



AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2024



HUN-REN
Magyar Kutatási Hálózat

NETWORKSHOP

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Korrektúra: Danyi Melinda

Angol nyelvi lektor: Cseresnyés Dorottya

Networkshop 2024 konferencia előadásainak közleményei

Eszterházy Károly Egyetem, Eger

2024. április 3–5.

ISBN 978-615-82243-2-1

DOI: <https://doi.org/10.31915/NWS.2024>

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2024

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

TARTALOMJEGYZÉK

Előszó	5
Ungváry Rudolf A MARC21 formátum kettős szerkezete és a formátum felhasználói szintjének fordításai	7
Holl András, Andódy Katalin Adatbányászati gyakorlatok repozitóriumra és MTMT-re.....	16
Simon András Mesterséges intelligenciával támogatott adatgazdagítás a Nemzeti Levéltárban.....	22
Soós Gábor, Rövid András, Ormos Pál V2X – A járművek közötti kommunikáció kihívásai	29
Csernai Zoltán Egy online tanulást támogató portál kurzusának vizsgálata Big Data adatelemző módszerekkel.....	36
T. Nagy László, Németh Áron A mesterséges intelligencia (MI) teológiai kompetenciái	45
Mészáros Erika Kodaktól a jövőig – Egy könyvtári digitalizálás szinterei.....	55
Frankó Máté, Sándor Ákos Adatvizualizáció a könyvtári menedzsmentben: fejlesztések az SZTE Klebelsberg Könyvtár döntéstámogató rendszerében.....	63
Hernek István A felhasználóképzés szintjei az SZTE Klebelsberg Könyvtárban: az elsőévesektől a kutatókig	72
Némethi-Takács Margit, Borbély Mária Bibliográfiai kapcsolatok az általános megjegyzés adatmezőben.....	78
Dobás Kata, Tüskés Anna A magyar irodalomtörténet bibliográfiájának migrációja az ITIdata szemantikus adatbázisba	87
Horváth Péter A kanonikus magyar költészet versformakeresője.....	96
Sebestyén Ádám, Sárközi-Lindner Zsófia Történeti források szemantikus feldolgozása – Az ELTEdata adatbázis új gyűjteményei	105

Bolya Mátyás	
Lyukkártya és népdalrendezés – Egy mechanikus népzenei adatbázis digitális rekonstrukciójának lehetőségei.....	112
Kovácsházy Tamás	
Az idő, mint alapvető infrastruktúra, az idő szerepe az adatközpontban.....	121
Albert Ágota Katalin	
A mesterséges intelligencia használatának követelményei az oktatási szektorban, különös tekintettel a mesterséges intelligencia használatáról szóló rendelethez.....	129
Varga Emese	
Digitális szövegszerkesztés a dHUpa keretrendszerében	135
Nemoda Zsuzsanna, Héjja Balázs, Nagy Andor, Tóth Máté	
A Pest Megyei Digitális Könyvtár fejlesztése	141
Nagy Dóra, Sándor Ákos	
Voice2text: a hanganyagátírás lehetőségei MI segítségével.....	149
Kalcsó Gyula	
Képek és metaadataik gyűjteményezése scrapingtechnológiával közösségi képmegosztó oldalról	157
Péter Róbert, Szántó Zsolt, Biacsi Zoltán, Kocsis Zoltán, Berend Gábor, Bilicki Vilmos	
Az AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) többnyelvű kutatási eszköz munkafolyamata és új funkciói	163
Máray Tamás	
Kvantum-számítástechnika: ez már a „jövő”?	171
Fellegi Zsófia	
Digitális kiadások migrációja: gépi és emberi intelligencia együttműködése.....	177
Palkó Gábor	
Posztmodern intertextualitás és digitális szövegkiadás	184
Antal Dániel	
A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei.....	192
Vass Johanna	
Kutatási adatok megosztása a gyakorlatban – Adatrepozitóriumok használata az Ökológiai Kutatóközpont publikációiban	199
Mihály Eszter, Micsik András, Nagy Kadosa	
Irodalmi levélváltások nyomában TEI-vel és térképpel.....	208

