEI kétségkívül nagy népszerűsége tett szert, valamint a legtöbb textológiai projekt számára kézenfekvőbb megoldás a kitaposott ösvényeken haladni, mint az úttörés

²⁰ Bíró Szabolcs, „A szövegfeldolgozás modern eszközei – az SGML és XML nyelvek”, *Tudományos és Műszaki Tájékoztatás* 51, 10. sz. (2004): 453–459, 454.

²¹ PALKÓ GÁBOR, „Mit jelent a digitális filológia a szemantikus web korában?”, *Magyar Tudomány* 177, 11. sz. (2016): 1316–1322, 1320.

fáradtságos munkáját felvállalni. Ugyanakkor emlékeztetnünk kell Debreczeni Attila felvetésére is:

Abban a tekintetben vita folyik, hogy milyen mértékű lehet és legyen a TEI-XML standardizáció. Vannak vélemények, amelyek olyannyira túlzottnak tartják e keretrendszer rugalmasságát, hogy az már a standard működését veszélyezteti, én inkább azokkal tartanék, akik ezt a rugalmasságot a TEI-XML standard lényegének tekintik.²²

Valójában arról van itt szó, hogy mennyire kell a TEI-nek részletkérdések szabályozásába bonyolódnia. Véleményem szerint is közelebb áll az XML mint jelölőnyelv filozófiájához a rugalmasság, különben megvan a veszélye annak, hogy a TEI az SGML-hez hasonló szörnyeteggé fejlődik, ahol az apróságokat szabályozó részletes elemleírások áttekinthetlenné és nehézkesen alkalmazhatóvá tennék az ajánlásokat.

A kéziratok tudományos elektronikus kiadásában sem választhatunk tehát mást, mint valamely szabványos jelölőnyelvet, célszerűen a TEI-t. A kiadás elkészítésének a megtervezésekor azonban még egy kérdésre választ kell adnunk: mit fogunk pontosan kiadásnak tekinteni? Maróthy Szilvia fentebb már idézett cikkében az alábbi dilemmákat veti föl:

Mit nevezhetünk tehát kiadásnak az elektronikus közegben? A magában álló TEI XML fájl kiadás-e? A szakirodalom erősen megoszlik a tekintetben, a megjelenítés, felhasználás kritériuma-e a szövegkiadás kiadás voltának. Amennyiben elfogadjuk, hogy a magában álló XML fájl is lehet kiadás, mi húz határt kézirat és publikáció között? Ha pedig az XML fájl megjelenítése(i) által válik publikációvá, hol húzódnak a kiadás határai? E kérdések mind az eredeti problémára vezetnek vissza: a kiadást lezárt egészként képzeljük-e el (mint ami ki van adva). A problémára a digitális filológiai közösségben sem született még konszenzus.²³

A definíciós nehézségeket az okozza, hogy a digitális filológiában határozottan elkülönítik a procedurális és a logikai markupot egymástól. Azaz a megjelenítés csak lehetőség, nem szükségszerűség, vagyis egy XML-fájl publikálása nem a tartalmának és struktúrájának a megjelenítésével egyenlő. Persze könyvtártudományi (vagy ad absurdum tudománymetriai) szempontból fontos kérdés lehet, hogy mire rakhatunk azonosítószámot, azaz hogy mely informatikai entitás jelentheti az egyértelműen azonosítható digitális objektumot, mondjuk egy kézirat forrás tudományos elektronikus kiadását. Erre talán megadható a válasz:

²² DEBRECZENI, „Az elektronikus kiadás színe és visszája”, 53.

²³ MARÓTHY, „Elektronikus szövegkiadások a könyvtárban”, 299.

a verziószámmal rendelkező szoftver, amelyet fejlesztője bizonyos jól meghatározott kritériumok szerint – mindig ideiglenesen! – lezártnak (jellemzően például új főverzióknak) tekint. Ez a digitális filológiai objektumok esetében akár egy fájl is lehet, de általában egy könyvtár vagy több könyvtár strukturált halmaza. A legjobb példa erre maga a TEI Guidelines, melynek aktuális verziója²⁴ megtekinthető a TEI hivatalos kódrepositóriumában (githubos verziószámmal), és amelynek saját Zenodo DOI-száma van.²⁵ Tehát a Debreczeni Attila által is hangsúlyozott „mediális különeműség”²⁶ egyik fontos útjelzője lehet, ha az elektronikus kiadásokat a szoftvermérnökök által használt (több közreműködő esetén ún. osztott) verziókövető rendszerek²⁷ segítségével verziószámokkal látjuk el (amelyhez a repositóriumban publikálása pillanatában DOI-szám rendelhető). Debreczeni fel is hívja a figyelmet arra, hogy az elektronikus kiadás nem képernyőn megjelenő könyv:

