

**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 repositóriumra 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 rendeleltre.....	129
Varga Emese	
Digitális szövegszerkesztés a dHUpla 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

Irodalmi levélváltások nyomában TEI-vel és térképpel

Tracking literary letters on the map using TEI

Mihály Eszter

*MNM KK Országos Széchényi Könyvtár,
Digitális Bölcsészeti Központ (OSZK DBK)*
mihaly.eszter@oszk.hu

Micsik András, Nagy Kadosa

*HUN-REN Számítástechnikai és Automatizálási Kutatóintézet,
Elosztott Rendszerek Osztály (SZTAKI DSD)*
micsik@sztaki.hu, nagy.kadosa@sztaki.hu

Absztrakt

Egy digitális szövegkiadás készítése során szükségszerűen keletkeznek különböző típusú adatok. Az előállt adatok közötti összefüggések feltárása, illetve azok láthatóvá, értelmezhetővé és kutathatóvá tétele ugyanúgy hozzátartozik a digitális bölcsészeti munkához, mint a szövegek gondozása. A tanulmány fókuszában ezúttal a digitális levelezéskiadásokhoz fejlesztett térképes vizualizáció áll, amely jól illusztrálja, hogyan lehet egy strukturált adathalmazt átláthatóbbá tenni, akár több adathalmazt összekötni, s mit tud megmutatni a vizualizáció, amit az adatok maguk nem tudnak.

Egy szerzői levelezés kiadása során a legalapvetőbb metaadatok a levél szerzője és írója, a címzett, a megírás, a feladás és átvétel helye, illetve ideje. Ezek az adatok a források feldolgozása, valamint az ahhoz kapcsolódó kutatói munka során keletkeznek, s aprólékos adat tisztító, majd adatstrukturáló munkafázison kell átesniük ahhoz, hogy egy vizualizáció bázisát képezzék. Térképre helyezve e metaadatok mögötti dinamizmus az ún. levelezési irányokban fogható meg, amely a személyes kapcsolatok mellett a földrajzi helyek és dátumok viszonylatában élővé teszi a szerzők kommunikációjának idő- és térbeli vonatkozásait. Az egymással kortárs szerzői hagyatékokban lapuló levelezések metaadatai egyetlen közös vizualizációban összekapcsolva pedig olyan összefüggéseket hozhatnak felszínre, amelyek eddig teljességgel rejtve voltak az önmagukban álló adatok mögött.

Kulcsszavak: vizualizáció, térkép, levelezéskiadás, TEI XML

Abstract

The Digital Humanities Centre (DBK) of the Hungarian National Library has developed a platform called dHUpla (Digital Humanities Platform - dhupla.hu) for publishing digitized text editions. This platform contains a growing collection of literary documents in TEI XML format, including letters. Based on this data, SZTAKI (Institute for Computer Science and Control) developed a mashup to show the routes of these letters on the map. The environment contains a set of scripts to extract place name tags from TEI XML and map them to geographical namespaces and geocoordinates. The visualization displays the locations of creation, posting and delivery for the letters with single points, and the routes of the letters with arrows. Filtering of letters is possible by collection, location, type of location and date of creation. The interface is responsive so it is usable on tablets and

mobile phones as well. It is easy to create dedicated configurations to focus on collections or eras and thus to serve educational purposes, exhibitions and scholarly needs.

Keywords: visualization, geodata, digital scholarly editions, TEI XML

Bevezetés

Az Országos Széchényi Könyvtár Digitális Bölcsészeti Központ (DBK) egyik elsődleges feladata egy korszerű online platform fejlesztése a közgyűjteményekben őrzött szöveges források kezelésére, amely egységes kutatói környezetet jelent az irodalomtudomány, a nyelvtudomány, és más humán tudományok számára. A dHUpla¹ (Digital Humanities Platform) [1] 2021 óta üzemel (2021 decemberéig a Petőfi Irodalmi Múzeum, majd az Országos Széchényi Könyvtár szolgáltatásaként), és számos szövegkiadást tesz nyilvánosan elérhetővé, amelyhez ún. entitástár, valamint kreatív tartalmak, vizualizációk is társulnak.

