

VALÓS TÉRBEN – AZ ONLINE TÉRÉRT

Networkshop 31: országos konferencia

2022. április 20–22.
Debreceni Egyetem

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2022



A kötet megjelenését támogatta az
Energiaügyi Minisztérium

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Workshop

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

ISBN 978-615-82243-0-7

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

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

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

TARTALOMJEGYZÉK

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



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

Mi az a dHUpla? A Digitális Bölcsészeti Platform bemutatása

Mihály Eszter
Országos Széchényi Könyvtár, Digitális Bölcsészeti Központ
mihaly.eszter@oszk.hu

Abstract

One of the main tasks of the National Széchényi Library's Digital Humanities Centre is to develop an online platform for textual resources and manuscripts held in public collections, providing a common research environment for literary studies, linguistics and other humanities disciplines. It enables the digital publication of a text corpus, the content analysis using natural language processing and other computer tools, the creation of data visualisations, and the processing and publication of born digital text content.

The service also makes available the wide knowledge of other institutions through the databases linked to the platform (namespaces, bibliographic databases) and facilitates the international integration of the field (Europeana).

The dHUpla platform (dhupla.hu) makes a part of the Hungarian cultural heritage accessible by digitising and computerising the national data assets and archiving them in a most advanced way for long-term preservation. The contemporary research and user interface also includes digital philological recommendations for publishing the content of scientific projects, responds to the challenges of the 21st century transformation of public collection functions.

This paper presents the development of the platform so far and its ongoing projects.

Keywords: digital humanities, digital philology, digital text edition, searching, metadata, text-image linking, manuscript, text corpus, data visualisation, corpus search, annotation.

Absztrakt

Az Országos Széchényi Könyvtár Digitális Bölcsészeti Központ egyik elsődleges feladata egy korszerű online platform fejlesztése a közgyűjteményekben őrzött szöveges források számára, amely egységes kutatói környezetet jelent az irodalomtudomány, a nyelvtudomány, és más humán tudományok számára. Lehetőséget teremt digitális szövegtörzsek publikálására, a platformon elérhető tartalmakon nyelvtechnológiai és egyéb számítógépes eszközökkel történő elemzések elvégzésére, adatvizualizációk létrehozására, illetve ún. born digital (digitálisan született) szöveges tartalmak feldolgozására és közzétételére is.

A szolgáltatás biztosítja a más intézményekben fellelhető tudás széles körű újrahasznosítását is a platformmal kapcsolatban álló adatbázisok révén (névterek, bibliográfiai adatbázisok), illetve elősegíti a szakterület nemzetközi integrációját (Europeana).

A dHUpla platform (dhupla.hu) hozzáférhetővé teszi a magyar kulturális örökség eddig elérhetetlen rétegét a nemzeti adatvagyon digitalizálásával és gépi feldolgozásra alkalmassá tételével, továbbá a legkorszerűbb módon



archiválja azt a hosszútávú megőrzés céljából. A digitális filológiai ajánlásokat is tartalmazó, tudományos projektek tartalmainak publikálására is lehetőséget nyújtó korszerű kutatói és felhasználói felület a közgyűteményi funkciók átalakulásának 21. századi kihívásaira reagál.

Jelen tanulmány a platform fejlesztésének eddigi eredményeit és folyamatban lévő projektjeit mutatja be.

Kulcsszavak: digitális bölcsészet, digitális filológia, digitális szövegkiadás, keresés, metaadatok, text-image linking, kézirat, szövegtörzs, adatvizualizáció, korpuszlekérdezés, annotálás.

Elsődleges profil



1. ábra: A dhUpla nyitóoldala

A szolgáltatás kialakításának alapvető célja eddig kiadatlan, nehezen hozzáférhető vagy valamilyen szempontból kiemelt jelentőségű, a tudományos-kulturális érdeklődés középpontjában álló, illetve éppen méltatlanul elhanyagolt *kézírtos* források digitális feldolgozása, közzé- és kutathatóvá tétele. A kézírtatok újszerű, pontos definícióját mellőzve¹ azt mondhatjuk, jelenleg ide soroljuk többek között a művek szerzői kéziratait, fogalmazványokat, naplókat, levelezéseket, feljegyzéseket stb., amelyek ma már természetesen szülehetnek eleve digitális formában is. A platform alkalmas nyomtatásban már megjelent szövegek digitális kiadására is, de nem ez a szöveges források elsődleges feldolgozandó köre. Vannak természetesen kivételek, mint például Kassák Lajos folyóiratainak publikálása: <https://dhupla.hu/collection/kassak-lajos-a-tett>.

Publikáció

Már tavalyi előadásomban is felvettem, hogy a források digitális másolata létrehozásának általános gyakorlatával az is kérdésessé vált, hogy a közzététel mely szintjét nevezhetjük

1 Meglepő módon sem a magyar nyelvű, sem a nemzetközi szakirodalomban nem találtam közmegegyezésen alapuló meghatározást. E téma kibontása természetesen külön tanulmány keretein belül lehetséges.

publikációnak. Annyi bizonyos, hogy a szövegkiadás egyes szintjei mára a közgyűjtemények hatáskörébe (is) kerültek. A GLAM szektor és a felhasználók részéről is – teljes joggal – egyre gyakoribb az igény, hogy a szöveges tartalmak ne képként (digitális faksimile), hanem digitális szöveggént legyenek elérhetőek az online térben. A Digitális Bölcsészeti Központ (DBK) éppen ezért elsősorban digitális forráskiadásokat készít, illetve kiemelt projektek esetében digitális kritikai kiadásokat.

Infrastruktúra főbb elemei

A kiépített infrastruktúra középpontjában egy verziókövető rendszer áll (GitLab),² amelyben a tárolt anyagok minden változása visszakövethető, s innen indul ki minden automatizált folyamat, ezzel áll kapcsolatban közvetlenül vagy közvetve az összes használt eszköz.