nem indokolatlan azt állítani, hogy míg a nyomtatott kiadásnak olvasója, addig az elektronikus kiadásnak felhasználója van, s magát e kiadást is joggal nevezhetjük inkább *working place*-nek vagy *research site*-nak. Az így felfogott elektronikus kiadás nem könnyebben használható könyv immár, hanem olyan új kutatási lehetőségek tárháza, amelyek a lineárisan rögzített nyomtatott médiumban elképzelhetetlenek.²⁸

Visszakanyarodva Tószegi fogalmához, az „interpretatív szintű” digitalizáláshoz: a kiadás első lépése az annotálandó szöveg vagy szövegek (szövegváltozatok) előállítás, hiszen az „interpretációnak” (az annotációnak) erre kell épülnie. Ez máris számos elméleti és gyakorlati kérdést felvet. Debreczeni Attila írja: „A szöveg három reprezentációja, azaz a kézírathű átírás vagy kritikai szöveg (*critical text*), az emendált, megtisztított olvasószöveg (*reading text*) és a digitális faksimile napjainkban már az elektronikus kritikai kiadások meghatározó és megszokott elemei.”²⁹ Tehát az általa „megszokottnak” tekintett digitális fotómásolaton túl két alapvető szövegváltozat felmutatása jellemzi az „elektronikus kritikai kiadásokat”: a kézírathű átírás és az ún. olvasószöveg. A kézírathűség kritériuma régi vitatéma a textológiában, Orlovsky Géza például *A régi magyar*

²⁴ T. E. I. CONSORTIUM, *TEI P5: Guidelines for Electronic Text Encoding and Interchange* Version 3.6.0. (TEI Consortium, 2019).

²⁵ A TEI P5 Guidelines a Zenodón, hozzáférés: 2019.12.19, <https://doi.org/10.5281/zenodo.3413525>.

²⁶ DEBRECZENI, „Az elektronikus kiadás színe és visszája”, 49.

²⁷ A szemantikus verziószámozás egy lehetséges sémája, hozzáférés: 2019.12.19, <https://semver.org/lang/hu/spec/v2.0.0.html>.

²⁸ DEBRECZENI, „Az elektronikus kiadás színe és visszája”, 54.

²⁹ Uo., 52.

textológia helyzetéről című cikkében részletesen foglalkozik a pro és kontra érvekkel.³⁰ Következtetése jó irányelv az elektronikus kiadásokra nézve:

Látható, hogy mindkét szövegkiadási eljárás rendelkezik erényekkel és hiányosságokkal. A kettő közötti választást az anyag természetéből fakadó megfontolásokon túl elméleti meggyőződések is motiválhatják. A döntés sokszor feloldhatatlannak tűnő dilemmáit egy olyan szemlélet haladhatja meg, amely elválasztja egymástól a kritikai kiadás műfajában hagyományosan együtt jelen lévő két funkciót: az archiválást és a szövegprezentációt. Egy ideális szövegkiadás rendelkezésre bocsátaná egyrészt egy mű valamennyi szövegforrásának jó minőségű faksimiléjét, ezek szigorúan betűhű átíratát, és ezek mellett egy (vagy akár több!) szükség szerint rekonstruált, hangzashű átíratot.³¹

Matthew Driscoll a TEI *Electronic Textual Editing* című kézikönyvében kifejti, hogy a szigorú diplomatikai hűség és a teljesen modernizált átíratok között széles a spektrum,³² a két véglet közötti kevert változatok inkább csak az egyes szövegsajátságok kezelésében mutatnak különbségeket (feloldják-e a rövidítéseket, megtartják-e a nagybetűsítést stb.).

