

Pázmány Péter Katolikus Egyetem
Bölcsészet- és Társadalomtudományi Kar

**A digitális filológia Magyarországon:
elvek és gyakorlatok**

Doktori (PhD) értekezés
Fellegi Zsófia

Irodalomtudományi Doktori Iskola

Témavezető: Dr. Almási Zsolt tanszékvezető egyetemi docens

A doktori iskola vezetője: Dr. Dobos István tanszékvezető egyetemi tanár

Budapest

2024

Tartalomjegyzék

Bevezető	3
Átmenet az analógból a digitális filológiába	27
A genetikus kritika megvalósulása papíron és a digitális térben: Madách Imre: Az ember tragédiája – Szinoptikus kritikai kiadás	27
Nyomtatott kritikai kiadások informatizálása	45
Arany János kéziratának és a szerző életművének kritikai kiadásainak közzététele	45
Digitálisan született tudományos szövegkiadások	64
Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása	64
A born digital kritikai kiadások térnyerése	73
Digitális filológia és a szemantikus web	81
Forráskiadások a szemantikus weben	91
Digitális kritikai kiadások a szemantikus weben. A digitális filológia új infrastruktúrái	101
A digitális kiadásokon túl: Jelölőnyelvi átiratok elemzése	113
Digitális kiadások a szemantikus weben. Metaadat-aggregáció	113
Jelölőnyelvi átiratok elemzése gépi olvasás segítségével	132
A digitális filológia oktatása	148
Összefoglalás	166
Summary	168
Bibliográfia	170

Bevezető

A digitális médiumváltás hatással van a jelen kor minden kulturális gyakorlatára, a tudományra, azon belül a bölcsészetre, az irodalomtudományra, a filológiára is. A nemzeti kulturális örökség filológiai szempontból legmagasabb színvonalú szövegkiadásai a kritikai kiadások, a filológia, azáltal, hogy ellenőrzött, „stabil” szövegkorpuszokat állított elő a bölcsészeti kutatások számára, mindig is formálta azt, hogy mi válik a kutatás tárgyává. Az elmúlt évtizedekben a számítástechnika és az internet fejlődése magával hozta a tudományos módszertanok változását is. Az irodalomtudomány és a filológia területén is felmerült az igény olyan kutatási infrastruktúra kifejlesztésére, amely lehetővé teszi a fizikai határokon átívelő kutatási hálózatok kiépítését.

Jelen disszertáció nem vállalkozik a magyar digitális filológia történetének megírására, ez a vállalkozás önmagában egy nagymonográfia témája lehetne. A kutatás során az is egyértelművé vált, hogy a magyarországi digitális filológiai műhelyek és projektek genezise csupán töredékesen dokumentált, számos kulcsfontosságú információ nem áll rendelkezésre, az ehhez szükséges feltáró munkálatok meghaladják a disszertáció kereteit. A disszertáció szintén nem nyújtja a digitális filológia rendszeres elméleti áttekintését. Ez a feladat talán még az előzőnél is nagyobb, már csak a technológia döbbenetesen gyors változása miatt is lehetetlen lenne monográfiászerűen egy ilyen problémát tárgyalni. A disszertáció ugyanakkor érinti a magyar digitális filológia történetének egyes fejezeteit és elméleti szempontokat is tárgyal, de csak egy konkrét digitális filológiai műhely, a DigiPhil (A magyar irodalomtudomány filológiai portálja) gyakorlatának kontextusában.¹ A disszertáció elsődleges célja az, hogy segítsen a magyarországi filológia területén felmerülő gyakorlati és elméleti problémák áttekintésében, a jó gyakorlatok kiemelésével. Ennek háttérében az a törekvés áll, hogy a DigiPhil körül az elmúlt egy évtizedben összegyűlt tapasztalatokat más filológiai műhelyek és, különösen, a jövő digitális filológusai számára közvetítse. A disszertációban tehát arra vállalkozom, hogy a DigiPhil genezisének és jelen gyakorlatának problémaorientált áttekintését nyújtsam. Ehhez, egyrészt, elengedhetetlen a DigiPhil-t megelőző, illetve a kialakulása körüli magyarországi és nemzetközi kontextus vázlatos bemutatása, valamint azoknak az elméleti kérdéseknek, vitáknak, kurrens diskurzusoknak a megidézése, amelyek közvetlenül hatással voltak és vannak a DigiPhil kutatási módszertanának mikro- és makrodöntéseire. A disszertáció egyaránt tárgyalja a DigiPhil 1.0 és 2.0 gyakorlatát és azokat az elméleti kontextusokat, amelyek szerepet

¹ A DigiPhil szolgáltatás elérhetősége: <https://digiphil.hu/>.

játszottak ezek kiépítésében. Mielőtt azonban bemutatnám a disszertáció felépítését, elengedhetetlen azoknak az alapfogalmaknak a bevezetése, amelyekre az érvelés támaszkodik, kezdve magával a *digitális filológia* fogalmával.

Digitális filológia és a digitális szövegkiadás fogalma

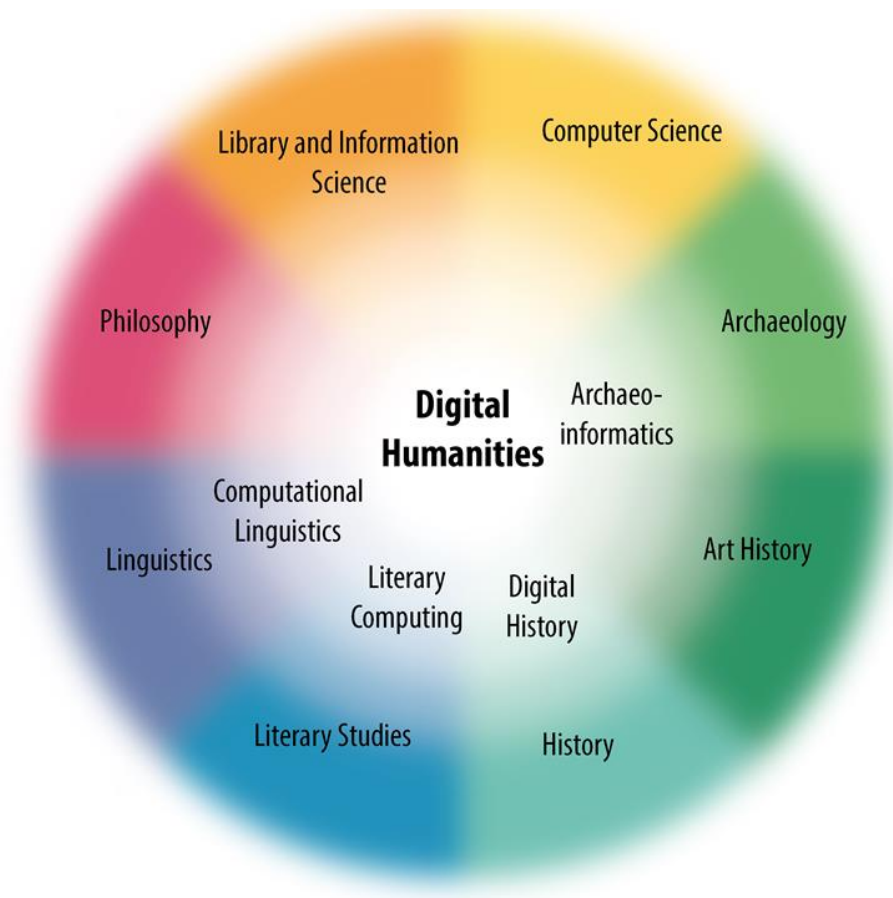
A digitális filológia egy szakdiszciplína, amely a digitális bölcsészet alá sorolható. A digitális bölcsészet viszonylag fiatal tudományág, elnevezése – ahogyan Matthew G. Kirschenbaum írja² – eleinte csupán néhány kutató közös megegyezésén alapult (a *digital humanities* nyert az *ehumanities* és a *humanities computing* fogalmakkal szemben), majd 2005-re terjedt el, amikor megjelent a *Companion to Digital Humanities* című tanulmánykötet,³ valamint ekkor véglegesítették a legnagyobb vált digitális bölcsészeti ernyőszervezet, az Alliance of Digital Humanities nevét. A 2008 óta létező THATCamp (The Humanities and Technology Camp) 2010-ben adta ki kiáltványát (*Manifesto for the Digital Humanities*), amelyben megfogalmazták a csoport irányelveit, céljait. A kiáltvány a digitális bölcsészetet „transzdiszciplínának” nevezi, amely a társadalom- és bölcsészettudományokon alapul és hasznosítja a digitális technológiákat és szemléletmódokat.⁴ Patrick Sahle egyik előadásában⁵ a digitális bölcsészetet úgy definiálta, mint hidat a hagyományos tudomány és a modern, számítógépes technológiák között. Ennek mentén alkotta meg a digitális bölcsészet három szférájának modelljét.

² Matthew KIRSCHENBAUM, „What Is Digital Humanities and What’s It Doing in English Departments?”, in *Debates in the Digital Humanities*, szerk. Matthew K. GOLD, 3–11 (Minneapolis: University of Minnesota Press, 2012).

³ Susan SCHREIBMAN, Ray SIEMENS és John UNSWORTH, szerk., *A Companion to Digital Humanities* (Malden Blackwell Publishing, 2004).

⁴ A kiáltvány 2010-ben jelent meg a szerző személyes weboldalán: Marin DACOS, „Manifesto for the Digital Humanities”, *Marin Dacos* (blog), 2010. március 31., <https://marin.dacos.org/manifesto-for-the-digital-humanities/>. A teljes szöveg a Hypothesis oldalán olvasható: Marin DACOS, „Manifesto for the Digital Humanities”, *Hypotheses* (blog), 2011. március 16., <https://tcp.hypotheses.org/411>.

⁵ Patrick SAHLE, „On the Ongoing Formation of Digital Humanities - A 2018 Snapshot” (DH_BUDAPEST_2018, Eötvös Loránd Tudományegyetem, 2018. május 28.).



[1. ábra] Patrick Sahle, A digitális bölcsészet szférái.⁶

Sahle modellje szerint a digitális bölcsészet felfogható a hagyományos diszciplínákon belüli speciális részterületek összességéként. A középső szféra azt mutatja meg, hogy a digitális bölcsészet új kutatási témák, önálló szakdiszciplínák csoportja, míg a modell központjában a digitális bölcsészet már mint önálló tudományág áll. Sahle az előadásában felhívta a figyelmet a modell szférái közötti feszültségekre: a külső szféra kutatói szerint nincs szükség a digitális bölcsészetre mint önálló tudományágra, míg a belső szférában elhelyezkedő digitális bölcsészek szerint az, ami a külső szférában zajlik, nem digitális bölcsészet. A középső szféra kutatói – Sahle szerint – néha a digitális bölcsészet, néha más tudományágak részeként tekintenek magukra. Ez a modell – vagy ahogyan Sahle nevezte: térkép – segíthet eligazodni az egyes tudományterületek közötti „határvitákban”. Arra, hogy a digitális bölcsészet önálló

⁶ Patrick SAHLE, „Digital Humanities? Gibt’s doch gar nicht!”, *Zeitschrift für digitale Geisteswissenschaften (ZfdG)*, 2015, https://zfdg.de/sb001_004.

tudományág, Sahle szerint – azon túl, hogy számos ismert tudományos folyóiratban jelennek meg publikációk és rengeteg konferenciát rendeznek a témában – a legerősebb érv az oktatás. A digitális filológia mint szakdiszciplína szintén elhelyezhető Sahle térképén a digitális bölcsészet részeként. A digitális filológia elnevezése, hasonlóan a digitális bölcsészethez, sokáig nem rögzült, különböző formákban hivatkoztak rá, mint például *számítógépes* vagy *elektronikus filológia*, ennek ellenére látható, hogy már évtizedekkel a fogalom megjelenése előtt megkezdődött a tudományterület módszertanának kidolgozása, az 1980-as években már zajlottak digitális filológiai kutatások, a *Helikon* 1989-ben megjelent számában Magyarországon is olvashatók voltak azok az értekezések,⁷ amelyekre a mai napig mint a digitális filológia kezdeti dokumentumaira hivatkozunk. De egy évtizeddel korábban Robert R. Dyer már megfogalmazott olyan gondolatokat, amelyek, ahogyan Fiormonte, Numerico és Tomasi *The Digital Humanist: A Critical Inquiry* című könyvükben rámutattak, akár egy „modern digitális bölcsészeti könyvből is származhatnának”:⁸

Úgy vélem, az emberiség történetében a legizgalmasabb tudomány küszöbén állunk, amit az ember eddig elképzelt. [...] [A]z írott és kimondott szó által működő elme tanulmányozása nehezen illeszkedik a nyelvészetbe vagy a pszichológiába. Filológia ez is, én azt mondom, az ember tanulmányozásának eme nagy ága. [...] Javaslatom szerint szükség van, még hozzá sürgősen, olyan nagy kutatóközpontokra, ahol a gondolatok verbális kifejezésének hagyományát és eredetiségét a számítógép segítségével lehet vizsgálni. [...] Ezekben a központokban viselkedéskutatókkal, nyelvészekkel, pszichológusokkal, statisztikusokkal, elektromérnökökkel, neurofizikusokkal és biokémikusokkal fogunk együtt dolgozni; a diákjaink pedig ezekben a diszciplínákban szereznek tudományos gyakorlatot. A filológia és az embernek mint szavak és eszmék alkotójának a vizsgálata itt válik majd valódi tudománnyá, és a humanizmus végre megtanulja, hogy alaptörekvései ugyanúgy nyitottak a

⁷ V.ö.: Andrew OLIVER, „Mikroinformatika és textológia”, ford. Ildikó FARKAS, *Helikon* 35, 3–4. sz. (1989): 412–420., Hans Walter GABLER, „A kiadói szöveg születése: számítógép bába-szerepben”, ford. Ildikó FARKAS, *Helikon* 35, 3–4. sz. (1989): 421–428., Jerome J. MCGANN, „Az Ulysses mint posztmodern szöveg: a Gabler-féle kiadás”, ford. Judit FRIEDRICH, *Helikon* 35, 3–4. sz. (1989): 429–452.

⁸ Domenico FIORMONTE, Teresa NUMERICO és Francesca TOMASI, *The Digital Humanist: A Critical Inquiry*, szerk. Desmond SCHMIDT, Christopher FERGUSON és Geoffrey ROCKWELL (punctum books, 2015), 161, <https://doi.org/10.21983/P3.0120.1.00>.

természettudományos formalizálásra, mint az emberi tapasztalat meghódítására irányuló bármely más kutatás.⁹

A szerzőhármás által idézett, 1969-ben megfogalmazott vízió talán ebben a formában soha nem valósul meg, de a digitális filológia történetéből egyértelműen kiolvasható az a törekvés, hogy a mindenkori legkorszerűbb számítógépes eszközök alkalmazásával, más diszciplínák képviselőivel együttműködve vizsgálja és publikálja a szövegeket. A digitális filológiai projektek végső célja a kutatási adataik közzététele, digitális tudományos, vagy digitális kritikai szövegkiadás formájában. A két fogalmat párhuzamosan alkalmazza a nemzetközi szakirodalom, de a digitális tudományos kiadás (Digital Scholarly Edition – DSE) fogalom értelmezése tágabb, egyaránt érthetünk alatta annotált vagy kritikai kiadást is. A szakterminológiában sokáig a *digitális* és *elektronikus* (ritkábban *hálózati*) fogalmakat szinonimaként használták, az elmúlt években azonban inkább a *digitális* forma terjedt el, a disszertáció alapvetően ezt a terminust alkalmazza, azokban az esetekben tér el, amikor a hivatkozott szakirodalom, vagy kiadás valamely fent említett szinonimát használja.

Magyarországon a digitális tudományos szövegkiadás terminus nem terjedt el, nálunk, egyedülálló módon, a Magyar Tudományos Akadémia Textológiai Munkabizottsága definiálja, hogy mely kiadások tekinthetők kritikai kiadásnak. A Munkabizottság viszonylag korán, már a 2004-ben megjelentetett *Alapelvekben*¹⁰ ajánlásokat tett az elektronikus kritikai kiadások publikálásának módszertanára. Kecskeméti Gábor, az *Alapelvek* előszavában foglalta össze az elektronikus kritikai kiadások ismérveit.¹¹ Ahogyan Kecskeméti is kiemeli, az elektronikus szöveg fő jellemzőjeként sokáig a hiperlinkek által lehetővé tett dinamikus olvasást tekintették, holott nem feledkezhetünk meg a mögöttes szemantikus struktúráról sem, amely kizárólag jelölőnyelvi rendszerekkel teremthető meg. Az *Alapelvek*, bár nem javasol konkrét szabványt, leszögezi, hogy „egy alkalmas jelölőnyelv szerinti struktúra-kódolásban kell rögzíteni [a kiadásokat], vagy lehetővé kell tenni, hogy ilyenné konvertálva legyenek kinyerhetők.”¹² A digitális tudományos és kritikai kiadások ma széles körben elfogadott meghatározását a nemzetközi szakirodalomban Patrick Sahle fogalmazta meg 2016-ban:¹³

⁹ Idézi Uo., 161–162.

¹⁰ DEBRECZENI Attila és KECSKEMÉTI Gábor, „Alapelvek az irodalmi szövegek tudományos kiadásához”, *Irodalomtörténet* 35, 3. sz. (2004): 328–330.

¹¹ KECSKEMÉTI Gábor, „A textológia egyes problémáiról – az új textológiai alapelvek közrebocsátásakor –”, *Irodalomtörténet* 35, 3. sz. (2004): 317–327.

¹² DEBRECZENI és KECSKEMÉTI, „Alapelvek az irodalmi szövegek tudományos kiadásához”, 329.

¹³ Patrick SAHLE, „About a Catalog Of Digital Scholarly Editions, 2008-2011”, <https://v3.digitale-edition.de/vlet-about.html>.

Az általam használt definíciót, „Edition ist die erschließende Wiedergabe historischer Dokumente”, nem lehet közvetlenül lefordítani angolra. Jó megközelítése lehet a definíciónak, „A tudományos szövegkiadás a történeti dokumentumok kritikai reprezentációja”. [...] Ez a „tudományos szövegkiadás” tág értelmezése. De mi a „digitális tudományos szövegkiadás”? Ezek nem csupán tudományos szövegkiadások a digitális médiumban. Különbséget teszek a digitális és a digitalizált között.

Egy digitalizált nyomtatott kiadás szigorúan véve nem „digitális kiadás”. Egy digitális kiadás nem nyomtatható ki az információk és/vagy a funkcionalitás elvesztése nélkül. A digitális kiadást más paradigma vezeti. Ha egy kiadás paradigmája az „oldal” kétdimenziós terére és az információ reprezentációjának tipográfiai eszközeire korlátozódik, akkor az nem digitális kiadás.

Sahle szerint már a hiperlinkek használata is olyan mértékben rendezi újra a kiadás szerkezetét, hogy létrejön a határvonal a nyomtatott vagy digitalizált, valamint a valódi digitális kiadások között. A hiperlinkek használata, szemben az implicit hivatkozásokkal olyan moduláris szemléletet hoz létre, amely elmosza a határokat a kiadás és az apparátus között, ez pedig nem valósítható meg papíron.¹⁴ Sahle azonban nem tér ki a nyomtatott kiadások alapján készített jelölőnyelvi átiratokra, így érdemes ebből a szempontból is megvizsgálni a kérdést.

Az angol nyelvben a digitalizálás kifejezésére két eltérő jelentésű szót alkalmaznak: *digitalization* és *digitization*.¹⁵ A *digitalization* az *Oxford English Dictionary* szerint a digitális technológiára való áttérést jelenti,¹⁶ míg a *digitization* a digitalizálás folyamatát jelöli.¹⁷ A magyar *digitalizál* szó azonban nem érzékelteti olyan jól a különbséget a két folyamat között. Maróthy Szilvia a *digitális átültetés* fogalmat javasolja tanulmányában,¹⁸ azonban ez a

¹⁴ Patrick SAHLE, „What is a Scholarly Digital Edition”, in *Digital Scholarly Editing: Theories and Practices*, szerk. Matthew James DRISCOLL és Elena PIERAZZO, 19–39 (Cambridge: Open Book Publishers, 2016).

¹⁵ A két fogalom közötti különbségről bővebben: J. Scott BRENNEN és Daniel KREISS, „Digitalization”, in *The International Encyclopedia of Communication Theory and Philosophy*, szerk. Klaus Bruhn JENSEN és mtsai., 556–566 (Malden: John Wiley & Sons, 2016).

¹⁶ „the adoption or increase in use of digital or computer technology by an organization, industry, country, etc.”

¹⁷ „the action or process of digitizing; the conversion of analogue data [esp. in later use images, video, and text] into digital form”

¹⁸ MARÓTHY Szilvia, „Tudományos szövegkiadások a hálózaton”, *Irodalomtörténeti Közlemények* 122, 5. sz. (2018): 617–633.

megfogalmazás véleményem szerint túl metaforikus, ezért Palkó Gábor nyomán¹⁹ az *informatizálás* kifejezés használatát preferálom, mivel ez közelebb áll a médiumváltás folyamatának kifejezéséhez. Sahle-nak abban kétségtelenül igaza van, hogy egy kiadás PDF reprezentációja nem tekinthető digitális kiadásnak, azonban, amennyiben hiszünk Elena Pierazzónak,²⁰ és a jelölőnyelvi kódolást a szöveg interpretációjának tekintjük,²¹ a jelölőnyelv implementálása inkább informatizálás, mint pusztán digitalizálás. Egy nyomtatott kiadásra épülő „digitalizált” kiadás, amely jelölőnyelvi átiratként készül el, olyan komplex rendszer, amelynek „el sem képzelhető papírváltozata”.²² A *born digital* kritikai kiadás fogalma viszonylag fiatal, a fent idézett definíciók nem is alkalmazzák a kifejezést, ezért érdemesnek tartom az idézett definíciókat részben átfogalmazni és kiegészíteni. Általánosságban elmondható, hogy minden, számítógépes eszköz segítségével készült kritikai kiadás *born digital* kiadás. A digitális filológia területén a fogalmat mégsem minden tudományos, vagy kritikai kiadásra alkalmazzák, hanem csak azokra, amelyek eleve jelölőnyelvi átiratként jönnek létre. Definiálhatnánk úgy a *born digital* kiadásokat, hogy ezek elkészítésük első pillanatától kezdve jelölőnyelvi átiratban készülnek, céljuk a digitális közzététel, ha készül is nyomtatott változat, az csupán pillanatnyi, töredékes lenyomata a teljes és autentikus változatnak, a digitális kiadásnak.

A digitális filológia mint adatszerkezet

A nyomtatott és digitális kiadások közötti különbségek elemzésekor a szakirodalom nem érint egy olyan kérdést, amely mentén szintén meg lehet ragadni a digitális szemléletű szövegkiadások lényegi jellemzőit, ez pedig az adatok, adathálózatok kérdése. A filológia alapvetően, függetlenül az adott irányzat szövegközlési elveitől, adatvezérelt. Az adatok komplex kezelésére valójában a nyomtatott médium korlátozottan alkalmas, a kiadások előkészítési munkálatai során felgyűlt adathalmaz számos esetben szinte nyom nélkül eltűnik a publikálást követően. Nem véletlen, hogy az elmúlt két évtizedben rengeteg

¹⁹ Szóbeli közlés alapján, a fogalom első publikált szövegben való használata tőlem származik, lásd: FELLEGI Zsófia, „Átmenet az analógból a digitális filológiába: médiumváltás? Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”, *Helikon* 67, 1. sz. (2021): 84–102.

²⁰ Pierazzo felveti ugyan a kérdést a digitalizált kiadások kapcsán, azonban ő egy-egy szolgáltatás definiálása felől közelít. Mi a különbség egy digitális archívum, könyvtár és kiadás között? Hol húzhatók meg a határok? Hogyan definiáljunk egy hibrid szolgáltatást, ahol digitalizált és digitálisan született kiadások egyaránt megtalálhatók? Pierazzo nem ad ezekre egyértelmű válaszokat, azonban megjegyzi, hogy a digitalizálni kívánt kiadások kiválasztásának szempontjai és archívumba emelésük is interpretációs aktus. Elena PIERAZZO, *Digital Scholarly Editing: Theories, Models and Methods*, 2. (New York: Routledge, 2016), 193–201.

²¹ Uo., 99.

²² KECSKEMÉTI, „A textológia egyes problémáiról...”, 322.

irodalomtudományi, illetve filológiai kutatásokhoz kötődő, egymástól független adatbázis született, a tudásmegosztás, az adathálózatok közzététele természetes igénye a filológiának. Az azonban egy egészen új kihívás, hogy ezeket az adatbázisokat hogyan lehet a kritikai kiadások szolgálatába állítani. Szintén kiemelt fontosságú, hogy mit tekintünk adatbázisnak: a számítástechnikában az adatok jól strukturált halmazát nevezik így. Az adatbázisok létrehozásának célja az adatok kezelése, valamint az információkinyerés, ehhez a célnak megfelelő számítógépes eszközöket, adatbáziskezelő-rendszereket kell használni, amelyek lehetővé teszik az adatok karbantartását, az adatbázis lekérdezését. Az adatbázisok struktúrája eltérő lehet, ennek megfelelően az adatbáziskezelő szoftverek is eltérően működnek.

Az adatbázisok mellett a disszertációban a különböző, nemzetközileg széles körben használt metaadat-szabványokat is érintem, melyek a filológiai szövegekre vonatkozó információk, leírások tárolására alkalmasak lehetnek, továbbá vizsgálom ezek használatának előnyeit és hátrányait.

Az elmúlt években egyre nagyobb a nyomás, hogy a rendelkezésre álló technológiák alkalmazása mellett a digitális filológia kiszolgálja az olyan jogos kutatói igényeket, mint a stabil hozzáférhetőség biztosítása, a verziókövetés és a hosszú távú fenntarthatóság. Ehhez hasonló elveket fogalmazott meg Mark D. Wilkinson és szerzőtársai a FAIR adatkezelési alapelvek (Findable, Accesible, Interoperable, Reusable) közzétételekor 2016-ban.²³ Az irányelvekben a kutatási adatok kapcsán többek között az alábbi kritériumokat rögzítették:²⁴

1. Megtalálható

- emberek és gépek számára egyaránt
- egyedi azonosítókkal rendelkezzen
- legyen metaadatokkal gazdagon ellátott
- a metaadatok repozitóriumban legyenek tárolva

2. Elérhető

- emberek és gépek számára egyaránt könnyű hozzáférés vagy letöltés biztosítása
- szabványos kommunikációs protokollon keresztül legyenek elérhetőek az adatok
- ha az adatok megsemmisülnek, a metaadatok akkor is legyenek elérhetőek

3. Együttműködő

- lehetőséget ad az adatok változtatására, cseréjére

²³ Mark D. WILKINSON és mtsai., „The FAIR Guiding Principles for Scientific Data Management and Stewardship”, *Scientific Data* 3, 1. sz. (2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

²⁴ Forrás: <https://openscience.hu/f-a-i-r-kutatasi-adatkezeles/>.

- a (meta)adatok a FAIR alapelveit követő szókészleteket használják
 - a (meta)adatok tartalmazzák a más (meta)adatokra vonatkozó minősített hivatkozásokat
4. Újra felhasználható
- lehetőséget ad az adatok jövőbeni felhasználására
 - a (meta)adatok egyértelmű és hozzáférhető adathasználati engedélyekkel kerülnek kiadásra
 - a (meta)adatok megfelelnek a domain-specifikus közösségi szabványoknak.

Bár a digitális filológia esetében bizonyos kritériumoknak nem könnyű megfelelni (pl. a szabványos protokoll biztosítása a jelölőnyelvi átiratokhoz), ma már nem képzelhető el olyan digitális kiadás, amely nem próbál meg legalább részben igazodni ezekhez az elvekhez. A digitális kiadások alapját képező jelölőnyelvi átiratok kiemelt fontosságúak a digitális filológia területén, az alkalmazott technológia nagyban befolyásolja, többek között, a hosszú távú megőrzést, kereshetőséget, az egyes formátumok közötti átjárhatóságot, általánosságban: a megfelelést a FAIR alapelveknek. A problémakör megértéséhez nem tekinthetünk el a jelölőnyelvi kódolás kérdéseinek érintésétől. Ez egyszerre foglalja magában az egyes jelölőnyelvi átiratok gyakorlati jellemzőinek értékelését és a mögöttes elméleti kereteket.

A Text Encoding Initiative 1987-ben alakult, azzal a céllal, hogy olyan ajánlást dolgozzon ki, amelynek segítségével a számítógép számára is olvasható módon leírhatók a szövegek és a hozzájuk tartozó metaadatok. Eleinte az akkoriban használatos SGML (Standard Generalized Markup Language) nyelvre, majd az abból kifejlesztett, azt lassan leváltó, 1998-ban megjelent XML (eXtensible Markup Language) nyelvre dolgoztak ki ajánlást. A nemzetközi együttműködést célzó TEI Consortium 1999-ben alakult meg, de addigra már a tudósok széles köre foglalkozott a kérdéssel, az ajánlás első verziója 1990-ben jelent meg.²⁵ Azóta a TEI-t egy szakértőkből álló csoport fejleszti, és jelenleg az ajánlás ötödik verziója érhető el, amelynek PDF változata közel 2000 oldalt tesz ki.²⁶ A TEI azonban nem csak egy ajánlás, ahogyan Fotis Jannidis fogalmazott, valójában három dolog: intézmény, kutatói közösség és fogalmak, jelölők halmaza.²⁷ Elena Pierazzo szerint a TEI nagyszerűsége részben ebből a hármasságból fakad, hiszen nem csak jelölési útmutatót nyújt, hanem egy olyan

²⁵ C. Michael SPERBERG-MCQUEEN – Lou BURNARD, *TEI P1 Guidelines for the Encoding and Interchange of Machine Readable Texts* (szoftver dokumentáció), Zenodo, 1990.
<https://zenodo.org/doi/10.5281/zenodo.3459202>.

²⁶ Lásd: <http://www.tei-c.org/Guidelines/P5/>.

²⁷ Fotis JANNIDIS, „TEI in a Crystal Ball”, *Literary and Linguistic Computing* 24, 3. sz. (2009): 253–265.

tudósközösség szakértelmét, amely páratlan a filológia területén.²⁸ A TEI segítségével egyaránt leírhatunk verset, egy-egy prózai művet, drámát, töredékes, javításokkal, áthúzásokkal, olvashatatlan szavakkal teli kéziratot is, valamint tökéletesen alkalmas egy teljes kritikai kiadás megszerkesztésére, a szerkesztői tevékenységek jelölésére, vagy épp az egyes szövegváltozatok együttes kezelésére. Számos grandiózus digitalizálási projekt (pl. Project Gutenberg,²⁹ Oxford Text Archive,³⁰ Beckett Archive³¹) használja az ajánlást, de létezik olyan vállalkozás is, amely továbbfejlesztette a TEI ajánlást a saját igényeinek megfelelően (pl. Van Gogh-levelezéskiadás³²). A TEI ajánlásának megjelenését követően megszülettek az első kritikák is, amelyek az XML hierarchikus struktúrájából fakadó problémákra mutattak rá. Az XML felépítése fa szerkezethez hasonlítható, az XML dokumentum egy gyökérellemmel rendelkezik, melynek bármennyi „gyermeke” lehet, és ez érvényes minden további elemre is. Az XML nyelv hátránya, hogy a szintaktikai szabályok nem teszik lehetővé az átfedések jelölését, tehát ha például egy bekezdéshatáron átível egy szövegjellemző, mondjuk egy törlés, akkor az nem jelölhető az XML hierarchia szerint helyesen. Erre számos egyedi megoldás születhet, például a pontszerű jelölők használata. Az OHCO (Ordered Hierarchy of Content Objects) modell problémái ellenére a TEI mára megkerülhetlenné vált a digitális szövegi kiadások területén,³³ nincs még egy ilyen részletesen kidolgozott, nemzetközileg elismert ajánlás a tudományos szövegek digitális közlésére vagy a kéziratok átíratának reprezentálására. Az 1990-es években zajló vitákat követően születtek is megoldási javaslatok, egy ideig az XML nyelvénél jóval „fiatalabb” JSON (JavaScript Object Notation) adatsere formátum tűnt a legesélyesebb „kihívónak”. A JSON, akár csak az XML, nyelvfüggetlen és ember által is jól olvasható adatformátum. Az XML struktúrával szemben a JSON kulcs-érték párok listájából áll, amelyek egymásba ágyazhatók. Az XML fájlok esetében az információkat „tag”-ekkel jelöljük, amelyeket attribútumokkal pontosíthatunk az alábbi módon:

²⁸ PIERAZZO, *Digital Scholarly Editing: Theories, Models and Methods*, 129.

²⁹ A Project Gutenberg szolgáltatásban különböző más formátumokban (pl. PDF, EPUB, HTML) is elérhetők szövegi kiadások. A projekt történetéről bővebben: <https://www.gutenberg.org/about/background/>.

³⁰ Az Oxford Text Archive projektről bővebben: <https://ota.bodleian.ox.ac.uk/repository/xmlui/page/about>.

³¹ A Samuel Beckett Digital Manuscript Project-ről bővebben: <https://www.beckettarchive.org/home>

³² A Vincent van Gogh The Letters kiadásról bővebben: https://vangoghletters.org/vg/about_6.html. A weboldalon a levelek jelölőnyelvi átírata szabadon hozzáférhető.

³³ A hierarchikus dokumentumleíró jelölőnyelvek hátrányairól bővebben lásd: Allen RENEAR, Elli MYLONAS és David DURAND, „Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies”, in *Research in Humanities Computing*, szerk. Nancy IDE és Susan HOCKEY, 263–280 (Oxford: Oxford University Press, 1996).

```

<name type="person">
  <surname>Petőfi</surname>
  <forename>Sándor</forename>
</name>

```

A fenti példa a JSON esetében az alábbi módon képezhető le:

```

{„name”:
  {„surname”: „Petőfi”,
  „forename”: „Sándor”}
}

```

Információtechnológiai szempontból a JSON előnye az XML-el szemben a számítógépes feldolgozáshoz szükséges kisebb memóriaigény volt, a digitális szövegkiadások szempontjából a formátum rugalmasságát és a szövegtest leírásától független, úgynevezett stand-off annotációk alkalmazási lehetőségeit emelték ki. A stand-off annotációk lényege, hogy egy-egy szövegszegmentumhoz az adatstruktúra egy más pontján további információkat rendelünk, korábban ebben látták a megoldást az egymást átfedő hierarchiák jelölésének problémájára. A nemzetközi gyakorlatban rendkívül ritka, amikor digitális filológiai vállalkozás JSON adatformátumot használ,³⁴ a digitális tudományos szövegkiadásokat gyűjtő katalógusban (*Catalogue Digital Editions*) jelenleg nem található JSON formátumban rögzített kiadás.³⁵ A JSON adatszerkezetben kulcs-érték párok esetében nincs megkötés a kulcsok elnevezésével kapcsolatban, az ismert digitális filológiai kísérletekben a TEI XML ajánlást próbálták implementálni. Zajlottak olyan kísérletek, amelyek a két módszert egyesíteni próbálták,³⁶ de a TEI ajánlásban is implementálták már az XML alapú stand-off annotációk használatát.³⁷ Bár úgy tűnt, hogy végül a TEI egyeduralkodó „szabvánnyá” vált a digitális filológiában, a mai napig születnek az XML leváltását célzó kísérletek, ezeket némelyikét és az alternatív jelölőnyelveket is elemzi később a disszertáció. A JSON mint adatszerkezet ugyan nem terjedt el a digitális szövegkiadások gyakorlatában, de a formátum jól alkalmazható a TEI XML

³⁴ Elie DANNAOUI, „Digital Arabic Gospels Corpus”, in *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, szerk. Claire CLIVAZ, Andrew GREGORY és Gregory HAMIDOVIĆ, 61–70 (Brill, 2014), https://doi.org/10.1163/9789004264434_006. Az idézett tanulmányon túl máshol nem tesznek említést az adatformátumról. A korpusz az alábbi linken érhető el: <https://pavone.uob-dh.org/>.

³⁵ A katalógus elérhetősége: <https://dig-ed-cat.acdh.oeaw.ac.at/>.

³⁶ Maróthy Szilvia részletesen beszámolt ezekről a kísérletekről: Szilvia MARÓTHY, „Beszámoló a bolognai ReIReS workshopról és néhány gondolat a szövegekódolás módszertanáról” (Digitális módszerek az irodalomtudomány támogatására, MTA BTK Irodalomtudományi Intézet, 2019. 12).

³⁷ James CUMMINGS, „A world of difference: Myths and misconceptions about the TEI”, *Digital Scholarship in the Humanities* 34, Supplement_1. sz. (2019. december 1.): 158–179, <https://doi.org/10.1093/llc/fqy071>.

alapú szolgáltatások informatikai megoldásaiban, mint például a vizualizációs szoftver működésében, vagy a keresőmotorok optimalizálásában. Említést kell még tennünk a digitális kiadások publikálására használt másik, sokáig népszerű formátumról is. Az XML formátum és a TEI ajánlás elterjedése előtt számos kiadás jelent meg HTML reprezentációban. A HTML a webes tartalmak leírására kidolgozott jelölőnyelv, amelyet felhasználva a web böngészők képesek megjeleníteni a szöveges, audiovizuális és egyéb tartalmakat. A HTML elsősorban a vizuális megjelenítésre szolgál, így nem teszi lehetővé a filológiai szempontból releváns szemantikai információk rögzítését. Ennek ellenére az XML alapú digitális tudományos, vagy kritikai kiadások esetében a felhasználó általában a webes megjelenítés során az XML fájl HTML reprezentációjával találkozik.

A disszertáció felépítése

1.) A disszertáció első fejezete (*Átmenet az analógból a digitális filológiába*) az analógból a digitális filológiába való átmenet metafilológiájával foglalkozik (elméleti szempontok figyelembevételével), de konkrét, a DigiPhil csoport szakmai támogatásával közzétett szövegkiadások összefüggésében. A bemutatott digitális tudományos és kritikai kiadások sokfélék, a fejezet arra törekszik, hogy a különböző filológiai iskolák, illetve szövegtörzsek számbavételével és elemzésével áttekintést nyújtson a digitális filológia legfontosabb kérdéseiről. Elsőként nyomtatott kiadások informatizálására láthatunk példát, itt a digitalizálás és informatizálás különbségére kerül a hangsúly. Ebben a részben az Arany János kritikai kiadások és Madách Imre *Az ember tragédiája, Szinoptikus kritikai kiadás* retrokonverzióval történő informatizálásának bemutatásán keresztül kaphatunk képet az egyes kiadástípusok eltérő szövegközlési elveinek egy rendszerben történő leképezéséről. Ezt követi a *born digital* kiadások elemzése. A *born digital* kiadások esetén nem tekinthetünk el a digitális filológia a szövegkiadás elveire gyakorolt hatásának, valamint a filológusi szerep megváltozásának elemzésétől. Az elemzés több kiadástípust vizsgál, az első esetben (*Móricz Zsigmond levelezésének kritikai kiadása*) a *born digital* kiadás nem egy sorozat részeként készült, a szövegközlési elveknek nem kellett egy sorozat elveihez igazodni. A második esetben (*Oláh Miklós levelezésének kritikai kiadása, 2. kötet*) a *born digital* kiadás előzménye, az első kötet, egy célszoftver segítségével készült, amelynek segítségével a sajtó alá rendező a kiadás készítésének teljes folyamatát, az átírástól kezdve az kritikai apparátus beillesztésén át a névmutató és a tördelt nyomdai formátum előállításáig egy rendszerben végezte el. Az első kiadás elkészítéséhez használt célszoftverben lehetőség lett volna szabványos jelölőnyelvi átírat

készítésére, azonban digitális filológiai kompetenciák nélkül az ehhez szükséges beállítások nem végezhetők el. A második kiadás egy, a kiadás igényeire szabott irodai szoftvert imitáló környezetben készült, ahol a felhasználóbarát felület segítségével jelölőnyelvi átirat jött létre, anélkül, hogy a sajtó alá rendezőnek ismernie kellett volna a TEI szabványt és az XML nyelvet. A kiadás a *Bibliotheca Scriptorum Medii Recentisque Aevorum* sorozat része, így a jelölőnyelvi átirat kialakításakor mind a sorozat, mind a DigiPhil szövegközlési elveit figyelembe kellett venni. Az utolsóként bemutatott vállalkozás, *Babits Mihály Összes Művei* sorozat Magyarországon biztosan, de valószínűleg európai viszonylatban is egyedülállónak számít. A genetikus kritika elveit követő kiadássorozat most készülő köteteinek XML reprezentációja, az anyag természetéből fakadóan, rendkívül bonyolult, a legújabb, a szemantikus web szemléletmódjának elterjedésével létrejött technológiák alkalmazása nélkül lehetetlen vállalkozás lenne elkészíteni. A szemantikus web, azaz a Web 3.0 koncepciójának alapját az a jelenség adta, hogy a Web 1.0 és 2.0 elterjedésével már a 2000-es évek elején is kezelhetetlen mennyiségű, rendezetlen adat és dokumentum vált elérhetővé a világhálón. A koncepció szerint a problémát intelligens gépi megoldásokkal lehet megoldani, azonban ehhez az adatoknak gépek által is értelmezhető formátumban kell rendelkezésre állniuk.

2.) A *Digitális filológia és a szemantikus web* című fejezet ezen technológiák és a szemantikus web szemléletmódjának a digitális filológiára gyakorolt hatását vizsgálja. Ez a hatás egyaránt tetten érhető a digitális filológia elméletében és a gyakorlatában is. A disszertáció a nemzetközi kísérletek és új elméletek elemzésén keresztül mutatja be a szemantikus web elvét és működését. A DigiPhil projekt a digitális kiadásokat a kezdetektől bibliográfiai és biográfiai adatokat tartalmazó háttéradatbázissal kötötte össze, az adatbázis könyvtári rendszerben épült, amely számos kompromisszummal járt. A szemantikusweb-technológiának alkalmazásával olyan gráf formájú tudástár hozható létre, amely már nem csak a könyvtári rendszerben tárolt adatokat képes kezelni, hanem képes a kutatási adatokat szemantikus hálózatba rendezni és az adatokból további következtetéseket levonni.

3.) A *digitális kiadásokon túl: Jelölőnyelvi átiratok elemzése* című fejezet azt vizsgálja, hogy a kritikai kiadások XML reprezentációi milyen számítógépes eszközökkel elemezhetők. A digitálisan reprezentált irodalmi szövegek esetében a kutatás sikerességéhez szükség van a metaadatok és a szövegek együttes kezelésére, illetve a szöveg több szempontú gépi feldolgozására. A nemzetközi szabványok és ontológiák, névterek használata révén lehetőség nyílt az egymástól – fizikailag – távol eső adatbázisokban való értelmes és kifinomult közös

keresésre. A fejezet első része, a kurrens web teóriák figyelembevételével, a metaadat-aggregáció folyamatát mutatja be. A fejezet második része jelölőnyelvi átiratok hálózati vizualizációinak lehetőségeivel foglalkozik. Az elemzések elvégzésére különböző eszközök állnak rendelkezésre, a fejezet ezen részében az *Arany János Levelezése 1866–1882*³⁸ informatizált kiadásának, azon belül is a levelek metaadatainak feldolgozását mutatom be, majd a metaadatok különböző eszközökkel történő vizualizációs lehetőségeit hasonlítom össze. A DigiPhil indulása óta közel 10.000 XML fájlból álló korpusz épült. A gyűjtemény jelentős részét 19–20. századi szerzők munkái teszik ki. A korpusz méretéből és a szerzők időbeli közelségéből adódik a lehetőség, hogy a rendelkezésre álló kritikai kiadásokon nyelvstatisztikai elemzéseket végezzünk. A szövegstatisztikai módszerekkel olyan, adott szerzőre jellemző mintázatokat deríthetünk fel, melyek korábban sokszor láthatatlanok maradtak a kutatók számára. A fejezet végén Kosztolányi Dezső kritikai kiadásainak vizsgálatával mutatom be a rendelkezésre álló módszerek némelyikét.

4.) A digitális filológia térnyerése az elmúlt évtizedekben vitathatatlan, azonban hiába az évtizedes tapasztalatok, a digitális filológia oktatása kiforratlan még, nem áll rendelkezésre átfogó képzés a leendő szakértők számára. Kifejezetten időszerű kérdés, tekintve, hogy a digitális, illetve digitalizált kiadásokkal foglalkozó kutatócsoportok száma egyre növekszik, de a magyarországi képzések jelenleg nem gondoskodnak szakértői „utánpótlásról”. A disszertáció záró fejezete, reflektál erre a problémára és bemutatja, milyen lehetőségek álltak rendelkezésre az elmúlt években és manapság a digitális filológia oktatása terén, illetve milyen szempontokat érdemes figyelembe venni egy komplex képzés kialakítása során. A javasolt szempontok egyaránt érvényesíthetők a graduális és a posztgraduális képzésben.

Történeti kontextus

Bár a disszertáció célja nem a magyarországi digitális filológia történetének megírása, ahhoz, hogy a későbbi fejezetek, elsősorban a 2014-ben induló DigiPhil egyes kiadásainak példái köré épülő érvelés történeti kontextusba kerüljön szükség van egy vázlatos és szelektív áttekintésre az érvelés szempontjából legjelentősebb digitális filológiai fejleményekről az elmúlt két évtizedből.

³⁸ KOROMPAY H. János, szerk., *Arany János, Levelezése 5.: 1866–1882*, Arany János Összes Művei 19 (Budapest: Universitas Kiadó–MTA BTK Irodalomtudományi Intézet, 2015).

A hálózati kritikai kiadások helyzetéről Tóth Tünde írt részletes ismertetőt 2005-ben, majd Palkó Gábor,³⁹ Maróthy Szilvia⁴⁰ és jómagam⁴¹ vizsgáltuk különböző szempontok mentén a hazai digitális filológiai helyzetét. A *Helikon* folyóirat 2020-ban megjelent *Számítógépes Irodalomtudomány* számában Szénási Zoltán közölt átfogó tanulmányt a digitális filológia magyarországi történetéről és helyzetéről.⁴² A következőkben mégis érdemesnek tartom a fellelhető információk összefoglalását, illetve olyan új szempontok felvetését, amelyek az említett tanulmányokban nem merültek fel.

Az első digitális filológiai kísérlet a digitális bölcsészet geneziséét tárgyaló érvelések szerint 1949-ben indult és Roberto Busa jezsuita szerzetes és az IBM együttműködéséhez (*Index Tomisticus*) köthető. Magyarországon viszonylag hamar, már 1976-ban megkezdődött egy olyan számítógépes irodalmi adatbázis építése a József Attila Tudományegyetemen,⁴³ amely a 17. századi magyar versekre vonatkozó elemző és textológiai adatokat (például metrum, nótajelzés, datálás, idegenkezűség stb.) tartalmaz. 1993-ban az Eötvös Loránd Tudományegyetemen működő munkacsoport versszövegekkel bővítette az adatbázist, így született meg az RPHA (*Répertoire de la Poésie Hongroise Ancienne*). Ebben az adatbázisban már igen kifinomult kereséseket tettek lehetővé a bevitt metrikai és irodalomtörténeti, bibliográfiai adatok.⁴⁴ Az adatbázist többször újraépítették és egyedülálló módon a mai napig folyamatosan fejlesztik és bővítik, szabadon elérhető.⁴⁵

Az 1990-es években és a 2000-es évek elején az Eötvös Loránd Tudományegyetem két programja, a Centre des Hautes Études de la Renaissance (CHER) és a Horváth Iván által vezetett Bölcsészettudományi Informatika Önálló Program⁴⁶ (BIÖP) műhelyek munkája révén számos, *born digital*, internetes megjelenésre szánt kritikai kiadás jelent meg. A *Balassa-kódex* átíratra volt az első hálózati forráskiadás, 1993-ban jelent meg faksimilékkel együtt. Három évvel később, az 1995-ös CD-változat után, tanulmányokkal bővítve újra kiadták, azonban

³⁹ PALKÓ Gábor, „Mit jelent a digitális filológia a szemantikus web korában? A DigiPhil projektről.”, *Magyar tudomány* 177, 11. sz. (2016): 1316–1322.

⁴⁰ MARÓTHY, „Tudományos szövegkiadások a hálózaton”.

⁴¹ FELLEGI Zsófia, „A digitális filológia infrastruktúrái. A DigiPhil megújulásáról”, in *Valós térben - az online térért, Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TÍCK József, KOKAS Károly és HOLL András, 338–344 (Budapest: Hungarnet Egyesület, 2022), <https://doi.org/10.31915/NWS.2022.43>.

⁴² SZÉNÁSI Zoltán, „Digitális átalakítás. A tudományos szövegkiadás a weben.”, *Helikon*, 66. sz. (2020): 99–112.

⁴³ BARTÓK István és MONOK István, „Szegedi kísérlet a XVI. századi magyar vers gépi feldolgozására”, *Irodalomtörténeti Közlemények* 84, 5–6. sz. (1980): 630–638.

⁴⁴ TÓTH Tünde, „Irodalomtörténészek a bábeli könyvtárból (A régi magyar vers repertórium)”, *Irodalomtörténeti Közlemények* 1997, 1–2. sz. (1997): 133–145.

⁴⁵ Répertoire de la poésie hongroise ancienne 7.4. változata: <https://f-book.com/rpha/v7/index.php>.

⁴⁶ Bölcsészettudományi Informatika Önálló Program: <http://magyar-irodalom.elte.hu>.

1998-ban megsemmisült.⁴⁷ Még ugyanabban az évben jelent meg Zrínyi Miklós *Adriai tengernek syrénája*⁴⁸ című műve online formában, ahol osztott képernyőn olvasható egymás mellett a szöveg két kiadása, a Syrena-kódex és a bécsi kiadás 1651-ből. 1998-ban tették közzé az online Balassi-kritikai kiadást, mely a mai napig elérhető. 1999-ben megjelent még József Attila összes tanulmánya és cikkének 1.0 verziója,⁴⁹ az *Effectus Amoris*⁵⁰ és Wathay Ferenc összes versének hálózati kritikai kiadásának „demó változata”,⁵¹ valamint A 16–17. századi gyülekezeti és graduális éneklésről szóló hosszabb szövegelemek kritikai kiadása.⁵² 2001-ben Kosztolányi Dezső *Édes Anna*,⁵³ az *Érdy-kódex*,⁵⁴ Rimay János ifjúkori versgyűjteménye⁵⁵ és Angerianus szerelmi epigrammakötetének „demó verziója”.⁵⁶ A BIÖP kiadványai a mai napig elérhetők, azonban a programot az egyetem megszüntette.⁵⁷ Bár már a projekt kezdetekor elérhető volt a Text Encoding Initiative ajánlása SGML-re, majd XML formátumra is, a kiadásokat mégis HTML nyelven kódolták. A választás érthető: a kor technológiája nem tette lehetővé sem az SGML, sem az XML formátumból automatikusan generált HTML oldalak létrehozását, a BIÖP célja pedig az interneten való széleskörű hozzáférés biztosítása volt.⁵⁸ A kiadások főszerkesztője, Horváth Iván a korszellemet, amikor a digitális publikálás elterjedni látszó formátuma még az e-book volt, meghaladva már 2001-ben így fogalmazott az „f-book”-ot bemutató szövegében: „Az »e-book« túl korán jelent meg a piacon. A hálózati képzetes könyv feltehetőleg csak évtizedek múlva fogja megtalálni végleges alakját. Az f-book »free«,

⁴⁷ TÓTH Tünde, „Online kritikai szövegkiadás Magyarországon az ezredfordulón”, *Helikon* 50, 3. sz. (2004): 417–442.

⁴⁸ ORLOVSZKY Géza, szerk., *Adriai tengernek syrénája, gróf Zrínyi Miklós. Párhuzamos forráskiadás* (Budapest: Gépeskönyv Kft., 1998), <https://magyar-irodalom.elte.hu/gepesk/zrinyi/>.

⁴⁹ HORVÁTH Iván és TVERDOTA György, szerk., *József Attila összes tanulmánya és cikke. v 1.0* (Eötvös Loránd Tudományegyetem, 1999), <http://magyar-irodalom.elte.hu/ja/eloind.htm>, A 2.0 verziót 2001-ben tették közzé: HORVÁTH Iván és TVERDOTA György, szerk., *József Attila összes tanulmánya és cikke. v 2.0* (Eötvös Loránd Tudományegyetem, 2001), <http://magyar-irodalom.elte.hu/ja/eloind.htm>.

⁵⁰ SZEGEDI Béla, szerk., *Effectus Amoris széphistória a XVI. századból. Digitális kritikai kiadás* (Budapest: Gépeskönyv Kft., 1998), <https://magyar-irodalom.elte.hu/effectus/>.

⁵¹ TÓTH Csilla és TÓTH Tünde, szerk., *Wathay Ferenc összes verse hálózati kritikai kiadás* (Budapest–Zalaegerszeg: Gépeskönyv Kft., 1999), <http://magyar-irodalom.elte.hu/gepesk/wathay/>.

⁵² H. HUBERT Gabriella, szerk., *Lelki énekekről régi magyarok. A XVI–XVII. századi gyülekezeti és graduális éneklésről szóló hosszabb szövegelemek. Hálózati kritikai kiadás* (Budapest: Gépeskönyv Kft., 1999), <http://magyar-irodalom.elte.hu/gepesk/eloszo/>.

⁵³ VERES András, szerk., *Kosztolányi Dezső Édes Anna hálózati kritikai kiadás* (Budapest: Gépeskönyv Kft., 2001), <http://magyar-irodalom.elte.hu/gepesk/kd/edesanna/>.

⁵⁴ Az Érdy-kódex elérése: <http://magyar-irodalom.elte.hu/erdy-kodex/>.

⁵⁵ TÓTH Tünde, szerk., *Rimay János verseinek fragmentumai. Kritikai kiadás* (Budapest: Gépeskönyv Kft., 2001), <http://magyar-irodalom.elte.hu/gepesk/rimay/>.

⁵⁶ A demó verzió elérése: <http://magyar-irodalom.elte.hu/gepesk/hac/>.

⁵⁷ Forrás: F-book nyitóoldal: <https://f-book.com/>.

⁵⁸ MARÓTHY Szilvia, „A tudományos elektronikus szövegkiadások feldolgozása”, in *Valóságos könyvtár – könyvtári valóság: Könyvtár- és információ tudományi tanulmányok 2018*, 351–356 (ELTE BTK Könyvtár- és Információ tudományi Intézet, 2018), 353, <https://doi.org/10.21862/vkky2018>.

azaz: szabadon fejleszthető, mert nincsenek mögötte szabványoknál merevebb ajánlások, miként az e-book esetében. A legjobb képzetes könyvek már ma is gazdagabbak, mint az e-book-ok. Néhány év múlva összehasonlíthatatlanul jobbaknak kell lenniük.”⁵⁹ A kiadások közül egyedül a Balassi kiadás további sorsáról vannak információink, Maróthy Szilvia összefoglalója szerint⁶⁰ a kiadás 2004 folyamán jelölőnyelvi és technológiai megújításon esett át, félautomatikus módszerekkel a HTML fájlokból a TEI P4-nek megfelelő formátumot állítottak elő, azonban sem a jelölőnyelvi átiratot, sem a szoftveres frissítést nem tették közzé. A jelölőnyelvi átiratot még ebben az évben Parádi Andrea a P5 ajánlásának megfelelően átkódolta, azonban ez sem hozzáférhető. A Gépeskönyv Kft., a digitális kiadások kiadója az elmúlt években új fejlesztésekbe kezdett, ezeknek révén jött létre az F-book portál. A fejlesztők célja, hogy a régi kiadásokat restaurálják és kompatibilissá tegyék a mai böngészőkkel, a régi kiadások honlapját „múzeumként változatlan formában hagyták.”⁶¹ Az oldal szerkesztőinek tájékoztatása szerint a Balassi kiadás átdolgozásának munkálatai jelenleg is zajlanak.

A BIÖP megszűnését követő évtized digitális filológiai kezdeményezéseiről keveset tudunk, de ebben az évtizedben is hozzáférhetővé váltak kritikai kiadások digitalizált változatban. Az 1990-es évek elején alapított Arcanum Adatbázis Kft. kiterjedt digitalizálási tevékenysége meghatározó. A cég folyamatosan digitalizál és tesz közzé közgyűjteményi anyagokat online, korábban számos offline, DVD-ROM alapú digitális kiadást is közzétettek. A 2001-ben kiadott Arcanum DVD könyvtár I. tartalmazza például *Mikszáth Kálmán összes művei* 86 kötetes kritikai kiadás-sorozatát. A DVD anyagának jelentős részét később online is közzétették.⁶² Szolgáltatásaik közé tartozik a 2014-ben indult Arcanum Digitális Tudománytár, amely számos tudományos és szakfolyóirat (pl. *Alföld, Életképek*), valamint heti- és napilap (pl. *Pesti Hírlap, Népszabadság*), lexikonok (pl. *A Pallas nagy lexikona*), mellett kritikai kiadások (pl. *Régi magyar költők tára*) digitalizált változatát is közzéteszi. A szövegek közötti böngészés ingyenes, de a szövegek megtekintése és letöltése regisztrációhoz és előfizetéshez kötött, illetve az EISZ (Elektronikus Információs Szolgáltatás) tagjai számára biztosít a hozzáférést a cég. A webszolgáltatásból kiadványok kétrétegű PDF formátumban tölthetők le.⁶³ A közzétett

⁵⁹ HORVÁTH Iván, „F-BOOK, F-CONTENT”, in *Új interaktivitás felé: f-book* (Budapest: Gépeskönyv Kft., 2001), <http://magyar-irodalom.elte.hu/vita/thi.html>.

⁶⁰ MARÓTHY, „A tudományos elektronikus szövegi kiadások feldolgozása”.

⁶¹ Forrás: F-book nyitóoldal.

⁶² <https://www.arcanum.com/hu/product/ADVDK1/>.

⁶³ Az Arcanum a közgyűjteményektől lapra vágható formában kéri be az anyagokat, majd lapadagolós szkennert és karakterfelismerő program segítségével elkészítik a PDF-eket: NAGY Gyula, „Hírlapok digitalizálása és szolgáltatása a világban és Magyarországon A Délmagyarország című napilap digitalizálása és szolgáltatása”, *Könyvtári Figyelő* 2014, 4. sz. (2014): 439–458, 449.

anyagokban az Arcanum saját fejlesztésű keresőjével végezhetünk összetett kereséseket, a teljes szövegben és a metaadatokban is. Ők üzemeltetik továbbá a Hungaricanat, amely közgyűjteményi digitalizálások⁶⁴ eredményeit publikálja és lehetővé teszi a különböző adatbázisokban való közös keresést.⁶⁵ Mivel rendszerük teljesen zárt, jelenleg nincsenek pontos adataink arra vonatkozólag, hogy milyen technológiákkal dolgozzák fel a szövegeiket.

Bár a Magyar Elektronikus Könyvtár (MEK) kiadványainak döntő hányada nem tudományos igényű kiadás, elérhetővé tettek néhány digitalizált kritikai kiadást is.⁶⁶ A MEK ötlete 1994-ben merült fel először, és a már online elérhető szolgáltatást az OSZK vette szárnyai alá 1999-ben.⁶⁷ Alapjául a MEK által digitalizált dokumentumok, a kötelezpéldányok és az osztály számára önkéntesen beszolgáltatott digitalizált dokumentumok szolgálnak, melyek érkezhettek akár magánszemélytől, akár intézménytől. A szövegeket különböző formátumban szolgáltatják (pl. PDF, RTF, XML, HTML stb.) ennek hátránya, hogy a szövegekre csak nehezen alakítható ki egy, a szemantikus web elveit követő keresőrendszer, valamint a közzétett szövegek kevés, illetve eltérő mennyiségű és minőségű metaadattal rendelkeznek.

A MEK-hez hasonlóan szintén nem kritikai kiadásokat publikál, de a tudományos szövegkiadások irányába mozdul el az egyik legnagyobb jelentőségű szolgáltatás, a Digitális Irodalmi Akadémia. 1998 tavaszán indult a DIA közvetlen elődje, a Digitális Halhatatlanok program a Petőfi Irodalmi Múzeumban (a Múzeum abban az évben Magyar Irodalom Háza néven működött), az Akadémia az alakuló ülés döntése alapján kapta nevét.⁶⁸ A DIA célkitűzése egyedülálló volt: kortárs magyar irodalmi művek digitalizálása és közzététele, valamint a szerzők életrajzának, bibliográfiájának, másodlagos bibliográfiájának, és a róluk szóló digitalizált szövegeknek a közzététele. Együttműködő partnerként a Neumann-ház felelt a megfelelő szabvány kiválasztásáért és a szolgáltatás informatikai hátteréért. A DIA szövegkiadásai 2000 októberétől érhetőek el és 2007 óta olvashatók online formában, előtte oktatási, kutatási és magáncélú használatra biztosították a letöltést.⁶⁹ A Neumann-ház döntése alapján a DIA szövegeit SGML nyelven kódolták, ezeket 2008-ban XHTML formátumra

⁶⁴ pl. Magyar Levéltári Kiadványok Adatbázisa, Képeslap adatbázis

⁶⁵ Az Arcanum Adatbázis Kft. szolgáltatja a Mapire nevű oldalt is, mely a Habsburg birodalom térképeit tartalmazza, melyeket különböző vizualizációs technológiák segítségével tettek különösen látványossá.

⁶⁶ Például a Vitályos László által gondozott *Ady Endre levelezése I.*: <http://mek.oszk.hu/05500/05565/html/>, vagy a Kerényi Ferenc szerkesztette *Madách Imre Ember tragédiája szinoptikus kritikai kiadás*: <http://mek.oszk.hu/13700/13789/13789.pdf>.

⁶⁷ A MEK történetét lásd: <http://mek.oszk.hu/html/tortenet.html>.

⁶⁸ TÓSZEGI Zsuzsanna, „A Digitális Irodalmi Akadémia első két évtizede”, in *Valóságos könyvtár – könyvtári valóság. Könyvtár- és információtudományi tanulmányok*, szerk. KISZL Péter és CSÍK Tibor, 155–165 (Budapest: ELTE BTK Könyvtár- és Információtudományi Intézet, 2018), 157.

⁶⁹ Uo., 159.

alakították át.⁷⁰ Fontos kiemelni, hogy a DIA a szövegek digitalizálása során nem a nyomtatott forrásdokumentum betűhű digitális kópiáját teszi közzé, hanem szakértők közreműködésével javított, önálló digitális szövegkiadást készít, és bár megfogalmazásuk szerint „a szövegek feldolgozása, szükség szerinti kritikai kiadás előkészítésére alkalmas javítása a tagokkal, örökösökkel vagy az irodalmi szakértőkkel együttműködésben zajlik. Ennek köszönhetően a honlapon publikált szövegek szándékuk szerint megbízható, az eredeti, nyomtatásban megjelent változathoz pontosabb kiadást jelentenek.”⁷¹ A DIA valójában, a legjobb szándéka ellenére, olyan új szövegváltozatot hoz létre, amely semmiképp nem lehet alapja egy modern kritikai kiadásnak. A jelöletlen emendációk valójában veszélyeztethetik egy kritikai kiadás megbízhatóságát, a filológusnak betűszinten össze kell vetnie a DIA által közölt szövegvariánst a többi szövegvariánssal, különösen igaz ez azon kortárs szerzők műveire, akiknek még életük során közzéteszik a kiadásait. Az Esterházy Péter Kutatócsoport, a *Termelési regény* kritikai kiadásának előkészítési munkálatai során talált olyan szöveghelyeket, ahol a DIA a kiadásában az első két kiadásban szereplő alak helyett a harmadik kiadástól megjelenő formát emelte át és olyat is, ahol a második kiadástól megváltozott szövegeket elvetve az első kiadásban szereplő formát közlik.⁷² A szövegkiadás elveire vonatkozóan a DIA csak egy rövid ismertetőt közöl, nem derül ki, hogy az egyes művek esetében milyen filológiai elvek mentén határozzák meg az adott kiadásban közölt változat szövegét, illetve kik az adott kiadás készítésében részt vevő szakértők.

Szintén érdemes kiemelni azt, az elmúlt néhány évben rögzült gyakorlatot, hogy a nyomtatásban megjelent kritikai kiadások nyomdai PDF-eit a kiadást követően meglehetősen hamar, ha nem azzal egyidőben közzéteszik online. Ebben a gyakorlatban úttörőnek számított a Reciti kiadó, ami kritikai igényű és tudományos szövegkiadásait 2009 óta tesz közzé letölthető PDF formában. A Reciti számos kiadványa LaTeX formátumban készült,⁷³ amely valójában szövegszerkesztésre és nyomdai tördelésre fejlesztett nyílt forráskódú formátum, célja

⁷⁰ A DIA befogadó intézménye 2000 és 2007 között a Neumann-ház volt, 2007-ben került vissza a PIM alá. Ezt követően a Neumann-házzal kötött együttműködés megszűnt.

⁷¹ Forrás: A DIA szövegkiadási elvei: <https://dia.hu/dia-szovegkiadasi-elvei>.

⁷² A regény 63. oldalán szereplő szövegrészletben: „(néha Thaisz Elek főkapitány úr „hordárjai”; de hiszen szegénynek magának neki több a főfájása ebből, mint annak, akit bekísértet)” a „de hiszen” az első két kiadásban „dehiszen” formában szerepel, a DIA szövege a harmadik kiadástól megjelenő alakot használja, míg a 128. oldalon szereplő „Jaj, nagy a baj, édes vezérigazgató elvtárs?” esetében a DIA az első kiadás gyakorlatát követte, az idézett rész a regény második kiadásától kezdve így szerepel: „Jaj, nagy a baj, édes vezérigazgató elvtárs.” V.ö. ESTERHÁZY Péter, *Termelési-regény (kissregény)*, I. (Budapest: Magvető, 1979), 63., Uo., 128. és ESTERHÁZY Péter, *Termelési-regény (kissregény)*, V. (Budapest: Digitális Irodalmi Akadémia - Petőfi Irodalmi Múzeum, 2011), <https://reader.dia.hu/document/Esterhazy-Peter-Termelési-regeny-430>.

⁷³ A LaTeX formátumot Leslie Lamport fejlesztette az 1980-as években, a formátumról bővebben lásd: <https://www.latex-project.org/about/>.

alternatívát nyújtani az olyan, zárt szövegszerkesztőkkel szemben, mint például a Microsoft Office Word. A formátum konvertálható például HTML vagy PDF formátummá. A PDF szabvány publikálásra és archiválásra való használata számos problémát vethet fel. Ennek megoldására fejlesztette ki az Adobe a PDF/A-t, mely megkísérli alkalmassá tenni a PDF-et – mint alapvetően közzétételi céllal létrejött digitális formátumot – archiválási célokra is. A PDF-ekben való keresés azonban továbbra is problémát jelent. A rendkívül elterjedt kétrétegű PDF-ek esetében, amennyiben nem megfelelő a képi réteg mögötti szövegréteg (pl. a javítatlan OCR-ezett szövegek esetében), hibás, hiányos találati halmazokat kapunk. A PDF-ben publikált szövegek esetében további nehézséget jelent az adatgazdagítás (pl. nevek azonosítása), illetve a szövegek egyes részeinek megjelölése azonosítókkal. A technológia elméletileg alkalmas ilyen feldolgozásra (három- és négyrétegű PDF), de nagyon szűk körben alkalmaznak csak ilyen eljárásokat. További probléma, hogy a népszerű publikációs szoftverek jó része nem tartalmaz tudományos célra alkalmas, kifinomult keresőrendszereket, amelyek egységesen keresnének a metaadatok között és a szövegekben.

Mint említettem a magyarországi digitális filológiai tevékenységről több mint 10 évig nagyon kevés forrás áll rendelkezésre. Kabdebó Lóránt 2004-ben számolt be a készülő Szabó Lőrinc kritikai kiadás előkészítő munkálatairól,⁷⁴ a Miskolci Egyetemen működő Szabó Lőrinc Kutatóhely akkoriban még CD alapú digitális kiadást tervezett készíteni, a versek kritikai igényű kiadása végül online formában jelent meg 2011-ben, amelynek jelölőnyelvi kódolásáról nem áll rendelkezésre információ, de a honlap szerkezetét vizsgálva levonható a következtetés, hogy a kiadás HTML formátumban készült.⁷⁵ Szintén ebben az időben jelent meg Mátyus Norbert sajtó alá rendezésében az Babits Mihály Dante *Isteni színjáték* fordításának kritikai kiadásának részletei hálózati kiadás formájában.⁷⁶ A kiadás, tudomásom szerint, HTML formátumban készült, a készülő kritikai kiadás részleteit tartalmazta,⁷⁷ az eredeti tervek szerint a teljes kritikai kiadás nyomtatásban jelent volna meg, azonban ez finanszírozás hiányában nem valósult meg. A kritikai kiadás nyomtatásban való megjelentetésére a Szénási Zoltán által vezetett *Babits Mihály verseinek és műfordításainak kritikai kiadása* (NKFIH-138529) pályázat vállalkozott, ezzel egyidőben a kiadás digitális, új, teljes változata a DigiPhil oldalán fog

⁷⁴ KABDEBÓ Lóránt, „A Szabó Lőrinc-szövegkiadás múltja és jövője”, in *Az V. Nemzetközi Hungarológiai Kongresszus (Jyväskylä, 2001. augusztus 6-10.) előadásai II.*, szerk. JANKOVITS József és NYERGES Judit, 91–101 (Budapest: Nemzetközi Magyarológiai Társaság, 2004).

⁷⁵ <http://krk.szabolorinc.hu/>.

⁷⁶ A kiadás a mai napig megtekinthető: <http://www.dantealighieri.hu/>. A dolgozat lezárásakor a kiadás weboldala szerverhibát jelez, az Internet Archive másolata nem nyitható meg.

⁷⁷ MÁTYUS Norbert, „Babits Dante-fordításának kritikai kiadása és Dante életműve”, Dokumentáció, OTKA Kutatási Jelentések (OTKA, 2012), <https://real.mtak.hu/12143/>.

megjelenni. A másik, ekkoriban megjelent digitális kiadás helyzete és utóélete, ahogyan Szénási is rávilágít,⁷⁸ egyedülállónak számít. Horváth Iván és Fuchs Anna szerkesztésében jelent meg 2012-ben *József Attila összes tanulmánya és cikke, 1930–1937* 1.0 verziója.⁷⁹ A József Attila Társaság kutatócsoportja a kutatások újabb eredményeit azonban már nyomtatott kiadásban publikálta, ahogyan Szénási is megjegyzi, „[a]z új nyomtatott kritikai kiadás szerkesztői előszava pusztán a szövegkritikai problémákkal indokolja az új kiadás szükségességét, és szó nélkül hagyja a médium(vissza)váltás problematikáját.”⁸⁰

A felsorolt HTML alapú kritikai kiadásoknak közös vonása, hogy a választott jelölőnyelvre rendkívül stabil környezetet építettek, a BIÖP kiadásai, néhány kisebb technológiai hibát leszámítva ugyanúgy működnek, mint amikor közel 25 éve közzétették őket. Egyaránt alkalmasak a párhuzamos szövegolvasásra és a faksimilék megjelenítésére, a kritikai jegyzeteket egyszerű felugró ablakokban jelenítik meg a felhasználó számára. Mégsem tekinthetők valójában informatizált, vagy digitális kritikai kiadásnak, ugyanis a HTML nyelv alkalmazása miatt nem volt lehetőségük a filológiai jelenségek szemantikai jelölésére és megjelölésére. Az egyes szövegeltérések jelölésére ugyanúgy, egy láb- vagy végjegyzethez hasonló formátumot használnak, mint egy nyomtatott kiadás esetében.

Az MTA-DE Textológiai Kutatócsoport egyaránt publikál korábban nyomtatásban megjelent kritikai kiadásokat digitális formában, valamint eleve *born digital* kiadványokat.⁸¹ A Text Encoding Initiative ajánlásának megfelelő XML jelölőnyelvi átiratokat készítenek, lehetővé teszik a szövegváltozatok párhuzamos olvasását. A szövegeltérés tényéről a HTML alapú kiadásokhoz hasonlóan, a lábjegyzet elektronikus megfelelője, egy felugró ablak ad tájékoztatást. Az autográf kéziratok jelölőnyelvi átirataiban az olyan filológiai jelenségek, mint a szerző törlése, vagy javítása szintén hasonló módon van jelölve.