Digitális szövegszerkesztés a dHUpa keretrendszerében

Digital text editing in the dHUpa framework

Varga Emese

Magyar Nemzeti Múzeum Közgyűjteményi Központ
Országos Széchényi Könyvtár Digitális Bölcsészeti Központ (Budapest)
varga.emese@oszk.hu

Abstract

The Digital Humanities Platform (dHUpa), developed at the Digital Humanities Centre (DHC) of the National Széchényi Library, is an online platform for publishing textual resources held in public collections. This service is based on an editorial system that allows the editing and encoding of digital texts in a user-friendly way. Text is encoded in XML format annotated according to the TEI (Text Encoding Initiative) international standard.

This paper aims to introduce the editorial interface, which includes custom toolbars, menus, templates, specialised displays, support functions and various AI-based, automated functions. Through examples of text editing, annotation and data enrichment, it presents the TEI XML specification, tailored by DHC for different text types, which has a built-in validation function (error checking system).

The resulting text corpus is published on dHUpa, which displays the encoded text in a user-friendly way. It allows the creation of different filters, thus supporting the researchability of content, easy searching and different ways of approaching the text.

Keywords: digital humanities, digital scholarly editions, TEI XML, digital text editing, annotation

1. Bevezetés

Az OSZK Digitális Bölcsészeti Központjában fejlesztett Digitális Bölcsészeti Platform, a dHUpa¹, egy korszerű online publikálási felület, amely elsősorban a közgyűjtemények szöveges forrásainak digitális közreadására jött létre. A platform alapja és felépítése már korábban részletesen bemutatásra került², kihangsúlyozva annak jelentőségét a kulturális örökség hozzáférhetővé tétele, és a hosszú távú megőrzés szempontjából. A közzétett anyagok ugyanakkor adatbázis jellegükből adódóan új kutatási lehetőségeket hordoznak magukban³. Az itt megjelenő gyűjtemények eltérő tudományos szintű, illetve célú publikációk (digitális forráskiadások, digitális népszerű kiadások és digitális kritikai kiadások)⁴,

1 Digital Humanities Platform. Hozzáférés: 2023.07.02., <https://dhupla.hu/>

2 Mihály Eszter, *Mi az a dHUpa? A Digitális Bölcsészeti Platform bemutatása*. Valós térben – Az online térért, Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem, 345–358.

3 Varga Emese, *Digitális szövegkiadások és kutatástámogatás a bölcsészettudományokban: A dHUpa platform bemutatása*. CELISR, 139–140.

4 Lásd: dHUpa Szövegkiadások. Hozzáférés: 2023.07.02., <https://dhupla.hu/collection/>

ezért átírási és textológiai jelölési elveik is különböznek egymástól, viszont feldolgozási folyamatuk egységesen illeszkedik a dHUpa infrastruktúrához.

A szövegkiadások publikációs folyamatának első lépésében a digitalizált kéziratképeket a Transkribus⁵ szoftverbe töltjük fel, ahol a képek szegmentálása, a szövegátírás és a textológiai jelölések egy része történik. A Transkribus a szövegkorpuszok előállítása mellett automatikus gépi kézírás-felismerítésre alkalmas, illetve saját gépi kézírás-felismerető modell építését is lehetővé teszi. A DBK által fejlesztett első magyar gépi kézírás-felismerető modell 2021-ben vált publikussá, alkalmazása jelentősen felgyorsítja az átírás folyamatát.⁶ A Transkribusban előállított anyagot TEI XML-formátumban exportáljuk, majd további annotálást és adatgazdagítást végzünk az Oxygen XML Editor programban.⁷

Az infrastruktúra középpontjában a GitLab verziókövető szoftver áll, amely a facsimile képeket, az előállított TEI XML-fájlokat és a programkódokat tartalmazza, hozzájárulva a verziókövetéshez, a publikáláshoz szükséges transzformációhoz, a rendszer karbantartásához és fejlesztéséhez.⁸ A szerkesztés alatt álló szövegek megjelenítése, működése egy belső felhasználású, ún. staging-szerveren történik. Ez után kerülhet sorra a publikáció az éles oldalon.