A kézírathűség vagy betűhűség kritériuma a képet is tartalmazó elektronikus kiadások esetében kissé okafogyottnak tűnik (lásd lentebb), illetőleg nincs egyértelmű konszenzus a tekintetben, hogy pontosan milyen transzkripciósi eljárásokkal állíthatók elő a fenti szövegvariánsok (különösen elektronikusan!), a betűhűségnek ugyanis számos szintje létezik. Szentgyörgyi Rudolf a *Tihanyi alapítólevélről* szóló monográfiájának első kötetében foglalkozik a problémával, és megkülönbözteti az ún. betű szerinti és a betűhű átírást (valamint ezektől a paleográfaiilag hűt):

Betűhű – és még nem paleográfaiilag hű – a forrásközlés, ha a forrás szövegét pontosan a forrásban található betűkkel, jelekkel, központozással, egybe- és különírással közöljük. Betű szerinti a közlés, ha a forrásban található betűket a nekik megfelelő mai betűkkel, a forrásnak megfelelő központozással, egybe- és különírással tesszük közzé, a rövidítéseket pedig feloldjuk, de zárójelezéssel vagy más tipográfiai eljárással megkülönböztetjük.³³

³⁰ ORLOVSKY Géza, „A régi magyar textológia helyzetéről”, *Irodalomtörténet* 85, 3. sz. (2004): 331–344, 333–337.

³¹ Uo., 337. Hozzátehetjük, hogy a „hangzashű átírat” egy újabb átírási mód, az alábbiakban még lesz róla szó.

³² Matthew DRISCOLL, „Levels of Transcription”, in *Electronic Textual Editing*, eds. Lou BURNARD, Katherine O'BRIEN O'KEEFE and John UNSWORTH, 254–261 (New York: Modern Language Association of America, 2006), <https://tei-c.org/Vault/ETE/Preview/driscoll.html>.

³³ SZENTGYÖRGYI Rudolf, *A tihanyi apátság alapítólevele: Az alapítólevél szövege, diplomatikai és nyelvi leírása* (Budapest: ELTE Eötvös Kiadó, 2014), 55–56.

Tehát a Debreczeni által olvasószövegnek nevezett változat betű szerinti átírás, ennek elveit a kiadás készítői lefektetik. A nagyobb gond a betűhűséggel (a kritikai szöveggel) van. Ez esetben a paleográfiai hűség és a diplomatikai hűség szintjei között az elektronikus kiadások esetében koncepciónk szerint egy célszerűen megválasztott kritérium alapján dönthetünk: addig megyünk el a betűhűségben, ameddig a szabványos eszközök azt lehetővé teszik, vagyis minden olyan grafematikai sajátosságot visszaadhatunk, amelynek kódolását a Unicode lehetővé teszi. Azért okafogyott valamelyest a Unicode által biztosítottól nagyobb mértékű betűhűséget várni, mert az átíratok egyébként összekapcsolhatók a ma már kellően nagy felbontású digitális képpel, így az eredeti paleográfiai sajátosságai közvetlenül azon tanulmányozhatók. Ennek ellenére vannak olyan projektek, amelyek a Unicode keretein túllépő betűhűséggel dolgoznak: nemzetközi konzorciális keretek között működik például egy olyan, középkori kéziratokkal foglalkozó közösség, a Medieval Unicode Font Initiative (MUFI),³⁴ amely rendszeresen ajánlásokat tesz a Unicode konzorciumnak újabb grafémák kódolására, valamint a Unicode által még nem kódolt grafémáknak a Unicode kódtáblán PUA-nak (Private Use Area) nevezett kódtartományban való kódolására. Ez utóbbihoz a konzorcium saját fontkészleteket is fejleszt. Ezzel azonban sajnos épp a Unicode eredeti célját (az univerzális egységességet és szabványosságot) szüntetik meg. Egyetlen cél indokolhatja a paleográfiailag hű szövegváltozat előállítását: ha kifejezetten ilyen jellegű kutatást szeretnénk támogatni vele, ekkor ugyanis az egyes paleográfiai sajátosságok is kereshetővé vál(hat)nak.