⁷⁸ SZÉNÁSI, „Digitális átalakítás. A tudományos szövegkiadás a weben.”, 105.

⁷⁹ HORVÁTH Iván és FUCHS Anna, szerk., *József Attila összes tanulmánya és cikke: Szövegek 1930–1937 v 1.0* (Eötvös Loránd Tudományegyetem, 2012), <http://textus.elte.hu/ja/>.

⁸⁰ SZÉNÁSI, „Digitális átalakítás. A tudományos szövegkiadás a weben.”, 105.

⁸¹ Arra vonatkozóan, hogy a kutatócsoport mikor publikálta első digitális kritikai kiadását, nem találtam információt. A kutatócsoport vezetője, Debreczeni Attila alábbi szövegében hivatkozik először a műhelyre: DEBRECZENI Attila, „Kritikai kiadás papíron és képernyőn”, in *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*, szerk. CZIFRA Mariann és SZILÁGYI Márton, Csokonai Könyvtár: Bibliotheca Studiorum Litterarium 55, 26–42 (Debrecen: Debreceni Egyetemi Kiadó, 2014). Az Internet Archive adatai sem nyújtanak további támpontot, az biztos, hogy 2011-ben a kutatócsoport honlapján még nem szerepelt az „Elektronikus kiadások” menüpont: <https://web.archive.org/web/20110120223338/http://irodalom.arts.unideb.hu/kutatas/textologia/>.

szeméi vezetője emiatt a csokonai
 Tsak most adott kezet egy szép Kis-Asszonnak.*
 Mátkája' portréját a' mellyén v
 Szája mosolygással,* szeme tűz
 Kardja brilliántos, strikkolt a' ruhája,
 Kellemes lejtőt jár kis ded paripája.

Jav. ebből: *Kis-Asszonyinak*

[2. ábra] Részlet Csokonai Vitéz Mihály [*Dorottya, vagyis a' dámák' diadalma a' Fárságon*] autográf kéziratának kiadásából.⁸²

A szolgáltatásban elérhető kiadások szerkezeti felépítése túlmutat a hagyományos nyomtatott kiadások megszokott felépítésén. A kutatócsoport a kiadásokhoz adattárakat készít, a kiadásokat ezekkel az adattárakkal összekapcsolja, így a személy- és helynevek, valamint a mitológiai alakok előfordulásai kereshetők.

A DigiPhil létrehozásának ötlete 2012-ben merült fel, a szolgáltatás az Irodalomtudományi Intézet és a Petőfi Irodalmi Múzeum együttműködésében jött létre. A DigiPhil kezdeti éveiben alakult ki az a szolgáltatás, amelyre ma DigiPhil 1.0-ként hivatkozunk. A DigiPhil 1.0 kifejezés egyaránt magába foglalja azokat a digitális filológiai gyakorlatokat, amelyeket a DigiPhil követett, valamint a szoftveres komponenseket, amelyekre a szolgáltatás épült. 2021-ben megszakadt a két intézmény közötti együttműködés, az Irodalomtudományi Intézet vált a DigiPhil kizárólagos fenntartójává. Szintén ebben az évben indult a Digitális Örökség Nemzeti Laboratórium, amelynek munkájában az Irodalomtudományi Intézet konzorciumi partnerként vesz részt. A projekt keretében vált lehetővé a szolgáltatás megújítása, a DigiPhil 2.0 kialakítása. A DigiPhil által közzétett első kiadás, *A Tett* című folyóirat annotált kiadása 2014-ben jelent meg a portálon. 2015-től elérhető a Labádi Gyergely XML specifikációjára épülő *Mikszáth Kálmán összes művei* 39.⁸³ és 42.⁸⁴ kötete. Csak néhány példát kiemelve, a projekt gondozásában jelennek meg a Kosztolányi Dezső kritikai kiadás-sorozat kötetei,⁸⁵ amelyek közül az *Édes Anna* és az *Aranysárkány* a nyomtatásban megjelent

⁸² Forrás: https://deba.unideb.hu/deba/csokonai_muvei/text.php?id=csokonai_vers_0672_k.

⁸³ HAJDU Péter, szerk., *Mikszáth Kálmán, Elbeszélések 13, 1888*, Mikszáth Kálmán összes művei 39 (Budapest: Argumentum, 2001).

⁸⁴ HAJDU Péter, szerk., *Mikszáth Kálmán, Elbeszélések 16, 1893–1897*, Mikszáth Kálmán összes művei 42 (Budapest: Argumentum, 2007).

⁸⁵ A DigiPhil oldalán jelenleg az alábbi kritikai kiadások informatizált változata érhető el: VERES András, szerk., *Kosztolányi Dezső. Édes Anna*, Kosztolányi Dezső Összes Művei. Kritikai kiadás (Pozsony: Kalligram, 2010); BENGI László és PARÁDI Andrea, szerk., *Kosztolányi Dezső. Aranysárkány*, Kosztolányi Dezső Összes Művei.

kötetekhez képest átdolgozott, javított kiadás. Ugyanígy, az Arany kritikai kiadás-sorozat egyik legújabb kötete, a *Kisebb költemények 3.*⁸⁶ informatizált kiadása javított kiadás.

Az elmúlt évtizedek nemzetközi gyakorlata hasonlóan vegyes képet mutat a digitális filológia területén, mint Magyarországon. Jelentős különbség azonban, hogy számos, nagy intézmények által támogatott kezdeményezést láthattunk az elmúlt évtizedekben. A kiadások, a kutatócsoportok, a digitális filológiai eszközök és a szolgáltatások száma és változatossága olyan nagy, hogy itt csak néhány, a disszertációban kiemelt szempontok alapján releváns példát említek. Az egyik legrégebbi kezdeményezés a Tübingeni Egyetemhez köthető, ahol a '60-as években kezdték el fejleszteni, majd 1978-ban tették elérhetővé a TuStep-et (Tübingen System of Text Processing Programs), ami különböző szövegfeldolgozó programok összessége.⁸⁷ A TuStep segítségével készült a '80-as években közzétett Hans Walter Gabler-féle híres *Ulysses*-kiadás is.⁸⁸ Ez a kezdeményezés azért kiemelkedő, sőt egyedülálló, mert a textológiai munka teljes spektrumát lefedi, a gyűjtőmunkától a szövegműveleteken át a publikálásig és a szöveganalízisig, az Egyetem azóta is folyamatosan fejleszti, és bár saját leíró szabványt fejlesztettek, az XML nyelvet implementálták a szoftverekbe.

Általánosságban azonban elmondható, hogy a digitális tudományos szövegkiadásokat támogató műhelyek a TEI XML alapú átiratokat preferálják. A Deutsches Text Archive (DTA), melynek keretében nem csak szövegek publikációjára van lehetőség, hanem a projekt TEI XML-ek elkészítéséhez fejlesztett ki eszközöket, melyek közül az egyik egy online elérhető űrlap a metaadatok rögzítésére,⁸⁹ melynek segítségével TEI XML fejléccet hozhatnak létre. A szövegek annotálására kifejlesztették a DTAoX-ot, amely az Oxygen XML Editor „WYSIWYG” („What You See is What You Get”) szerkesztőjéhez készült applikáció.⁹⁰ Ezzel a DTA belső, a TEI ajánlást specifikáló szabványát követve, programozói tudás nélkül

Kritikai kiadás (Pozsony: Kalligram, 2014); TÓTH-CZIFRA Júlia és VERES András, szerk., *Kosztolányi Dezső. Esti Kornél*, Kosztolányi Dezső Összes Művei. Kritikai kiadás (Pozsony: Kalligram, 2011); TAKÁCS László, szerk., *Kosztolányi Dezső. Nero, a véres költő*, Kosztolányi Dezső Összes Művei. Kritikai kiadás (Pozsony: Kalligram, 2011); BUCSICS Katalin, szerk., *Kosztolányi Dezső. Pacsirta*, Kosztolányi Dezső Összes Művei. Kritikai kiadás (Pozsony: Kalligram, 2013); GYÖREI Zsolt és LOVAS Borbála, szerk., *Kosztolányi Dezső. A szegény kisgyermek panaszzai*, Kosztolányi Dezső Összes Művei. Kritikai kiadás (Pozsony: Kalligram, 2014).

⁸⁶ S. VARGA Pál, szerk., *Arany János, Kisebb költemények 3. (1860–1882)*, Arany János Munkái (Budapest: Universitas Kiadó–MTA BTK Irodalomtudományi Intézet, 2019).

⁸⁷ Kuno SCHÄLKLE és Wilhelm OTT, *TUSTEP. Tübinger System von Textverarbeitungs-Programmen. Version 2023. Handbuch und Referenz* (Tübingen: pagina GmbH, 2023), <https://www.tustep.uni-tuebingen.de/pdf/handbuch.pdf>.

⁸⁸ GABLER, „A kiadói szöveg születése: számítógép bába-szerepben”.

⁸⁹ A *teiHeader*-t automatikusan előállító űrlap elérhetősége: <https://www.deutschestextarchiv.de/dtae/submit/clarin>.

⁹⁰ Susanne HAAF, Alexander GEYKEN és Frank WIEGAND, „The DTA “Base Format”: A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources”, *Journal of the Text Encoding Initiative [Online]* 2014–2015, 8. sz. (é. n.), <https://journals.openedition.org/jtei/1114>.

hozhatunk létre XML-ben kódolt szövegeket. Egy másik kiemelkedő projekt a TextGrid, mely 2006 óta fejleszt úgynevezett WYSIWYM („What You See Is What You Mean”) kutatói infrastruktúrát, többek között a szövegek jelölőnyelvi átíratának elkészítésére, az átíratok repozitóriumban való publikálására. A publikálásnak, szemben a DTA gyakorlatával, nem része az XML átíratok vizualizációja, tehát a felhasználók a nyers XML átíratokat érik el, viszont kifinomult keresőfelületet biztosítanak, amely egyszerre képes keresni a metaadatokban és a teljes szövegben. Ezekben a szolgáltatásokban közös, hogy nem csak kritikai kiadások publikálását teszik lehetővé, hanem általánosságban támogatják a jelölőnyelvi átíratban való szövegkiadást és a DigiPhil gyakorlatának mintáit alkotják.

Ezeken a nagy szolgáltatásokon kívül vannak olyan digitális filológiai műhelyek, amelyek csak egyetlen korszak vagy életmű jelölőnyelvi átíraton alapuló feldolgozására vállalkoznak, mint például az Antwerpen-i Egyetemen készülő Beckett Archive, amelyben Samuel Beckett műveinek digitális genetikai kiadásai érhetőek el, vagy a Bergen-i Egyetem által fenntartott Wittgenstein Archive,⁹¹ amely az 1990-es évektől kezdve a digitális textológia egyik vezető elméleti és gyakorlati központja, és amely elsőként alkalmazta a szövegkiadás minden elemében a szemantikus web technológiáját.⁹² E két projekt azért érdemel említést jelen kontextusban, mert a DigiPhil projekthez hasonló módszertani kérdések technikai megoldására tettek kísérletet. A Beckett Archive vezető kutatója, a Beckett-szakértő Dirk van Hulle 2016-ban a krakkói DH2016 konferencián a disszertáció vizualizációs vizsgálatai szempontjából meghatározó érvelést mutatott be Beckett-szövegek jelölőnyelvi korpuszára építve. A Wittgenstein Archive pedig, szintén a harmadik fejezetben érintett kérdéskörben, a szemantikus web filológiai gyakorlatának módszertana szempontjából jelentett inspirációt a disszertáció érvelésének, és a DigiPhil technológiai gyakorlatának kialakításában.

⁹¹ The Wittgenstein Archives at the University of Bergen: <https://wab.uib.no/>.

⁹² Alois PICHLER és Amélie ZÖLLNER-WEBER, „Sharing and Debating Wittgenstein by Using an Ontology”, *Literary and Linguistic Computing* 28, 4. sz. (2013): 700–707, <https://doi.org/10.1093/lc/fqt049>.

Átmenet az analógból a digitális filológiába

A genetikus kritika megvalósulása papíron és a digitális térben: Madách Imre: Az ember tragédiája – Szinoptikus kritikai kiadás⁹³

A bevezetőben megfogalmazott állítás, miszerint a filológia, a „stabil” szövegtörzsek előállításával formálja, hogy mi válik a kutatás tárgyává, egybecseng Cerquiglini szavaival, a szerző így forgalmaz elhíresült könyvében, *A variáns dicséretében*: „[a] filológia ezért félelmetes: a jámborság álarca alatt a tényeket manipulálja.”⁹⁴ A metafilológia erre a manipulatív gesztusra reflektál: az egyes filológiai iskolák, korszakok szövegfogalma és gyakorlati szövegműveletei közötti különbségeket mutatja be és azokat értelmezi. Jelen alfejezet fókuszában egy (analóg és digitális térben is megvalósuló) kiadási gyakorlat, illetve az amögött álló filológiai szemlélet áll.

A genetikus kritika a mű keletkezési folyamatának rögzítésére törekszik. Bár a genetikus elvek szerint létrehozhatóak nyomtatott kiadások – erre később számos példát mutatok be –, azáltal, hogy az egyes szövegforrások közzétételére a digitális tér számos új megoldást biztosít, a kiadások, illetve a filológusok előtt egészen új lehetőségek nyílnak meg.

Bernard Cerquiglini 1989-ben megjelent könyvében, miután aprólékosan végigvezeti az olvasót a genetikus kritika és a francia medievisztika kérdésein, arra a felismerésre jut, hogy a kiadási gyakorlat nem képes kielégítően prezentálni a francia középkori szövegek varianciáját. Hiába a jó szándék, sem a kéziratok egymás melletti mutatása, sem a művek felaprózása és többkötetes kiadása, sem a faksimile csatolása nem oldja meg a könyv kétdimenziós teréből fakadó hátrányokat. Képtelenség minden fellelhető variánst,⁹⁵ illetve kommentárt egy, vagy akár több könyvbe belesűríteni. Cerquiglini javaslata szerint a megoldás egészen új megközelítésmódot igényel, ő már ’89-ben méltatta az informatikai és a számítógép által nyújtott lehetőségeket a szövegkiadásban:

⁹³ Jelen alfejezet korábbi változata 2021-ben jelent meg a *Helikon* folyóirat *Genetikus kritika* című számában: FELLEGI, „Átmenet az analógból a digitális filológiába: médiumváltás?. Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”.

⁹⁴ CERQUIGLINI Bernard, „A variáns dicsérete”, in *Metafilológia 1. Szöveg – variáns – kommentár*, szerk. DÉRI Balázs és mtsai., ford. KESZEG Anna, 219-295. (Budapest: Ráció Kiadó, 2011), 263.

⁹⁵ A *szövegvariáns, szövegváltozat, szövegállapot* fogalmak használata és értelmezése a genetikus kritika képviselői között is komoly változásokon ment keresztül az évtizedek során. A jelenséget aprólékosan dolgozta fel Tóth Réka: TÓTH Réka, *A szöveggenetika elmélete és gyakorlata*, Csokonai könyvtár 52 (Debrecen: Debreceni Egyetemi Kiadó, 2012). Jelen alfejezetben, az idézett szövegek kapcsán nincs lehetőség ezen különbségek kibontására, a fogalmakat minden esetben úgy értem, ahogyan az általam idézett szerző.

A számítógép, amennyiben egy óriási emlékezetnek, gyors felhasználásának és sokszoros vizualizációjának artikulációja, éppen azoknak a kiadói szükségleteknek felel meg, melyeket feltártunk. Használata – a nyomtatásra, annak következményeire és kötöttségeire való hivatkozás nélkül – lehetővé teszi, hogy a középkori mű kéziratának összességéhez, sőt az empirikus, lachmannista, bédier-ista kiadásokhoz is, melyeknek tárgyát képezték, hozzáférjünk. Sőt mi több, a járulékos információk nagy része – melyeknek virtuálisaknak kell lenniük, hogy az olvasást ne zavarják, de mégis feltalálhatók legyenek – ilyenformán szintén megadható: konkordanciák, előfordulási gyakoriságot mérő listák, rímtáblázatok, mindenféle számítások, kodikológiai és paleográfiai információk stb.: mindaz, amit a nyomtatott kiadás általában elhagy vagy amelyek közül nagy fájdalommal választ, a hipertudományos kiadás pedig az olvashatatlanságig erőltet.⁹⁶

Nemcsak szövegkiadási szempontból tartja kiemelkedőnek a számítógép szerepét a filológus számára, hanem sokakhoz hasonlóan⁹⁷ úgy véli, a számítógép megváltoztathatja a szövegfelfogásunkat: a számítógép képernyőjén egy szöveg utáni filológia körvonalai jelennek meg; csábító a lehetőség, hogy a szövegutániség eszközt arra használjuk fel, hogy képet adjunk arról, ami a szöveg előtt lehetett.⁹⁸

A számítógépes lehetőségek nemcsak a nyomtatott kiadás hiányosságait orvosolják, hanem a szöveg megértését is elősegítik, Cerquiglioni ebben látja a legnagyobb előnyüket: az eddig elrejtett, illetve sugalmazott kapcsolatokat a számítógép élénk tárja. A szövegével szinte egy időben megjelent *Mikroinformatika és textológia* című tanulmányában⁹⁹ Andrew Oliver már sokkal óvatosabban fogalmaz a számítógép és filológia kapcsolatáról:

Folynak ugyan az elektronikus szöveg felhasználására irányuló kutatások, s egyesek már jó ideje magasztalják a szövegvariánsok számítógépes tárolásával

⁹⁶ CERQUIGLINI, „A variáns dicsérete”, 295. Eredeti megjelenés: Bernard CERQUIGLINI, *Eloge de la variante. Historie critique de la philologie*, Des Travaux (Paris: Seuil, 1989).

⁹⁷ Lásd ehhez: MCGANN, „Az Ulysses mint posztmodern szöveg: a Gabler-féle kiadás”, valamint Almuth GRÉSILLON, „Irodalmi kéziratok a technikai sokszorosíthatóság korában”, in *Metafilológia I. Szöveg – variáns – kommentár*, szerk. Balázs DÉRI és mtsai., ford. Anna KESZEG, 338-359. (Budapest: Ráció Kiadó, 2011).

⁹⁸ CERQUIGLINI, „A variáns dicsérete”, 297.

⁹⁹ Eredeti megjelenés: Andrew OLIVER, „A la recherche du texte: micro-informatique et manuscrits”, in *La naissance du texte.*, szerk. Louis HAY, 41–50 (Paris: José Corti, 1989). Fordítása még abban az évben a *Helikonban*: OLIVER, „Mikroinformatika és textológia”.

feltároló lehetőségeket, szembeállítva őket a béklyónak tekintett hagyományos módszerekkel. Egyik szószólójuk az amerikai Todd Bender, aki több ízben is pálcát tört a szövegváltozatok rangsorolása fölött, ő ezzel ellentétben olyan, változtatható elektronikus szövegről álmodik, amelyben a különböző lehetséges változókat egy billentyűnyomással be lehetne helyettesíteni a „kétes” eset helyébe.¹⁰⁰

Oliver kételkedik abban, hogy a számítógép használata a kiadások elkészítésében alapjaiban változtatná meg a textológus feladatát, ugyanakkor azt elismeri, hogy

[a] számítógép, a kutatás értékes eszköze, igazi, intellektuális szerepbe helyezte vissza a textológust. Azzal, hogy megkönnyíti a feladatát és szigorú önfegyelemre szorítja – hiszen a számítógép nem tűr pontatlanságot –, a gondolkodás szabadságával ajándékozza meg [...] a számítógép a textológus számára azzá válik, ami az író számára a tinta és a papír volt: egyfajta fegyelmezettség és esztétika kifejezőeszközevé.¹⁰¹

Ez egybecseng Hans Walter Gabler erőteljes kijelentésével a híres Ulysess-kiadás kapcsán: „A számítástechnika alkalmazása nélkül esztelen vállalkozás lett volna ez, ugyanis menthetetlenül a hordszerkezet újabb torzulásával járt volna. A számítógép jóvoltából viszont az újjáépítés munkája nem fenyegetett az alapanyag minőségromlásával.”¹⁰² Bár kiadása már az első pillanattól a TuStep nevű szoftver segítségével készült, Gabler nem számolt azzal, hogy az általa szerkesztett kiadás megjelenhet digitális formátumban, mivel akkoriban még nem állt rendelkezésre a megfelelő technológia, azonban a kiadás készítése kapcsán már Gabler is felveti, hogy „elképzelhető lenne adatbankokat is hozzájuk [az apparátushoz] csatolni, amelyek biztosítanák az utalások folytonosságát”.¹⁰³

Mindössze tíz évvel később Almuth Gréssillon már nemcsak a filológusi munka fordulópontjának örvend és annak, hogy a variánsok egyidejű vizsgálata lehetővé vált, hanem úgy látja, a szöveg születését filmszerűen vizualizálva követhetjük majd a képernyőn. Ez a

¹⁰⁰ OLIVER, „Mikroinformatika és textológia”, 413.

¹⁰¹ Uo., 420.

¹⁰² GABLER, „A kiadói szöveg születése: számítógép bába-szerepben”.

¹⁰³ Uo., 428.

gondolat jelenik meg később Edward Vanhoutte 2010-es tanulmányában is, amelyben Aarseth modelljét¹⁰⁴ adaptálva elemzi az elektronikus kiadásokat.¹⁰⁵ Vanhoutte, az átmenetiség kategóriája kapcsán úgy fogalmaz, hogy

Ha a felhasználói idő pusztán előre haladása révén jelennek meg a szkriptonok [scriptons], a kiadás múló [transient]; ha nem, akkor állandó [intransient]. A legtöbb, szinte az összes kiadás állandó, semmi sem történik a felhasználó beavatkozása nélkül. Mindenesetre elképzelhetünk egy lejátszás-módot, amely a kiadás tartalmait úgy mutatja be a felhasználónak, mint egy filmfelvételt.¹⁰⁶

Elena Pierazzo Vanhoutte modellje kapcsán hívja fel a figyelmet arra, hogy az átmenetiség funkciója kulcsfontosságú a genetikus kiadások kapcsán, ez a funkció világítja meg, hogy miért érdemes digitális genetikus kiadásokat készíteni.¹⁰⁷ A digitális kiadások lehetővé teszik, hogy a kézirat időbeli rétegeit elkülönítsük és rögzítsük az átírás során,¹⁰⁸ így elméletben létrehozhatók ilyen filmszerű kiadások. Maga Pierazzo is részt vett egy olyan kísérleti kiadás készítésében,¹⁰⁹ amely, merítve a számítógépes játékok eszköztárából, Proust jegyzetfüzetét a felhasználó interakcióit figyelembe vevő módon próbálta ábrázolni. Az elképzelés szerint így, anélkül, hogy feladnák a kutatói szemléletet, a kézirat felfedezése élvezetesebbé válik a felhasználó számára, a kiadás szélesebb közönséget érhet el.¹¹⁰

A kézirat születésének hasonló ábrázolására azóta is voltak törekvések, azzal a különbséggel, hogy már eleve számítógépen született kéziratokról készültek szimulációk. Ennek tudományos értéke nem magától értetődő, azonban a szélesebb közönség számára mindenképp érdekes lehet. Mi sem jobb példa erre, mint a Petőfi Irodalmi Múzeum A

¹⁰⁴ Espen J. AARSETH, „Nem-linearitás és irodalomelmélet”, ford. MÜLLNER András, *Helikon* 50 (2004): 313–348, 323.

¹⁰⁵ Edward VANHOUTTE, „Defining Electronic Editions: A Historical and Functional Perspective”, in *Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviours, Products and Institutions.*, szerk. Willard MCCARTY, 119–144 (Cambridge: Open Book Publishers, 2010), 142.

¹⁰⁶ Uo., 143. A *Helikon*-ban megjelent, Müllner András által készített terminológia-fordítást módosítottam.

¹⁰⁷ Elena PIERAZZO, *Digital Scholarly Editing: Theories, Models and Methods* (New York: Routledge, 2016), 35.

¹⁰⁸ Elena PIERAZZO, „Digital Genetic Editions: The Encoding of Time in Manuscript Transcription”, in *Text Editing, Print and the Digital World*, szerk. Marilyn DEEGAN és Kathryn SUTHERLAND, 169–186 (Aldershot: Ashgate, 2009).

¹⁰⁹ Elena PIERAZZO et Julie ANDRÉ, „Le Codage en TEI des Brouillons de Proust: Vers l'Édition Numérique”, *Genesis* 36, 13 (2013): 155–161. A tanulmányban hivatkozott linken már nem érhető el a kiadás, új domain-re költözött, http://peterstokes.org/elena/proust_prototype/.

¹¹⁰ PIERAZZO, *Digital Scholarly Editing...*, 31.

*műfordítógép*¹¹¹ című kiállításában megtekinthető képernyőfelvételek, amelyeket az egyes műfordítók szövegszerkesztőben végzett munkájuk során rögzítettek. Ennél tovább ment az a kurrens kutatás,¹¹² amely az Huygens ING és az Antwerpeni Egyetem közreműködésében valósult meg. A kutatáshoz az Antwerpeni Egyetem kutatói által fejlesztett Inputlog nevű szoftvert használták,¹¹³ amely figyeli a szerző számítógépét írás közben és olyan adatokat képes rögzíteni, mint például a szerző visszaugrásai és javításai a szövegben, kijelölései, módosításai stb.; ezeket a cselekvéseket időbélyeggel látja el, így kiderülhet számunkra például, hogy a szerző mennyi idő alatt írt meg egy bekezdést, mennyi idő alatt olvasta át azt és az újonnan megírt szegmentum fényében módosított-e a korábbi szövegrészeken.

A 21. századból visszatekintve egyértelműen elmondható, hogy Cerquiglini felismerése, miszerint a képernyőszerű elrendezésben és a szinte korlátlan adat(kapcsolat) tárolásában látja a filológia és a szövegkiadás jövőjét, beigazolódni látszik.

Jóllehet a kutatók 30 éve hangsúlyozzák a digitális médium utolérhetetlen előnyeit a nyomtatott kiadásokhoz képest, ez sokáig mégsem vezetett a digitális kritikai kiadások általánossá válásához. Ennek több oka is lehet. A kutatóközösség részéről érzékelhető egyfajta bizonytalanság a digitális kiadások lezáratlansága miatt, ami ezeknek a kiadásoknak egyszerre az egyik legnagyobb előnye és hátránya, hiszen egy tanulmányban hogyan hivatkozhatunk egy olyan kiadásra, amely a szöveg megjelenése után bármikor megváltozhat? Megbízhatunk-e egy ilyen kiadásban? Az *illékonyság* problematikája a mai napig foglalkoztatja a szakembereket, pedig az elmúlt években a digitális filológia területén számos ajánlás született a probléma megoldására: Digital Object Identifier (DOI) azonosító használata, verziókövetés alkalmazása és az időbélyegek közzététele. Technikailag azonban nem triviális ezeknek az ajánlásoknak az alkalmazása, a verziókövetéshez például olyan repozitóriumszoftver szükséges, amely ezt támogatja, illetve előfordulhat, hogy egy kis költségvetésű filológiai műhelynek nincs forrása DOI azonosítók vásárlására. A számítógépes technológiák rendkívül gyors fejlődése is hátrálthat egy olyan lassú folyamatot, mint egy kritikai kiadás elkészítése. Számos CD ROM-on publikált kiadás vált használhatatlanná, sok esetben a régi adatbázisokat már nem lehet új környezetbe konvertálni, weblapok tűntek el. Másfelől reális az esély arra, hogy a kiadás készítésekor kiválasztott eszközök elavulnak a kiadás elkészülte után néhány évvel.

¹¹¹ *A Műfordítógép: A műfordítás műhelyei*, Petőfi Irodalmi Múzeum, Budapest, 2016. november 25. – 2017. május 28., <https://pim.hu/hu/kiallitas/muforditogep>.

¹¹² Lamyk BEKIUS, „Following the Writer through the Document: Encoding Contemporary Writing Processes in TEI-XML” (Textual Scholarship in the 21st Century. ESTS 2019, University of Málaga, 2019. 28), https://drive.google.com/file/d/1_Tb6xxA941OcProU09aIgf_eHG8h42zj/view.

¹¹³ Inputlog: <https://www.inputlog.net/>.

Más szempontból érdemes Elena Pierazzo gondolait is figyelembe venni, miszerint egy digitális kiadás készítése és publikálása komoly informatikai szaktudást igényel, amellyel az egyéni kutatók, kisebb filológiai műhelyek nem, vagy csak ritkán rendelkeznek.¹¹⁴ Ezeknek a komplex ismereteknek az elsajátítása évekre telik és még a mai napig ritka az átfogó digitális filológiai képzés. A másik, szintén fontos szempont, hogy általában nem áll rendelkezésre megfelelő anyagi támogatás szakemberek és informatikai rendszerek finanszírozására.

Jelen alfejezetben a DigiPhil szolgáltatás egy informatizált kiadását mutatom be, mely érzékelteti a digitális kiadásokban rejlő lehetőségeket, a jelölőnyelvi átiratok komplexitását. A következőkben látni fogjuk, ahogy egy genetikus textológiai elvek mentén készült kiadás esetén a digitális filológus a jelölőnyelvi átirat készítése során újraértelmezi a kiadás komponenseit, a filológiai jelenségeket és kiszabadítja a kiadást annak eredeti, kétdimenziós teréből.

A DigiPhil 2016-ban látott hozzá a Kerényi Ferenc által szerkesztett *Madách Imre Az ember tragédiája. Szinoptikus kritikai kiadás*¹¹⁵ informatizálásához. A kötet 2005-ben jelent meg nyomtatásban az Argumentum Kiadónál, ez a mű második kritikai kiadása és ezzel együtt az első magyar szinoptikus kiadás.¹¹⁶ A kötet mintájául a már idézett Hans Walter Gabler által szerkesztett *Ulysses* kritikai kiadás szolgált. Gabler kiadásának megjelenésétől eredeztethetjük a genetikus textológia megjelenését,¹¹⁷ amely épít a genetikus kritika eszközeire, azonban nem mond le a kritikai szöveg létrehozásáról. Mindkét kiadásban a kötet bal oldalán a szövegváltozatok olvashatók, a jobb oldalán pedig a szerkesztő által megállapított szöveg, azonban míg Gabler a szinoptikus szöveg mellett „olvasói szöveget” publikált, Kerényi ezt „megállapított szövegnek” nevezte el. Jerome J. McGann elemzésében kiemeli, hogy *Ulysses* „olvasói szövege”, szemben a szinoptikus szöveggel, „csak fakó, hűvös és sivár dokumentum – kiábrándítóan elvont, egyszerű és egydimenziós, míg a másik gazdag, összetett és sokrétű”, a szinoptikus átirat „örökre szét kellene [hogy] foszlassa az irodalmi művek stabilitásának illúzióját.”¹¹⁸ S. Varga Pál szerint¹¹⁹ azonban a megállapított szöveg Kerényinél éppen

¹¹⁴ Elena PIERAZZO, „What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter”, *International Journal of Digital Humanities* 1 (2019): 209–220.

¹¹⁵ MADÁCH Imre, *Az ember tragédiája. Drámai költemény. Szinoptikus kritikai kiadás.*, szerk. KERÉNYI Ferenc (Budapest: Argumentum, 2005).

¹¹⁶ Tóth Réka úgy véli, hogy a Gablertól származó *szinoptikus* kifejezés a „fentebb genetikus szövegnek nevezett szövegfajta megfelelője nála.” TÓTH, *A szöveggenetika...*, 91.

¹¹⁷ Uo., 92.

¹¹⁸ MCGANN, „Az Ulysses mint posztmodern szöveg: a Gabler-féle kiadás”, 432.

¹¹⁹ S. VARGA Pál, „Textológiaiak között: Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”, *Holmi* 18, 3. sz. (2006): 410–421. A kritikai kiadásról viszonylag kevés recenzió született az elmúlt tizenöt évben, ezek közül Fried István és S. Varga Pál elemzései a legalaposabbak. Szilágyi Márton az okokat kutatva így fogalmaz:

ellentétesen működik, a szinoptikus szöveg valójában a megállapított szöveg háttéréül szolgál, amely végeredményében sokkal közelebb áll a klasszikus kritikai szövegkiadáshoz, mint Gabler „olvasói szövegéhez”. Így Kerényi, bár Gablert tünteti fel mintaként, az ő módszereit követi, S. Varga szerint kivételesen magas színvonalon, mégis olyan hibrid kiadást hozott létre, amely ötvözi a klasszikus kritikai kiadás és a genetikus kritikai kiadás jellemzőit.

Az ember tragédiája kiadástörténetét sajátos módon a folyamatos szövegromlás jellemzi, hosszú ideig nem állt rendelkezésre megbízható szöveg. Kerényi maga is amellet érvel, hogy a megállapított szöveg elkészítése feltétlenül szükséges, „mert a krk. az alapszövege a népszerű, iskolai, szemelvényes kiadásoknak, a színházi előadásoknak.”¹²⁰ A szinoptikus szövegközlés emellett teret enged az olvasó „fantáziájának”, lehetőséget ad az eredeti Madách-kézirat rekonstruálására és a szövegfejlődés kibontására. S. Varga részletekbe menő elemzésében, bár méltatja Kerényi vállalkozását, erősen bírálja a szerkesztési elveit, miszerint Kerényi ragaszkodik az *ultima manus* elvének érvényesítéséhez.

Madách esetében ez különösen égető kérdés, hiszen a „helyes” szöveg megállapításának szándéka nem Kerényinél jelent meg először. S. Varga és Fried István is¹²¹ rámutatnak arra, hogy Madách, miután az egyetlen kéziratot eljuttatta Aranyhoz, a javítási javaslatokról Arany leveleiből értesült, emlékezetből hagyta jóvá azokat. S. Varga több olyan pontot is idéz, ahol az olvasónak vitatkoznia kell a „megállapított szöveggel”, egyik példájában rámutat arra, hogy Arany félreértette Madách szövegét, így javítása szövegromlást eredményezett. A második kiadás előkészítésekor Madách szintén nem látta a javított kéziratot, saját javításait az első kiadás egy példányán végezte el, így Arany javításait nem revideálhatta, S. Varga szerint ez Kerényi feladata lett volna. Nem véletlen tehát, hogy nem mindenki értett egyet Madách

„A tanulmányként való interpretálása valamely textológiai kérdésnek ugyanis alapvetően tér el attól, ahogyan ezt a filológus a számára hagyományosan adott apparátussal, a textológiai értelmezés szövegtípusaival megpróbálja kezelni. Ez abban is megmutatkozik, hogy az utóbbi éveknek azok a jelentős textológiai kísérletei (mint például Kerényi Ferenc szinoptikus Madách-kiadása vagy Orbán László Kazinczy-edíciója), amelyeknek újításai saját innovatív, olykor kifejezetten szubverzív módon alkalmazott textológiai szövegtípusaikban valósultak meg, jóval kevesebb szakirodalmi figyelmet kaptak, pontosabban a kritikai reflexiók teljes mértékben a textológusok világán belül maradtak. Annak ellenére is, hogy mindkét vállalkozás olyan vitára és továbbgondolásra érdemes kötet volt, amelyről a nem textológus szakembereknek is lehetett volna mondandója. Mivel azonban a markáns szövegértelmezések nem tanulmány formájában mutatkoztak meg, az ő számukra nem bizonyult észrevehetőnek a filológia elmélete szempontjából is lényeges újdonság.” Márton SZILÁGYI, „Textológia, filológia, értelmezés”, in *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*, szerk. Mariann CZIFRA és Márton SZILÁGYI, Csokonai Könyvtár: Bibliotheca Studiorum Litterarium 55, 15–25 (Debrecen: Debreceni Egyetemi Kiadó, 2014), 22.

¹²⁰ KERÉNYI Ferenc, „Előszó”, in *Az ember tragédiája. Drámai költemény. Szinoptikus kritikai kiadás.*, 5–9 (Budapest: Argumentum, 2005), 8.

¹²¹ S. VARGA, „Textológiák között: Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”, 411. illetve FRIED István, „A Madách-kutatás hétköznapi ünnepei (Kerényi Ferenc Madách-köteteiről)”, *Forrás* 39, 2. sz. (2007): 81–89, 84.

„döntésével”, mely szerint Arany javításait „vakon” elfogadta, és szorgalmazták az „eredeti” szöveg helyreállítását, amely nem tartalmazza Arany János javításait. A kritikai kiadás ezért is komoly jelentőségű, mert bár Kerényi egyes szerkesztési döntései kétségbe vonhatók, az írásszakértői vizsgálat, amelyet Wohlrab József végzett a kéziraton, egyértelműen bizonyítja, hogy Arany javításainak mindössze 2,97 százaléka vonatkozott Madách stílusára, és csupán négy olyan sort írt, amelynek nincs közvetlen szövegelőzménye.

A szinoptikus kritikai kiadásban Kerényi négy szövegállapotot különít el:¹²²

- (1) Madách kézírata (K),
- (2) ugyanezen a példányon Madách Imre és Arany János javításaival módosított szöveg (K1),
- (3) az első kiadás (61) és
- (4) a Madách Imre, valamint Szász Károly javaslatait is részben érvényesítő második, javított edíció (63).

S. Varga Pál szerint Kerényi ugyan megteremti a lehetőséget a szinoptikus olvasásra, mégis irányítja ezt azáltal, hogy Madách javításait nem a K, hanem a K1 részeként jelöli, azaz egy szinten jeleníti meg Arany javításaival és szembeállítja Madách saját kéziratával.¹²³ A szinoptikus szöveget Kerényi – Gabler nyomán – olyan jelrendszer segítségével kódolta, amely tömöríti (talán túlságosan is) a filológiai információkat. McGann, Gabler szinoptikus szövege kapcsán, ahogyan arra Palkó is felhívja a figyelmet,¹²⁴ éppen ebben a jelrendszerben és a szöveg „olvashatatlanságában” látja a kiadás erejét:

A Gabler-féle kiadás mellékjelei egy mesterséges nyelv nyelvtani szabályai, és nem jelenthetnek komoly nehézséget a fantázia szülte művek olvasóinak, minthogy azok minden esetben valamilyen mesterséges szabályrendszer alapján működnek. Mire megtanuljuk, hogyan „olvassuk” a szinoptikus szöveget úgy, ahogy az olvasói szöveget, már komoly lépéseket tettünk általában a szöveg természetének megismerése felé. Egy ilyen élmény örökre szét kellene foszlassa

¹²² KERÉNYI, „Előszó”, 8. A kézirat két szövegállapotként való elkülönítésének gondolata már Waldapfel Józsefnél megjelent.

¹²³ S. VARGA, „Textológiák között: Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”, 412.

¹²⁴ PALKÓ GÁBOR, „Digitális filológia: számítógép anyaszerepben”, *Filológiai közlöny* 61, 2. sz. (2015): 187–199, 194.

az irodalmi művek stabilitásának illúzióját, amelyet általában elfogadunk, mivel oly gyakran találkozunk velük az állandóság álarcában.¹²⁵

A szövegállapotok közti változások követése azonban nem egyszerű feladat, magas szintű koncentrációt és gyakorlatot igényel, és folyamatosan fennáll a tévedés és félreértelmezés veszélye, ugyanis a Kerényi (és korábban a Gabler) által kialakított kódrendszerben használt jelek sűrűsége és bonyolultsága miatt nehéz a jelölni kívánt eltérések dekódolása.

A csillagok védszellemei → CSILLAGOK VÉDSZELLEMEI kü-
lö<m>Nböző nagyságú, szí→inű, egyes, || ket- ||| tes | csillaggöm-
bök(kel)et, üstökösöket és ködcsillagokat gör- || getve rohan- ||| nak
| el a trón előtt. S<v>ZFérák ° zenéje halkan.⁶

<Angyalok> kara⁷ → ANGYALOK KARA.

Milyen → Mílyen →63 Milyen büszke láng-golyó jó
Önfényében elbizottan,
S egy szerény <csillag csoportnak> csillagcsoportnak
Ép ő szolgál öntudatlan. –

[25] Pislog e parányi csillag → csillag,
Azt hinnéd, <hogyan dőre szikra>, EGY <ÉGI> <MESSZI>
GYÖNGE LÁMPA, °

'S → S mégis millió teremtés
Mérhetetlen nagy világa. °
Két golyó küzd egymás ellen

[30] <Szét szakadni, ö(szv)szsze esni> ÖSSZEHULLNI,
SZÉTSIETNI: ° 2

[3. ábra] Kerényi Ferenc szinoptikus átírata (részlet).

¹²⁵ MCGANN, „Az Ulysses mint posztmodern szöveg: a Gabler-féle kiadás”, 439.

THE ^(A)[^(A)⟨PROCURESS⟩ Ø^(A)] BAWD^(A)

Jewman's melt!

BLOOM

(*in an oatmeal sporting suit, a sprig of woodbine in the lapel, ^(A) tony^(A) buff*
 5 *shirt, shepherd's plaid^(O) ^(A)⟨tie⟩ Saint Andrew's cross scarf^(A), white spats,*
fawn dustcoat on his arm, tawny red brogues, fieldglasses in bandolier and
a grey billycock hat) Do you remember a long long ⁽⁵⁾[time] time,⁽⁵⁾ ^(A) years
 and^(O) years^(A) ⁽⁵⁾[ago] ago,⁽⁵⁾ just after Milly,^(O) ^(A)⟨marionette⟩ Marionette we
 10 *called her,^(C) was weaned when we all went together to Fairyhouse races, was*
it?

MRS BREEN

^(A)(*in smart Saxe tailormade, white velours hat and^(O) spider veil*)^(A)
 Leopardstown.

[4. ábra] Hans Walter Gabler szinoptikus átirata (részlet).

Erre a felismerésre jutott Debreczeni Attila is: bár a posztmodern filológiai elméletek „meggyőző erővel pozicionáltak újra a klasszikus és az azt reflektáltabbá tévő modern textológia megközelítésmódjait és szövegfelfogását”,¹²⁶ azonban ezeknek az elméleteknek a gyakorlati megvalósulása nem alkalmas az olvasásra, csak a kutatásra, azaz a genetikus kritika „nem képes olvasható szöveget létrehozni”. Erre nyújtana megoldást a könyv alkotta korlátok közül való kilépés, a digitális kiadás.

Amikor a DigiPhil csoport hozzálátott a kiadás informatizálásához, már komoly tapasztalattal rendelkezett szinoptikus kritikai kiadások közreadásában, hiszen a Kosztolányi-kutatócsoport által készített kiadások egy részét (*Aranysárkány, Édes Anna, Esti Kornél*) már publikálta. A szöveg jelölőnyelvi kódolása mégis komoly kihívást jelentett. A jelölőnyelvi átirat a TEI ajánlása szerint készült a korábbi kódolási gyakorlatnak megfelelően.

A DigiPhil kiadásai közül ez volt az első dráma, a digitális filológusoknak a specifikáció kialakításakor ki kellett választaniuk a TEI legfrissebb ajánlásából a drámakódolásban

¹²⁶ DEBRECZENI, „Kritikai kiadás papíron és képernyőn”, 27.

használatos jelölőket, valamint megtervezni ezek grafikus megjelenítését. A DigiPhil rendszerében egyetlen megjelenítő réteg alkotja meg a jelölőnyelvi átirat vizualizációját, így különös figyelmet kell fordítani a jelölők kiválasztására és azok megjelenítésére – ha egy jelölő megjelenítése megváltozik, az hatással van a korábbi kiadások vizualizációjára is.

A következő probléma a szövegállapotok elkülönítésének kérdése. A DigiPhil kiadásában, Kerényiével ellentétben nem négy szövegállapot került kódolásra, hanem csupán három. A jelölőnyelvi átiratban a K és a K1 egy szövegállapotnak felel meg, amelyen mindkét szerző autográf javításait jelöltük. Ennek oka, hogy a digitális kiadás szempontjából nem tűnt logikusnak a két szövegállapot elválasztása, mivel a javítások ugyanazon a kézíraton szerepelnek. Ha elválasztottuk volna a K1-et a K-tól, a szövegvizualizáció során azt az illúziót keltettük volna, hogy két külön kézirat reprezentációja látható. Ez a megoldás ráadásul feloldotta az S. Varga által is jelzett ellentmondást is.

Egy digitális kiadás készítésekor minden esetben az elsődleges feladat a digitális objektum megállapítása. A Madách-kiadásnak különleges szerkezete van, ugyanis a digitális kiadásban külön kellett összefogni a kötet bal oldalán elhelyezkedő szinoptikus és a jobb oldalán található megállapított szöveget. A DigiPhil csoport végül úgy határozott, hogy a szinoptikus és a megállapított szöveg esetén az egyes színeket tekinti egységnek, a jegyzetek esetében pedig az egyes alfejezeteket. Ez megoldotta az egyes objektumok állandó azonosítóinak kiosztását, valamint jelentősen növelte a szolgáltatás sebességét, korábban ugyanis azt tapasztaltuk, például az *Édes Anna* esetében, hogy amikor egy fájl tartalmazta a teljes szinoptikus szöveget, a jelölők mennyisége miatt a fájl mérete túl nagy volt ahhoz, hogy a felhasználó elvárásainak megfelelő sebességgel töltsön be a teljes kiadás. A jelölőnyelvi átiratban azok a metaadatok, amelyek a könyv címlapján találhatóak (szerző, cím, kiadó), minden digitális objektum fejlécébe bekerültek, a DigiPhil oldalán pedig egy külön menüpontban érhetők el.

A TEI egyértelmű ajánlást ad a szövegállapotok jelölésére.¹²⁷ A TEI fejlécében, amely a digitális objektumra vonatkozó metaadatokat tartalmazza, egy úgynevezett **witness** (listWit) listát kell megadni, ahol felsoroljuk és azonosítóval látjuk el az egyes szövegtanúkat, majd a főszövegben az **app** (apparatus entry) jelölővel egybefogva jelöljük a megfelelő szöveghelyen a lemmát¹²⁸ (ha van ilyen), illetve az egyes szövegvariánsokat, **lem** és **rdg** elemekkel, ahol

¹²⁷ 2010-ben készült egy részletesebb jelölőkészlet genetikus kiadások kódolására, ami kiegészíti a TEI ajánlását, azonban ez nem vált a TEI hivatalos ajánlásának részévé. Lou BURNARD, Fotis JANNIDIS, Gregor MIDDEL, Elena PIERAZZO and Malte REHBEIN, „An encoding model for genetic editions”, <https://tei-c.org/Vault/TC/tcw19.html>.

¹²⁸ A lemma a TEI ajánlásában az alapszöveget jelöli.

@wit attribútumban hivatkozunk a szövegatanú¹²⁹ azonosítójára. Mivel a szinoptikus átirat nem állapít meg alapszöveget, a **lem** jelölőt a jelölőnyelvi átiratban sem használtuk.

```
<head type="title">
  <app>
    <rdg wit="#K">
      <fw corresp="AJ" type="pageNum">1</fw>
      <hi rend="doubleunderline">
        <del corresp="AJ">1<hi rend="sup">ö</hi></del>
        <add type="correction" corresp="AJ">első</add>
        szín.</hi>
      </rdg>
      <rdg wit="#EK61 #MK63">ELSŐ SZÍN.</rdg>
    </app>
  </head>
  <stage>
    <p>A mennyekben.
      <app>
        <rdg wit="#K">Az úr</rdg>
        <rdg wit="#EK61 #MK63">AZ ÚR</rdg>
      </app>
      dicstől környezetten trónján.
      <app>
        <rdg wit="#K">Angyalok <milestone unit="p"/> serege</rdg>
        <rdg wit="#EK61">ANGYA-<milestone unit="p"/>LOK SEREGE</rdg>
        <rdg wit="#MK63">ANGYALOK SE-<milestone unit="p"/>REGE</rdg>
      </app>térden. A
      <app>
        <rdg wit="#K">négy főangyal</rdg>
        <rdg wit="#EK61 #MK63">NÉGY FŐANGYAL</rdg>
      </app> a trón mellett áll. <app>
        <rdg wit="#K #MK63">Nagy fényesség.</rdg>
        <rdg wit="#EK61">Nagy<milestone unit="p"/>fényesség.</rdg>
        <note type="critic">Arany nyomdai utasítása tollal a lapszegélyen: "pet." A kiadásokban a szerzői utasítás zárójelben.</note>
      </app></p>
    </stage>
```

[5. ábra] Az Első szín jelölőnyelvi átirata (részlet).

Bár a nem professzionális olvasó számára ezek a kódok nem hordoznak jelentést, összevetve Kerényi kiadásával, már ebben a formában is sokkal könnyebben követhető a szövegfejlődés, hiszen a kód jobban strukturálja az eltéréseket, mint a kiadásban alkalmazott jelek, a TEI kódban például tisztán látszanak a szerkesztői elválasztások (<**milestone unit="p"/>**

A kódolás során rögzíteni kellett, hogy a kéziraton az autográf javítás kitől származik. A törlés esetében például a szokásos **del** jelölőt olyan attribútummal láttuk el, amely a számítógép és az olvasó számára egyértelműen elkülöníti Arany és Madách törléseit: **@corresp="AJ"** és **@corresp="MI"**. Ugyanezt az attribútumot használtuk a javítások megkülönböztetésére. Mivel a kritikai kiadás nem jelölte a (4) szövegállapotban a Szász Károly, illetve a Madách által végrehajtott javításokat, a digitális kiadásban mi sem jelöltük, ezek külön, a szövegkritikai dokumentumokban található, amelyet összekötöttük az adott szöveghellyel. Madách a második kiadásra szánt javításait az első kiadás egy szerzői példányán

¹²⁹ A szövegatanú kifejezés alapja a TEI ajánlása szerinti 'witness' szó.

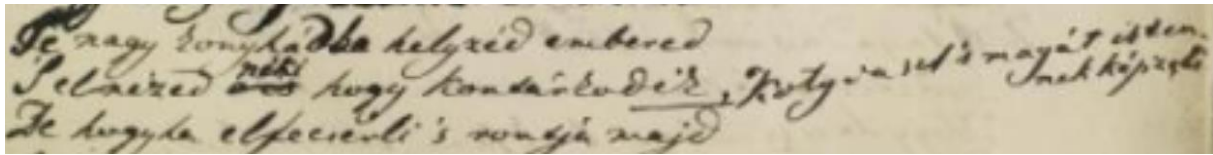
jelölte, amely elveszett, így nem lehet része a szinoptikus átiratnak, azonban ezen szöveghelyek szövegkritikai jegyzetekkel való összekötése nagyban megkönnyíti az olvasó számára a tájékozódást.

A korábban felsorolt jellemzőkön túl komoly kihívást jelentett a szöveg elrendezésének jelölése. A digitális filológia gyakorlatában az az általánosan elfogadott nézet, hogy a szövegelrendezésre vonatkozó információkat csak különösen indokolt esetben jelölnek a kiadások, ha az elrendezés jelentésszerű.¹³⁰ Ennek egyik fő oka, hogy a digitális médium lehetőséget teremt arra, hogy az olvasó különböző platformokon olvassa a szöveget, amelyek különbözőképpen jelennek meg asztali gépen, táblagépen vagy az okostelefon képernyőjén. Ennek megfelelően a TEI elemkészletében is csak néhány jelölőt találunk ezeknek a jelenségeknek a leírására. A Madách-kiadás esetében azonban voltak olyan jellemzők, amelyek jelölésétől mi sem tekinthetünk el. Ilyen például a prózai szerzői utasításban megjelenő elválasztás (mely gyakran eltért az első és második kiadás elválasztásától), az Arany által jelölt új kéziratoldal, az első és második kötet lapszámozása, amely csak néhány ritka esetben különbözött, valamint a szerkesztő által kiosztott sorszámozás jelölése. Az oldalszámozások esetében csak az adott szövegállapotban adtuk meg az oldalszámot, megjelölve azt, hogy kihez kapcsolódik: `<fw corresp="AJ" type="pageNum">`. Az `fw` („forme work”) jelölő,¹³¹ a TEI meghatározása szerint alkalmazható bármilyen főszövegtől elkülöníthető paratextusra (például sorszám, oldalszám, fejléc), amely az adott oldalon található, ezért a Kerényi által kiosztott sorszámokat is ezzel a jelölővel láttuk el.

A DigiPhil specifikációja szerint a sorszámokat automatikusan osztjuk ki és csupán az adott sor egyedi, állandó azonosítójaként használjuk, így biztosítva a kiadások pontos hivatkozhatóságát, azonban a Madách-kéziratban feltüntetett sorszámok számos esetben megtörik az egymásutániséget, például új sor beékelődése esetén:

¹³⁰ A DigiPhil kiadásai közül ilyen például Móricz Zsigmond világháborús naplójának kiadása, ahol a szerző többféle oldalszámozást is megadott a naplók szövegeihez, illetve a *Termelési-regény* informatizálása során a sortörések, amelyek Esterházy akaratának megfelelően, minden kiadásban pontosan ugyanott találhatók.

¹³¹ A TEI névhasználata utalás a nyomdászatban használatos zárókeretre, amelyben az oldalkép már rögzítésre került. Az *Oxford English Dictionary 2nd Edition* így határozza meg: „form, n. – 20. Printing. A body of type, secured in a chase, for printing at one impression. (Often spelt forme.)”



- 6 [90] Vegy<... n ...>konyhájában szintén megteszi. –
 <E> Te nagy konyhá<ban>dba helyezéd embered
 'S → elnézed <ő is> néki → néki, hogy kontárkodik,
 [92a] Kotyvaszt 's → s magát istennek képzele → képzele.
 De hogyha elfecsérli 's → s rontja majd
 A főztet, akkor gyúlsz késő haragra.

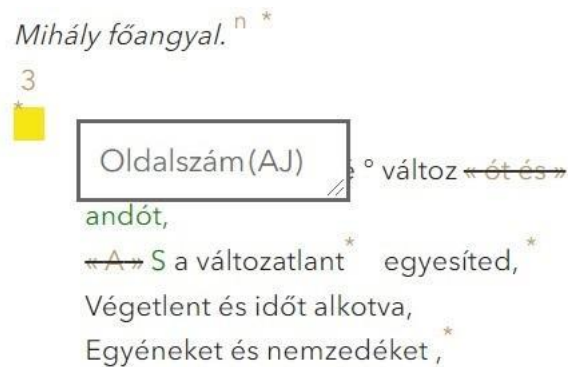
[6. ábra] Új sor beékelődése a kéziratban és annak átírata a kritikai kiadásban.

Ezért a DigiPhil csoport az automatikusan kiosztott sorszámok mellett nem tekinthetett el a szerkesztő által adott sorszámok megjelenítésétől, az automatikusan kiosztott azonosítók így helyenként eltérnek Kerényi sorszámaitól.

A digitális kiadás első változatában, amely a DigiPhil régi oldalán jelent meg, a szövegvariánsok külön ablakokban voltak megtekinthetők, a megfelelő szöveghelyeken kiemeléssel mutatva az eltéréseket, és az olvasó választhatta ki, hogy egyszerre hány szövegvariánst szeretett volna látni a képernyőn.



[7. ábra] A szövegvariánsokat tartalmazó szövegdobozok egymás mellett.



[8. ábra] Arany János oldalszámozása a kézíraton.



[9. ábra] Madách Imre törlése a kézíraton.

A DigiPhil digitális filológusai, mivel nem az adott korszak és életmű kutatói, szakértő filológus nélkül nem javítják a kritikai kiadás hibáit, de a lehetőség adott egy későbbi, javított kiadás publikálására. A jelenlegi kiadás esetén is nyíltak azonban olyan lehetőségek, amelyek jelentősen növelhetik annak kutatási értékét.

Bár a DigiPhil technikai infrastruktúrájának megújulását a disszertáció egy későbbi fejezetében tárgyalom részletesen, néhány jellegzetességet már jelen kiadás kapcsán is érdemesnek tartok bemutatni. A digitális kritikai kiadás ismérve a folyamatos javíthatóság, a Madách-kiadás esetében sincs ez másként. A kiadás első változatát a DigiPhil 1.0 felületén 2018-ban tette közzé. A DigiPhil 2.0-ra való áttérés részeként a kiadás jelölőnyelvi kódját a DigiPhil 2023-ban újra lektorálta, a korábban hiányzó részeket (pl. táblázatok) pótolta. Az előkészítés részeként megtörtént a személy- és földrajzi nevek névtérre fektetése, adatgazdagítása. A Kerényi-féle kiadás kritikusan kiemelik Kerényi pontatlanságát a

szakirodalmi és világirodalmi hivatkozásokat illetően.¹³² A DigiPhil a kötet bibliográfiai hivatkozásait feldolgozta és szemantikus adatbázisba rendezte, így jelentősen javult a kiadás tanulmányainak használhatósága, anélkül, hogy módosítottunk volna a szövegén. Bár Kerényi felhívja a figyelmet arra, hogy a Madách-filológia által eddig felfedezett, több mint negyven „rokonhangzású szövegelőzményt” nem szabad túlbecsülnünk, Madách-monográfiájában maga is bemutat néhány konkrét példát.¹³³ Izgalmas vállalkozás lenne ezeket külső annotációk formájában összekötni a digitális kritikai kiadás szövegével. Az Arany kritikai kiadássorozat informatizálásának köszönhetően lehetőségünk nyílt, hogy a kiadás jegyzeteiben hivatkozott leveleket összekössük a kiadással, így a kutatók egy gombnyomással elérhetik a teljes levélszöveget. A DigiPhil 2.0 megjelenítőjében a kiadás vizualizációja jelentősen eltér a korábbiaktól, az új megjelenítő két típusú nézetet tesz lehetővé. Az első esetben egy választott szövegváltozatot lehet tanulmányozni:

1

<<10>>Első szín.*

* A mennyekben. Az úr* dicstől környezetben trónján. Angyalok

serege* térden. A négy főangyal* a trón mellett áll. Nagy fényesség. ^c * *

Angyalok kara. ^c *

Dicsőség a* magasban istenemnek,* [↔]
 Dicsérje őt a* föld és a* nagy ég,
 Ki egy szavával hív életre mindent
 'S* pillantásától függ ismét a* vég.

[5]

Ó az erő, tudás, gyönyör egésze,
 Részünk csak az árny, mellyet ránk vetett,
 Imádjuk őt a* végetelen kegyért, hogy [↔]
 Fényében i<<|>>ly* osztályrészt engedett. [↔]
 <<Meg testesült>>Megtetesült* az örökös nagy eszme,

[10]

Ím a* teremtés befejezve áll,* [↔]
 'S* az úr* mindentől, mit leheni enged*
 Méltó adót szent számolyra vár.

Az <<ú>>Úr. ^c *

Be van fejezve* a* nagy <<mű>>mű,* igen <<,>>.*
 <<'S úgy ő<<szv>>ssze vág minden, hogy azt hiszem>>A gép forog, az alkotó<<ás>>pihen. ^c *

[10. ábra] A *Tragédia* a DigiPhil új megjelenítőjében.

¹³² S. Varga Pál és Fried István is számos konkrét példát emel ki.

¹³³ KERÉNYI Ferenc, *Madách Imre (1823–1864)* (Pozsony: Kalligram Kiadó, 2006), 160–163.

A második nézetben két választott szövegváltozatot lehet összehasonlítani:

<p><< ő>>Első szín.*</p> <p>* A mennyekben. Az úr* dicstől környezetten trónján. Angyalok serege* térdén. A négy főangyal* a trón mellett áll. Nagy fényesség. C * *</p> <p>Angyalok kara. C *</p> <p>Dicsőség a* magasban istenemnek,* Dicsérje őt a* föld és a* nagy ég, Kí egy szavával hív életre mindent S* pillantásától függ ismét a* vég.</p> <p>[5]</p> <p>Ó az erő, tudás, gyönyör egésze, Részünk csak az árny, melyet ránk vetett, Imádjuk őt a* végtelen kegyért, hogy Fényében i<< >>ly* osztályrészt engedett. <<Meg testesült>>Megtetesült* az örökös nagy eszme,</p> <p>[10]</p> <p>Ím a* teremtés befejezve áll,* S* az úr* mindentől, mit leheni enged* Méltó adót szent számolyra vár.</p> <p>Az <<ú>>úr. C *</p> <p>Be van fejezve* a* nagy <<mű>>mű,* igen <<:>>.* <<S úgy ő<<szv>>ssze vág minden, hogy azt hiszem>>A gép forog, az alkotó<<ás>>pihen. C *</p> <p>[15]</p> <p>Év-millióig <<szépen el forog>>eljár tengelyén,* * Mig egy kerékfogát ú* jítani kell. Fel hát* világim véd-nemtői, fel <<:>>.* Kezdjétek végtelen pályákat.</p>	<p>AZ EMBER TRAGÉDIÁJA. C *</p> <p>ELSŐ SZÍN.*</p> <p>(* A mennyekben. AZ ÚR* dicstől környezetten trónján. ANGYALOK SEREGE* térdén. A NÉGY FŐANGYAL* a trón mellett áll. Nagy fényesség.*)*</p> <p>ANGYALOK KARA.*</p> <p>Dicsőség a* magasban istenemnek,* Dicsérje őt a* föld és a* nagy ég, Kí egy szavával hív életre mindent S* pillantásától függ ismét a* vég.</p> <p>[5]</p> <p>Ó az erő, tudás, gyönyör egésze, Részünk csak az árny, melyet ránk vetett, Imádjuk őt a* végtelen kegyért, hogy Fényében ily* osztályrészt engedett. Megtetesült* az örökös nagy eszme,</p> <p>[10]</p> <p>Ím a* teremtés befejezve áll,* S* az Úr* mindentől, mit leheni enged,* Méltó adót szent számolyra vár.</p> <p>AZ ÚR.*</p> <p>Be van fejezve* a* nagy mű,* igen.* A gép forog, az alkotó pihen.*</p> <p>[15]</p> <p>Év-millióig eljár tengelyén,*</p>
---	--

[11. ábra] A *Tragédia* két szövegforrása párhuzamos megjelenítéssel.

A Madách-kiadás szempontjából azonban a „Minden szövegváltozat” opció lehet talán a legizgalmasabb, amely nézet az egyes szövegeltéréseket az adott szöveghelyen, egymást követően mutatja. Ez a nézet áll legközelebb a szinoptikus szövegközléshez és talán ez mutatja a legjobban a szövegfejlődés folyamatát.

1

<<1ő>>Első szín.* | ELSŐ SZÍN.*

* | (* A mennyekben. Az úr* | AZ ÚR* dicstól környezetben trónján. Angyalok

serege* | ANGYA-

LOK SEREGE* | ANGYALOK SE-

REGE* térden. A négy főangyal* | NÉGY FŐANGYAL* a trón mellett áll. Nagy fényesség.^c* | Nagy

fényesség.* * |)*

Angyalok kara.^c* | ANGYALOK KARA.*

Dicsőség a* | a* magasban istenemnek,* | istenünknek,*

Dicsérje őt a* | a* föld és a* | a* nagy ég,

Ki egy szavával hív életre mindent

'S* | S* pillantásától függ ismét a* | a* vég.

[5]

Ő az erő, tudás, gyönyör egésze,

Részünk csak az árny, mellyet ránk vetett,

Imádjuk őt a* | a* végetelen kegyért, hogy

Fényében i<<1>>ly* | ily* osztályrészt engedett.

<<Meg testesült>>Megtetesült* | Megtetesült* az örökös nagy eszme,

[10]

[12. ábra] A *Tragédia* „Minden szövegváltozat” nézetben.

A megállapított szöveg, ugyanúgy, mint a kiadás előző változatánál, külön jelölőnyelvi átiratokban érhető el, így, bár a kutatónak lehetősége nyílik a szinoptikus átirattal való összevetésre, a két átirat, a kutatói szándéktól függően külön-külön is vizsgálható. A kiadás szövegközléseinek ilyen jellegű elválasztása orvosolja az S. Varga által jelzett problémát, a szinoptikus átirat már nem csak háttérül szolgál a megállapított szövegnek, hanem azzal egyenrangú szövegközlésként jelenik meg.

Nyomtatott kritikai kiadások informatizálása

Az előzőekben bemutatott eljárás, egy szinoptikus kritikai kiadás szemantikus feldolgozása és adatgazdagítása, lassú folyamat, amely nagy odafigyelést és magas szintű digitális filológiai szakértelmet kíván meg. Mindeközben egyre nagyobb a nyomás, hogy a már megjelent kritikai kiadásokat informatizáljuk és a nagyközönség számára szabadon hozzáférhetővé tegyük. Az Irodalomtudományi Intézet kísérletet tett a nemzeti klasszikusok kritikai kiadásának közzétételére, textológiai portálján jelenleg 364 kötet digitalizált változata érhető el PDF formátumban. A kiadások szövegrétegeit automatikus karakterfelismeréssel állították elő, ezért az egyes kötetek szövegrétege eltérő minőségű.¹³⁴ A jelentős vállalkozás ellenére az lenne a kívánatos, hogy ha ezek a kritikai kiadások a Tragédiához hasonló feldolgozási folyamaton mennének át. Bár a DigiPhil folyamatosan képez digitális filológiai szakértőket, számuk máig csekély egy ekkora vállalkozás elvégzésére, a félautomatikus megoldásokra egyelőre csak kezdetleges kísérletekre van példa. A következőkben egy olyan projektet mutatok be, amelynek során olyan környezetet alakítottunk ki, amelynek segítségével a kritikai kiadások informatizálása jelentősen felgyorsítható, miközben a folyamathoz szükséges kompetencia radikálisan csökkenthető. A bemutatott projekt hatása kiemelkedő: megváltoztatta mind a kiadások informatizálásának, mind a *born digital* kritikai kiadások készítésének gyakorlatát.

Arany János kéziratainak és a szerző életművének kritikai kiadásainak közzététele

Arany János születésének 200. évfordulóján, 2017. március 2-án Nagyszalontán vette kezdetét az Arany János Emlékév, amelynek keretei között számtalan tudományos tanácskozás, kiállítás, kötetkiadás, egyéb, tudományos és népszerűsítő program valósult meg. Az emlékév adott alkalmat annak a nagyszabású digitalizálási projektnek a megkezdésére is, amely magába foglalta az Arany-kéziratok és az Arany János Összes Művei kritikai kiadások online publikálását a nagyközönség számára. Az Arany-kéziratok digitális fakszimiléit összegyűjtő és egységes platformon kereshetővé tevő szolgáltatás (Arany János kéziratainak repozitóriuma), valamint a kritikai kiadások jelölőnyelvi átírata a DigiPhil projekt koordinálásában valósult meg. Palkó Gábor, a fenti projektet és az annak végrehajtására fejlesztett kutatókörnyezetet *A digitális bölcsészet kultúrtechnikái* című tanulmányában helyezte elméleti horizontba.

¹³⁴ Forrás: <https://szovegtar.iti.mta.hu/>.

A Petőfi Irodalmi Múzeum és az MTA BTK Irodalomtudományi Intézet a DigiPhil projekt (digiphil.hu) keretei között vállalkozott az Arany János-levelezés kritikai kiadás öt nyomtatott kötetének digitális átírására és kutathatóvá tételére. A munkában részt vevő szerkesztők számára irodai szoftverkörnyezetet imitáló felhasználói felület készült. Egy ilyen felület nem enged bepillantást a fekete dobozba, a kötet jelölőnyelvi kódolását végző szerkesztő nem látja az ismerős/családias felület mögött azt a kódot és bonyolult munkamenetet, amelynek specifikációja digitális filológusok és Aranyszakértők együttműködésével jött létre. Latouri értelemben itt a cselekvők nem egyidejű együttműködésének példáját azonosíthatjuk, ahol a kutatás helyének, a digitális környezet egymáshoz kapcsolódó fekete dobozainak a kreatív megalkotása, mint De Certeau-i értelemben véve tudományos kutatótevékenység, illetve a környezetben belül folyó munka elválik egymástól. Ha – mintegy megakasztva a munkafolyamatot, a dobozfedelet felemelve – rápillantunk a kutatókörnyezet mint apparátus működésére, valóban emberi és nem emberi, közeli és távoli, múltbéli és jelenkori páratlanul bonyolult összjátékát látjuk magunk előtt.¹³⁵

A DigiPhil minden tevékenység típusában igyekszik a különmemű, egymástól időben/térben/személetben távoli humán és technikai aktorok együttműködésének, vagyis a Palkó által is idézett, latouri fekete doboznak az átlátható megszervezésére, reflektív feltárására.¹³⁶

A kritikai kiadások műhelyeiben a filológiai munkát előkészítendő és segítő, forrásjegyzékek, primer és másodlagos irodalmi bibliográfiák készülnek. Jóllehet az elmúlt évtizedekben ezek szerkesztése már mindenhol számítógéppel történik, a transzmedializáció valójában többnyire még várat magára: a papíron készült és a nyomtatásban megjelentetett (irodai szövegszerkesztő programmal készített) bibliográfia egyaránt a nyomtatott médium keretei között marad. Ezzel szemben a könyvtári rendszerek önmagukban, a szemantikus web

¹³⁵ PALKÓ Gábor, „A digitális bölcsészet kultúrtechnikái.”, in *A humán tudományok és a gépi intelligencia*, szerk. TOLCSVAI NAGY Gábor, 33–47 (Budapest: Gondolat Kiadó, 2018), 41.

¹³⁶ Bruno LATOUR, „A Collective of Humans and Nonhumans: Following Daedalus’s Labyrinth”, in *Pandora’s hope: essays on the reality of science studies*, 183 (Cambridge, Mass: Harvard University Press, 1999), 183. Idézi PALKÓ, „A digitális bölcsészet kultúrtechnikái.”, 139.

személetét is megelőzve, egy adathálózat elemeiként tekintettek rekordjaikra. Helyi névtereket építettek, amelyek publikálásával és egymáshoz kapcsolásával jöttek létre a ma digitális kulturális örökségének alapját képező névterek, mint amilyen például a VIAF.¹³⁷ A digitális filológia, szemben a nem szöveg típusú digitális objektumokra épülő archívumi gyakorlatokkal, nemcsak a szövegek egészét leíró metaadatokat felelteti meg névtérelmeknek, de a szövegszegmentumok szintjén is végez adatgazdagítást, vagyis hivatkozik a saját személynévterében és a máshol publikált névterekben fellelhető besorolási rekordokra. Ilyen jellegű adatgazdagítás nemcsak személy- és helynevek esetében történhet, hanem műfaji fogalmak esetén is érdemes publikált teauruszokra, ontológiákra hivatkozni. (A DigiPhil például, követve az Europeana példáját, kiadásában a műfaji fogalmakat a Getty Art & Architecture¹³⁸ névteréből azonosítja.)

Miért van szükség adatgazdagításra? Az egyik nyilvánvaló indok az *aggregáció*, amely a digitális kulturális örökség talán legradikálisabb és legelőíróbb – de biztosan a legtöbb heterogén aktort integráló – gyakorlata. Olyan nagy hatású intézmények propagálják, mint az Europeana,¹³⁹ amely jelenleg a világ legnagyobb volumenű és hatású aggregátora, közel 60 millió digitális objektummal. Ennek mintájára hozták létre például a Digital Public Library of America-t (DPLA), amely 48 millió digitális objektumot tartalmaz.¹⁴⁰ Az aggregációs tevékenység során a kisebb helyi gyűjtemények anyagait nemzeti és nemzetközi szolgáltatásokban összegyűjtik és – adatgazdagítás után – kereshetővé és megtekinthetővé teszik.

Az *aggregáció* azonban nem az egyetlen, talán nem is az elsődleges oka, hogy a digitális filológusok adatmegfeleltetést, szemantikus adatgazdagítást végeznek a digitális objektum, illetve a szövegszegmentumok szintjén is. Ezáltal ugyanis a szabad szavas keresésnél jóval pontosabb és kiterjedtebb kapcsolatrendszer jön létre a szövegek között, amelynek különösen a nem szépirodalmi szövegek (levelezés, naplók) esetén van kiemelt jelentősége. Egy adott kulturális közeg kapcsolati hálózatának mintázatait csak ilyen sűrűségű adatrendszerben érdemes ma vizsgálni. A DigiPhil két projektjében jött létre kulturális mintázatok feltárására alkalmas minőségi adathálózat, a Móricz-naplók és -levelek feldolgozása során, illetve az Arany János kritikai kiadás munkálatainak eredményeképpen. A következőben e munkálatok négy területét ismertetem, előbb a nyomtatott kritikai kiadások jelölőnyelvi kódolásának

¹³⁷ VIAF: The Virtual International Authority File, <http://viaf.org/>.