Az eredetileg tervezett komponensek közül az XML-adatbázist végül egyelőre mellőztük, ugyanis a gyakorlati tapasztalatok azt mutatták, hogy a jelenleg elvárt funkciók enélkül is mind megoldhatóak, s így tovább egyszerűsödik az informatikai háttér felépítése, ami fejlesztői és felhasználói szempontból is öröndetes.

Nagy előnye a git alapú rendszernek, hogy gyakorlatilag teljesen szoftverfüggetlen, így a közgyűjteményi szféra különböző infrastruktúráival könnyen kompatibilissé tehető. A szövegek elvárt formátuma TEI XML, amely nemzetközileg is a legáltalánosabban használt szövegjelölő nyelv, így a legtöbb projekt befogadhatóvá válik. A moduláris kialakítás továbbá lehetővé teszi, hogy a munkafolyamat különböző pontjain is be lehessen csatlakozni: ha például egy projekt nem a dHUpla szerkesztőségi környezetében, hanem más módon vagy más szoftverekkel szeretné előállítani a publikálandó tartalmakat, akkor megteheti, hogy a kész XML fájlokat adja át a Git-be való betöltésre.

Szerkesztő program 1. Transkribus - Digitális átírat

A kéziratok kézi és automatikus gépi átírásához, valamint az ún. text-image linking (kép és szöveg összekötése) elvégzéséhez a Transkribus³ programot használjuk, amely TEI XML formátumban is exportálható. A kézi átírás ingyenes, a mesterséges intelligencián alapuló automatikus kézírásfelismerő (Handwritten Text Recognition) modell építése is, a létrehozott modellel pedig 4-500 kézírásos oldal ingyenesen felismertethető, utána fizetnünk kell érte. Ennek fejében viszont készen kapunk egy eszközt az automatikus gépi kézírásfelismertetés eléréséhez, a betanító szövegek előállításán túl csak a modell paraméterezését kell elvégeznünk. Az egyes modellek egymásba is építhetők, így egyre általánosabb érvényű HTR-modell hozható létre az adott nyelven. Az első publikus magyar nyelvű kézírásfelismerő modellt a Digitális Bölcsészeti Központ 2022. március 4-én tette nyilvánossá.⁴

Ahhoz, hogy a jövőben ezen a szakterületen tömeges feldolgozás vehesse kezdetét, a különböző magyar nyelvű projekteknek, amelyek automatikus kézírásfelismertést használnak vagy használnának, mindenképp össze kell fogniuk. Egyrészt azért, hogy a saját szövegtörzsekön betanított *modelleket egymásba építve* egyre általánosabb érvényű eszköz jöjjön létre a magyar kéziratok digitalizációjához, másrészt azért, hogy a tömeges adatfeldolgozás lebonyolításához megfelelő szakmai, személyi és anyagi feltételek

2 GitLab, hozzáférés: 2022. 09. 19., <https://about.gitlab.com/>.

3 Transkribus, ReadCoop, hozzáférés: 2022. 09. 19., <https://readcoop.eu/transkribus>.

4 *Hungarian handwriting 19th–20th cent.*, ReadCoop, hozzáférés: 2022. 09. 19., <https://readcoop.eu/model/hungarian/>. A Transkribus, illetve a HTR használatáról és eredményeiről dr. Szűcs Kata kötetben szereplő tanulmánya számol be részletesen.



mellett megvalósítható *intézményközi vagy akár nemzetközi megállapodásokat* tudjanak kötni az olyan fejlesztői műhelyekkel, mint a Transkribus.

1.1. Biztonság

A Transkribus programmal lehetséges a saját gépen való szerkesztés is, de az összes funkció kihasználásához a szerverre is fel lehet tölteni a dokumentumokat, amely szintén verziókezelővel biztosítja a tárolásukat. A számos felhasználói funkció teljes igénybevétele miatt a projektek során a szerveren dolgozunk, ezért saját fejlesztés keretében megoldottuk a Transkribus-gyűjtemények automatikus napi mentését saját szerverre is.

Szerkesztő program 2. Oxygen XML Editor - TEI XML

A dHUpla oldalára szánt szövegeket a DBK által készített TEI XML specifikációk szabályai szerint kell létrehozni. Mivel a TEI igencsak rugalmas XML formátum, az általános elvekhez képest meglehetősen szigorú szabályrendszer létrehozására volt szükség ahhoz, hogy az annotációk alapján a platformon való megjelenítés, kereshetőség, lekérdezhetőség stb. jól működjön. Az útmutató külön foglalkozik a metaadatok, illetve a szövegtest tagelési módjával. A magyar nyelvű digitális szövegkiadások, azon belül a különböző kategóriákba sorolt szövegtípusok jelölési útmutatója hamarosan bárki számára elérhető lesz a DBK oldalán.

ATEI XML-ek további szerkesztésére az Oxygen XML Editor⁵ programot használjuk. Ezen belül a szerkesztőségi környezet optimalizálása céljából kialakítunk egy ún. frameworköt, amely a dHUpla igényeire van szabva, annak minden szempontját figyelembe véve. A framework egyrészt kiterjedt textológiai-filológiai jelölésrendszer alkalmazására nyújt lehetőséget anélkül, hogy a szerkesztőnek a technológiai hátteret és az XML világot ismernie kéne, másrészt az adatgazdagítás, metaadatok részletes kitöltése, külső adatbázisok bekapcsolása, szerkezeti és szemantikus annotációk alkalmazása, alapszinten a text-image linking mind felhasználóbarát módon megoldható ebben a környezetben. A létrejövő TEI XML fájl végül tulajdonképpen önmagában egy strukturált adathalmazt képez, amely a legkülönbözőbb módokon lesz feldolgozható (ld. kreatív tartalmak),⁶ lekérdezhető.

A framework- és a git alapú funkciók teszik lehetővé a dHUpla rendszerének egyik legfontosabb elemét, a különböző automatizált műveletek beépítését (pl. automatikus transzformációk, azonosítókiosztás stb.), vagy éppen a szerkesztőségi folyamatok felügyeletét, irányítását.