A – Debreczeni kifejezéseivel élve – „emendált és megtisztított” olvasószöveget sokszor normalizálnak is nevezik. A normalizált szövegváltozatok nagyobb mértékben tükröztetnek szerkesztői interpretációt. Az interpretatív mozzanat minden átírás esetében jelen van, és minden esetben szelektálást jelent (Elena Pierazzo megfogalmazásában: „a kiadott szöveg tehát egy az interpretáló-szerkesztő által létrehozott modell, amely a dokumentumban található bizonyos jellegzetességeket kombinál a kutatási célnak megfelelő rendezőelvek alapján”),³⁵ lényeges szempont azonban, hogy ennek eredményeképpen milyen mértékben torzul a szöveg. A különféle normalizálási műveletek elvégzésekor ennek nagyobb a veszélye. Egy ómagyar kori szöveg esetében például gyakran előfordul, hogy egy betű vagy betűkapcsolat olyan hangot jelöl, amely a mai magyarban nincs meg. Ekkor a normalizálás bármilyen formája jelentős torzítást eredményez: ha csak a betűt másoljuk, akkor elsikkad, hogy nem a mai hangértékében szerepel, ha pedig a hangértéket próbáljuk visszaadni, akkor az szükségképpen csak valamely mai betűvel vagy betűkapcsolattal lehetséges egy normalizált szövegben (máskülönben az nem is normalizált), ekkor viszont a betűk mai hangértékével

³⁴ A Medieval Unicode Font Initiative honlapja, hozzáférés: 2019.12.19, <https://folk.uib.no/hnooh/mufi/>.

³⁵ Elena PIERAZZO, *Digital Scholarly Editing: Theories, Models and Methods*, Digital Research in the Arts and Humanities (Farnham: Ashgate, 2015), 44–45. Saját fordítás. K. Gy.

való téves azonosítás torzítja az átírást. Nem teljesen világos, hogy Orlovsky Géza fentebb idézett szövegében mit értett „hangzashű átíraton”, de a nyelvészeti forrásközlések olvasati része, amely valamely fonetikus átírással történik, ezt a problémát hivatott orvosolni – igaz, hogy „népszerű” olvasószövegnek aligha alkalmas. Kérdés, hogy egy tudományos kiadásnak van-e szüksége ez utóbbira. Gyakori még a tudományos szövegkiadások – különösen a nyelvészeti célú közlések – esetében egy szemantikai interpretációt tartalmazó szövegváltozat (értelmezés) előállítását, amely tulajdonképpen mai nyelvre írja át a szöveget (minél régebbi a forrás, annál gyakrabban lehet szükség kihalt szavak, grammatikai jelenségek, maitól eltérő, megváltozott jelentések stb. magyarázatára).

Mindent összevetve valójában a szövegváltozatok materiális és immateriális összetevőinek a kontinuumáról van itt szó:³⁶ az eredeti kéziratot különféle szinteken hűen tükröző átírásoktól (paleográfaiilag hű, betűhű, betű szerinti) a különböző, az interpretációnak nagyobb teret engedő átírásokig (olvasószöveg, normalizált szöveg, olvasat, értelmezés stb.). Mindezek a szöveg különböző dimenzióit ragadják meg (a paleográfai, ikonológiai, kodikológiai a nyelvészeti, szemantikai és irodalmi dimenzióig). A középkori kéziratok tudományos elektronikus kiadásakor célszerűnek látszik a kontinuum minél teljesebb lefedése: fakszimilét is tartalmazó kiadás esetén elegendő lehet egy Unicode-kompatibilis betűhű szöveg az eredeti rövidítésekkel, írásjelekkel stb., egy kidolgozott szabályrendszer alapján emendált betű szerinti szöveg, amelyben a rövidítések feloldva szerepelnek, egy fonetikus átírat, amely akár tükrözhet variációkat is, valamint egy értelmező átírat (ugyancsak célszerűen akár variációkkal). Nagyon fontos szempont, hogy a felhasználó milyen műveleteket végezhet az előállított szövegekkel. A publikált tudományos elektronikus kiadások egyik fontos funkciója például a kereshetőség, amelyet érdemes szem előtt tartani a szövegváltozatok megtervezésekor is. Mivel a speciális Unicode-karakterek gépelése kicsit körülményes lehet, valamint az ortográfiai sokszínűség miatt egy-egy lemma példáinak az előállítása nem szerencsés betűhív vagy betű szerinti szövegben kereséssel (bár az is megoldható). Ezért lehet nagyon hasznos az értelmező átírat, mert abban a szavak mai alakjára kereshetünk.