¹³⁸ Getty Art & Architecture Thesaurus: <https://www.getty.edu/research/tools/vocabularies/aat/>.

¹³⁹ Europeana Collections: <https://www.europeana.eu/portal/hu>.

¹⁴⁰ Digital Public Library of America, <https://dp.la/>.

munkamenetét, majd a kézirat repozitórium technikai fejlesztését és eredményét, végül a mesterséges intelligencia egy kísérleti alkalmazását a digitális filológia területén.

Az Arany János kritikai kiadás digitalizálása a különműveletű aktorok együttműködésének egy komplex példája. A korpusz terjedelme (mintegy tizenháromezer könyvoldal) és az Arany- emlékévként időkerete nem tette lehetővé, hogy a kódolás munkáját az Irodalomtudományi Intézetben (ITI) működő kutatócsoport Arany-szakértői végezzék, olyan külső munkatársak bevonására volt szükség, akiknek a többsége nem foglalkozott korábban szövegkiadásokkal (esetleg irodalomtudománnyal sem). A munkamenet kialakítása során tehát figyelembe kellett venni, hogy a transzmedializációba bevont munkatársak egy része nem rendelkezett filológiai gyakorlattal, a TEI XML használatát, illetve a DigiPhil munkamenetét és specifikációját nem ismerték. A DigiPhil számára kihívást jelentett azon feltételek megteremtése, amelyek révén úgy állhattak elő jelölőnyelvi átíratok, hogy az Intézet által bevont munkatársaknak¹⁴¹ nem kellett több hónapnyi digitális filológiai oktatáson részt vennie. Ennek érdekében a digitális bölcsész munkacsoport ergonomikus, egy irodai szoftver felhasználói felületét imitáló virtuális kutatókörnyezetet¹⁴² fejlesztett. A grafikus felület mögé „rejtették el” azt az apparátust, amely a voltaképpen jelölőnyelvi átíratot létrehozta.¹⁴³ Az a törekvés, hogy a kutatók elől „elrejtjük” a számítógépes kódokat, nem példa nélküli, a korábban már említett TextGrid¹⁴⁴ szolgáltatáson belül is elérhető olyan modul, amely támogatja a TEI XML-ek WYSIWYM („What You See Is What You Mean”) módon történő szerkesztését. Mivel a TEI séma jelenleg közel kétezer oldalt tesz ki, és számos filológiai jelenség kódolható többféleképpen, használata műhelyenként eltérő, a WYSIWYM szerkesztők pedig igyekeznek a teljes TEI elemkészletet lefedni, a használatukhoz elengedhetetlen a TEI alapos ismerete. A DigiPhil, ami akkor már hat éve működött, hosszú évek alatt fejlesztette ki azt a TEI XML specifikációt, amely az összes digitális kiadásban érvényesül, és ami magában foglalja a szövegjellemzők széles körének kódolási mechanizmusát. Ez a specifikáció biztosítja egyrészt a szolgáltatás filológiai és

¹⁴¹ A projekt teljes időtartama alatt az alábbi kollégák dolgoztak a kiadásokon: Csonki Árpád, Horváth-Márjánovics Diána, Hunyadi-Zoltán Emese, Káli Anita, Móre Tünde, Roskó Mira, Sárközi-Lindner Zsófia, Varga Szilvia, Vétek Bence. A DigiPhil csoport (Palkó Gábor, Fellegi Zsófia, Bobák Barbara) munkáját segítették még: Czinkos Zsolt és Metzger Réka (virtuális környezet fejlesztése); Kundráth Péter (megjelenítő fejlesztése); Vincze Tímea és Eőri Lilla (adatgyűjtés, adattisztítás).

¹⁴² A kifejezéshez lásd: PALKÓ, „A digitális bölcsészet kultúrtechnikái.”

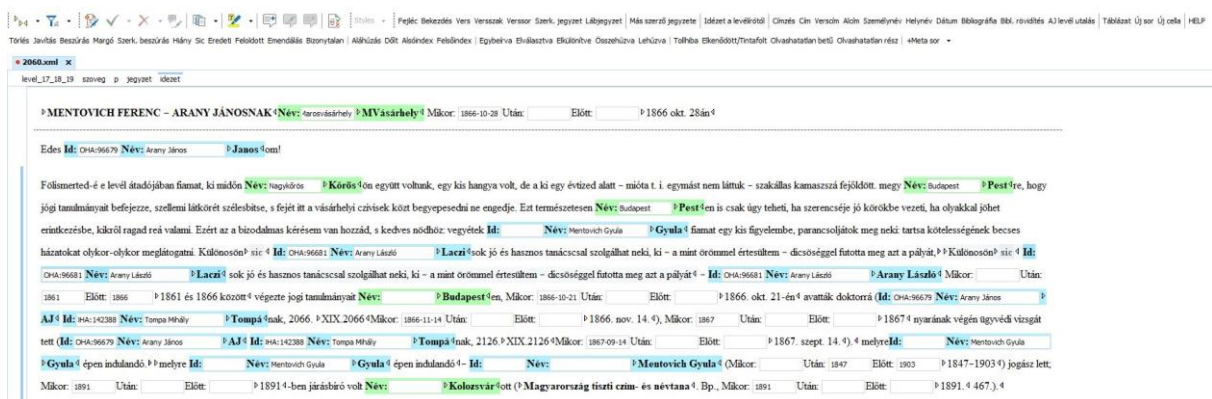
¹⁴³ A DigiPhil korábban, a Móricz Zsigmond-levelezéskiadás kapcsán már próbálkozott olyan leíró környezetet teremteni a Microsoft Office Word szoftverben, amely kiváltja a TEI kódolást, azonban erről a módszerről kiderült, hogy számos előre nem látható hibalehetőséget tartalmaz. A tapasztalatokról a részletes leírást lásd a következő fejezetben.

¹⁴⁴ TextGrid: Virtual Research Environment, <https://textgrid.de/>.

technikai egységét, illetve ez teszi lehetővé, hogy a DigiPhil összes kiadásánál ugyanazt a megjelenítő réteget használja, amely lefedi a különböző kiadástípusok eltérő jellemzőit. Ha a specifikációban rögzítettől eltérő módon használunk egy elemet, akkor az nem, vagy hibásan jelenik meg a DigiPhil oldalán. (Komoly kihívást jelent ugyanakkor, ha egy új kiadás miatt módosítunk a specifikáción, hiszen az hatással lehet a már kész kiadásokra is, illetve ügyelni kell a megjelenítő réteg globális jellegéből adódó hibákra is.)

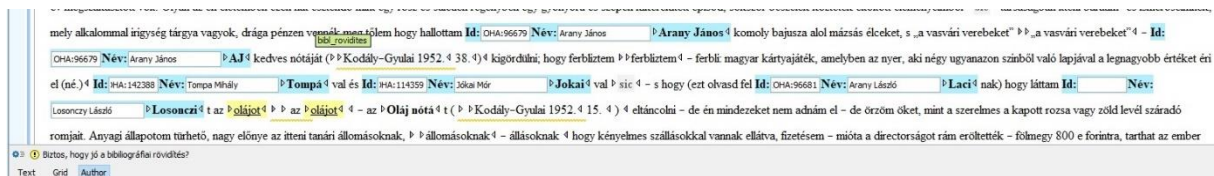
A DigiPhil a kutatókörnyezet implementálására a digitális filológusok körében az egyik legelterjedtebb szerkesztő eszközt, az Oxygen XML Editor-t használta, amely biztosítja azokat a technikai jellemzőket, amelyek elengedhetetlenek a munkakörnyezet gyors és stabil kialakításához. Az Oxygen eleve rendelkezik alapszintű beépített környezetekkel, úgynevezett keretrendszerekkel (framework), a TextGrid-hez hasonlóan a TEI-hez is biztosít egyszerű grafikus szerkesztői felületet, amely azonban nem alkalmas komplex kódok generálására felhasználóbarát módon. A DigiPhil ezért saját keretrendszert hozott létre a szerkesztőn belül, amelyben csak a DigiPhil által meghatározott specifikáció érhető el, és az előre definiált szabályok mentén ellenőrzi, hogy az elkészült fájlok megfelelőek-e. A grafikus réteg mögött lévő XML nem a TEI sémán alapult, hanem egy annál jóval egyszerűbb, magyar nyelvű jelölőket tartalmazó készleten. A munkamenet arra épült, hogy a kódolók a lehető legegyszerűbb felületen dolgozhassanak, a környezetben létrejövő kódot automatizált algoritmusok ellenőrizték szintaktikai, illetve, ahol ez lehetséges, szemantikai szinten is, majd ezen kódból automatikusan előállítható legyen a publikálásra szánt, a filológiai munka eredményeként értelmezett TEI XML. Erre azért volt szükség, mert egy TEI XML sémának megfelelő dokumentum – különösen annak metaadatokat tartalmazó része – bonyolult szintaxissal rendelkezik, számos olyan kötelező elemet és hierarchikus beágyazást tartalmaz, amelyek használata jelentősen lassította volna a kódolás folyamatát. Olyan köztes, leegyszerűsített kódolási sémát dolgoztak ki, amely automatikusan átfordítható a komplex TEI sémára. A specifikáció kialakításakor a DigiPhil először meghatározta a tizenkilenc kritikai kiadáson belül az egyes kiadástípusokat. Különböző elemkészlettel írható le ugyanis például egy drámafordítás, ahol jelölni kell a színpadi utasítás és a beszélő személyét, vagy a levelezés, ahol a feladót és a címzettet kell leírni. A DigiPhil az egyes kiadástípusokhoz külön környezetet fejlesztett, így a projekt folyamán összesen hét különálló keretrendszer jött létre (drámafordítások, lírai művek, prózai művek, hivatali iratok, levelezés I–II, levelezés III–V, apparátus és jegyzetek). A munka a hivatali iratok és a levelezéskötetek keretrendszereinek kidolgozásával kezdődött, ezeknek a felépítése ugyanis egyszerűbb, mint a szövegváltozatokat is jelölő köteteknek. A kódolók már készségszerűen használták a munkakörnyezetet, mire a

bonyolultabb jegyzetapparátussal és több szövegréteggel operáló kötetek leírását megkezdték. A levelezéskötetek esetén azért volt szükség két önálló környezet kialakítására, mert a II. és a III. kötet megjelenése között több, mint húsz év telt el, a kötetek szerkesztési elvei teljesen eltérők voltak, nem lehetett ugyanazzal az specifikációval információvesztés nélkül leírni őket. A kutatókörnyezet ergonómiáját, az irodai szoftverek grafikai felületéhez hasonlóan menüpontok és gombok teremtik meg: egy gomb vagy menü elem – akár a Microsoft Word makró funkciója – olyan műveletsort generál, ami a specifikációnak megfelelő kódrészletet hoz létre. Mindez láthatatlan marad a kódoló munkatárs számára, ugyanis a felhasználói felület grafikus formázott (programkódokat nem tartalmazó) szöveget jelenít meg, a kód jellemzőit grafikai jelekkel, színekkel, elrendezéssel vizualizálva. (Azt a kérdést, hogy az irodai szoftverek grafikus felületei milyen archívumi praxisok metaforái, Palkó, Cornelia Vismann híres kötetét idézve, elemzi, erre jelen fejezetben nem térek ki.)

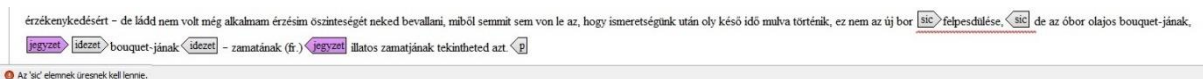


[13. ábra] A kutatókörnyezetben elérhető műveletek

A kutatókörnyezet a DigiPhil által előre meghatározott szintaktikai és szemantikai szabályok alapján folyamatosan ellenőrizte a munkatársak által bevitt adatokat, jelölte a kötelezően kitöltendő metaadatmezőket, az egyes jelölők egymásba ágyazhatóságát, a jelölők attribútumainak értékeit, illetve az egyes mezőkbe írt adatok (például a dátumok) formátumát. A leírás során, ha a rendszer kódolási hibát észlelt, jelezte azt a felhasználó számára hibaüzenet vagy figyelmeztetés formájában.

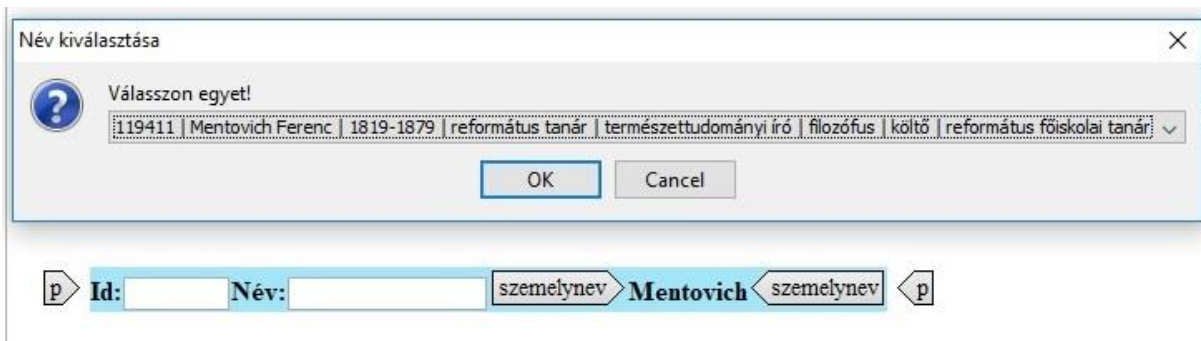


[14. ábra] A kutatókörnyezet üzenete szintaktikai hiba esetén.



[15. ábra] A kutatókörnyezet figyelmeztetése, amely a beírt érték ellenőrzésére szólítja fel a kutatót.

A komplex adathálózatok létrehozásának egyik legfőbb akadálya a munkamenet szempontjából, hogy az adatgazdagítás (nevek, bibliográfiai adatok azonosítása és az azonosítók kódba illesztése) radikálisan lassíthatja a kiadások szövegtárolási folyamatát. Ennek a problémának a megoldására félautomatikus névazonosító funkció került beépítésre a keretrendszerekbe.



[16. ábra] A félautomatikus névazonosító a környezetben külön ablakban mutatja a találati halmazt, amelyből a kutató választhatott

A kritikai kiadások névmutatói alapján a DigiPhil akkor használt névtérben¹⁴⁵ elérhető személynevek azonosítóit és a legfontosabb azonosítást segítő adatait integrálták a rendszerbe. A kódolás során a névazonosító algoritmus személynevek beírásakor azokat automatikusan megfeleltette a névtér elemeinek, a kódoló a rendszer által felajánlott lehetséges egyezések alapján választotta ki a megfelelő besorolási névrekordot, amelynek azonosítója automatikusan íródott be a kódba, ezzel is csökkentve a gépelési hibák esélyét. A kódolás során a bibliográfiai rövidítések jelölésre kerültek.

Egy 21. századi digitális filológiai szolgáltatásnak – a digitális médium *illékonyságával* szemben – meg kell felelnie az adatbiztonság legmagasabb elvárásainak. Bár a digitális bölcsészet (nem is olyan távoli) hőskorában Magyarországon is történt komoly adatvesztés,¹⁴⁶ a repozitóriumtechnológia és a verziókövető szolgáltatások általánossá válásával ennek esélye minimalizálható. A transzmedializáció során nagy számú digitális objektum jött létre,¹⁴⁷ amelyek archiválása és változáskövetése volt a munkamenet egyik legkiemeltebb feladata. Mindezt a DigiPhil két szinten kellett, hogy megoldja. Egyrészt el kellett kerülni, hogy a munkamenet ezen része a kódolók heterogén számítógépes környezetétől függjön, másrészt biztosítani kellett a kódolás első (kódolók, munkakörnyezet) és második (a DigiPhil digitális filológusainak feladatai) munkafázisának szétválasztását. A kódolók által kezelt keretrendszer lehetővé tette, hogy a kódokat tartalmazó fájlokat közvetlenül a munkakörnyezetből az erre a célra létrehozott verziókövető szolgáltatásba mentsek, így a munka során használt saját számítógépek esetleges biztonságtechnikai vagy egyéb hibái nem veszélyeztették az adatbiztonságot. A verziókövető rendszer követi a szerverre feltöltött kódok módosításait, így a helyi adatvesztés esetén visszaállíthatók a fájlok, a kódoló számítógépén ugyanis csak egy úgynevezett munkapéldány (working copy) található. Az eszköz csoportmunka esetén elvégzi a kódokon történt változtatások adminisztrációját. Mivel mindenki saját felhasználói profillal rendelkezik, elkerülhetők a kódolás során egymástól távol „együttműködő” aktorok által

¹⁴⁵ A DigiPhil a Magyar Életrajzi Index (Petőfi Irodalmi Múzeum) névtérét építette tovább saját integrált könyvtári rendszerében, amely a Koha szoftvert használta. Ez a gyakorlat később módosult, lásd a fejezet utolsó alfejezetében. A régi rendszer már nem elérhető.

¹⁴⁶ Az első hálózati betűhű forráskiadás, a Balassa-kódex átírata, faksimilékkel együtt 1993-ban jelent meg online, kísérleti jelleggel, majd 1995-ben CD-változatban is. 1996-ban már tanulmányokkal bővítve jelent meg ismét, azonban 1998-ban megsemmisült. TÓTH, „Online kritikai szövegkiadás Magyarországon az ezredfordulón”. A kiadás további történetét lásd: MARÓTHY, „Tudományos szövegkiadások a hálózaton”.

¹⁴⁷ Az egyes köteteknél eltérő, hogy egy digitális objektum, azaz egy TEI XML fájl a kötet mekkora egységét fedi le. A levelezéskötetknél például egy fájl egy levelet és a hozzá tartozó jegyzeteket tartalmazza, míg a drámafordításoknál egy fájl egy felvonást tartalmaz, a teljes darabra vonatkozó jegyzetek külön digitális objektumot képeznek. A projekt során összesen több, mint 4200 XML dokumentum jött létre.

módosított kódok okozta adatvesztés. A DigiPhil munkatársai a szerverre feltöltött kódok átalakítását egy attól független rendszerben archiválta.

A munkamenet függetlenítését a kódolók egyéni számítógépes környezetétől nemcsak a kódok verziókövetése és archiválása érdekében kellett elvégezni, hanem a munkakörnyezet fejlesztése és egységes használata érdekében is. A kódoló munkatársak az Oxygen XML Editor szoftveren keresztül, egy szerverre mutató link segítségével telepítették a keretrendszert, ami ezt követően minden újraindításkor ellenőrizte az elérhető csomag verziószámát, és amennyiben változtatás történt, automatikusan frissítette a rendszert. Ez lehetővé tette, hogy minden munkatárs – függetlenül attól, hogy milyen számítógép(ek)en dolgozott – mindig ugyanazt a kutatókörnyezetet használja, költséges és időigényes hardveres támogatás igénybevétele nélkül.

A projektbe bevont munkatársak a kutatókörnyezet segítségével közel másfél év alatt dolgozták fel az Arany kritikai kiadások köteteit. Ezt követően a DigiPhil munkatársai a létrejött fájlokból egyedi fejlesztésű stíluslapok segítségével előállították a publikálásra szánt TEI XML fájlokat, és feltöltötték lektorálásra a szolgáltatást biztosító repozitóriumba. A munkatársak az „éles” szolgáltatással mindenben megegyező tesztkörnyezetben ellenőrizték a kiadások szövegét, míg a digitális filológusok a TEI XML fájlokat vizsgálták át, a javításokra már a TEI XML fájlokban került sor.

A kötetek publikálásához szükség volt a specifikáció implementálására: a megjelenítő réteg fejlesztésére, hiszen a kritikai kiadássorozat kötetei több, mint fél évszázad (hatvanöt év) filológiai munkáját – és az azalatt történt szemléletváltozást – foglalják magukba. A szolgáltatás ünnepélyes bemutatására 2019 novemberében került sor az Irodalomtudományi Intézetben.¹⁴⁸

Bár a digitalizálási munka lezárult, a virtuális kutatókörnyezetek építése és alkalmazása folytatódik. A DigiPhil szakértői a projekt során szerzett tapasztalatokra építve olyan környezetet fejlesztenek, amelynek az alapját már nem egy leegyszerűsített köztes XML séma adja, hanem a TEI specifikációnak mindenben megfelelő, az adott kiadásra alkalmazott végleges séma. Ezzel lehetőség nyílik arra, hogy a digitális kritikai kiadások a munkafolyamat kezdőpontjától egységes és szabványos formátumban készüljenek, és publikálás után is folyamatosan javíthatók és bővíthetők legyenek a létrehozás folyamatában felhasznált keretrendszerrel azonos környezetben. A folyamatban lévő Arany kritikai kiadások már ebben a környezetben készülnek, az S. Varga Pál által szerkesztett *Arany János Kisebb Költemények*

¹⁴⁸ GÖNCZY Monika, „Az Arany János-bibliográfia” (Értékteremtő tudomány. Az Arany-gyűjtemény bemutatkozása, Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet, Budapest, 2019. 20).

3. (1860–1882) jelölőnyelvi átírata,¹⁴⁹ a nyomtatott kiadás javított, digitális változata is az új környezetben született meg. A DigiPhil esetében ez a második olyan kiadás, amelynek autentikus változata a digitális.¹⁵⁰ Az első ilyen a Veres András által szerkesztett Kosztolányi Dezső *Édes Anna* kritikai kiadás, amelyet 2018 márciusában egy sajtótájékoztató és konferencia keretében mutatott be a kutatócsoport.¹⁵¹

Az Arany kritikai kiadás során Arany-filológusok, digitális bölcsészek és kódoló munkatársak működtek együtt egy olyan kutatókörnyezet biztosította digitális térben, amely maga is különféle forrású, korú és elhelyezkedésű, heterogén technikai aktorok (kliens és szerver számítógépek, szoftverkomponensek, szabványok, sémák, kódnyelvek) összjátékaként írható le. A következőkben egy térben és időben még heterogénebb intézményközi kollaboráció felépítését és munkamenetét ismertetem. Az Arany-kéziratok online publikálását célzó projekt kezdeményezője és koordinálója az Irodalomtudományi Intézet (ITI) volt, a résztvevő intézmények pedig a nagyszalontai Arany János Emlékmúzeum (AJE), a nagykőrösi Arany János Kulturális Központ (AJKK), a Magyar Tudományos Akadémia Könyvtár és Információs Központ (MTA KIK), a Petőfi Irodalmi Múzeum (PIM) és az Országos Széchényi Könyvtár (OSZK) voltak.

Az Arany-kéziratkereső létrehozásának alap gondolata az volt, hogy a fizikailag szétszórt, különböző intézményekben fellelhető Arany János kéziratok egyetlen online térben váljanak elérhetővé a kutatók és a szélesebb közönség számára egyaránt, a metaadatok szerint kereshető formában. A kezdeményezésnek különös hangsúlyt ad a tény, hogy a magyar irodalom kéziratok örökségén (számos kisebb gyűjtemény mellett) három nagy intézmény osztozik (OSZK, MTA-KIK, PIM), de sem a tömeges digitalizálás, sem a kéziratok le- vagy átírása, sem pedig az online közzététel szintjén nem működtek eddig együtt.

Az Arany kézirat-repozitórium létrehozásához meg kellett teremteni azt az intézményi együttműködést, amelynek során a kéziratokat gondozó intézmények közös elvek mentén készítik el a kéziratok digitális másolatait és összehangolt módszerekkel végzik el a kéziratok leírását. A kéziratkereső specifikációját a DigiPhil-csoport alakította ki az együttműködő intézmények könyvtárinformatikus szakértőinek segítségével.

¹⁴⁹ S. VARGA, *Arany János, Kisebb költemények 3. (1860–1882)*.

¹⁵⁰ A kiadás az alábbi linken érhető el: <https://digiphil.hu/gallery/kisebb-koltemenyek-3/>.

¹⁵¹ PALKÓ Gábor, VERES András és DOBÁS Kata, „A PIM DigiPhil műhelyének keretében megvalósuló digitális Kosztolányi-kiadások, valamint a Kosztolányi-adatbázis sajtóbemutatója” (Kosztolányi-nap - konferencia és koncert. Kosztolányi 2.0, Petőfi Irodalmi Múzeum, 2018. március 27.). A kritikai kiadás az alábbi linken érhető el: <https://digiphil.hu/edes-anna/>.

A digitalizálás első lépéseként az Országos Széchényi Könyvtár és az Irodalomtudományi Intézet munkatársai elvégezték az Arany János Emlékmúzeumban és az Arany János Kulturális Központban található kéziratok digitalizációját. A folyamat során nagy felbontású képfájlok jöttek létre, majd ezekből a fájlokból készültek el a publikálásra szánt nézőképek. Az OSZK, a PIM és az MTA KIK a saját gyűjteményeikben fellelhető kéziratok digitalizálását a belső szabályzataiknak megfelelően végezték el, a képfájlokat a saját rendszereikben tárolták.

A közös szolgáltatás kialakításának egyik sarkalatos pontja az aggregált tartalmak egységesítése, illetve a képfájlok elhelyezése volt. Követendő példának az Europeana aggregációs modellje kínálkozott: az Europeana csak a metaadatok összegyűjtését végzi el, a memóriaintézmények gyűjteményi elemeinek digitális szurrogátumait maguk az intézmények teszik közzé: az aggregált metaadatok egyike a digitális objektumra mutató link. Az Europeana a beérkező metaadatok egységesítésével sem foglalkozik, azokat saját formátumában fogadja csak el (Europeana Data Model, EDM), az adatszolgáltató intézmények saját felelőssége a metaadatok konverziója EDM formátumra.

A fenti modell sem a digitális objektumok, sem pedig a metaadatok kezelése szintjén nem volt alkalmazható a projekt számára. A határon túli gyűjtemények nem rendelkeztek repozitóriumi kapacitással, vagyis a digitális objektumok szolgáltatását is meg kellett oldani. A magyarországi közgyűjtemények ugyanakkor eltérő megoldásokat alkalmaztak a digitális faksimilék közzétételére, amelytől nem akartak eltérni, emellett különböző intézményi szabályzatok alapján kezelték a digitális objektumaikat. A probléma heterogén technikai megoldásokat követelt meg mind a metaadatok, mind a faksimilék kezelésében.

A digitális objektumok szolgáltatásához a repozitóriummal nem rendelkező határon túli partnerek számára a Bölcsészettudományi Kutatóközpont virtuális szerverén a DigiPhil egy olyan repozitórium szoftvert telepített, amely képes a digitális faksimilék tárolására. Ehhez a repozitóriumhoz nem készített nyilvános felhasználói felületet, az abban található objektumok közzététele egyedül a közös keresőfelületen történik. Ugyanezt a technikai megoldást (és ugyanezt a DSpace nevű szoftvert) alkalmazta az OSZK is saját szerverén. A faksimilét az MTA KIK az EPrints szoftverre épülő saját REAL rendszerében helyezte el, a partnerek közül egyedül kézirategységként egyetlen PDF-be egyesítve, a PIM pedig a HunTéka részeként működő JaDoX dokumentumkezelő szoftverben helyezte el a képeket. Az MTA KIK és a PIM kéziratára az átadott rekordokban tüntette fel a képek elérési útvonalát.

Mivel a projekt elsődleges feladata a kéziratokhoz készülő közös metaadatkereső-felület kialakítása volt, figyelembe kellett venni azt a tényt, hogy bár a résztvevő intézmények a

könyvtári rendszereikben szabványos formátumban írják le a rekordokat, az egyes könyvtári rendszerekben használt metaadatsémák eltérők lehetnek. A keresőfelület kialakításakor a résztvevők a Marc21 könyvtári adatsere formátum használata mellett döntöttek. A DigiPhil ekkor még a Koha integrált könyvtári rendszert használta az írói bibliográfiák építésére. Mivel a két határon túli intézmény nem rendelkezett megfelelő technikai háttérrel, kézírataikat a DigiPhil könyvtári rendszerében írták le az OSZK (Feke Eszter és Rózsafalvi Zsuzsa) és az ITI (Antal Alexandra és Bátor Anna) munkatársai. A Koha által használt metaadatformatum ugyan Marc21, azonban a DigiPhil korábban még nem írt le kéziratokat, ezért új metaadatspecifikáció kialakítására volt szükség. Az OSZK, annak érdekében, hogy a kéziratok leírása egységes legyen, nem a saját, HUNMARC alapú rendszerükben, az Amicusban, hanem egy új, erre a célra telepített Koha példányban dolgozta fel a kézírataikat. A DigiPhil metaadatspecifikációját a PIM könyvtárinformatikus szakértője, Mohay Anikó készítette elő, majd ez került kiegészítésre az OSZK munkatársa, Fejes Judit javaslataival. A DigiPhil és az OSZK Koha rendszereiben a kéziratok leírása ugyanazon specifikáció alapján zajlott. Mivel a feldolgozást végző munkatársak közül nem mindenki rendelkezett tapasztalattal a könyvtári rendszerek használatában, az adatfelvitel megkönnyítésének érdekében Mohay Anikó magyar nyelvű űrlapokat hozott létre, amelyek nem mutatták a teljes Marc21 adatmezőkészletet, csak a leíráshoz szükséges mezőket, emellett sok mező esetében (például proveniencia) pedig előre meghatározott listából választhattak a kollégák. A két Koha rendszerben leírt rekordok esetében, ahol lehetséges volt, a leíró kollégák adatgazdagítást végeztek. Egy előzetesen, a kritikai kiadások helynevei alapján készített listából felvitték a helynevekhez a GeoNames névterében található azonosítót, tárgyszavakkal látták el a rekordokat, valamint azokat a kéziratokat, amelyek valamely kritikai kiadásban megjelentek, összekötötték a DigiPhil oldalán publikált digitális kiadással. Az MTA KIK már 2016 folyamán átállt a Marc21 adatformatumra,¹⁵² a PIM saját HunTéka rendszeréből szintén ebben a formátumban nyerte ki és adta át a publikálásra szánt rekordokat. Az MTA KIK a saját gyűjteményéhez tartozó anyagok metaadatolását az Aleph könyvtári rendszerben végezte.


A közös adatsere formátum használatának ellenére felmerültek olyan jelölési eltérések, amelyeket egységesíteni kellett a rekordok közös rendszerbe való betöltése előtt. Az MTA KIK által szolgáltatott rekordokban a személyneveket leíró 100-as (szerző), 600-as (róla szóló) és 700-as (egyéb személynév) mezőkben a vezetéknevet és keresztnévet vesszővel választották el,

¹⁵² GYURICZA Andrea, „Szabványos metaadatok jelentősége a kooperatív szolgáltatásokban. Átállás HUNMARC-ról a MARC21 szabványra a WorldCat-csatlakozás előkészítéseként.”, *Könyvtári Figyelő* 63, 2. sz. (2017): 218–225.

míg a többi intézmény által szolgáltatott rekordokban csak a nem magyar személynevek esetén alkalmaztak hátravetést. Annak érdekében, hogy a keresőfelületen ezek az értékek egységesen szerepeljenek, az MTA KIK-től kapott rekordokban félautomatikus módon, reguláris kifejezések segítségével a magyar személynevekből el kellett távolítani a szeparátorokat. Ugyanezen mezők \$d almezőjében, amelyben a személy születési és halálozási dátuma adható meg, a többi intézménnyel ellentétben a DigiPhil leírásában az évszám zárójelbe került, ezeket szintén egységesítették. Az egységesítésnél kiemelt szempont volt, hogy minimális változtatások történjenek a rekordokon és azokat a lehető legrészletesebben dokumentálják, hogy a műveletek bármikor megismételhetők legyenek, hiszen ezek a változtatások nem kerültek vissza az eredeti rendszerekbe, egy esetleges frissítés során újra végre kell hajtani őket. A mezők kitöltésének eltérésein túl előfordultak olyan esetek is, amikor a gyűjteményekben ugyanazt az információt más adatmezőbe vitték fel. Ilyen például az egyedi azonosító, ami a DigiPhil és az OSZK által leírt rekordok esetében a 999-es mezőbe került, míg az MTA KIK a 001-es mezőt használta ugyanerre a célra. A különbség a használt könyvtári rendszerekből fakad, a Koha sajátossága, szemben az Aleph-el, hogy a kiosztott azonosítót nem a 001-es mezőbe helyezi. Hasonló a helyzet a tárgyszavakkal: míg a DigiPhil és az OSZK a leírás során a tárgyszavakat a 650-es mezőbe vitte fel, addig az MTA KIK ugyanerre a 653-as mezőt használta. Az ilyen jellegű eltéréseket a kereső algoritmus finomításával oldották meg, például a tárgyszavak esetében a rendszer egyszerre figyeli mindkét mezőt. A fenti munkamenet leírása jól érzékelteti azt a problémát, hogy bizonyos esetekben hiába az azonos rendszer, vagy szabvány alkalmazása, ezek rugalmassága és az egyéni kutató értelmezése következtében eltérő adatstruktúrák jönnek létre. Erre szolgálhat megoldásul, ha az egyes intézmények közös irányelveket követve írják le az adataikat.

Az Arany-kéziratok repozitórium keresőfelülete egyszerre támogatja a szabad szavas és facettás keresést a metaadatok között. A „Részletes keresés” menüpont alatt található azok a keresőmezők, amelyek az egyes metaadatmezőket reprezentálják, ezekben szabad szavasan és legördülő listából való választással egyaránt kereshet a felhasználó.

DIGIPHIL
PROJEKTEK SZERZŐK MŰVEK



Kézirat kereső

Rendezés: Dátum (növekvő) ▼

Részletes keresés
Filterek törlése

Szerző/Közreműködő

Cím/Megnevezés

Dátum -

Dedikáció

Hely

Megjegyzés

Műfaj

Róla szóló

Proveniencia/Származás

Tárgyszó

ID

Részletes keresés
Filterek törlése

[17. ábra] Keresőmezők az Arany-kéziratok repozitóriumában

A facettás kereső használatakor a felhasználó egyszerre több metaadatmezőben végezhet szűkítést, továbbá az egyes listák esetében is egyszerre több értéket jelölhetünk ki, ekkor a szűkítés során a keresőmotor minden kiválasztott értéket figyelembe véve jeleníti meg a találatokat.

Szerző/Közreműködő	Arany János 1817-1882
Cím/Megnevezés	Bármely
Dátum	ÉÉÉÉ - ÉÉÉÉ
Dedikáció	Bármely
Hely	Bármely
Megjegyzés	
Műfaj	vers
Róla szóló	Bármely
Proveniencia/Származás	Arany János Emlékmúzeum, Nagyszalo
Tárgyszó	<input type="text" value="Search"/>
ID	<input checked="" type="checkbox"/> Arany János Emlékmúzeum, Nagyszalonta <input type="checkbox"/> Magyar Tudományos Akadémia Könyvtár és Információs Központ <input checked="" type="checkbox"/> Nagykőrösi Arany János Kulturális Központ <input type="checkbox"/> Országos Széchényi Könyvtár <input type="checkbox"/> Petőfi Irodalmi Múzeum

Összesen: 8 találat

[18. ábra] Szűrés az Arany János Nagykőrösön és Nagyszalontán található verskézirataira

A keresőrendszer kialakításakor elsődleges szempont volt, hogy olyan nem robusztus, nyílt forráskódú eszközt válasszunk, amelynek hosszú távú fenntartása, fejlesztése biztosított, emellett az adatbetöltés, frissítés felhasználóbarát módon történhet. A választás az eXistDB nevű adatbázis-kezelő szoftverre esett, amely képes nagy mennyiségű XML formátumú fájl tárolására és kezelésére, miközben a felhasználói felület egyszerűen testre szabható. A szolgáltatás kialakításakor fontos szempont volt, hogy a magyarországi közgyűjtemények maguk szolgáltatassák a digitális faksimiléiket, a keresőbe csak a metaadatokat töltsük fel. A felhasználói felület tervezésekor sikerült megoldani, hogy a DSpace rendszerekben tárolt fájlok magán a közös keresőfelületen kerüljenek megjelenítésre.

A programban résztvevő intézmények a projekt végére összesen 2003 rekordot szolgáltatottak. A projekt lezárását követően a rendszerben található rekordok nem frissültek, új rekordok nem kerültek a rendszerbe. Bár a projekt véget ért, az MTA KIK folyamatosan javítja és egységesíti a katalóguscédulákból létrejött rekordjait, valamint feldolgozza a még digitalizálatlan kéziratosokat. A projekt során kialakított munkafolyamat alkalmasnak tűnt arra, hogy a később hasonló módon dolgozzuk fel és publikáljuk más szerzők kéziratos hagyatékát.

A kialakított heterogén aggregációs modell azt is lehetővé tette volna, hogy a szolgáltatáshoz különféle technológiai és intézményi feltételekkel csatlakozhassanak memóriaintézmények és kutatócsoportok. A bevezetőben említett infrastruktúra váltás a fent bemutatott gyűjteményt is érinti, a rekordok egy új, szemantikus web alapú rendszerbe kerülnek át, a jövőbeni bővítésük már abban a rendszerben fog zajlani.

Az eddigiekben olyan szolgáltatásokat ismertettem, amelyek különmemű aktorok együttműködésében jöttek létre, és amelyekben a humán közreműködők kollaborációját valamely magában is heterogén virtuális kutatókörnyezet vagy technológiai platform kanalizálta. Az ismertett szolgáltatásokban tevékeny (emberi és nem-emberi) aktorok köre ugyanakkor a megszokott vagy szabályozott keretek között maradt mind a humán, mind a technikai oldalon: digitális bölcsészek, szabványok, széles körben használt szoftvereszközök működtek közre. Ugyanakkor a digitális kulturális örökség, illetve filológia területén számos olyan gyakorlat létezik ma, amely a fentieknél radikálisabban nyitja meg az aktorok körét. Az egyik ilyen, mind szélesebb körben alkalmazott praxis a kéziratok átírásának közösségi (crowdsourcing) kiterjesztése. A már klasszikusnak számító Bentham Project¹⁵³ mellett érdemes az Europeana hasonló, még szélesebb kört megszólító kezdeményezését is szóba hozni.¹⁵⁴

Még radikálisabb azonban az a gyakorlat, amelyik a gépi intelligenciát a korábban kizárólag emberi szereplők kompetenciakörébe utalt feladatokba integrálja. Ilyen terület a mesterséges intelligencia felhasználása a kéziratok átíratainak elkészítésében.

Az optikai karakterfelismerés (OCR) gyakorlatait már jó ideje alkalmazzák a kulturális örökség digitalizálásában, legtöbbször olyan hibrid, a digitális és az analóg médium határán egyensúlyozó formátumok létrehozásában, mint amilyen a kétrétegű dokumentum, amelyik a nyomtatott médium elrendezését (layout) a kereshetőség igényével párosítja. Köztudott, hogy az OCR technológia kevésbé alkalmas kézírásos dokumentumok átírására, az azonban kevésbé ismert tény, hogy már az '50-es évek közepétől, párhuzamosan a nyomtatott szövegek gépi átírásának fejlesztésével, a kéziratok anyagai gépi átírására is történtek kísérletek.¹⁵⁵ A kézírás-felismerés (handwritten text recognition, HTR) alkalmazása a pénzügyi és banki ágazatokra korlátozódott (például levélcímzés felismerése, csekkek elemzése). A tömeges gyűjteményi digitalizálás hozta magával azt az igényt, hogy a kéziratok kikerülve, vagy legalábbis

¹⁵³ A Bentham Project oldala: <https://www.ucl.ac.uk/bentham-project/>.

¹⁵⁴ Enrich Europeana: <http://transcribathon.eu/>.

¹⁵⁵ Guenter MUEHLBERGER és mtsai., „Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study”, *Journal of Documentation* 75, 5. sz. (2019. szeptember 9.): 954–976, <https://doi.org/10.1108/JD-07-2018-0114>.

előkészítve a humán aktorok végezte átírás lassú és fáradságos folyamatát, ugyanúgy kereshetővé és gépek számára is feldolgozhatóvá váljanak, mint a nyomtatott és OCR-ezett szövegek.¹⁵⁶

A HTR technológia legnagyobb kihívása a kézírás egyedisége, erre jelent megoldást a gépi tanulás friss trendje, a neuronháló (Artificial Neural Network, ANN) technológia alkalmazása a HTR folyamatában. Amennyiben egy szerzőtől megfelelő mennyiségű átírat és faksimile áll rendelkezésre, a gép képes lehet megtanulni a „kézre” jellemző stílusjegyeket és létrehozni egy olyan modellt, amely felismeri, kiolvassa a szerző kézírását. Minél több átírat és nagyobb számítási kapacitás áll rendelkezésre, a modell annál pontosabb lesz.

2016-ban hozta létre az Innsbrucki Egyetem Digitalisierung & Elektronische Archivierung¹⁵⁷ kutatócsoportja a Horizon 2020 támogatásból megvalósult READ (Recognition and Enrichment of Archival Documents)¹⁵⁸ projekt keretében a Transkribus szolgáltatást.¹⁵⁹ A Transkribus volt az első olyan platform, amely nyílt hozzáférést biztosított egy felhasználóbarát felülettel rendelkező és neuronhálón alapuló HTR eszközhöz.

Az ELTE Digitális Bölcsészeti Központ az ITI-vel együttműködve kezdett kísérletezni Arany János hivatali iratainak felismertetésével a Transkribus szolgáltatásban. A Transkribus rendelkezett ugyan kész kézírás-felismerő modellekkel, azonban azokat angol és német nyelvű kéziratokon tanították be, így magas hibaszázalékkal működtek az Arany-kéziratokon. A Transkribus lehetőséget biztosít arra, hogy saját modellt hozzunk létre, így a Központ munkatársai száz kéziratlap átíratát készítettek el, amelyből a Transkribus algoritmusai létrehozták az első Arany-kézírás modellt. Az első modell ellenőrzése során kiderült, hogy karakterek szintjén mintegy 30%-os hibaszázalékkal (Character Error Rate, CER) működött. A második modell előkészítésében és tesztelésében már a DigiPhil munkatársa, Bobák Barbara és az ITI munkatársa, Gábori Kovács József vettek részt.¹⁶⁰ Újabb száz, Arany által írt kéziratlap került átírásra, így az új modell már kétszáz kéziratlap átírata alapján készült el. A Transkribus a második modell elkészülte előtt javított a modellt készítő algoritmuson, a Google által

¹⁵⁶ Ehhez lásd: PALKÓ Gábor, „Mesterséges intelligencia, digitális bölcsészet, kulturális örökség: trendek és eredmények”, in *Valós térben - az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TICK József, KOKAS Károly és HOLL András, 235–240 (Budapest: Hungarnet Egyesület, 2022), <http://real.mtak.hu/155513/>.

¹⁵⁷ Digitalisierung & Elektronische Archivierung (DEA):

<https://www.uibk.ac.at/germanistik/einrichtungen/dea.html>.

¹⁵⁸ Recognition and Enrichment of Archival Documents: <https://readcoop.eu/about/>.

¹⁵⁹ Transkribus: <https://readcoop.eu/transkribus/>.

¹⁶⁰ BOBÁK Barbara és GÁBORI KOVÁCS József, „Kézírásfelismerés Arany János levelein”, in *Networkshop 2019*, szerk. TICK József, KOKAS Károly és HOLL András, 38–44 (Budapest: Hungarnet Egyesület, 2019), <http://ocs.mtak.hu/index.php/nws/2019/paper/view/23>.

fejlesztett Tensorflow alkalmazásával.¹⁶¹ Ennek következtében jelentősen javult a neuronháló teljesítménye. Ahogyan Bobák és Gábori írják,¹⁶² a második modell hatékonysága erős javulást mutatott, a CER a felére csökkent. A Transkribus saját tapasztalatai szerint a modell tanításával a CER akár 5% alá csökkenthető.¹⁶³ Mivel a modellt csak az Arany által írt kéziratokon tanították be, nem volt alkalmas arra, hogy más szerző kézírataiból jó minőségű átíratokat készítsen, azonban a modell tanítóanyagának bővítésével lehetővé válhatott általánosabb érvényű modellek megalkotása is.

Az Európai Unió által finanszírozott Horizon 2020 támogatás megszűnése miatt a Transkribus fejlesztésére és fenntartására céget alapítottak (READ-COOP SCE) és az addig korlátlanul hozzáférhető szolgáltatást jelentősen korlátozták, használatát előfizetéshez köthették.¹⁶⁴ Az új finanszírozási modell részeként továbbra is 500 kreditet ingyen biztosítanak a felhasználók számára, így a kis terjedelmű kéziratok ingyenes, automatikus átírása továbbra is lehetséges.¹⁶⁵ Ezen felül nem korlátozták a korábban fejlesztett kliensben és az online szerkesztőben a lap elrendezés gépi felismertetését (szegmentálását), valamint a kézi átírást. Az Irodalomtudományi Intézet az Arany-modell továbbfejlesztéséhez és a további modellek építéséhez krediteket vásárolt. Az Intézet a Miskolci Egyetemmel és az ELTE-vel együttműködésben kezdte meg a harmadik Arany-modell fejlesztését. A modell tanításához, az első két tanítóanyag mellett, Arany titkárának, Ring Adorjának a kezétől származó, valamint további, közel 30 kéz írását tartalmazó korpuszt készítettek a munkatársak. A harmadik modellt összesen 874 (Arany: 200, Ring: 101, Más: 573) átírt kéziratlapon tanították. Fekete Norbert elemzéséből látható, hogy az így létrejött modell segítségével elenyésző hibaarányal (CER: 6%) ismertethetők fel a hivatali iratok.¹⁶⁶ Hasonlóan kiemelkedő eredményeket sikerült elérni Johannes Michael Brutus 2020-ban megtalált, addig elveszettnek hitt történeti munkájának

¹⁶¹ MUEHLBERGER és mtsai., „Transforming Scholarship in the Archives through Handwritten Text Recognition”, 966.

¹⁶² BOBÁK és GÁBORI KOVÁCS, „Kézírásfelismerés Arany János levelein”, 6.

¹⁶³ MUEHLBERGER és mtsai., „Transforming Scholarship in the Archives through Handwritten Text Recognition”, 962.

¹⁶⁴ A modellváltást 2020. október 20-i, „Transkribus: A new Chapter begins” tárgyú körlevelükben jelentették be. A körlevél honlapjukon közzétett változata a dolgozat írásának idejében már nem elérhető, az Internet Archive sem mentette a felületet.

¹⁶⁵ A szolgáltatásban jelenleg 50 kézírás-felismerő modell érhető el a felhasználók számára. Az egyetlen magyar nyelvű modellt, amelyet közzétettek az OSZK Digitális Bölcsészeti központja készítette, a modellt Kiss József és levelezőpartnerei kezén tanították: <https://readcoop.eu/model/hungarian/>.

¹⁶⁶ FEKETE Norbert, „HTR-modellépítés és kézírásfelismerés nagyméretű, többszerzős szövegtörzseten A Transkribus alkalmazása az Arany János hivatali iratokon”, in *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TÍCK József, KOKAS Károly és HOLL András, 271–275 (Budapest: Hungarnet Egyesület, 2022), 275, http://real.mtak.hu/155518/1/34_NWS_2022_Fekete.pdf.

(*Rerum Hungaricum Libri*) kéziratán is.¹⁶⁷ A 2200 lapból álló, Brutus írnoka által jegyzett kézirat felismertetéséhez 200 lap átírásával készült tanítóanyag. A modell mindössze 1,9%-os karakterhiba-aránnyal olvasta ki a teljes szöveget.

A szolgáltatás egyik nagy előnye, hogy az átírás eredményét, legyen az gépi közreműködés eredménye vagy humán aktor által javított, a hétköznapi formátumok mellett (TXT, PDF) olyan TEI XML formátumban is exportálhatjuk, amely tartalmazza az átírat és a faksimile kép kapcsolatát. Ezáltal lehetőség nyílik arra, hogy a mesterséges intelligencia közreműködésével meglepően rövid idő alatt olyan digitális forráskiadásokat készítsünk, amelyek – emberi közreműködéssel – kritikai kiadások alapjául szolgálhatnak, hiszen a feldolgozás az első lépésektől szabványos környezetben történik.

Továbbra is problémát jelentettek azonban a szolgáltatás használatának költségei, az Irodalomtudományi Intézet által vásárolt kreditek egy része a modellkészítések költségeit fedezték. Nem csak a filológiai kutatások, hanem a közgyűjteményi digitalizáció terén is jelentős igény tapasztalható a kéziratos korpuszok automatikus felismertetésére. A Digitális Örökség Nemzeti Laboratórium vezetésével kezdődött meg az a projekt,¹⁶⁸ amelynek célja egy, a Transkribus-t helyettesíteni képes, ingyenesen hozzáférhető szolgáltatás fejlesztése. Mivel a Transkribus nem teszi közzé a tanítóanyagok alapján készített modelleket, ezért saját modelleket kellett létrehozni. A modellek létrehozásához a Transkribus-ban átírt tanítóanyagokat exportálták az OCR és a HTR területén használt, szabványos ALTO XML formátumban. Az eszköz fejlesztése folyamatos,¹⁶⁹ de a Brutus kézirat esetében már látható, hogy közel azonos hatékonysággal ismeri fel a szöveget, míg a saját Arany-modell hatékonysága jelentősen felülmúlja a Transkribus-szal elért eredményeket.¹⁷⁰ A korábbiak mellett az Intézet és a konzorciumi partnerek folyamatosan együttműködnek új modellek készítésében, az ahhoz szükséges tanítóanyagok előállításában.

¹⁶⁷ BOBÁK Barbara és KASZA Péter, „Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus *Rerum Ungaricarum libri* kéziratának digitális kiadása (esettanulmány)”, in *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TICK József, KOKAS Károly és HOLL András, 154–160 (Budapest: Hungarnet Egyesület, 2022), http://real.mtak.hu/155493/1/19_NWS_2022_Bobak_Kasza.pdf.

¹⁶⁸ PALKÓ Gábor, BOBÁK Barbara és SZEKRÉNYES István, „A Digitális Örökség Nemzeti Laboratórium webszolgáltatásai automatikus kézírás-felismertetéshez”, in *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé*, szerk. TICK József, KOKAS Károly és HOLL András, 164–169 (Budapest: Hungarnet Egyesület, 2023), <https://real.mtak.hu/182935/>.

¹⁶⁹ A tanítási folyamat technikai részleteihez lásd: Uo.

¹⁷⁰ Uo., 166.

Digitálisan született tudományos szövegkiadások

Az előző alfejezetekben a korábban nyomtatásban megjelent kritikai kiadások informatizálásának lehetőségeit mutattam be. Láthatóvá vált, hogy még az alapfokú szemantikai feldolgozás, amelynek kimenete a jelölőnyelvi átírat, is mennyivel komplexebb kiadás közzétételét teszi lehetővé egy digitalizált, azaz egy automatikusan előállított kétrétegű PDF formátumban publikált kiadással szemben. Egy nyomtatott, például genetikus kritikai kiadás magas szintű szemantikai értelmezése lehetővé teszi a nyomtatott kiadás hátrányainak kiküszöbölését, jelentősen megkönnyíti annak használatát, a szövegtanúk összevetését. A következőkben eleve digitálisan született, *born digital* kritikai kiadásokat elemzek. Az elemzés célja kettős: rámutatni a digitális technológia filológiára gyakorolt hatására, valamint bemutatni a technológia fejlődésén keresztül a szövegkiadói gyakorlat és a szövegközlési elvek formálódását.

A *born digital* kiadás fogalmát a bevezetőben definiáltam, lényegében olyan kiadás, amely elkészítésének első pillanatától kezdve jelölőnyelvi átíratban készül, ám a következőkben látni fogjuk, hogy mennyire törekeny ez a definíció, az a kiadás, amelyet DigiPhil történetében az első, valóban *born digital* kritikai kiadásaként tartunk számon, a kezdetekben például irodai szövegszerkesztőben készült.

Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása

A Petőfi Irodalmi Múzeumban 2016-ban indult, az NKFIH támogatásával, Móricz Zsigmond levelezésének kritikai kiadását előkészítő, Cséve Anna által vezetett kutatási projekt.¹⁷¹ Már a kutatási pályázat is tartalmazta, hogy a kutatócsoport digitális kritikai kiadást készít, a kutatási terv kiadásra vonatkozó részének kialakításában a DigiPhil munkatársai is közreműködtek.¹⁷² Fontos kiemelni, hogy, nem sokkal a levelezéskiadás munkálatainak megkezdése előtt indult, szintén Cséve Anna vezetésével és a DigiPhil szakmai felügyeletével, Móricz világháborús naplójának, az 1914 és 1917 között írt *Tükörnek* az annotált digitális

¹⁷¹ A kutatócsoportban résztvevő kutatók listája a kiadás előszavában érhető el: CSÉVE Anna, „Előszó”, in *Móricz Zsigmond levelezésének digitális kritikai kiadása (1882–1913): 2.0*, szerk. PALKÓ Gábor és mtsai. (Budapest: Petőfi Irodalmi Múzeum, 2023), https://dhupla.hu/text/mzs-lev_eloszo.

¹⁷² A projekt első tárgyévéről szóló beszámolót lásd: CSÉVE Anna, FELLEGI Zsófia és KÓMÁR Éva, „Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása”, *Digitális Bölcsészlet* 1 (2018. szeptember 19.): 159–174, <https://doi.org/10.31400/dh-hun.2018.1.227>.

forráskiadását készítő kutatási projekt.¹⁷³ Bár a fejezet elsősorban a kritikai kiadásra fókuszál, a két projekt tapasztalatai hatottak egymásra, ezért a levelezés-kiadás bemutatása mellett, ahol szükséges, kitérek a Tükörben alkalmazott megoldásokra is.

A DigiPhil, ahogyan láthattuk, leszámítva *A Tett* folyóirat forráskiadását, korábban csak nyomtatásban megjelent kritikai kiadásokat informatizált. A digitális kiadás munkafolyamatainak előkészítése különböző, egymással összefüggő szinteken okozott kihívásokat. Az első és legfontosabb lépés a jelölőnyelvi átírat specifikációjának kidolgozása volt. A projekt indulásakor a digitális levelezés-kiadások száma csekély volt,¹⁷⁴ az akkor elérhető specifikációk és a TEI ajánlásában szereplő egyszerű példák nem nyújtottak kellő támpontot. A specifikáció kidolgozása, Palkó Gábor szakmai irányítása mellett, az én feladatom volt.¹⁷⁵ Az első és talán legfontosabb lépés a digitális objektum meghatározása volt. A kutatócsoporttal egyetértésben az a döntés született, hogy a digitális objektum nem csak a levelet, hanem a borítékot és a levéllel együtt elküldött mellékleteket is tartalmazza, ugyanis az ezeken található információk (pl. újságkivágatok, versek, fényképek) relevánsak a teljes korpusz szempontjából. Ennek leképezésére olyan egyedi struktúrát alakítottunk ki a TEI specifikációban, amely alkalmas mind a metaadatok, mind a szövegek közlés szintjén megjeleníteni a három egységet. A TEI ajánlása lehetővé teszi, hogy több kézirat metaadatait és azok szövegek közlését egy TEI objektumban rögzítsük:

¹⁷³ A kutatócsoportban résztvevő kutatók listája a forráskiadás előszavában érhető el: Anna CSÉVE és mtsai., szerk., „Előszó”, in *Móricz Zsigmond: Tükör* (Budapest: Petőfi Irodalmi Múzeum, 2018), <https://dhupla.hu/collection/moricz-zsigmond-tukor>.

¹⁷⁴ Az akkor elérhető legnagyobb ilyen korpusz a Vincent Van Gogh-levelezés-kiadás volt, ők a TEI ajánlását saját jelölőkkel egészítették ki, amelyek egy részét a Digital Archive of Letters in Flanders, szintén a TEI-t kiegészítő, saját elemeket is tartalmazó sémájából vették. Bővebben lásd: Leo JANSEN, Hans LUIJTEN és Nienke BAKKER, szerk., *About This Edition* (Amsterdam & The Hague: Van Gogh Museum & Huygens ING, 2009), https://vangoghletters.org/vg/about_6.html.

¹⁷⁵ A projekt során csatlakozott a DigiPhil-hez Bobák Barbara. A DigiPhil-ben a finanszírozás jellege miatt a projektek egy részében több munkatárs is dolgozott, a kiadások kapcsán szeretném kiemelni Vincze Tímea Brigitta, Metzger Réka, Czinkos Zsolt és Kundráth Péter közreműködését.

```

<msDesc>
  <msIdentifier>
    <country>Magyarország</country>
    <settlement>Budapest<idno type="KOHA_GEO">KOHA_GEO:9227</idno>
    </settlement>
    <institution>Petőfi Irodalmi Múzeum</institution>
    <repository>Petőfi Irodalmi Múzeum Kézirattár</repository>
    <idno>PIM M. 130</idno>
    <msName>Móricz Zsigmond-hagyaték</msName>
  </msIdentifier>
  <msPart style="letter">
    <msIdentifier/>
    <physDesc>
      <objectDesc>
        <supportDesc>
          <extent>
            <measure type="quantity" unit="folio">2</measure>
            <dimensions unit="mm">
              <height>211</height>
              <width>170</width>
            </dimensions>
          </extent>
          <condition>
            <p>Fekete tintairás az első fólió mindkét oldalán és a
              második fólió rektóján. A második fólió
              verzóján<persName>Móricz Zsigmond<idno
                corresp="Móricz Zsigmond" type="KOHA_AUTH"
                >KOHA_AUTH:120256</idno>
              </persName>utólagos kék ceruzairású datálása, fekete
              tintairású rájegyzése és aláíráspróbája. A hajtásoknál
              kissé szakadozott papíron.</p>
          </condition>
        </supportDesc>
      </objectDesc>
    </physDesc>
  </msPart>

```

```

<msPart style="attachment">
  <msIdentifier/>
  <physDesc>
    <objectDesc>
      <supportDesc>
        <extent>
          <measure type="quantity" unit="folio"/>
          <measure type="quantity" unit="piece">4</measure>
          <dimensions unit="mm">
            <height/>
            <width/>
          </dimensions>
        </extent>
        <condition>
          <p>A mellékletek nincsenek a levél mellett.
            <persName>Pallagi Gyula<idno corresp="Pallagi Gyula"
              type="KOHA_AUTH">KOHA_AUTH:121464</idno>
            </persName>négy ajánlólevelet mellékel:
            <persName>Kovács Lajosnak<idno
              corresp="Kovács Lajos" type="KOHA_AUTH"
              >KOHA_AUTH:184318</idno>
            </persName>,<persName>Sinka Sándornak<idno
              corresp="Sinka Sándor" type="KOHA_AUTH"
              >KOHA_AUTH:215177</idno>
            </persName>,<persName>Erdős Józsefnek<idno
              corresp="Erdős József" type="KOHA_AUTH"
              >KOHA_AUTH:107932</idno>
            </persName>, valamint <persName>Bujdosó Lajosnak<idno
              corresp="Bujdosó Lajos" type="LOK"
              >LOK:00308</idno>
            </persName>cimezve.</p>
          </condition>
        </supportDesc>
      </objectDesc>
    </physDesc>
  </msPart>
</msDesc>

```

[19. ábra] Részlet Pallagi Gyula Móricz Zsigmondnak írt, keltezetlen levelének jelölőnyelvi átíratából.¹⁷⁶

A kézirat leírására szolgáló egységekben a kutatócsoport nem csak a kézirat lelőhelyére vonatkozó adatokat rögzítette, hanem teljes kéziratleírást is készített, a kéziratlapok pontos méretével és a kéziratok állapotának leírásával.

A levelezés aktusainak leírására a **creation** jelölőbe ágyazott **correspDesc** modul szolgál. Itt a levél írásának körülményeinek (szerző, dátum, hely) leírásán túl, amennyiben rendelkezésre álltak további adatok, a küldésre és az átvételre vonatkozó információk is

¹⁷⁶ A levél átíratának jelenlegi verziója: <https://dhupla.hu/text/o:mzs-lev.tei.22>.

rögzítésre kerültek. A digitális technológia hatása a filológiai munkálatokra a szövegek közlésén is tetten érhető. A jelölőnyelvi átírat lehetőséget biztosít a filológiai jelenségek kifinomult jelölésére és ez kitágította a szövegek közlés lehetőségeit is. A kutatócsoport párhuzamosan készített betűhív átíratot és emendált olvasói szöveget. Ezek egy TEI XML fájlban, az adott szöveghelyen kerültek rögzítésre. A kiadásban számos olyan filológiai szempontból releváns jelenséget is rögzítettek, ami a nyomtatott kritikai kiadások gyakorlatában ritka, mint például az aláírás jelölővel történő ellátása, vagy a pecsét szövegének és helyének rögzítése. Bár a jelölőnyelvi átírat megfelel a TEI ajánlásának, bizonyos esetekben el kellett térnünk a TEI javaslataitól. Erre jó példa az előbb is említett aláírás jelölése. Az erre szolgáló jelölő, a **signed** a TEI szintaktikája szerint a **closer** jelölőbe kerül, amelyet azonban az adott szövegegységben (**div**), nem követhet további, bekezdések alá rendelt szöveg. A kézirategyüttes jellegéből fakadóan ezt a megoldást nem alkalmazhattuk, hiszen számos helyen követte az aláírást egyéb szöveg, utóirat, rájegyzés stb. Ennek megoldására vezettük be a **seg** elemet, amelynek funkciója, hogy bizonyos elvek alapján kijelölt szegmentumokat azonosítsunk vele. A szegmentum típusát a **@type** attribútumban definiáltuk, jelen esetben **@type="signature"**. A kiadásban a filológiai jelenségek jelölése mellett a kutatók elvégezték a személy- és földrajzi nevek, valamint az említett művek azonosítását és a DigiPhil könyvtári rendszerében található névterekkel és bibliográfiai rekordokkal történő összekapcsolását is. Az azonosításon túl a kutatócsoport folyamatosan építette, javította a könyvtári rendszerben található rekordokat. A korpuszban rengeteg utalás van olyan személyekre, akiknek az azonosítására nem állt rendelkezésre elegendő adat (pl. a Móricz család alkalmazottai), így nem lehetett önálló rekordot létrehozni a névtérben. Ezeknek a személyeknek az elkülönítésére úgynevezett lokális azonosítókat hoztunk létre, a korpuszban minden azonosítható előfordulásukat jelölte a kutatócsoport, a rájuk vonatkozó leírásokat pedig egy XML fájlban rögzítették. A források feldolgozása során természetesen új jelenségek kerültek elő, ezért a specifikáción is folyamatosan alakítottunk.

A kutatási projekt másik kihívása a munkamódszer kidolgozása volt. Bár nemzetközi szinten már rengeteg *born digital* kiadás volt elérhető, a kiadások elkészítésének folyamatára vonatkozó információk töredékesek voltak. Általánosságban elmondható, hogy, néhány kivételtől eltekintve, ezeknek a kiadásoknak a jelölőnyelvi átíratát egy XML szerkesztő segítségével digitális filológiai ismeretekkel rendelkező kutatók állították elő. Mivel a projektben részt vevő kutatók nem rendelkeztek digitális filológiai szakértelemmel, olyan munkafolyamatot dolgoztunk ki, amelyben a kutatóknak nem kellett közvetlenül jelölőnyelvi kódokat előállítaniuk. Az első kísérlet még nem a levelezéskiadáshoz kapcsolódik, hanem a

Tükör forráskiadásához. Az volt az alapgondolat, hogy a kutatócsoport olyan, jól ismert környezetben végezze a munkát, amelyből aztán automatikus módszerekkel előállítható a jelölőnyelvi átírat. Kézenfekvőnek tűnt a Microsoft Word irodai szövegszerkesztő program alkalmazása.¹⁷⁷ A *Tükör* esetében olyan sémafájlt hoztunk létre,¹⁷⁸ amelynek importálásával új menüsor jött létre, amely gombokra helyezve tartalmazta az előre rögzített műveleteket. Például a beszúrás menüpontra kattintva a kijelölt szövegrész formázása megváltozott, egy előre meghatározott színkóddal jelölt színárnyalatú lett a szövegrész. Ez a megoldás elméletben lehetővé tette, hogy a rögzített formázásokat XML jelölőkre cseréljük Visual Basic szkriptek segítségével. Mivel a projekt nem sokkal a levelezéskiadás előtt kezdődött, még nem rendelkezünk megfelelő tapasztalatokkal az átalakítást illetően, ezért a levelezés-kiadást készítő kutatók hasonló rendszerben kezdték meg a munkálatokat.

A levelezés-kiadás digitális objektumai egy levelet és ahhoz tartozó borítékot, illetve mellékleteket tartalmazták, szemben a *Tükörrel*, ahol a digitális objektum egy kézírásos könyvet reprezentált, tehát végül ott csak néhány TEI XML jött létre kimenetként. A levelek esetében egy-egy digitális objektumhoz különféle metaadatok tartoztak, a mintegy 1500 levél esetében ezek rögzítését is a Word-ben kellett elvégezni. Az egyes levélleírások két részből álltak, a kutatóknak egy előre definiált mezőket tartalmazó táblázatot kellett kitölteniük, majd a szövegátírásban a jelölni kívánt szegmentumokat és a névazonosítást egy speciális karaktereket tartalmazó specifikáció mentén kellett elvégezniük.

¹⁷⁷ Akkoriban a DigiPhil partnere, a Graz-i egyetemen működő Zentrum für Informationsmodellierung munkatársai hasonló kísérletekről számoltak be, ők előre definiált stílusokat alkalmaztak.

¹⁷⁸ A sémafájl első változatát, amely úgynevezett makró parancsokat tartalmazott, Mihály Eszter készítette.

levél azonosítója	PKEL.M.130-pallagigyulato1-0011_a	\$ levél azonosítója
lelőhely	PIM M. 130/pallagigyulato1/0011	\$ lelőhely
proveniencia	Kiss Ferenc tulajdonából (2005)	\$ proveniencia
hagyaték	Móricz Zsigmond-hagyaték	\$ hagyaték
levélíró, azonosítója	Pallagi Gyula KOHA_AUTH:121464	\$ levélíró
levélíró testület		\$ levélíró testület
egyéb levélíró, azonosítója		\$ egyéb levélíró
megírás helye, azonosítója		\$ megírás helye
megírás dátuma	after:1900-11-18 ¹	\$ megírás dátuma
feladás helye, azonosítója		\$ feladás helye
feladás dátuma		\$ feladás dátuma
címzett, azonosítója	Móricz Zsigmond KOHA_AUTH:120256	\$ címzett
címzett testület		\$ címzett testület
egyéb címzett, azonosítója		\$ egyéb címzett
átvétel helye, azonosítója		\$ átvétel helye, azonosítója
átvétel ideje		\$ átvétel ideje
nyelv	hu	\$ nyelv
levél fólió száma	2	\$ levél fólió száma
levél típus leírás	levél	\$ levél típus leírás
levél leírás	Fekete tintairás a fóliók mindkét oldalán. A második fólió <u>verzóján</u> [személy] [@ Móricz Zsigmond @ KOHA_AUTH:120256] Móricz Zsigmond [személy vége] utólagos, kék ceruzairású datálása, valamint a levélben említett versének két szakasza fekete tintairással, javításokkal.	\$ levél leírás
levél szélessége	143	\$ levél szélessége
levél magassága	220	\$ levél magassága
levél kézírás/gépirás	K	\$ levél kézírás/gépirás

[20. ábra] Táblázat a metaadatok leírására.

Látod [személy] [@ Móricz Zsigmond @ KOHA_AUTH:120256] Zs [személy vége]. én egy fél óra, – nem [felsőindex:3]/[alsóindex:4] alatt mennyit tudok írni, és oly ritkán írtok. Hogy kérjek – kéni szabad, adni nem muszály – a dok.val küldj 30 ^Krt^ ×Koronát× , tedd bele, kellene vennem egy pár harizsnját, a lábom ^ujján^ ×ujján× szétment már ez egészen. Tehát [személy] [@ Móricz Sándor @ KOHA_AUTH:313739] Sándorék [személy vége] ne menjenek inasnak, [...] [---] **Ezt a megjegyzést** [személy] [@ Pallagi Gyula @ KOHA_AUTH:121464] Gy [személy vége]. ^b.^ ×bátyám× **tétette oda.** ^Ép^ ×Épp× **most jött, kérdezte mit írok, mondtam, hogy** [személy] [@ Móricz Sándor @ KOHA_AUTH:313739] S.ékról [személy vége], **hát felfogadtatta velem, és ide íratva, hogy ti is lássátok. Különbem magamtól jöttem rá erre.**³⁴ lelkiismeretesen fogom őket minden vizsgára elkészíteni a ^gymn.ból.^ ×gymnáziumból.× Ezt azért tettem hozzá, hogy magamat is megszilárdítsam ebben a [személy] [@ Bessenyei György @ KOHA_AUTH:102170] Bessenyei [személy vége] féle „Jámbor szándék”ban.³⁵ Ó akadémiát sürgetett, én az öcsém tanítását. Ha mi tanulunk, nekik is ^ép^ ×épp× oly joguk van ahhoz. és **nem szép volna szep** tőlünk, ha elnéznénk, hogy az öcsünk neveletlen inasok legyenek. Ha tanultak, úgy **lehetnek megélnék** bármily kenyér kereseti ágon. Szervus. Elég már a lelkesedésből.

[21. ábra] Részlet egy levél leírásából.

A projektek előrehaladtával bebizonyosodott, hogy a köztes platform használata túl sok hibalehetőséget rejt a szövegtranszformáció során. A Microsoft Word-fájlokból nem lehet egy lépésben TEI XML-eket kinyerni, csak bonyolult, többlépcsős folyamattal. A Microsoft Word-ben történő átalakítás Visual Basic (Visual Basic for Applications) kóddal és reguláris kifejezések segítségével végeztem, majd a kinyert (még nem hierarchikus) XML-eket Oxygen XML Editor program segítségével alakítottam XML kóddá a projekt számára írt egyedi stíluslap segítségével. A Microsoft Word-ből való átmásolás problémás a karakterkódolás miatt is (idézőjelek, rövid és hosszú kötőjelek keveredése), nem beszélve a hibás jelölésekről, gépelési hibákról. A többféle ellenőrző program futtatása, hibajavítások, újabb ellenőrzések beiktatása jelentősen megnövelték a szövegtranszformációra fordított időt. Jellemző hibajelenség volt például, amikor a kutató megjelölt egy szövegrészt és akaratlanul is a kijelölésbe került a bekezdéshatár, így a transzformáció során egymást átfedő hierarchiák jöttek létre, amelyet a Microsoft Word belső, egyébként XML alapú specifikációja engedett, a TEI XML azonban nem. A bonyolult ellenőrző algoritmusok ellenére is történt adatvesztés, ezeket a kutatócsoport és a DigiPhil munkatársai a Word fájlokból állították helyre. Egyértelmű tanulságnak tűnt, hogy a projekt folytatása során csak szabványos környezetben történhet mind a levelek, mind a naplók kódolása a célnak megfelelő eszközzel. A hibák kiküszöbölésére a DigiPhil csapata új leíró környezetet fejlesztett az Oxygen XML Editor programban, amely felváltotta a Microsoft Word-ot mint adatbeviteli felületet.

A DigiPhil a virtuális kutatókörnyezetet először az Arany János kritikai kiadások digitalizálási projektjén tesztelte.¹⁷⁹ A környezet kialakítása és tesztelése után a Móricz-kutatócsoport is áttért az új környezet használatára a levelek leírásához.¹⁸⁰ A kutatókörnyezet a Microsoft Word-del szemben számos előnnyel rendelkezik. Azon túl, hogy eleve szabványos kimenetet biztosít, számos hibalehetőséget is megelőz. A kutatókörnyezet szintaktikai ellenőrző algoritmusokat tartalmaz, amelyek figyelmeztetnek a formalizálható szintaktikai hibákra. A DigiPhil a Microsoft Word-fájlok átalakításából átörököltette a köztes, kevés hierarchiát tartalmazó XML struktúrát a levelek leírásához. Ezekből a „sík” XML-ekből a stíluslapok segítségével állítottuk elő a publikálásra szánt TEI XML-eket. A kutatókörnyezet használatához

¹⁷⁹ Nem egyedülálló, a digitális filológiában elterjedt Oxygen szoftverhez való kutatókörnyezet fejlesztése, a Berlin-Brandenburg Academy of Sciences and Humanities TELOTA projektje a 2013-as TEI konferencián mutatta be az általuk fejlesztett környezetet (framework). Stefan DUMONT és Martin FECHNER, „Bridging the Gap: Greater Usability for TEI Encoding”, *Journal of the Text Encoding Initiative [Online]* December 2014 – December 2015, Issue 8. sz. (2015), <https://doi.org/10.4000/jtei.1242>. Ennek a forráskódját elérhetővé tették egy nyilvános GitHub repozitóriumban: <https://github.com/telota/ediarum>.