2. A dHUpa framework-je

Az előállított szövegkorpuszokat a TEI nemzetközi standardjának megfelelő XML-formátumban tároljuk. A TEI standardja keretet biztosít a szövegek kódolásának, hogy azok elemezhetőek, kereshetőek és megjeleníthetőek legyenek a digitális kiadásokban. A TEI többször emlegetett előnye és egyben hátránya is a rugalmassága, amely lehetővé teszi, hogy a jelölés az adott forrásanyag sajátosságaira legyen szabható, ugyanakkor hátrányt jelent, hogy egy jelenséget többféleképpen is lehet tegelni. Ahhoz, hogy az előállított szövegek a dHUpa-platfomon megjeleníthetőek, kereshetőek, lekérdezhetőek legyenek, az általános elvekhez képest egy kötöttebb szabályrendszer létrehozására volt szükség.⁹ A DBK TEI-XML specifikációja szövegtípusok szerint határozza meg a tegelési szabályokat. Külön specifikáció készül a levél-, a monografikus- és a folyóirat-típusú szövegkorpuszokra, amely tartalmazza a metaadat-struktúrát, illetve a szövegben előforduló jelenségek jelölési szabályait. Ez egy meghatározott tegkészletet jelent, attribútumokkal, specifikálva, hogy ebből mi kötelező, mi opcionális, milyen értékek választhatóak, és melyek a szabad-szavasan kitölthetőek. Bár így egy kötöttebb, szigorúbb szabályrendszer áll fel, a szöveg-

5 Transkribus, Hozzáférés: 2023.07.02., <https://www.transkribus.org/>

6 Szűcs Kata Ágnes, *Automatikus kézírás-felismerítés Kiss József levelezésén*. Online térben az online térért: Networkshop 30: országos online konferencia. 2021. április 6–9. Eötvös Loránd Tudományegyetem, 73–80.

7 Oxygen XML Editor, Hozzáférés: 2023.07.02., https://www.oxygenxml.com/xml_editor.html

8 Vö.: Mihály, *Mi az a dHUpa? A Digitális Bölcsészeti Platform bemutatása*. Valós térben - Az online térért, Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem, 350–351.

Mihály Eszter és Micsik András, *Szerkesztői környezet TEI-alapú szövegkiadásokhoz*. Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé, Networkshop 32: országos konferencia. 2023. április 12–14. Pannon Egyetem, Veszprém, 187–188.

9 Mihály Eszter, *Mi az a dHUpa? A Digitális Bölcsészeti Platform bemutatása*. Valós térben – Az online térért, Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem, 348.