2.1. Framework-funkciók

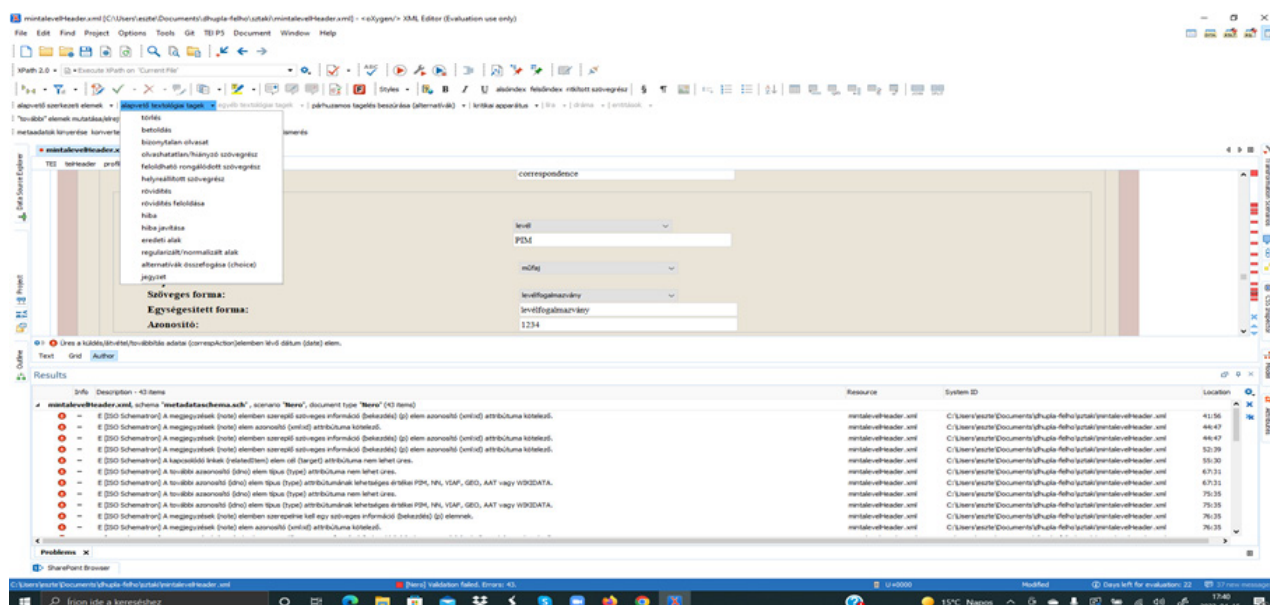
Az alább ismertetett funkciók jelenleg tesztelési fázisban vannak.

1. A szerkesztőfelületet igyekszünk úgy kialakítani, hogy a folyószöveg minél jobban elkülönüljön a TEI tagektől és azok attribútumaitól, hogy jól olvasható és áttekinthető legyen, ugyanakkor az annotáció hozzáadása vagy szerkesztése is egyszerű maradjon. Ezért a tagek csak akkor látszódnak, ha a felhasználó szeretné, továbbá a szerkesztő magyar terminusokkal dolgozhat, legördülő listákból választhat, gombnyomással annotálhat, felajánlott opciókból válogathat. Az ún. TEI fejléctet, tehát a metaadatokat jól strukturált űrlap kitöltésével lehet létrehozni.

⁵ XML Editor, Oxygenxml, hozzáférés: 2022. 09. 19., https://www.oxygenxml.com/xml_editor.html.

⁶ Kreatív tartalmak, dHUpla, hozzáférés: 2022. 09. 19., <https://dhupla.hu/page/kreativ/>.

2. A validáláshoz, illetve a felhasználó számára felajánlott segítő funkciók meghatározásához ún. sémafájlokat (RELAX NG, Schematron) alkalmazunk, amelyek az aktuális nemzetközi TEI szabvány mellett saját szabályrendszerünk szerint is definiálják és ellenőrzik a dokumentumokat. A Schematron sémával való ellenőrzéskor a felhasználó általunk megfogalmazott szöveges hibaüzenetet kap, ha nem megfelelő a tagelés (pl. „Hiányzik a bekezdés azonosítója”), vagy figyelmeztetheti olyan gyanús esetre is, amely adott szituációban lehet helytálló is (pl. „A levél mérete nem mm-ben van megadva. Biztos jó így?”). Mindez nagyban megkönnyíti a helyes formátumú TEI XML fájl kialakítását, hiszen kiszűri az emberi szemmel rendkívül nehezen, nagyobb mennyiségű szövegfeldolgozás esetén pedig egyáltalán nem felfedezhető tévedéseket.



2. ábra: A szerkesztőfelület

3. A különböző szövegtípusokhoz létrehozunk ún. template-fájlokat, amelyek egy új dokumentum készítéséhez szükséges alapvető elemeket tartalmazzák, ebből indulhat ki a szerkesztő minden új szöveges objektum létrehozásakor, így soha nem kell nulláról kezdenie a munkát, a keret már eleve adott lesz hozzá.