¹⁸⁰ A kutatókörnyezetet Metzger Réka fejlesztette az előre meghatározott XML specifikáció alapján.

elég az Oxygen szerkesztőt egyszer telepíteni és importálni a leíráshoz fejlesztett komponenseket. A levelek leírása során új jelenségek bukkantak fel, amelyek kódolására új jelölőket kellett bevezetni, illetve ezeknek a környezetbe való beillesztését a környezet kis módosításával elvégezni. A kutatói környezet egy másik kiemelkedő tulajdonsága, hogy minimalizálja az adatvesztés kockázatát, és ezzel együtt biztosítja a kutatócsoport számára, hogy különböző munkaállomásokon dolgozzanak, a környezet ugyanis összeköttetésben áll egy változáskövető szerverrel, amelyre csak szintaktikailag helyes fájlokat menthetett a kutatócsoport. A DigiPhil meghatározott időközönként archiválta a fájlokat egy repozitóriumban, ahonnan a változáskövető szerver esetleges leállása esetén is visszaállíthatók voltak a fájlok.

A kutatókörnyezet kialakításakor a DigiPhil figyelembe vette azt az igényt, hogy a környezetet felhasználóbarát, irodai szoftvereket imitáló grafikus megjelenítéssel rendelkezzen, amely hasonlít a már megszokott Microsoft Word környezethez (menürendszer, magyar feliratú gombok), melynek segítségével a megfelelő jelölők automatikusan a kijelölt szöveghelyre kerültek.

Bár a kutatócsoport már a Microsoft Word-ben való leírás során is végzett adatgazdagítást, az automatikus átalakításhoz kidolgozott szintaxis rendkívül bonyolultnak bizonyult, a legtöbb szintaktikai hiba, elgépelés itt jelentkezett, ami jelentős utólagos ellenőrzést és javítást igényelt a szövegtranszformáció elvégzése után. A megfelelő elem beillesztése után a kutatókörnyezet automatikusan létrehozta a kitöltendő mezőket, illetve előre kitöltötte a névterek DigiPhil-ben használt azonosítójának prefixét.

Az új kutatókörnyezet kialakításán túl átalakult a levelek (és a hozzájuk tartozó XML-ek) ellenőrzésének folyamata is. Ahhoz, hogy a lektorálást végző kutatók is hozzáférjenek az átiratokhoz, anélkül, hogy a kutatói környezet használatát meg kelljen tanulniuk,¹⁸¹ olyan, ExistDB alapú online megjelenítő felületet fejlesztett a DigiPhil, amely alkalmas volt az egyszerű XML megjelenítésére, a kutatói környezetben megszokott vizualizációval, valamint lehetővé tette a fájlokban való kereséseket. Így a kiadást készítő kollégák és a lektorok ugyanolyan megjelenítési formában találkoztak az átiratokkal, ami jelentősen megkönnyítette az együttműködést. A tesztfelületen történő ellenőrzés után a kutatócsoport a javításokat a

¹⁸¹ Bár a DigiPhil törekedett a minél inkább felhasználóbarát környezet kialakítására, a szoftver sajátosságaiból fakadóan így is időbe telt, mire egy-egy kutató magabiztosan használta a felületet. Emellett, mivel az Oxygen XML Editor előfizetéshez kötött, nem elvárható, hogy egy projekt az összes résztvevő számára biztosítsa a megfelelő licenst.

kutatói környezetben végezte el, majd ezt követően alakítottuk át a fájlokat szabványos TEI XML fájllokká.

A projekt finanszírozása 2019-ben lezárult, ekkor tette közzé a DigiPhil az elkészült kritikai kiadás „kézirátát”, a 0.9-es verziót. A lektorálási folyamat és a javítások ezt követően is folytatódtak. Látható, hogy bár a projekt második felében a kiadás már szabványos környezetben zajlott, továbbra is jelen volt egy köztes réteg, az egyszerűsített XML formátum. Ez további átalakítási feladatokat igényelt a publikálás előtt, lassítva ezzel a javítások közzétételét. A DigiPhil szakmai vezetése úgy döntött, hogy szükséges egy olyan környezet fejlesztése, amelyben már a TEI sémának és a DigiPhil saját specifikációjának megfelelő jelölőkkel, közvetlenül lehet az átírást végezni. A közel egy évig tartó fejlesztési folyamat eredményeképpen elkészült az első olyan magyarországi kutatókörnyezet, amely lehetővé teszi, hogy billentyűparancsok és gombok használatával TEI XML átírat készüljön, ezzel megújult a DigiPhil közreműködésével készülő kritikai kiadások munkafolyamata is. Az egyes kiadásokhoz testre szabjuk a kutatókörnyezetet, így az csak azokat a jelölőket tartalmazza, amelyek az adott kiadás elkészítéséhez szükségesek. A kritikai kiadás első, lektorált verzióját 2021 februárjában tettük közzé, ebben a változatban már minden technikai és szövegközlési hibát javítottunk, ekkorra készült el a levelezéskiadáshoz fejlesztett keresőfelület is. 2021 májusában megszűnt az együttműködés a Petőfi Irodalmi Múzeum és az Irodalomtudományi Intézet között, ekkor került a kritikai kiadás és a forráskiadás a PIM-ben működő Digitális Bölcsészeti Központba, a megállapodás értelmében a DigiPhil eltávolította a kiadásokat a szolgáltatásból.¹⁸²

***A born digital* kritikai kiadások térnyerése**

A Móricz kiadások úttörő jellege és a DigiPhil fejlesztéseinek hatása Intézményi keretek között is mérhető. Az Irodalomtudományi Intézet vezetősége döntésének értelmében az Intézetben készülő összes új kritikai kiadást eleve szabványos környezetben állítják elő, tehát a készülő kiadások *born digital* kiadások lesznek. Elsőként a nagy múltra visszatekintő sorozat, a *Bibliotheca Scriptorum Medii Recentisque Aevorum* sorozat fordult a *born digital* kiadások felé. A szabványos formátum használatát az Intézet vezetősége már régóta szorgalmazta, azonban akkoriban még nem állt rendelkezésre a megfelelő infrastruktúra. Az első ilyen kísérlet még nem a DigiPhil szakmai vezetésével indult, az ott szerzett tapasztalatokat azonban a

¹⁸² CSÉVE, „Előszó”.

DigiPhil projektben is hasznosítottuk. Szilágyi Emőke Rita 2013-ban kezdte meg Oláh Miklós levelezésének kritikai kiadásának előkészítését. A kezdetektől az a megfontolás vezette a szerkesztőt, hogy a kiadást egyaránt publikálja online és nyomtatott formában.¹⁸³ Az első kötet előkészítő munkálataihoz a Classical Text Editor célszoftvert választotta abban a reményben, hogy az eszköz segítségével egyszerre készülhet el a kiadás TEI XML átirata és a nyomdai PDF formátum.¹⁸⁴ A szerkesztőbizottság és az Intézet is nagy reményeket fűzött az eszközhöz, az Intézet több licenst is vásárolt. A kiadás elkészültét követően kezdődött meg az együttműködés a szerkesztő és a DigiPhil között. Az eredeti tervek szerint a szoftverből exportáltuk volna a TEI XML formátumot és azt, a DigiPhil specifikációjával való összehangolást és az adatgazdagítást követően tettük volna közzé a DigiPhil oldalán. Az exportot követően a kimenet vizsgálata során azonban kiderült, hogy a szoftver által előállított XML formátum nem csak hogy nem felel meg a TEI ajánlásnak, hanem a szemantikai jelöléseket sem tartalmazza. A rendszer XML kimenete kizárólag formázási információkat tartalmazott, a szerkesztő által jelölt filológiai jellemzők, nem szerepeltek az adott szöveghelyeken, hanem a rendszerben előállított PDF-hez hasonlóan lábjegyzetek formájában tartalmazta az információt. Az XML formátum vizsgálata során világossá vált, hogy nem lehet egyértelműen dekódolni a jelöléseket és automatikusan egy olyan TEI XML-t előállítani, amely megfelel a DigiPhil specifikációjának és a digitális filológia elvárásainak. Mivel a kiadást támogató pályázat feltételei között szerepelt az online kiadás közzététele, a helyzet megoldására félautomatikus módszerekkel létrehoztunk a levelekből egy olyan, egyszerű TEI XML formátumot, amit közzétettünk a DigiPhil oldalán. Ezzel azonban egy olyan kiadás jött létre, amely valójában érdemben nem különbözik a nyomtatott kiadás PDF reprezentációjától. A filológiai jelenségek megfelelő átvezetése a fájlokba a második kötet előkészítésével párhuzamosan zajlottak egy, már a sorozat igényeire fejlesztett kutatókörnyezetben.

A közös munka és az Intézeti döntés hatására a BSMRAe új kiadásai már eleve egy testreszabott környezetben készülnek, jelenleg tíz, különböző fázisban lévő kritikai kiadás munkálatai zajlanak az együttműködés keretében. A BSMRAe kiadások DigiPhil-be való integrálása hatással volt a DigiPhil XML specifikációjára is, míg a digitális filológia lehetőségei megváltoztatták a készülő kritikai kiadások szerkesztési elveit. Az egyik legjobb példa erre a

¹⁸³ Emőke Rita SZILÁGYI, „Editorial Procedures and Types of Censure: On the Upcoming Critical Edition of Nicolaus Olahus’ Correspondence”, in *Nicolaus Olahus 450: Proceedings of the International Conference on the 450th Anniversary of Nicolaus Olahus’ Death*, szerk. Emőke Rita SZILÁGYI, 193–203 (Wien: Institut für Ungarische Geschichtsforschung in Wien, Balassi Institut Collegium Hungaricum, Wien, Ungarische Archivdelegation beim Haus-, Hof- und Staatsarchiv, Wien, 2019).

¹⁸⁴ A Classical Text Editor használata előfizetéshez kötött.

sajtó alá rendező emendációinak jelölése.¹⁸⁵ A sorozat nyomtatásban megjelent kötetekben hagyományosan nem jelölték azokat a típusú emendáció csoportokat, amelyeket rögzítettek a kiadás előszavában. Így bizonyos esetekben azok rejtve maradtak az olvasó elől. A szerkesztőbizottság ülésein ezeket az elveket kiadásonként döntötték el, de a jelölőnyelvi átiratban lehetőség nyílik minden emendáció szöveghelyen való jelölésére, így gyakorlatilag egyszerre közölhetnek betűhív és emendált átiratot a sajtó alá rendezők.

Az eddigieknél talán még érzékletesebben bemutatható az az egymásra ható folyamat, amelyet a BSMRAe kapcsán már érintettem. 2021-ben kezdődött az NKFIH támogatásával Babits Mihály verseinek és műfordításainak kritikai kiadását előkészítő projekt. A projekt célja, hogy Babits 1911 és 1920 között keletkezett verseinek, valamint az Isteni színjáték fordításának kritikai kiadása nyomtatott és digitális formában megjelenjen. A versek kritikai kiadás-sorozatának első kötete 2017-ben jelent meg,¹⁸⁶ a második kötet 2022-ben.¹⁸⁷ Az első kötet, az Ady kritikai kiadások mintájára, „felépítése, szövegkritikai koncepciója és gyakorlata az úgynevezett hagyományos kritikai kiadás elveit követi,¹⁸⁸ tehát a sajtó alá rendezők a főszöveg megállapítása után a vers közlését követően táblázatos formában rögzítették a szövegeltéréseket.¹⁸⁹ Kelevéz már az első kritikai kiadás előszavában felhívja a figyelmet arra, hogy Babits írói gyakorlata sokat változott, az 1910-es évek közepétől sokkal több töredékes kézirat, fogalmazvány áll rendelkezésre, amely a jövőbeni kötetek szöveggözlési elveire is hatással lesz.¹⁹⁰ A második kötetben Kelevéz már elmozdul a hagyományos kritikai kiadás logikájától a genetikus kiadási gyakorlat felé, az előszóban így fogalmaz: „A genetikus szemlélet hatására az egyetlen főszöveg alá beosztott előzmények merev rendszerét korszerű szempontokkal igyekeztem kiegészíteni [...] Ugyanakkor túl nagy áldozatokkal járt volna genetikus módon mindig teljesen egyenrangúként kezelni az összes szövegváltozatot, ugyanis éppen Babits e korszakának hevesen vitatott versei esetében nem feledkezhetünk meg a döntő különbségről, mely a csak kéziratban létezett korai, illetve a rendkívül viszontagságos

¹⁸⁵ Enikő BÉKÉS és Sándor BENE, „Módszertani ajánlás a Bibliotheca Scriptorum Medii Recentisque Aevorum új sorozatának (Series Nova) szöveggondozásához”, *Irodalomtörténeti Közlemények* 118, 5. sz. (2014): 698–713, 705.

¹⁸⁶ Ágnes SOMOGYI és Zoltán HAFNER, szerk., *Babits Mihály összes versei, 1, 1890–1905* (Budapest: Argumentum – MTA BTK Irodalomtudományi Intézet, 2017).

¹⁸⁷ Ágnes KELEVÉZ, szerk., *Babits Mihály összes versei, 2, 1906–1910* (Budapest: Argumentum, 2022).

¹⁸⁸ Ágnes KELEVÉZ, „A Babits-versek kritikai kiadásának sorozata elé”, in *Babits Mihály összes versei, 1, 1890–1905*, szerk. Ágnes SOMOGYI és Zoltán HAFNER, 13–47 (Budapest: Argumentum, 2017), 28.

¹⁸⁹ Ahogyan a kritikai kiadás előszavában is fogalmaznak, a szöveggözlés elvei követik az *Ady Endre Összes versei* kritikai kiadások elveit, azzal a különbséggel, hogy az Ady kiadások esetén a szövegkritikai jegyzetek a kötet végén, egybefüggő szöveggé helyezkedtek el. Uo.

¹⁹⁰ Uo., 29.

hatástörténet részévé váló, publikált változatok között mutatkozik.”¹⁹¹ A harmadik kötet sajtó alá rendezői, 2019-ben publikálták az új kötetek textológiai elveit bemutató értekezésüket.¹⁹² Major Ágnes és Buda Attila koncepciója az volt, hogy az „első kötet szövegeitől eltérően a kéziratos forrásanyag a befejezettségéből a köztes szövegállapot felé mozdul el, a fogalmazványok és a töredékek nagy száma következtében nem lehet oly mértékben egy végleges változat egyedüli közlésére és annak kizárólagos szövegkritikájára építeni”.¹⁹³ Ezt a köztes szövegállapotot úgy akarták megjeleníteni, hogy az alapszöveg kiválasztásának gyakorlatát megváltoztatták volna; a keletkezés folyamatából kiragadva két szövegállapotot: az nyomtatott megjelenés előtti utolsó tisztázatot, vagy annak hiányában a legutolsónak tekinthető kéziratot és az *ultima editio* alapján megállapított szöveget. Ezzel a módszerrel tulajdonképpen két alapszöveget állapítottak volna meg. Emellett a kötetek sajtó alá rendezői az alapszövegek alapján olvasói szövegeket közöltek volna. A kutatócsoport 2019-ben kezdte meg az együttműködést a DigiPhil-el. Az együttműködés kezdetére kidolgozott szövegközlési tervet a Textológiai Bizottság 2020. december 16-i ülésének döntése alapján több szempontból is újra kellett gondolni. A Bizottság három fókuszpontot jelölt ki, amelyek mentén az új közlési terv kialakítását elvégeztük. A nyomtatott kiadás kapcsán kiemelték, hogy az olvasói szöveg fogalmát pontosítani kell, illetve a szövegkiadási intervenció mértékén csökkenteni kell,¹⁹⁴ illetve a genetikus jelkészletet túl bonyolultnak ítélték, kérték annak egyszerűsítését. Emellett a közlési tervet össze kellett hangolni a DigiPhil gyakorlatával. Az új közlési terv kidolgozásakor visszatértünk az alapokhoz, nem őriztük meg az első terv elveit. Az első két kritikai kiadás elveit alapul véve, a genetikus szemlélet érvényesítése volt a fő cél. Elsőként a digitális kiadás specifikációját készítettük el, a nyomtatott kiadás közlési tervét ehhez igazítottuk. A kritikai kiadás sajtó alá rendezői minden vers esetében egy főszöveget állapít meg, a csupán kéziratban, illetve gépiratban meglévő versek esetében az *ultima manus*, a nyomtatásban megjelent versek esetében az *ultima editio* elvét követve, a szövegközlés betűhű lesz. A nyomtatott kiadásban a versek kronológiai sorrendben fogják követni egymást. A nyomtatásban megjelent versek szövegközlése során az eltérések lábjegyzetben szerepelnek majd. Ezt követi az előszövegek közlése, az egyes előszövegek közötti eltérések széljegyzet

¹⁹¹ KELEVÉZ, *Babits Mihály összes versei, 2, 1906–1910*, 56–57.

¹⁹² MAJOR Ágnes és BUDA Attila, „A Babits-versek kritikai kiadásának átalakulása, avagy a kéziratos hagyaték változásának következményei”, *Irodalomtörténet* 100, 4. sz. (2019): 435–447.

¹⁹³ Uo., 440.

¹⁹⁴ Az eredeti terv szerint az olvasói szöveg nagy mértékben „javított” változatot közölt volna, nem őrizték volna meg például Babits írásmódjának sajátosságait, tehát bizonyos esetekben modern helyesírás szerint módosították volna a magán- és mássalhangzók hosszúságát, ezzel módosítva a versek poétikai jellemzőit.

formájában lesznek feltüntetve. A kiadás tipográfiája újszerű lesz, kihasználva a modern nyomdatechnológiát, a filológiai jelenségeket színekkel, szimbólumokkal jelenítjük majd meg.

A digitális kiadás alapja a jelölőnyelvi átírat, a Babits-kiadások esetében is meg kellett határozni a digitális objektumot. Ahhoz, hogy minél inkább érvényesíteni tudjuk a genetikus szemléletet, három típusba soroltuk az objektumokat. Az első típusba a nyomtatásban megjelent versek tartoznak, ezeknél a verseknél a TEI XML fájl tartalmazza a megállapított főszöveget és a szövegeltéréseket. A második típusba azok az előszövegek tartoznak, amelyek általában tisztázatai vagy gépiratai a műnek, Babits általában kisebb módosításokat végzett csak. A harmadik, és talán digitális filológiai szempontból legérdekesebb csoportba azok a fogalmazványok, verstördékek tartoznak, amelyekben oly mértékű szövegeltérések találhatók, hogy nem lehet egyértelműen egy-egy műhöz kapcsolni, vagy egy XML fájlban reprezentálni. Erre hoznám példaként a negyedik, Szénási Zoltán által sajtó alá rendezett kötetben megjelenő a *Jóság dala* című vers digitális reprezentációját. A verset a Nyugat közölte először 1917 augusztus 1-jén,¹⁹⁵ majd Babits életében megjelent a *Nyughatatlanság völgyében*,¹⁹⁶ a *Versekben*,¹⁹⁷ az *Összes verseiben*,¹⁹⁸ valamint a *Válogatott versekben*.¹⁹⁹ Egyrészt tehát egy TEI XML fájl tartalmazza a megjelenéseket és az egyes megjelenések közötti eltéréseket az adott szöveghelyen rögzítve:

```
<l>Most csönd van és <app>
    <lem wit="#KK #V #OV #VV">szívem</lem>
    <rdg wit="#Ny #NyV">szívem</rdg>
</app> mint a süket föld, erjedő, sötét, meleg,</l>
```

[22. ábra] Részlet a szövegváltozatok jelöléséből.

A vershez két előszöveg tartozik, **k1**²⁰⁰ a vers 2. részének ceruzairású fogalmazványa, **k2**²⁰¹ a vers 4. részének, szintén ceruzairású fogalmazványa. Mivel a fogalmazványok, amelyek maguk is töredékesek, csupán a vers részleteit tartalmazzák, ezért nem lehetett őket egy XML átíratban

¹⁹⁵ BABITS Mihály, „A jóság dala (Versek, darócban.)”, *Nyugat*, 15. sz. (1917. augusztus 1.): 175–179.

¹⁹⁶ BABITS Mihály, *Nyughatatlanság Völgye* (Budapest: Táltos, 1920), 35–39.

¹⁹⁷ BABITS Mihály, *Versek 1902–1927* (Budapest: Athenaeum, 1928), 281–284.

¹⁹⁸ BABITS Mihály, *Babits Mihály összes versei. 1902-1937*, Babits Mihály összegyűjtött munkái, 1. (Budapest: Athenaeum, 1937), 256–258.

¹⁹⁹ *Babits Mihály válogatott versei*, Atheneum olvasótábora (Budapest: Athenaeum, 1941), 47–50.

²⁰⁰ OSzK Fond III/1969. 23. fólió verzóján.

²⁰¹ OSzK Fond III/1969. 79. fólió verzóján.

reprezentálni. A kéziratokon ezen felül olyan filológiai jelenségek is találhatóak, amelyekre a DigiPhil gyakorlatában nem volt korábban példa. Például a **k1C**-en szereplő töredékben (**k1C**), amely a vers 2. rész 3. versszakának fogalmazványa, két versszakra bontva. A töredék második versszaka több helyen hiányos, a harmadik sort Babits áthelyezte az első versszak ötödik sorába. Ilyen jellegű áthelyezéseket korábban nem jelöltünk, ezért ebben az esetben új jelölési rendszert kellett kidolgozni, az eredeti helyén a törlési aktus jelölésekor azonosítóval láttuk el (@xml:id), míg az új helyén a beszúrási aktusnál erre az azonosítóra utaltunk vissza egy új attribútumban (@sameAs). A nyomtatott kiadásban színkódok fogják jelölni ezeket az áthelyezéseket.

```

<l>Már elhagy, elhagy az ifjuság,</l>
<l>a jóság, szépség és hiuság</l>
<l>kis érzésmagvak, csúf<del rend="overtyped">,</del><add type="corr"
  >-</add>makacsok</l>
<l>sugják hogy rossz, hogy öreg vagyok!</l>
<l><add sameAs="#move.1" type="insert"/>A szívemnek csúf salakja</l>
<l>Ó istenem, hát ez a világ</l>
<l>akinek van, nem irigykedik</l>
<l>könnyű jó lenni az ifjuncak </l>
</lg>
<lg>
  <l>Irigység, önzés, rossz akarat</l>
  <l>ó istenem hát <gap extent="unknown" reason="damaged" unit="char"/></l>
  <l>
    <del xml:id="move.1">A szívemnek csúf salakja</del>
  </l>
  <l>Hát <gap extent="unknown" reason="damaged" unit="char"/> minden erő csak <gap
    extent="unknown" reason="damaged" unit="char"/></l>
  <l>minden nemesség csak <gap extent="unknown" reason="damaged" unit="char"/></l>
</lg>

```

[23. ábra] Részlet **k1C** jelölőnyelvi átíratából.

A DigiPhil gyakorlatában korábban kerültük az úgynevezett stand-off annotációk és az úgynevezett összekapcsolt horgonyelemek használatát. Azon túl, hogy el akartuk kerülni a szöveg karaktereinek indexén alapuló hivatkozási rendszer bevezetését,²⁰² a DigiPhil által használt megjelenítő szoftver sem volt alkalmas az ilyen típusú annotációk vizualizálására,

²⁰² A stand-off annotációk az XML-ek esetében lehetnek szegmentumokra mutató azonosító alapúak is, ám ez a gyakorlat kevésbé elterjedt.

illetve az általunk fejlesztett TEI alapú kutatókörnyezetek nem voltak felkészítve az ilyen típusú annotációk rögzítésére. A Babits-kiadások esetében azonban nem tudtuk elkerülni ezek használatát. Egymásra utaló horgonyelemek bevezetésére azért volt szükség, mert számos kézirat esetében nem lehetett másképp jelölni a vers nagyobb szegmentumát érintő törléseket (pl. a vers egy részének, vagy egészének áthúzását), különösen abban az esetben, amikor két vagy több szövegváltozatot rögzítünk egy XML fájlban. Ezt a TEI ajánlásával összhangban a delSpan és anchor elemekkel jelöljük, kiegészítve egy-egy attribútummal, ami, hasonlóan az áthelyezéshez, meghatározott szintakszis szerint egyedi azonosítókat tartalmaz. Ahhoz, hogy ne keltsük azt az illúziót, hogy a törlés mindkét szövegen megtalálható, azt is specifikálni kellett, hogy a törlés mely kézirathoz kapcsolódik, ezt a @source attribútumban, a szövegtanú azonosítójára történő hivatkozással jelöljük. A stand-off annotációk bevezetésére a töredékek és a teljes kéziratok szemantikus összekapcsolása miatt volt szükség. Minden fájl esetében a verssorok egyedi azonosítót kapnak, amelyet a digitális objektum állandó azonosítójából és a sorszámból generálunk (pl. bm-lira4.regi-kert.gen.tei#1.4). A stand-off annotációk segítségével a töredékek minden sorát egyértelműen össze lehet kapcsolni egy másik XML fájlban található vers adott sorával, az egyszerű linkelésnél több információ feltüntetésével.

Az új jelölők bevezetése miatt a kutatókörnyezetet is sokkal kifinomultabbá kellett tenni, és, bár a keretrendszer alkalmas a jelölők használatára, valamint automatikus módszerekkel is ellenőrizzük a jelölőnyelvi átiratokat, ez a jelölési metódus a kutatóktól is nagy odafigyelést követel meg.

A sorozat már megjelent köteteit szintén közzétesszük a DigiPhil oldalán, az első kötet informatizált, javított és bővített második kiadása már elérhető a szolgáltatásban, az adatgazdagítás jelenleg is zajlik. A második kötet informatizálását is megkezdtük. Egy digitális kiadás esetén elvárás, hogy a jelölőnyelvi átiratok mellett a kéziratok fakszimilái is elérhetőek legyenek, a Babits kutatócsoport, Kelevéz Ágnes felvetésére úgy döntött, hogy nem csak a kéziratok, hanem a folyóiratos és kötetbeli megjelenésekről készített fotókat is elérhetővé teszik, hogy láthatóvá váljanak a kutatók számára az egyes közlések közötti tipográfiai különbségek is.

Ahhoz, hogy ezt, az európai viszonylatban is kiemelkedőnek számító kritikai kiadás-sorozatot ilyen kifinomult jelölőnyelvi átiratban közzé teheszük, számos fejlesztésre volt szükség mind a DigiPhil specifikációjában, mind az infrastruktúrájában. Az eddig tárgyalt metódusok válaszul szolgáltak olyan kérdésekre, amely az „analóg” filológusokban felmerül a digitális publikálás kapcsán. Azonban nem tekinthetünk el az elmúlt évtizedben kialakult új digitális technológiáktól, amelyek a digitális filológia korábban megválaszoltak hitt kérdéseit

új fénybe helyezik. Az új technológiák új látásmódot igényelnek, új kérdéseket vetnek fel a digitális filológia elméletében és gyakorlatában, a következőkben ezeket a kérdéseket vizsgálom.

Digitális filológia és a szemantikus web

Az elmúlt évtizedekben a technológia dinamikus fejlődése új lehetőségeket teremtett a digitális filológia területén is, a nemzetközi gyakorlatban sokáig mégsem vált általánossá ezeknek a lehetőségeknek a kiaknázása. Barbara Bordalejo 2018-ban megjelent tanulmányában egyenesen vitatja, hogy egyáltalán létezik-e olyan, hogy digitális tudományos szövegkiadás.²⁰³ A szerző vitatja, hogy a számítógép, illetve a digitális módszerek alkalmazása forradalmi változásokat hoztak volna a tudományos szövegkiadások gyakorlatába, tételmondata: „nem létezik olyan dolog, hogy tudományos szövegszerkesztés.”²⁰⁴

A szerző módszeresen halad a digitális filológia két nagy hatású gondolkodójának, Patrick Sahle-nak és Elena Pierazzo-nak az elméletein és sorra cáfolni próbálja a világhálón elérhető kiadások alapján azokat. Míg Pierazzo úgy fogalmaz, hogy „[a] tudományos szövegszerkesztési gyakorlat mély strukturális és elméleti változásokon megy keresztül”²⁰⁵, addig Bordalejo vitatja, hogy megváltozott volna a szövegek vizsgálata, csak mert számítógépes eszközöket használunk a szövegkritikai és szövegszerkesztői munkavégzés támogatására. A szerző úgy látja, hogy a digitális médium mellett szóló érvek, mint a hozzáférhetőség, vagy a korlátlan rendelkezésre álló tárhely, csupán mítoszok, amelyek valójában nem hozták el a várt radikális változásokat. Bordalejo, kissé naivan úgy fogalmaz, hogy „ugyanazokhoz az anyagokhoz férhetünk hozzá, amelyekhez mindig is hozzáférhettünk, csak gyorsabban,”²⁰⁶ és valójában ezt a mítoszt az anglofón és északi országok terjesztik, ahol rendelkezésre áll a megfizethető és gyors internetelérés. Az természetesen igaz, hogy a hozzáférhetőséget befolyásolja a rendelkezésre álló internetelérés, de a finanszírozhatóságtól való függés kérdése ugyanúgy igaz a kéziratokat őrző memóriaintézmények látogatására is. Bordalejo érvelésének minden pontján hangsúlyozza, hogy a digitális médiumban alkalmazott szerkesztői gyakorlatok mindegyike támaszkodik a nyomtatott hagyományban kialakult gyakorlatokra. Ilyen például a Sahle által megfogalmazott, digitális szemléletű szövegszerkesztés egyik kulcsfontosságú ismérve, a hipertextualitás,²⁰⁷ amely Bordalejo szerint semmivel sem forradalmibb, mint a nyomtatott médiumban lábjegyzetek, glosszák és hivatkozások formájában használt utalások, sőt, azt, az általánosan elfogadott gondolatot is

²⁰³ Barbara BORDALEJO, „Digital versus Analogue Textual Scholarship or The Revolution is Just in the Title”, *Digital Philology: A Journal of Medieval Cultures* 7, 1. sz. (2018): 7–28, <https://doi.org/10.1353/dph.2018.0001>.

²⁰⁴ Uo., 24.

²⁰⁵ Idézi Bordalejo, Uo., 8.

²⁰⁶ Uo., 10.

²⁰⁷ SAHLE, „What is a Scholarly Digital Edition”, 29.

megkérdőjelezi, hogy a „digitális kiadásokban hirtelen számtalan befogadási útvonal létezne. Úgy tűnik, hogy ilyen utak mindig is léteztek a különféle kiadásokban.”²⁰⁸

Bár Bordajelo következtetései első pillantásra helyesnek tűnhetnek, azonban ha a távolabbról vizsgáljuk a kérdést, hamar világossá válik, hogy valójában téved. Arra már Andreas Kuczera is rámutatott, hogy az új módszertani lehetőségek megjelenésével, a papírtól, mint médiumtól való függetlenedés révén rengeteg új interpretációs megközelítésmódra nyílik lehetőség és ez ellentmond Bordajelo kijelentésének.²⁰⁹ Kuczera nem vizsgálja, hogy Bordajelo miért téved, azonban érdemes megvizsgálni, hogy mi az oka annak, hogy digitális filológusként Bordajelo úgy látja, hogy semmilyen forradalmi újítást nem hozott a digitális technológia a filológia és a textológia területére. A digitális tudományos szövegkiadások kapcsán két nagyon fontos szempontot nem vesz figyelembe, a struktúra és az adatkezelés radikális megváltozását. Egy digitális kiadás struktúrája természeténél fogva nem hasonlítható egy nyomtatott kiadáséhoz. A digitális kiadások esetén a kiadás egységei a digitális objektumok, mint például a jelölőnyelvi átiratok, a faksimilék stb., a kiadás valójában ezen objektumok kapcsolati rendszere, amely különböző rendezési lehetőségeket jelent, tehát egy digitális kiadás számos, eltérő felépítésű kiadás összességként is értelmezhető. Egy digitális kiadás, legyen az informatizált kiadás, vagy digitálisan született kiadás nyomtatásban való megjelenítése során a könyv felépítése csupán egy interpretációja lesz a digitális kiadásnak. A másik szempont, amely nem független az elsőtől, az adatok kezelése. A digitális objektumok viszonyrendszerébe illeszkednek a kiadás adatgazdagítását támogató belső és külső adattárak. Adatgazdagítás során például a VIAF egy rekordja kapcsolódik a kiadáshoz, így az ott szereplő adatok is a digitális kiadás részévé válnak, míg nyomtatott kiadásoknál a kifelé mutató referenciák esetén ez a folyamat nem megy végbe, csupán abban az esetben, ha a referált adattár rendelkezésünkre áll. Az adattárak szerepe is átalakul. Míg a nyomtatott kiadások esetében a szerkesztők, ha készítettek is adatárat,²¹⁰ a feltüntetett információk csupán a kiadás kontextusában voltak értelmezhetők, míg a digitális kiadásokat támogató adattárak létrehozásakor a cél a minél nagyobb kutatói közönséget kiszolgáló, újra felhasználható adatok leírása, a kiadás

²⁰⁸ BORDALEJO, „Digital versus Analogue Textual Scholarship or The Revolution is Just in the Title”, 19.

²⁰⁹ Andreas KUCZERA, „TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs”, in *Graph Technologies in the Humanities - Proceedings 2020*, szerk. Tara ANDREWS és mtsai., 3110. köt., CEUR Workshop Proceedings (Graph Technologies in the Humanities 2020, Vienna, Austria: CEUR, 2020), <http://ceur-ws.org/Vol-3110/#paper6>.

²¹⁰ Vö. SÁFRÁN Györgyi, BISZTRAY Gyula és SÁNDOR István, szerk., *Arany János Levelezése 1.: 1828–1851*, Arany János Összes Művei 15 (Budapest: Akadémiai Kiadó, 1975). SÁFRÁN Györgyi, BISZTRAY Gyula és SÁNDOR István, szerk., *Arany János Levelezése 2.: 1852–1856*, Arany János Összes Művei 16 (Budapest: Akadémiai Kiadó, 1982).

kontextusára vonatkozó adatok csak kiegészítésként készülnek. Látható, hogy az adatgyűjtésnél megváltozik a kiadást (és az adattárat) készítő szerkesztők célja, amelyet Bordajelo hiányolt a forradalmi újdonságok közül.

Bordajelo érvelése mégsem teljesen elítélendő, hiszen az általa hivatkozott digitális kiadások nem aknázzák ki a technológia nyújtotta lehetőségeket. Ahogyan Joris van Zundert is rámutatott, amíg a digitális kiadást készítők olyan kiadásokat tesznek közzé, amelyek a könyv digitális metaforái, addig nem várható paradigmaváltás.²¹¹ Zundert nyomán, Jeffrey C. Witt és Andreas Kuczera is ezt a paradigmaváltást sürgetik. Míg Zundert (Robinsont követve) elmozdult a szöveg mint dokumentum ábrázolástól a szöveg mint mű, illetve szöveg mint folyamat ábrázolásának irányába, Witt már a szöveg mint hálózat paradigmája felől tekint a digitális tudományos szövegkiadásokra.²¹² Kuczera azonban felhívja a figyelmet, hogy bár Witt tulajdonképpen többszereplős hálózatként²¹³ – ami gráfként könnyedén leírható – gondolkodik a digitális tudományos szövegkiadásokról, nem tekinti a kiadás készítésében résztvevő szereplőket a hálózat részeként, míg Kuczera elképzelése szerint a kiadásban közreműködők minden egyéni döntése a digitális tudományos szövegkiadás, mint mű részévé válik, ezért ő már nem csak többszereplős gráfként képzel el egy kiadást, hanem olyan származási tudásgráfként [provenance knowledge graph],²¹⁴ amely tartalmazza az átiratokat, a kritikai apparátust az apparátusok közötti kapcsolatokat és az egyes állítások eredetét. Kuczera elképzelése szerint a származási tudásgráf formájában közzétett kutatási adatok így, a kutatók által hozott, szubjektív döntések halmazává válnak. Ennek következtében a felhasználó feladata, hogy a kutató tudományos profilja alapján eldöntse, bíz-e a kutató döntéseiben. Bár Kuczera itt nem reflektál arra az elképzelésre, hogy a digitális tudományos szövegkiadások egyik fontos hozadéka a tudományos demokrácia, azonban itt feltétlenül vissza kell utalnunk Bordajelo-ra, aki jól látja, hogy a várva várt demokrácia, ami egyelőre nem következett be, természetszerűen ellentmond a tudományosság kívánalmainak.²¹⁵ Hiszen egy könyv megjelentetése során még ha intézményes garancia nincs is minden esetben, a kiadói folyamat elméletben szolgál bizonyos fokú minőségbiztosítással. Ezzel szemben az online térben, elméletben, bárki közzé tehet bármit, Kuczera víziója viszont részben megoldást jelenthetne

²¹¹ Joris VAN ZUNDERT, „Barely Beyond the Book?“, in *Digital Scholarly Editing: Theories, Models and Methods*, szerk. Matthew James DRISCOLL és Elena PIERAZZO, 83–106 (Cambridge: Open Book Publishers, 2016), 105.

²¹² Jeffrey C. WITT, „Digital Scholarly Editions and API Consuming Applications“, szerk. Roman BLEIER és mtsai., 12. köt., 219–247 (Norderstedt: BoD, 2018), 222, <https://kups.ub.uni-koeln.de/9118/>.

²¹³ lsd. többszereplős gráf

²¹⁴ KUCZERA, „TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs“, 103.

²¹⁵ BORDALEJO, „Digital versus Analogue Textual Scholarship or The Revolution is Just in the Title“, 23.

ezen ellentmondás feloldására. A származási tudásgráffal kapcsolatos elképzelései a Hildegraph projekt keretei között öltöttek testet, azonban a kiadás jelenleg nem elérhető,²¹⁶ így csupán az általa jegyzett tanulmányra támaszkodhatunk. Az első és talán legfontosabb döntés az XML formátum elvetése. A Hildegraph²¹⁷ projekt keretei között megkíséreltek olyan jelölőnyelvi átíratot előállítani, amely követi a Text Encoding Initiative ajánlását, de nem XML alapon, hanem a Codex projekt során kidolgozott formátumban.²¹⁸ Az általuk használt modell alapját az úgynevezett stand-off annotációk adják. Az XML átírat használata helyett nyers szövegformátum tartalmazza az átíratokat, az apparátust pedig egy erre a célra fejlesztett szerkesztőfelület²¹⁹ segítségével stand-off annotációként illesztették a szöveg meghatározott indexeihez. A szövegek és az annotációk kapcsolatait a Neo4j gráfadatbázis segítségével modellezték. A Hildegraph projekt készítése során megkísérelték az annotációkat megfeleltetni a TEI ajánlásában megfogalmazott kategóriáknak. A rendszer hat annotációs réteget különböztet meg, amelyek egy rendszerben egyszerre léteznek: tipográfia, stílus (formázások), entitások, szintaxis, morfológia, nyelv.²²⁰ Kuczera azzal érvel, hogy ezzel a megoldással elkerülhetőek az XML-ben nehezen kezelhető, egymást átfedő hierarchiák, azonban az általa hozott példák mindegyike valójában könnyedén kódolható XML-ben is,²²¹ ha elfogadjuk, hogy az egyes annotációs szinteket nem egy XML fájlban tároljuk. Számos példát láthatunk olyan megoldásokra, ahol a kiadás alapját az XML formátum adja, a nyelvi elemzés annotációit pedig

²¹⁶ Az Internet Archive mindössze egyszer, 2021. november 7-én készített pillanatfelvételt, azonban ekkor már a kiadás nem volt elérhető: <https://web.archive.org/web/20211207042743/http://hildegraph.org/>. A dolgozat lezárásakor az átírási link szerverhibát jelez: <https://hildegard.adwmainz.de/>.

²¹⁷ A projekt Bingeni Szent Hildegárd (Hildegard von Bingen) levelezésének kritikai kiadásának közzétételét tűzte ki célul, feldolgozva a szerző életében, a *Riesen-Codex* leveleskönyvbe (*Liber Epistolarum*) rendezett változatot is.

²¹⁸ Az eredeti tanulmány a dolgozat írásának pillanatában nem elérhető: Andreas KUCZERA és Lian NEILL, „The Codex – An Atlas of Relations.”, szerk. Andreas KUCZERA, Thorsten WÜBBENA és Thomas KOLLATZ, *Sonderband der Zeitschrift für digitale Geisteswissenschaften*, 4. sz. (2019), https://doi.org/10.17175/sb004_008. [az online folyóirat nem használ paginációt] A tanulmány elérhető az egyik szerző, Lian Neill Academia.edu profilján keresztül: <https://vfg-mz.academia.edu/LianNeill>.

²¹⁹ A *Codex* projektben ugyanezt a szerkesztőfelületet, a SPEEDy-t használták a stand-off annotációk leírására. Lian NEILL és Desmond SCHMIDT, „SPEEDy. A Practical Editor for Texts Annotated With Standoff Properties”, in *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerk. Elena SPADINI, Francesca TOMASI és Georg VOGELER, 55–71 (Norderstedt: Books on Demand, 2021).

²²⁰ KUCZERA, „TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs”, 107.

²²¹ Az egyik példa jól mutatja ezt. A kódexben az egyes leveleket nagyobb betűméretű, vörös tintával jegyzett karakterek választják el, ez egyrészt strukturális jelölés, másrészt retorikai funkcióval bír, azonosítja a levél feladóját és címzettjét. A projekt résztvevői TEI szakértőkkel egyetemben arra jutottak, hogy kétféleképpen kódolható a jelenség: **emph** jelölő alkalmazása a retorikai hangsúlyosságát emelné ki, a **hi** jelölő pedig a kiemelés, azon belül is a szint jelölne. Kuczera fő problémája, hogy ha a kettős funkciót együtt szeretnénk jelölni, akkor el kell döntenünk, hogy a hierarchiában melyik jelölő foglalja magába a másikat. Valójában a szakértők által javasolt első megoldás kiváltja ezt a kérdést, hiszen az **emph** jelölőt a színre vonatkozó attribútummal látták el, további attribútumok hozzáadása is lehetséges. A strukturális jelölésre szintén alkalmas a **div** jelölő, amennyiben egy digitális objektumként (tehát egy TEI XML fájlként) tekintünk a leveleskönyvre, ennek bevezetése egyértelműen elválasztaná az egyes leveleket egymástól.

egy független TEI XML-ben,²²² vagy adott esetben más, a számítógépes nyelvészek által széles körben használt formátumban tárolják (pl. TSV).

A jelölőnyelvi kódolás elvetésén túl a tudásgráfhoz használt adatszerkezet is szokatlan választásnak tűnhet. A Hildegraph projekt során nem a szemantikus web területén elterjedt RDF (Resource Description Framework) szabványos adatsere formátumot alkalmazták, hanem egy úgynevezett címkézett tulajdonsággráfot (LPG: Labeled Property Graph).

Ahhoz, hogy megértsük a döntés előnyeit és hátrányait, ismernünk kell a szemantikus web elméletét és technológiai megvalósításait és össze kell hasonlítanunk a két adatmodellt. A szemantikus web, más néven a Web 3.0 gondolata a World Wide Web és a HTML atyjához, Tim Berners-Lee-hez kötődik, aki szerzőtársaival még 2001-ben publikálta a szemantikus web várható fejlesztéseit bemutató írásukat.²²³ A szemantikus web kifejezés egyaránt jelöl elméletet és technológiát. A koncepció alapját az a jelenség adta, hogy a Web 1.0 és 2.0 elterjedésével már a 2000-es évek elején is kezelhetetlen mennyiségű, rendezetlen adat és dokumentum vált elérhetővé a világhálón. Ebben az információhalmazban szinte lehetlenné vált kiigazodni, akkoriban is látható volt, hogy „magas szinten, eredményesen keresni egyfajta művészetnek is tekinthető”.²²⁴ A Berners-Lee-ék által felrajzolt vízió szerint a problémát intelligens gépi megoldásokkal lehet megoldani, azonban ehhez az adatoknak gépek által is értelmezhető formátumban kell rendelkezésre állniuk. Elképzelésük szerint a megoldás két részre bontható, az első, hogy az internetes erőforrásokhoz metainformációkat kell kapcsolni, a másik, hogy ezen információk segítségével a számítógép következtetéseket tudjon levonni. Ehhez kapcsolódnak szorosan az ontológiák, amelyek alkalmazása lehetővé teszi a számítógépes feldolgozást, a következtetéseket pedig az úgynevezett leíró logikai formalizmusok alkalmazása teszi lehetővé.²²⁵

Az RDF modellt gráfszerkezet számítógépes leképezésére hozták létre még 1999-ben.²²⁶ Az RDF alapú gráfok úgynevezett hármásokra épülnek, a hármások minden esetben tartalmazzák az állítás alanyát, tárgyát és magát az állítást, a gráfszerkezetben az alany és a tárgy képzik a csomópontokat, az állítás pedig az őket összekötő éleket. Az alanyt és a tárgyat

²²² Vö. HORVÁTH Péter, „A vershangzás jellemzőinek automatikus feltárása József Attila verseiben”, *Digitális Bölcsészlet*, 3. sz. (2020. július 15.): M:3-M:27, <https://doi.org/10.31400/dh-hun.2020.3.422>.

²²³ Tim BERNERS-LEE, James HANDLER és Ora LASSILA, „The Semantic Web”, *Scientific American*, 5. sz. (2001): 29–37.

²²⁴ FÜLÖP Csaba, KOVÁCS László és MICSIK András, „A metaadatsémák és a szemantikus web: egységesítés és specializáció a metaadatok világában”, *Tudományos és Műszaki Tájékoztatás* 51, 7. sz. (2004): 276–284, 276.

²²⁵ SZEREDI Péter, LUKÁCSY Péter és BENKŐ Tamás, *A szemantikus világháló elmélete és gyakorlata* (Budapest: Typotex, 2005), 9. Uo., 62.

²²⁶ Ora LASSILA és Ralph R. SWICK, „Resource Description Framework (RDF) Model and Syntax Specification”, 1999, <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/#model>.

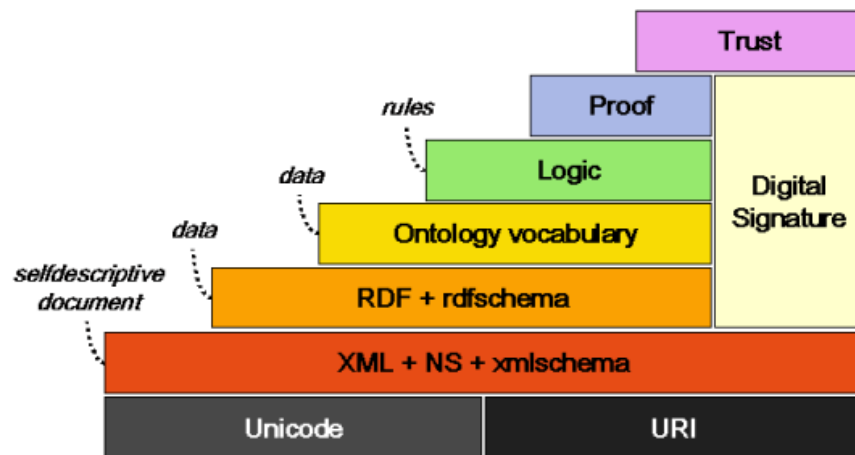
egy-egy URI (Uniform Resource Identifier) azonosítja,²²⁷ az állítás, ami pontosan meghatározza a két entitás közötti kapcsolatot szintén egy URI, ami az állítás típusát jelöli. A tárgy emellett rendelkezhet szöveges értékkel is (például dátum), azonban ebben az esetben nem tehetünk további állításokat. Az URI-k, amelyek az erőforrást azonosítják, lehetővé teszik, hogy ugyanarról az erőforrásról (azaz alanyról és tárgyról) különböző, egymástól független rendszerekben jöhessenek létre állítások. Az egyes entításokra vonatkozóan a rendelkezésre álló hármassokból az algoritmusok további következtetéseket vonnak le az alábbi módon:

1. hármass: A folyóirat [alany] címe [állítás] *A Tett*. [tárgy]

2. hármass: *A Tett* [alany] szerkesztője [állítás] Kassák Lajos. [tárgy]

Következtetés: A folyóirat szerkesztője Kassák Lajos.

A szemantikus web alapelveinek implementálására az RDF használata csupán egy eszköz, a teljes rendszer felépítését egymásra épülő, együttműködő technológiák összességéként kell elképzelni:²²⁸



[24. ábra] A szemantikus web modellje.

Az Unicode és az URI-k használata biztosítja az egységes nemzetközi karakterkódolást és erőforrásazonosítást, az XML réteg névterek és sémák alkalmazásával lehetővé teszi, hogy a szemantikus web definíciói más XML alapú szabványokba is beépülhessenek. Az RDF és az RDF Schema, ami egy konkrét megvalósítási modell,²²⁹ teszi lehetővé, hogy állításokat

²²⁷ Az elmúlt években látható, hogy az URI használata mellett elterjed az IRI (Internationalized Resource Identifier). Az URI protokoll az ASCII karakterkódoláson alapul, míg az IRI ezt kiterjeszti az Universal Character Set karakterkódolással, így az URI-val szemben alkalmas például a cirill, vagy kínai karakterek használatára is.

²²⁸ Marja-Riitta KOIVUNEN és Eric MILLER, „W3C Semantic Web Activity”, in *Semantic Web Kick-Off in Finland. Vision, Technologies, Research, and Applications*, 27–43 (Helsinki: Helsinki Institute for Information Technology, 2002), 34, <https://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/proceedings.pdf>.

²²⁹ Az RDF Schema: <https://www.w3.org/TR/rdf-schema/>.

fogalmazzunk meg, míg az ontológiaréteg segítségével hozhatunk létre kifinomultabb szótárakat, ahol már a fogalmak közötti összefüggések is jelölhetők. A felső rétegek a következtetésekre vonatkozó logikai szabályok alkotását, a számítógép által generált következtetések kiértékelését és az eredmények hitelességének ellenőrzését reprezentálják. A szemantikus web elmélete kapcsán megfogalmazott elképzelésekre reflektáló modellt a World Wide Web Consortium dolgozta ki, a modell publikálása óta folyamatosan ajánlásokat fogalmaz meg. A gyakorlatban az RDF leírásokat tartalmazó hármastokat úgynevezett triplestore-okban tárolják, az állítások lekérdezőnyelve pedig a SPARQL, amelynek első verzióját a W3C 2004-ben tette közzé.²³⁰

A címkézett tulajdonsággráf-modellt csomópontok, kapcsolatok, tulajdonságok és címkék halmaza reprezentálja. Mind az adatok csomópontjai, mind a kapcsolataik tárolhatnak kulcs-érték párok által reprezentált tulajdonságokat. Az RDF modellel szemben, ahol például egy állítás forrásának megadásához új hármast kell létrehozni, az LPG modellben ezek tulajdonságok, így a gráf sokkal kompaktabbá válik. A csomópontokat elláthatjuk egy vagy több címkével, amelyek révén csoportosíthatóvá válnak. A kapcsolatokat reprezentáló éleknek két tulajdonságuk van: mindig van egy kezdő- és egy végcsomópontjuk, és irányítottak, tehát a gráf irányított gráf.²³¹ Az LPG gráfadatbázisokban általában JSON formátumban tárolják az adatokat, minden csomópont és él egy-egy JSON dokumentum. A legfontosabb különbség, hogy míg az RDF URI-kat használ az azonosításhoz, az LPG adatmodell csak karakterláncokat használ, amelyek így csupán a gráfon belüli lokális azonosításra alkalmasak, nincsenek ajánlások, illetve szabványok. Ennek következtében az LPG alapú gráfadatbázis, bár teljes mértékben testre szabható, zárt rendszert alkot, a modell sajátosságai nem teszik lehetővé, hogy kapcsolódjon a szemantikus webhez. Az LPG modell implementálása sem triviális, nincs hivatalos lekérdezőnyelve sem, a legelterjedtebb a Neo4j-ben alkalmazott Cipher/Cypher.²³² Egyelőre úgy tűnik, hogy az LPG adatmodellt a bölcsészettudományok területén csak egy szűk réteg használja,²³³ ők is a Neo4j rendszerét használják. A Neo4j legnagyobb hátránya, az LPG modell sajátosságaiból fakadó zártságon túl, hogy míg az RDF technológia alkalmazásához

²³⁰ Eric PRUD'HOMMEAUX és Andy SEABORNE, „SPARQL Query Language for RDF. W3C Working Draft 12 October 2004.”, 2004, w3.org/TR/2004/WD-rdf-sparql-query-20041012/.

²³¹ Ian ROBINSON, Jim WEBBER és Emil EIFREM, „Data Modeling with Graphs”, in *Graph Databases. New Opportunities for Connected Data*, 2. (O'Reilly Media, 2015).

²³² Kevin FEENEY, „Graph Fundamentals — Part 2: Labelled Property Graphs”, *TerminusDB Community* (blog), 2019. október 7., <https://medium.com/terminusdb/graph-fundamentals-part-2-labelled-property-graphs-ba9a8edb5dfe>.

²³³ Vö. Elena SPADINI, Francesca TOMASI és Georg VOGELER, szerk., *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing* (Norderstedt: Books on Demand, 2021).

számos nyílt forráskódú, ingyenes eszköz érhető el, a Neo4j rendszerét egy cég fejleszti, a felhőszolgáltatásban az ingyenes hozzáférés korlátozott, legfeljebb ötvenezer csomópont tölthető be a rendszerbe, míg a saját fenttartású rendszer ugyanolyan informatikai tudást és eszköztárat igényel, mint egy RDF tripletstore, vagy egy SQL adatbázis kiépítése és fenntartása.²³⁴ Bár az LPG modellnek számos hátránya van, az RDF technológia nem képes kiszolgálni azt az igényt, hogy a csomópontokat összekötő éleket (állításokat) további tulajdonságokkal lássuk el. A probléma megoldásán már dolgozik az Ontotext és a W3C, bár az RDF* (és a továbbfejlesztett lekérdezőnyelv, a SPARQL*) még nem hivatalos szabvány, de 2021 elején már közzétették az első ajánlást.²³⁵

Látható, hogy bár az RDF technológia, mint minden más szabványos formátum, kompromisszumokat követel, mégis, a kidolgozott szabványok révén ez a formátum biztosítja a különböző rendszerek és más szabványos adatsereformátumok közötti átjárhatóságot. A technológiai hátránya, miszerint minden állításhoz külön hármast kell létrehozni, így „kezelhetetlen” méretű adathalmazok jönnek létre, a jelenkor számítógépes kapacitásával könnyedén áthidalható. A Hildeglyph projekt, bár innovatív vállalkozás, jelen formájában nem képes a digitális kulturális örökség elvárásainak megfelelni, ahhoz pedig, hogy a projekt során kidolgozott módszertan elterjedjen, az általuk létrehozott adatmodell széles körben történő megvitatására lenne szükség.

A projekt másik döntése, az XML formátum elvetése, szintén vitatható. Ahogyan a későbbiekben látni fogjuk, nem minden digitális szövegkiadás esetén van feltétlenül szükség TEI XML átírat használatára, azonban kijelenthetjük, hogy a filológiai igényű szövegkiadások esetén elengedhetetlen. Míg Kuczera arra a következtetésre jut, hogy az XML struktúra nem alkalmas a szofisztikált annotálásra és a „TEI közösséget az a veszély fenyegeti, hogy zsákutcába jut, hacsak nem talál hamarosan életképes alternatívát az XML helyére”.²³⁶

Georg Vogeler rámutat, hogy az XML és a gráf adatmodell szembeállítása valójában téves, hiszen tekinthetünk az XML-re úgy is, mint gráfok szerializációjára, vagy pont fordítva, az XML-re, mint gráfra, ugyanis az XML struktúra maga is természetéből fakadóan egy gyökérellemmel ellátott, rendezett gráfot alkot.²³⁷ Vogeler az asszertív kiadások kapcsán elemzi

²³⁴ Forrás: <https://neo4j.com/pricing/>.

²³⁵ Olaf HARTIG és Pierre-Antoin CHAMPIN, „RDF-star and SPARQL-star Draft Community Group Report 18 February 2021”, 2021. február 18., <https://w3c.github.io/rdf-star/cg-spec/2021-02-18.html>.

²³⁶ KUCZERA, „TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs”, 106.

²³⁷ Georg VOGELER, „Standing-off Trees and Graphs”: On the Affordance of Technologies for the Assertive Edition”, in *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerk. Elena SPADINI, Francesca TOMASI és Georg VOGELER, 73–94 (Norderstedt: Books on Demand, 2021), 73.

az XML, az RDF és a stand-off annotációk előnyeit és hátrányait. Az assertív kiadás, illetve assertív szerkesztés fogalma is Vogeler-től származik.²³⁸ Ahogyan Vogeler fogalmaz, a történészek és az irodalmárok számára mindig is a szöveg volt a fő forrás, azonban míg az irodalmárokat a minél pontosabb átirat érdekelte, a történészek számára a szövegből kinyerhető adat volt fontos, ami a digitális előtti szövegszerkesztési gyakorlatban olyan megoldásokhoz vezetett, mint például a szövegrészletek parafrázálása, vagy a személynevek normalizálása, amelyek az irodalmárok számára elfogadhatatlanok voltak. Vogeler szerint az assertív szerkesztési mód magában hordozza a különböző megközelítési módokat, így ezek a kiadások megfelelnek az eltérő tudományterületek igényeinek. Az ilyen kiadások képesek egyszerre kiszolgálni azt az igényt, hogy az adott szöveg filológiai értelemben pontos átiratként legyen elérhető, emellett a szövegből kinyerhető információkat strukturált, lekérdezhető formában szolgáltatassák a felhasználók számára. Azt Vogeler is elismeri, hogy az assertív kiadás fogalma még nem határozható körül megfelelően, de tény, hogy ilyen kiadások már léteznek.²³⁹ Az általa bemutatott példák mindegyike TEI XML átiratot használ és azt kombinálja RDF technológiával, vagy stand-off annotációkkal. Vogeler rámutat, hogy az XML formátum eleve magában hordozza a többretegű megközelítés elvét: az attribútumok és beágyazott elemek használatával újabb és újabb rétegekkel gazdagíthatjuk a tartalmat.²⁴⁰ Azonban elemzésében Vogeler legfontosabb megállapítása mégis az, hogy a digitális bölcsészet közösségének az olyan fogalmakra, mint szekvencia, hierarchia, állítás, hálózat és réteg, valójában metaforákként kellene gondolni, amelyek „a szerkesztő és az adatok közötti kölcsönhatásokat írják le egy assertív kiadás létrehozása során.”²⁴¹ Az egyes kiadástípusok eltérő metaforákat preferálhatnak, a stemmatológia egyik fő metaforája például a fa, de összességében elmondható, hogy a jelenleg rendelkezésre álló szerializációs technológiák, mint eszközök és adatmodellek összessége, nem támogatják egyformán az összes létező metaforát. Egy assertív kiadás létrehozásakor arra kell törekednünk, hogy a szerkesztési feladathoz szükséges metaforát, illetve metaforákat legjobban támogató eszközöket alkalmazzuk. Eközben a tudományos közösségnek olyan eljárások kidolgozására kell fektetni a hangsúlyt, amelyek segítségével az egyes formátumok egymásba konvertálhatók, vagy olyan eszközt kell fejleszteni, ami lehetővé teszi az eltérő adatmodellek egy platformon való szerkesztését. Az

²³⁸ Georg VOGELER, „The ‘Assertive Edition’”, *International Journal of Digital Humanities* 1, 2. sz. (2019. július 1.): 309–322, <https://doi.org/10.1007/s42803-019-00025-5>.

²³⁹ Uo., 318.

²⁴⁰ VOGELER, „‘Standing-off Trees and Graphs’: On the Affordance of Technologies for the Assertive Edition”, 82.

²⁴¹ Uo., 88.

elmúlt évtizedek tapasztalatai azt mutatják, hogy sem egy univerzálisnak gondolt eszköz, sem egy adatmodell nem képzelhető el kompromisszumok nélkül, ahogyan az például a TEI XML esetén is látható, inkább arra kellene helyezni a hangsúlyt, hogy a digitális filológus (bölcész) az általa használt technológiákat kellően ismerje, azokhoz reflexíven és kritikusan álljon hozzá. A TEI ajánlás példájánál maradva, ahogy James Cummings is rámutat, eredetileg SGML nyelvre dolgozták ki, amennyiben olyan új, a digitális szöveg szerializációját lehetővé tevő jelölőnyelv születik, a TEI ajánlásban implementálni fogják azt.²⁴² Ehhez azonban meg kell felelni a hosszú-távú megőrzéshez szükséges alapfeltételeknek, úgy mint kifejezőképesség [expressiveness], ellenőrizhetőség [validation], integráció [integration], és széles körű felhasználás [mass adoption].

A Hildegaph projekten túl az elmúlt években számos példát láthattunk a tudományos szövegkiadások adatainak szemantikus web technológiával, illetve gráf adatmodellel történő leképezésére. Olyan példákat is láthatunk, ahol, a Hildegaph projekthez hasonlóan, a TEI XML formátumot próbálják kiváltani, azonban más projektek inkább a TEI és a gráf adatmodellek kombinációjával kísérleteztek. Az Online Cultural and Historical Research Environment (OCHRE)²⁴³ rendszerben például egy saját fejlesztésű XML formátumot használnak, amely egy kiadás legkisebb értelmezhető egységét, általában egy-egy szót, vagy grafémát tartalmaz. Az általuk fejlesztett gráfadatbázis alapú rendszert eredetileg régészeti adatok leírására készítették, ezért azon túl, hogy a rendszerben előre definiált állítás- és elemtípusok érhetők el, a kutatóknak lehetőségük nyílik saját kategóriák definiálására is. Ez a szemlélet érvényesül a később bevezetett, digitális szövegkiadások létrehozását és publikálását lehetővé tevő szolgáltatásukban is. Bár az általuk használt XML formátum nem felel meg a TEI ajánlásának, fejlesztettek egy olyan eszközt, amely XSLT stíluslapok segítségével lehetővé teszi a kutatók számára, hogy a szövegeket TEI formátumban exportálják az adatbázisból.

Szintén egy friss kutatásban arra tesznek kísérletet, hogy a kritikai apparátust gráfadatbázisban modellezzék a TEI szemantikájának megfelelően, azonban RDF nyelven.²⁴⁴ Az általuk készített ontológia, a Critical Annotation Ontology (CAO) megfelel a Web Annotation

²⁴² CUMMINGS, „A world of difference”, 59.

²⁴³ Miller C. PROSSER és Sandra R. SCHLOEN, „The Power of OCHRE’s Highly Atomic Graph Database Model for the Creation and Curation of Digital Text Editions”, in *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerk. Elena SPADINI, Francesca TOMASI és Georg VOGELER, 55–71 (Norderstedt: Books on Demand, 2021).

²⁴⁴ Francesca GIOVANNETTI, „The Critical Apparatus Ontology (CAO): Modelling the TEI Critical Apparatus as a Knowledge Graph”, in *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerk. Elena SPADINI, Francesca TOMASI és Georg VOGELER, 125–139 (Norderstedt: Books on Demand, 2021).

szabványnak,²⁴⁵ valamint implementálták az FRBRoo ontológia egy részét is.²⁴⁶ Az elképzelés szerint a stand-off annotációk egy TEI jelölőnyelvi átírat meghatározott pontjaira mutatnak, azonban nem a stand-off annotációknál megszokott módon, az annotált szegmentum kezdő és végindexére mutatva, hanem az átíratban **seg** jelölővel ellátott szegmentum `@xml:id` attribútumában megadott azonosítóra. Ennek a módszernek számos előnye lehet, például megoldhatja olyan jelenségek jelölését, mint például a transzpozíciók, amelyeket csak nehezen, vagy egyáltalán nem lehetne jelölni a TEI szintaktikája szerint.

Ezek a megoldási kísérletek, bár számos előnyük van, nem reflektálnak a szélesebb kutatóközönség és a felhasználók igényeire. Sok esetben nem derül fény a szöveg és a gráf vizualizáció kapcsolatára, a keresési lehetőségekre. Ilyen kiadások létrehozása, illetve az ilyen kiadások „olvasása” magas informatikai szaktudást igényel,²⁴⁷ ez számos kutatót elrettenthet a használatuktól. Azzal, hogy ezek a projektek egyéni adatstruktúrákat és jelölési rendszert fejlesztenek, növelik a digitális kiadásokkal kapcsolatos bizalmatlanságot.

A következőkben egy olyan komplex, felhasználóbarát infrastruktúrát mutatunk be, amelyben a kritikai kiadások alapját továbbra is a TEI XML reprezentáció adja, azonban integrálja a szemantikus web technológiáit és a stand-off annotációkat.

Forráskiadások a szemantikus weben

Tudományos szövegkiadások digitális közzététele az elmúlt évtizedekben egyre inkább elterjedt gyakorlat, általános tapasztalat, hogy a kiadások a TEI ajánlásának megfelelő jelölőnyelvi átíratban készülnek. Azonban, ahogyan arra már a korábbi fejezetekben többször rámutattam, a TEI XML alapú kiadások készítése magas informatikai szakértelmet kíván. Egy kritikai kiadás esetében ez nem kerülhető meg, de milyen lehetőségei vannak egy annotált forráskiadást készítő kutatócsoportnak, ha nem tudja finanszírozni a komplex infrastruktúra kiépítését és fenntartását? A következőkben egy olyan kísérleti kiadást mutatok be, amely szabványos formátumban, de nem a TEI ajánlása szerint készült egy olyan rendszerben, amelyet eredetileg a szemantikus web technológiáit alkalmazva, adatok tárolására fejlesztettek ki.

²⁴⁵ Uo., 129.

²⁴⁶ Uo., 132.

²⁴⁷ Egy gráfadatbázisban történő kifinomult kereséshez például el kell sajátítani az adatbázis lekérdezőnyelvét és ismerni kell az adatbázis struktúráját. Ha ez egyes projektek esetében eltérő, akkor minden esetben egy tanulási folyamat előzi meg a kutatást.

2016-ban indult A tudományos tudás áramlásának mintázatai a *Magyar Királyságban, 1770-1830 (Tudásáramlás)* című NKFIH finanszírozású projekt,²⁴⁸ amely arra vállalkozott, hogy a vizsgált időszakban, amikor a tudományok fokozatosan új rendszerbe szerveződtek, megvizsgálja ezeket a folyamatokat, leírja a tudásáramlás magyarországi mintázatait és rekonstruálja a tudományfejlődés történetét.²⁴⁹ A projektben részt vevő kutatók (két történész, egy klasszika-filológus-irodalomtörténész, egy filozófiatörténész, egy történész-antropológus és egy politológus) hét olyan tudományterületet vizsgál, amelyek instrumentalizálódása nyomon követhető Magyarországon: filozófia, történettudományok, esztétika/klasszikus filológia, politikatudomány/statisztika, orvostudomány, etnológia/antropológia és agrártudomány.

A kutatók különböző forrásszövegeket (levelezés, útleírások, gyűjteményjegyzékek, egyetemi tankönyvek, képzési tervek, egyetemi matrikulák) tártak fel az egyes ismeretkörökhez kapcsolódóan. Ezeken keresztül kerülnek bemutatásra a „hat tudásterület magyarországi szervezetirendszerének kialakulása tekintetében meghatározó kapcsolati hálózatok, episztémikus szinterek, mediális és materiális eszközök komplex használatai.”²⁵⁰

A forrásszövegeken kívül a kutatócsoport számos értékes történeti adatot, valamint személy- és földrajzi nevet azonosított. A projekt célkitűzései között szerepelt a szövegek digitális közzététele, a források közötti kapcsolatok feltérképezése, valamint a kutatás során gyűjtött adatokból egy szemantikus hálózat felépítése, amely a forrásszövegekhez kapcsolódik. A kutatócsoporttal az ELTE BTK akkori Digitális Bölcsészeti Központjával együtt,²⁵¹ 2019-ben kezdtem meg az együttműködést.

Az első lépés a publikációs modell kidolgozása volt. Hamar világossá vált, hogy a szabványos megoldások a projekt jellege miatt nem alkalmazhatók.²⁵² Egy XML alapú kiadás készítése nagy gyakorlatot és digitális filológusi szakértelmet kíván, a jelölőnyelvi átiratok biztonságos archiválása, szolgáltatása és vizualizációja meglehetősen költséges. Bár léteznek kész vizualizációs eszközök, azok testre szabása további informatikai fejlesztést igényel. A TEI

²⁴⁸ A kiadást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal K_16 119577. számú projektje támogatta.

²⁴⁹ A kiadást lásd: BALOGH Piroska és mtsai., *A tudományos tudásáramlás mintázatai a Magyar Királyságban, 1770-1830: DigiTudásáramlás (DITU) /DigiCirculation of Knowledge (DICIKO)*. (ELTE, 2022), <https://eltdata.elte-dh.hu/wiki/Tud%C3%A1s%C3%A1raml%C3%A1s>. A projektről bővebben lásd: KRÁSZ Lilla, „Tudásképzés és tudásközvetítés Magyarországon 1770–1830”, *Századok* 2017, 5. sz. (2017): 949–952, 950.

²⁵⁰ KRÁSZ, „Tudásképzés és tudásközvetítés Magyarországon 1770–1830”, 950.

²⁵¹ Az ELTE BTK-n 2020-ban alakult meg a Történettudományi Intézet alá tartozó Digitális Bölcsészeti Tanszék, amely a Központ utódjaként működik.

²⁵² A kutatási projekt során készült forráskiadás kiadója az ELTE volt, emellett nem kritikai kiadás, így nem jelenhetett meg a DigiPhil felületén.

ajánlás egyik legnagyobb hátránya, hogy nem alkalmas kifinomult adatleírásra, az adatokat mindenképpen egy külső rendszerben kell tárolni, tehát a kutatás szempontjából releváns adatok tárolására további háttéradatbázis kiépítésére lett volna szükség. Emellett az XML struktúra nem támogatja az egymást átfedő jelöléseket, így például a forrásszövegeket összekapcsoló fogalmak megjelölése a jelölőnyelvi átiratban nem lett volna lehetséges. A háttéradatbázis szerkezetének kidolgozása, az adatok felvitele és tárolása, valamint az adatok összekapcsolása a szövegkiadással, ezek egy platformon történő megjelenítése szintén komoly fejlesztést igényelt volna. Sem egy ilyen kifinomult infrastruktúra kifejlesztése, sem annak hosszú távú megőrzése és karbantartása nem volt lehetséges. Ezen szempontok figyelembevételével döntöttünk úgy, hogy olyan eszközt keresünk, amely egyszerre támogatja a szövegkiadás publikálását és biztosítja a megfelelő háttéradatbázist.

A digitális szövegkiadás a nyílt forráskódú Wikibase szoftver keretein belül készült, a szolgáltatáshoz a szükséges tárhelyet a Digitális Bölcsészet Tanszék biztosítja. A szoftver kiválasztásánál lényeges szempont volt a fenntarthatóság, a Wikimedia csoport üzemelteti és fejleszti, nemzetközi kutatócsoportok is előszeretettel használják prozopográfiai kutatások eredményeinek online publikálásához.²⁵³ Az akkori Digitális Bölcsészet Központ az MTA-ELTE Humanizmus Kelet-Közép-Európában Lendület Kutatócsoporttal (MTA-ELTE MTA-ELTE, Humanism in East Central Europe, HECE) közösen tesztelte a rendszert és fejlesztett adatbázist az ELTEdata-ban.²⁵⁴

A Wikimedia²⁵⁵ csoport fejleszti nem csak a világszerte ismert és használt, szabadon szerkeszthető enciklopédiát, a Wikipédiát, de ők hozták létre a mára már több mint 100 millió entitás tartalmazó Wikidatát, valamint az azt meghajtó szoftvert, a Wikibase-t. A Wikidata egy olyan, dokumentumorientált adatbázis, amely a szemantikus web technológiáját valósítja meg.