jelölési opcióknak továbbra is széles eszköztára áll rendelkezésre, amely a feldolgozandó forrásszöveg sajátosságaira szabható. A specifikáció nemcsak a dHUp!a platformjára szánt szövegtörzsek standardjaként szolgál, hanem más, magyar nyelvű források digitális szövegkiadásának keretrendszereként is használható, ezért hamarosan szabadon elérhető lesz a platformon is.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
		Jelenlég (Framework - szakmai terminológia)			TEI tag									Attribútumok	Megjegyzés
															A kötelező elemek az attribútumok közöttgel jellel (pozitív a TEI szabvány szerint, negatív a dft-ja szintén kötelező elemek és attribútumok). A kötelező elemek és attribútumok kötelező kötelező.
		Gyökerelem			TEI									@xmlns="http://www.tei-c.org/ 1.0"	attribumok
		Fájlfejlés	teiHeader/teiDesc	<teiDesc>											
		Cím információk	teiHeader/teiDesc/iti	<title>										-	
		Fájl címe	teiHeader/teiDesc/iti	<title>										-	lehető - címet (vagy nap)
		Kiadási információk	teiHeader/teiDesc/iti	<edition>										-	
		Kiadási típusa	edition/teiDesc/iti	<edition>										-	lehető értékek: Digitális forráskezelés, Digitális kritikai kiadás, Digitális kiadás, Tudományos digitális kiadás, Digitális forráskezelés kiadás, Digitális népszerűsítő kiadás
		Kiadási kézikönyv kiadás információk	edition/teiDesc/iti	<resp>										-	ismételhető
		A kiadásban betöltött szerep leírása	edition/teiDesc/iti	<resp>										-	kötelező a resp/iti elemben; lehető értékek: Szerkesztő, Gyűjtő alá rendező, Jegyzetkészítő, Általános, Kiegészítő, Kiegészítő, Digitális kiadás, Tudományos kiadás, Kiegészítő, Információs kiadás, Forráskezelés és Fájlkezelés (nem lesz, csak az Impresszumban).
		Személynév	edition/teiDesc/iti	<persName>										@xmlns="http://www.tei-c.org/ 1.0"	egyszerűsített név (név) (név), azonosító (név) (ID), ismétlődő, kötelező a resp/iti elemben attribum kötelező ismétlődő jellel további azonosító megadása (a kiadás) típus (type) PFM, NN, VAF, GEO, AAT, WIKIDATA, CSZK
		További azonosító	edition/teiDesc/iti	<idno>										@type="ADATBÁZIS"	

1. ábra: Specifikáció a levéltípusú források metaadataira

A specifikáció alapját képezi a dHUpLa frameworkjének, amely az Oxygen XML Editor szoftver beépített bővítményeként használható. A framework felhasználóbarát módon teszi lehetővé a szövegek szerkesztését és jelölőnyelvi kódolását, amelyet bárki könnyen használhat a TEI XML-struktúra és tegkészlet mélyebb ismerete nélkül. Fontos, hogy rendszerünkben egy frameworkon belül, külön modulokban érhetőek el a különböző szövegtípusokra specializált fejlécek (template-ek) és az erre vonatkozó validáló funkciók.

A framework-höz tartozik egy személyre szabott, magyar nyelvű eszköztár, amely az összes használható teget (szerkezeti és textológiai elemet) tartalmazza, közvetlenül innen elérhetőek a fejléc-sablonok (pl. levél header), automatizált műveleteket végrehajtó- (azonosítók kiosztása, konverziók, MI-alapú névelim-felismerés, a névterekkel, adatbázisokkal való összekapcsolás), illetve feldolgozási parancsok (metaadat-kinyerés, PDF-előállítás). A specializált megjelenítés és különböző nézetek segítenek eligazodni a szerkesztés folyamatában. A szövegen végzett jelölések részleteit kinyíló fűleken (» jelre kattintva) láthatjuk, majd szerkesztés után becsukhatjuk. Az egyes entitások, illetve a szövegek közötti jegyzetapparátus színben is elkülönül a forrásszövegtől, ezzel is megkönnyítve a jelölésekkel ellátott szövegben való eligazodást.

Személy

▷ ▶ Tekintetes Nemes, nemzetes és Vitéz ▷ ▶ ◁ ▶ ▶ Személynév: Fgységessített név: Petőfi Sándor
Entitástípus: dHUpla azonosító: PL_szemely_001 Petőfi Sándor ▶ További azonosító: 665958 Típus: NN ✓ 44 úrnak, több tekintes ▶ ▶ nemes, nemzetes és vitéző vármegyék
rdemes ▶ ▶ Táblalábjának etc. etc. megkülönböztetett
▶ ▶ hisztelettel.
▷ ▶ ◁ ▶ ▶ Földrajzi név: Egségessített név: Szatmárnémeti Entitástípus: említett_hely dHUpla azonosító: PL_hely_0003 Szatmárnémetiben ▶ ▶ További azonosító: 504519 Típus: NN ✓
A hely
▷ ▶ ▶ Nyedámon
▶ ▶ A keltezésben a helyeként tréfás fantázián: Nyedám, ténylegesen valószínűleg Pest az élettrajzi adatok alapján.
1847 Aug. 18. reszel.