4. Automatizált műveletek

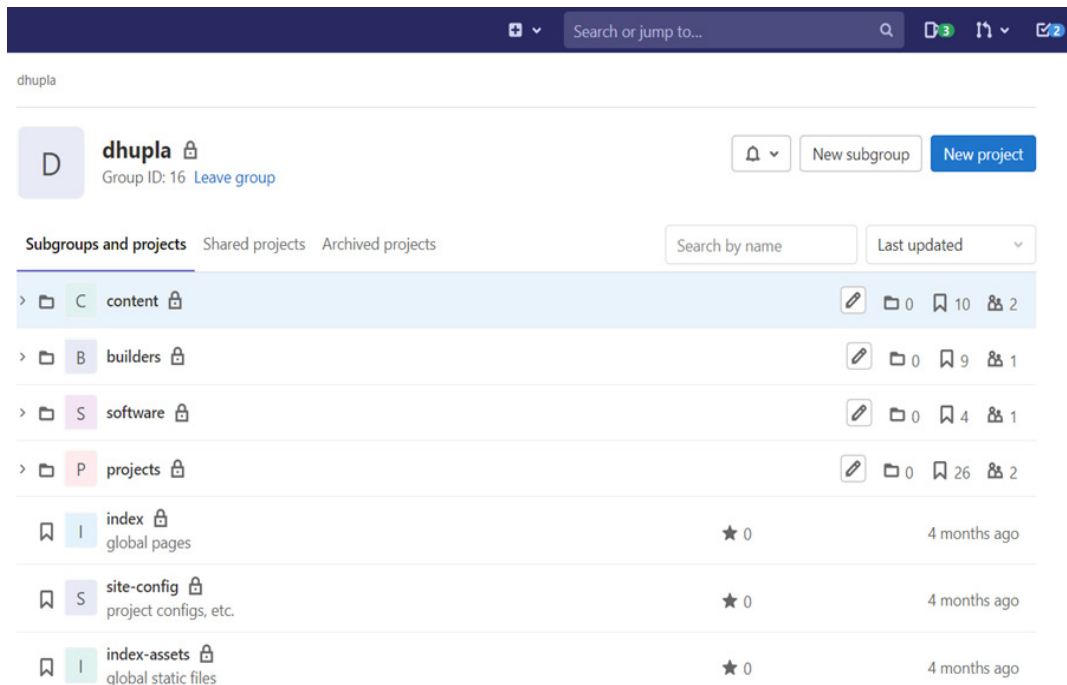
- Automatikus névelemfelismerés: egy adott szövegben vagy szövegrészben egy gombnyomásra lefuttathatjuk a mesterséges intelligencián alapuló ún. Named Entity Recognition⁷ funkciót, amely felismeri az abban szereplő tulajdonneveket (személynév, helynév, testületi név, műcím), és el is látja azokat a megfelelő annotációkkal⁸.
- Entitásazonosítás: az ún. entitásokat (ld. előző pont, azonkívül ún. terminusok: pl. szakkifejezések, köznevek), valamint akár bibliográfiai tételeket is beazonosíthatjuk külső névterekben, adatbázisokban való közvetlen kereséssel, s egy kattintással elláthatjuk a megfelelő egyedi azonosítóval, valamint megadhatjuk az egységesített névformát is.

7 Named-entity recognition, Wikipedia, hozzáférés: 2022. 09. 19., https://en.wikipedia.org/wiki/Named-entity_recognition.

8 A jelenlegi funkció a huSpacy (<https://github.com/huspace/huspace>) NER moduljára épül, mivel a magyar nyelvű írott normához közeli szövegtípusok esetén jelenleg ez bizonyult a leghatékonyabbnak, ugyanakkor folynak párhuzamos fejlesztések e téren, tehát a modul felépítése még változhat.



- c) Automatikus transzformációk: jelenleg a Transzkribusban létrehozott anyagokhoz készült ilyen funkció, amely az onnan exportált szövegeket automatikusan átalakítja a dHUpla számára valid TEI XML fájlkká.
 - d) PDF előállítás: a TEI XML-ekből egyedileg és tömegesen is létrehozhatunk PDF formátumú fájlokat, amelyek a born digital szövegkiadásból nyomtatott kiadás előkészítésére alkalmasak. A PDF elrendezését, formázását a DBK határozza meg.
 - e) Automatikus metaadatkinyerés adatvizualizációkhoz: jelenleg a digitális levelezéskiadásokhoz készült funkció, amellyel egy gombnyomással kinyerhető egy teljes szövegtörzs összes releváns metaadata, s máris felhasználható a legkülönbözőbb adatvizualizációk, statisztikák készítéséhez.
5. Az Oxygen rendelkezik ún. Git-klienssel is, amelynek segítségével a szerkesztők közvetlenül a verziókezelő rendszerbe dolgozhatnak, illetve megoldható a szerkesztőségi környezetből való közvetlen publikáció is.
- 6.



3. ábra: A GitLab-repozitórium

A Git verziókezelő rendszer lát el gyakorlatilag minden olyan feladatot, amelyre a publikáció menedzseléséhez szükség van.

1. Repozitóriumként (repository) funkcionál, tehát nincs szükség külön adatbázis használatára és karbantartására, ami rendkívüli módon leegyszerűsíti az üzemeltetési és supportálási munkát és a munkafolyamatot is.
2. A verziókezeléssel megoldható, hogy egy szöveg életét a kezdetekig visszakövessük, megtaláljunk egy folyamatban elkövetett hibát, vagy akár visszaállítsunk egy korábbi verziót. Hosszútávú terveink közé tartozik, hogy a felhasználók számára is elérhetővé tesszük a korábbi ún. release⁹-ek verzióit.

⁹ Egy teljes szövegtörzs újabb kiadásnak tekintett változata.