Az adatbázisban minden entitás egyedi azonosítóval rendelkezik ('QID'), az entitásokat úgynevezett tulajdonságok ('property') kapcsolják össze, minden így létrejött állítás ('statement') tehát három elemű. A Wikidata fejlesztésekor megvalósították azt a technológiát, amelynek szabványosításával a W3C is foglalkozik (Isd. RDF*, SPARQL*), ami lehetővé teszi, hogy az állításokhoz és a kapcsolatokhoz további, úgynevezett minősítőket rendeljünk.

²⁵³ Az egyik legnagyobb ilyen jellegű projekt a FactGrid (https://database.factgrid.de/wiki/Main_Page), amelynek célja kutatási platformot és adatbázist biztosítani történészek számára.

²⁵⁴ SEBESTYÉN Ádám, „Az ELTEdata szemantikus adatbázis legújabb fejlesztései”, in *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TICK József, KOKAS Károly és HOLL András, 179–184 (Budapest: Hungarnet Egyesület, 2022). Az ELTEdata elérhetősége: https://eltedata.elte-dh.hu/wiki/Main_Page.

²⁵⁵ A Wikimedia további fejlesztési az alábbi linken érhetők el: <https://www.mediawiki.org/wiki/MediaWiki>.

A Wikibase szoftver egyszerre volt képes támogatni a szövegkiadást és az adatgazdagításhoz szükséges háttéradatbázist, vizuális felülete a Wikipediával és a Wikidata-val megegyező, felhasználóbarát felület. A Wikibase-ben az adattípusokat a felhasználó határozza meg, de a rendszer összekapcsolható és együttesen lekérdezhető más szemantikus web-alapú rendszerekkel. A szövegkiadás a Wikimedia saját formátumában, az úgynevezett 'wikitext' formátumban készült. Ez a formátum könnyedén archiválható, később, ha szükséges, XML formátumra konvertálható. Az adatbázis tartalmazza a kutatás során gyűjtött adatokat, méghozzá a legkorszerűbb technológiával, a szemantikus web gyakorlatát használva. A rendszer lehetőséget biztosít adatvizualizációk készítésére, amelyek feltárhatnak eddig ismeretlen mintázatokat, alátámaszthatják a kutatási eredményeket, új szempontokat vetíthetnek előre.

Az ELTEdata-ban korábban nem készült szövegkiadás, így első lépésként ki kellett dolgozni a jelölőnyelvi átírat specifikációját. A wikitext formátum főként formázási jelöléseket tartalmaz, azonban, akárcsak a Wikipedia esetében, kifinomultan helyezhetők el a szöveghelyekhez tartozó linkek, amelyek egyaránt mutathatnak entitásra, másik 'oldal'-ra, vagy külső forrásra.

```

Április negyedikén költ levelét július másodikán vettem, - mondhatom - egy leírhatatlan örömmel. De, hogy az ahhoz mellékelt supplement, melyet legalábbis
egy fertály esztendővel előbbre várhattam, nem tovább folytatandó, hanem hátra hagyott utamat tárgyzta, csak elküldtem. [https://eltdedata.elte-
dh.hu/wiki/Item:Q2543 Hollandia] isméti beutazására rendet (sic) időm eltellett, előbb, hogysen instrukcióm megérkezett volna. Június végére halasztott
instrukcióm se érkezett még meg. Mindent szívesen, s készséggel megtennék, valamit tisztem pontosan teljesíteni parancsol. Alig határozhatom el, mit tegyek,
midőn egyszerre annyi egymással összeütköző környüállások adják magokat elő! - '''<big>[https://eltdedata.elte-dh.hu/wiki/Item:Q11283 :]</big>''' Sok
tanakodás után végre abban állapotam meg, hogy még vagy egy hétig [https://eltdedata.elte-dh.hu/wiki/Item:Q2400 Brüsszel] környékén várakozzak, ha talán ez
idő alatt Kegyelmes Uram további rendelése, melyet utolsó levelem, de még inkább Kegyelmes Uram utolsó rendjeinek következésében minden órára remélek,
megérkezne, azután pedig - ha hasztalan várakoztam volna - Hollandia felé meginduljak. Ez az út, terheinek ellenére is, kellemes leszén, a haszon miatt,
melyet általa remélek. De [https://eltdedata.elte-dh.hu/wiki/Item:Q298 Angliába] teendő utam majd olyan idő szakaszra fog esni, amelyben más idegen onnét
kikívánkozik. '''<big>[https://eltdedata.elte-dh.hu/wiki/Item:Q11283 :]</big>''' Légyen az akármint azonban, Kegyelmes Uram utolsó levelére haladék nélkül
értekezni kötelességem.

```

[25. ábra] Részlet a kiadás wikitext átíratából.

A wikitext rugalmasságának köszönhetően nem okozott nehézséget a tudományágak kulcsszavainak az annotált szövegekben való megjelenítése, valamint a szöveges annotációk beillesztése.

Az első annotált szövegek, amelyeken teszteltük a specifikációt és a munkafolyamatot, a korábban nyomtatásban megjelent, Kurucz György által szerkesztett *Technológiai utazás a*

modern kor hajnalán Válogatás Gerics Pál és Lehrmann József georgikoni professzorok nyugat-európai jelentéseiből és naplóiból (1820–1825) című kötet kéziratából származtak.

A Georgikont 1797-ben gróf Festetics György alapította a magyarországi nemzeti mezőgazdaság fejlesztése érdekében az intézmény több mint fél évszázadon át működött. Az intézmény jogutódja a Magyar Agrártudományi és Élettudományi Egyetem Georgikon Campus. Festetics halála után, fia, apja végakarátát követve, Festetics László két fiatal tanárt, Gerics Pált és Lehrmann Józsefet küldte hosszú nyugat-európai tanulmányútra. A két Georgikon-tanár nemcsak azt az utasítást kapta, hogy látogassák meg a tudásközpontokat és lépjenek kapcsolatba a kor szellemi elitjével, hanem azt is, hogy tapasztalataikat minél pontosabban rögzítsék. Mindketten pontos útvonaltervet és részletes utasításokat kaptak, többek között arra vonatkozóan, hogy hogyan és miként készítsenek jegyzeteket. A naplóbejegyzések mellett hetente rövid jelentést kellett küldeniük a Direciónak.²⁵⁶ Gerics és Lehrmann meglátogatták a legnagyobb akadémikusokat, egyetemeket és nemesi gazdaságokat Morvaországtól és Hollandiától át Angliáig, és dokumentálták az útjukat. A kéziratgyűjtemény többnyelvű, Gerics magyarul, Lehrmann József pedig németül írt.

A forrásszövegek adatgazdagításához először létre kellett hozni az ELTEdata rendszerében az entitásokat. Az adatok a kiadás névmutatójából származtak, számos olyan földrajzi- és személynevet azonosított Kurucz, amelyek a későbbi forrásszövegekben is szerepeltek. Az adatok automatikus importálásához CSV formátumot használtunk, amely exportálható bármilyen táblázatszerkesztő szoftverből (pl. Google Sheet, Microsoft Excel). A CSV, vagyis a „Comma Separated Value” (vesszővel elválasztott érték) egy széles körben használt formátum egyszerű adatszerkezetek tárolására. Először meg kellett tervezni a CSV szerkezetét. A személynevek esetében automatizáltuk névmutató egyes értékeinek szétválasztását, aminek eredményeként olyan adatmezőket kaptunk, mint a személy neve, születési dátuma, halálozási dátuma és foglalkozása táblázatos formában elrendezve. A következő lépésben az ELTE Történettudományi Intézet két hallgatója olyan további adattípusokkal egészítette ki az adatsorokat, mint például az entitás neve, WikiData azonosítója, a nevek angol nyelvű formája, és elkülönítették a foglalkozásra, tisztségre, nemesi címre, uralkodók esetében pedig az uralkodás kezdetére és végére vonatkozó információkat. Eközben összevetették az összegyűjtött adatokat a nyomtatott kiadás lábjegyzeteiben megadott életrajzi adatokkal, és a jegyzetekből átvették a szerkesztői leírásokat. Bár kezdetben úgy tűnt,

²⁵⁶ György KURUCZ, „A »technological journey« normái: georgikoni tanárok utazási instrukciói a 19. század elején”, *Művelődés-, Tudomány- és Orvostörténeti Folyóirat* 9, 17. sz. (2018): 261–274, 267, <https://doi.org/10.17107/KH.2018.17.261-274>.

hogy a CSV formátumból történő importálás teljesen automatizálható, a folyamat során olyan rendszerhibák derültek ki, amelyek megnehezítették a betöltést, emellett az adatszerkezetet is át kellett alakítanunk.

Mivel a Tanszék által telepített és karbantartott Wikibase rendszer különböző gyűjteményekből áll, a személynevek leírása esetében fontos az adott információ forrásának meghatározása. Személyi entitások esetében egyes információk csak egy adott kiadással összefüggésben relevánsak. Ezért úgy döntöttünk, hogy az ilyen jellegű konkrét információkat nem fogjuk bevonni a szemantikai állításokba, amelyek a személy általános leírásából állnak, hanem létrehoztunk egy új adatmezőt, amelyben a leíráson túl egy minősítővel jeleztük, hogy az információ mely mű kontextusában releváns, az információ a 'állítás forrása' ('stated in') minősítő alkalmazásával kapcsolódik az entitás rekordjához.

FESTETICS GYÖRGY (Q3413)

Nincs leírás

[További nyelveken](#)
ÁLLÍTÁSOK

osztály, amelynek példánya	ember	▼ 0 hivatkozás
Wikidata-azonosító	Q1000935	▼ 0 hivatkozás
nem	férfi	▼ 0 hivatkozás
gyűjtemény	Tudásáramlás	▼ 0 hivatkozás
születési idő	1755	▼ 0 hivatkozás
halálzási idő	1819	▼ 0 hivatkozás
nemesi cím	gróf	▼ 0 hivatkozás

[26. ábra] Festetics György rekordja az ELTEdata-ban.²⁵⁷

Az adatbázisban a hálózatépítés alapjául szolgálhat például a pozíciók, foglalkozások, nemesi címek és rangok azonosítása. Ilyen típusú adatokból a Gericz-kiadás jegyzeteiből 180 egyedi típust emeltünk ki. (Csak néhány példa a felsorolásból, amely mutatja e tételek sokszínűségét: üvegművész, zoológus, tájépítész.)

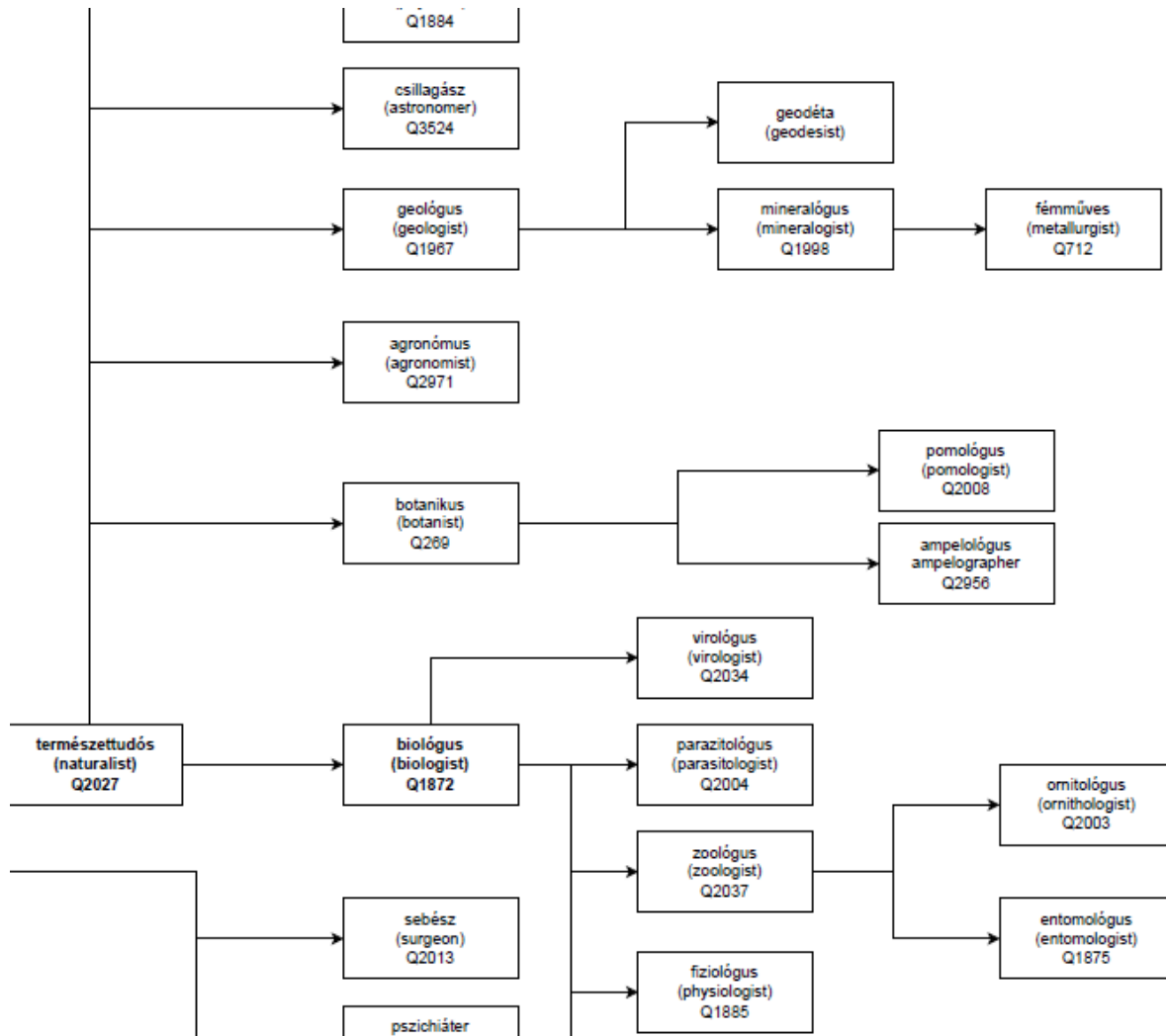
Ezen a ponton rájöttünk, hogy ezeket a tulajdonságokat rendszereznünk kell, és hierarchiába kell rendeznünk, különben nem tudnánk általános kritériumok alapján keresni, egy

²⁵⁷ A teljes rekord megtekinthető: <https://eltdata.elte-dh.hu/wiki/Item:Q3413>.

listában megjeleníteni minden katonai ranggal rendelkező személyt a korszakból. Ezek a tulajdonságok univerzálisak egy Wikibase rendszerben, így a mi gyűjteményünk és a Wikibase más gyűjteményei ugyanazokat az entitásokat használják egy személy a nemesi rangjának, beosztásának és foglalkozásának jelölésére. A hierarchia létrehozása során figyelembe kellett vennünk az adatbázisban épülő összes algyűjtemény (HECE, ELITdata,²⁵⁸ Tudásáramlás) jellegét.

Ezeknek a pozícióknak a rendszere kiterjedt, a hierarchia több rétegből áll. A foglalkozások esetében a hálózat még összetettebb; a foglalkozások hierarchiájának egy kis része, a természettudósok (lásd a 6. ábrát.). A hierarchia rendezésekor igyekeztünk a Wikidata modelljét követni, de néhány fogalmi nehézséggel is meg kellett küzdenünk. A Wikidata-ban például a „bárónő” nem létezik önálló entitásként, csupán a „báró” elem női formája, névvariánsként szerepel. Másrészt a magyarban két különböző formával jelöljük ezt a női rangot, a „bárónő” születése révén viseli a címet, „báróné” pedig házassága révén. Ez szinte az összes női nemesi címnél következetes probléma volt, ezért el kellett térnünk a Wikidata-modelltől, így egyedi tételeket hoztunk létre, és nem tudtuk a tételünket egyetlen Wikidata-elemhez kapcsolni.

²⁵⁸ Az ELITdata az ELTE TáTK Prozopográfiai és családtörténeti kutatócsoportja által épített adatbázis. A projektről bővebben: <https://tatk.elte.hu/kutatokozpontok/prozopografiai>.



[27. ábra] A természettudósok hierarchiája Tudásáramlás projektben.

A táblázat véglegesítését követően indulhatott el az automatikus betöltés. Legjobb megoldásnak az adatok importálására a Wikibase saját beépített eszközének, a QuickStatements-nek a használata tűnt. Kiderült azonban, hogy a rendszer, az akkori beállításokkal egyszerre mindössze 100 rekord betöltését tette lehetővé, és a folyamat meglehetősen lassú volt. Ebben a projektben közel 700 személynév és ugyanennyi földrajzi név állt rendelkezésünkre. Emellett a szoftververzió sajátossága volt, hogy nem tudott minden adatot kiolvasni a CSV fájlból, mert nem fogadta el a felhasználói útmutatóban megadott szintakszis szerint leírt szöveges adatokat.²⁵⁹

²⁵⁹ A Wikibase rendszert később más beállításokkal újra telepítette a Monguz Kft., a problémát ezzel kiküszöbölték.

E nehézségek fényében új módszereket kerestünk az adatok tömeges importálására, és a digitális bölcsészek által széles körben használt Python programnyelvre írt, a Wikibase szoftverrel API protokollon keresztül kommunikáló könyvtárra esett a választás.

A projekt első fázisában a Gerics-kiadás szövegeit tettük közzé és annotáltuk. A digitális kiadás szerkezete követi a nyomtatott kiadás eredeti szerkezetét, egy-egy Wikipedia-szerű oldal egy-egy fejezetet tartalmaz.

A kiadás részei a következők:

- Gerics angliai levelei Festetics Lászlónak.
- Gerics franciaországi jelentései a Directio részére
- Szemelvények Gerics naplójának ausztriai, morvaországi, sziléziai és szászországi fejezeteiből
- Szemelvények Gerics naplójának Berlinben és több német fejedelemségben szerzett tapasztalatairól szóló fejezeteiből
- Szemelvények a Gerics Pál naplójának Hollandiában szerzett tapasztalatait tárgyaló fejezetekből.
- Gerics naplójának fejezeteiből vett részletek, amelyek Anglia különböző részein szerzett tapasztalatait tárgyalják.
- Részlet Lehrmann morvaországi, poroszországi, valamint több német fejedelemségben, Hollandiában, Angliában és Franciaországban szerzett tapasztalatait összefoglaló leírásából.

A kiadás szövegeinek felvitele és szerkesztése egyszerű folyamat volt, a Wikibase beépített Visual Editor szerkesztője felismeri a Microsoft Word dokumentum szerkezetét, és automatikusan létrehoz egy speciális jelölőnyelvi átíratot, magát a wikitext-et.

Kezdetől fogva az volt a célunk, hogy a kutatás során feltárt szemantikai egységeket hozzákapsoljuk a kiadás szövegrészleteihez. Bár a Digitális Bölcsészet Tanszék fejleszt olyan eszközt, amely alkalmas lehet személy- és földrajzi entitások automatikus felismerésére és névterekkel való megfeleltetésre, a Tudásáramlás szövegeinek jellegéből adódóan ez a folyamat nem lehet teljesen automatizált. Ennek oka a szövegekben található történeti névalakok nagy száma. A jelenleg elérhető, mesterséges intelligencia alapú eszközöket modern szövegeken tanították,²⁶⁰ az egyszerűbb módszerek, mint például a keresések és cserék kombinációjával

²⁶⁰ PALKÓ Gábor és NEMESKEY Dávid Márk, „Szemantikus névelim-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása)”, in *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TICK József, KOKAS Károly és HOLL András, 359–364 (Budapest: Hungarnet Egyesület, 2022).

történő szövegmanipulációs műveletek az agglutináló nyelvek esetén jóval kevésbé hatékonyak. A szövegkiadásokban található rövidítések szintén megnehezítik az automatikus felismerést, például Gerics Pál a jelentésekben előszeretettel használta a személynevek rövidített formáját, gyakran csak a keresztnévén hivatkozott egy-egy személyre. Ez a korpusz emellett többnyelvű, a szövegek egy részét Lehrmann németül írta, de Gerics gyakran idéz különböző nyelvű szövegeket is.

A kutatócsoport 2022 folyamán a hat másik tudományterület forrásait is közzétette, ezek a szövegek nem jelennek meg nyomtatásban. A Gerics-féle kiadás a nyomtatotthoz képest javított, így ebben az esetben is a digitális kiadás az autentikus. A forrásszövegek közzétételével párhuzamosan folyt az ELTEdata-ban a szemantikus adatbázis építése, a szövegek adatgazdagítása.

A kutatócsoport a kutatás részeként olyan absztrakt, hierarchiába rendezett fogalmakat definiált, amelyek alkalmasak arra, hogy a szövegek közötti összefüggéseket megjelenítse, például

- kultúrtáj (urbanitás, lakókörnyezet, vadak és civilizáltak különbsége stb.)
- társadalom (erkölcs, oktatás, társadalmi rendek stb.)
- technológia (növénytermesztés, állattenyésztés, élelmiszer-feldolgozás stb.)
- rítusok (kannibalizmus, szertartások, szentségek stb.)

Ezzel az annotációs módszerrel létrehozták a korszakban írt szövegek fogalmi térképét, így vizualizálhatóvá és kereshetővé válik a szövegek és a különböző tudományterületek közötti kapcsolati háló. Egyelőre ez a fajta fogalmi térkép egyedülálló, különösen azért, mert az entitások hálózatát, a kísérőtanulmányok és a forrásszövegek is egy platformon érhetők el.

Digitális kritikai kiadások a szemantikus weben. A digitális filológia új infrastruktúrái

Az előzőekben bemutatott eljárás nem alkalmazható kritikai kiadások esetében, a wikitext formátum a kritikai apparátus leírásánál nem helyettesítheti a TEI ajánlást és az XML formátumot, azonban a szemantikus web szemléletmódja egyértelműen hatással van a digitális filológia gyakorlatára is. Látható, hogy egy digitális filológiai vállalkozás, legyen az egy kritikai kiadás vagy egy teljes szolgáltatás, nem nélkülözheti a megfelelő háttéradatbázis használatát, illetve létrehozását és biztosítását. Korábban úgy tűnt, hogy a könyvtári rendszerek jó megoldást nyújtanak a digitális filológusok számára, hiszen egyaránt alkalmasak személy- és földrajzi névtér építésére és bibliográfiai adatok tárolására, emellett támogatják a nemzetközi

szabványokat (pl. Marc21), azonban a könyvtári rendszereket alapvetően egy fizikai könyvtár állományának kezelésére tervezték, ezért számos olyan funkcióval rendelkeznek, amelyek a filológusok számára nem relevánsak, például a kölcsönzési funkciók, azonban a rendszer részeként megkerülhetetlenek, így a digitális filológusnak könyvtárinformatikusi kompetenciákat kell elsajátítania. Emellett az ilyen rendszerekben, az alkalmazott szabványok jellege miatt, bizonyos, a filológiai szempontból kulcsfontosságú kapcsolattípusok nem képezhetők le.

Ennek a problémának a feloldására a TEI XML fájlok úgy épülnek fel, hogy a fájlok fejléce számos olyan adatot tartalmaz, amelyeket ideális esetben egy szabványos környezetben kellene leírni, mint például a kéziratok leírására szolgáló mezők (pl. méret, hagyaték, őrző intézmény) és az adatgazdagítás csupán a személy- és helynevekre, bibliográfiai adatokra és a teauruszból azonosítható műfaji besorolásokra terjedt ki. A korábban bemutatott Móricz és Oláh levelezés-kiadások specifikációiból egyértelműen látható, hogy egy-egy XML fájl fejlécében milyen gazdag adatstruktúra található – szöveges formában. Bár ezek a mezők a TEI ajánlásának megfelelnek, sőt, a nemzetközi gyakorlatban hasonlóan alkalmazzák, valójában csak az emberi szem számára adatszerű leírások, a számítógépes algoritmusok számára csupán karakterláncok. Így hiába a rengeteg feltárt adat és kapcsolati hálózatok, ezek valójában nem válnak kereshetővé, lekérdezhetővé és megjeleníthetővé. A kompromisszum oka valójában triviális, sokáig nem állt rendelkezésre olyan nyílt forráskódú eszköz, amely alkalmas lett volna filológiai adatok leírására és szemantikai összekapcsolására, speciálisan filológiai adatok leírására fejlesztett, publikált szabvány a mai napig nem érhető el.²⁶¹

2021 májusától a DigiPhil az Irodalomtudományi Intézet szakmai felügyeletével és keretei között működik tovább, a két intézmény közötti megállapodás megszűnését követően a Petőfi Irodalmi Múzeum által gondozott kiadásokat²⁶² új szolgáltatásban, az először a PIM, majd az OSZK-hoz tartozó Digitális Bölcsészeti Platformon (dHUpla)²⁶³ tették közzé. Ezzel egy időben a Bölcsészettudományi Kutatóközpont konzorciumi partnerként csatlakozott a

²⁶¹ Egy szabványkialakítási kísérletről tudunk a Digitised Manuscript to Europeana projektről, amelynek célja az volt, hogy olyan, szemantikus web alapú ajánlást fejlesszen kéziratok leírására. A projekt végeztével megszűnt a szabvány fejlesztése.

²⁶² Anna CSÉVE és mtsai., szerk., *Móricz Zsigmond levelezésének digitális kritikai kiadása (1882–1913): 2.0* (Budapest: Petőfi Irodalmi Múzeum, 2021), <https://dhupla.hu/collection/moricz-zsigmond-levelezes>, Anna CSÉVE és mtsai., szerk., *Móricz Zsigmond: Tükör* (Budapest: Petőfi Irodalmi Múzeum, 2018), <https://dhupla.hu/collection/moricz-zsigmond-tukor>, *Kassák Lajos: A Tett (folyóirat). Digitális forráskiadás.* (Petőfi Irodalmi Múzeum, 2014), <https://dhupla.hu/collection/kassak-lajos-a-tett>.

²⁶³ <https://dhupla.hu/page/impreszum>.

Digitális Örökség Nemzeti Laboratóriumhoz, és az együttműködés keretein belül megkezdődött a projekt szakmai és technikai átalakítása.

A DigiPhil infrastruktúrájának megújítása égetővé vált, részben a DigiPhil indulása óta bekövetkezett technológiai fejlődések miatt, részben a digitális kritikai kiadások számának drasztikus és folyamatos emelkedése miatt, hiszen a filológusok közül egyre többen szeretnék digitális kiadást készíteni. Az ITI-ben készülő minden új kritikai kiadásnál elvárás, hogy eleve digitálisan, TEI XML alapon készüljön, a nagy *múltú Bibliotheca Scriptorum Medii Recentisque Aevorum* (BSMRAe) sorozat minden új kiadása szintén eleve digitálisan születő kiadás és a DigiPhil szakmai segítségével készül, a kiadásokat a DigiPhil oldalán publikálják. A digitális kiadások népszerűsége mellett továbbra is elvárás, hogy nyomtatott formában is megjelenjenek a kiadások. A DigiPhil erre az igényre válaszul olyan algoritmusok fejlesztésébe kezdett, amelyek a kritikai kiadásokat reprezentáló TEI XML átiratokból automatikusan előállítja a nyomdai előkészítésre alkalmas formátumot. Az első ilyen nyomtatásban már megjelent kiadás a korábban már más szempontból bemutatott *Nicolaus Olahus: Epistulae. Pars II*²⁶⁴ kritikai kiadás.

Bár a DigiPhil, az Europeana aggregátoraként már a kezdetektől foglalkozott szemantikus web technológiákkal,²⁶⁵ eddig a pontig nem volt lehetőség és gyakorlat a szemantikus webtechnológia szolgáltatásba történő integrálására. Az infrastruktúra átalakítása során megújult a TEI XML fájlok megjelenítésére szolgáló eszköz, a fájlok repozitálására szolgáló szoftver, az adatgazdagításhoz használt könyvtári rendszer, valamint a kereső szolgáltatás. Az infrastruktúrán túl megújult a DigiPhil munkaszervezési gyakorlata és a kiadások készítése során alkalmazott munkamódszere is. Ahhoz, hogy megértsük a technológiai fejlesztések jelentőségét és a kritikai kiadásokat készítő műhelyekre gyakorolt hatásukat, ismernünk kell a DigiPhil 1.0 felépítését. A rendszer egyes elemeit a disszertáció korábbi fejezeteiben elszórtan már említettem, a következőkben az egyes komponensek egymással való kapcsolatát mutatom be, kitérve az egyes a szoftverek kiválasztásának okaira is.

A szolgáltatás kialakításakor, 2013-ban is számos olyan kritérium fogalmazódott meg, amelyek a mai napig hatással vannak a szolgáltatás működésére és fejlesztésére. Az egyik legfontosabb kritérium, hogy a DigiPhil infrastruktúrája megfeleljen a digitális kulturális örökség elvárásainak, amelyeket olyan nemzetközi intézmények, mint például az Europeana

²⁶⁴ SZILÁGYI Emőke Rita, szerk., *Nicolaus Olahus: Epistulae. Pars II* (Budapest: r e c i t i, 2023), https://www.reciti.hu/wp-content/uploads/olahus2_vn.pdf.

²⁶⁵ Erről részletesebben lásd a következő fejezetet.

fogalmaztak meg. Ezek az elvárások a következők: platformfüggetlenség, nyílt forráskódú eszközök használata, metaadatszabványok használata és a nyílt hozzáférés biztosítása a kulturális örökséghez.

A szolgáltatás kialakítása többlépcsős folyamat volt. Első lépésként ki kellett alakítani a kiadások alapját képző XML fájlokra egy olyan specifikációt, amely a Text Encoding Initiative ajánlását követi, annál azonban szűkebb. Ennek oka, hogy a TEI ajánlása rendkívül szerteágazó, egyes filológiai jelenségekre különböző jelölési lehetőséget biztosít, egy szolgáltatás fejlesztésekor azonban a választani kell a lehetőségek közül. A DigiPhil specifikációja folyamatosan fejlődik, annak érdekében, hogy ki tudja szolgálni a szolgáltatásban elérhető kritikai kiadások eltérő igényeit.

A kezdeti kísérleteket és teszteket²⁶⁶ követően esett a választás a szolgáltatás alapját képző, a repozitálást is lehetővé tevő infrastruktúrára. A Grazi Egyetemhez tartozó Zentrum für Informationsmodellierung, az Austrian Center for Digital Humanities (ACDH) tagja 2014. augusztus 27-én, az azóta a digitális bölcsészet és filológia nélkülözhetetlen eszközévé vált GitHub kód repozitóriumban tette közzé,²⁶⁷ az akkor már évek óta saját, digitális filológiai céljaira fejlesztett eszközt, a GAMS (Geisteswissenschaftliches Asset Management System) alapját adó Cirilo-t szoftvert.

A Cirilo egyszerre alkalmas TEI XML átiratok, az arra vonatkozó és azokból kiolvasott metaadatok kezelésére, a vizualizációs stíluslapok implementálására.²⁶⁸ Az eszköz kiválasztása befolyásolta az infrastruktúra kiépítését, valamint maguknak a TEI XML fájloknak a felépítését is. A Cirilo szoftver valójában egy olyan, humán tudósok számára fejlesztett felhasználói felület, amely egy, a szolgáltatás alapjául szolgáló Fedora 3.5 repozitórium rendszerben történő műveletek elvégzését teszi lehetővé. A rendszerben a digitális objektumok úgynevezett „context-object” viszonyban állnak egymással, az egyes „context”-ek, amelyek az egyes „object”-eket összekapcsolják pedig egymásba ágyazhatók. Az egyes kiadások egészét, vagy egy-egy részét reprezentáló TEI XML fájlok egy „context” alá sorolva alkottak egységet. A DigiPhil 1.0 szolgáltatásban ezért az egyes digitális kritikai kiadások úgy épültek fel, hogy a digitális objektumok, ami lehet egy teljes regény, vagy csupán egy vers jelölőnyelvi átirata, összessége rendelődött a kiadást jelentő kontextus alá, ez meghatározta a kiadások felépítését

²⁶⁶ A DigiPhil elsőként a SADE (Scalable Architecture for Digital Editions) akkoriban fejlesztett eszközét tesztelte, azonban akkor még nem állt a készültség megfelelő fokán.

²⁶⁷ A Cirilo dokumentációja az alábbi linken érhető el: <http://gams.uni-graz.at/archive/objects/o:gams.doku/methods/sdef:TEI/get?locale=de>.

²⁶⁸ A TEI XML fájlok vizualizációjára gyakran XSLT (Extensible Stylesheet) stíluslapokat alkalmaznak, amely automatikusan generálja az XML fájlból a megjelenítésre alkalmas HTML-t.

is, tehát nem volt lehetőség különböző típusú rendezésekre a rendszeren belül. Ezen felül, minden olyan egyéb, a felhasználó informálására szolgáló weblapot, mint például a kezdőlap, vagy az impresszum is ebben a rendszerben helyeztük el, ami megnehezítette az oldalak közötti navigációt. A Cirilo szoftverben nem csak a repozitórium elérésére volt lehetőség, hanem egyéb, rendkívül hasznos funkciót fejlesztettek a rendszerbe, lehetőség volt tömeges adatszűrésre stíluslapok alkalmazásával, különböző export lehetőségek álltak rendelkezésre. A Cirilo szoftvernek az előnyeinek túl számos hátránya is volt, amik kompromisszumos megoldásokra kényszerítették a filológust. Az egyik ilyen, a jogosultságkezelés hiánya, így minden felhasználó adminisztrátori jogokkal rendelkezett, minden funkcióhoz hozzáfért (pl. törlés). A felhasználói felület kezelése bonyolult volt, nagy gyakorlatot igényelt a helyes beállítások alkalmazása, bizonyos funkciók, mint például az egy „context” alá tartozó objektumok sorrendezése néha nem megbízhatóan működött, ilyenkor manuálisan kellett javítani, ami több száz objektum esetében időigényes feladat volt.

Bár a rendszer a mai napig stabil, nem lehet elvonatkoztatni attól a tényről, hogy a magját adó repozitóriumszoftver ma már elavultnak számít, a fejlesztők nem tudták úgy átalakítani a szoftvert, hogy az újabb verziókhoz is használható legyen,²⁶⁹ mivel a következő szoftververziók teljesen más adatszerkezetet használtak, már nem „context-object” alapon épültek fel, ezzel azonban törekennyé és támadhatóvá váltak az erre a verzióra épülő szolgáltatások, kiadások.

A TEI XML fájlok vizualizációjára a, szintén digitális bölcsészek által fejlesztett, Versioning Machine, bonyolult, sok helyen hiányosan dokumentált XSLT stíluslapokból álló eszközt használta a szolgáltatás. A DigiPhil hosszú éveken át fejlesztette és szabta testre a megjelenítő eszközt, azonban a Versioning Machine egyik legnagyobb hibáját nem sikerült orvosolni. Az eszköz, bár rendkívül stabilnak bizonyult, egy nagy méretű TEI XML fájlt rendkívül lassan volt képes csak megjeleníteni, ha egyáltalán meg tudott.

A munkafájlok és a publikált kiadások TEI XML fájljainak verziókövetése és archiválása kódrepozitóriumokban történt. A kiadás készítésében részt vevő kollégák a munkafájlokat a DigiPhil saját SVN (subversion) szerverére mentették,²⁷⁰ onnan, az automatikus feldolgozást (például az azonosítók kiosztása) követően, a DigiPhil zárt Bitbucket Git technológiát²⁷¹ alkalmazó repozitóriumába töltötte át. Mindkét technológia olyan

²⁶⁹ Jelenleg a Fedora 6.x a legfrissebb verzió: <https://fedora.lyrasis.org/download/>.

²⁷⁰ Az SVN technológiáról bővebben lásd: <https://subversion.apache.org/>.

²⁷¹ A Git technológiáról bővebben lásd: Scott CHACON és Ben STRAUB, *Pro Git*, 2nd kiad. (Apress, 2014), <https://git-scm.com/book/en/v2>.

verziókezelő számítógépes protokoll, amely nyomon követi a számítógépes fájlok bármely halmazában bekövetkezett változásokat, és a szoftverfejlesztés során a forráskódot közösen fejlesztő programozók munkájának koordinálására szolgál. Az SVN technológia úgynevezett központosított verziókezelő rendszer, az összes fájlt egy központi szerveren tárolja, az adott munkaállomáson csupán a legutolsó munkapéldány érhető el, tehát a felhasználónak folyamatosan online kell lennie a változtatások elmentéséhez, illetve a fájltörténet eléréséhez. Ezzel szemben a Git egy elosztott verziókezelő, ami azt jelenti, hogy a szerveren tárolt fájlokról, azok teljes változástörténetével együtt másolat (klón) készül az adott munkaállomásra, a változtatások elmentéséhez nem szükséges a szerverkapcsolat. Mindkét rendszerhez elérhetőek szerverek és kliensek, azonban ezek jórészt korlátozott hozzáférést biztosítanak. A saját SVN szerver telepítése, kezelése és fenntartása informatikai szaktudást igényelt, így csak rövid ideig használtuk ezt a módszert. A Bitbucket rendszerében az ingyenes hozzáférés korlátozott volt, egy-egy gyűjteményt mindössze öt személlyel lehetett megosztani. A két rendszer együttes használata rendkívül nagy odafigyelést és szakértelmet igényelt, magas volt az adatvesztés lehetősége.

A TEI XML struktúrája, mivel rendkívül széles kutatóréteg igényeire alakították ki, nem alkalmas az igazán strukturált adatkezelésre, így a személy- és helynevek, valamint bibliográfiai adatok leírására. A szolgáltatást ezért egy integrált könyvtári rendszer, a Koha egészítette ki, amely tartalmazta a Petőfi Irodalmi Múzeum névterét, valamint a Geotauruszt. Emellett a DigiPhil külső adatbázisokból (pl. OSZK, Humanus) automatikusan töltött be bibliográfiai adatokat, amelyeket a munkatársak a rendszerben bővítettek és javítottak (pl. Kosztolányi-forrásjegyzék). A TEI fájlokban a felismert entitások a Koha-ban található besorolási rekordokkal lettek összekötve.

A bonyolult és eltérő felépítésű TEI XML fájlokban történő keresés nem triviális feladat. A DigiPhil szolgáltatásban eleinte egy informatikus kolléga által fejlesztett Solr alapú, facettás keresőfelület tesztelése zajlott, azonban miután az együttműködés megszűnt, lehetetlenné vált megfelelő és megfizethető, a Solr-hoz értő szakembert találni. Ezt követően egy jóval egyszerűbb, és pont ezért kevésbé kifinomult rendszert épített a DigiPhil, amelynek az alapját egy ExistDB adta. Az ExistDB azonban nem kezeli jól a nagy mennyiségű XML fájlokat, a DigiPhil fejlődésével pedig a kezelendő XML fájlok száma lassan elérte a tízezres nagyságrendet.

Az infrastruktúra egyik pillére a repozitórium, a Cirilo szolgáltatást a CERN által fejlesztett Invenio RDM repozitóriumszoftver váltotta. A DigiPhil 2.0-ban az Invenio RDM felel a TEI XML fájlok tárolásáért és vizualizációjáért. Az Invenio RDM-ben, szemben a Fedora

3.5 repozitóriummal, nem szükséges a fájlok közötti hierarchiák kialakítása, így valóban lehetőség nyílik egy-egy digitális kiadás különböző módon történő reprezentálására. A repozitórium rendelkezik vizuális szerkesztőfelülettel, amelynek magyar fordítását a Digitális Örökség Nemzeti Laboratórium munkatársai készítették, a fejlesztők pedig integrálták a rendszerbe. Az Invenio rendelkezik API végponttal, így lehetőségünk van automatikus fájlbetöltésre a protokollon keresztül. Elkülöníthetők a publikált fájlok a munkafájloktól, így különböző tesztelési lehetőség adódik a rendszeren belül, míg korábban a DigiPhil erre a célra két szolgáltatást működtetett, egy nyilvános felületet és egy tesztoldalt – ez két, egymástól független Fedora repozitórium fenttartását jelentette. Az Invenio repozitórium egy másik nagy előnye, hogy kidolgozott metaadatsémával rendelkezik, szemben a Cirilo szoftverrel, ahol csupán az alap Dublin Core adatokat fogadta.

Az új keresőszolgáltatás, amit a DigiPhil szakmai vezetésével a Digitális Örökség Nemzeti Laboratórium partnere, a Monguz Kft. fejlesztett, szintén a repozitóriumban található fájlokat indexeli, szemben a korábban használt Solr alapú keresővel és az azt követően használt ExistDB-vel, ahol a fájlokat két helyen kellett tárolni. A munkafájlok verziókövetésében is új szoftvertechnológia váltotta le a korábbi gyakorlatot, a GitHub ingyenessé tette a felhasználók számára a korlátlan privát repozitórium létrehozását, a DigiPhil minden gyűjteményét migrálta a rendszerbe, így a munkafájlok és a publikált kiadások egy helyen érhetők el. Az Invenio RDM és a GitHub összekapcsolása lehetővé teszi a kritikai kiadást készítő munkatársak számára, hogy a munkafájlok vizualizációját egy zárt rendszerben lássák, ami jelentősen megkönnyíti a lektorálási folyamatot.

Az új infrastruktúra része a DigiPhil és a Monguz által fejlesztett új vizualizációs eszköz, amely a korábban használt Versioning Machine-t váltotta fel. Az új eszközzel számos olyan problémát megoldottunk, amelyeket a korábbi eszközökkel nem lehetett megoldani. A Versioning Machine, amelyet elsősorban az egyes szövegváltozatok egymás melletti megjelenítésére fejlesztettek, a TEI kódolás természetéből fakadóan olyan szövegvizualizációkat generált, amelyek valójában abban a formájukban sosem léteztek. Az új megjelenítő eszköz több, különböző vizualizációs lehetőséget kínál a felhasználó számára. Két olyan új funkcióval rendelkezik, ami nemzetközi szinten is egyedülálló, az egyik az úgynevezett „Minden verzió nézet”, amely nem bontja szét a szövegváltozatokat, hanem azokat egymás mellé helyezve, az adott szöveghelyen mutatja. Egy másik fontos újítás a „Minden szövegváltozat (egyszerűsített nézet)” megjelenítése opció, ami lehetőséget ad arra, hogy egyes szövegváltozatokat, amelyek adott esetben zavaróak lennének az átlagos felhasználó számára, de komoly kutatási értékkel bírnak, nem releváns szövegváltozatként jelöljük meg. Ennek

segítségével olyan szövegeltérések kódolására is lehetőségünk nyílik, amelyek egy nyomtatott kiadásban nem jelölhetők, mint például egy adott nyomda betűkészlet hiányosságából fakadó szövegeltérések. Az egyes kiadásokat, a kutatócsoportokat bemutató és a DigiPhil-hez kapcsolódó egyéb információkat tartalmazó weblapokat egy, a repositóriumtól független, rugalmasan szerkeszthető WordPress alapú weboldalon helyeztük el. A DigiPhil 2.0 rendszerre történő áttérés folyamatos, amíg a kritikai kiadások jelölőnyelvi átíratainak migrációja lezárul, a még áttöltésre váró kritikai kiadások a DigiPhil 1.0 rendszerében érhetők el. Az újonnan készült informatizált és *born digital* kiadásokat már eleve a DigiPhil 2.0-án tesszük közzé. Az első két, újonnan informatizált kiadás a Régi Magyar Költők Tára sorozat két kötete 2023 végén jelent meg.²⁷² A kiadások informatizálását még Vadai Istvánnal (1960-2018) együttműködésben kezdtük meg, váratlan halála évekre felfüggesztette a kötetek informatizálását. Vadai István kiemelkedő szerepet töltött be az RMKT sorozat történetében, ahogy Hernády és Etlinger tudósításában is olvashatjuk,²⁷³ a sorozat 16. és 17. századi köteteinek digitalizált változatait évtizedes munkával készítette elő,²⁷⁴ az Irodalomtudományi Intézet Vadai hagyatékát megvásárolta, az ő munkásságára alapozva folytatja a sorozat informatizálását.

Az eddig bemutatott fejlesztések részben a munkafolyamatokat érintő változtatások, a fájlok archiválása és verziókezelése sokkal gyorsabbá és biztonságosabbá vált, míg a szolgáltatás technológiai fejlesztései a felhasználó által látható, lekérdezhető és kereshető felületeket érintették. A következőkben egy olyan vállalkozást mutatok be, amelynek segítségével sikerült a szemantikus webtechnológiát integrálni a DigiPhil rendszerébe, és ez, a nemzetközi szinten is egyedülálló fejlesztés, hatással van nem csak a digitális kritikai kiadások alapját adó TEI XML fájlok felépítésére, hanem mind az informatizálás gyakorlatára, mind az új kritikai kiadások szövegszerkesztési elveire és a kiadás készítésének folyamatára. A digitális kritikai kiadások adatgazdagításának támogatására szolgáló könyvtári rendszert egy saját, szemantikus web alapú rendszerre cseréltük. A kezdetektől az volt a célunk, hogy túllépjünk a kritikai kiadásokhoz szorosan kapcsolódó adatok kezelésén és egy olyan szemantikus adatteret

²⁷² STOLL Béla, szerk., *Régi magyar költők tára: XVII. század, 6, Szenci Molnár Albert költői művei* (Budapest: Akadémiai Kiadó, 1971). KOMLOVSZKI Tibor és S. SÁRDI Margit, szerk., *Régi magyar költők tára: XVII. század, 16, Rozsnyai Dávid, Koháry István, Petrőczy Kata Szidónia és Kőszeghy Pál versei* (Budapest: Balassi Kiadó, 2000).

²⁷³ ETLINGER Mihály és HERNÁDY Judit, „A kiadás hagyatéka / a hagyatéka kiadása: A Régi Magyar Költők Tárának hálózati kiadásáról”, in *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerk. TICK József, KOKAS Károly és HOLL András, 199–203 (Budapest: Hungarnet Egyesület, 2022), <https://doi.org/10.31915/NWS.2022.25>.

²⁷⁴ A 17. századi sorozat digitalizált változata a DigiPhil oldalán érhető el: <https://digiphil.hu/context:rmkt-17>.

alakítsunk ki, amely képes kiszolgálni az irodalomtudomány különböző részdiszciplínáinak különböző kutatási projektjei igényeit. Ehhez olyan rugalmas rendszerre volt szükségünk, amelyben sajátos adatstruktúrákat tudunk kialakítani, amelyek továbbra is könnyen összekapcsolhatók a meglévő névterekkel és ontológiákkal. Jelenlegi ismereteink szerint az egyetlen ilyen eszköz, az ELTE Digitális Bölcsészeti Tanszék által már tesztelt, korábban bemutatott Wikibase szoftver. Mivel jelenleg nem áll rendelkezésre olyan szemantikus adatstruktúra, amely minden lehetséges filológiai cselekményt vagy munkafolyamatot képes lenne reprezentálni, 2021-ben saját specifikációt kezdtünk el kidolgozni. Az adattér külső forrásokkal való összekapcsolhatóságának fenntartása érdekében a meglévő példákra támaszkodtunk, figyelembe véve a TEI ajánlását és a Wikidata különböző, gyakran egymásnak ellentmondó gyakorlatát. Figyelembe vettük továbbá a schema.org adatszerkezetet és a Marc21 könyvtári adatszerkezet-szabványt is. Világossá vált, hogy lehetőségünk van egy általános irodalomtudományi adatbázis létrehozására, amely nemcsak a kritikai kiadások metaadatait, hanem más típusú adatokat is képes leírni, például olyan archívumi anyagokat, mint kéziratosok vagy más egyedi dokumentumokat, elektronikus és analóg levelezéseket, vagy prozopográfiai adathálózatokat. Sajnos az elmúlt évtizedekben a magyar irodalomtudományban számtalan olyan egyedi építésű relációs adatbázis jött létre, amelyek nem felelnek meg a szemantikus web és a FAIR adatkezelés szabályainak, létük veszélyeztetett, mert többnyire korlátozott életciklusú kutatási projektek keretében jöttek létre.

A DigiPhil projekt az Irodalomtudományi Intézet keretein belül működik, így logikus volt, hogy az Intézet adatbázisai, a különböző témákkal és struktúrákkal, beépüljenek az új adattérbe, az ITIdata-ba. A korábban egymástól független projektek, mint a *Magyar Irodalomtörténet Bibliográfiája*, a most épülő magyar avantgárd kapcsolati hálózat és a narratológiai enciklopédia mostantól egy, egységes és fenntartható környezetben épülnek. A rendszerbe első lépésként betöltöttük a PIM névtérét, majd az ITI Bibliográfiai Osztálya által, részben retrokonverzióval,²⁷⁵ részben új gyűjtéssel létrehozott *Magyar Irodalomtörténet Bibliográfiáját*, ami mintegy 130.000 rekordból áll. A MIB névtérre fektetése, a szemantikus kapcsolatok kialakítása félautomatikusan zajlik. A DigiPhil korábbi könyvtári rendszeréből a szükséges adatkonverziókat követően áttöltjük a *Kosztolányi-forrásjegyzék*, valamint az *Arany János Bibliográfia*, valamint az Arany-kéziratkeresőben található adatokat.

²⁷⁵ A Magyar Irodalomtörténet Bibliográfiája tíz kötetes sorozatából két kötet feldolgozása történt meg eddig, a további kötetek egyelőre PDF formátumban érhetők el a honlapon: <http://www.iti.mta.hu/mib/>.

A heterogén adatokat tartalmazó szemantikus adatterek felülírhatják a kritikai kiadások készítésének folyamatát, illetve a digitális kiadás szerkesztési elveit. Egy genetikus kiadás esetén lehetőségünk nyílik minden kapcsolódó dokumentumot, vagy más anyagot, a kéziratoktól a bibliográfiákon át a szövegváltozatok különböző kombinációjáig egy rendszerben megjeleníteni, a közöttük lévő kifinomult kapcsolati rendszert láthatóvá és kereshetővé tenni. A következőkben egy olyan vállalkozást ismertetek, amely jól prezentálja, hogy a szemantikus adattér hogyan változtatta meg az informatizálás gyakorlatát.

2023-an ünnepeltük Petőfi Sándor születésének bicentenáriumát. Életművének első kritikai kiadássorozatát 1951 és 1964 között jelent meg az Akadémiai kiadó gondozásában, Varjas Béla, V. Nyilassy Vilma és Kiss József sajtó alá rendezésében. 1973-ban új filológiai elvek mentén újakezdték a sorozatot, 2008-ig öt kötet jelent meg, most készül a hatodik. A projekt során az utolsó, 2008-ban megjelent kritikai kiadást dolgoztuk fel, amely Petőfi 1847-ben írt verseit tartalmazza. A kötet különlegessége, hogy az apparátus nem csak a szövegforrások bibliográfiai adatait tartalmazza, hanem a jegyzetek támaszkodnak a Petőfi életét feldolgozó korábbi munkákra is. Az első prozoprográfiai munka az 1911-ben megjelent *Petőfi napjai a magyar irodalomban 1842-1849*,²⁷⁶ a további három 1987 és 1992 között megjelent Petőfi-adattárak rengeteg,²⁷⁷ különböző típusú adatot tartalmaznak. Az adattárak szerzője, Kiss József, összegyűjtötte a szerző életében megjelent versek jegyzékét, az erről szóló kritikákat, a kortársak naplóbejegyzéseit, a róla szóló leveleket és a hozzá kapcsolódó okmányokat (anyakönyvi kivonatok, jegyzőkönyvek, iskolai iratok, kémjelentések stb.), a dokumentumokat jegyzetekkel látta el. A három kötetes adattár bizonyos esetekben szövegrészleteket is közöl, máshol csupán tartalmi összefoglalót közöl a forrásról. Az adattárak belső utalásokat is tartalmaznak, valamint hivatkoznak az 1911-es kötetre is. Bár a két említett életrajzi gyűjtemény fontos forrása és hivatkozási pontja a Petőfi-filológiának, semmiképpen sem felel meg a mai filológiai vagy adatbázis-építési elveknek. Az eredeti források közzététele sokszor pontatlan, hol közelítő átiratok, hol szerkesztői összefoglalók formájában, gyakran korábbi újraközlésekre támaszkodva, az eredeti dokumentumot figyelmen kívül hagyva.²⁷⁸ Ezen munkák javítása, pótlása és újrakiadása elkerülhetetlen, azonban erőforrás hiányában ez

²⁷⁶ ERDŐDI Sándor, *Petőfi napjai a magyar irodalomban 1842-1849* (Budapest: Fővárosi Szabó Ervin Könyvtár, 1972).

²⁷⁷ KISS József, szerk., *Petőfi-adattár I.* (Budapest: Akadémiai Kiadó, 1987); OLVÁNYI Ambrus és KISS József, szerk., *Petőfi-adattár II.* (Budapest: Akadémiai Kiadó, 1987); KISS József, *Petőfi-adattár III.* (Budapest: Akadémiai Kiadó, 1992).

²⁷⁸ Természetesen ennek lehetett az az oka, hogy az eredeti dokumentum nem volt hozzáférhető, esetleg elveszett.

a munka még nem kezdődött el. Ettől függetlenül rengeteg értékes adatot tartalmaznak a kötetek, a bennük kirajzolódó adathálózat azonban jelen formájukban, a kritikai kiadást is ideértve, átláthatatlan. Az informatizálási feladatok megváltoztak. Míg a korábbi gyakorlat során a nyomtatott kiadásban megjelent kritikai kiadás apparátusát szemantikusan értelmeztük, azokat a lehető legpontosabban megfeleltettük a TEI ajánlásnak, majd ezeket rögzítettük a jelölőnyelvi átiratban és a rendelkezésre álló névterekből azonosítottuk az entitásokat, itt a feladat az volt, hogy a kritikai kiadásban megőrizzük az életrajzi forráskiadványokra való hivatkozásokat, és azok hivatkozott adatait feltöltsük az adattérbe, anélkül, hogy azt az illúziót keltenénk, hogy ezek ellenőrzött, filológiaiag pontos adatok lennének. A kötetben található összes bibliográfiai hivatkozás az ITIdata adattárban szerepel. Ha a kritikai kiadás hivatkozik Petőfi korábbi életrajzi forráskiadványaira, akkor az e forráskiadások által szolgáltatott adatokat is létre kell hoznunk, de egyértelműen jeleznünk kell, hogy a DigiPhil nem az eredeti forrásokra, hanem egy korábbi, vitatott hitelességű kiadásra támaszkodik. A másik nehézséget Kerényi Ferenc pontatlansága jelentette, az apparátus számos helyen téves hivatkozásokat tartalmaz. Mivel a jelölőnyelvi átirat a kritikai kiadást képviseli, annak emendálása nem lehetséges, hiszen a projekt célja nem a kritikai kiadás újrakiadása. Bár a kritikai kiadás jelölőnyelvi átiratában nem emendáljuk ezeket a tévedéseket, az adattér már a javított adatokat tartalmazza, hiszen az ITIdata adatainak hitelességeért nem a korábbi kritikai kiadások szerkesztői, hanem a DigiPhil kutatói felelnek.

A kritikai kiadás és az életrajzi forráskiadások feldolgozása során nemcsak a bibliográfiai és életrajzi adatokat, hanem a személy- és földrajzi neveket is azonosítjuk, így azok is bekerülnek az ITIdata rendszerbe. Ez a heterogén forrásokra épülő szemantikus adattér, amely a nyomtatott kritikai kiadások, a régi és új életrajzi forráskiadások és a kortárs névterek között helyezkedik el, bizonyítja, hogy a digitális médium új szintre emeli a tudományos szerkesztést. Bordalejo tételmondatára a válasz tehát az, hogy téved, létezik digitális tudományos szövegszerkesztés.

Mivel az ITIdata különböző projektekből származó adatokat fogad be, amelyek összekapcsolódnak a különböző forrásokkal és névterekkel, így az általunk feldolgozott adatok komplex szemantikai környezetbe kerülnek. Az egyes részgyűjtemények egymásra épülnek, ami nagymértékben felgyorsítja az adatbázis építésének folyamatát.

A DigiPhil által gondozott, informatizált és digitálisan születő, kritikai kiadások összes metaadata (kéziratok, bibliográfiai hivatkozások, személy- és helynevek) szintén integrálásra kerül az ITIdata rendszerbe. A jelölőnyelvi átiratok, amelyek maguk tárolták ezeket az adatokat, a szemantikus entitásokra az adattér csomópontjaiként fog hivatkozni. Ez lehetővé teszi, hogy

a TEI visszatérjen eredeti szerepéhez, a kritikai apparátus kifinomult rögzítéséhez, míg a jelölőnyelvi átírat jelentésszegmensei egy komplex adathálózatba rendeződnek és egy kereshető és rugalmasan alakítható rendszerbe kerülnek.

A digitális kiadásokon túl: Jelölőnyelvi átiratok elemzése

A disszertáció utolsó fejezete azt vizsgálja, hogy a kritikai kiadások XML reprezentációi milyen számítógépes eszközökkel elemezhetők. A digitálisan reprezentált irodalmi szövegek esetében a kutatás sikerességéhez szükség van a metaadatok és a szövegek együttes kezelésére, illetve a szöveg több szempontú gépi feldolgozására. A nemzetközi szabványok, ontológiák használata révén lehetőség nyílt az egymástól – fizikailag – távol eső adatbázisokban való értelmes és kifinomult közös keresés. A fejezet első része a metaadat-aggregáció folyamatát mutatja be, valamint olyan alternatív kísérleteket, amelyek során a digitális kiadások egészen más kontextusba kerültek. Az Arany János levelezés-kiadás kapcsán már említettem a jelölőnyelvi átiratok hálózati vizualizációjának lehetőségeit. A vizualizációk elkészítésére különböző eszközök állnak rendelkezésre, ezek közül az egyik legelterjedtebb, amely a kutatók rendelkezésére áll (és nem igényel programozói tudást), a Nodegoat.²⁷⁹ A fejezet második részében az *Arany János Levelezése 1866–1882*²⁸⁰ informatizált kiadásának, azon belül is a levelek metaadatainak feldolgozását mutatom be, majd a metaadatok különböző eszközökkel történő megjelenítési lehetőségeit hasonlítom össze.

A DigiPhil-ben készült jelölőnyelvi átiratok lehetővé teszik, hogy rendelkezésre álló kritikai kiadásokon mind a szövegek szintjén, mind a kódolt filológiai jelenségek szintjén statisztikai elemzéseket végezzünk. A szövegstatisztikai módszerekkel olyan, adott szerzőre jellemző mintázatokat deríthetünk fel, melyek korábban, erőforrás hiányában, sokszor láthatatlanok maradtak a kutatók számára. A fejezet végén Kosztolányi Dezső kritikai kiadásainak vizsgálatával mutatom be a rendelkezésre álló módszereket.

Digitális kiadások a szemantikus weben. Metaadat-aggregáció

A disszertáció ezen alfejezete szorosan kapcsolódik az előző, a szemantikus webtechnológiát a digitális filológia nézőpontjából vizsgáló alfejezethez. Domenico Fiormonte, Teresa Numerico és Francesca Tomasi 2015-ben megjelent *Digital Humanist. A Critical Inquiry* című könyvükben rávilágítottak: „[a]z interneten a szövegek csak az őket körülvevő

²⁷⁹ Pim van BREE és Geert KESSELS, „nodegoat: a web-based data management, network analysis & visualisation environment, <http://nodegoat.net>”, LAB1100, 2013, <http://lab1100.com>.

²⁸⁰ KOROMPAY H., *Arany János, Levelezése 5.: 1866–1882*.

kapcsolathálózaton keresztül élnek.”²⁸¹ Amikor digitális kritikai kiadásokat készítünk és publikálunk, nem csak egy kiadást készítünk, hanem magát a „webet” is írjuk. A web írásával kapcsolatban állapítja meg a szerzőhármast, hogy a hiperszövegekre való áttérés nem a szöveg eltűnését jelenti, hanem a szöveg, három, egymással kölcsönhatásban álló rétegből álló, paratextuális átrendeződését. Ezek a rétegek a para- mikro- és makrotextusok. A paratextus Genette-i meghatározását²⁸² alapul véve a „tranzakció zónája: a pragmatika, a stratégia és a nyilvánosságra gyakorolt befolyás privilegizált helye,”²⁸³ a nyomtatott médium klasszikus paratextusai (címek, jegyzetek stb.) az interneten új formában születnek újjá. A mikortextus, vagy mikrotartalom²⁸⁴ olyan rövid terjedelmű szövegeket (maximum 60 karakter) jelent, amelyek „valamilyen szintetizáló, definiáló, irányító, vagy illusztrációs funkcióval rendelkeznek a weboldalakon belül, de amelyeknek [a weboldaltól] függetlenül is van (valójában kell, hogy legyen) értelme.”²⁸⁵ A szerzők úgy vélik, hogy a mikrotartalom fogalma átformálhatja a paratextus Genette-i értelemben vett klasszikus funkcióit a nyomtatott médiában.²⁸⁶ A harmadik, és jelen fejezet szempontjából legfontosabb réteg a metatextus, amely tulajdonképpen a web „mély” szövege és két típusát különböztetik meg a szerzők: az a) típusba tartoznak maguk a metaadatok, amelyek sokszor, a felhasználó elől elrejtett információkat tartalmaznak a keresőmotorok számára, illetve b) fájlnevek, jelölők vagy szkriptek.²⁸⁷ A para- mikro- és metatextusok (PMM) nem csak e három kategória mentén írhatók le, hiszen az egyes textusok akár egyszerre több kategória alá is besorolhatók, hanem kognitív és szemiotikai funkcióik szerint is csoportosíthatók. A deskriptív funkció a felhasználó irányítását könnyíti meg, számára nem látható módon. A legtöbb metaadat egyébként a PMM leíró formája. A normatív funkció arra szolgál, hogy a weben elérhető tartalmak között rendet teremtsenek, ezek tulajdonképpen azok az ajánlások és szabványok, amelyek a mikrotartalomra összpontosítva a felhasználhatóságra és olvashatóságra. A harmadik funkció a dialogikus-

²⁸¹ FIORMONTE, NUMERICO és TOMASI, *The Digital Humanist*, 112.

²⁸² Gerard GENETTE, *Paratexts: Thresholds of Interpretation*, ford. Jane E. LEWIN (Cambridge: Cambridge University Press, 1997).

²⁸³ FIORMONTE, NUMERICO és TOMASI, *The Digital Humanist*, 112.

²⁸⁴ A fogalom Jakob Nielsen-től származik: Jakob NIELSEN, *Designing Web Usability: The Practice of Simplicity* (Indiana: New Riders, 1999).

²⁸⁵ FIORMONTE, NUMERICO és TOMASI, *The Digital Humanist*, 113.

²⁸⁶ A szerzők a definíció megalkotásakor nem részletezik a mikrotartalom jelentését ennél konkrétan, Nielsen 1999-ben megjelent könyvében sem találunk pontosabb magyarázatot, azonban a szerző által írt cikk ad további támpontokat a fogalom megértéséhez. A mikrotartalom olyan rövid, jól megírt szövegeket takar, amelyek a felhasználót cselekvésre ösztönzik, mint például egy cikk rövid összefoglalója. Hoa LORANGER és Jakob NIELSEN, „Microcontent: A Few Small Words Have a Mega Impact on Business”, *Nielsen Norman Group logoNielsen Norman Group* (blog), 2017. január 29., <https://www.nngroup.com/articles/microcontent-how-to-write-headlines-page-titles-and-subject-lines/>.

²⁸⁷ FIORMONTE, NUMERICO és TOMASI, *The Digital Humanist*, 113.

transzformatív funkció, amely a PMM interaktív formája, olyan beágyazott tartalmakra vonatkozik, amelyek bővíthetők, vagy módosíthatják a tartalmat (pl. JavaScript).²⁸⁸ A szemantikus web technológiái ebben az értelmezésben olyan metatextusok, amelyek deskriptív jellegűek.

A szerzőhármast, a digitális bölcsészetet, azon belül is a webet részletesen elemző könyvében a szemantikus web kérdéseire is kitér. Úgy vélik, hogy a szemantikus webnek akkor van értelme, ha „a mögöttes technológiák és elvek alkalmazhatók a meglévő adattárakra és a webes valóságra.”²⁸⁹ Ez a gyakorlatban azt jelenti, hogy ha metaadatszabványokra, illetve a szemantikus web technikai alapjait adó szabványokra (URI, RDF stb.) épül egy digitális gyűjtemény, akkor létrejön egy hatékony, erőforrásmegosztásra képes rendszer, ahhoz pedig, hogy maga a Web 3.0, a szemantikus web globálisan is létrejöjjön, a korábban bemutatott technológiai modellek minél szélesebb körben való használatára van szükség.

Barabási Albert László elhíresült, 2002-ben megjelent művében²⁹⁰ írja le azt, a Notre Dame-i Egyetemmel közös projektet, melynek során a webet mint hálózatot vizsgálták. A vizsgálat során a web egy részhálózatát elemezték, majd az eredményeiket a teljes hálózatra érvényesítették.²⁹¹ Az általuk vizsgált 325.000 weblap 82 százalékára mindössze két vagy három link mutatott, míg 42 esetben több mint tízezer mutató linket találtak, a maradék weboldalakra nagyjából 1000 link mutatott.²⁹² A kísérlet során arra a megállapításra jutottak, hogy a webről teljes mértékben hiányzik a demokrácia, méltányosság és az egyenlőség. A web topológiája gyakorlatilag megakadályozza, hogy a milliárdnyi elérhető dokumentumból egy maroknyinál többet lássunk.²⁹³ A *Digital Humanist: A Critical Inquiry* szerzői rámutatnak, hogy bár a demokrácia fogalom alkalmazása a web esetében kockázatos lehet, az egyértelműen elmondható, hogy a web struktúrája hajlamos aszimmetrikus rendszert létrehozni, „amely továbbra is jutalmazza azokat, akik már jól kapcsolódnak a piacon, és károsítja az újonnan érkezőket vagy azokat a csomópontokat, amelyek valamilyen oknál fogva kevés kapcsolatot vonzanak. Az a könnyedség, amellyel a linkekben már gazdag oldalak kattintásokat vonzanak, egy olyan rendszerre utal, amely igazságtalan, legalábbis az új linkek létrehozását illetően.”²⁹⁴

²⁸⁸ Uo., 113–114.

²⁸⁹ Uo., 159.

²⁹⁰ Albert László BARABÁSI, *Linked: The New Science of Network* (Cambridge: Perseus Publishing, 2002).

²⁹¹ Uo., 57.

²⁹² Uo., 58.

²⁹³ Uo., 56.

²⁹⁴ FIORMONTE, NUMERICO és TOMASI, *The Digital Humanist*, 171.

Így, ahelyett, hogy egy olyan rendszer alakulna ki, ami egy alkalmi konstrukció, ahol nagyjából azonos mennyiségű kapcsolattal rendelkezik minden szereplő, egy úgynevezett *Power Law Distribution* jön létre, ahol néhány csomópont sok kapcsolattal rendelkezik, míg a csomópontok nagy része csak kevéssel. Ennek következtében a hozzáférés az online információkhoz nem egyenletes. A szerzők rámutatnak, hogy a webre, mint az „online közzétett összes információhoz hozzáférést biztosító eszköz”²⁹⁵-re gondolni manapság már naivítás. A hálózat jelentős része nehezen járható csak be, tehát az átlagos felhasználó csak kis részeit éri el a webnek, a könnyedén elérhető tartalmak a hálózatot uraló alcsoportok.

A web ilyen típusú elemzése tartalomtól függetlenül vizsgálja a hálózatot, azonban a digitális kulturális örökség, és így a digitális filológia szempontjából a web dinamikájának megértése kulcsfontosságú. A kritikai kiadások online közzétételének fontossága tagadhatatlan, azonban reflektálnunk kell arra, hogy az online publikált kiadások ebbe, az igazságtalan hálózatba rendeződnek, ahol piaci szereplőkkel kell „versengeni,” amennyiben minél szélesebb körben hozzáférhetővé akarjuk tenni a tartalmakat. A következőkben olyan projekteket és kísérleteket mutatok be, amelyek erre a problémára kerestek megoldást.

Ahogy Palkó fogalmaz, az „aggregáció a digitális kulturális örökség legmeghatározóbb trendje, melynek során a kisebb helyi gyűjtemények digitális anyagait, képeket, hangokat, videókat és szövegeket, illetve az ezeket leíró adatokat nagyobb, nemzeti vagy nemzetközi szolgáltatásokban összegyűjtik, feldolgozzák, gazdagítják, kereshetővé teszik.”²⁹⁶ Az aggregáció szerepe többes. Egyrészt a bevitt metaadatok szintaktikai ellenőrzésére szolgál, másrészt a metaadatokban található leírások alapján kapcsolati hálót hoz létre a gyűjteményekben szereplő objektumok, szövegek között. Noha az aggregáció célja mindenhol ugyanaz, a konkrét munkafolyamat környezetről, környezetre, gyűjteményről-gyűjteményre változik. Az ezredfordulót követően az Európai Unió már szorgalmazta a kulturális örökség online publikálását és elérhetővé tételét. 2001 és 2004 között zajlott a TEL projekt (The European Library), majd 2005-ben indult online portáljuk, ahol mintegy két tucat nemzeti könyvtár tette elérhetővé gyűjteményi katalógusát.²⁹⁷ Ehhez a programhoz csatlakozott az Országos Széchényi Könyvtár 2006-ban. A TEL portálja, bár 2016 óta nem frissült a tartalom, jelenleg is üzemel,²⁹⁸ és nemcsak a katalógusokban való keresést teszi lehetővé, hanem bizonyos digitalizált tartalmak is elérhetők. Az Európai Bizottság 2006-ban adta ki az Európai

²⁹⁵ Uo., 172.

²⁹⁶ PALKÓ, „Mit jelent a digitális filológia a szemantikus web korában? A DigiPhil projektről.”, 1320.

²⁹⁷ NÉMETH Márton, „Az Europeana és a TELPlus projekt”, *Könyv, Könyvtár, Könyvtáros* 2009, 3. sz. (2009): 3–7.

²⁹⁸ <https://www.theeuropeanlibrary.org/>.

Digitális Könyvtár szükségességéről szóló közleményét, célként az európai kulturális örökség reprezentálhatóságát tűzték ki.²⁹⁹ A tervezett aggregációs felület, az Europeana így már nem csak könyvtári adatok, hanem általában az európai közgyűjtemények közös platformja kívánt lenni. Fontos kiemelni, hogy az Europeana infrastruktúráját a kezdetektől szemantikus web alapon fejlesztették. Ezt a TELPlus projekt keretei között készítették elő, az Országos Széchényi Könyvtár három munkafolyamatban is aktív résztvevőként szerepelt.³⁰⁰ A projekt célja egy olyan adatbázis kifejlesztése volt, melyben az Európai Unió tagállamaiban digitalizált és digitálisan született kulturális örökség metaadatai egy rendszerben érhetők el. Az Europeana a könyvtári forrásokon túl múzeumi és levéltári forrásokra is támaszkodik és tartalomszolgáltató oldalán dokumentumtípustól független, szabad kereséseket tesz lehetővé. Annak érdekében, hogy a különböző adatszolgáltatók metaadatai egy platformon váljanak kereshetővé, a különböző adatleírási gyakorlatokat össze kellett hangolni, új metaadatstruktúrát kellett kidolgozni. Az Europeana az európai digitalizált és digitálisan született kulturális örökség közzétételén túl az európai identitás erősítésére és bemutatására tesz kísérletet. Ezt bizonyítja, hogy az Europeana aktívan szerepel a közösségi oldalakon (Facebook, Tumblr, Twitter) és számos olyan alprojektet indított (pl. Europeana 1914–1918, Europeana Fashion), amely ezt hivatott reprezentálni.³⁰¹

2008 végén indult el az Athena projekt, szorosán együttműködésben az Europeana-val, céljai között szerepelt egy olyan infrastruktúra fejlesztése és kiépítése, amely segíti a múzeumok digitalizálási folyamatait és az adatok eljuttatását az Europeana-ba. Az Athena projekt, felmérve, hogy az európai múzeumok jelentős része nem áll készen az adatok szolgáltatására, a metaadat-szabványosításon túl, aggregátorként jelölte meg magát, majd kialakították az ehhez szükséges infrastruktúrát.³⁰² Az Athena folytatásaként induló AthenaPlus, melyben a Petőfi Irodalmi Múzeum is részt vett, pedig számos szoftvert fejlesztett a múzeumok számára, melyek megkönnyítették a tartalomszolgáltatást, az adatgazdagítást és az aggregációt. Ilyen szoftverek voltak például a MINT, mely a múzeumi gyűjtemények metaadatainak szabványosítását és adatgazdagítását segítette elő,³⁰³ valamint a MOVIO, melynek segítségével virtuális kiállításokat hozhattak létre a gyűjteménykezelő intézmények.