2. ábra: A framework megjelenítője

3. Metaadat-struktúra

A metaadatok rögzítése (TEI Header) részletesen kidolgozott űrlapokon történik, amelyben meg vannak határozva a kötelezően kitöltendő, az opcionális és ismételhető elemek és attribútumok, illetve az egyes adatok rögzítésére vonatkozó szabályok.

Például:

- egy levél címének (<title>) és azonosítójának (idno @type=PID) kitöltése kötelező;
- több levélíró esetén az erre vonatkozó adatmezőt (correspAction @type=sent/persName) ismételten beilleszthetjük az űrlap megfelelő helyére;
- A forrásleírás mint a kézirat azonosítására szolgáló információk, fizikai leírások, külön egységekre (boríték, levél, melléklet) bontott leírások a kiadás igényeinek megfelelő részletességgel rögzíthetőek.

A szerkesztési folyamatot automatikus validáció (Schematron¹⁰) támogatja, ami nemcsak a TEI általános szabályai szerint ellenőrzi a tegelést, hanem a framework-specifikáció alapján is. Ahogy az ábrán is látható, a piros aláhúzás hibás vagy hiányos információt jelez, az ehhez tartozó magyarázat, hibaüzenet a lap alján olvasható.

Korábbi megjelenések
Szöveges információ (bekezdés):
Korábbi kiadásában is olvashatjuk a szöveget.
Bibliográfiai hivatkozás **Forrás:**
PETŐFI Sándor. *Összes művei*. összeáll. KISS József, V. NYILASSY Vilma. 7. köt. Petőfi Sándor levelezése. Budapest: Akadémiai Kiadó, 1964. 75.
Linkhivatkozás
Cél:
Szöveg:
Megjegyzések **Tipus:**
Szöveges információ (bekezdés):
Kapcsolódó linkek: petofi-levelezes **Tipus:** isPartOf

10 A linkhivatkozás (ref) elem cél (target) attribútuma nem lehet üres.

3. ábra: A framework validáló funkciója

4. Az annotálás és adatgazdagítás gyakorlata

A személyre szabott eszköztár mind a metaadatok (TEI Header), mind a szövegek (body) jelenségek jelölését szolgálja. Csoportokba rendezve választhatunk a szerkezeti elemek, az ismételhető meta-elemek, a textológiai tegek, entitások stb. közül.

A jelölendő szövegrész kijelölése után egyszerűen kiválasztjuk az eszköztárból a megfelelő elemet, ezzel be is kerül a tag a szövegre. Ez után további információkkal gazdagíthatjuk a jelölést az attribútumok kitöltése által. A 4. ábrán látható példán egy törölt szövegrésznél megadhatjuk a felelős személyt, az íróeszközt, a törlés típusát, módját, illetve a jelenség egyedi azonosítót is kaphat. Az attribútumoknál is piros aláhúzás és hibaüzenet jelzi, ha a kitöltés hibás, vagy hiányos.

¹⁰ Schematron. Hozzáférés: 2024.07.02., <https://schematron.com/>

▶ **Áthúzás, törles:** Felelős személy: Jókai Mór **Írászközp:** tinta **Felelős egységesített név:** Jókai Mór **Azonnali:** **Típus:** **Mód:** **Művelet azonosító:**
 id_yy3_wb1_ybc **Hatá** egyen meg, a mi meg meg nem evett, mit szídd te az én ▶ **4**apám teremtetű? he? Azt meg csak elszívelem: hogy Júlia ▶ **4**miatt összezártál, arra rászöglültm, de azért ne bízza el ma- ▶ **4**gát kigyemel, nett noha most messze van innen, de majd ▶ **4**hehnye megyek én kendepek Szeretnem hetedikn, s völgenyü id ▶ **4**völgenyü oda, felszemetem kennek a szüz hátán! – Majd adok én ▶ **4**neked káitromkodni ▶ **4** ▶ **4**Artan nem igaz a: hogy én Egresnyem nem küldöm a ▶ **4**lapot hogy a haragos kemköbe ne küldésim! De az csak nem ▶ **4**viárod tán, hogy magam ▶ **4**herdiam? ki is herdiam? (Kedvem ▶ **4**völna olyan orthographiával írni ezt a szót levetel a szá- ▶ **4**smódra hogy tótágat állt tőle) aztán ha a kihórdóval ba- ▶ **4**van, akkor is nekem kellene szólnam, nem egy harmadik ▶ **4**akasztani valónak.