A következő probléma az előállított szövegváltozatok összekapcsolása. Erre azért van szükség, mert a variációk egyik nagy haszna épp az lehet, hogy a felhasználó által megadott lekérdezések együttesen jeleníthetők meg az egyes átíratokban (illetve akár egy jó minőségű fakszimilén).³⁷ Az együttes lekérdezhetőséghez meg kell határoznunk azt a nyelvi szintet, amelynek a kereshetőségét biztosítani kíván-

³⁶ Uo., 52.

³⁷ Az együttes lekérdezhetőség a fontosabb szempont: nagyon óvatossá kell lennünk, mert alapvetően az elektronikus tudományos kiadásokban, hogy a megjelenítés nem rendelhető alá a strukturális kódolásnak. (Maga az XML épp azért jött létre, mert a weboldalak megjelenítését kódoló HTML jelölőnyelv nem tudta szétválasztani a megjelenítésre és a logikai struktúrára vonatkozó információkat.) A „mediális különműködés”, a hagyományos papíralapú kritikai kiadásoktól való elkülönülés talán egyik leglényegesebb mozzanata ez: a megjelenítő eszköz másodrangúvá válik. Ugyanakkor a majdani felhasználás praktikuma sem kell, hogy elsikkadjon, csak nem válhat a strukturális kódolást vezérlő elvvé.

juk. A leggyakoribb megoldás erre a szószintű összekapcsolás, amely alapján a többszavas kifejezésekre történő keresés is biztosítható. Természetesen elő fog fordulni, hogy az egyes szövegváltozatokban a szóalak spaciúm(ok)kal tagolt kifejezés lesz: ha a scriptor (egységes ortográfia hiányában) nem írt egybe összetett szavakat, ha a toldaléket – vagy éppen toldalékká grammatikalizálódó morfémát – nem írta egybe a szótóvel, ha az értelmezésben csak több szóval lehet kifejezni a forrás szóalakját stb., az a lényeges, hogy ezek a kód szintjén megfelelően legyenek összekapcsolva egymással. Valójában a tudományos elektronikus kiadások nemzetközi gyakorlatában az ilyen szintű összekapcsolásra egyelőre kevés példát találunk.

Az alábbi három egyszerűsített kódrészlet³⁸ a *Halotti beszéd* egy mondatának háromféle szövegváltozatát mutatja be.³⁹

Betűhű:

```
<w xml:id="1">Num</w>
<w xml:id="2">heon</w>
<w xml:id="3">muga nec</w>.
<w xml:id="4">ge</w>
<w xml:id="5">mend</w>
<w xml:id="6">vv</w>
<w xml:id="7">foianec</w>
<w xml:id="8">halalut</w>
<w xml:id="9">evéc</w>.
```

Betű szerinti:

```
<w xml:id="1">Num</w>
<w xml:id="2">heon</w>
<w xml:id="3">muga nec</w>.
<w xml:id="4">ge</w>
<w xml:id="5">mend</w>
<w xml:id="6">vv</w>
<w xml:id="7">foianec</w>
<w xml:id="8">halalut</w>
<w xml:id="9">evéc</w>.
```

³⁸ A <w> címke a szóalakokat jelzi (word, lásd: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-w.html>). Ezek sorszámot kaptak (xml:id; az itt közöltek nem a valódi kiadás sorszámjai). A sorszám kapcsolja össze a három változat megfelelő szóalakjait. Az írásjelek kódolását a jobb áttekinthetőség kedvéért kivettem.

³⁹ A fonetikus átírás bemutatásától ezúttal az egyszerűség kedvéért eltekintettem. A betűhű és a betű szerinti átírat ez esetben csak a pont nélküli *i* jelölésében tér el egymástól. Vö. Kis Tamás, „A *Halotti beszéd* paleográfája és betűhű átíratai”, *Magyar Nyelvtudományok* 56 (2018): 37–56, 40.

Értelmező:

```

<w xml:id="1">Nem</w>
<w xml:id="2">csupán</w>
<w xml:id="3">magának</w>,
<w xml:id="4">hanem</w>
<w xml:id="6">
  <app>
    <choice>
      <orig>ő</orig>
      <reg></reg>
    </choice>
  </app>
</w>
<w xml:id="5">
  <app>
    <rdg>egész</rdg>
    <rdg>minden</rdg>
  </app>
</w>
<w xml:id="7">
  <app>
    <rdg>fajtájának</rdg>
    <rdg>ivadékának</rdg>
  </app>
</w>
<w xml:id="8">halált</w>
<w xml:id="9">evett</w>.