²⁹⁹ NÉMETH, „Az Europeana és a TELPlus projekt”, 3.

³⁰⁰ Németh Márton (2009): *Az Europeana és a TELPlus projekt*. Könyv, könyvtár, könyvtáros. 3, 3-7.

³⁰¹ Zsolt ALMÁSI, „Europeana: The European Identity Transfigured for and through the Digital”, in *Transfigurations of the European Identity.*, szerk. Zsolt ALMÁSI és Bulcsu BOGNÁR, 61–84 (Newcastle upon Tyne: Cambridge Scholar Publishing, 2014).

³⁰² BÁNKI Zsolt István, „Múzeumok az Europeanában - az ATHENA”, *Tudományos és Műszaki Tájékoztatás* 2010, 9. sz. (2010), <https://tmt.omikk.bme.hu/tmt/article/view/7759/10704>.

³⁰³ A MINT szoftver fejlesztését már az Athena elkezdte, az AthenaPlus pedig ezt folytatta.

Az Athena használta először a LIDO (Lightweight Information Describing Model)³⁰⁴ múzeumi szabványt, mely mára olyannyira elterjedt, hogy a Google Art projekt is ezt az XML alapú metaadatsémát használja. A 2012-ben indult Europeana Inside projekthez, melynek célja még több ország bevonása az Europeana-ba, pedig három magyar intézmény is csatlakozott aggregátorként, a Petőfi Irodalmi Múzeum, a Magyar Nemzeti Múzeum és a Szépművészeti Múzeum.³⁰⁵ A projekt célja – az adataggregáción túl – a gyűjteménykezelő rendszerek (CMS) (ezek fejlesztőinek bevonásával) közvetlen összekapcsolása az Europeana infrastruktúrájával.

Mivel az Europeana projekt keretében gyűjteményi ágazattól függetlenül gyűjtik a metaadatokat, ezért egy egyszerű metaadat-struktúrát alakítottak ki, így a különböző gyűjtemények adatainak leírása csak absztrakt keretben lehetséges; ám így a keresések, illetve a reprezentált adatok sokszor nagyon távol kerülnek az eredeti gyűjtemény nyújtotta gazdag magyarázó kontextustól. Ezen probléma megoldása lehet, hogy ágazatonként, illetve tématerületenként specializált metaadatolási struktúrát dolgoznak ki a gyűjtemények számára. Több projekt indult ilyen céllal, mint például a már korábban említett Athena, mely a múzeumok számára fogalmazott meg javaslatokat, vagy az ECLAP, amely az előadóművészeti ágak kulturális objektumaira vonatkozó metaadatok közzétételét segítette elő. A digitális filológia területén kiemelt fontosságú volt a 2012 és 2015 között futó Digitised Manuscripts to Europeana (DM2E) projekt,³⁰⁶ amelynek eredeti célja az volt, hogy a kéziratok digitalizálására, metaadatolására és a metaadatok szemantikus webben való közzétételére tegyen ajánlásokat. Az általuk kidolgozott metaadat-séma³⁰⁷ részletgazdag és kifinomult, szemantikus web alapú metaadat leírást tett lehetővé, emellett a projekt aggregátori szerepet is betöltött, az általuk nyújtott szolgáltatás része volt a metaadatok összegyűjtése és eljuttatása az Europeana-ba.

A digitális kulturális örökség magyarországi aggregációs tevékenységének első képviselői között említhetjük az Országos Széchényi Könyvtárat (a Magyar Digitális Képkönyvtár, majd a Digitális Könyvtár), a Neumann Kht.-t (Nemzeti Digitális Archívum, Nemzeti Audiovizuális Archívum), az Országos Levéltárat és a Magyar Nemzeti Múzeumot.³⁰⁸

³⁰⁴ A LIDO-ról bővebben lásd: <https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/about-lido/what-is-lido/>.

³⁰⁵ BERGER Ágoston, NÉMETH Márton és SIMON András, „Az Europeana Inside és magyar megvalósítása”, *Tudományos és Műszaki Tájékoztatás* 2014, 9. sz. (2014): 342–349.

³⁰⁶ „Digitised Manuscripts to Europeana: Final Public Report”, 2015, https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Digitising_Manuscripts_to_Europeana/Documents/DM2E%20Public_Final_Report.pdf.

³⁰⁷ Evelyn DRÖGE, Julia IWANOWA és Steffen HENNICKE, „A Specialisation of the Europeana Data Model for the Representation of Manuscripts: The DM2E Model”, *Libraries in the Digital Age (LIDA) Proceedings* 2014, 13. sz. (2014).

³⁰⁸ BÁNKI Zsolt István, „Felelet a Mandalatra”, *Tudományos és Műszaki Tájékoztatás* 2012, 8. sz. (2012): 311–319.

Az OSZK 2010-ben, Európában az elsők között tette közzé törzsgyűjteményi olvasói katalógusát és a Digitális Könyvtár teljes tartalmát és metaadat állományát a szemantikus webben, valamint névrekordjait eljuttatta a VIAF (Virtual International Authority File) adatbázisába, belépve ezzel a szemantikus web világába.

Ahogy egyre több, heterogén kulturális örökséget őrző közgyűjtemény csatlakozott az aggregációs folyamatokhoz, egyértelművé vált, hogy egy országos aggregátor helyett közgyűjteményi áganként épüljenek ki az aggregátorok. Ezzel a folyamattal párhuzamosan, 2011-ben a magyar kormány létrehozta a MaNDA-t (Magyar Nemzeti Digitális Archívum), melynek célja, hogy a magyar kulturális objektumokat egységesen digitalizálják, metaadatulják és tegyék közzé, majd a metaadatokat eljuttassa az Europeana-ba. 2016-ban az Europeana portálján majdnem 50.000, a MaNDA által aggregált adat volt elérhető, többek között térképek, fényképek, videók és folyóiratok. A MaNDA jogutódja a Forum Hungaricum, jelenleg ők a legnagyobb magyarországi aggregátor, közel félmillió rekordot tettek közzé.³⁰⁹ A Forum Hungaricum széles szolgáltatási körrel rendelkezik, többek között online filmkölcsonzóval, közös metaadatkeresővel. A MaNDA-val szinte egy időben indult el első magyar múzeumi aggregációs szolgáltatásként a MuseuMap projekt, mely 2012-ben csatlakozott az Europeana-hoz. A MuseuMap szolgáltatása magába foglalja gyűjteménykezelő-rendszer szolgáltatását, aggregációs adatbázisként funkcionál, adatkonverziót hajt végre az egyes adatformátumok között, valamint a keresést elősegítő adatgazdagítást végez (pl. tárgyszavak angolra fordítása, többnyelvű helységnévtár építése). Ezek mellett a MuseuMap segíti a múzeumok közösségi médiában való részvételét, a portálon bemutatja az egyes intézményeket, lehetővé teszi virtuális kiállítások közzétételét, illetve az intézmények saját programjainak népszerűsítését. A különböző típusú szolgáltatások közül a partnerintézmények kiválaszthatják a számukra megfelelőket.³¹⁰ A MuseuMap pilot projektje 2014-ben indult, amelynek tapasztalatai alapján 2015-ben indult az éles szolgáltatás, ami azóta is dinamikusan fejlődik.

A magyarországi aggregációs tevékenység sokáig európai szinten jelentős volt. Bár levéltári, könyvtári és múzeumi szféra úgynevezett domain specifikus aggregátorokon keresztül rendkívül nagy adattömeget tett közzé, azonban filológiai jellegű aggregáció sokáig nem volt

³⁰⁹ A rekordok listája az Europeana-ban:

<https://www.europeana.eu/hu/search?page=1&qf=PROVIDER%3A%22Forum%20Hungaricum%20Non%5C-profit%20Ltd.%22&query=&view=grid>

³¹⁰ FEJES Ildikó, LENGYEL Mónika és SIMON András, „MuseuMap, az első magyar múzeumi aggregációs portál” (Networkshop 2015, Sárospatak, 2015. április 2.), <https://kifu.videotorium.hu/recordings/10582/museumap-az-első-magyar-muzeumi-aggregacios-portal>.

Magyarországon. A DigiPhil, jellegéből fakadóan nem csak digitális kritikai kiadásokat tesz közzé, hanem egyben aggregációs szolgáltatás is. Az már kiderült a szolgáltatás kialakításakor, hogy „önmagában a jelölőnyelvi szabvány használata nem elégséges, mert a szövegtest és a szövegjelenségek kódolásán a szövegekre vonatkozó adatok, a metaadatok szabványosítása is elengedhetetlen. Erre vonatkozóan a TEI ajánlása nem elég szigorú és szisztematikus.”³¹¹ Ezért alig egy évvel indulását követően, a Wittgeinstein Source példáját követve csatlakozott a Digitised Manuscript to Europeana projekthez.³¹² A metaadatok közzétételében a DigiPhil korai szakaszában a DM2E által kidolgozott metaadat-sémát követte. Elsőként *A Tett* folyóirat XML átírataiból kinyert metaadatokat alakította át a DM2E által fejlesztett XSLT stíluslapok segítségével az általuk meghatározott sémát követő, a szemantikus webtechnológia alapjául szolgáló RDF-ekké.

Az XSLT (eXtensible Stylesheet Language Transformations)³¹³ egy úgynevezett Turing-kompatibilis szkriptnyelv. Ez azt jelenti, hogy bármilyen számítógép által elvégezhető feladat kódolására alkalmas. Ezen tulajdonsága ellenére mégis a legjobb abban az alkalmazásban, amire kifejlesztették, ez pedig az XML típusú fájlok transzfomálása. Az XSLT sokat merített a korábban kifejlesztett szövegmintázat-illesztés alapú szövegfeldolgozó nyelvekből, mint például az AWK (Unix shell alapú szövegfeldolgozó programnyelv, illetve program). Az XSLT, ellentétben az elterjedt programnyelvekkel, amelyek imperatívák, deklaratív, azaz a programozónak csak azt kell megadni, hogy mit szeretne végrehajtani, magát az algoritmust az értelmező-, illetve fordítóprogram állítja elő. Az XSLT egy XML fájl feldolgozása közben az XSLT stíluslapokban általunk meghatározott sablonok illesztésével keresi a feldolgozandó XML csomópontot, XPath és XQuery kifejezéseket használva. Amennyiben találkozik azokkal, úgy a sablonban leírt műveleteket a csomóponton végrehajtva kiírja azt a kimenő fájlban. Egy stílusban több sablon meghatározása is lehetséges. A sablonok mindig a legtágabbtól a legszűkebb illesztési kritériummal rendelkező sorrendben futnak le. Tehát a teljes dokumentumra vonatkozó sablon mindig hamarabb fut le, mint egyetlen elemre vonatkozó sablon. Ezzel a hierarchizált viselkedéssel ugyan sokszor redundáns transzformációk készülnek, de elkerülhető, hogy egy hibásan megírt XSLT szkript véletlenül már egy kész transzformáció eredményét írja felül. A sablonokon belül lehet változókat, ciklusokat és feltételes elágazásokat használni. A fent említett két tulajdonság alkalmassá teszi gyakorlatilag két bármilyen XML alapú formátum egymásba való átalakítására.

³¹¹ PALKÓ, „Mit jelent a digitális filológia a szemantikus web korában? A DigiPhil projektről.”, 1320.

³¹² Uo.

³¹³ Michael KAY, *XSLT 2.0 and XPath 2.0 Programmer's Reference* (Indianapolis: Wiley Publishing Inc., 2008).

Elsőként a DigiPhil A *Tett* folyóirat, mint a digitális kiadás „kézirata”, metaadatait aggregálta a DM2E és az Europeana szolgáltatásokba. A DM2E sémájának kidolgozásakor az Europeana által alkalmazott szabványt vették alapul. Az Europeana Data Model³¹⁴ (EDM) egy RDF nyelven alapuló metaadat-modell, amely szigorúan meghatározza mind az RDF-ek, mind a metaadatok szintaxisát. A modell nem kizárólag saját elemekből építkezik, számos már létező ontológiát használtak fel a specifikáció során (pl. ORE, DCTERMS, RDAGR2). Az Europeana a modell megalkotása során osztályokat definiált és azokhoz tulajdonságokat rendelt. Ezek a tulajdonságok további hierarchiákat alkotnak. Például az „edm:aggregatedCHO” elem, amely az aggregált objektum URI-ját tartalmazza és az „edm:hasView” elem, amely az aggregált kulturális örökség digitális reprezentációjának URI-ját tartalmazza, alárendeltjei az „ore:aggregates”³¹⁵ elemnek, ezzel mutatva, hogy az aggregátor az eredeti kulturális örökség metaadatait és annak digitális reprezentációját szolgáltatja. Az EDM teljes specifikációja számos osztályt tartalmaz,³¹⁶ de ezek közül a jelenlegi aggregációs gyakorlatban csak hét osztályt fogadnak be. A törzsosztályok közül az „edm:ProvidedCHO” tartalmazza a szolgáltatott kulturális örökségre jellemző kötelező elemeket, mint az objektum típusa, címe és leírása, valamint olyan választható tulajdonságokat, mint a variánsainak megjelölése, vagy az egyéb kapcsolódó kulturális örökségek URI-ja. Az „ore:Aggregation” törzsosztályba az aggregációs tevékenységre vonatkozó metaadatok kerülnek, az aggregátor neve, az objektum közzétevője, a felhasználási jogokra vonatkozó információk, valamint az adatszolgáltató neve. Az utolsó osztály, melyet az Europeana implementált, az „edm:WebResource”, amely az aggregált objektum webes elérhetőségét és a megtekinthető digitális objektum típusát (pl. szöveg, vagy kép) reprezentálja.

A törzsosztályok mellett az Europeana öt kontextuális osztályt implementált, amelyek a törzsosztályokban szereplő adatok további gazdagítására szolgálnak: „edm:Agent”, „edm:Place”, „edm:TimeSpan”, „skos:Concept”, „cc:License.”

Mára már az adatok leírására a kontextuális osztályok alárendelt állításaiban, amellet, hogy szövegesen rögzíthetjük az adatot, lehetőségünk van az Europeana által támogatott névterek és teauruszokkal való megfeleltetésre is. Az Europeana maga is végez szemantikus adatgazdagítást, különböző algoritmusok segítségével próbálják a törzsosztályok szöveges adatait megfeleltetni külső forrásokkal, valamint a rendszer által nem támogatott névterekre

³¹⁴ Az EDM előzménye az Europeana Semantic Elements, egy XML alapú séma, amely még nem felelt meg a szemantikus web kívánalmainak: <https://pro.europeana.eu/page/ese-documentation>.

³¹⁵ Ennek az elemnek a domain-je az „ore:Aggregation”.

³¹⁶ „EDM - Mapping guidelines”, 2023. március 21., <https://europeana.atlassian.net/wiki/spaces/EF/pages/987791389/EDM+-+Mapping+guidelines>.

mutató URI-kat próbálják megfeleltetni a támogatott névterekkel.³¹⁷ Ez azonban nem volt mindig így, sokáig nem implementálták az adatgazdagítást, így a kontextuális osztályok alárendeltjeiben csak kevés névteret fogadtak el, a legtöbb mezőt csak szöveges értékkel lehetett megadni.

Az adatmodell egyik nagy újdonsága az elődjéhez, az ESE modellhez képest, hogy támogatja a digitális objektumok hierarchiájának a megjelenítését. Mivel az RDF egyszerre csak egy digitális objektumot reprezentál, a hierarchiában feljebb vagy lejjebb található objektum adatait az RDF nem tartalmazza, csupán az URI-ját, és az Europeana kapcsolja össze a megjelenítésnél az objektumokat. Az objektumok összekapcsolására az „edm:ProvidedCHO” törzsosztály tulajdonságai között a „dcterms:isPartOf”, és a „dcterms:hasPart” jelölők használatával van lehetőség. Emellett a modell lehetőséget ad arra is, hogy egyes objektumok között sorrendet definiáljunk, erre szolgál, az „edm:NextInSequence”. Egyéb kapcsolattípusokat is létre lehet hozni az egyes objektumok között, meg lehet jelölni általános kapcsolatot, az eredeti objektum leszármazottait, illetve egyéb reprezentációit is. A filológiai objektumok esetében ez utóbbi viszonyok jelölése kulcsfontosságú, hiszen kéziratlapon egységeit, vagy folyóiratcikkek lineáris és hierarchikus viszonyait másképpen nem lehetne reprezentálni.

Bár a DM2E-t eredetileg kéziratok (és azok átiratainak) kifinomult metaadatolására tervezték, a projekt során a kézirat fogalmát bővítették, minden olyan dokumentum-átiratot le lehetett írni, amelynek van faksimiléje. Ennek megfelelően a kézirat létrehozói között nem csupán a szerző és a digitális átiratban közreműködők leírására van lehetőség, hanem számos speciális jelölőt emeltek át más névterekből, illetve hoztak létre a saját ontológiájukban. Így lehetőséget biztosítottak például a szerkesztő, vagy az illusztrátor metaadatainak megadására. A DigiPhil esetében ez tette lehetővé, hogy a *Tett* adatait a sémának megfelelően, adatvesztés nélkül írja le.

A DM2E kidolgozása során az EDM-et tekintették kiindulási alapként, a modellbe számos, már létező névteret integráltak, melyeket kiegészítve az új, saját elemekkel új ontológiát hoztak létre.³¹⁸ Az ontológia kialakítása során az új elemeket a felhasznált ontológiák

³¹⁷ A támogatott névterek listáját egy Google táblázat formájában tették közzé, az alábbi linken érhető el: https://docs.google.com/spreadsheets/d/1BoDNolkcp_qfvVShdOZYGcf61XslcwKF2MdGcJgYs20/edit#gid=0. Az automatikus szemantikus adatgazdagítás folyamatáról itt írnak bővebben: <https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit>.

³¹⁸ Evelyn DRÖGE és Julia IWANOWA, „DM2E Model V 1.2 Specification”, 2015. január 20., https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Digitising_Manuscripts_to_Europeana/Deliverables/DM2E_Model_V1.2.pdf.

egyes osztályainak tulajdonságaiként definiálták. Az egyes integrált névterekből a modell meghatározza, hogy mely elemeket lehet felhasználni az RDF elkészítése során, továbbá az egyes elemeknél, ha szükségesnek tartották, újradefiniálták az elem jelentését, illetve kiegészítették azt a speciális célnak, a kéziratleírásnak megfelelően. Az EDM-hez hasonlóan itt is három törzsosztályt lehet használni, a modell további osztályokat is definiál. Az egyik legjelentősebb különbség az EDM és a DM2E között, hogy egészen másként jelenítik meg a hierarchikus viszonyokat. Ha a leírni kívánt objektum egy hierarchia részét képezi, nem kell a hierarchiában feljebb álló alkotásra hivatkozni és létrehozni annak egy külön RDF-et. A DM2E-ben egy RDF tartalmazhatja nem csak az adott kézirat adatait, hanem a hierarchiában felette álló objektum összes metaadatát. Ennek következtében az összes törzs- és kontextuális osztály annyiszor ismétlődik meg, ahány hierarchiaszintet megkülönböztetünk. Ezek az osztályok az RDF-en belül azonos szinten állnak, a megkülönböztetésükről maguk az URI-k gondoskodnak, hiszen a megegyező URI-val rendelkező osztályok tartoznak azonos objektumhoz. A leírni kívánt egység hierarchiában elfoglalt helyét az „edm:ProvidedCHO” törzsosztályban, a „dm2e:levelOfHierarchy” elembe lehet jelölni. A dokumentáció meghatározása szerint úgy, hogy a legmagasabb és egyben legtágabb szintet a legkisebb számmal kell leírni. Ha például két szintet határozunk meg, melyben a legmagasabb szinten egy könyv áll, alatta pedig annak a fejezetei, akkor a könyvet leíró törzsosztályban az 1-es értéket kell használni, a fejezeteknél pedig a 2-est. Mindezek mellett jelölni kell, hogy az adott RDF által leírt objektum a hierarchia mely szintjén áll. Erre az „ore:Aggregation” osztályon belül a „dm2e:displayLevel” elembe van lehetőség, ahol a „true” érték jelöli az aktuálisan leírt objektum szintjét, a „false” érték pedig azt, amelyhez kapcsolódik.

A DM2E aggregátorként olyan infrastruktúrát fejlesztett, amelynek segítségével a digitális objektumok metaadatait egy felhasználói felület segítségével lehetett átalakítani az általuk támogatott sémának megfelelő RDF-ekké. Olyan adatformátumokat tudott a rendszer kezelni, mint pl. a TEI, vagy a Marc21. Emellett a projektben résztvevőknek közös repozitóriumot biztosítottak, ahol mindenki kialakíthatta a saját algyűjteményét, ennek megfelelően a modellben meghatározták az URI-k szintaxisát is, az „ore:Aggregation” törzsosztályhoz tartozó URI például így volt definiálva: `http://data.[aggregator].[domain]/data/aggregation/[provider]/[dataset]/[identifier]`

A modell minden osztályhoz kötelező elemeket rendel, melyek az integrált osztályok esetében megegyeznek az EDM-ben kötelezőként jelöltekkel, illetve kiegészülnek a hierarchia jelölésére szolgáló elemekkel. Emellett olyan elemeket határoztak meg, amelyek használatát

nem tették kötelezővé, de alkalmazásukat erősen ajánlották, mint például a kézirat digitális reprezentációjának a felhasználására vonatkozó jogi feltételek.

A modell kifinomultságát jelzi, hogy olyan adatokra dolgoztak ki jelölési rendszert, amelyeket például a Marc21 szabvány szerint, vagy a TEI fejléc szabályai mentén nem lehet adatszerűen jelölni, mint például az adott kézirat tulajdonosainak leírása, vagy egy levelezés teljes kapcsolatrendszerének reprezentálása. A modell emellett rendkívül rugalmas volt, egy RDF tartalmazhatta a teljes kéziratot, vagy annak akár csak egy kis részletét (pl. illusztráció, vízjel).

A DM2E modell legnagyobb újítása az akkor érvényben lévő EDM-hez képest, hogy támogatta, sőt helyenként megkövetelte az adatgazdagítást, beágyazott elemek formájában. Az RDF-ben például a nevet leíró elemeknél a meghatározott szintakszisú URI nem közvetlenül a jelölő attribútum értékébe kerül, hanem egy beágyazott elembe, melyben jelölhetjük, hogy személyt, intézményt vagy helynevet írunk le. Ezen belül egy újabb beágyazott elembe lehetett megadni szóvegesen magát a nevet, illetve egy másikban a névtérből azonosított leírás elérhetőségét. Általánosságban elmondható, hogy a modell sokkal részletgazdagabb metaadatolást tett lehetővé az EDM-mel, vagy a TEI-vel szemben, így a hasonló elvek mentén leírt adatok közti kifinomult keresés is lehetségessé vált.

A DigiPhil *A Tett* annotált forráskiadását 2014-ben tette közzé a szolgáltatásban. A metaadat-aggregáció során először előállította a DM2E sémájának megfelelő RDF-eket, majd az EDM-eket. Mivel a séma támogatja a hierarchiák jelölését, egy-egy RDF felépítése meglehetősen bonyolult volt. A DigiPhil *A Tett* esetében 3 hierarchiaszintet határozott meg. Az első szinten állt maga a folyóirat, a második szinten az adott folyóiratszám, a harmadik szinten pedig a konkrét cikk. Minden hierarchiai szinten mindhárom törzsosztályt ki kell tölteni, az RDF-ben a „dm2e:levelOfHierarchy” elem tartalmazta az elem hierarchiában betöltött szerepét. Mivel a folyóiratra vonatkozó adatok minden RDF-ben megegyeznek, ezeket az adatokat az XSLT állította elő, ahelyett, hogy a TEI fájllokból olvasta volna ki. Ugyanez volt az eljárás az aggregációs tevékenységet leírók adatoknál, ugyanis ezek a hierarchia szintjétől függetlenül megegyeznek.

```

<edm:ProvidedCHO rdf:about="http://data.digiphil.hu/data/item/pim/kassak/o:atett-1.tei.3">
  <edm:type xml:lang="en">TEXT</edm:type>
  <dc:type rdf:resource="http://purl.org/spar/fabio/Article"/>
  <dc:title xml:lang="hu">Novella és egyéb kísérletek</dc:title>
  <dc:language xml:lang="hu">hu</dc:language>
  <bibo:pages xml:lang="hu">4-6</bibo:pages>
  <dcterms:isPartOf rdf:resource="http://data.digiphil.hu/data/item/pim/kassak/context:atett-1"/>
  <dm2e:levelOfHierarchy rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">3</dm2e:levelOfHierarchy>
  <edm:isNextInSequence rdf:resource="http://data.digiphil.hu/data/item/pim/kassak/o:atett-1.tei.2"/>
  <pro:author>
    <foaf:Person rdf:about="http://data.digiphil.hu/data/agent/pim/kassak/Vilmos_Rozv%C3%A1nyi">
      <skos:prefLabel>Vilmos Rozványi</skos:prefLabel>
    </foaf:Person>
  </pro:author>
</edm:ProvidedCHO>

```

[28. ábra] Részlet Rozványi Vilmos Novella és egyéb kísérletek című cikkének DM2E reprezentációjából.

A fenti példa, amely a teljes RDF csupán egy részlete, a hierarchia legalsó szintje, jól illusztrálja a séma bonyolult felépítését. A DM2E projektben elkezdtek egy olyan program fejlesztését, amely képes lett volna azt az RDF konverziót elvégezni, melynek során a DM2E RDF-ből automatikusan előállt volna az EDM szintaxisának megfelelő RDF, azonban időközben a program fejlesztése leállt és az előállított fájlok nem feleltek meg az elvárt formátumnak, ezért saját algoritmust fejlesztettünk.

Az Europeana nem támogatja a hierarchiák közzétételét, így egy RDF csupán egyetlen digitális objektumot reprezentálhat. A *Tett* esetében ezért a folyóiratszámok és a maga a folyóirat is önálló objektumként jelenik meg. A folyóiratcikkek esetén itt is egy XSLT segítségével készültek az RDF-ek a TEI XML fájlokból, míg a folyóiratot és az egyes számokat leíró RDF-eket manuálisan készültek. Az egyes objektumok közötti kapcsolatot itt is jelölni lehet, minden cikk RDF-ébe bekerült, hogy mely folyóirat része, az „isPartOf” elem segítségével, illetve a folyóiratszámokba is bekerültek a benne található cikkek URI-jai a „hasPart” elem alá, valamint a folyóiratnál megjelennek a folyóiratszámok, a szolgáltatásban is látszódnak ezek a kapcsolatok. Mivel az Europeana, az EDM modellnek csak egy részét tudta implementálni, az előállított RDF fájlok is jóval egyszerűbbek, mint a DM2E esetében.

```

<edm:ProvidedCHO rdf:about="http://digiphil.hu/o:atett-1.tei.3">
  <dc:creator>Rozványi Vilmos</dc:creator>
  <dc:contributor>Kassák Lajos</dc:contributor>
  <dc:contributor>http://viaf.org/viaf/100269972/</dc:contributor>
  <dcterms:issued>1915-11-01</dcterms:issued>
  <dc:title>Novella és egyéb kísérletek</dc:title>
  <dc:description>4-6 p.</dc:description>
  <dc:language>hu</dc:language>
  <dc:type rdf:resource="http://vocab.getty.edu/aat/300048715"/>
  <edm:type>TEXT</edm:type>
  <dcterms:isPartOf rdf:resource="http://digiphil.hu/context:atett-1"/>
</edm:ProvidedCHO>

```

[29. ábra] Részlet Rozványi Vilmos *Novella és egyéb kísérletek* című cikkének EDM reprezentációjából.

Az RDF-eket többféleképpen lehet eljuttatni az Europeana számára, a DigiPhil az úgynevezett OAI PMH (Open Archive Initiative-Protokoll for Metadata Harvesting) protokollt³¹⁹ használta erre a célra. Az Europeana a szintaktikai ellenőrzést követően lehetőséget biztosít az aggregátoroknak a szemantikai ellenőrzésre, amelyet egy, az Europeana éles oldalához hasonló tesztfelületen végezhetik el.

A DigiPhil akkoriban fejlesztett egy keresőfelületet, amely valójában egy facettált böngésző volt, a keresési feltételek és a találatok esetében az úgynevezett facetták segítségével szűkíthetők, illetve bővíthetők azok a halmazok, melyekben a felhasználó keresni kívánt. A keresések az RDF-ekben történtek SPARQL lekérdezések segítségével. A DM2E projektbe aggregált RDF-ek az általuk biztosított adatböngészőben voltak elérhetők. A DM2E projekt lezárultát követően nem állt rendelkezésre forrás a szolgáltatás fenttartására, az általuk kiépített rendszer lassan leállt, így a DigiPhil *A Tett*-en kívül nem aggregált más adatokat. Ennek ellenére a projekt során szerzett tapasztalatok rendkívül értékesnek bizonyultak a szemantikus web és a digitális filológia kapcsolatának megértésére és ezek a tapasztalatok nagyban hozzájárultak az előző fejezetben bemutatott fejlesztésekhez.

Az Europeana ezzel szemben a mai napig dinamikusan fejlődik, a szolgáltatás minőségének javításán folyamatosan dolgoznak. A rendszerben található több mint 50 millió rekordból 1 millió közvetve, vagy közvetlenül magyar aggregátoron keresztül érkezett, a

³¹⁹ Az OAI egy kommunikációs protokoll, amely lehetővé teszi, hogy egy szolgáltatás több adatforrásból is azonos módon jusson adatokhoz. Az architektúra lehetővé teszi, hogy az egyes adatszolgáltatók, illetve aggregátorok egy hálózati protokoll mentén, tehát fizikai kapcsolat és kapcsolatfelvétel nélkül legyenek képesek együttműködni. A szabvány HTML és XML alapú és kizárólag a kommunikáció formáját írja le, a metaadat-szabványt nem.

DigiPhil több mint 200 ezer rekordot aggregált különböző magyarországi intézmények gyűjteményeiből.

Az Europeana szolgáltatása széles körben használt, azonban az látható, hogy a rendszerben található tengernyi adat között nehéz eligazodni, ebben a rendszerben, ha valaki nem megfelelően használja a sémát, elveszhetnek az adatok a találati halmazban, ennek elkerüléséhez pedig magas fokú informatikai ismeretekre van szükség. Az Europeana maga is változtatott az elmúlt években a gyakorlatán, az aggregált adatokat, azok gazdagsága, vagy „minősége” szerint meghatározott kategóriákba sorolja, minél magasabb kategóriába tartozik az adott rekord, annál előrébb kerül a találati listákban. Az egyik legfontosabb kritérium a kiváló minőségű digitális szurrogátum szolgáltatása, ez a filológiai jellegű aggregációs tevékenységnél kulcsfontosságú: ha jó minősítést akarunk elérni, akkor nagyfelbontású faksimilét kell szolgáltatnunk a szövegek mellett, ez pedig nem minden esetben triviális feladat. Nem csak az Europeana-ba történő aggregáció növelheti egy kiadás láthatóságát, más népszerű tartalmakkal való összekapcsolás is szolgálhatja ezt a célt, nem kisebbítve ezzel a kiadás tudományos jellegét. A következőkben egy olyan projektet mutatok be, ami egy annotált szövegkiadást kapcsolt külső, népszerű tartalomhoz.

A Kassák Múzeum kutatási projektjének célja, hogy a múzeum a folyóiratkutatás terén Közép- és Kelet Európa nemzetközileg is elismert avantgárd kutatóközpontjává váljon és közben feltárja Kassák Lajos teljes életművét, illetve annak nemzetközi kontextusait.

A projekt Csaplár Ferenc munkáira támaszkodott³²⁰ (Csaplár 1987, Csaplár 2006), illetve a Sasvári Edit és Csatlós Judit által rendezett állandó kiállítás, a *Kassák!*,³²¹ eredményeire, melyek a hangsúlyt a Kassák életműre, nemzetközi kapcsolatrendszerének kutatására is helyezték. Ez a szemléletmód tette lehetővé, hogy a múzeum több, kisebb, korábban alig ismert avantgárd gyűjteményből álló kiállítást hozzon létre.³²²

A kutatócsoport a kiállításokon túl azt tűzte ki célul, hogy összegyűjti és annotált formában publikálja a Kassák Lajos által szerkesztett folyóiratokat (*A Tett, MA, 2x2, Dokumentum*). Kísérletet tettek arra, hogy a kurrens irodalomtudományi, képzőművészeti és

³²⁰ CSAPLÁR Ferenc, *Kassák körei* (Budapest: Szépirodalmi, 1987); CSAPLÁR Ferenc, *A Kassák múzeum kiállításai 1976-2006* (Budapest: Petőfi Irodalmi Múzeum - Kassák Múzeum, 2006).

³²¹ SASVÁRI Edit és CSATLÓS Judit (kurátor): „Kassák! A Kassák Múzeum állandó kiállítása” (Kiállítás, Kassák Múzeum, Budapest, 2011. április 19.).

³²² „Az avantgárd magángyűjteményekben I. – A Kiss Ferenc-gyűjtemény” (Kiállítás, Kassák Múzeum, Budapest, 2011. november 26.); MARINKO SUDAC, SASVÁRI Edit (kurátor): „Az avantgárd magángyűjteményekben II. A kölcsönös hatások körei. A MA és a Zenit a zágrábi Marinko Sudac-gyűjteményben.” (Kassák Múzeum, Budapest, 2012. január 28.).

történelemtudományi megközelítésmódokkal helyezték kontextusba a folyóiratokat. Ennek első lépéseként három nemzetközi workshopot rendeztek,³²³ amelyeken bemutatásra kerültek azok az avantgárd kutatások, amelyek szorosan kapcsolódnak a Kassák tematikához. A kutatócsoport 2015. januárjában indította el az Online Avant-Garde Database-t, azzal a céllal, hogy összekösse az európai avantgárd kutatói közösséget.³²⁴ A honlap szerkesztői folyamatosan frissítik az avantgárddal foglalkozó legújabb publikációk listáját, a látogatható kiállításokat, konferenciákat és az egyéb kutatási eredményeket Európában.

A kutatási projekt első feldolgozott folyóirata *A Tett* volt, amelynek eredményeiből a *Jelzés a Világba – Háború ∩ Avantgárd ∩ Kassák* címen rendeztek kiállítást. A folyóirat annotált forráskiadását a faksimilékkel együtt, a DigiPhil közreműködésével jelölőnyelvi átiratban publikálták.³²⁵ A kutatócsoport tagjai, Dobó Gábor és Szeredi Merse Pál, a Visegrádi Alap támogatásával pedig nemzetközi konferenciát szerveztek *Local Contexts / International Networks – Avant-Garde Magazines in Central Europe (1910–1935)* címmel.³²⁶ A konferencia központi témája a közép- és kelet európai avantgárd folyóiratok voltak, amelyek 1910 és 1935 között fontos felületei voltak az irányzat irodalmi és művészeti alkotásainak.

A kutatócsoport a kiadás és a kiállítás készítésekor is követte a legújabb nemzetközi trendeket. *A Tett* Kassák Lajos szerkesztésében megjelent első avantgárd folyóirat, amely 17 lapszámot élt meg 1915-1916 között. A kiállítás apropóját, amellet, hogy Kassáknak ez egy kevésbé kutatott folyóirata, a folyóirat megjelenésének 100 éves évfordulója adta. A kiállítás koncepciója az volt, hogy bemutassa, hogy a folyóirat hogyan próbált szembe szegülni korának háborúpárti attitűdjével, a „háború kultúrájával”. A kiállítás elkészültét alapos kutatómunka

³²³ Gábor ANDRÁSI, szerk., „...fejünkéből töröljük ki a regulákat.” *Kassák Lajos az író, képzőművész, szerkesztő és közszereplő* (Budapest: PIM-Kassák Alapítvány, 2010)., „A magyar avantgárd és a szomszédos országok avantgárd irodalmi és művészeti termése a Kárpát-medencében 1919 és 1929 között. Közös témák, közös kutatási tervek, pályázatok.” (Kassák Múzeum, Budapest, 2011. július 1.)., „Kassák és Kassák 2.” (Kassák Múzeum, Budapest, 2014. április 9.).

³²⁴ Az adatbázis (<http://avantgardedb.org/>) már nem elérhető, az Internet Archive szerint 2018. május 27-én még elérhető volt: <http://web.archive.org/web/20180527012654/http://avantgardedb.org/>.

³²⁵ A jelölőnyelvi átirat lehetővé tette volna más nemzetközi projektekben közzétett avant-garde folyóiratokkal való összekapcsolást. A Ryerson University a Yellow Nineties Online szolgáltatása TEI átiratban teszi közzé faksimile kiadásait, a metaadatokat RDF formátumban, kritikai jegyzeteket pedig TEI XML formátumban teszik elérhetővé: <https://1890s.ca/about/>. A Princeton-on futó Blue Mountain projekt szintén faksimile kiadásban szolgáltatja a folyóiratokat, azonban a kifinomult keresőfelületük a jelölőnyelvi átiratokban keres, amelyeket az OCR-ezett faksimilékéből állítottak elő: <https://github.com/Princeton-CDH/bluemountain-transcriptions>. Az adatbázisukban egyébiránt három, Kassáktól származó szöveg található: <https://bluemountain.princeton.edu/bluemtn/?a=q&hs=1&r=1&results=1&txf=txIN&txq=Kass%C3%A1k&e=---en-20--1--txt-txIN-kass%C3%A1k----->.

³²⁶ Gábor DOBÓ és Merse Pál SZEREDI, szerk., *Local Contexts / International Networks: Avant-Garde Journals in East-Central Europe* (Budapest: Petőfi Literary Museum – Kassák Museum Kassák Foundation, 2018), https://monoskop.org/images/b/b2/Local_Contexts_International_Networks_Avant-Garde_Journals_in_East-Central_Europe_2018.pdf.

előzte meg, amely a múzeumban található forrásdokumentumokra és a korabeli recepcióra épült.³²⁷ Emellett feltárták a hazai és nemzetközi kontextusokat és kísérletet tettek a folyóirat elhelyezésére a korabeli kultúra viszonyrendszerében. A kiállítás tervezésében nem törekedtek egy előre meghatározott, lineáris bejárési útvonal kiépítésére, a kiállítást négy teremben helyezték el, „[...] mindegyik önálló és a többi nélkül is érthető, s mindegyik egészen más metszetét adja a középpontba állított folyóiratnak.”³²⁸ A különböző termeket a látogató saját érdeklődése szerint járhatta be, a falakon különböző színű vonalak kötötték össze a kiállított dokumentumokat, amelyek egyszerre képezték a dizájn részét és jelöltek kapcsolatrendszereket. Az infografika alkalmazása egyre népszerűbb a múzeumi világban, segítségével a kiállításon lehetővé vált a bonyolult összefüggések vizualizációja, illetve, hogy ezek felfedezését, értelmezését a látogató maga végezze el.³²⁹ Ebben a kiállításban azonban a Rudas Klára által tervezett, újszerű infografika „egyenrangú minőséget képvisel a kurátori / kutatói koncepcióval azzal, hogy vizualizálni képes azt.”³³⁰ Maguk a kurátorok így fogalmazzák az infografikával kapcsolatban:

A kiállításban fontos szerepet játszik az infografika mint az elemzés sajátos vizuális eszköze. A kurátorok ezzel természetesen nem mondanak le az értelmezés munkájáról, de tágas teret szeretnének nyitni az egyéni interpretációk létrehozásában. Az infografikák bonyolult összefüggéseket közvetítenek anélkül, hogy a kutatások eredményeit előre összegeznék és tetszetős narratívákba csomagolnák. Ugyanakkor a kiállítás ambíciója csak látszólag szerény: azoknak az árnyalatoknak és ellentmondásoknak az érzékeltetésére vállalkozik, amelyek miatt ma is izgalmas *A Tett* és kora.³³¹

A kutatócsoport az időszak kiállításból virtuális kiállítást készített, ami a kiállítás „kutatói és oktatási segédanyagaként készült”.³³² A virtuális kiállítás részben megőrzi, archiválja az eredeti kiállítást, azonban jellegéből fakadóan bizonyos kontextusok elvesznek, miközben új információk érhetőek el. Míg az időszak kiállítás főként a magyarországi, azon

³²⁷ Interjú Szeredi Merse Pállal és Dobó Gáborral: http://itadokt.hu/jelzes_a_vilagba_-_a_kassak_muzeum_kiallitasa.

³²⁸ KÁLMÁN C. György, „Vicc, novella, tanulmány. Jelzés a világba. A Tett. Háború, avantgárd, Kassák. A Kassák Múzeum (Budapest) kiállítása”, *Jelenkor*, 2015. augusztus 1., <https://www.jelenkor.net/visszhang/467/vicc-novella-tanulmany>.

³²⁹ A digitális bölcsészletben az adatvizualizáció, a kutatási eredmények grafikus megjelenítése az egyik legaktuálisabb gyakorlat és probléma.

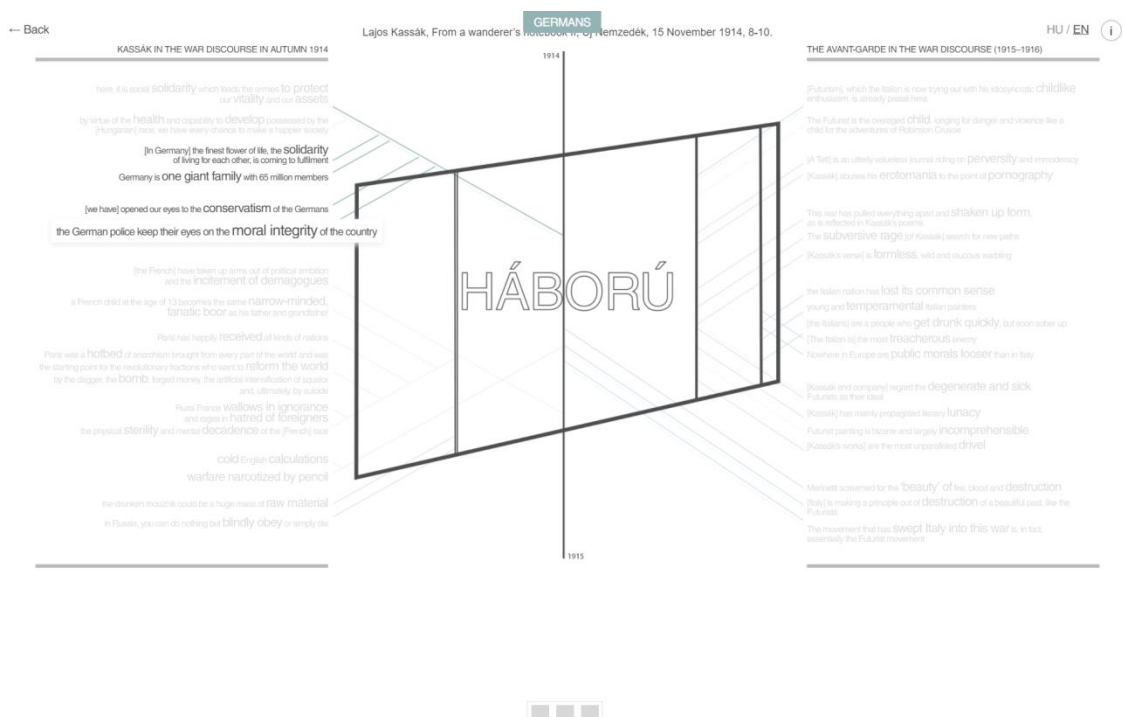
³³⁰ SASVÁRI Edit és DOBÓ Gábor, „Kassák Múzeum: a magyar avantgárd kortárs múzeuma”, *Tiszatáj* 71, 12. sz. (2017): 155–164, 161.

³³¹ A teljes leírás az információ gombra kattintva érhető el: <https://atett.kassakmuzeum.hu/>.

³³² Impresszum: <http://atett.kassakmuzeum.hu>.

belül is a budapesti közönséghez jutott el, a kétnyelvű virtuális kiállítás a nemzetközi közönséget is elérheti.

A virtuális kiállításon három „fal” tekinthető meg, amelyek az eredeti kiállítás egy-egy termét reprezentálják és amelyeknek vizualizációja – ellentétben a fizikai kiállítással – dinamikus, miközben megőrizte a fizikai kiállítás dizájnját. Az első falon, hasonlóan a fizikai kiállításhoz, *A Tett* címlapja a korabeli folyóiratok címlapja mellett látható, azonban itt, a címlapokra kattintva egy lapozóhoz jutunk, elolvashatjuk a kiállított számot. A második falon a fizikai kiállításon is látható idézetek olvashatók, Kassáknak a háborúról, valamint a háborúpárti országokról és mások Kassákról, az avantgárdról alkotott véleményét láthatjuk. Az egyes idézetek – a kurzor mozgatására - kiemelkednek a többi közül, a kapcsolatokat szimbolizáló vonalakkal együtt, kontextusuk feltárul a képernyőn.



[30. ábra] Részlet a virtuális kiállításból.³³³

A harmadik fal *A Tett*-ben megjelent külföldi irodalmi és képzőművészeti alkotásoknak a válogatási koncepcióját mutatta be. A honlapon egymás mellett olvashatók az irodalmi alkotások eredeti nyelvű szövegei és *A Tett*-ben megjelent fordításuk, a képzőművészeti alkotások esetében az eredetiről készült fotó és a folyóiratban megjelent változat tekinthető meg. Az eredeti kiállítás negyedik terme nem került be a virtuális kiállításba, ott *A Tett* körüli

³³³ Forrás: <https://atett.kassakmuzeum.hu/wall2.html>.

vitákat mutatták be, megjelenésének, majd betiltásának dokumentumait. A DigiPhil-en megjelent annotált kiadásra a virtuális kiállítás több helyen is utalt, így a kiállítást böngésző felhasználók akaratlanul is eljuthattak a DigiPhil szolgáltatás felületére, ahol a további kiadásokat is böngészhették.

A múzeum látogatói nem csak *A Tett*, illetve a virtuális kiállítás kapcsán találkozhattak a DigiPhil kiadásaival. A Kassák Múzeum 2017-ben Kassák Lajos születésének 130. és halálának 50. évfordulójára *Kassákizums* címmel kiállítássorozatot rendezett.³³⁴ A sorozat második része, *Az új Kassák – a ló meghal és a madarak kiröpülnek*³³⁵ 2017. október 13. és 2018. február 25. között volt megtekinthető. Fókuszában a „Kassák talán legismertebb és legtöbbet idézett önéletrajzi ihletésű remekműve, *A ló meghal és a madarak kiröpülnek* című költemény áll.”³³⁶ A kiállítás előkészítésekor a művet és annak fordításait³³⁷ a DigiPhil jelölőnyelvi átiratban feldolgozta és közzétette az oldalán.³³⁸ A szolgáltatásban kétféle nézetet biztosítottunk, a mű és fordításait egyenként (tehát önálló digitális objektumként) és soronként egymáshoz rendelve is közöltük. A kiállítás terében a látogatók szabadon böngészhették a művet és a DigiPhil további kiadásait, a digitális kiadás a tárlat részévé vált, ezzel olyan látogatókhoz is eljuthatott a DigiPhil híre, vagy akár maguk a kritikai kiadások is, akik máshonnan nem szerezhettek volna tudomást róluk.³³⁹

Ezeknél a projekteknél a digitális filológusok és a művészettörténészek együttműködésével olyan kontextusba kerülhettek a kiadások, amelyek már túlmutatnak a digitális filológia szűkebb értelemben vett keretein. A szövegek, fakszimilék és metaadatok szabványos közzétételén túl lehetőségünk van számítógépes eszközökkel elemezni a jelölőnyelvi átiratokat, az eredmények pedig alátámaszthatják, vagy cáfolhatják a korábbi kutatások eredményeit, illetve új szempontokat nyújthatnak egy-egy szerző életművének kutatásában. Ehhez kapcsolódóan érdemes megvizsgálni a távoli olvasás, azaz a distant reading és a stilometria technikáit is, hiszen ezek a módszerek kritikai kiadásokra egyaránt alkalmazhatók. A szövegstatisztikai módszerekkel olyan, adott szerzőre jellemző mintázatokat

³³⁴ A kiállítássorozatról bővebben lásd: <https://pim.hu/hu/kiallitas/kassakizmus>.

³³⁵ SASVÁRI Edit, SZEREDI Merse Pál (kurátorok): „Kassákizmus 2. Az új Kassák – a ló meghal és a madarak kiröpülnek” (Kiállítás, Kassák Múzeum, Budapest, 2017. október 13.).

³³⁶ Idézet kiállítás programjából: <https://kassakmuzeum.hu/hu/kiallitas/kassakizmus-2?p=kiallitas&id=356>.

³³⁷ A korábban említett Blue Mountain adatbázisban elérhető a mű eredeti német megjelenése: <https://bluemountain.princeton.edu/bluemtn/?a=d&d=bmtnabg19230301-01.2.4&srpos=3&e=-----en-20--1--txt-txIN-Kass%c3%a1k----->.

³³⁸ A Petőfi Irodalmi Múzeummal kötött megállapodás értelmében a művet eltávolítottuk a szolgáltatásból. A kiadás megtekinthető az Internet Archive-ban: <http://web.archive.org/web/20220813020844/https://digiphil.hu/context:alomeghal>.

³³⁹ A kiállításból készült interaktív online kiállítás elérhető a Kassák Múzeum oldalán: <https://nikkelszamovar.kassakmuzeum.hu/>.

deríthetünk fel, melyek korábban, erőforrás hiányában, sokszor láthatatlanok maradtak a kutatók számára.

Jelölőnyelvi átiratok elemzése gépi olvasás segítségével

A Moretti által bevezetett distant reading, magyarul távoli (egyreszerzőknél távolságtartó) olvasás fogalma jól ismert, először egy 2000-ben publikált tanulmányában jelent meg,³⁴⁰ ekkor még az összehasonlító irodalomtudománnyal összefüggésben kapott jelentést, később, ahogyan Labádi Gergely Moretti *Distant Reading* könyvéről írt kritikájában rámutat, a fogalom „látens alakulástörténete”³⁴¹ olvasható ki a tanulmánykötetből. Ahogyan Labádi fogalmaz, a distant reading egy olyan fogalom,

amellyel Moretti a nagy mennyiségű adat beszerzését és értelmezését legitimálni tudja: elveszíti ugyan magát a szöveget, de megnyeri a szöveg alatti (témák, trópusok) és feletti szinteket (műfajok, rendszerek). Az egyes tanulmányok segítségével aztán a regények terjedésének, meghonosodásának korokon és kultúrákon átívelő törvényszerűségeit tárja fel.³⁴²

Moretti hatása nehezen túlbecsülhető, a kvantitatív irodalmi szövegvizsgálatok alkalmazásának szimbólumává vált és sokat tett a módszerek elfogadtatásáért az irodalomtudomány területén. A dolgozat szempontjából azonban érdemesnek tartom kiemelni azokat a kritikákat, amelyek az elmúlt két évtizedben születtek, és amelyek a Moretti és az általa alapított Stanford Literary Lab kutatói által vizsgált korpuszok filológiai minőségére vonatkoznak. Arra már Labádi is rámutatott, hogy Moretti nem számolt a digitális filológia újításaival, a jelölőnyelvi átiratokkal, amelyek megváltoztathatják a distant reading által alkalmazott módszereket.³⁴³ Morettit, a méltatások mellett, számos kritika érte, ezek egy része az Association of Literary Scholars and Critics (ma: Association of Literary Scholars, Critics, and Writers) által fentartott *The Valve* blogon jelent meg 2006-tól kezdve.³⁴⁴ Az ott megjelent kritikákat végül 2011-ben adták ki

³⁴⁰ Franco MORETTI, „Conjectures on World Literature”, *New Left Review* 2000, 1. sz. (2000): 54–68.

³⁴¹ LABÁDI Gergely, „Franco Moretti: Distant Reading”, *Irodalomtörténet* 95, 4. sz. (2014): 561–564, 562.

³⁴² Uo.

³⁴³ Uo., 564.

³⁴⁴ A blog már nem elérhető, az Internet Archive-ban azonban megtekinthetők a bejegyzések.

könyv formában, amelyben Moretti három válasza is helyet kapott.³⁴⁵ Azonban ezekben a kritikákban, ahogyan Joyeux-Prunel írja 2024-ben, „valójában nem vonták kérdőre a korpusz és a módszerek megválasztásával kapcsolatban és a megközelítéseinek korlátait sem vizsgálták.”³⁴⁶ Joyeux-Prunel azonban nem említi meg a későbbi kritikákat, Moretti (és kollégája, Matthew L. Jockers)³⁴⁷ módszereit például Katherine Bode több művében is bírálta.³⁴⁸ A módszertani szempontokon túl Bode számonkéri Morettin – és a Stanford Literary Lab által közzétett munkák³⁴⁹ szerzőin – hogy nem osztják meg az adataikat,³⁵⁰ így az elemzések és következtetések nem ellenőrizhetők és nem reprodukálhatók. A másik szempontja pedig egybecseng Labádi saját kutatásai során megfogalmazott gondolataival, Bode bírálja a Moretti által használt forrásokat, szerinte azok nem felelnek meg a tudományosság kívánalmainak, így a következtetések sem pontosak. Bode amellet érvel, hogy a distant reading kutatások során tudományos szövegkiadásokat kellene használni, mert azok, azon túl, hogy filológiai pontossággal készültek, tudományos és történeti kontextussal is szolgálnak.³⁵¹ Fontos kiemelni, hogy distant reading kutatásokat nemcsak magukon a szövegeken lehet végezni, hanem azok metaadatain is,³⁵² tehát jelen esetben nem csupán a szövegek filológiai értelemben vett pontosságáról beszélünk, hanem a felhasznált adatbázisok megbízhatóságáról is. Természetesen ennek történeti kontextusát is érdemes figyelembe venni, sokáig nem is voltak elérhetőek digitális tudományos szövegkiadások, vagy nagy mennyiségű, megbízható adatot tartalmazó bibliográfiai adatbázisok, az mindenesetre érdekes, hogy Moretti nem reflektált erre a problémára. Labádi *Az olvasó gép. Berzsenyi Dániel versei távolról* című tanulmányában a

³⁴⁵ Jhon HOLBO és Jonathan GOODWIN, szerk., *Reading Graphs, Maps, Trees – Critical Responses to Franco Moretti* (Parlor Press, 2011).

³⁴⁶ Béatrice JOYEUX-PRUNEL, „Digital Humanities in the Era of Digital Reproducibility: Towards a Fairest and Post-Computational Framework”, *International Journal of Digital Humanities*, 2024. január 3., <https://doi.org/10.1007/s42803-023-00079-6>. [a kiadásban nem használnak paginációt]

³⁴⁷ Jockers a distant reading fogalma helyett a makroanalízis fogalmat részesíti előnyben, ezzel kapcsolatos érveléséhez lásd a személyes weboldalán közzétett gondolatait: Matthew L. JOCKERS, „On Distant Reading and Macroanalysis”, *Matthew L. Jockers* (blog), 2011. július 1., <https://www.matthewjockers.net/2011/07/01/on-distant-reading-and-macroanalysis/>.

³⁴⁸ Bode érvélését követve született például az alábbi tanulmány is: James DOBSON és Scott SANDERS, „Distant Approaches to the Printed Page”, *Digital Studies / Le Champ Numérique* 12, 1. sz. (2022. május 5.), <https://doi.org/10.16995/dscn.8107>.

³⁴⁹ A Stanford Literary Lab a kutatások eredményeit úgynevezett pamfletlek formájában teszi közzé weboldalán: <https://litlab.stanford.edu/pamphlets/>.

³⁵⁰ Katherine BODE, „Abstraction, Singularity, Textuality: The Equivalence of “Close” and “Distant” Reading”, in *A World of Fiction: Digital Collections and the Future of Literary History*, 17–36 (University of Michigan Press, 2018), 24.

³⁵¹ Katherine BODE, „The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”, *Modern Language Quarterly* 78, 1. sz. (2017): 77–106.

³⁵² Magyarországon ehhez lásd például: PÉTER Róbert, „A big data kihívás és lehetőség a bölcsészettudományokban: digitális szövegek és metaadatok távoli olvasása”, *Magyar Tudomány* 177, 11. sz. (2016): 1323–1330.

Magyarországon rendelkezésre álló szövegek minőségével kapcsolatban így fogalmaz: „[...] a kétrétegű PDF például korántsem a digitális mint olyan kvintesszenciája; a DIA anyaga nem könnyen hozzáférhető, a MEK szövegeinek megbízhatósága pedig problematikus, azaz a magyarországi digitalizált szövegek nem egyenletes minőségűek, és nem egységes elvek szerint készültek.”³⁵³ Jobb alternatíva híján Labádi a MEK-en található korpuszt használta, akárcsak későbbi, Mikszáth prózáját vizsgáló kutatásában,³⁵⁴ a szövegeket manuális javítás után a *Magyarlanc*³⁵⁵ nyelvi elemzővel készítette elő a vizsgálataihoz. Kiemelendő, hogy Labádi tanulmányaiban pontosan részletezte a szövegek forrását, az előkészítés lépéseit és a vizsgálati módszereit, a mai napig elérhető blogján pedig további módszertani részleteket is közölt.³⁵⁶

Moretti hatására mára megszámlálhatatlan kutatás született a témában, ezek közül kettőt emelnék ki, amelyeknek módszereit ismerem és amelyeket jelölőnyelvi átiratok alapján készültek. A Distant Reading for European Literary History projekt 2018-ban indult azzal a céllal, hogy olyan, többnyelvű európai korpuszt hozzon létre, amely alkalmas távoli olvasással végzett kutatások támogatására. A projekt elsődleges célja 10 európai, egyenként 100-100, 1840 és 1920 között megjelent, meghatározott szempontok szerint válogatott,³⁵⁷ regény jelölőnyelvi átiratban való összegyűjtése és közzététele, a European Literary Text Collection (ELTeC) szabadon hozzáférhető a GitHubon.³⁵⁸ A regények jelölőnyelvi átiratának specifikációja a Text Encoding Initiative ajánlása alapján készült, de a korpusz előállításának felgyorsítására, az elemző algoritmusok futtatásának, valamint a későbbi felhasználásnak a megkönnyítésére egy jelentősen szűkített elemkészletből álló sémát készítettek.³⁵⁹ A regényeket nyelvi elemzőkkel vizsgálták, amelynek eredményeit a jelölőnyelvi átiratban is

³⁵³ LABÁDI Gergely, „Az olvasó gép: Berzsenyi Dániel versei távolról”, *Digitális Bölcsészet* 1 (2018. szeptember 19.): 17–34, 19, <https://doi.org/10.31400/dh-hun.2018.1.126>.

³⁵⁴ LABÁDI Gergely, „Géppel mért irodalom: a mikszáthi élőbeszédyszerűség”, *Digitális Bölcsészet* 2 (2019. október 11.): M:3-M:19, <https://doi.org/10.31400/dh-hun.2019.2.390>.

³⁵⁵ János ZSIBRITA, Veronika VINCZE és Richárd FARKAS, „Magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian”, in *International Conference on Recent Advances in Natural Language Processing*, szerk. Ruslan MITKOV, Galia ANGELOVA és Kalina BONTCHEVA, 763–771 (Shumen: Incoma Ltd., 2013).

³⁵⁶ Ebben a bejegyzésében lépésről-lépésre mutatja be a Mikszáth-korpusz előkészítését: <https://labadigergely.github.io/szovegek/2017/05/17/R6/>.

³⁵⁷ Ilyen szempontok voltak például terjedelem, a nemek közti eloszlás, illetve a kanonicitás. Ezek részletes elemzéséhez lásd: BAJZÁT Tímea Borbála, SZEMES Botond és SZLÁVICH Eszter, „Az ELTE DH Regénykorpusz és lehetőségei”, in *Online térben az online térért: Workshop 30: országos online konferencia. 2021. április 6-9. Eötvös Loránd Tudományegyetem*, szerk. KOKAS Károly, TICK József és HOLL András, 63–72 (Budapest: Hungarnet Egyesület, 2021), https://real.mtak.hu/132253/1/63_NWS_2021_v3.pdf.

³⁵⁸ Az ELTeC korpusz elérhetősége: <https://github.com/COST-ELTeC>.

³⁵⁹ A séma kidolgozásáról bővebben lásd: Lou BURNARD, Christof SCHÖCH és Carolin ODEBRECHT, „In Search of Comity: TEI for Distant Reading”, *Journal of the Text Encoding Initiative*, 14. sz. (2021), <https://doi.org/10.4000/jtei.3500>.

kódolták. Az ELTE Digitális Bölcsészet Tanszéke képviselte Magyarországot a projektben. A magyar nyelvű regényeket főként a MEK adatbázisából gyűjtötték és félautomatikus módszerekkel alakították át az ELTeC sémájának megfelelő jelölőnyelvi átíráttá, majd az e-magyar³⁶⁰ nyelvi elemző segítségével vizsgálták.³⁶¹ Az elemzés során kiemelt szempont volt a nemzetközi projekt elvárása, hogy a nyelvi elemzés eredményei egységesen kerüljenek leírásra a jelölőnyelvi átíratokban, ezért a *Universal Dependencies and Morphology* nevű nemzetközi sémát alkalmazták.³⁶² A korpusz része az ELTeC korpuszának, de a Tanszék további regények beemelésével bővebb, 400 regényből álló korpuszt hozott létre és olyan keresőfelületet fejlesztett,³⁶³ amely lehetővé teszi az irodalomtörténészek számára a különböző szempontú elemzések végrehajtását. Bajzát Tímea Borbála, Szemes Botond és Szlávich Eszter elemzésükben a 19. század során fellelhető tendenciát, a mondatok hosszúságának rövidülését vizsgálták.³⁶⁴ Ugyanilyen típusú elemzéseket végezhetünk az ELTE által készített további korpuszokban, mint például a Verskorpuszban,³⁶⁵ illetve a Drámakorpuszban.³⁶⁶ Ezek a korpuszok hasonló módszertannal és elvek mentén készültek, mint a Regénykorpusz, azonban a Drámakorpusz esetén egy kivételt ki kell emelnem, Az ember tragédiája a DigiPhil által informatizált kiadás megállapított szövegből került átemelésre, tehát, szemben a korpuszokban található, internetes forrásokból félautomatikus módszerekkel importált szövegekkel, ez ellenőrzött forrásból származik. A drámák esetében a korpusz létrehozásakor figyelembe vettük a DraCor projekt³⁶⁷ gyakorlatát és a jelölőnyelvi átíratokat eljuttattuk a DraCor korpuszába (HunDraCor), így nem csak nyelvt statisztikai vizsgálatokat lehet végezni, hanem a DraCor felületén a drámaszövegeken belüli kapcsolati hálózatokat is lehetőségünk van kutatni.³⁶⁸

³⁶⁰ Az e-magyar és az annak továbbfejlesztett változata az emTSV, valamint a Magyarlanc közötti különbségekről lásd: INDIG Balázs és mtsai., „emtsv – Egy formátum mind felett”, in *XV. Magyar Számítógépes Nyelvészeti Konferencia*, 235–247 (Szeged: Szegedi Tudományegyetem Informatikai Intézet, 2019), <http://real.mtak.hu/99685/>.

³⁶¹ BAJZÁT, SZEMES és SZLÁVICH, „Az ELTE DH Regénykorpusz és lehetőségei”, 67.

³⁶² A séma specifikációja az alábbi linken érhető el: <https://universaldependencies.org/>

³⁶³ A keresőfelület az alábbi címen érhető el: <https://regenykorpusz.elte-dh.hu/?lang=hu-HU>

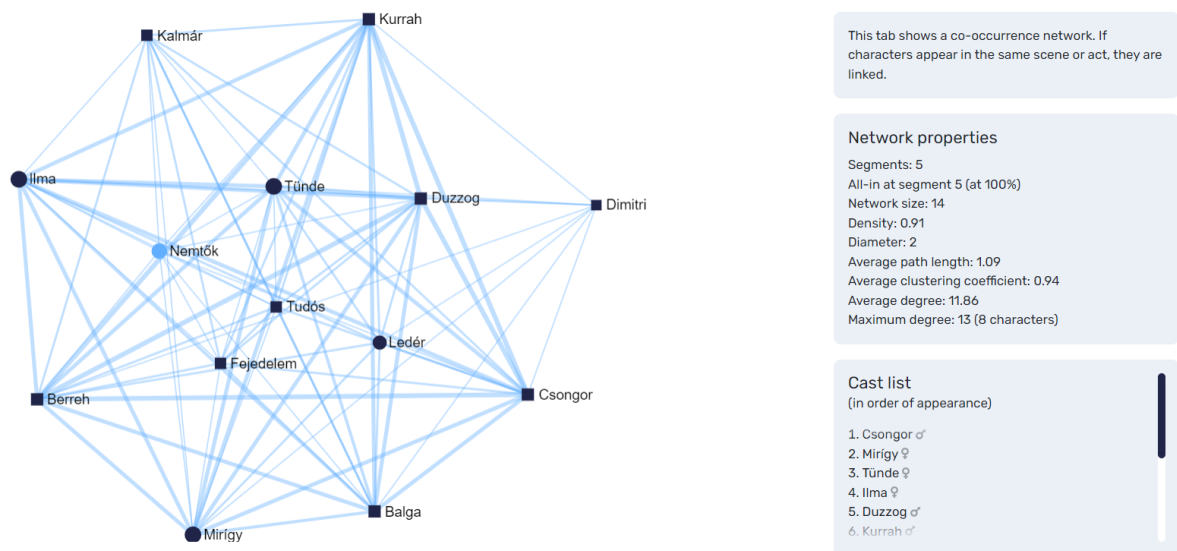
³⁶⁴ BAJZÁT, SZEMES és SZLÁVICH, „Az ELTE DH Regénykorpusz és lehetőségei”, 70.

³⁶⁵ A Verskorpuszban jelenleg 13362 vers található: <https://verskorpusz.elte-dh.hu/?lang=hu-HU>

³⁶⁶ A Drámakorpusz jelenleg 76 drámát tartalmaz: <https://dramakorpusz.elte-dh.hu/?lang=hu-HU>

³⁶⁷ A projektről bővebben lásd: <https://dracor.org/doc/what-is-dracor>. A DraCor a TEI ajánlásának megfelelő jelölőnyelvi átíratokat tartalmaz, a kód fejlécében a dráma szereplőinek a listája is szerepel, illetve a szöveg kódolása során jelöltük, hogy a megszólalások mely szereplőkhöz köthetők. A DraCor algoritmusai többek között ezen jelölők elemzésével készíti a kapcsolati hálózati modelljeit. A DraCor korpuszán végzett vizsgálatokról bővebben lásd: SZEMES Botond és VIDA Bence, „Tragikus és komikus hálózatok. Drámai műfajok csoportosítása szerkezeti tulajdonságok alapján”, *Helikon* 68, 2. sz. (2022): 345–368.

³⁶⁸ SZEMES Botond és mtsai., „Az ELTE Drámakorpuszának létrehozása és lehetőségei”, in *Valós térben - az online térért*, szerk. TICK József, KOKAS Károly és HOLL András, NIIF Networkshop konferenciák; 31., 170–178 (Budapest: Hungarnet Egyesület, 2022), 176–177, <https://real.mtak.hu/155495/>.



[31. ábra] A Csongor és Tünde kapcsolatrendszerének hálózati modellje a DraCor felületén.

A fent részletezett módszerek egyaránt érvényesíthetők a digitális kritikai kiadások jelölőnyelvi átíratain. A következőkben két példát mutatok be, az első, Arany János leveleinek metaadatainak elemzési lehetőségeit vizsgálja különböző eszközökkel, majd egy olyan, saját kísérletet mutatok be, amely során szövegstatistikai módszerekkel vizsgáltam, hogy vajon a rendelkezésre álló informatizált Kosztolányi kritikai kiadások elemzésével a szerző, a kéziratokon végzett utólagos módosításai mögött felsejlik-e valamilyen tendencia.

Arany János levelezésének hálózati vizualizációjának lehetőségei

Arany János kritikai kiadásainak informatizálása lehetővé tette a kritikai kiadások számítógépes vizualizációját. Az informatizálás folyamatát tárgyaló alfejezetben már kiemelttem, hogy a nem szépirodalmi szövegek esetében különös szerepe van a szemantikus adatgazdagításnak, hiszen olyan kiterjedt kapcsolati hálózat jön itt létre, amely számítógépes eszközökkel tovább vizsgálható. Az vizsgálatokhoz és az eredmények prezentálásához is egyre népszerűbb az adatvizualizációk használata. Az informatizálás során a levelezéskiadások esetében különös figyelmet szenteltünk a metaadatstruktúra kialakítására és az adatgazdagításra. A DigiPhil akkor követett gyakorlata azóta megváltozott, az adatgazdagítás a korábban bemutatott ITIdata szemantikus adatbázis segítségével történik. A korábban

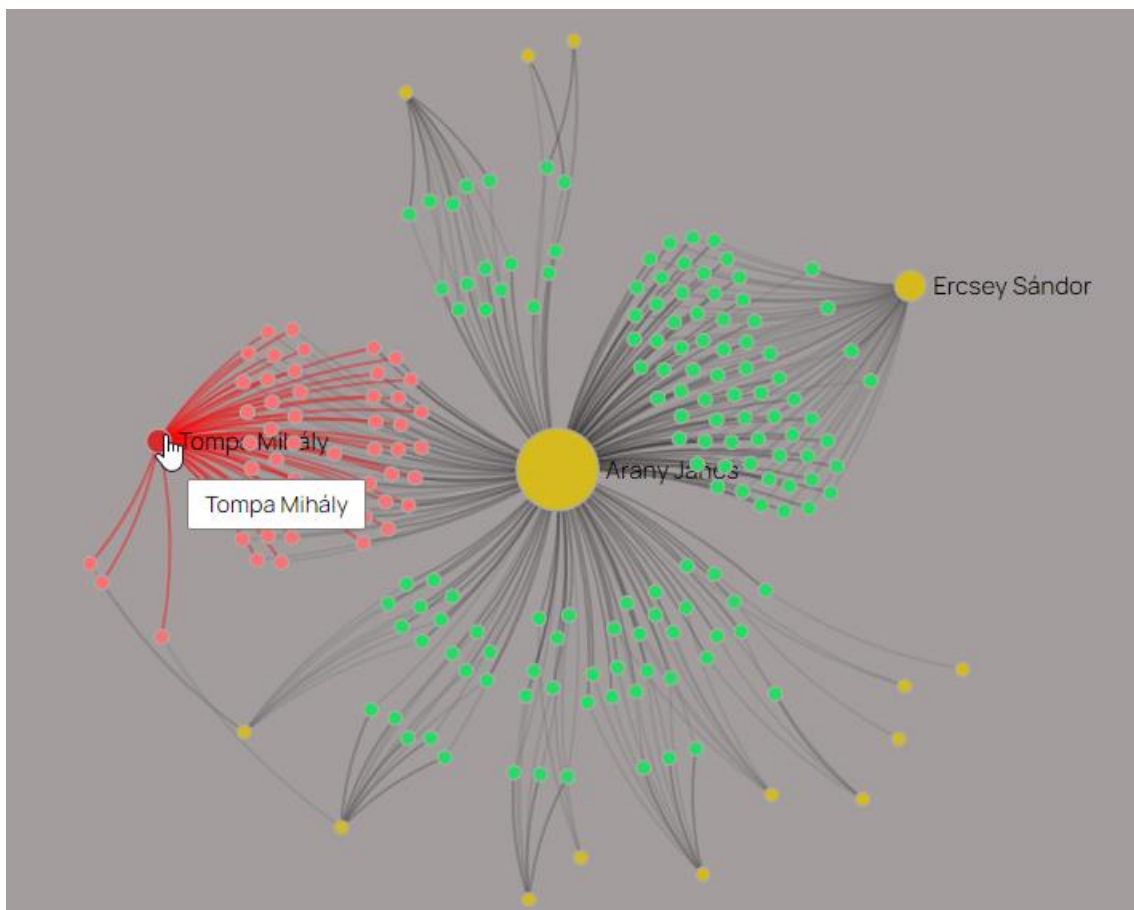
informatizált kiadások metaadatai szintén bekerülnek az új adatbázisba.³⁶⁹ A levelezéskiadások esetében egy levelet két ITIdata rekord reprezentál, az első a kritikai kiadásban közölt levél adatait tartalmazza, a második a levél kéziratának metaadatait, az adatok feltöltése a disszertáció lezárásakor még nem zárult le. Amikor a levelek metaadatainak gépi elemzési lehetőségeit mérlegeltük, még nem álltak rendelkezésre adatbázisba rendezett, strukturált adatok, az első, kísérleti vizsgálatokhoz a TEI XML átiratokat használtuk fel. A kutatás célja a kutatók által széles körben használt Nodegoat hálózatelemzést és adatvizualizációt támogató eszköz által nyújtott lehetőségek vizsgálata volt.³⁷⁰ A Nodegoat, hasonlóan a korábban már bemutatott Neo4j rendszeréhez, ingyenes, de funkcionalitásában korlátozott hozzáférést biztosít a webszolgáltatáshoz, a teljes körű hozzáférés, amely már a kollaborációt is lehetővé teszi, előfizetéshez kötött. A webszolgáltatásban saját adatmodell alakítható ki, az adatok összeköthetők olyan külső névterekkel, mint a Wikidata, a VIAF vagy a GeoNames. A Wikidata esetében a rendszeren belül futtathatunk SPARQL lekérdezéseket, ezzel is gyorsítva az adatgazdagítást. Az ingyenes szolgáltatás legnagyobb hátránya, hogy nincs lehetőség a kutatási adatok közzétételére, azonban a rendszer kifinomult import és export funkciókkal rendelkezik, különféle beépített vizualizációs lehetőségeket kínál, így egy kisebb kutatási projekthez tökéletes választás lehet. Az előfizetéshez kötött változatok esetében a szolgáltatás saját szerveren, vagy felhőszolgáltatásban futtatható, létrehozható nyilvános portál a kutatási adatok közzétételéhez. Ez a szolgáltatás ugyan lehetővé teszi nagyobb kutatási projektek adatainak kezelését, elemzését és vizualizációját, hátránya ugyanaz, mint a Neo4j esetében, a kidolgozott adatmodell nem szabványos és nem felel meg a szemantikus web filozófiájának.