Az egyes vizualizációk egészen különböző céllal, így eltérő módszerekkel, eszközökkel készültek. Az alább bemutatott térképes vizualizáció olyan átfogó szemlélettel közelít az irodalmi szövegekhez, azon belül is a szerzői levelezésekhez, amely az érdeklődő felhasználók, illetve a szakértő kutatók számára egyaránt újdonságot nyújthat.

A célkitűzésekről

A levelezések az emberi kapcsolatok lenyomatai. Anélkül, hogy belemennénk a levél mint műfaj jellemzőinek tárgyalásába, néhány sajátosságát mégis ki kell emelnünk: egyrészt ezekből a szöveges alkotásokból biztosan nem zárható ki a szerző, szerves részüket képezi, ún. egodokumentumként kell kezelni, természetesen a folyamatos kritikai szemléletet is fenntartva, illetve a fikció-nonfikció spektrumát is figyelembe véve. Ugyanígy szorosan hozzátartozik egy levél szövegéhez, hogy ki a címzettje, a megszólított személy gyakorlatilag szintén a szöveg alakítójává válik. A levél megírásának, feladásának és fogadásának helyszínei folytonos mozgásban tartják a szöveget, ugyanezen cselekedetek különböző időpontja pedig dinamizmussal, saját történettel tölti fel őket.

Ezek a dimenziók a levelek olvasása közben csak kevésbé láthatóak, érzékelhetőek. Az időbeli, térbeli, illetve relációk szintjén megfogható rétegek viszont rendkívül jól megjeleníthetőek, sőt egymásra vetíthetőek egy vizualizáció formájában. Nem beszélve arról, hogy a különböző hagyatékokban őrzött levelezések adatai egy felületen összeköthetőek, s így az egohálózatokból kapcsolati hálók keletkeznek.

A vizualizáció nem szemfényvesztés, hanem segítségével egyrészt olyan mintázatok, összefüggések, valóság rétegek válnak láthatóvá, amelyeket az önmagukban álló száraz adatok nem tudnak felfedni, másrészt tükrözni képes a levélváltások sajátos dinamikus jellegét is.

A térképes megjelenítés emellett számos szűrési, keresési opciót kínál fel a szövegtörzsekben és a bennük lévő adatokban egyaránt, amely a kutatók számára is új utakat nyit meg.

1 dhupla.hu

A szolgáltatás elsődleges célja nem feltétlenül az, hogy válaszokat adjon feltett kérdésekre (bár sokszor gyors válaszokat is adhat), hanem a digitális bölcsészet legfőbb törekvését követve lehetőséget teremt új szemléletű kérdések feltevésére, sőt ő maga is új megválaszolható kérdéseket generál.

Az adatok kinyerése

Tavalyi előadásunkban részletesen ismertettük azt a környezetet, amelyben a TEI XML² formátumú szövegek készülnek [2]. Ebben a környezetben lehetőség van a földrajzi- és személynevek névtérhez kapcsolására, akár több névteret is felhasználva. Továbbá a TEI *nymRef*³ attribútumában az egységes névforma rögzíthető, amely alapján a település vagy személy a névtérben könnyebben megtalálható.

Első lépésként kinyerjük a levelek megírásának, postára adásának és kézbesítésének adatait, illetve további általános metaadatokat a megjelenítés céljára. Itt még a kinyert lista nem tartalmaz földrajzi koordinátákat, ezt a következő lépésben adjuk hozzá. Amennyiben a földrajzi név rendelkezik névtér-azonosítóval, akkor az alapján a megfelelő névtérből (Geonames⁴, Nemzeti Névtér⁵) a koordinátákat le tudjuk tölteni. A régebben digitalizált leveleknél előfordul, hogy csak ún. KOHA-azonosítót tartalmaznak (a Petőfi Irodalmi Múzeum egy régebben használt, belső adatbázisából⁶), amelynél gyakran hiányoznak a koordináták. Ezt az esetet ugyanúgy kell kezelnünk, mintha hiányozna a névtér-azonosító.