Az 5. ábrán az írás módja van jelölve. Ezt a teget abban az esetben használjuk, ha a szövegben idegen kézírás van a forráson, az attribútumokban pedig jelölhetjük a felelős személyt, az íróeszközt, az írás típusát.

a titv van Frankenburg és tüstöl (ha kedve tartja) az a nyomorék naplója a Pakh gyerek egy betűb. > y^{4D} = t^{4D} = ^{4S}em írt ez ideig senki fiának. Tompa Miska is írt vala- mit, de már nem tudom mit, úgy tartom hogy kiromkodott valamiért, de e tekintetben meg kell vallanom minden ügyesség mellett is, hogy messze áll ...

> a háttad mögött. Kerényi küldöttek verseket, egy meg, fordítások. Pálffy szerelme. Lauka octoberben esküszik. Degre köszál. Imre vígan van, felanny publicuma van mint nekünk. Keményi küld neked valamit majd általam. P. H. L. úr hírem szabadalom folyományod, kapott - kuckert. > Írás módja: Felelős személy: Arany János Íróeszköz: tinta > Felelős egységesített név: Arany János Típus: írás Altípus: idegen...>

Hire van, hogy mit eddig ő felsége sem mert elkövetni. István herczeg Metternichet ma- ga elé hívatta s tudtára adá hogy ezt és ezt követeli tőle. > El' ei^{4D} = ti^{4D} = re^{4D} a bérkocsisok gorbomók. 4

5. ábra: Idegenkezűség jelölése

5. Támogató automatizált funkciók

Generate IDs Metaadatok kinyerése Transkribus kliens export konverzió Transkribus szerver export konverzió Fájlnév beszúrása PID-ként

Keresés a PIM névtérben Keresés a Geonames névtérben Keresés a Vialf névtérben Keresés a Wikidata névtérben Névelem felismerés PDF készítés

6. ábra: Automatizált funkciók az eszköztáron

Az eszköztárból különböző automatizált, illetve feldolgozási parancsokat lehet indítani. A névelem-felismerő (NER) automatikusan felismeri az entitásokat a teljes szövegen, és kategóriák szerint (helynév, személynév, intézménynév) beilleszti a megfelelő TEI XML-tegeket (pl. <persName>, <placeName>) a talált nevekhez. Ezt a műveletet a metaadat-mezőben is elvégezhetjük. Ezek ellenőrzése után manuálisan javíthatjuk a jelölést, ha az automatikus felismertetés helytelenül, vagy nem azonosította be az összes tulajdonnevet.

Ez után következik a névelem összekapcsolás és egyértelműsítés egy szintén beépített, félautomatizált funkcióval. A tegelt névelem kijelölése után valamelyik névtérben (PIM, GeoNames, Vias, Wikidata) indítunk keresést, majd a találatokból kiválasszuk a megfelelő névelemet, amivel az azonosító és az egységesített név is automatikus kitöltésre kerül a tegelésben. (Például: <persName nymRef="Egry József">Egry József<idno type="PIM">PIM52426</idno></persName>) Ez annyiban könnyíti meg az adatgazdagítás folyamatát, hogy a névelemek azonosítóit nem kell külön a névtér oldalán megkeresni, és bemásolni, hanem az adatok a framework infrastruktúráján belül kerülnek be a szövegbe. Mindig az adott projekt határozza meg, hogy melyik névtérrel kötjük össze a kiadással, esetenként az adott entitásnak legalkalmasabb névtérrel is használjuk (ahol azok az információk érhetőek el, amelyek a kiadásunk számára lényegesek), ugyanakkor egyszerre több névtérrel is összeköthetjük az entitásokat. Bizonyos esetekben a kiadásokban szereplő névelemekhez szeretnénk olyan projekt- vagy gyűjteményspecifikus jegyzeteket hozzáfűzni, amelyek nem érhetőek el egyik adatbázisban sem (pl. lásd a Mórícz Zsigmond-levelezéskiadás említett személyneveit, ahol a jegyzetekben az adott személy Móríczal való