Az Arany levelezés kötetek adatainak vizsgálati lehetőségeit 2020-ban, nem sokkal az Arany-évhez kapcsolódó informatizálási folyamat lezárultát követően kezdtük meg. A rendszer teszteléséhez a regisztrációhoz kötött webszolgáltatást használtuk. Az adatok kinyerése és rendszerezése a jelölőnyelvi átiratokat felhasználva egy Python szkript segítségével történt. A program kiolvasta a jelölőnyelvi átiratokban rögzített metaadatokat (levél írója, levél címzettje, megírás dátuma és megírás helye) és CSV formátumba rendezte azokat. Ahhoz, hogy az

³⁶⁹ A levelek adatait félautomatikus módszerekkel állítja elő a DigiPhil. Az adatokat a TEI XML fájlok fejlécéből Python szkriptek segítségével nyertük ki CSV formátumban. Az így kapott listák alapján fény derült az XML fájlokban található következetlenségekre, hibákra, amelyeket a DigiPhil munkatársai javítanak. Ezt követően a javított fájlokból előállított új CSV fájlok segítségével töltjük be a rekordokat az ITIdata rendszerébe, majd a rekordokat manuálisan egészítjük ki azokkal az adatokkal, amelyeket az XML fájlok nem tartalmaznak, mint például a levélhez kapcsolódó kritikai jegyzet oldalszáma a nyomtatott kiadásban.

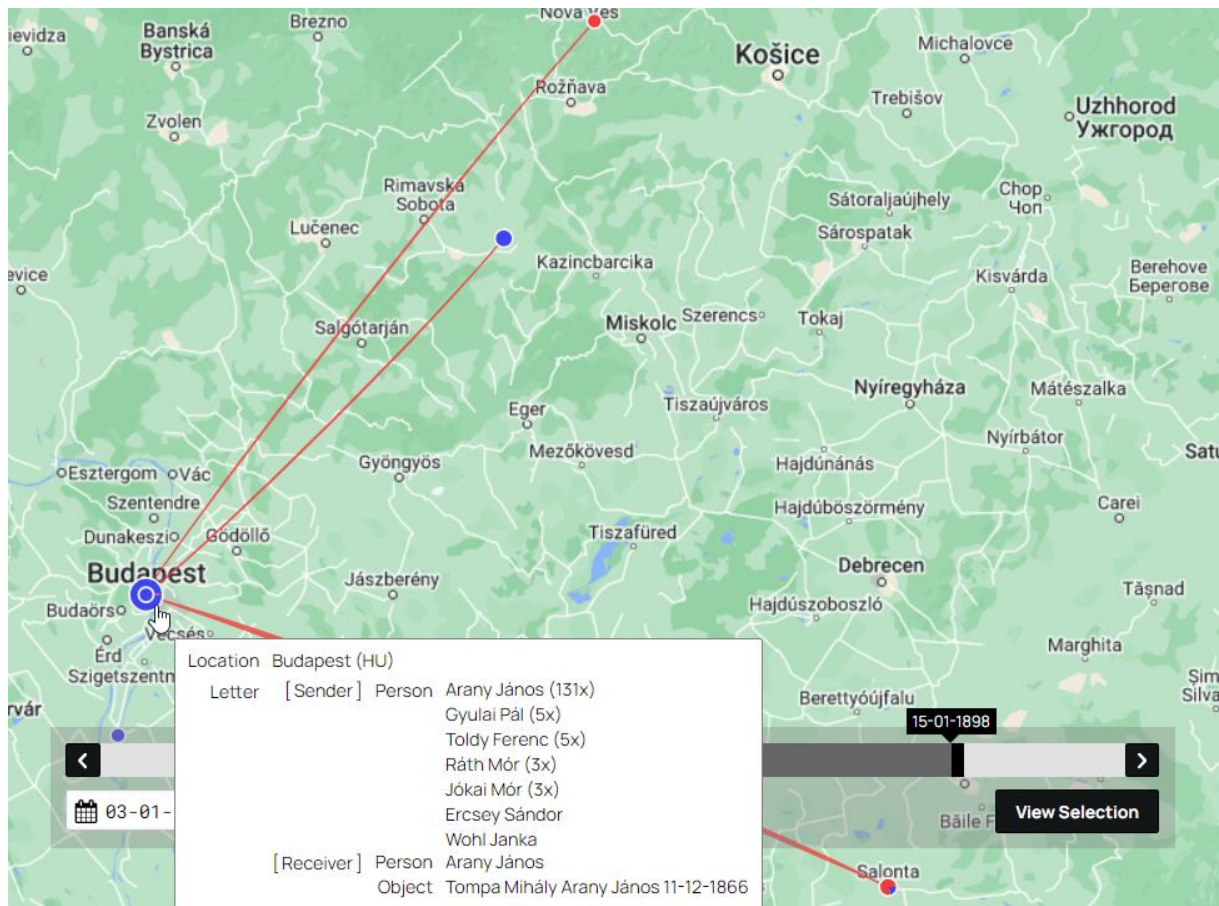
³⁷⁰ A Nodegoat felhasználói között neves egyetemek és kutatóközpontok is megtalálhatók, a teljes lista elérhető a szolgáltatást bemutató weboldalon: <https://nodegoat.net/about>.

adatokat a rendszer kezelni tudja, kialakítottuk a szükséges adatstruktúrát, meghatároztunk fő objektumokat (személy, levél, város), valamint al-objektumtípusokat (szerző, címzett). Ezt követően a metaadatokat a CSV formátumból importáltuk a Nodegoat rendszerbe. Az adatok előkészítése során egyértelművé vált, hogy az automatikusan kinyert adatsoroknak csak töredéke importálható további javítások nélkül. Mivel a kritikai kiadás, és így a jelölőnyelvi átiratok metaadatai is hiányosak voltak, csak azoknak a leveleknek a rekordjait használtuk, amelyekhez minden szükséges információ rendelkezésre állt. A metaadatok rendszerezése rávilágított olyan hibákra is, amelyek az informatizálás során keletkeztek. Az importálást követően manuálisan adatgazdagítottuk a személyneveket a VIAF és a Wikidata rendszeréből. Bár az általunk a Nodegoat rendszerében kialakított adatstruktúra rendkívül egyszerű, a betöltött adatok segítségével már így is olyan vizualizációk készíthetők, amelyek segíthetnek egy levelezéskorpusz vizsgálatában. Az alábbi ábrán látható, ahogyan a levelek metaadataiból kirajzolódik az egyes személyek levélváltásai alapján egy kapcsolati hálózat.



[32. ábra] A tesztkorpusz kapcsolati hálóját a Nodegoat rendszerében.

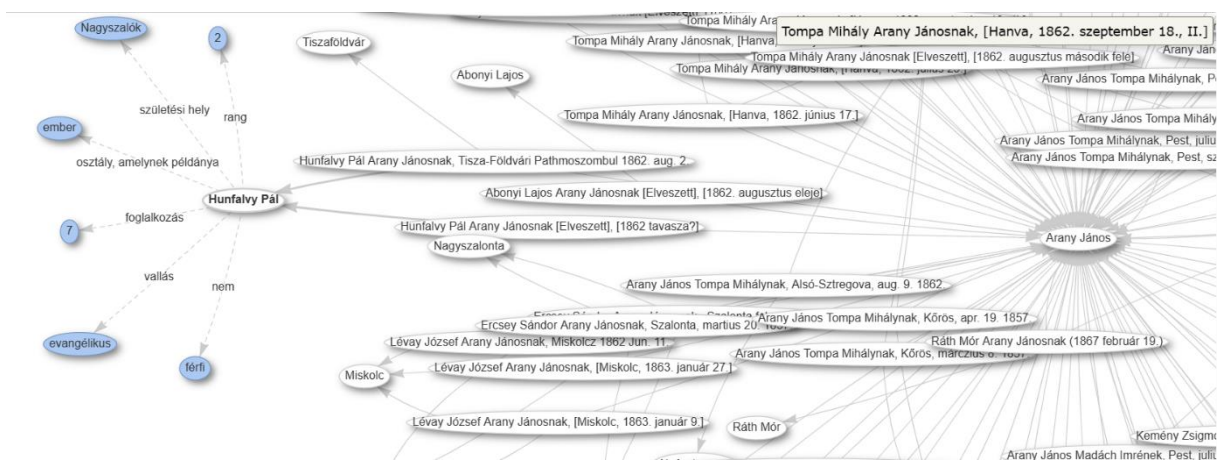
A rendszerben lehetőségünk nyílik a levelek térképes vizualizációjára is. A metaadatokban szereplő településneveket félautomatikusan azonosítottuk a Wikidata rendszeréből, kinyertük a geokordinátákat és az importálásra előkészített CSV fájlban rögzítettük azokat.



[33. ábra] A tesztkorpusz leveleinek térképes vizualizációja a Nodegoat rendszerében.

A Nodegoat rendszer tesztelése során nem állt rendelkezésre kellő humánerőforrás az összes levél metaadatának ellenőrzésére és javítására, ezért nem folytattuk a kísérletet, azonban a DigiPhil 2.0 egyik pillérének számító ITIdata mögötti szoftver, a Wikibase kiválasztásakor a beépített vizualizációs eszközök megléte is szempont volt. A levelek ITIdata reprezentációinak kialakításakor már a jövőbeni vizualizációs szempontokat is figyelembe vettük. Fontos

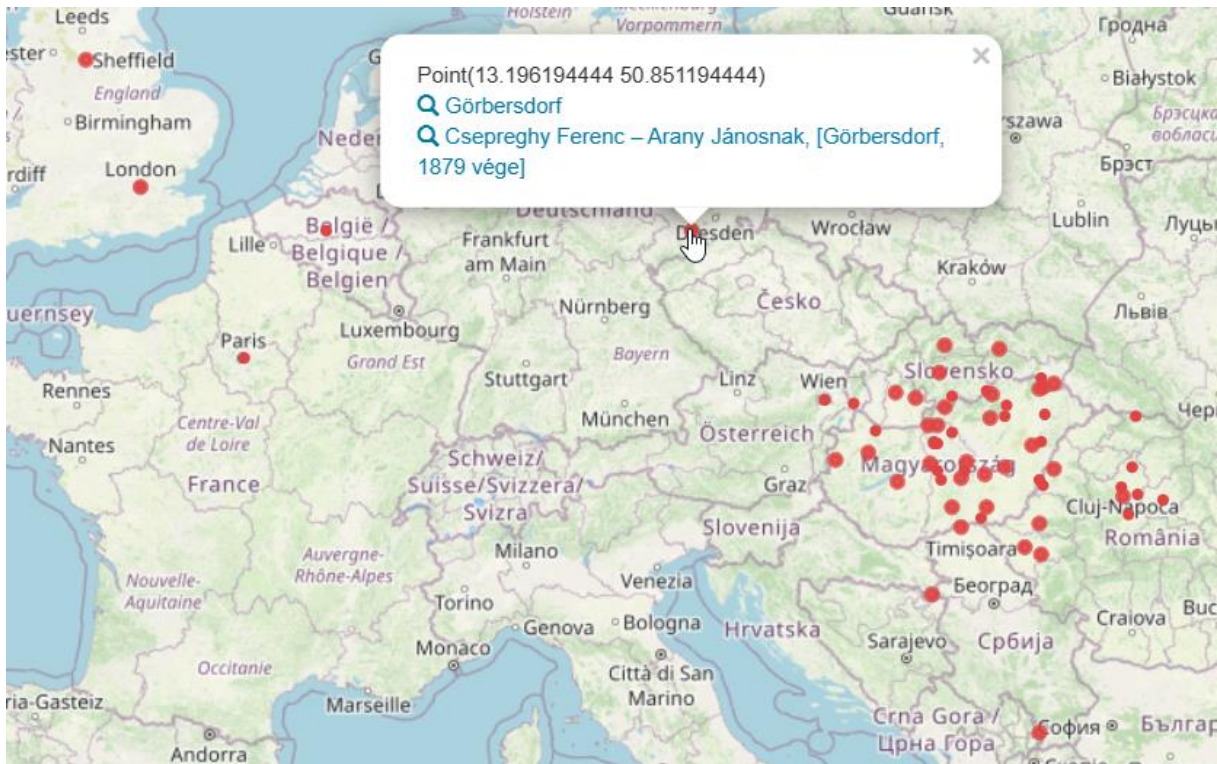
kiemelni, hogy sem az adathalmazok mérete, sem az adatsrúktúrák közötti különbségek nem teszik lehetővé a két rendszer összehasonlítását, itt csupán az eszközök beépített funkcióinak és alkalmazási lehetőségeinek bemutatása a cél. Az Arany-levelek migrációja során a kódolási és azonosítási hibákat javítjuk, az adatokat kiegészítjük, az adatok feltöltése jelenleg is zajlik. A migráció első lépéseként a három legújabb levelezés-kiadás adatainak, az *AJOM 17*, *18* és *19*-es kötetek a feldolgozását kezdtük meg.³⁷¹ A dolgozat írásának pillanatában a három kötetben szereplő leveleknek mintegy a fele, közel 900 levél került feldolgozásra. A feldolgozás mértéke még nem teszi lehetővé az adatok elemzését, de arra már most alkalmas, hogy a különböző vizualizációs eszközöket bemutassam. A Wikibase rendszerekben az adatvizualizáció mindig a rendszerben futtatott SPARQL lekérdezés eredményét jeleníti meg, így a megírt lekérdezések alapozzák meg a vizualizációt. Az alábbi ábrán a lekérdezés eredményéből rajzolódott ki a kapcsolati háló. Mivel a lekérdezések publikált adathalmazokon futnak le, szemben a Nodegoattal, nem csak a vizualizációk exportálhatók (kép formátumban), hanem a lekérdezés eredménye is.



[34. ábra] A tesztkorpusz kapcsolati hálójának részlete az ITIdata rendszerében.

³⁷¹ Ezek a kötetek a Korompay H. János szerkesztésében megjelent kötetek: KOROMPAY H. János, szerk., *Arany János, Levelezése 3.: 1857–1861*, Arany János Összes Művei 17 (Budapest: Universitas Kiadó, 2004); Új Imre Attila, szerk., *Arany János, Levelezése 4.: 1862–1865*, Arany János Összes Művei 18 (Budapest: Universitas Kiadó–MTA BTK Irodalomtudományi Intézet, 2014); KOROMPAY H., *Arany János, Levelezése 5.: 1866–1882*.

Akárcsak a Nodegoat, a Wikibase is rendelkezik beépített térképes vizualizációval. Az alábbi ábrán a levelek megírásának helyére szűkítettem a lekérdezést:



[35. ábra] A tesztlevelek keletkezési helyei az ITIdata térképes vizualizációján.

Mindkét rendszerben lehetőségünk nyílik az adatokat idővonalra helyezni, azonban a Wikibase rendszer további vizualizációs lehetőségeket is biztosít, a megjeleníteni kívánt adatok típusától függően lehetőségünk van különböző grafikonok készítésére, hierachiák megjelenítésére, buborékdiagrammok előállítására. A két rendszer teljeskörű összehasonlítására is lehetőségünk lesz, amikor a levelezés-korpusz feldolgozása lezárult, hiszen az ITIdata-ból a metaadatokat importálhatjuk a Nodegoat rendszerébe. Az azonban már a fenti összehasonlításból is egyértelmű, hogy a Wikibase alapú rendszerek legnagyobb előnye,

hogy a vizualizációk alapjául szolgáló kutatási adatok és a lekérdezések is megoszthatók, így transzparenssé reprezentálják a kutatás folyamatát és az adatvizualizációk alapján levont következtetéseket. Ugyanakkor, mivel mindig az adatbázis adott pillanatban érvényes állapotát tükrözik, el is változhatnak a publikációkor érvényes állapottól, ezért szükség lehet pillanatszerű archiválásra is.

Kosztolányi Dezső kritikai kiadásainak elemzése

A kutatás előzményének tekinthető a Beckett Archive kutatása, ahol a jelölőnyelvi átiratok vizsgálatával vizualizálták Samuel Beckett kéziratain található módosításokat.³⁷² A kutatás során az *Aranysárkány*, az *Édes Anna*, a *Pacsirta* és a *Nero, a véres költő* jelölőnyelvi átiratait vizsgáltam. Bár a DigiPhil-ben elérhető ezeken kívül az *Esti Kornél* és a *Szegény kisgyermek panaszzai* informatizált kritikai kiadásai, amelyek azonban nem szinoptikus átiratok, az *Esti Kornél*-ből mindössze egyetlen kézirat maradt fenn. A vizsgált korpusz két részre osztható: a) az *Aranysárkány* és az *Édes Anna* esetében a sajtó alá rendezők három-három szövegállapotot különítettek el, a kézirat legkorábbi, közbenső és végső állapotát; b) míg a *Pacsirta* és *Nero, a véres költő* esetében 1-1 szövegállapotot rögzítettek. A jelölőnyelvi átiratok ennek megfelelően eltérő felépítést mutatnak, ami befolyásolja a vizsgálatot is. A vizsgálat során egyaránt alkalmaztam szövegstatistikai módszereket, amelyekkel a törlések és javítások arányát figyeltem, valamint nyelvi elemzőt, amellyel a szófaji arányok változását próbáltam megragadni és ezekben mintázatokat felismerni. A nyelvi elemző számára a Python programnyelv BeautifulSoup könyvtárának alkalmazásával, egyszerű szkriptek segítségével nyertem ki a jelölőnyelvi átiratokból a szövegeket.

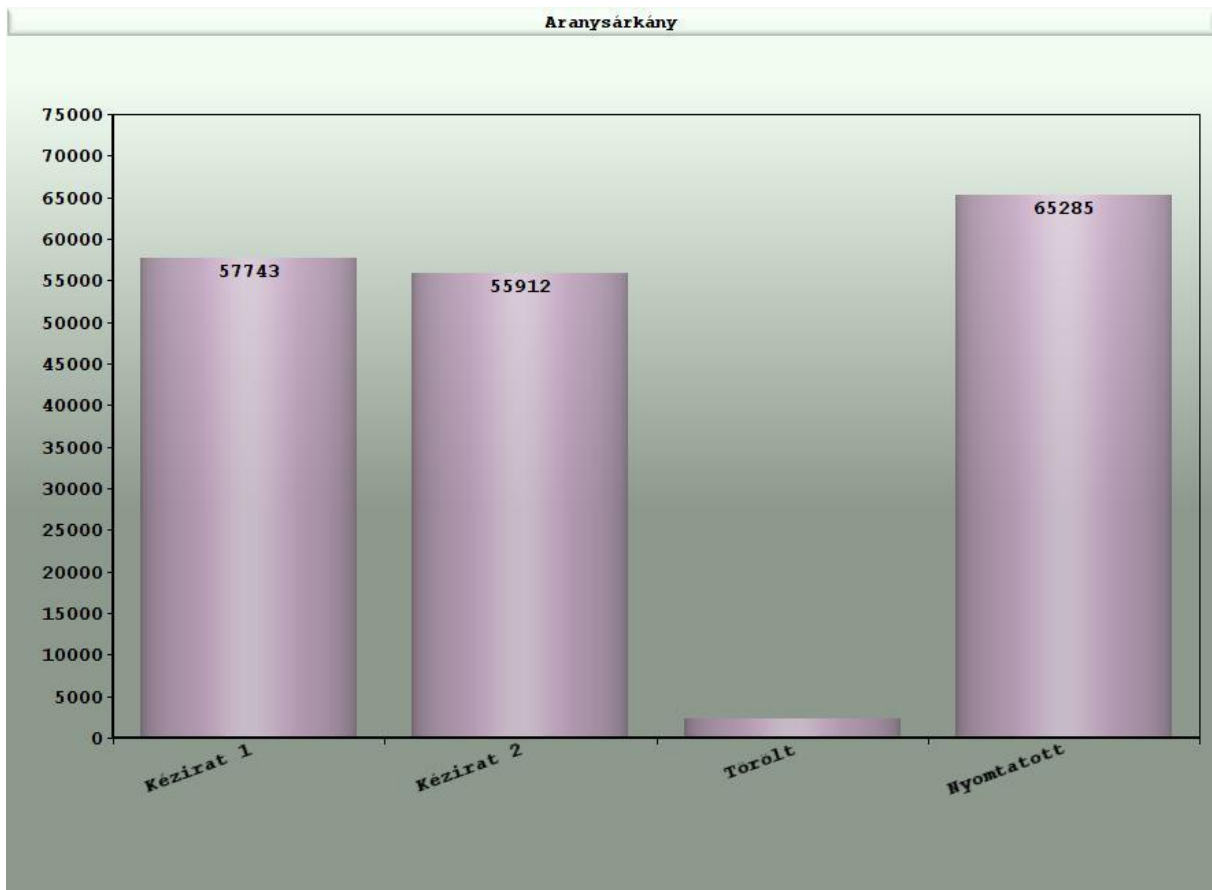
Az a) csoport esetében csak a törlési aktusok vizsgálatára volt lehetőség, ugyanis a kódolás jellegéből fakadóan a beszúrások, javítások az adott szöveghelyen nem voltak jelölve, megtörténtükre, amint az alábbi példán is látható, az adott szövegállapot értelmezésével lehet következtetni:

```
<app>
  <rdg change="#stage1" hand="#t" wit="#K">teszed. Ez a gerince,</rdg>
  <rdg change="#stage3" hand="#t" wit="#K">teszed, ide ni,</rdg>
</app> hogy megrögzítse a <del instant="true" status="unremarkable">vázat. Ez a</del> vázat s
feszesen <del instant="true" status="unremarkable">összekötözöd</del> összekötöd a szitakéreg e
két pontjával.
```

³⁷² Az eredményeket lásd: <https://www.beckettarchive.org/statistics>.

[36. ábra] Példa az *Aranysárkány* kéziratának jelölőnyelvi átiratából.

A fenti példán látható, hogy a legkorábbi szövegállapotban a „teszed” szólaakat új mondat követi „Ez a gerince,” kezdettel, míg a végső szövegállapotban a mondat „ide, ni”-vel folytatódik. A teljes átirat vizsgálata során tehát csak a **del** jelölővel felcímkézett szólaakat

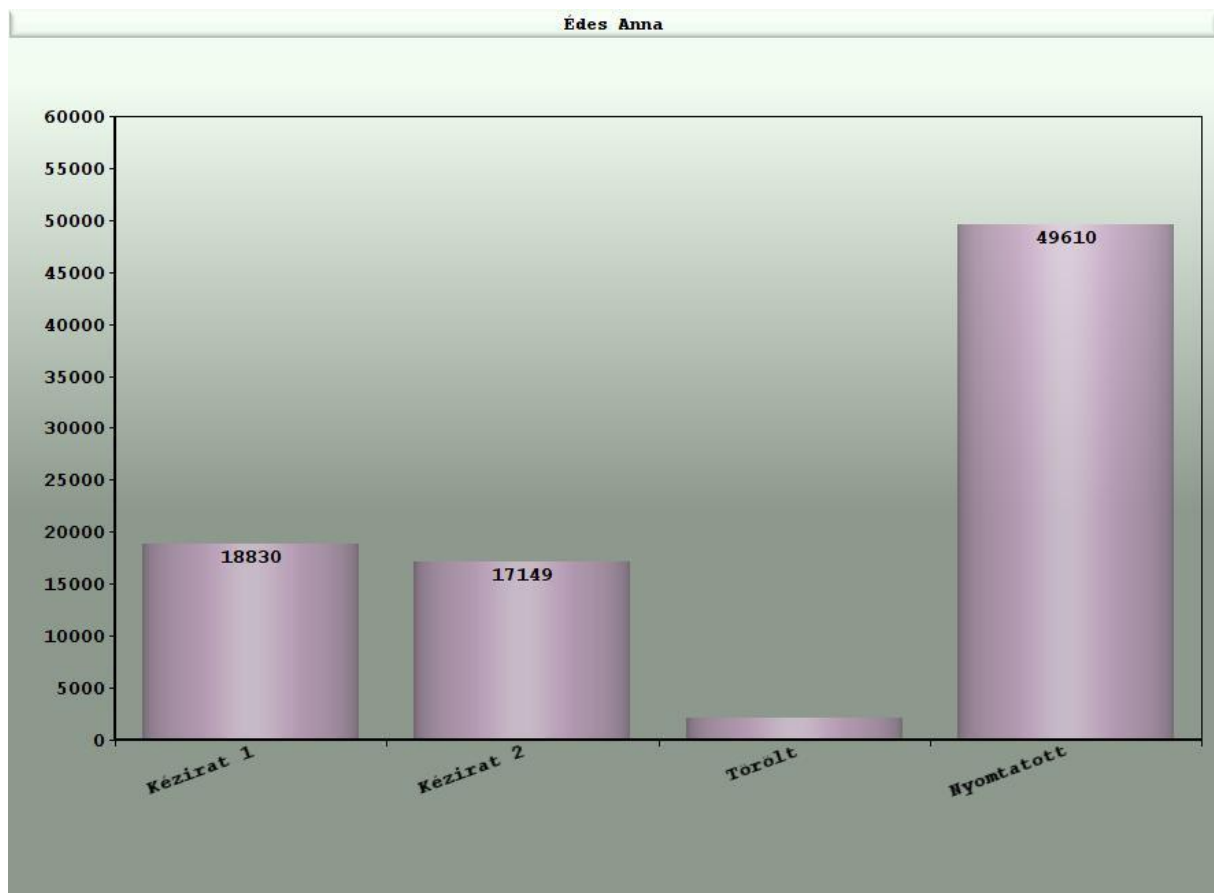


tudtam vizsgálni. Az *Aranysárkány* kéziratának esetében a kézirat legelső állapotához képest, a szólaakat számolva látható, hogy a legvégső szövegállapot 2232 szóalakkal rövidebb, a törlések közül 401 esetben a sajtó alá rendezők nem tudták kiolvasni a törölt szót. A regény első, nyomtatásban megjelent szövege azonban majd 10.000 szóalakkal hosszabb, mint az elérhető kéziratoké.

[37. ábra] Szövegstatistikai vizsgálatok eredménye az *Édes Anna* kéziratának jelölőnyelvi átiratán.

Az *Édes Anna* vizsgálata során érdekes adat, hogy bár a kézirat sokkal rövidebb (tehát kevesebb szóalakat tartalmaz), a törlések száma (2109) és ezekből a kiolvashatatlanok száma

(797) is sokkal magasabb. Az ábrán látható, hogy ebben az esetben a kézirat sokkal töredékesebb, az első kiadásnak mindössze a 38 százaléka maradt fenn kéziratban.



[38. ábra] Szövegstatistikai vizsgálatok eredménye az *Édes Anna* kéziratának jelölőnyelvi átiratán.

A számszerű elemzéseket követően az egyes szövegállapotokat az e-magyar nyelvi elemzőjével vizsgáltam a szövegállapotok közötti szófaji eltérések után kutatva.³⁷³

	Kézirat 1		Kézirat2		AS2	
Token	57743		55912		65285	
NOUN + PROPN	14577	25,24	14319	25,61	17158	26,28
VERB	11009	19,07	10678	19,10	12415	19,02
ADV	6756	11,70	6451	11,54	7481	11,46
DET	5854	10,14	5605	10,02	6671	10,22
PRON	5849	10,13	5650	10,11	6310	9,67
ADJ	5677	9,83	5483	9,81	6191	9,48

[39. ábra] Az *Aranysárkány* jelölőnyelvi átiratának nyelvi elemzésének eredménye.

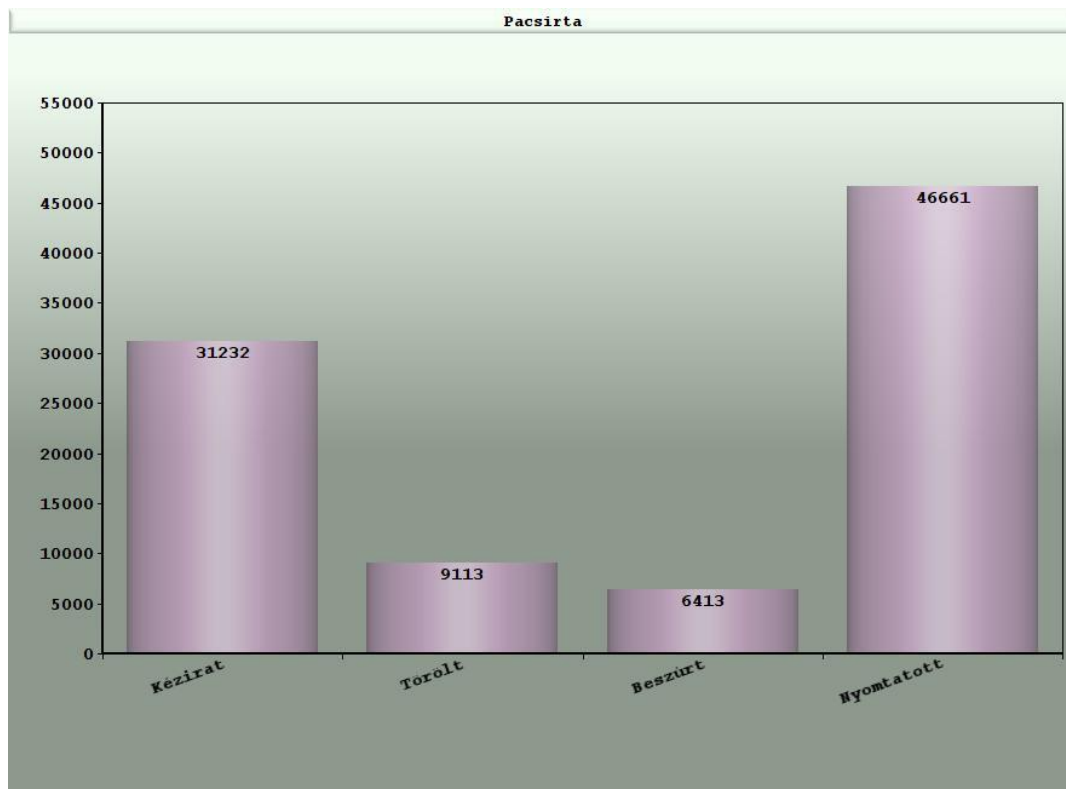
	Kézirat 1		Kézirat2		EA1	
Token	18830		17149		49610	
NOUN + PROPN	4689	24,90	4283	24,98	11955	24,10
VERB	3612	19,18	3238	18,88	9566	19,28
ADV	2241	11,90	2067	12,05	6018	12,13
DET	2023	10,74	1826	10,65	5630	11,35
PRON	1806	9,59	1681	9,80	5149	10,38
ADJ	1788	9,50	1618	9,43	4711	9,50

[40. ábra] Az *Édes Anna* jelölőnyelvi átiratának nyelvi elemzésének eredménye.

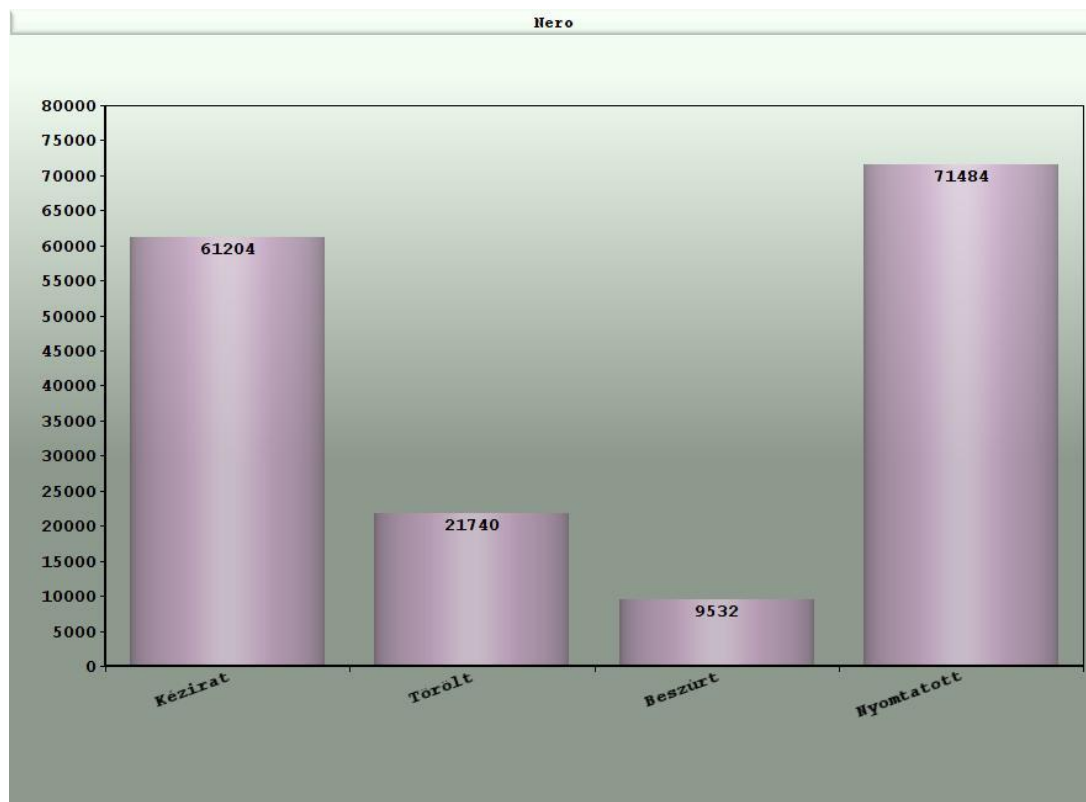
Az eredmények a vizsgált szövegállapotokon nem mutattak szignifikáns különbségeket. A b) csoport vizsgálata során nem csak a törlési aktusok elemzésére nyílt lehetőség, hanem a törölt szavak javításainak, illetve az utólag beszúrt szövegrészek vizsgálatára is. A b) csoportban található kéziratok szintén töredékesek, de teljesebbek, mint az a) csoportban vizsgáltak, a *Pacsirta* kézírata 66 százaléka az első kiadás szövegének, míg a *Nero* esetében az első kiadás 85 százaléka elérhető. Mindkét esetben elmondható, hogy a törölt szakaszok száma jóval magasabb, mint a javított, illetve utólag beszúrt részeké. A b) csoport esetében, a kódolás jellegéből fakadóan nem lehetett nyelvi elemzést végezni, ugyanis a jelölőnyelvi átiratban

³⁷³ Ezúton is köszönöm Indig Baláznak, az e-magyar, illetve annak továbbfejlesztésében az emTSV-ben fejlesztőként részt vevő kollégámnak a segítséget.

minden jellemzőt a lehető legkisebb szegmentumon jelöltünk, ami sokszor csak néhány betűt jelentett.



[41. ábra] Szövegstatistikai vizsgálatok eredménye a *Pacsirta* kéziratának jelölőnyelvi átiratán.



[42. ábra] Szövegstatistikai vizsgálatok eredménye a *Nero* kéziratának jelölőnyelvi átiratán.

A fentiekben bemutatott elemzések összességében nem szolgáltak pozitív eredménnyel, de a kutatás mégsem volt haszontalan, ugyanis rávilágított arra, hogy a jelölőnyelvi kódolás jellege és a mögötte álló filológiai és digitális filológiai gyakorlat meghatározza, hogy milyen elemzések hajthatók végre egy-egy korpuszon. Egy készülő *born digital* kritikai kiadás sajtó alá rendezőinek és a digitális filológus szakértőnek érdemes lehet ezeket a szempontokat is figyelembe venni, amikor a kiadás szövegközlési elveit és a jelölőnyelvi átirat specifikációját előkészítik.

A digitális filológia oktatása

A digitális filológia térnyerése az elmúlt évtizedekben vitathatatlan, azonban hiába az évtizedes tapasztalatok, a digitális filológia oktatása kiforratlan még, nem áll rendelkezésre átfogó képzés a leendő szakértők számára. Kifejezetten időszerű kérdés, tekintve, hogy a digitális, illetve digitalizált kiadásokkal foglalkozó kutatócsoportok száma egyre növekszik, de a magyarországi képzések nem gondoskodnak szakértői „utánpótlásáról”.³⁷⁴

Elena Pierazzo, a University of Grenoble, Italian Studies, Digital Humanities professzora oktatóként aktívan vesz részt digitális filológiai kurzusokon, olyan nagyszabású projektek vezető kutatója, mint a *Jane Austen's Fiction Manuscripts: A Digital Edition*. Egy 2016-ban tartott előadásában felvetette,³⁷⁵ majd az egyik legfrissebb tanulmányában részletesen körüljárta azt a problémát, hogy a kisebb filológiai műhelyeknek, PhD hallgatónak nincs esélyük komolyabb programozási tudás vagy a megfelelő szakember finanszírozására szánt forrás nélkül színvonalas kiadás publikálására. Tapasztalata szerint túl magas a belépési küszöb egy egyedi digitális kritikai kiadás elkészítéséhez. A nagyívű digitális filológiai projektek egyedi megoldásokat alkalmaznak, Pierazzo ezeket Elise Leclerc nyomán az „Haute Couture”-hoz hasonlítja, és emellett kiemeli, olyan „ready to wear” (viselésre kész) kiadásokra is szükség lenne, amelyek elkészítéséhez nincs szükség a fent sorolt digitális kompetenciákra.³⁷⁶ Érvéle szerint a rendelkezésre álló kész eszközök jelentősen korlátozzák a kutatók kreativitását, például az olyan megjelenítőeszközök, mint a Versioning Machine, a TEI elemkészletének csak kis százalékát képes megjeleníteni.

Nem Elena Pierazzo az első, aki felismeri, hogy mennyire összetett kompetenciák elsajátítása szükséges egy digitális kiadás elkészítéséhez, Shillingsburg 2006-ban így fogalmazott az ilyen jellegű kiadások készítői kapcsán:

³⁷⁴ Jelen fejezet korábbi változatát lásd: FELLEGI Zsófia, „A digitális filológia oktatása: Gyakorlati tapasztalatok”. in *Digitális eszközök a középiskolai irodalomtatásban*, szerk. MOLNÁR Gábor Tamás és FEJES Richárd, 157–178. (Budapest: Prae.hu, 2021.)

³⁷⁵ Elena PIERAZZO, „Of Digital Scholarly Editions and Building Tools” (Digital Scholarly Editing: Theory, Practice, Methods, Antwerpeni Egyetem, Antwerpen, 2016. október 5.).

³⁷⁶ „Ha ezt a metaforát a digitális tudományos kiadásokra alkalmazzuk, észre kell vennünk, hogy a kiadások, amelyeket ma készítünk, inkább az 'Haute Couture'-re emlékeztetnek, mint a 'viselésre kész' (ready to wear) termékekre: a digitális kiadások többnyire egyediek. Mindegyik egy sor dedikált eszköz felhasználásával készült, és mindegyik innovatív, kreatív, drága, és csak arra a szövegre szabott, amelynek kapcsán létrejött, és általában mások nem is használhatják fel. Az a kihívás itt, hogy elképzelnünk, hogyan nézne ki egy 'viselésre kész' kiadás, vagyis hogy modellezzük a jövő digitális kiadásait, és azok szerkesztőit, vagy még inkább azokat a kompetenciákat, amelyeket el kell majd sajátítaniuk.” (ford. PALKÓ Gábor, „Kritikai kiadások a digitális térben” (Móricz Zsigmond levelezésének digitális kritikai kiadása (1892–1913), Petőfi Irodalmi Múzeum, 2019. január 14.).)

Elektronikus kiadást létrehozni nem egyszemélyes művelet. Olyan képességeket követel, amelyek csak ritkán találhatók meg egy személyben, ha egyáltalán megtalálhatók. A szerkesztők elsősorban szövegkritikusok, de emellett bibliográfusok, képesek irodalmi és történelmi kutatásokat végezni. Ám általában nem könyvtárosok, betűszedők, nyomdászok, kiadók, könyvtervezők, programozók, webmesterek vagy rendszerelemzők. A nyomtatott kiadások idejében egyes szerkesztők vállalták ezen produkciós szerepek egy részét, a számítógép korában egyes szerkesztők megpróbálnak programozni és interfészeket tervezni.³⁷⁷

Ahogy Rehbein és Fritze is rámutatnak Shillingsburg idézett szövege kapcsán, a digitális filológia területén (és ez kétségtelenül igaz más digitális bölcsészeti területekre is) olyan, speciális képzésre van szükség, amely a szükséges kompetenciák elsajátításán alapul, a tantervnek ehhez mérten képesnek kell lennie a szükséges, széles skálán mozgó készségek összefogására és átadására.³⁷⁸ Jelen fejezetben azt kívánom bemutatni, hogy milyen lehetőségek álltak rendelkezésre az elmúlt években és manapság a digitális filológia oktatása terén, illetve milyen szempontokat érdemes figyelembe venni egy komplex képzés kialakítása során. Az általam javasolt szempontok egyaránt érvényesíthetők a graduális és a posztgraduális képzésben. Mielőtt azonban rátérnék a terület oktatási gyakorlataira, nem tekinthetek el a digitális filológia oktatás nemzetközi és magyarországi történetének rövid bemutatásától.

A digitális filológia oktatásának magyarországi történetének vizsgálatakor azt látjuk, hogy a humáninformatikai oktatás már a '90-es években elindult, az 1997-ben alapított BIÖP keretében például a hallgatók 46 kredités képzés során ismerkedtek meg az akkoriban korszerű digitális filológiai gyakorlatokkal.³⁷⁹ Az elmúlt két évtized folyamán a BIÖP programján túl, a magyarországi egyetemeken elszórta zajlottak szemináriumok a graduális képzésben részt vevők számára.³⁸⁰ Ezt követően a 2010-ben akkreditált Digitális Bölcsészet Mesterképzés

³⁷⁷ Peter SHILLINGSBURG, *From Gutenberg to Google: Electronic Representations of Literary Texts* (Cambridge: Cambridge University Press, 2006), 94.

³⁷⁸ Christiane FRITZE és Malte REHBEIN, „Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing”, in *Digital Humanities Pedagogy: Practices, Principles and Politics*, szerk. Brett D. HIRSCH, 47 (Cambridge: Open Book Publishers, 2012), 47.

³⁷⁹ A képzés programja az alábbi oldalon tekinthető meg: <http://magyar-irodalom.elte.hu/regi.html>.

³⁸⁰ Erre kiváló példa Labádi Gergely egy szemináriuma, amelynek keretében a hallgatók Mikszáth-elbeszélések XIII. (1888) kritikai kiadásának jelölőnyelvi átiratát készítették el. A kiadás később a DigiPhil oldalán jelent meg: <http://digiphil.hu/context:mkk39>.

nyújtott rövid ideig, egyetemi hallgatók számára, átfogó digitális filológiai képzést a Digitális kultúra szakirány keretein belül.

Az elmúlt évtizedben digitális bölcsészeti alap- és mesterszakok jöttek létre, tanszékek, kutatócsoportok és központok foglalkoznak a digitális bölcsészet oktatásával. Sula, Hackney és Cunningham 2017-es tanulmányukban³⁸¹ összegyűjtötték az angolszász területeken működő digitális bölcsészeti képzéseket és programokat, majd ezeket az adatokat összehasonlították az európai országok adataival. Az eredményekből látható, hogy 2008-tól radikálisan emelkedett a digitális bölcsészeti képzések és projektek száma. Míg az európai képzések nagy részén mesterszakos diplomát szerezhetnek a hallgatók (pl. Universität Trier), illetve több egyetemen lehetőség van BA (pl. Universiteit Utrecht) és PhD (pl. University of Zagreb) képzés elvégzésére is, az angolász képzések általában nem adnak diplomát, a képzéseknek nagyjából a kétharmadát nem az egyetemek szolgáltatják. Az egyetemek képzései is nagyon eltérő képet mutatnak: van olyan egyetem, ahol a digitális bölcsészet minor szakként (pl. Australian National University), specializációként (pl. Michigan State University) és szakirányként (pl. University of Georgia) végezhető el; több egyetem szakirányú továbbképzést nyújt (pl. Stanford University, University of Iowa) és csak néhány egyetemen van lehetőség BA, MA, illetve PhD képzés elvégzésére (pl. King's College London, National University of Maryland). A Magyar Akkreditációs Bizottság 2010-ben engedélyezte a digitális bölcsészet mesterképzés alapítását két szakiránnyal: számítógépes nyelvészet és digitális kultúra. A következő évben, 2011-ben két egyetemen indult digitális bölcsészet mesterképzés: a Debreceni Egyetemen és a Pázmány Péter Katolikus Egyetemen. A mesterképzés a 2016/17-es tanévben indult utoljára, miután a képzést a 2015-ös kormányrendelettel megszüntették.³⁸² A PPKE-n a mesterképzés keretein belül 2014-től a számítógépes nyelvészet specializáció mellett a digitális kultúra specializáció is elérhető volt.³⁸³

³⁸¹ Chris A. SULA, Sarah E. HACKNEY és Phillipp CUNNINGHAM, „A Survey of Digital Humanities Programs”, *The Journal of Interactive Technology and Pedagogy*, 11. sz. (2017), <https://jitp.commons.gc.cuny.edu/a-survey-of-digital-humanities-programs/>.

³⁸² 139/2015. (VI. 9.) Korm. rendelet a felsőoktatásban szereshető képesítések jegyzékéről és új képesítések jegyzékbe történő felvételéről, <http://www.kozlonyok.hu/nkonline/mkpdf/hiteles/mk15079.pdf>. A Digitális Bölcsészet Mesterképzést az ELTE BTK és a PPKE BTK ismételt akkreditáltatta, a 2020. 08. 12-én megjelent kormányrendelet alapján a kurzus visszakerült a felsőoktatásban szereshető képesítések jegyzékébe, így a 2021/2022-es tanévtől indítható újra. Ehhez lásd: 393/2020. (VIII. 12.) Korm. rendelet a felsőoktatásban szereshető képesítések jegyzékéről és új képesítések jegyzékbe történő felvételéről szóló 139/2015. (VI. 9.) Korm. rendelet módosításáról, <http://www.kozlonyok.hu/nkonline/MKPDF/hiteles/MK20188.pdf>.

³⁸³ *Digitális Bölcsészet mesterképzési szak mintatanterve 2014/2015. tanévtől*, Pázmány Péter Katolikus Egyetem:

<https://btk.ppke.hu/uploads/articles/172326/file/digit%C3%A1lis%20b%C3%B6lcs%C3%A9szet%20MA%20mintatanterv%281%29.pdf>. A PDF fájl a dolgozat lezárásakor nem elérhető, az Internet Archive nem készített mentést.

A digitális kultúra meghatározása korántsem olyan könnyű feladat, mint a számítógépes nyelvészeté. A digitális kultúra egyaránt foglalkozik például adatvizualizációval, digitális művészettel, digitális filológiával, a kommunikáció és médiatudomány digitális vonatkozásaival (pl. közösségi média), szövegbányászattal, valamint a digitális kulturális örökség kezelésével. Ez a sokszínűség tapasztalható a magyarországi és külföldi képzésekben is.³⁸⁴ A PPKE digitális kultúra szakirányán a kötelező informatikai alapozó képzésen túl³⁸⁵ a hallgatók különböző kurzusokon sajátíthatták el a digitális technológiák alkalmazásait az egyes bölcsészettudományi területeken. Az olyan szemináriumok és előadások, mint például az Információ- és tudásszervező rendszerek, Könyvtári digitális adatbázisok, Funkcionális metaadat-modellek könyvtárinformatikai alapismereteket nyújtottak, a PPKE és a Petőfi Irodalmi Múzeum együttműködésében meghirdetett kurzusokon a hallgatók bepillantást nyertek a múzeuminformatika világába és a digitális kulturális örökség kezelésének módszereibe. A gyakorlati ismereteken túl a PPKE az elméleti képzésre is nagy hangsúlyt fektetett, kurzuskínálatban a bölcsészettudományok különböző területeiről szerepeltek olyan előadások, amelyek kapcsolódnak a digitális kultúrához, mint például *A multimédia elméletei*, *Kritikai kultúrakutatás*, *Digitális adatbázisok és szövegtörzsek: történet és materialitás*, *Kulturális emlékezet* és *Vizuális kultúra*. A mesterképzés elvégzése után a hallgatóknak, akik nem a munkaerőpiaci lehetőségek közül választanak, lehetőségük volt doktori tanulmányok elkezdésére és így a kutatói pálya választására.

A PPKE mellett a Debreceni Egyetem Általános és Alkalmazott Nyelvészeti Tanszékén is elérhető volt a digitális bölcsészet mesterképzés számítógépes nyelvészet és digitális kultúra szakiránnyal, ám a képzés csak 2011-ben és 2016-ban indult el. Általánosságban elmondható, hogy a Debreceni Egyetem tanterve³⁸⁶ kisebb hangsúlyt fektetett a programozási ismeretek átadására, ehelyett elméletibb informatikai képzést nyújtott. A digitális kultúra szakirány

³⁸⁴ Néhány kiragadott példa a nemzetközi képzésekből: King's College London, Digital Culture BA: <https://www.kcl.ac.uk/study/undergraduate/courses/digital-culture-ba.aspx>; Arizona State University, Digital Culture BA: <https://artsmediaengineering.asu.edu/content/digital-culture>; The University of Sydney, Digital Cultures MA: <https://sydney.edu.au/courses/subject-areas/major/digital-cultures0.html>; Utrecht University, New Media & Digital Culture MA: <https://www.uu.nl/masters/en/new-media-digital-culture>; University of Bergen, Digital Culture MA: <https://www.uib.no/en/studies/MAHF-DIKUL>.

³⁸⁵ A képzésen résztvevők magas óraszámban vettek részt programozási, matematikai és logikai szemináriumokon és előadásokon, amelyeket a Pázmány Információs és Technológiai Kara tartott. A digitális bölcsészet, különösen a számítógépes nyelvészet oktatásáról részletesebben lásd: FELLEGI Zsófia és VADÁSZ Noémi, in „Digitálisbölcsészet-oktatás Magyarországon”, szerk. L. VARGA Péter, MOLNÁR Gábor Tamás és PALKÓ Gábor, *Digitális (szöveg)kultúrák a bölcsészsképzésben*, 168–180, (Budapest: ELTE-BTK, 2018).

³⁸⁶ A digitális bölcsészet mesterszak leírása, Debreceni Egyetem: http://lingua.arts.unideb.hu/doc/digitalis_bolcseszeti_kepzesi_program.pdf. A link a dolgozat lezárásakor nem elérhető, az Internet Archive egyetlen mentése szerverhibát jelez.

esetében a digitális bölcsészeti ismeretek alkalmazását segítették elő az egyes bölcsészettudományi területeken.

A terület iránt érdeklődő PhD hallgatók és kutatók (a szak megszűnése óta az egyetemi hallgatók is) magyar, illetve nemzetközi workshopokon, nyári egyetemeken, előadássorozatok keretében,³⁸⁷ online kurzusokon sajátíthatják el a szükséges kompetenciákat. Ez alól kivétel az ELTE BTK, ahol a 2019/2020-as tanévtől kezdően az alapszakos hallgatók számára kötelező elvégezni a Digitális bölcsész kompetenciák című kurzust, a PhD hallgatóknak kötelező egy Digitális bölcsészet workshop online kurzus elvégzése,³⁸⁸ az oktatók számára Digitális Bölcsészet Tanszék tart digitális bölcsészeti témájú workshopokat.³⁸⁹

A digitális filológia oktatásának nemzetközi jó gyakorlatai

Európai viszonylatban kétféle gyakorlat vált általánossá. Egyrészt, ahogyan a digitális bölcsészet oktatás is egyre elterjedtebbé vált, a digitális filológia oktatása beépült az egyetemi képzésbe,³⁹⁰ másrészt Európa-szerte nyári egyetemeket, többnapos intenzív kurzusokat,³⁹¹ konferenciaworkshopokat tartanak.³⁹² A legrégebbi és talán leghíresebb kurzust a University of Oxford szervezi, az első TEI Summer Schoolt 1997 júniusában tartották Lou Burnard, a TEI

³⁸⁷ Maróthy Szilvia és Simon Zsolt például négyalkalmas előadást tartottak a Pesti Bölcsész Akadémián. Lásd: https://pestibolcseszakademia.blog.hu/2016/10/04/digitalis_bolcseszeti_szovegen_innen_es_tul_marothy_szilvia_es_simon_zsolt_kurzusa#more11763277.

³⁸⁸ A *Digitális bölcsészet workshop* választható kurzusokból épül fel. A digitális filológiai kurzus alapját a DARIAH Teach által készített TEI oktatási anyag adja, amelyhez magyar fordítás is elérhető. BOBÁK Barbara, FELLEGI Zsófia és PALKÓ Gábor, ford., „Szövegkódolás és a Text Encoding Initiative”, 2018, <https://teach.dariah.eu/course/view.php?id=40>.

³⁸⁹ Erről részletesebben lásd Palkó Gábor és Smrz Ádám tanulmányát: PALKÓ Gábor és SMRCZ Ádám, „A digitális bölcsészet oktatása az ELTE BTK-n”, in *Digitális eszközök a középiskolai irodalomoktatásban*, szerk. MOLNÁR Gábor Tamás és FEJES Richárd, 135–156 (Budapest: Prae.hu, 2021).

³⁹⁰ A számos esetből csupán két példát kiragadva: a Nosov Magnitogorsk State Technical University-n például önálló Digitális Filológia Mesterképzés zajlik (<https://www.masterstudies.com/Master-of-Digital-Philology/Russia/Nosov-Magnitogorsk-State-Technical-University/>), az Universität zu Köln Institut für Digital Humanities alap- és mesterszakon is kínál digitális filológiai képzéseket. Lásd: <https://dh.phil-fak.uni-koeln.de/studium-lehre/studieren-am-idh/informationsverarbeitung>.

³⁹¹ 2017-ben zárult a Marie Curie Actions által támogatott nagyszabású projekt, a DiXiT (Digital Scholarly Editions Initial Training Network), amely a széles körű nemzetközi hálózata révén összehangolt képzési és kutatási programot kínált fiatal és tapasztalt kutatók számára. Négy év alatt 25 oktatási eseményt rendeztek 11 európai országban. Lásd: <https://cordis.europa.eu/project/id/317436/reporting/de>.

³⁹² A DigiPhil és a Zentrum für Informationsmodellierung Universität Graz partnerségében került megrendezésre a 2016-os Digital Humanities konferencián egy önálló workshop, amelynek keretében a grazi partnerek által fejlesztett Cirilo szoftvert mutattuk be. A Cirilo alkalmazás egy multifunkciós *asset management rendszer*, amelyet kifejezetten TEI XML alapú kiadások kanalizálására és publikálására fejlesztettek. A résztvevők a workshopon megismerkedtek a rendszer kínálta lehetőségekkel, a TEI XML fájlok kezelésével, és bemutattuk a partneri együttműködésünket. Martina BÜRGERMEISTER és mtsai., „GAMS and Cirilo: research data preservation and presentation” (DH 2016, Krakkó, 2016). Az absztrakt elérhető itt: <https://dh2016.adho.org/abstracts/116>.

P1 Guidelines egyik szerzőjének szervezésében és vezetésével.³⁹³ Ettől kezdve minden évben megrendezésre kerül a TEI Summer School, a szervezők folyamatosan bővítették a képzési palettát (például XSLT oktatással), majd 2012-ben a TEI Summer Schoolt átnevezték Digital.Humanities@Oxford Summer Schoolra, a képzés kiterjedt egyéb digitális bölcsészeti témákra (például szemantikus web technológiák), párhuzamos workshop szekciókat és előadásokat tartanak.

A képzés komponensei

A következőkben, támaszkodva az elmúlt évek saját oktatási tapasztalataira, megkísérlem összefoglalni, milyen komponensekből állhat egy komplex, átfogó filológiai képzés, amely összhangban van a nemzetközi oktatási praxissal. Rehbein és Fritze az általuk tartott, eltérő időtartamú és mélységű kurzusok terveiben hasonló célokat fogalmazznak meg. A következőkben bemutatott tantervi javaslat előfeltevése, hogy a résztvevő rendelkezik hagyományos filológiai ismeretekkel, illetve kutatói tapasztalattal.

a. Digitális filológiai bevezető

A kurzus első része egy általános digitális filológiai bevezető, melynek során érdemes bemutatni a digitális filológia és a jelölőnyelvek történetét, kiemelve a nemzetközi filológiai műhelyeket, jó gyakorlatokat. Ennek részeként kerülnek bemutatásra azok, a korábban már bemutatott nemzetközi elvárások a digitális kulturális örökségek kezelésével kapcsolatban, mint platformfüggetlenség, az open source eszközök használata, metaadat-szabványok követése és a szabványosság, a hosszútávú megőrzés és a nyílt hozzáférés biztosítása a kulturális örökséghez. A szemantikus web szemléletmódjának megfelelően kitérünk az adatgazdagítás gyakorlatára, a személy-, intézmény- és földrajzi nevek névterekkel való összekötésére, valamint a bibliográfiai adatok könyvtári rendszerben történő leírására, az ott létrejött azonosítók kiadással való összekapcsolására. Amennyiben az időkeret lehetőséget biztosít erre, külön figyelmet szentelhetünk az adatbázis és XML alapú digitális filológiai projektek összehasonlítására, valamint a nemzetközi intézmények által használt repozitóriumszoftverek bemutatására is. A digitális tudományos szövegkiadások publikálása összetett folyamat, amely számos tényezőtől függ, többek között például attól, hogy önálló

³⁹³ DHOxSS Archive: <https://digital.humanities.ox.ac.uk/dhoxss-archive>.

szolgáltatást építünk-e, vagy csatlakozunk egy olyan szolgáltatáshoz, amely saját szabályzattal rendelkezik, mint például a TextGrid. A digitális objektum kiválasztásának főbb szempontjai egy regény vagy levelezéskiadás esetében jelentősen eltérhetnek, így ez a kérdés szintén befolyásolhatja a kiadás elkészítését. Feltétlenül szükséges ezeknek a gyakorlatoknak a bemutatása, hogy a kurzuson részt vevők megértsék a döntéshozatal komplexitását, és bepillantást nyerjenek a különböző publikálási módszerek előnyeibe és hátrányaiba is. A digitális filológiai kurzus jelentős részét az XML nyelv és a TEI ajánlás elsajátítása teszi ki. Amennyiben minden feltétel adott (erről részletesen később), érdemes bemutatni a digitális filológia eszközeit, a különböző online és offline szerkesztőprogramokat, valamint a filológus munkáját segítő egyéb szoftvereket, mint például a karakterfelismerő (OCR és HTR) eszközöket. Ezt követően a választott keretrendszer segítségével kezdődik el a hallgatók bevezetése, Elena Pierazzonál kevésbé elegáns megfogalmazással élve, a kacsacsőrök világába.³⁹⁴

b. Az XML és a TEI

Az első lépés az XML nyelv alapjainak elsajátítása és tudatosítása egyszerű példákon keresztül. Gyakori feladat egy versszak, bekezdés, illetve egy adott személy adatainak leírása egyszerű, magyar nyelvű jelölőkkel. Ezt követik a bonyolultabb kódok, amelyekben a jelölők már úgynevezett attribútumokat is tartalmaznak jelentésük további specifikálásaként. Ha a kurzus hallgatói már készségi szinten használják az XML nyelv jelölőit, következik a TEI ajánlás alaposabb ismertetése. A TEI Guidelines, ahogyan már korábban említettem, majdnem kétezer oldal teszt ki, az ajánlásban való eligazodás nem egyszerű. Először – a tájékozódást elősegítendő – érdemes bemutatni az egyes modulokat, a jelölők közötti keresési lehetőségeket, valamint a jelölők egyértelműsítő lapjainak felépítését, ugyanis ott definiálják a szerzők a jelölők pontos használatát, az XML hierarchiában való elhelyezkedését.

³⁹⁴ Pierazzo így fogalmaz: „world of angle brackets”.


```

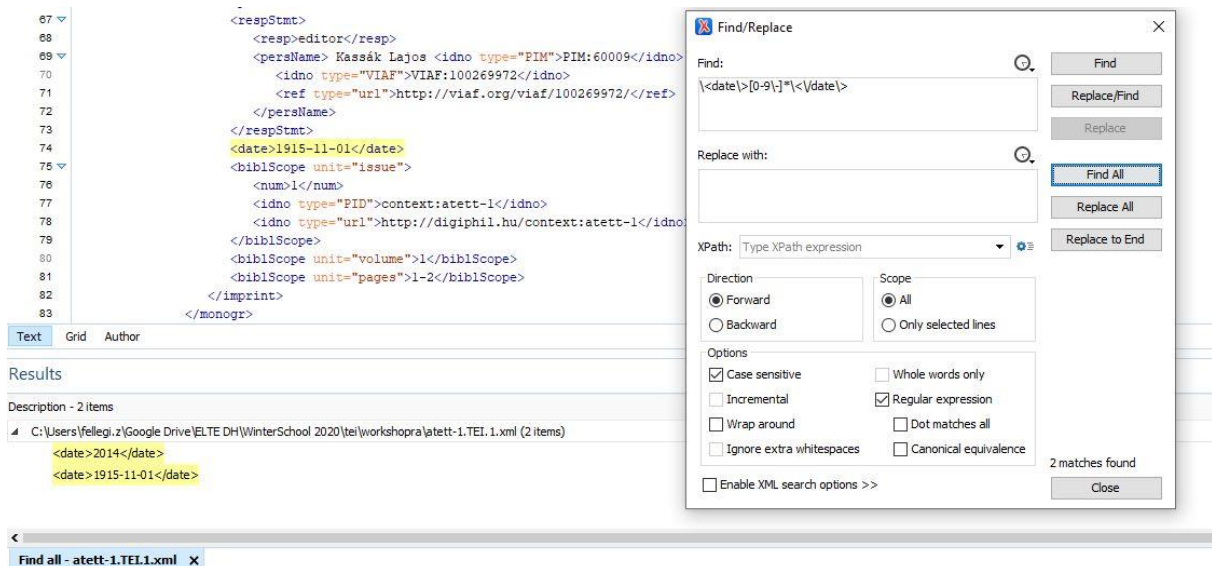
<biblStruct>
  <analytic>
    <title level="a">Keresztelőre</title>
    <author>
      <persName>
        <surname>Szabó</surname>
        <forename>Dezső</forename>
        <idno type="PIM">PIM:70684</idno>
        <idno type="VIAF">VIAF:37040595</idno>
        <ref type="url">http://viaf.org/viaf/37040595</ref>
      </persName>
    </author>
  </analytic>
  <monogr>
    <title level="j">A Tett</title>
    <title level="j" type="sub">kéthetenként megjelenő lap</title>
    <idno type="ISSN">0200-514X</idno>
    <idno type="PID">context:atett</idno>
    <idno type="url">http://digiphil.hu/context:atett</idno>
    <imprint>
      <pubPlace/>
      <publisher/>
      <respStmt>
        <resp>editor</resp>
        <persName>
          <surname>Kassák</surname>
          <forename>Lajos</forename>
          <idno type="PIM">PIM:60009</idno>
          <idno type="VIAF">VIAF:100269972</idno>
          <ref type="url">http://viaf.org/viaf/100269972</ref>
        </persName>
      </respStmt>
      <date>1915-11-01</date>
      <biblScope unit="issue">
        <num>1</num>
        <idno type="PID">context:atett-1</idno>
        <idno type="url">http://digiphil.hu/context:atett-1</idno>
      </biblScope>
      <biblScope unit="volume">1</biblScope>
      <biblScope unit="pages">1-2</biblScope>
    </imprint>
  </monogr>

```

[44. ábra] Részlet a Kassák Lajos által szerkesztett *A Tett* folyóirat jelölőnyelvi átiratának teiHeader részéből, amely a cikk bibliográfiai adatait írja le.

c. Műveletek XML fájlokon

A kurzus következő lépéseként, az elsajátított ismereteket felhasználva kezdődhet meg az XML kezelési technikák bevezetése, mint például a reguláris kifejezések³⁹⁶ és az XPath lekérdezőnyelv³⁹⁷ használata.



[45. ábra] Keresés dátumokra reguláris kifejezés segítségével.

Ez utóbbi módszer nagyon hasznos lehet akár a szerkesztés során automatikus javítások elvégzésére, bevitt adatok ellenőrzésére, akár egy XPath alapú keresőfelület fejlesztésekor.

³⁹⁶ A reguláris kifejezések segítségével szövegmintázatok kereshetők az XML fájlokban, valamint a találatokat ugyanezen keresőnyelv használatával automatikusan le lehet cserélni.

³⁹⁷ Az XPath (XML Path Language) egy olyan lekérdezőnyelv, amellyel akár nagy mennyiségű XML fájlban futtathatók lekérdezések.

The screenshot displays the Oxygen XML Editor interface. The top pane shows an XML document with the following structure:

```

47 <sourceDesc>
48 <bibliStruct>
49 <analytic>
50 <title level="a">Rezepteleőr</title>
51 <author>
52 <persName> Szabó Dezsa</persName> <idno type="PID">P2M70684</idno>
53 <idno type="VIAF">VIAF:37040595</idno>
54 <ref type="url">http://viaf.org/viaf/37040595</ref>
55 </persName>
56 </author>
57 </analytic>
58 <monogr>
59 <title level="j">A Tett</title>
60 <title level="j"> type="sub">kéthetenként megjelenő lap</title>
61 <idno type="ISSN">0200-5140</idno>
62 <idno type="PID">conncs:rects/ridno</idno>
63 <idno type="url">http://digipii.hu/contentatett</idno>
64 < imprint>
65 <pubPlace/>
66 </publisher/>

```

The bottom pane shows the results of an XPath query: `//persName[idno]`. The results table is as follows:

Description - 8 items	XPath location	Resource	Location	System ID
Szabó Dezsa	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	11:26	C:\Users\Felleg-z\Google Drive\B
Sálemán Teodora	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	20:26	C:\Users\Felleg-z\Google Drive\B
Szabó Dezsa	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	52:32	C:\Users\Felleg-z\Google Drive\B
Kassák Lajos	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	69:35	C:\Users\Felleg-z\Google Drive\B
Hendenburg	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	119:29	C:\Users\Felleg-z\Google Drive\B
Hackensen	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	120:42	C:\Users\Felleg-z\Google Drive\B
Juffe	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	121:42	C:\Users\Felleg-z\Google Drive\B
Szabó Dezsa	/TEI[1]/XslHeader[1]/XslDesc[1]/XslText[1]/XslAuthor[1]/XslPersName[1]/XslName[1]	atett-1.TEI.1.xml	168:29	C:\Users\Felleg-z\Google Drive\B

[46. ábra] Személynevek kinyerése A *Tett* folyóirat XML átirataiból XPath segítségével.

Az XML manipulációs műveletek elsajátítása és végrehajtása a kurzus végső, gyakorlati és egyben legnehezebb állomása. Első lépésként érdemes az XSLT stíluslapok készítésének menetét bemutatni. Az XSLT alkalmazásával lehetőség nyílik az XML formátumok közötti konverzióra, illetve dinamikus weboldalak készítésére, azonban ilyen stíluslapok készítéséhez már komolyabb informatikai kompetencia szükséges. A digitális filológia kurzuson érdemes olyan automatikus, egyszerűen elvégezhető folyamatokkal megismertetni a hallgatókat, amelyeket önállóan is el tudnak később végezni, ilyen például az XML fájlban található verssorok vagy bekezdések automatikus számozása. Tudásukat később, online tananyagok segítségével bővíthetik. Azonban az XSLT egyik legnagyobb hátránya, hogy futtatásához értelmező- és fordítóprogram szükséges. Az Oxygen XML Editor bár rendelkezik ilyen funkcióval, a tapasztalataim azt mutatják, hogy nagy mennyiségű fájl feldolgozása esetében a folyamat rendkívül lassú, nagy forrásigényű és olykor nem megbízható. Ennek alternatívája lehet az elmúlt években egyre elterjedtebb XML manipulációs eljárás a Python programnyelv BeautifulSoup könyvtárának alkalmazása.³⁹⁸ Amennyiben lehetőség nyílik rá, érdemes bevezetni a hallgatókat a könyvtár használatának lehetőségeibe is. Minden olyan művelet, amelyet XSLT-vel végre lehet hajtani, végrehajtható Pythonnal is, de az XSLT esetében tapasztalt hátrányok nélkül.

³⁹⁸ A könyvtár dokumentációja az alábbi linken érhető el: <https://pypi.org/project/beautifulsoup4/>.

d. Az elsajátított ismeretek alkalmazása

Ideális esetben lehetősége nyílik a résztvevőknek saját kiadás elkészítésére, illetve bekapcsolódni egy kutatócsoport (a mi esetünkben a DigiPhil) munkájába. A kurzus utolsó lépéseként, az elsajátított ismereteket felhasználva kerülhet sor az egyes résztvevők által tervezett digitális kiadással kapcsolatos konzultációra, a publikálási lehetőségek vizsgálatára.

e. Összegzés

Az általam bemutatott átfogó digitális filológiai képzés időtartama legalább egy, de inkább két egyetemi szemeszter időtartamát tenné ki. Még egy többnapos intenzív kurzus esetén is csupán felületes bemutatásra nyílik lehetőség, ugyanakkor ezek a kurzusok, de még a másfél-két órás előadások is nagyban elősegíthetik egy digitális kiadást tervező kutató, illetve kutatócsoport munkáját, hiszen útmutatást adhatnak az első lépésekhez.