3. A szerkesztők párhuzamosan is végezhetik a munkájukat, akár ugyanazon a dokumentumon is, ilyenkor a változtatások merge-ölésével (összegyűrésével), illetve az ún. konfliktusok feloldásával lehet elmenteni a közös változtatásokat.
4. Mivel az Oxygen programnak van Git kliense, a felhasználók közvetlenül dolgozhatnak a Git-be, minden változást azonnal látnak magában az XML-szerkesztő programban.
5. A különböző szerepkörökhöz különböző szintű jogosultságokat osztunk ki, így meghatározhatunk pl. szerkesztői, főszerkesztői, lektori tevékenységekhez tartozó engedélyeket.
6. Az általunk használt Gitlab ún. issue tracker (feladatkövető) használatával írjuk össze és követjük nyomon az egyes anyagokkal kapcsolatos teendőket, hibajavításokat, tesztelési feladatokat, megfelelő személyhez rendeljük, határidőt szabunk a feladathoz.
7. Ún. pipeline-ok¹⁰ alkalmazásával megoldhatóak az olyan automatizált eljárások, amelyeknek mindig a munkafolyamat egy adott pontján kell lezajlaniuk. Ilyen például az XML well-formedness (jól formáltság) ellenőrzése betöltéskor, vagy az állandó azonosítók kiosztása objektum szinten. Ha az automatikus validáció hibát jelez, fejlesztésünk segítségével a projekt-admin email-ben értesítést kap, s a hibás tartalmat nem lehet betölteni, amíg ki nem javítják.
8. A Git repositoryban egy helyen tárolhatjuk a tartalmakat és a programkódokat, mindent ugyanott kezelhetünk, átlátható struktúrába rendezve.
9. Az ún. build-művelet, amelynek során a TEI XML-ből előáll a honlapon publikált HTML tartalom, teljes egészében automatizált, semmilyen manuális munkára nincs hozzá szükség.
10. A docker-technológiával¹¹ való összehangolásnak köszönhetően lehetővé vált a projektszintű kezelhetőség. Ez azt jelenti, hogy egy adott projekt bármilyen szoftverrel, bármilyen módon előállíthatja a TEI XML fájlokból a megfelelő HTML-eket, s beépítheti az adott projekt build-folyamatába.
11. A különböző típusú szövegkiadásokat (gyűjteményeket) kategóriákba soroltuk, amelyek beállításai testreszabhatóak egyszerű szöveges fájlok (YAML) kitöltésével. Itt adhatjuk meg például, hogy az adott kiadásban milyen szűrési lehetőségek jelenjenek meg, vagy milyen szempontok alapján lehessen rendezni az objektumokat. A konfigurációs fájlokban bármikor új kategóriákat is meghatározhatunk.

¹⁰ CI/CD pipelines, GitLab, hozzáférés: 2022. 09. 19., <https://docs.gitlab.com/ee/ci/pipelines/>.

¹¹ Docker (software), Wikipedia, hozzáférés: 2022. 09. 19., [https://en.wikipedia.org/wiki/Docker_\(software\)](https://en.wikipedia.org/wiki/Docker_(software)).

A feldolgozás státuszai



4. ábra: A staging server

A szerkesztés alatt álló szövegkiadások megjelenésének és működésének ellenőrzésére létrehoztunk egy belső felhasználású ún. staging-szervert, amely az éles szolgáltatás másolata azzal a különbséggel, hogy itt nem csak a publikusra beállított tartalmak,¹² hanem minden a repozitóriumban található tartalom megtekinthető. Így a szövegkiadásban résztvevő munkatársak már a tartalomelőállítás folyamata közben is láthatják eredményeiket, a javítandó hibákat, illetve jelezhetnek informatikai teendőket is. Ezen a felületen tesztelhetünk újabb fejlesztéseket is akár olyan szövegkiadásban, amely célzottan csakis erre lett létrehozva (lásd: Demo gyűjtemény).

¹² A gyűjtemények, illetve a bennük szereplő objektumok státusza lehet draft (szerkesztés alatt), withdrawn (visszahívott), finished (kész) és published (publikus).

A publikáció alapegységei

Digitális kritikai kiadás

Móricz Zsigmond levelezése (1892–1913)

2.0 verzió

Az alap kutatás célja Móricz Zsigmond levelezésének (1892–1913) forrásfeltárása és közzététele digitális kritikai kiadás formájában. A projekt az NKFIH támogatásával valósult meg. [Tovább olvasom.///](#)

elejére vissza [tovább](#) [végére](#) 1-30/1526

[régebbiek elől](#) [újabbak elől](#) [ábécé szerint](#)

keresés

keresés a gyűjteményben

szűrési feltételek törlése

| feladó-címzett | típus | keletkezés |
|--|--------------|------------|
| Móricz Zsigmond – A Debreceni Református Kollégium tanári kara (1892-06-20) | levél | 1892-06-20 |
| Pallagi Erzsébet – Móricz Zsigmond (1894-03-23) | levél | 1894-03-23 |
| Móricz Zsigmond – Hajdu István (1894-09-10) | levél | 1894-09-10 |
| Móricz Zsigmond – Pallagi Gyula (1896-09-03) | verses levél | 1896-09-03 |
| Heeger János – Móricz Zsigmond (1898-06-16) | levél | 1898-06-16 |
| Heeger János – Móricz Zsigmond (1898-06-19) | levél | 1898-06-19 |
| Heeger János – Móricz Zsigmond (1898-07-16) | levél | 1898-07-16 |
| Móricz Bálint, Móricz István, Pallagi Erzsébet – Móricz Zsigmond, Móricz Dezső (1899 után) | levél | 1899 után |
| Heeger János – Móricz Zsigmond (1899-03-18) | levél | 1899-03-18 |
| Balázs József – Móricz Zsigmond (1899-08-25) | levelezőlap | 1899-08-25 |

TÍPUS ^

- levél (903)
- képeslap (249)
- levelezőlap (227)
- levélfogalmazvány (40)
- boríték (24)
- névjegy (16)

SZERZŐ / LEVÉLÍRÓ ^

- Móricz Zsigmond (495)
- Holics Janka (121)
- Móricz Dezső (77)
- Szklanár Teréz (69)
- Móricz István (58)
- Pallagi Gyula (43)

FELADÁS HELYE ^

5. ábra: A Móricz Zsigmond-levelezése gyűjtemény nyitóoldala

A dHUpla kiadásai alapvetően háromféle szerkezeti egységből épülnek fel. A legnagyobb egység az ún. gyűjtemény, amely egy teljes szövegkiadást foglal magában. Magához a gyűjteményhez is tartozik egy TEI XML fájl, amely tartalmazza a gyűjtemény metaadatait, leírását, állandó azonosítóját stb. Egy gyűjtemény lehet egyetlen regény, de akár egy teljes levelezés kiadása is. Ebben találhatóak az ún. digitális objektumok, azaz a kiadáshoz tartozó szöveges egységek TEI XML formátumban, valamint az ún. információs oldalak, amelyek Markdown vagy HTML formátumúak lehetnek.¹³