```

Amennyiben szeretnénk biztosítani a 6-os sorszámmal rendelkező *ő* kereshetőségét, akkor meghagyhatjuk a fenti kódolási megoldással, vagy ha a betűhú/betű szerinti átiratunk tartalmaz nyelvészeti annotációt (például a lemmák megadását), akkor az értelmező kódrészletből ki is hagyhatjuk. Ha valaki például az ómagyar E/3. személyes névmásokra akar keresni egy leendő korpuszban, amelyben a *HB*. is szerepelni fog, akkor a névmás mai alakját megadva mindegyik átiratban rátalálhat (vagy az értelmező átiratban szereplő *ő* lesz a találat, vagy a nyelvészeti annotációban megadott lemmája), és akár együttesen (legjobb esetben a képen is) megjelenítheti. A kiadás készítőjének (kódolójának) arra kell ügyelnie, hogy valamilyen módon biztosítsa az eredeti szöveg nyelvi megoldásának a kereshetőségét. Tehát ha nincs nyelvészeti annotáció, azaz nincs megadva a betűhú vagy a betű szerinti átiratban, hogy az *vv* az *ő* névmás alakja, akkor az értelmező részletből sem hagyható el, mert különben nem lenne megtalálható. Látható, hogy variá-

ciók is megadhatók, a fenti példában az értelmező átiratban szerepelnek olyan változatok, amelyek teljesen egyenértékűek egymással: így lehet megadni, hogy az ő egyébként a mai magyarban az ómagyar korival ellentétben nem kötelező (sőt: névelő nélkül agrammatikus, erre utal a kódban a <reg> címke, amely a *regularized* angol szóból ered).

Az elkészítendő kiadás alapvető strukturális kérdése, hogy a szövegváltozatok hogyan jelenjenek meg a kód szintjén. Egy fájlban szerepeljenek? Mindegyik átirat külön fájlba alkosson? Mivel ezek a szövegváltozatok nem a kiadott szöveg genetikai példányai, ezért a TEI ún. párhuzamos szegmentációs eljárása⁴⁰ itt nem megfelelő megoldás. A különböző átiratok egy fájlban kódolására vannak példák a nemzetközi gyakorlatban, de legalább olyan gyakori az is, hogy egy-egy tudományos elektronikus kiadás könyvtárstruktúrába szervezett fájlok együtteseként jelenik meg. Élünk a gyanúperrel, hogy az egy fájlhoz ragaszkodás a hagyományos papíralapú kiadások hatása. Hogy milyen hátrányokkal járhat, azt jól mutatja a MENOTA⁴¹ esete. A Medieval Nordic Text Archive egyike a legszakoszerűbb nemzetközi projekteknek, bizonyos megoldásaik azonban mégis kérdéseket vetnek fel. Ahogyan kézikönyvük⁴² 4.6. fejezete írja, többszintű átirással készítének középkori északi germán kéziratokból tudományos elektronikus kiadásokat. Az általuk alkalmazott eljárást Lavrentyev és munkatársai hibrid transzkripciónak nevezik.⁴³ Ennek az a lényege, hogy az átirási rétegeket egy fájlban tartják, a „hibrid” jelző arra utal, hogy egymás mellett szerepelnek a különböző típusú átiratok. Ez azonban általában nem lehetséges szabványosan a TEI keretei között. Az egyik lehetséges megoldása a problémának a TEI saját névtérben alkalmazott kiegészítése, ez viszont szembemegy a konzorcium létrehozásának, működésének egyik legfontosabb céljával: a nemzetközi egységesség követelményével. A MENOTA például háromféle szövegváltozatot kódol, a <me: facs> címkével egy betűhű átiratot (ők „facsimilé”-nek nevezik), amelyben a rövidítések nincsenek feloldva, a <me: dipl> címkével egy betű szerintit (diplomatic, azaz „diplomatikai”), amelyben feloldják a rövidítéseket, továbbá bizonyos egyéb emendálásokat alkalmaznak, valamint a <me: norm>, ami normalizáltat jelent. Így a projekt egyébként igen magas filológiai színvonalát sajnos beárnyékolja, hogy nem tudnak a TEI P5 keretei között maradni.

⁴⁰ A *parallel segmentation method* (párhuzamos szegmentálási eljárás) a TEI Guidelinesban, hozzáférés: 2019.12.19, <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPPS>.