Nehézségek, problémák

Egy intenzív digitális filológia képzés, workshop sikerességét számos tényező befolyásolhatja. Az elsődleges probléma, hogy a résztvevők eltérő digitális kompetenciákkal rendelkezhetnek. Előfordul, hogy haladó kurzusra kezdők jelentkeznek, de az alapvető számítógépes ismeretek hiánya is hátráltathatja a kurzus sikerességét. A kurzuson részt vevők összetétele is fontos befolyásoló tényező, általánosságban elmondható, hogy ha kötelező a részvétel az előadáson, illetve kurzuson, akkor a hallgatók motiváltsága igen alacsony. Ahogyan Pierazzo is rámutat, „még az erősen motivált és elkötelezett hallgatók is gyakran küzdenek az ismeretlen környezettel, mivel a formális nyelvek (mint a jelölőnyelvek) a szabadbölcsest más, talán strukturáltabb gondolkodásra készítetik”.³⁹⁹ A kurzus előkészítése során figyelembe kell venni technológiai szempontokat is. Amennyiben egy gépteremben zajlik az oktatás, biztosítani kell, hogy a befogadó intézmény informatikusai előre telepítsék a szükséges programokat. Amennyiben a résztvevők saját laptopot használnak a kurzuson, stabil internetelérést kell biztosítani, valamint az eltérő operációs rendszerek is okozhatnak

³⁹⁹ Simon MAHONY és Elena PIERAZZO, „Teaching Skills or Teaching Methodology?“, in *Digital Humanities Pedagogy: Practices, Principles and Politics*, szerk. Brett D. HIRSCH, 6, 2012, 6.

nehézségeket. Ezeknek a szempontoknak az ismeretében kell dönteni arról, hogy online eszközöket, vagy előre telepített alkalmazásokat mutatunk be. A digitális filológiai munka során a legelterjedtebb eszköz az Oxygen XML Editor, amely folyamatosan ellenőrzi az XML, illetve a TEI XML fájl szintaktikai helyességét, illetve a beépített modulok segítségével lehetőséget biztosít arra, hogy a már korábban bemutatott lekérdezéseket és algoritmusokat egy rendszeren belül futtassuk. Az eszköz legnagyobb hátránya azonban az, hogy nem szabadon elérhető, és ingyenesen csupán 30 napos próbaverzió érhető el. Néhány kivételtől eltekintve azonban ezt az eszközt preferálom, mivel részben kiküszöbölhetők vele az idő rövidege okozta hátrányok. Az Oxygen XML Editor egyik funkciója, hogy automatikusan kiegészíti az XML jelölőket, így felgyorsítva a kódolás folyamatát. Mivel az intenzív kurzusok esetében nem marad idő a kellő számú ismétlésre, az ismeretek elmélyítésére, ez a funkció megkönnyíti a hallgatók és az oktatók dolgát is. Hasonló a probléma a karakterfelismerő eszközökkel. Az egyik legelterjedtebb OCR szoftver, az Abbyy Finereader szintén megvásárláshoz kötött, így a bemutatásához legalább egy licensszel kell rendelkezni. A kézírásfelismerés bemutatása esetében a Transkribus szolgáltatás, ugyan korlátozott mértékben, de biztosít ingyenes átírási lehetőséget az online felületen.

A DigiPhil és az Digitális Bölcsészet Tanszék képzései (2016–2024)

A következőkben, a DigiPhil munkatársaként, általam egyedül, illetve csoportban tartott digitális filológiai kurzusokat mutatom be. Először olyan kurzusokat mutatok be, amelyeken kizárólag kutatók vettek részt, majd egy olyan szemináriumot, amelyen graduális hallgatók voltak, végül olyan intenzív kurzusokat, amelyeken vegyesen vettek részt graduális és PhD hallgatók, valamint kutatók.

a. Képzések kutatók számára

A DigiPhil csoport 2016 tavaszán tartotta az első, több alkalmas digitális filológia képzést, a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetének munkatársai számára. A négy alkalom során a fókusz az XML jelölőnyelv, valamint a TEI ajánlás alapjain volt, a munkatársak ezeken túl közelebbről is megismerték a genetikus kritikai kiadások TEI ajánlás szerinti leképzésének lehetőségeit, majd a közösen elkészítettük az Intézet munkatársa, Török Zsuzsa munkája alapján Arany János Rózsa és Ibolya című költeményének szövegváltozatait tartalmazó TEI átíratát.

A BTK ITI szervezésében került megrendezésre a Digitális módszerek az irodalomtudomány támogatására⁴⁰⁰ című nyílt mesterkurzus, egy olyan előadássorozat, amelynek keretei közt a közönség különböző, digitális bölcsészeti megközelítésmódokkal ismerkedhetett meg. A másfél-két órás előadások, a rendelkezésre álló idő rövidege miatt, inkább ismeretterjesztő jellegűek voltak, az előadók azokat a digitális módszereket mutatták be, amelyek támogathatják, elősegíthetik a kutatók munkáját. A DigiPhil csoport munkatársai három alkalommal tartottak előadást. A mesterkurzusokon, bár minden érdeklődő számára nyitott, jórészt az Intézet kutató munkatársai vettek részt, jelentős részük komoly filológiai gyakorlattal rendelkezik. Az első, Tudományos szövegkiadások jelölőnyelvi átírása⁴⁰¹ című előadásban a digitális filológia történetét és eszközeit mutattam be, a korábban tartott, szűkebb körű képzéshez képest távolabbi, általános nézőpontból.⁴⁰² A második előadásban⁴⁰³ a DigiPhil által fejlesztett kutatói környezetet mutatta be a kutatócsoport, amellyel a digitális kritikai kiadások ellenőrzött környezetben, grafikus felületen, mégis jelölőnyelvi átirat formájában jöhetnek létre. Fontos azonban kiemelni, hogy a második előadás folyamán sem lehetett eltekinteni a jelölőnyelvi átirat bemutatásától és az XML jelölőnyelv alapjainak elsajátításától, ugyanis hiába áll rendelkezésre egy kész eszköz, amelynek segítségével a kutató eleve digitálisan készítheti a kritikai kiadást, ha nem érti meg a mögöttes logikát, nem fogja átlátni az eszközben rejlő lehetőségeket és korlátokat.

Az ELTE Digitális Bölcsészet Tanszék adatgazdász szakirányú továbbképzést indított 2023 szeptemberében.⁴⁰⁴ A három féléves képzés során a hallgatók két féléven keresztül a *Digitális szövegfeldolgozás* című kurzuson sajátíthatják el a már ismertetett OCR és HTR technológiákat, valamint TEI XML alapú kiadások közzétételének lehetőségeit, ami egy közgyűjteményben dolgozó kutató számára kiemelten fontos lehet, ha például tömeges kézírásfelismerést szeretnének végezni a közgyűjteményi anyagokon, vagy digitális forráskiadást szeretnének közzétenni. A képzés kiemelt hangsúlyt fektet továbbá a szemantikus web és az adatgazdagítás kérdéseire is.

⁴⁰⁰ Az előadások listája az alábbi linken érhető el: <https://iti.btk.mta.hu/hu/component/tags/tag/digitalis-modszerek-az-irodalomtudomany-tamogatasa>.

⁴⁰¹ FELLEGI Zsófia, „Tudományos szövegkiadások jelölőnyelvi átírása” (MTA BTK Irodalomtudományi Intézet, Budapest, 2017. március 28.).

⁴⁰² Hasonló felépítésű előadást hallgathattak meg a PPKE BTK Irodalomtudományi Doktori Iskola hallgatói 2020. október 9-én. BOBÁK Barbara és FELLEGI Zsófia, „Bevezetés a digitális filológiába PhD hallgatók számára” (PPKE BTK Irodalomtudományi Doktori Iskola, Budapest, 2020. október 9.).

⁴⁰³ BOBÁK Barbara, FELLEGI Zsófia és PALKÓ Gábor, „Filológia digitális munkakörnyezetben: a Móricz-levelezés kritikai kiadásának tapasztalatai” (MTA BTK Irodalomtudományi Intézet, Budapest, 2019. április 16.).

⁴⁰⁴ A képzés leírása elérhető a Tanszék weboldalán: <https://elte-dh.hu/adatgazdasz/>.

b. Képzés graduális hallgatók számára

Vaderna Gábor a DigiPhil munkatársaival együttműködésben digitális filológiai szemináriumot tartott az Eötvös Collegium hallgatói számára. Az itt alkalmazott módszerek állnak talán a legközelebb Rehbein és Fritze által megfogalmazott „learning-by-project”⁴⁰⁵ megközelítéshez. A szeminárium során a hallgatók három digitális filológiai bevezető előadáson vettek részt, amelyek révén megismerkedtek az XML jelölőnyelv és a TEI ajánlás alapjaival. A szeminárium következő lépéseként a hallgatók bekapcsolódtak a DigiPhil csoport munkájába, részt vettek a Kassák Lajos által szerkesztett Ma folyóirat digitalizálásában. A DigiPhil csoport előkészítette az egyes cikkekhez tartozó TEI XML fájlokat, előre kitöltötte az ismétlődő adatokat, majd minden hallgató, az útmutató alapján elkészítette a jelölőnyelvi átiratokat. Az útmutató tartalmazta a korábbi XML kurzus kulcsfogalmait, hasznos online anyagokat az ismétléshez, továbbá a TEI átirat elkészítéséhez szükséges információkat. Rögzítettük a DigiPhil által kidolgozott TEI specifikáció részleteit, mint például az állandó azonosító előállításának szabályait, illetve a jelölők használatát ragozott forma esetén. Leírtuk továbbá a személynevek VIAF (Virtual Authority Files) nemzetközi névteréből történő azonosításának pontos menetét. Az útmutató a részletes magyarázatokon túl informatív screenshotokat is tartalmazott. A folyamat során a hallgatók megismerkedtek egy kollaborációs munkaszervező eszközzel, valamint a fájlok szakszerű archiválásának menetével. A Ma folyóirat jelölőnyelvi átírata – mivel nyomtatásban jelent meg – nem tartalmaz kódolandó filológiai jellemzőket, például javításokat, azonban az XML fájl fejlécében szereplő metaadatok bonyolult hierarchiába rendeződnek, bibliográfiai adatokat és kereszthivatkozásokat tartalmaznak. A hallgatók adatgazdagítást végeztek a VIAF névtér segítségével. A szeminárium sikerességét mutatja, hogy a hallgatók által előállított jelölőnyelvi átiratok ellenőrzése és javítása kevés időt vett igénybe, valamint később a DigiPhil több, a szemináriumon részt vevő, kiváló munkát végző hallgató számára munkát ajánlott.

Az ELTE Digitális Bölcsészet Tanszék digitális bölcsészet BA minor programot indít, melynek képzési tervében az XML használat és a digitális filológia is hangsúlyos szerepet kap. Az első évfolyam várhatóan 2024 tavaszi félévében indul.

c. Intenzív kurzusok graduális és posztgraduális résztvevők számára

⁴⁰⁵ FRITZE és REHBEIN, „Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing”.

Az ELTE.DH központ, majd annak utódja a Tanszék három alkalommal szervezett nyílt, de előzetes jelentkezéshez kötött digitális filológia workshopot. Az első alkalom egy olyan sorozat részét képezte, amelynek során a kurzusokon részt vevők a digitális filológián túl az adatvizualizáció és a text-mining alapjaival ismerkedtek meg.⁴⁰⁶ A digitális filológiai workshopon főként az ELTE téma iránt érdeklődő kutatói vettek részt, de mesterszakos hallgatók is jelentkeztek. A bevezető előadást követően a résztvevők Petőfi Sándor egy versén keresztül ismerkedtek meg az XML kódolással, majd a TEI versekre vonatkozó ajánlásával. A workshop előkészítésekor kiderült, hogy nincs lehetőség a gépteremben a szükséges XML szerkesztő telepítésére, illetve az internetelés kimaradásai miatt az online szerkesztőprogramok sem voltak elérhetők. Ezért a gyakorlás egy egyszerű szövegszerkesztőben történt, ami nehezebb feladatot jelentett a kurzuson részt vevőknek.

A második alkalomra az ELTE.DH központ által szervezett Digital Philology Winter School 2020 a Visegrad Fund által támogatott Training Digital Scholars: Knowledge Exchange between V4 and Austria utolsó állomásaként került sor.⁴⁰⁷ A négynapos Winter School első napján Palkó Gábor és Véték Bence mutatta be a Distant Reading for European Literary History⁴⁰⁸ projektet, amelynek célja, hogy egy olyan többnyelvű, legalább tíz nyelvet magába foglaló, a TEI ajánlás szerinti, 1840 és 1920 között megjelenő regények jelölőnyelvi átiratából álló korpuszt hozzon létre, az European Literary Text Collectiont (ELTeC), amelyen különböző algoritmusok futtatásával elemzések végezhetőek. A projekt további célja, hogy jó gyakorlatokat dolgozzon ki a többnyelvű szövegtörzsekön végezhető elemzésekre, valamint az eredmények alapján reflektáljon az alkalmazott módszerek irodalomelméletre és irodalomtörténetre gyakorolt hatásaira. Az ELTeC korpuszát kitevő szövegek kiválasztásának szempontjai igen kifinomultak, a részt vevő országoknak figyelembe kellett venni például a beemelt szerzők nemi eloszlását, az egyes művek kanonizáltságát, illetve hosszát. A projekt bemutatásán keresztül vezették be az előadók a résztvevőket az XML jelölőnyelv alapjaiba. A második nap folyamán, építve az első nap során elsajátított ismeretekre, a hallgatók betekintést nyertek az Oxygen XML Editor használatába, megismerték a szoftver olyan alapfunkcióit, mint például a fájlok validálása a TEI ajánlása vagy saját, előre definiált szabályok szerint. Bemutattuk az Oxygen egyéb funkcióit is, az automatikus jelölőkiegészítést vagy például az

⁴⁰⁶ Lásd az *ELTE DH Yearbook 2018* című kiadvány 15–16. oldalát. Online elérhető itt: http://elte-dh.hu/wp-content/uploads/2019/08/ELTE_DH_Yearbook_2018.pdf.

⁴⁰⁷ Lásd: <https://ufal.mff.cuni.cz/training-digital-scholars-knowledge-exchange-between-v4-and-austria>.

⁴⁰⁸ A projekt honlapja: <https://www.distant-reading.net/>.

XML olvashatóságát elősegítő rendezési algoritmust. Ezt követően a résztvevők megismerkedtek az XPath lekérdezőnyelvvvel és az XSLT algoritmusokkal, majd a tanultakat önálló lekérdezések és algoritmusok írása révén ültették át a gyakorlatba. Az általuk készített algoritmusokat nagy mennyiségű XML fájlkon tesztelték. A Winter School harmadik napján a TEI XML alapú természetes nyelvfeldolgozás (NLP, Natural Language Processing) lehetőségeit mutatta be Horváth Péter, az ELTE.DH központ munkatársa az ELTE Verskorpuson keresztül. A képzés utolsó napján a résztvevők a korábban tanultakat felhasználva mutatták be a saját, digitális filológiai témájú kutatásukat, és az oktatókkal közösen mérték fel a továbblépési lehetőségeket.

2021 során az Irodalomtudományi Intézet a Digitális Bölcsészet Tanszékkal és a Digitális Örökség Nemzeti Laboratóriummal közösen nyári és őszi egyetemet hirdetett digitális bölcsészet témakörben,⁴⁰⁹ amelynek felépítése során támaszkodtunk a korábbi kurzusok tapasztalataira. A kurzus célja az volt, hogy bevezesse a résztvevőket a digitális bölcsészet részterületeibe. A négynapos kurzus hibrid formában zajlott, a résztvevők minden nap két, egyenként 2-3 órás workshopon vettek részt. Az eddig bemutatott témákon túl (XML nyelv alapjai, ELTeC, NLP) a hallgatók megismerkedtek a szemantikus webtechnológia és a webarchiválás alapjaival, stilometriai módszerekről tájékozódhattak, valamint elsajátították a LaTeX alapú szövegszerkesztés alapjait. A digitális filológiai workshop a program legvégére került, amikor a hallgatók már magabiztosan használták az XML nyelvet és az Oxygen XML Editort. A workshop egy történeti és elméleti bevezetővel kezdődött. Az első részben a hallgatók megismerték a digitális filológia alapfogalmait, az egyes kiadástípusok közötti különbségeket, a rendelkezésre álló kész eszközöket és a DigiPhil gyakorlatát. Ezt követően először egy kéziratról készített jelölőnyelvi átiraton, majd egy szövegváltozatokat tartalmazó jelölőnyelvi átirat segítségével sajátították el a TEI ajánlás alapjait. Minden workshop-hoz önálló feladat is társult, amelyet a teljes kurzus végéig volt lehetőségük a hallgatóknak elvégezni. A négynapos kurzust egy konzultációs nap követte, ahol a résztvevők kijelöltek egy-egy projektfeladatot, amely egyrészt kapcsolódott a saját kutatásukhoz, másrészt a kurzuson tanultakhoz. A résztvevőknek egy hetük állt rendelkezésre a feladat teljesítésére, amely során az adott terület szakértője mentorálta őket, majd egy zárókonferencia keretei között mutatták be eredményeiket. Ennek a rendkívül intenzív kurzusnak a sikerét mutatja, hogy a nyári

⁴⁰⁹ A nyári és őszi egyetem felhívásai az alábbi linkeken érhetők el: <https://elte-dh.hu/nyariegyetem/>, <https://elte-dh.hu/osziegyetem/>.

egyetemet követően négy résztvevőt alkalmaztunk a DigiPhil-ben, egyet pedig a Laboratóriumban.

Összefoglalás

A disszertáció célja az Irodalomtudományi Intézet digitális filológiai műhelyének, a DigiPhil genezisének és jelen gyakorlatának problémaorientált áttekintése volt. Ehhez, egyrészt, elengedhetetlen a DigiPhil-t megelőző, illetve a kialakulása körüli magyarországi és nemzetközi kontextus vázlatos bemutatása, valamint azoknak az elméleti kérdéseknek, vitáknak, kurrens diskurzusoknak a megidézése, amelyek közvetlenül hatással voltak és vannak a DigiPhil kutatási módszertanának mikro- és makrodöntéseire. A disszertációban egyaránt tárgyaltam a DigiPhil 1.0 és 2.0 filológiai gyakorlatát, és azokat az elméleti kontextusokat, amelyek szerepet játszottak ezek kiépítésében, a munkamenetek, a specifikációk kialakításában és a szolgáltatás komponenseinek kiválasztásában.

A disszertáció elején röviden bemutattam azokat a terminusokat, amelyekre az érvelés támaszkodott, kezdve a *digitális bölcsészet* és a *digitális filológia* meghatározásával. A digitális filológia a digitális bölcsészet alá sorolható fiatal szakdiszciplína, azonban már évtizedekkel a fogalom megjelenése előtt megkezdődött a tudományterület módszertanának kikristályosodása. A szakterminológia az elmúlt évtized során megszilárdult, ezért az olyan fogalmak, mint *digitális*, *digitalizált*, *informatizált* és *born digital* kiadás definiálására szintén szükség volt még a konkrét esettanulmányokat és problémafelvetéseket megelőzően.

A disszertáció első fejezete a digitális filológiába való átmenet metafilológiájával foglalkozott, konkrét, a DigiPhil által informatizált szövegkiadások összefüggésében. A fejezet első részében az Arany János kritikai kiadások és Madách Imre *Az ember tragédiája*, *Szinoptikus kritikai kiadás* retrokonverziójának, az informatizálás folyamatának a bemutatásán keresztül kaptunk képet az egyes kiadástípusok eltérő szövegközlési elveinek egy rendszerben történő leképezéséről. A fejezet második része a DigiPhil együttműködésében készült *born digital* kritikai kiadások készítésének módszertanát elemezte. Ezek elemzése révén láthatóvá vált, hogy milyen elméleti és technikai szempontok határozzák meg egy-egy digitális kritikai kiadás készítését és ezek a szempontok hogyan befolyásolják a kritikai kiadást készítő filológus módszereit.

A filológia alapvetően, függetlenül az adott irányzat vagy műhely szövegközlési elveitől, adatvezérelt. Az adatok komplex kezelésére a nyomtatott médium korlátozottan alkalmas, a kiadások előkészítési munkálatai során létrejövő adathalmaz számos esetben szinte nyom nélkül eltűnik a publikálást követően. A disszertáció második fejezete bevezette a szemantikus web fogalmát, nemzetközi kísérletek és új elméletek elemzésén keresztül mutatta be a szemantikus web elvét és működését a digitális filológia területén. A szemantikusweb-

technológiának alkalmazásával olyan gráf formájú tudástár hozható létre, amely a kutatási adatokat szemantikus hálózatba rendezi és így lehetségessé válik az adatokból további következtetéseket levonni. A disszertáció a DigiPhil által fejlesztett, a szemantikus web technológiájára épülő adatterét mutatta be mint jó gyakorlatot, melynek célja a kritikai kiadások adatkezelésének és adatgazdagításának támogatása.

A jelölőnyelvek és nemzetközi szabványok követése lehetővé teszi a kritikai kiadások gépi elemzését, a nemzetközi szabványok és ontológiák, névterek használata révén pedig lehetőség nyílik az egymástól távol eső adatbázisokban való értelmes és kifinomult közös keresésre. A disszertáció harmadik fejezetében a metaadat-aggregáció folyamatát mutattam be, majd konkrét kiadások számítógépes elemzési lehetőségeit, illetve a különböző adatvizualizációs lehetőségeket vizsgáltam.

A digitális filológia térnyerése az elmúlt évtizedekben vitathatatlan, azonban hiába az évtizedes tapasztalatok, a digitális filológia oktatása kiforratlan még, nem áll rendelkezésre átfogó képzés a leendő szakértők számára. Kifejezetten időszerű kérdés, tekintve, hogy a digitális, illetve digitalizált kiadásokkal foglalkozó kutatócsoportok száma egyre növekszik, de a magyarországi képzések jelenleg nem gondoskodnak szakértői „utánpótlásról”. A disszertáció záró fejezete erre a problémára reflektált annak bemutatásával, milyen lehetőségek álltak rendelkezésre az elmúlt években és manapság a digitális filológia oktatása terén.

Summary

The aim of the dissertation was to provide a problem-oriented overview of the genesis and current practice of DigiPhil, the digital philology working group at the Institute of Literary Studies. To achieve this, it was essential to outline the Hungarian and international contexts preceding and surrounding DigiPhil, as well as to evoke the theoretical questions, debates, and current discourses directly influencing the micro- and macro-decisions of DigiPhil's research methodology. The dissertation addressed both the philological practices of DigiPhil 1.0 and 2.0, as well as the theoretical contexts that played a role in their development, the formulation of workflows, specifications, and the selection of service components.

At the beginning of the dissertation, key terms relied upon in the argumentation were briefly introduced, starting with the definitions of *digital humanities* and *digital philology*. Digital philology is a young sub-discipline of digital humanities, but the crystallization of the methodology of the field began decades before the concept emerged. Over the past decade, specialized terminology has solidified, necessitating definitions for concepts such as *digital*, *digitized*, *computerized*, and *born-digital* editions before proceeding to specific case studies and problem formulations.

The first chapter of the dissertation dealt with the meta-philology of transitioning into digital philology, within the context of specific texts digitized by DigiPhil. The first part of the chapter provided an overview of the critical editions of János Arany and the retro-conversion of Imre Madách's *The Tragedy of Man*, demonstrating the mapping of different editorial principles within a system. The second part of the chapter analyzed the methodology of creating born-digital critical editions in collaboration with DigiPhil. Through this analysis, it became apparent how theoretical and technical considerations determine the methodologies of philologists involved in creating digital critical editions.

Fundamentally, philology, regardless of the editorial principles of a particular trend or research group, is data-driven. The printed medium is limited in its ability to handle complex data structures, and in many cases, the datasets generated during the preparation of editions disappear almost without a trace after publication of the edition. The second chapter of the dissertation introduced the concept of the semantic web, presenting its principles and operation within the field of digital philology through the analysis of international experiments and new theories. By applying semantic web technology, a graph-shaped knowledge base can be created,

organizing research data into a semantic network and enabling further deductions to be made from the data. The dissertation presented the data repository developed by DigiPhil as best practice, based on semantic web technology, aimed at supporting the data management and enrichment of critical editions.

Adhering to markup languages and international standards enables the automated analysis of critical editions, and through the use of international standards and ontologies, it becomes possible to conduct meaningful and sophisticated cross-database searches. The third chapter of the dissertation presented the process of metadata aggregation, followed by an examination of the computational analysis possibilities of specific editions and various data visualization options.

The rise of digital philology in recent decades is undeniable; however, despite decades of experience, the teaching of digital philology remains underdeveloped in Hungary, lacking comprehensive training for future experts. This is a particularly relevant issue considering the increasing number of research groups focusing on digital or digitized editions, yet current Hungarian education does not adequately address the need for expert succession. The final chapter of the dissertation reflected on this problem by presenting the opportunities available in recent years and today for training in digital philology.

Bibliográfia⁴¹⁰

„A magyar avantgárd és a szomszédos országok avantgárd irodalmi és művészeti termése a Kárpát-medencében 1919 és 1929 között. Közös témák, közös kutatási tervek, pályázatok.” Kassák Múzeum, Budapest, 2011. július 1.

AARSETH, Espen J. „Nem-linearitás és irodalomelmélet”. *Helikon* 50, 3. sz. (2004): 313–348.

ALMÁSI, Zsolt. „Europeana: The European Identity Transfigured for and through the Digital”. In *Transfigurations of the European Identity.*, szerkesztette Zsolt ALMÁSI és Bulcsu BOGNÁR, 61–84. Newcastle upon Tyne: Cambridge Scholar Publishing, 2014.

ANDRÁSI, Gábor, szerk. „...fejünkből töröljük ki a regulákat.” *Kassák Lajos az író, képzőművész, szerkesztő és közszereplő.* Budapest: PIM-Kassák Alapítvány, 2010.

„Az avantgárd magángyűjteményekben I. – A Kiss Ferenc-gyűjtemény”. Kiállítás, Kassák Múzeum, Budapest, 2011. november 26.

„Az avantgárd magángyűjteményekben II. A kölcsönös hatások körei. A MA és a Zenit a zágrábi Marinko Sudac-gyűjteményben.” Kassák Múzeum, Budapest, 2012. január 28.

BABITS Mihály. „A jóság dala (Versek, darócban.)”. *Nyugat*, 15. sz. (1917. augusztus 1.): 175–179.

———. *Babits Mihály összes versei. 1902-1937.* Babits Mihály összegyűjtött munkái, 1. Budapest: Athenaeum, 1937.

———. *Nyugalanság Völgye.* Budapest: Táltos, 1920.

———. *Versek 1902–1927.* Budapest: Athenaeum, 1928.

Babits Mihály válogatott versei. Atheneum olvasótábora. Budapest: Athenaeum, 1941.

BAJZÁT Tímea Borbála, SZEMES Botond és SZLÁVICH Eszter. „Az ELTE DH Regénykorpusz és lehetőségei”. In *Online térben az online térért: Networkshop 30: országos online konferencia. 2021. április 6-9. Eötvös Loránd Tudományegyetem,* szerkesztette KOKAS Károly, TICK József és HOLL András, 63–72. Budapest: Hungarnet Egyesület, 2021.

https://real.mtak.hu/132253/1/63_NWS_2021_v3.pdf.

BALOGH Piroska, BODNÁR-KIRÁLY Tibor, GURKA Dezső, KRÁSZ Lilla és KURUCZ György. *A tudományos tudásáramlás mintázatai a Magyar Királyságban, 1770-1830:*

⁴¹⁰ A disszertáció bibliográfiai hivatkozásainak összeállítása az Irodalomtudományi Intézet által készített stílusfájl és Zotero program segítségével készült. A stílusfájl elérhető az Intézet weboldalán: <https://iti.abtk.hu/hu/kiadvanyok/az-irodalomtudomanyi-intezet-cimleirasi-rendje>. A disszertációban hivatkozott internetes linkek utolsó megtekintése: 2024. március 31.

DigiTudásáramlás (DITU) /DigiCirculation of Knowledge (DICIKO). ELTE, 2022.

<https://eltdata.elte-dh.hu/wiki/Tud%C3%A1s%C3%A1raml%C3%A1s>.

BÁNKI Zsolt István. „Felelet a Mandalatra”. *Tudományos és Műszaki Tájékoztatás* 2012, 8. sz. (2012): 311–319.

———. „Múzeumok az Europeanaban - az ATHENA”. *Tudományos és Műszaki Tájékoztatás* 2010, 9. sz. (2010).

BARABÁSI, Albert László. *Linked: The New Science of Network*. Cambridge: Perseus Publishing, 2002.

BARTÓK István és MONOK István. „Szegedi kísérlet a XVI. századi magyar vers gépi feldolgozására”. *Irodalomtörténeti Közlemények* 84, 5–6. sz. (1980): 630–638.

BÉKÉS, Enikő és Sándor BENE. „Módszertani ajánlás a Bibliotheca Scriptorum Medii Recentisque Aevorum új sorozatának (Series Nova) szöveggondozásához”. *Irodalomtörténeti Közlemények* 118, 5. sz. (2014): 698–713.

BEKIUS, Lamyk. „Following the Writer through the Document: Encoding Contemporary Writing Processes in TEI-XML”. Előadás Textual Scholarship in the 21st Century. ESTS 2019, University of Málaga, 2019. 28.

https://drive.google.com/file/d/1_Tb6xxA94lOcProU09aIgF_eHG8h42zj/view.

BENGI László és PARÁDI Andrea, szerk. *Kosztolányi Dezső. Aranysárkány*. Kosztolányi Dezső Összes Művei. Kritikai kiadás. Pozsony: Kalligram, 2014.

BERGER Ágoston, NÉMETH Márton és SIMON András. „Az Europeana Inside és magyar megvalósítása”. *Tudományos és Műszaki Tájékoztatás* 2014, 9. sz. (2014): 342–349.

BERNERS-LEE, Tim, James HANDLER és Ora LASSILA. „The Semantic Web”. *Scientific American*, 5. sz. (2001): 29–37.

BOBÁK Barbara és FELLEGI Zsófia. „Bevezetés a digitális filológiába PhD hallgatók számára”. PPKE BTK Irodalomtudományi Doktori Iskola, Budapest, 2020. október 9.

BOBÁK Barbara, FELLEGI Zsófia és PALKÓ Gábor. „Filológia digitális munkakörnyezetben: a Móricz-levelezés kritikai kiadásának tapasztalatai”. MTA BTK Irodalomtudományi Intézet, Budapest, 2019. április 16.

BOBÁK Barbara, FELLEGI Zsófia, és PALKÓ Gábor, ford. „Szöveggódolás és a Text Encoding Initiative”, 2018. <https://teach.dariah.eu/course/view.php?id=40>.

BOBÁK Barbara és GÁBORI KOVÁCS József. „Kézírásfelismerés Arany János levelein”. In *Networkshop 2019*, szerkesztette TICK József, KOKAS Károly és HOLL András, 38–44. Budapest: Hungarnet Egyesület, 2019. <http://ocs.mtak.hu/index.php/nws/2019/paper/view/23>.

BOBÁK Barbara és KASZA Péter. „Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus Rerum Ungaricarum libri kéziratának digitális kiadása (esettanulmány)”. In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 154–160. Budapest: Hungarnet Egyesület, 2022.
http://real.mtak.hu/155493/1/19_NWS_2022_Bobak_Kasza.pdf.

BODE, Katherine. „Abstraction, Singularity, Textuality: The Equivalence of “Close” and “Distant” Reading”. In *A World of Fiction: Digital Collections and the Future of Literary History*, 17–36. University of Michigan Press, 2018.

———. „The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”. *Modern Language Quarterly* 78, 1. sz. (2017): 77–106.

BORDALEJO, Barbara. „Digital versus Analogue Textual Scholarship or The Revolution is Just in the Title”. *Digital Philology: A Journal of Medieval Cultures* 7, 1. sz. (2018): 7–28. <https://doi.org/10.1353/dph.2018.0001>.

BREE, Pim van és Geert KESSELS. „nodegoat: a web-based data management, network analysis & visualisation environment, <http://nodegoat.net> from LAB1100, <http://lab1100.com>”, 2013.

BRENNEN, J. Scott és Daniel KREISS. „Digitalization”. In *The International Encyclopedia of Communication Theory and Philosophy*, szerkesztette Klaus Bruhn JENSEN, Robert T. CRAIG, Jefferson D. POOLEY és Eric W. ROTHENBUHLER, 556–566. Malden: John Wiley & Sons, 2016.

BUCSICS Katalin, szerk. *Kosztolányi Dezső. Pacsirta. Kosztolányi Dezső Összes Művei*. Kritikai kiadás. Pozsony: Kalligram, 2013.

BURNARD, Lou, Fotis JANNIDIS, Gregor MIDDEL, Elena PIERAZZO és Malte REHBEIN. „An encoding model for genetic editions”, 2010.

BURNARD, Lou, Christof SCHÖCH és Carolin ODEBRECHT. „In Search of Comity: TEI for Distant Reading”. *Journal of the Text Encoding Initiative*, 14. sz. (2021).
<https://doi.org/10.4000/jtei.3500>.

BÜRGERMEISTER, Martina, Zsófia FELLEGI, Gábor PALKÓ, Schneider GERLINDE, Scholger MARTINA, Steiner ELISABETH és Vasold GUNTER. „GAMS and Cirilo: research data preservation and presentation”. Előadás DH 2016, Krakó, 2016.

CERQUIGLINI Bernard. „A variáns dicsérete”. In *Metafilológia 1. Szöveg – variáns – kommentár*, szerkesztette DÉRI Balázs, KELEMEN Pál, KRUPP József és TAMÁS Ábel, fordította KESZEG Anna, 219-295. Budapest: Ráció Kiadó, 2011.

———. *Eloge de la variante. Historie critique de la philologie*. Des Travaux. Paris: Seuil, 1989.

CHACON, Scott és Ben STRAUB. *Pro Git*. 2nd kiad. Apress, 2014. <https://git-scm.com/book/en/v2>.

CUMMINGS, James. „A world of difference: Myths and misconceptions about the TEI”. *Digital Scholarship in the Humanities* 34, Supplement_1. sz. (2019. december 1.): 158–179. <https://doi.org/10.1093/llc/fqy071>.

CSAPLÁR Ferenc. *A Kassák múzeum kiállításai 1976-2006*. Budapest: Petőfi Irodalmi Múzeum - Kassák Múzeum, 2006.

———. *Kassák körei*. Budapest: Szépirodalmi, 1987.

CSÉVE Anna. „Előszó”. In *Móricz Zsigmond levelezésének digitális kritikai kiadása (1882–1913): 2.0*, szerkesztette CSÉVE Anna, PALKÓ Gábor, MIHÁLY Eszter, TÖRÖK Sándor Mátyás, VÉTEK Bence, MAKKAI Csilla Tamara és VARGA Emese. Budapest: Petőfi Irodalmi Múzeum, 2023. https://dhupla.hu/text/mzs-lev_eloszo.

CSÉVE Anna, FELLEGI Zsófia és KÓMÁR Éva. „Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása”. *Digitális Bölcsészet* 1 (2018. szeptember 19.): 159–174. <https://doi.org/10.31400/dh-hun.2018.1.227>.

CSÉVE Anna, PALKÓ Gábor, MIHÁLY Eszter, TÖRÖK Sándor Mátyás, VÉTEK Bence, MAKKAI Csilla Tamara és VARGA Emese, szerk. *Móricz Zsigmond levelezésének digitális kritikai kiadása (1882–1913): 2.0*. Budapest: Petőfi Irodalmi Múzeum, 2021. <https://dhupla.hu/collection/moricz-zsigmond-levelezes>.

CSÉVE Anna, MODOR Bálint, VÉTEK Bence, MAKKAI Csilla Tamara és KAPPELLER Rita, szerk. „Előszó”. In *Móricz Zsigmond: Tükör*. Budapest: Petőfi Irodalmi Múzeum, 2018. <https://dhupla.hu/collection/moricz-zsigmond-tukor>.

———, szerk. *Móricz Zsigmond: Tükör*. Budapest: Petőfi Irodalmi Múzeum, 2018. <https://dhupla.hu/collection/moricz-zsigmond-tukor>.

DACOS, Marin. „Manifesto for the Digital Humanities”. *Marin Dacos* (blog), 2010. március 31. <https://marin.dacos.org/manifesto-for-the-digital-humanities/>.

———. „Manifesto for the Digital Humanities”. *Hypotheses* (blog), 2011. március 16. <https://tcp.hypotheses.org/411>.

DANNAOUI, Elie. „Digital Arabic Gospels Corpus”. In *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, szerkesztette Claire CLIVAZ, Andrew GREGORY és Gregory HAMIDOVIĆ, 61–70. Brill, 2014. https://doi.org/10.1163/9789004264434_006.

DEBRECZENI Attila. „Kritikai kiadás papíron és képernyőn”. In *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*, szerkesztette CZIFRA Mariann és SZILÁGYI Márton, 26–42. Csokonai Könyvtár: Bibliotheca Studiorum Litterarium 55. Debrecen: Debreceni Egyetemi Kiadó, 2014.

DEBRECZENI Attila és KECSKEMÉTI Gábor. „Alapelvek az irodalmi szövegek tudományos kiadásához”. *Irodalomtörténet* 35, 3. sz. (2004): 328–330.

„Digitised Manuscripts to Europeana: Final Public Report”, 2015. https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Digitising_Manuscripts_to_Europeana/Documents/DM2E%20Public_Final_Report.pdf.

DOBÓ, Gábor és Merse Pál SZEREDI, szerk. *Local Contexts / International Networks: Avant-Garde Journals in East-Central Europe*. Budapest: Petőfi Literary Museum – Kassák Museum Kassák Foundation, 2018.

https://monoskop.org/images/b/b2/Local_Contexts_International_Networks_Avant-Garde_Journals_in_East-Central_Europe_2018.pdf.

DOBSON, James és Scott SANDERS. „Distant Approaches to the Printed Page”. *Digital Studies / Le Champ Numérique* 12, 1. sz. (2022. május 5.). <https://doi.org/10.16995/dscn.8107>.

DRÖGE, Evelyn és Julia IWANOWA. „DM2E Model V 1.2 Specification”, 2015. január 20. https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Digitising_Manuscripts_to_Europeana/Deliverables/DM2E_Model_V1.2.pdf.

DRÖGE, Evelyn, Julia IWANOWA és Steffen HENNICKE. „A Specialisation of the Europeana Data Model for the Representation of Manuscripts: The DM2E Model”. *Libraries in the Digital Age (LIDA) Proceedings* 2014, 13. sz. (2014).

DUMONT, Stefan és Martin FECHNER. „Bridging the Gap: Greater Usability for TEI Encoding”. *Journal of the Text Encoding Initiative [Online]* December 2014 – December 2015, Issue 8. sz. (2015). <https://doi.org/10.4000/jtei.1242>.

„EDM - Mapping guidelines”, 2023. március 21. <https://europeana.atlassian.net/wiki/spaces/EF/pages/987791389/EDM+-Mapping+guidelines>.

ERDŐDI Sándor. *Petőfi napjai a magyar irodalomban 1842-1849*. Budapest: Fővárosi Szabó Ervin Könyvtár, 1972.

ESTERHÁZY Péter. *Termelési-regény (kissregény)*. I. Budapest: Magvető, 1979.

———. *Termelési-regény (kissregény)*. V. Budapest: Digitális Irodalmi Akadémia - Petőfi Irodalmi Múzeum, 2011. https://reader.dia.hu/document/Esterhazy_Peter-Termelesi_regeny-430.

ETLINGER Mihály és HERNÁDY Judit. „A kiadás hagyatéka / a hagyaték kiadása: A Régi Magyar Költők Tárának hálózati kiadásáról”. In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 199–203. Budapest: Hungarnet Egyesület, MTA Könyvtár és Információs Központ, 2022.

<https://doi.org/10.31915/NWS.2022.25>.

FEENEY, Kevin. „Graph Fundamentals — Part 2: Labelled Property Graphs”. *TerminusDB Community* (blog), 2019. október 7. <https://medium.com/terminusdb/graph-fundamentals-part-2-labelled-property-graphs-ba9a8edb5dfe>.

FEJES Ildikó, LENGYEL Mónika és SIMON András. „MuseuMap, az első magyar múzeumi aggregációs portál”. Előadás Networkshop 2015, Sárospatak, 2015. április 2. <https://kifu.videotorium.hu/hu/recordings/10582/museumap-az-első-magyar-muzeumi-aggregacios-portal>.

FEKETE Norbert. „HTR-modellépítés és kézírásfelismerés nagyméretű, többszerzős szövegtörzsen A Transkribus alkalmazása az Arany János hivatali iratokon”. In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 271–275. Budapest: Hungarnet Egyesület, 2022.

http://real.mtak.hu/155518/1/34_NWS_2022_Fekete.pdf.

FELLEGI Zsófia. „A digitális filológia infrastruktúrái. A DigiPhil megújulásáról”. In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 338–344. Budapest: Hungarnet Egyesület, 2022. <https://doi.org/10.31915/NWS.2022.43>.

———. „Átmenet az analógból a digitális filológiába: médiumváltás? Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”. *Helikon* 67, 1. sz. (2021): 84–102.

———. „A digitális filológia oktatása: Gyakorlati tapasztalatok”. In *Digitális eszközök a középiskolai irodalomoktatásban*, szerkesztette MOLNÁR Gábor Tamás és FEJES Richárd, 157–178. Budapest: Prae.hu, 2021.

———. „Tudományos szövegkiadások jelölőnyelvi átírása”. MTA BTK Irodalomtudományi Intézet, Budapest, 2017. március 28.

FELLEGI Zsófia és VADÁSZ Noémi. „Digitálisbölcészlet-oktatás Magyarországon”. In *Digitális (szöveg)kultúrák a bölcsészképzésben*, szerkesztette L. VARGA Péter, MOLNÁR Gábor Tamás és PALKÓ Gábor, 168–180. Budapest: ELTE-BTK, 2018.

FIORMONTE, Domenico, Teresa NUMERICO és Francesca TOMASI. *The Digital Humanist: A Critical Inquiry*. Szerkesztette Desmond SCHMIDT, Christopher FERGUSON és Geoffrey ROCKWELL. punctum books, 2015. <https://doi.org/10.21983/P3.0120.1.00>.

FRIED István. „A Madách-kutatás hétköznapi ünnepei (Kerényi Ferenc Madách-köteteiről)”. *Forrás* 39, 2. sz. (2007): 81–89.

FRITZE, Christiane és Malte REHBEIN. „Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing”. In *Digital Humanities Pedagogy: Practices, Principles and Politics*, szerkesztette Brett D. HIRSCH, 47. Cambridge: Open Book Publishers, 2012.

FÜLÖP Csaba, KOVÁCS László és MICSIK András. „A metaadatsémák és a szemantikus web: egységesítés és specializáció a metaadatok világában”. *Tudományos és Műszaki Tájékoztatás* 51, 7. sz. (2004): 276–284.

GABLER, Hans Walter. „A kiadói szöveg születése: számítógép bába-szerepben”. Fordította Ildikó FARKAS. *Helikon*, 35 3–4. sz. (1989): 421–428.

GENETTE, Gerard. *Paratexts: Thresholds of Interpretation*. Fordította Jane E. LEWIN. Cambridge: Cambridge University Press, 1997.

GIOVANNETTI, Francesca. „The Critical Apparatus Ontology (CAO): Modelling the TEI Critical Apparatus as a Knowledge Graph”. In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerkesztette Elena SPADINI, Francesca TOMASI és Georg VOGELER, 125–139. Norderstedt: Books on Demand, 2021.

GÖNCZY Monika. „Az Arany János-bibliográfia”. Előadás Értéktéremtő tudomány. Az Arany-gyűjtemény bemutatkozása, Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet, Budapest, 2019. 20.

GRÉSILLON, Almuth. „Irodalmi kéziratok a technikai sokszorosíthatóság korában”. In *Metafilológia 1. Szöveg – variáns – kommentár*, szerkesztette Balázs DÉRI, Pál KELEMEN, József KRUPP és Ábel TAMÁS, fordította Anna KESZEG, 338-359. Budapest: Ráció Kiadó, 2011.

GYÖREI Zsolt és LOVAS Borbála, szerk. *Kosztolányi Dezső. A szegény kisgyermek panasza*. Kosztolányi Dezső Összes Művei. Kritikai kiadás. Pozsony: Kalligram, 2014.

GYURICZA Andrea. „Szabványos metaadatok jelentősége a kooperatív szolgáltatásokban. Átállás HUNMARC-ról a MARC21 szabványra a WorldCat-csatlakozás előkészítéseként.” *Könyvtári Figyelő* 63, 2. sz. (2017): 218–225.

H. HUBERT Gabriella, szerk. *Lelki énekekről régi magyarok. A XVI-XVII. századi gyülekezeti és graduális éneklésről szóló hosszabb szövegemlékek. Hálózati kritikai kiadás.* Budapest: Gépeskönyv Kft., 1999. <http://magyar-irodalom.elte.hu/gepesk/eloszo/>.

HAAF, Susanne, Alexander GEYKEN és Frank WIEGAND. „The DTA “Base Format”: A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources”. *Journal of the Text Encoding Initiative [Online]* 2014–2015, 8. sz. (é. n.). <https://journals.openedition.org/jtei/1114>.

HAJDU Péter, szerk. *Mikszáth Kálmán, Elbeszélések 13, 1888.* Mikszáth Kálmán összes művei 39. Budapest: Argumentum, 2001.

———, szerk. *Mikszáth Kálmán, Elbeszélések 16, 1893–1897.* Mikszáth Kálmán összes művei 42. Budapest: Argumentum, 2007.

HARTIG, Olaf és Pierre-Antoin CHAMPIN. „RDF-star and SPARQL-star Draft Community Group Report 18 February 2021”, 2021. február 18. <https://w3c.github.io/rdf-star/cg-spec/2021-02-18.html>.

HOLBO, Jhon és Jonathan GOODWIN, szerk. *Reading Graphs, Maps, Trees – Critical Responses to Franco Moretti.* Parlor Press, 2011.

HORVÁTH Iván. „F-BOOK, F-CONTENT”. In *Új interaktivitás felé: f-book.* Budapest: Gépeskönyv Kft., 2001. <http://magyar-irodalom.elte.hu/vita/thi.html>.

HORVÁTH Iván és FUCHS Anna, szerk. *József Attila összes tanulmánya és cikke: Szövegek 1930–1937 v 1.0.* Eötvös Loránd Tudományegyetem, 2012. <http://textus.elte.hu/ja/>.

HORVÁTH Iván és TVERDOTA György, szerk. *József Attila összes tanulmánya és cikke. v 1.0.* Eötvös Loránd Tudományegyetem, 1999. <http://magyar-irodalom.elte.hu/ja/eloidnd.htm>.

———, szerk. *József Attila összes tanulmánya és cikke. v 2.0.* Eötvös Loránd Tudományegyetem, 2001. <http://magyar-irodalom.elte.hu/ja/eloidnd.htm>.

HORVÁTH Péter. „A vershangzás jellemzőinek automatikus feltárása József Attila verseiben”. *Digitális Bölcsészet*, 3. sz. (2020. július 15.): M:3-M:27. <https://doi.org/10.31400/dh-hun.2020.3.422>.

INDIG Balázs, SASS Bálint, SIMON Eszter, MITTELHOLCZ Iván, KUNDRÁTH Péter és VADÁSZ Noémi. „emsv – Egy formátum mind felett”. In *XV. Magyar Számítógépes Nyelvészeti Konferencia*, 235–247. Szeged: Szegedi Tudományegyetem Informatikai Intézet, 2019. <http://real.mtak.hu/99685/>.

JANNIDIS, Fotis. „TEI in a Crystal Ball”. *Literary and Linguistic Computing* 24, 3. sz. (2009): 253–265.

JANSEN, Leo, Hans LUIJTEN és Nienke BAKKER, szerk. *About This Edition*. Amsterdam & The Hague: Van Gogh Museum & Huygens ING, 2009.

https://vangoghletters.org/vg/about_6.html.

JOCKERS, Matthew L. „On Distant Reading and Macroanalysis”. *Matthew L. Jockers* (blog), 2011. július 1. <https://www.matthewjockers.net/2011/07/01/on-distant-reading-and-macroanalysis/>.

JOYEUX-PRUNEL, Béatrice. „Digital Humanities in the Era of Digital Reproducibility: Towards a Fairest and Post-Computational Framework”. *International Journal of Digital Humanities*, 2024. január 3. <https://doi.org/10.1007/s42803-023-00079-6>.

KABDEBÓ Lóránt. „A Szabó Lőrinc-szövegkiadás múltja és jövője”. In *Az V. Nemzetközi Hungarológiai Kongresszus (Jyväskylä, 2001. augusztus 6-10.) előadásai II.*, szerkesztette JANKOVITS József és NYERGES Judit, 91–101. Budapest: Nemzetközi Magyarságtudományi Társaság, 2004.

KÁLMÁN C. György. „Vicc, novella, tanulmány. Jelzés a világba. A Tett. Háború, avantgárd, Kassák. A Kassák Múzeum (Budapest) kiállítása”. *Jelenkor*, 2015. augusztus 1. www.jelenkor.net/visszhang/467/vicc-novella-tanulmany.

„Kassák! A Kassák Múzeum állandó kiállítása”. Kiállítás, Kassák Múzeum, Budapest, 2011. április 19.

„Kassák és Kassák 2.” Kassák Múzeum, Budapest, 2014. április 9.

Kassák Lajos: A Tett (folyóirat). Digitális forráskiadás. Petőfi Irodalmi Múzeum, 2014. <https://dhupla.hu/collection/kassak-lajos-a-tett>.

„Kassákizmus 2. Az új Kassák – a ló meghal és a madarak kiröpülnek”. Kiállítás, Kassák Múzeum, Budapest, 2017. október 13.

KAY, Michael. *XSLT 2.0 and XPath 2.0 Programmer's Reference*. Indianapolis: Wiley Publishing Inc., 2008.

KECSKEMÉTI Gábor. „A textológia egyes problémáiról – az új textológiai alapelvek közrebocsátásakor –”. *Irodalomtörténet* 35, 3. sz. (2004): 317–327.

KELEVÉZ Ágnes. „A Babits-versek kritikai kiadásának sorozata elé”. In *Babits Mihály összes versei, 1, 1890–1905*, szerkesztette Ágnes SOMOGYI és Zoltán HAFNER, 13–47. Budapest: Argumentum, 2017.

———, szerk. *Babits Mihály összes versei, 2, 1906–1910*. Budapest: Argumentum, 2022.

- KERÉNYI Ferenc. „Előszó”. In *Az ember tragédiája. Drámai költemény. Szinoptikus kritikai kiadás.*, 5–9. Budapest: Argumentum, 2005.
- . *Madách Imre (1823-1864)*. Magyarok emlékezete. Pozsony: Kalligram, 2006.
- KIRSCHENBAUM, Matthew. „What Is Digital Humanities and What’s It Doing in English Departments?” In *Debates in the Digital Humanities*, szerkesztette Matthew K. GOLD, 3–11. Minneapolis: University of Minnesota Press, 2012.
- KISS József, szerk. *Petőfi-adattár I*. Budapest: Akadémiai Kiadó, 1987.
- . *Petőfi-adattár III*. Budapest: Akadémiai Kiadó, 1992.
- KOIVUNEN, Marja-Riitta és Eric MILLER. „W3C Semantic Web Activity”. In *Semantic Web Kick-Off in Finland. Vision, Technologies, Research, and Applications*, 27–43. Helsinki: Helsinki Institute for Information Technology, 2002.
- <https://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/proceedings.pdf>.
- KOMLOVSZKI Tibor és S. SÁRDI Margit, szerk. *Régi magyar költők tára: XVII. század, 16, Rozsnyai Dávid, Koháry István, Petrőczy Kata Szidónia és Kőszeghy Pál versei*. Budapest: Balassi Kiadó, 2000.
- KOROMPAY H. János, szerk. *Arany János, Levelezése 3.: 1857–1861*. Arany János Összes Művei 17. Budapest: Universitas Kiadó, 2004.
- , szerk. *Arany János, Levelezése 5.: 1866–1882*. Arany János Összes Művei 19. Budapest: Universitas Kiadó–MTA BTK Irodalomtudományi Intézet, 2015.
- KRÁSZ Lilla. „Tudásképzés és tudásközvetítés Magyarországon 1770–1830”. *Századok* 2017, 5. sz. (2017): 949–952.
- KUCZERA, Andreas. „TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs”. In *Graph Technologies in the Humanities - Proceedings 2020*, 3110. kötet, szerkesztette Tara ANDREWS, Franziska DIEHR, Thomas EFER, Andreas KUCZERA és Joris van ZUNDERT, 101. CEUR Workshop Proceedings. Vienna, Austria: CEUR, 2020.
- <http://ceur-ws.org/Vol-3110/#paper6>.
- KUCZERA, Andreas és Lian NEILL. „The Codex – An Atlas of Relations.” Szerkesztette Andreas KUCZERA, Thorsten WÜBBENA és Thomas KOLLATZ. *Sonderband der Zeitschrift für digitale Geisteswissenschaften*, 4. sz. (2019).
- KURUCZ, György. „A »technological journey« normái: georgikoni tanárok utazási instrukciói a 19. század elején”. *Művelődés-, Tudomány- és Orvostörténeti Folyóirat* 9, 17. sz. (2018): 261–274. <https://doi.org/10.17107/KH.2018.17.261-274>.
- LABÁDI Gergely. „Az olvasó gép: Berzsenyi Dániel versei távolról”. *Digitális Bölcsészet* 1 (2018. szeptember 19.): 17–34. <https://doi.org/10.31400/dh-hun.2018.1.126>.

———. „Franco Moretti: Distant Reading”. *Irodalomtörténet* 95, 4. sz. (2014): 561–564.

———. „Géppel mért irodalom: a mikszáthi előbeszédyszerűség”. *Digitális Bölcsészet* 2 (2019. október 11.): M:3-M:19. <https://doi.org/10.31400/dh-hun.2019.2.390>.

LASSILA, Ora és Ralph R. SWICK. „Resource Description Framework (RDF) Model and Syntax Specification”, 1999. <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/#model>.

LATOUR, Bruno. „A Collective of Humans and Nonhumans: Following Daedalus’s Labyrinth”. In *Pandora’s hope: essays on the reality of science studies*, 183. Cambridge, Mass: Harvard University Press, 1999.

LORANGER, Hoa és Jakob NIELSEN. „Microcontent: A Few Small Words Have a Mega Impact on Business”. *Nielsen Norman Group logoNielsen Norman Group* (blog), 2017. január 29. <https://www.nngroup.com/articles/microcontent-how-to-write-headlines-page-titles-and-subject-lines/>.

MADÁCH Imre. *Az ember tragédiája. Drámai költemény. Szinoptikus kritikai kiadás*. Szerkesztette KERÉNYI Ferenc. Budapest: Argumentum, 2005.

MAHONY, Simon és Elena PIERAZZO. „Teaching Skills or Teaching Methodology?” In *Digital Humanities Pedagogy: Practices, Principles and Politics*, szerkesztette Brett D. HIRSCH, 6, 2012.

MAJOR Ágnes és BUDA Attila. „A Babits-versek kritikai kiadásának átalakulása, avagy a kéziratos hagyaték változásának következményei”. *Irodalomtörténet* 100, 4. sz. (2019): 435–447.

MARÓTHY Szilvia. „A tudományos elektronikus szövegkiadások feldolgozása”. In *Valóságos könyvtár – könyvtári valóság: Könyvtár- és információtudományi tanulmányok 2018*, 351–356. ELTE BTK Könyvtár- és Információtudományi Intézet, 2018. <https://doi.org/10.21862/vkkv2018>.

———. „Beszámoló a bolognai ReIReS workshopról és néhány gondolat a szövegkódolás módszertanáról”. Előadás Digitális módszerek az irodalomtudomány támogatására, MTA BTK Irodalomtudományi Intézet, 2019. 12.

———. „Tudományos szövegkiadások a hálózaton”. *Irodalomtörténeti Közlemények* 122, 5. sz. (2018): 617–633.

MÁTYUS Norbert. „Babits Dante-fordításának kritikai kiadása és Dante életműve”. Dokumentáció. OTKA Kutatási Jelentések. OTKA, 2012. <https://real.mtak.hu/12143/>.

MCGANN, Jerome J. „Az Ulysses mint posztmodern szöveg: a Gabler-féle kiadás”. Fordította Judit FRIEDRICH. *Helikon* 35, 3–4. sz. (1989): 429–452.

MORETTI, Franco. „Conjectures on World Literature”. *New Left Review* 2000, 1. sz. (2000): 54–68.

MUEHLBERGER, Guenter, Louise SEAWARD, Melissa TERRAS, Sofia ARES OLIVEIRA, Vicente BOSCH, Maximilian BRYAN, Sebastian COLUTTO, és mtsai. „Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study”. *Journal of Documentation* 75, 5. sz. (2019. szeptember 9.): 954–976.
<https://doi.org/10.1108/JD-07-2018-0114>.

NAGY Gyula. „Hírlapok digitalizálása és szolgáltatása a világban és Magyarországon A Délmagyarország című napilap digitalizálása és szolgáltatása”. *Könyvtári Figyelő* 2014, 4. sz. (2014): 439–458.

NEILL, Lian és Desmond SCHMIDT. „SPEEDy. A Practical Editor for Texts Annotated With Standoff Properties”. In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerkesztette Elena SPADINI, Francesca TOMASI és Georg VOGELER, 55–71. Norderstedt: Books on Demand, 2021.

NÉMETH Márton. „Az Europeana és a TELPlus projekt”. *Könyv, Könyvtár, Könyvtáros* 2009, 3. sz. (2009): 3–7.

NIELSEN, Jakob. *Designing Web Usability: The Practice of Simplicity*. Indiana: New Riders, 1999.

OLIVER, Andrew. „A la recherche du texte: micro-informatique et manuscrits”. In *La naissance du texte.*, szerkesztette Louis HAY, 41–50. Paris: José Corti, 1989.

———. „Mikroinformatika és textológia”. Fordította Ildikó FARKAS. *Helikon* 35, 3–4. sz. (1989): 412–420.

OLTVÁNYI Ambrus és KISS József, szerk. *Petőfi-adattár II*. Budapest: Akadémiai Kiadó, 1987.

ORLOVSZKY Géza, szerk. *Adriai tengernek syrénája, gróf Zrínyi Miklós. Párhuzamos forráskiadás*. Budapest: Gépeskönyv Kft., 1998. <https://magyar-irodalom.elte.hu/gepesk/zrinyi/>.

PALKÓ Gábor. „A digitális bölcsészet kultúrtechnikái.” In *A humán tudományok és a gépi intelligencia*, szerkesztette TOLCSVAI NAGY Gábor, 33–47. Budapest: Gondolat Kiadó, 2018.

———. „Digitális filológia: számítógép anyaszerepben”. *Filológiai közlöny* 61, 2. sz. (2015): 187–199.

———. „Kritikai kiadások a digitális térben”. Előadás. Móricz Zsigmond levelezésének digitális kritikai kiadása (1892–1913), Petőfi Irodalmi Múzeum, 2019. január 14.

———. „Mesterséges intelligencia, digitális bölcsészet, kulturális örökség: trendek és eredmények”. In *Valós térben - az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 235–240. Budapest: Hungarnet Egyesület, MTA Könyvtár és Információs Központ, 2022. <http://real.mtak.hu/155513/>.

———. „Mit jelent a digitális filológia a szemantikus web korában? A DigiPhil projektről.” *Magyar tudomány* 177, 11. sz. (2016): 1316–1322.

PALKÓ Gábor, BOBÁK Barbara és SZEKRÉNYES István, „A Digitális Örökség Nemzeti Laboratórium webszolgáltatásai automatikus kézírás-felismeréshez”. In *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé*, szerkesztette TICK József, KOKAS Károly és HOLL András, 164–169. Budapest: Hungarnet Egyesület, 2023. <https://real.mtak.hu/182935/>.

PALKÓ Gábor és NEMESKEY Dávid Márk. „Szemantikus névelem-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása)”. In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 359–364. Budapest: Hungarnet Egyesület, 2022.

PALKÓ Gábor és SMRCZ Ádám. „A digitális bölcsészet oktatása az ELTE BTK-n”. In *Digitális eszközök a középiskolai irodalomoktatásban*, szerkesztette MOLNÁR Gábor Tamás és FEJES Richárd, 135–156. Budapest: Prae.hu, 2021. <https://m2.mtmt.hu/api/publication/31936948>.

PALKÓ Gábor, VERES András és DOBÁS Kata. „A PIM DigiPhil műhelyének keretében megvalósuló digitális Kosztolányi-kiadások, valamint a Kosztolányi-adatbázis sajtóbemutatója”. Előadás. Kosztolányi-nap - konferencia és koncert. Kosztolányi 2.0, Petőfi Irodalmi Múzeum, 2018. március 27.

PÉTER Róbert. „A big data kihívás és lehetőség a bölcsészettudományokban: digitális szövegek és metaadatok távoli olvasása”. *Magyar Tudomány* 177, 11. sz. (2016): 1323–1330.

PICHLER, Alois és Amélie ZÖLLNER-WEBER. „Sharing and Debating Wittgenstein by Using an Ontology”. *Literary and Linguistic Computing* 28, 4. sz. (2013): 700–707. <https://doi.org/10.1093/llc/fqt049>.

PIERAZZO, Elena. „Digital Genetic Editions: The Encoding of Time in Manuscript Transcription”. In *Text Editing, Print and the Digital World*, szerkesztette Marilyn DEEGAN és Kathryn SUTHERLAND, 169–186. Aldershot: Ashgate, 2009.

———. *Digital Scholarly Editing : Theories, Models and Methods*. 2. New York: Routledge, 2016.

———. „Of Digital Scholarly Editions and Building Tools”. Előadás Digital Scholarly Editing: Theory, Practice, Methods, Antwerpeni Egyetem, Antwerpen, 2016. október 5.

———. „What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter”. *International Journal of Digital Humanities*, 1. sz. (2019): 209–220.

PIERAZZO, Elena és Julie ANDRÉ. „Le Codage en TEI des Brouillons de Proust: Vers l'Édition Numérique”. *Genesis* 36, 13. sz. (2013): 155–161.

PROSSER, Miller C. és Sandra R. SCHLOEN. „The Power of OCHRE’s Highly Atomic Graph Database Model for the Creation and Curation of Digital Text Editions”. In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerkesztette Elena SPADINI, Francesca TOMASI és Georg VOGELER, 55–71. Norderstedt: Books on Demand, 2021.

PRUD’HOMMEAUX, Eric és Andy SEABORNE. „SPARQL Query Language for RDF. W3C Working Draft 12 October 2004.”, 2004. w3.org/TR/2004/WD-rdf-sparql-query-20041012/.

RENEAR, Allen, Elli MYLONAS és David DURAND. „Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies”. In *Research in Humanities Computing*, szerkesztette Nancy IDE és Susan HOCKEY, 263–280. Oxford: Oxford University Press, 1996.

ROBINSON, Ian, Jim WEBBER és Emil EIFREM. „Data Modeling with Graphs”. In *Graph Databases. New Opportunities for Connected Data*, 2. O’Reilly Media, 2015.

S. VARGA Pál, szerk. *Arany János, Kisebb költemények 3. (1860–1882)*. Arany János Munkái. Budapest: Universitas Kiadó–MTA BTK Irodalomtudományi Intézet, 2019.

———. „Textológiák között: Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás”. *Holmi* 18, 3. sz. (2006): 410–421.

SÁFRÁN Györgyi, BISZTRAY Gyula és SÁNDOR István, szerk. *Arany János, Levelezése I.: 1828–1851*. Arany János Összes Művei 15. Budapest: Akadémiai Kiadó, 1975.

———, szerk. *Arany János, Levelezése 2.: 1852–1856*. Arany János Összes Művei 16. Budapest: Akadémiai Kiadó, 1982.

- SAHLE, Patrick. „About a Catalog Of Digital Scholarly Editions, 2008-2011”. <https://v3.digitale-edition.de/vlet-about.html>.
- . „Digital Humanities? Gibt’s doch gar nicht!” *Zeitschrift für digitale Geisteswissenschaften (ZfdG)*, 2015. http://www.zfdg.de/sb001_004.
- . „On the Ongoing Formation of Digital Humanities - A 2018 Snapshot”. Előadás DH_BUDAPEST_2018, Eötvös Loránd Tudományegyetem, 2018. május 28.
- . „What is a Scholarly Digital Edition”. In *Digital Scholarly Editing: Theories and Practices*, szerkesztette Matthew James DRISCOLL és Elena PIERAZZO, 19–39. Cambridge: Open Book Publishers, 2016.
- SASVÁRI Edit és DOBÓ Gábor. „Kassák Múzeum: a magyar avantgárd kortárs múzeuma”. *Tiszatáj* 71, 12. sz. (2017): 155–164.
- SCHÄLKLE, Kuno és Wilhelm OTT. *TUSTEP. Tübinger System von Textverarbeitungs-Programmen. Version 2023. Handbuch und Referenz*. Tübingen: pagina GmbH, 2023. <https://www.tustep.uni-tuebingen.de/pdf/handbuch.pdf>.
- SCHREIBMAN, Susan, SIEMENS, Ray és UNSWORTH, John, szerk., *A Companion to Digital Humanities*. Malden: Blackwell Publishing, 2004.
- SEBESTYÉN Ádám. „Az ELTEdata szemantikus adatbázis legújabb fejlesztései”. In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 179–184. Budapest: Hungarnet Egyesület, 2022.
- SHILLINGSBURG, Peter. *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press, 2006.
- SOMOGYI, Ágnes és Zoltán HAFNER, szerk. *Babits Mihály összes versei, 1, 1890–1905*. Budapest: Argumentum – MTA BTK Irodalomtudományi Intézet, 2017.
- SPADINI, Elena, Francesca TOMASI és Georg VOGELER, szerk. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*. Norderstedt: Books on Demand, 2021.
- STOLL Béla, szerk. *Régi magyar költők tára: XVII. század, 6, Szenci Molnár Albert költői művei*. Budapest: Akadémiai Kiadó, 1971.
- SULA, Chris A., Sarah E. HACKNEY és Phillipp CUNNINGHAM. „A Survey of Digital Humanities Programs”. *The Journal of Interactive Technology and Pedagogy*, 11. sz. (2017). <https://jitp.commons.gc.cuny.edu/a-survey-of-digital-humanities-programs/>.
- SZEGEDI Béla, szerk. *Effectus Amoris széphistória a XVI. századból. Digitális kritikai kiadás*. Budapest: Gépeskönyv Kft., 1998. <https://magyar-irodalom.elte.hu/effectus/>.

SZEMES Botond, BAJZÁT Tímea, FELLEGI Zsófia, KUNDRÁTH Péter, HORVÁTH Péter, INDIG Balázs, DIÓSSY Anna, és mtsai. „Az ELTE Drámakorpuszának létrehozása és lehetőségei”. In *Valós térben - az online térért*, szerkesztette TICK József, KOKAS Károly és HOLL András, 170–178. NIIF Networkshop konferenciák ; 31. Budapest: MTA Könyvtár és Információs Központ, 2022. <https://real.mtak.hu/155495/>.

SZEMES Botond és VIDA Bence. „Tragikus és komikus hálózatok. Drámai műfajok csoportosítása szerkezeti tulajdonságok alapján”. *Helikon* 68, 2. sz. (2022): 345–368.

SZÉNÁSI Zoltán. „Digitális átalakítás. A tudományos szövegkiadás a weben.” *Helikon*, 66. sz. (2020): 99–112.

SZEREDI Péter, LUKÁCSY Péter és BENKŐ Tamás. *A szemantikus világháló elmélete és gyakorlata*. Budapest: Typotex, 2005.

SZILÁGYI, Emőke Rita. „Editorial Procedures and Types of Censure: On the Upcoming Critical Edition of Nicolaus Olahus’ Correspondence”. In *Nicolaus Olahus 450: Proceedings of the International Conference on the 450th Anniversary of Nicolaus Olahus’ Death*, szerkesztette Emőke Rita SZILÁGYI, 193–203. Wien: Institut für Ungarische Geschichtsforschung in Wien, Balassi Institut Collegium Hungaricum, Wien, Ungarische Archivdelegation beim Haus-, Hof- und Staatsarchiv, Wien, 2019.

———, szerk. *Nicolaus Olahus: Epistulae. Pars II*. Budapest: r e c i t i, 2023. https://www.reciti.hu/wp-content/uploads/olahus2_vn.pdf.

SZILÁGYI Márton. „Textológia, filológia, értelmezés”. In *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*, szerkesztette CZIFRA Mariann és SZILÁGYI Márton, 15–25. Csokonai Könyvtár: Bibliotheca Studiorum Litterarium 55. Debrecen: Debreceni Egyetemi Kiadó, 2014.

TAKÁCS László, szerk. *Kosztolányi Dezső. Nero, a véres költő*. Kosztolányi Dezső Összes Művei. Kritikai kiadás. Pozsony: Kalligram, 2011.

TÓSZEGI Zsuzsanna. „A Digitális Irodalmi Akadémia első két évtizede”. In *Valóságos könyvtár – könyvtári valóság. Könyvtár- és információtudományi tanulmányok*, szerkesztette KISZL Péter és CSÍK Tibor, 155–165. Budapest: ELTE BTK Könyvtár- és Információtudományi Intézet, 2018.

TÓTH Csilla és TÓTH Tünde, szerk. *Wathay Ferenc összes verse hálózati kritikai kiadás*. Budapest–Zalaegerszeg: Gépekönyv Kft., 1999. <http://magyar-irodalom.elte.hu/gepesk/wathay/>.

TÓTH, Réka. *A szöveggenetika elmélete és gyakorlata*. Debrecen: Debreceni Egyetemi Kiadó, 2012.

TÓTH Tünde. „Irodalomtörténészek a bábeli könyvtárból (A régi magyar vers repertórium)”. *Irodalomtörténeti Közlemények* 1997, 1–2. sz. (1997): 133–145.

———. „Online kritikai szövegkiadás Magyarországon az ezredfordulón”. *Helikon*, 50, 3. sz. (2004): 417–442.

———, szerk. *Rimay János verseinek fragmentumi. Kritikai kiadás*. Budapestú: Gépeskönyv Kft., 2001. <http://magyar-irodalom.elte.hu/gepesk/rimay/>.

TÓTH-CZIFRA Júlia és VERES András, szerk. *Kosztolányi Dezső. Esti Kornél*. Kosztolányi Dezső Összes Művei. Kritikai kiadás. Pozsony: Kalligram, 2011.

ÚJ Imre Attila, szerk. *Arany János, Levelezése 4.: 1862–1865*. Arany János Összes Művei 18. Budapest: Universitas Kiadó–MTA BTK Irodalomtudományi Intézet, 2014.

VANHOUTTE, Edward. „Defining Electronic Editions: A Historical and Functional Perspective”. In *Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviours, Products and Institutions.*, szerkesztette Willard MCCARTY, 119–144. Cambridge: Open Book Publishers, 2010.

VERES András, szerk. *Kosztolányi Dezső. Édes Anna*. Kosztolányi Dezső Összes Művei. Kritikai kiadás. Pozsony: Kalligram, 2010.

———, szerk. *Kosztolányi Dezső Édes Anna hálózati kritikai kiadás*. Budapest: Gépeskönyv Kft., 2001. <http://magyar-irodalom.elte.hu/gepesk/kd/edesanna/>.

VOGELER, Georg. „“Standing-off Trees and Graphs”: On the Affordance of Technologies for the Assertive Edition”. In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, szerkesztette Elena SPADINI, Francesca TOMASI és Georg VOGELER, 73–94. Norderstedt: Books on Demand, 2021.

———. „The ‘Assertive Edition’”. *International Journal of Digital Humanities* 1, 2. sz. (2019. július 1.): 309–322. <https://doi.org/10.1007/s42803-019-00025-5>.

WILKINSON, Mark D., Michel DUMONTIER, IJsbrand Jan AALBERSBERG, Gabrielle APPLETON, Myles AXTON, Arie BAAK, Niklas BLOMBERG, és mtsai. „The FAIR Guiding Principles for Scientific Data Management and Stewardship”. *Scientific Data* 3, 1. sz. (2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

WITT, Jeffrey C. „Digital Scholarly Editions and API Consuming Applications”. 12. kötet, szerkesztette Roman BLEIER, Martina BÜRGERMEISTER, Helmut W. KLUG, Frederike NEUBER és Gerlinde SCHNEIDER, 219–247. Norderstedt: BoD, 2018. <https://kups.ub.uni-koeln.de/9118/>.

ZUNDERT, Joris VAN. „Barely Beyond the Book?” In *Digital Scholarly Editing: Theories, Models and Methods*, szerkesztette Matthew James DRISCOLL és Elena PIERAZZO, 83–106. Cambridge: Open Book Publishers, 2016.

ZSIBRITA, János, Veronika VINCZE és Richárd FARKAS. „Magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian”. In *International Conference on Recent Advances in Natural Language Processing*, szerkesztette Ruslan MITKOV, Galia ANGELOVA és Kalina BONTCHEVA, 763–771. Shumen: Incoma Ltd., 2013.