Megjelenítés, szűrés, keresés, rendezés, hálózatoság

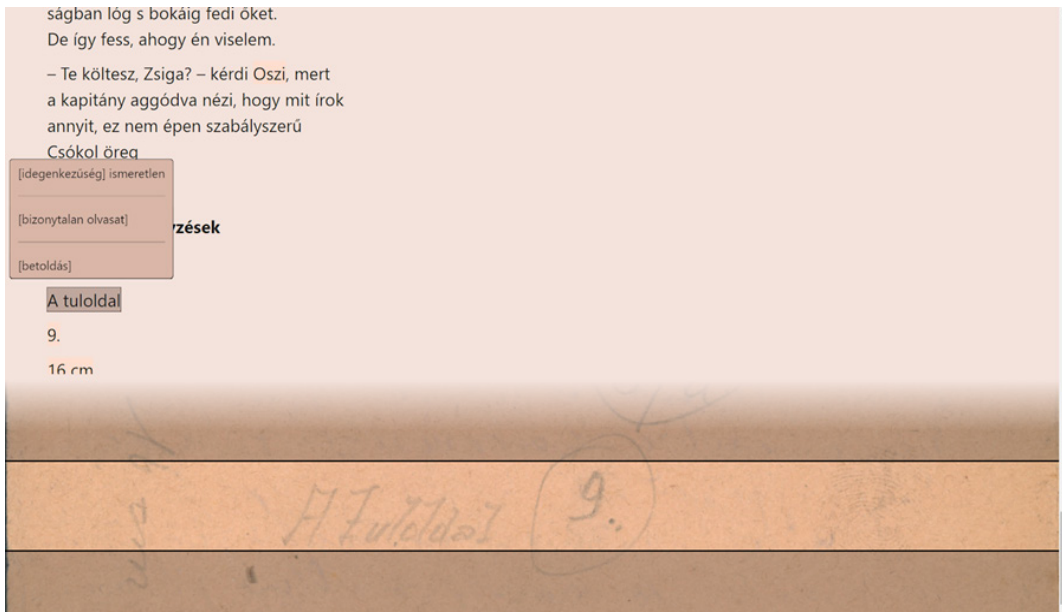
Az általunk alkalmazott text-image linking szövegfeldolgozó módszer rendkívül jól használható minden felhasználói réteg számára. Ennek lényege, hogy a digitális faksimile zónái össze vannak kapcsolva az átírt szöveggel, tehát a felhasználó együttesen láthatja a kettőt. A dHUplán többféle nézet állítható be, attól függően, hogy soronként vagy kisebb kontextusban szeretnénk látni az éppen olvasott szövegrész fotón megfeleltethető részét, de arra is lehetőség nyílik, hogy a képernyőt kettéosztva párhuzamosan olvassuk az eredeti kéziratot és az átíratát.

¹³ Ilyenek például a kreatív tartalmak oldalai is, de létre lehet hozni bármilyen kiegészítő jellegű, nagyobb terjedelmű járulékos tartalomhoz.



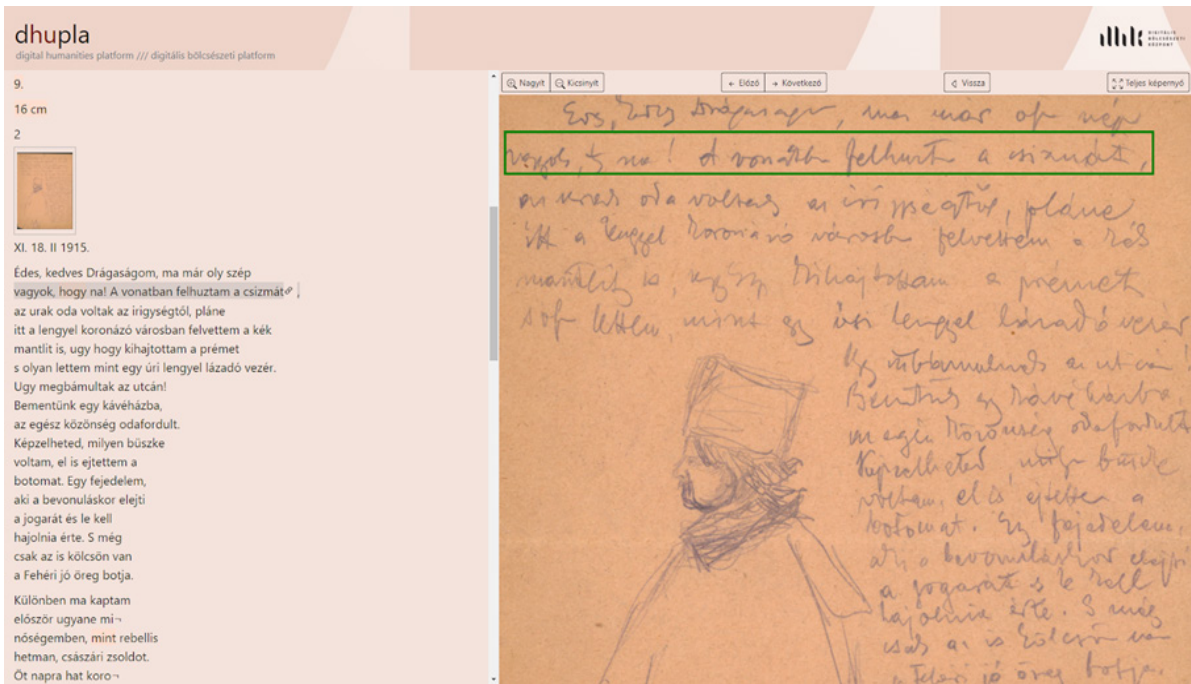
6. ábra: Kép és szöveg

A szövegben szereplő annotációk jól láthatóvá válnak azáltal, hogy kiemeltük az annotált szövegrészeket, emellett igyekeztünk jól olvashatóvá tenni magát az annotáció tartalmát is (pl. textológiai-filológiai annotációk, tárgyi és magyarázó jegyzetek, entitások beazonosítása).