⁴¹ A Medieval Nordic Text Archive angol nyelvű honlapja, hozzáférés: 2019.12.19, https://www.menota.org/EN_forside.xhtml.

⁴² Odd Einar HAUGEN, ed., *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources*, Version 3.0. (Bergen: Medieval Nordic Text Archive, 2019), <http://www.menota.org/handbook.xml>.

⁴³ Alexei LAVRENTIEV, Yann LEYDIER and Dominique STUTZMANN, „Specifying a TEI-XML Based Format for Aligning Text to Image at Character Level”, 2015, hozzáférés: 2020.02.29, <https://www.balisage.net/Proceedings/vol16/print/Lavrentiev01/BalisageVol16-Lavrentiev01.html>.

Bár szorosabban a megjelenítés kérdésköréhez tartozó téma (amely jelen tanulmányunknak nem tárgya), mégis szükséges szólnunk a képeknek a tudományos elektronikus kiadásokban betöltött szerepéről. A jó minőségű fényképpel való szinoptikus olvasás lehetősége nyilvánvalóan emeli a kiadás színvonalát. A legjobb megoldás viszont az, ha a kép és a szöveg összekapcsolása által megteremtődik annak a lehetősége, hogy a kereséseink és egyéb műveleteink (például a szövegben vagy a képen valamely metaadat eléréséhez kattintás) eredményeit egyszerre jeleníthessük meg a képen és tetszőleges szövegvariánson (akár valamennyin). Gyakoribb, hogy a forrásdokumentum valamely vizuálisan jól megragadható egysége, jellemzően például a kézirat sora van összekapcsolva az átiratokkal, azonban a szóalapú keresés lehetősége miatt jobb, ha szószinten valósul meg ugyanaz. Annak ellenére, hogy a szabványos technikai feltételei ennek adottak, még ma sem gyakoriak az ilyesféle kiadások.⁴⁴ A Greta Franzini által 2012 óta szerkesztett *Catalogue of Digital Editions*⁴⁵ keresőjében az alábbi beállításokkal szűrtem meg a több mint 300 kiadást: legyen tudományos (*scholarly*), legyen kiadás, például nem lehet adatbázis (*edition*), legyen TEI-XML, legyenek képek mellékelve, és legyenek a képek összekapcsolva a szöveg(ek)kel (*text-image linking*). A találati lista mindössze 9 tételből állt (ha a szöveg és a kép összekapcsolásának a követelményétől eltekintünk, akkor már 83 tételből áll a lista, vegyesen szerepelnek benne kéziratok és nyomtatványok kiadásai). A találatok között sajnos magyar kiadást nem találunk.

A kliensoldalon (például böngészőben) megjelenítés megvalósításával itt most nem foglalkozom, ki kell térnem viszont arra, hogy a szövegek hogyan kapcsolódnak össze a forrás képével. A képen megjelenő bármilyen vizuális egység kódolható. A TEI P5 módszerének lényege az, hogy a forrás digitális képét zónákra osztja, amelyeknek a koordinátáit összekapcsolja a szöveg megfelelő részeivel, illetőleg azokhoz metaadatokat rendelhet. A zónák lehetnek négyszögletesek, de bármilyen formájú poligon kódolása is megoldható a sarokpontjainak képkordinátáit rögzítve. A zónák manuális létrehozásához ma már rendelkezésre állnak online is használható eszközök.⁴⁶ Az automatizálás kérdését eddig szándékosan kikerültem (bár vannak ígéretes eredmények, meggyőződésem szerint a megfelelő minőségű digitális filológiai munkához, elsősorban a szövegváltozatok előállításához még sokáig gazdaságosabb módszer lesz szakembert alkalmazni), a nagyobb terjedelmű szövegek képfájljainak a zónákra osztása azonban bizonyos keretek között eredményesen automatizálható. Lavrentyev és munkatársai

⁴⁴ Szinoptikus olvasást lehetővé tevő kéziratkiadás már hazánkban is van: a Digiphil Móricz-levelei például ilyenek, a képek azonban nincsenek összekapcsolva a szöveggel. Hozzáférés: 2019.12.19, <http://digiphil.hu/context:moriczlevelezes>.

⁴⁵ Idézi: MARÓTHY, „Tudományos szövegkiadások a hálózaton”, 617. A katalógus a világhálón, hozzáférés: 2019.12.19, <https://dig-ed-cat.acdh.oeaw.ac.at/>.