Amennyiben hiányzik a névtér-azonosító, többlépéses keresési stratégiát kell végrehajtani, amely során kihasználjuk, hogy a Geonames és a Wikidata⁷ tartalmazza a települések magyar neveit is. Ilyen esetekben tehát először a Geonames API segítségével keresünk a helynévre, és a visszakapott találatokból kiválasztjuk a keresett helyet, kinyerjük a koordinátát, illetve a Geonames-azonosítóval bővítjük az XML-ekből kinyert azonosítók listáját. Esetenként nem találjuk a találatok között a keresett helynevet, ilyenkor a Wikidata névterében keresünk szintén név alapján. Amennyiben van találat, az itt megtalálható Geonames-azonosítóval bővítjük a névtér-azonosítót, és ez alapján keresünk a hozzá tartozó névtérben. A helyek géppel történő beazonosítását természetesen manuális ellenőrzési fázis követi a névváltozatok, illetve a történetiség figyelembe vétele miatt. A végleges, névtereken keresztül történő térképen való elhelyezésük ezután történik.

Az egyszerűség kedvéért nem területeket, hanem pontokat jelenítünk meg a térképen. Mivel az egyes névterek esetében ezek a pontok kissé eltérnek, ezért egységesíteni kell az egy településhez tartozó koordinátákat.

2 <https://tei-c.org/>

3 <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.naming.html>

4 <https://www.geonames.org/>

5 <https://magyarnemzetinevter.hu/>

6 Az elavult adatbázis kiváltása nemzetközi és hazai névterekre jelenleg folyamatban van.

7 https://www.wikidata.org/wiki/Wikidata:Main_Page

Az így feldolgozott adatok egy JSON adatstruktúrába kerülnek, amelyet a térképes megjelenítő könnyen be tud tölteni. A feldolgozó szkripteket Pythonban írtuk, és ahol lehetett, az XML-ben megbújó adatokhoz XPath segítségével férünk hozzá.