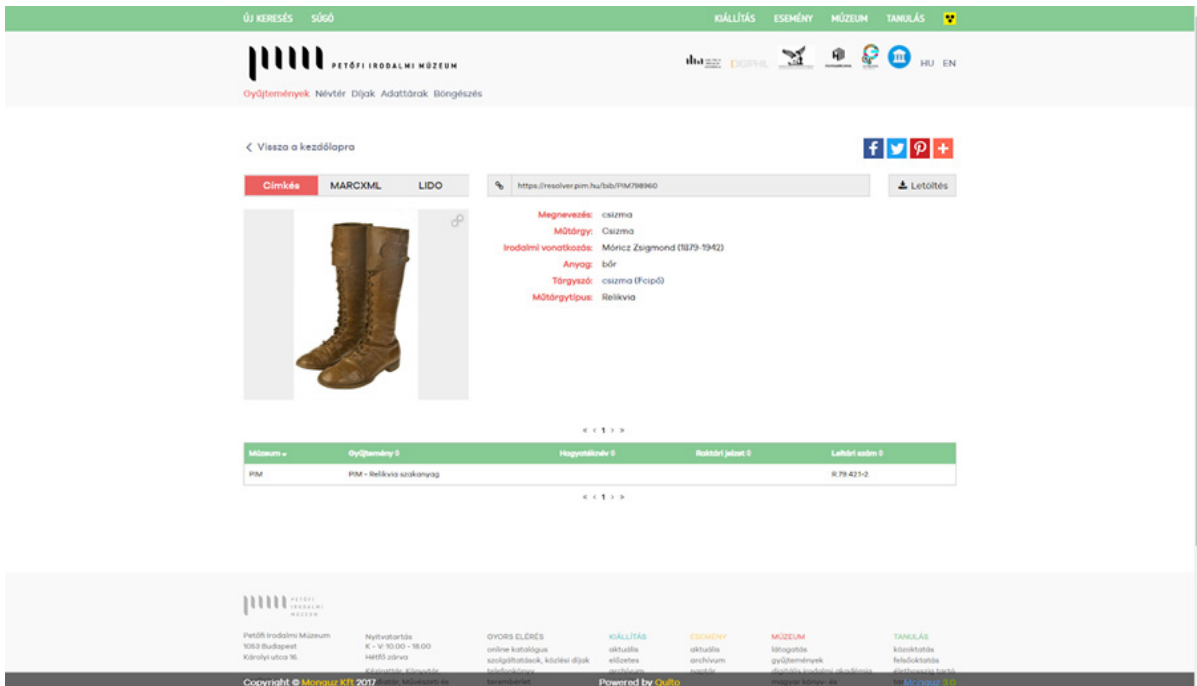
7. ábra: Az annotációk megjelenítése

A szöveg a digitális térben menthetetlenül az adatok hálójába kerül, ami ez esetben nem csapdát vagy megkötöttséget, hanem éppen a határtalan szabadságot teremti meg számára. Jó példa erre Móríc Zsigmond levele, amelyben csizmájáról mesél:



8. ábra: Móricz Zsigmond levele

Afelhasználó egy kattintással továbbis léphet a Petőfi Irodalmi Múzeum Relikviagyűjteményébe, és megtekintheti a szóban forgó csizmát:



9. ábra: Móricz Zsigmond csizmája a PIM Gyűjteményi katalógusból

Ehhez hasonló és még jóval komplexebb adatkapcsolatok építhetők ki a különböző adattárakkal, online tartalmakkal. A fent említett entitásazonosítások, bibliográfiai hivatkozások révén szintén szerteágazó adathalmazban találja magát az érdeklődő felhasználó.



Mindemellett a dHUpa oldalán belül is létrehoztunk egy ún. Entitástárat (<https://dhupla.hu/collection/nevter>), amely kifejezetten a platformon elérhető szövegekben szereplő nevek a kiadás szempontjából releváns adatait tartalmazza. Az egyes entitások itt is kapnak egyedi azonosítót, de természetesen összeköthetőek külső névterekkel, adatbázisokkal.



10. ábra: a dHUpa Entitástára

A keresési, szűrési műveleteket ún. facetek használatával tettük egyszerűvé, támogatóvá. A facetek a dokumentum metaadatai szerint is működhetnek, de a szövegben szereplő annotációk szerint is megadhatunk szűkítési lehetőségeket (pl. íróeszköz, említett műcímek szerint). Szabadszavas keresést egy gyűjteményen belül, de a teljes dHUpa tartalmaira is indíthatunk. A Solr¹⁴ alapú keresőbe szótövelő eszközt is beépítettünk, hogy a keresett kifejezés összes todalékolt változata is megjelenjen a találatok között.

14 Solr, hozzáférés: 2022. 09. 19., <https://solr.apache.org/>.

előjére vissza tovább végére 1-6/6 relevancia szerint időpontoktól újabbak elől cím szerint csizma

keresés a gyűjteményben

szűrési feltételek törlése

TÍPUS ^

levél (5)
képeslap (1)

SZERZŐ / LEVÉLÍRŐ ^

Móricz Zsigmond (2)
Ady Lajos (1)
Holics Janka (1)
Móricz István (1)
Szilágyi Julianna (1)

FELADÁS HELYE ^

n.a. (5)
Budapest (1)

CÍMZETT ^

Móricz Zsigmond (4)
Holics Janka (2)

ÁTVÉTEL HELYE ^

n.a. (5)
Budapest (1)

KELETKEZÉS HELYE ^

n.a. (1)
Barszentdéd (1)
Budapest (1)
Kisújszállás (1)
Kölköc (1)

képeslap
Ady Lajos – Móricz Zsigmond (1913-06-18)
Kézsók – az Asszonytól csók – meleg köszönet, bő szüret (majd őszre) és hozzá illő nagy **csizma!** Igaz szeretettel: Ady Lajosok

levél
Szilágyi Julianna – Móricz Zsigmond (1909-09-22)
Higye meg anyi bajom van – az a kis pénz a mit kapok ha kedves Apám a lakást és a háztartást nagy részbe nem fizetné nem tudom mit tennék, most megint kura alá kell jönnöm, a kislíknak ruha **csizma** kell, nekem sincs igazán meleg ruhám most azt beszerezni és miből, igazán nagyon szegénylenem de magát akarom megkérni ha kölcsönözne nekem januárig 100 vagy 150 frntot forintot a kamatját is fizetném decemberbe a mit kapok aból mindjárt le fog hatna lefoghathna 50 forintot.