kapcsolatáról olvashatunk). Ezek a névelemek a dHUpla-infrastruktúrán belül fejlesztett Entitástárba¹¹ kerülnek.

A Transkribusból kinyert TEI XML-konverziókat lehet tömegesen végezni, azonosítókat kiosztani, illetve PDF-generálás is lehetséges. A metaadatok kinyerése funkcióval táblázatosan kapjuk meg akár egy teljes levelezéskorpusz adatait, amelyek aztán különböző vizsgálatok, statisztikák és vizualizációk alapjául szolgálhatnak.¹²

6. Összefoglalás

A dHUpla frameworkje gazdag eszköztárat és felhasználóbarát felületet nyújt különböző szintű és típusú digitális szövegkiadások előállításához, lehetővé téve a digitális szöveg-előállítás korszerű gyakorlatát. Az elkészült korpuszokat a dHUplán publikáljuk, amely a kódolt szövegeket felhasználóbarát módon jeleníti meg, illetve különböző szűrők (ún. facet-ek) kialakítását teszi lehetővé, ezzel támogatva az előállt tartalmak kutathatóságát, a könnyű keresést, és összességében a szöveg különböző megközelítési módjait. Folyamatos teszteléssel és fejlesztésekkel tovább bővül a funkciók és a megjelenítés gazdagsága, amely a feldolgozandó gyűjtemények sajátosságait is figyelembe veszi.

Irodalomjegyzék

- Mihály Eszter, *Mi az a dHUpla? A Digitális Bölcsészeti Platform bemutatása. Valós térben - Az online térért*, Workshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem. Kiadja a HUNGARNET Egyesület az MTA Könyvtár és Információs Központ közreműködésével, Budapest (2022): 345–358. ISBN 978-615-82243-0-7 <https://doi.org/10.31915/NWS.2022.44>, <https://real.mtak.hu/155532/>
- Mihály Eszter és Micsik András, *Szerkesztői környezet TEI-alapú szövegkiadásokhoz. Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé*, Workshop 32: országos konferencia. 2023. április 12–14. Pannon Egyetem, Veszprém. Kiadja a HUNGARNET Egyesület az MTA Könyvtár és Információs Központ közreműködésével, Budapest (2023): 186–191. ISBN 978-615-82243-1-4 <https://doi.org/10.31915/NWS.2023>
- Szűcs Kata Ágnes, *Automatikus kézírás-felismertetés Kiss József levelezésén*. Online térben az online térért: Workshop 30: országos online konferencia. 2021. április 6–9. Eötvös Loránd Tudományegyetem. Kiadja a HUNGARNET Egyesület (2021): 73–80. ISBN 9786150129563 <https://doi.org/10.31915/NWS.2021.8>
- Varga Emese. „Digitális szövegkiadások és kutatástámogatás a bölcsészettudományokban: A dHUpla platform bemutatása”, *Közép-európai Könyvtár- és Információtudományi Szemle* (CELISR) 1. évf., 2. sz. (2024): 134–143. <https://doi.org/10.3311/celistr.37709>

¹¹ dHUpla Entitástár. Hozzáférés: 2024.07.02., <https://dhupla.hu/collection/nevter>

¹² Lásd az OSZK és SZTAI közös fejlesztésű Térképes vizualizációját. <https://dbk.dsd.sztaki.hu/terkep/> Hozzáférés: 2024.07.02.