⁴⁶ Ilyen pl. Marjorie Burghart és Chris Sparks TEI-zonere, hozzáférés: 2019.12.19, <http://teicat.huma-num.fr/zoner.php>, ahol tetszőleges fényképet tölthetünk föl, és azt szabványos TEI-annotációval zónákra oszthatjuk.

fentebb már idézett előadásában például egy olyan szoftvert mutatnak be, amely képes automatikus karakterszintű szegmentálásra. A sorokra, illetve szóalakokra bontás pedig egyre több képfeldolgozó szoftver beépített funkciója.⁴⁷ Léteznek olyan projektek is, amelyek képesek a kimeneteiket TEI-XML-ben előállítani.⁴⁸ A szóalakokra szegmentált kép például összekapcsolható a fentebb leírt módon, szószinten összekapcsolt szövegváltozatokkal. Így akár a szóalapú keresés eredménye is megtekinthető egyszerre bármely szövegváltozatban és a forrás képén.

A tanulmányban leírt módszer természetesen csak egyike a kéziratok tudományos elektronikus kiadásának. A legfontosabb alapelvei az alábbiakban foglalhatók össze. Az elektronikus kiadás minden eleme szabványos kell legyen. A szövegek szigorúan Unicode-kompatibilis karakterláncok lesznek. Négyféle szövegváltozatot célszerű előállítani: egy betűhű, egy betű szerinti, egy fonetikus és egy értelmező szöveget. A jó minőségű faksimile biztosíthatja a paleográfiai jellegzetességek tanulmányozását – amennyiben a kép össze van kapcsolva az átiratokkal –, így a betűhű változatban megengedhető bizonyos grafematikai egyszerűsítés, de mindent kódolni kell, amit a Unicode lehetővé tesz. Ebben a rövidítések, az összevonások, a javítások, a betoldások, az egybe- és különírás sajátosságai stb. egyaránt tükröződnek. A betű szerinti változat feloldhatja a rövidítéseket, összevonásokat, illetve alkalmazhat egyéb célszerű emendálást. A két, jelentősebb mértékben interpretatív szövegváltozat, az olvasat és az értelmezés a könnyebb olvasást hivatott elősegíteni, valamint – különösen az utóbbi – megkönnyíti a szövegekben keresést. A szövegeket a TEI-XML jelölőnyelvvvel kell annotálni. Az annotáció összekapcsolja a változatokat a szóalakok szintjén, valamint azokat a képpel. Az egyes szövegváltozatok külön fájlban szerepelnek. A TEI-XML-fájlok fejlesztését szigorúan dokumentálni kell valamely osztott verziókövető rendszer (például Git) használatával. A fejlesztés egyes fázisait meg kell tervezni, és meg kell határozni, hogy mit tekintünk alapkiadásnak (1.0 verzióknak). Ehhez DOI-azonosítót (Digital Object Identifier) kell rendelni. Nem alapkövetelmény, viszont kívánatos, hogy a munkálatokhoz nyílt forráskódú és nyílt hozzáférésű szoftvereket használjunk. A tudományos munka átláthatósága, verifikálhatósága érdekében az egész folyamat, valamint a létrehozott produktumok (beleértve a fájlok forráskódját is) ideális esetben nyíltan hozzáférhetők.

⁴⁷ Itt pl. a szerzők összegzik az általuk ismert képfeldolgozó szoftvereket azok funkcióit is felsorolva (Table 1): Nicholas JOURNET, Muriel VISANI, Boris MANSENCAL, Kieu VAN-CUONG and Antoine BILLY, „DocCreator: A New Software for Creating Synthetic Ground-Truthed Document Images”, *Journal of imaging* 3, no. 4. (2017): 62.

⁴⁸ Rory McNICHOLL and Timothy MILES-BOARD, „transcriptorium: Computer-Aided, Crowd-Sourced Transcription of Handwritten Text (for Repositories)”, in *10th International Conference on Open Repositories (OR2015)*, 2015; Transkribus. Tanítható HTR (Handwritten Text Recognition, azaz kéziratfelismerő) szoftver. Alkalmos például a kéziratok szöveg sorainak és szóalakjainak az elkülönítésére. Hozzáférés: 2019.12.19, <https://transkribus.eu/Transkribus/>.