levél
Holics Janka – Móricz Zsigmond (1904-12-24)
– Pali nem tud egyebet hirtelen kitalálni, hát Jenő : Egy csizmatalpat Pali : Ha még csak azt, egy egész **csizmat** huzzunk ki. – Hát ez nem furcsa?

levél
Móricz István – Móricz Zsigmond (1902-08-25)
De csak nekik rohan Dühös Varju Gábor Lazsnakolja őket Dühös Varju Gábor. Mikor oszt elfárad az útlegelesbe **Csizma** szára mellől Kést kap a kezébe. Dühös Varju Gábor Meg szurkálja őket Véres mán a kése De szurkálja őket.

levél
Móricz Zsigmond – Holics Janka (1904-08-29)
Édes Egyem, ne haragudj én rám, Soha, soha, Ha a sorsom, a mely ily bajt mért rám, – Oly mostoha: Irót, költőt, írka-fírka gyártót, Én belélem ilyen mestert gyártott: Hisz tudod, Jól tudod: A csizmadianak rongyos a **csizmája**. A kertésznek gyomos az ó háza tája, A kovácsnak nincsen patkó a **csizmáján**. A doktornak pedig baj lakozik máján, A szegény írónak, Rongyos firkálónak Soha sincs ideje sirni hogy ha fáj ám!

levél
Móricz Zsigmond – Holics Janka (1910-04-26)
Ma egész nap ezen ábrándoztam, hogy te kis gazdasszony volnál, én meg totyakos, terebélyes gazda. Vadász **csizmat** huznék, lovagolnék s zöld kalapom lenne gazdászosi! A tokámat három rófosé ereszteném s ha megunnánk az idillt, hip, hop – egy ugrás Pest , rögtön ott volnánk kultur embernek.

11. ábra: Keresés és szűrés

Egy gyűjtemény objektumai több szempontból rendezhetőek, s ahogy fentebb már szó volt róla, az aspektusok az igények szerint konfigurálhatóak:

Digitális kritikai kiadás

Móricz Zsigmond levelezése (1892–1913)

2.0 verzió

Az alapkutatás célja Móricz Zsigmond levelezésének (1892–1913) forrásfeltárása és közzététele digitális kritikai kiadás formájában. A projekt az NKFIH támogatásával valósult meg. [Tovább olvasom. ///](#)

előjére vissza tovább végére 1-30/1526 régebbiek elől újabbak elől ábécé szerint keresés

keresés a gyűjteményben

szűrési feltételek törlése

TÍPUS ^

levél (903)
képeslap (249)
levelezőlap (227)
levélfogalmazvány (40)
boríték (24)
névjegy (16)

SZERZŐ / LEVÉLÍRŐ ^

Móricz Zsigmond (495)
Holics Janka (121)
Móricz Dezső (77)
Szklenár Teréz (69)
Móricz István (58)
Pallagi Gyula (43)

FELADÁS HELYE ^

| feladó-címzett | típus | keletkezés |
|---|----------|--------------|
| A Jövendő – Móricz Zsigmond (1910-04-22) | levél | 1910-04-22 |
| Ács Emma – Móricz Zsigmond (1902-12-31) | képeslap | |
| Ács Emma – Móricz Zsigmond (1903-05-11) | levél | 1903-05-11 |
| Ács Vilma – Móricz Zsigmond (1900-09-27) | képeslap | 1900-09-27 |
| Ady Endre – Móricz Zsigmond (1913-07) | képeslap | |
| Ady Lajos – Móricz Zsigmond (1913-06-18) | képeslap | |
| Ady Lajos – Móricz Zsigmond (1913-07-02) | képeslap | 1913-07-02 |
| Ady Lajos – Móricz Zsigmond (1913-08-15) | képeslap | 1913-08-15 |
| Ady Lajos – Móricz Zsigmond (1913-12-15) | levél | 1913-12-15 |
| Ágai Béla – Móricz Zsigmond (1906)-08-18 | névjegy | [1906]-08-18 |
| Árkosi-Ferenczi Kálmán – Móricz Zsigmond (1911-03-25) | levél | 1911-03-25 |

12. ábra: A digitális objektumok rendezése



Adatvizualizációs lehetőségek

A metaadatok, illetve a szövegekben elhelyezett annotációk lekérdezésével és strukturált feldolgozásával statisztikák, vizualizációk végtelen sora állítható elő az egyes szövegkorpuszokról. Ilyenek többek között a dHUpla oldalán (<https://dhupla.hu/page/kreativ/>) található kapcsolati hálók ábrázolása, családfa-ábra, interaktív tartalmak (pl. virtuális séta), nyelvtechnológiai elemzések vizualizációja (pl. szófelhő), stilometriai vizsgálatok.

Legtöbb esetben a vizualizációk a dHUpla tartalmaira épülnek, de akadnak kivételek is, mint például az ukrán háborúról szóló online magyar sajtó szóhasználatának vizsgálata különböző módokon (<https://dhupla.hu/page/kreativ/ukrajna-hirek-szokeszlet>¹⁵).

Mindezek a kreatív tartalmak arra hivatottak, hogy a közzétett magyar kulturális kincseket a felhasználók, kutatók egészen újszerű perspektívából tudják megközelíteni a források digitális közegbe áthelyezett új élete során.

15 Lásd még: *Magyar hírportálok orosz-ukrán háborús tartalmainak interaktív vizualizációja*, dHUpla, hozzáférés: 2022. 09. 19., <https://dhupla.hu/page/kreativ/ukrajna-hirek-szokeszlet-interaktiv>; Általános statisztikák, dHUpla, <https://dhupla.hu/page/kreativ/ukrajna-hirek-szokeszlet-diagram>.