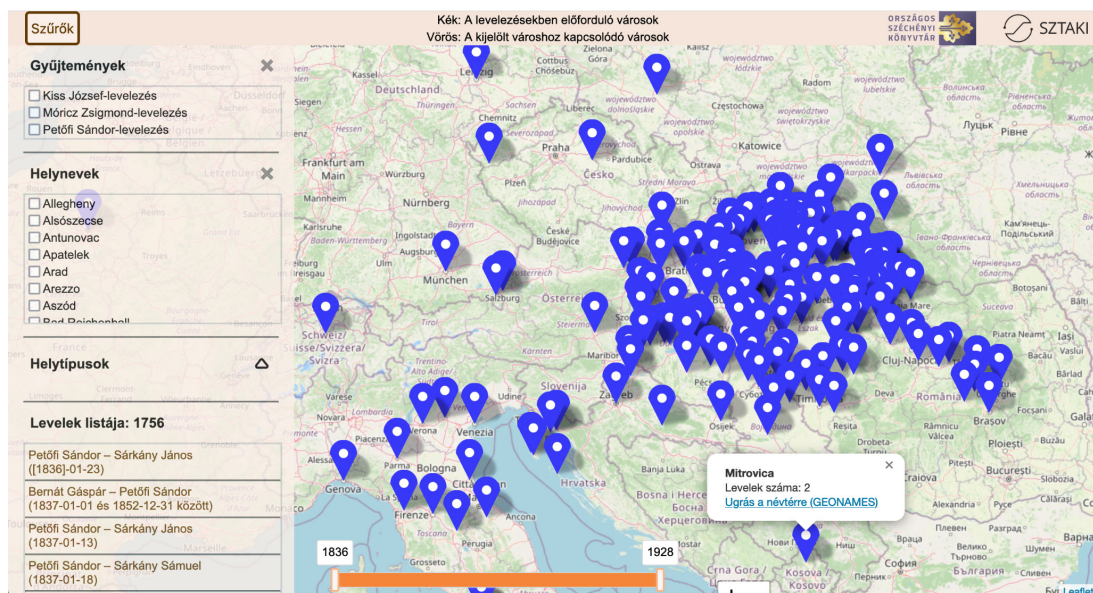
A térképes megjelenítő

A megjelenítő egy hagyományos mashup (1. ábra), amely az OpenStreetMap térképre helyezi fel a levelekkel kapcsolatos helyszíneket a Leaflet szoftver segítségével. A bal oldali sávban elhelyezett szűrőkkel települések, helytípusok (keletkezés, feladás vagy átvétel helye) és gyűjtemények szerint lehet szűrni. A keletkezés évét két csúszkával lehet behatárolni egy idővonalon, ezt kényelmi okokból a középső terület alján helyeztük el. A megszűrt leveleket bal oldalt listázzuk, egy levelet kiválasztva a jobb oldalon megjelennek a levél főbb adatai, és innen tovább tudunk ugrani a levél megtekintéséhez.

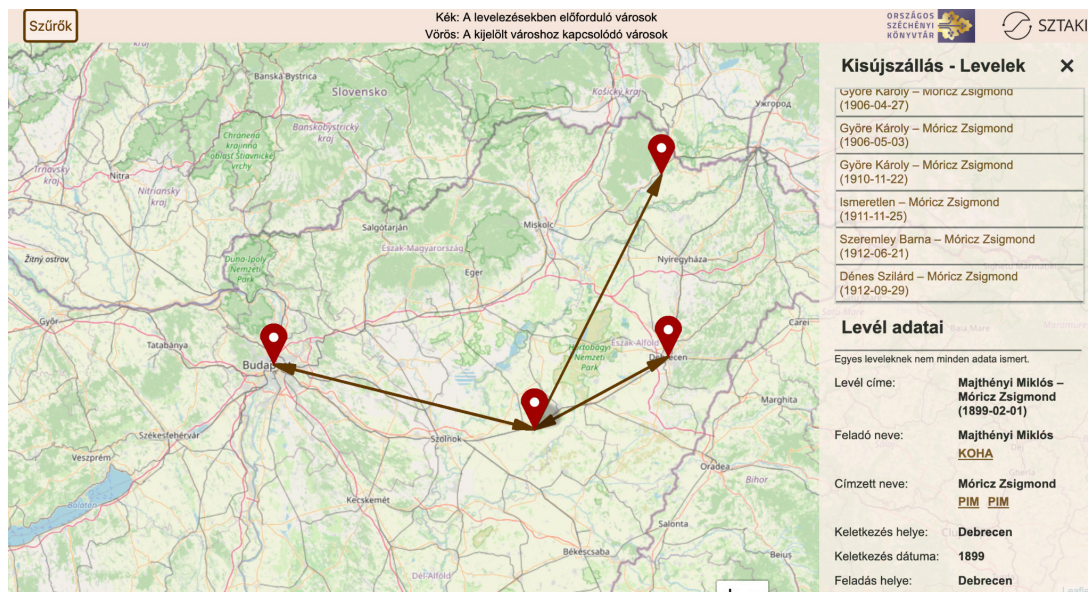
Egy településre kattintva is megkapjuk az oda kapcsolódó levelek listáját, illetve ezen levelek kézbesítési irányait nyilak mutatják a térképen (2. ábra). Sajnos a jelenleg látható gyűjteményekben nagyon sok esetben nem ismert a feladás vagy az átvétel helye és dátuma, mivel a levelet vagy fel sem adták, vagy a boríték elveszett.

A megjelenítő dizájnja reszponzív, vagyis különféle felbontásokon is működik, bár egy kis méretű mobiltelefonon a használata nem kényelmes. A Szűrők panel bezárásával a teljes képernyőt a térkép tölti ki, viszont a kiválasztott településhez kapcsolódó levelek felugró listája eltakarhatja a térkép nagy részét.

Természetesen több ezer levél adatainak vizualizációja számítási és használhatósági problémákat eredményez, azonban a gyűjtemények betöltése konfigurálható, így könnyen kialakíthatók gyűjteményekre vagy korszakokra fókuszált kisebb egyedi telepítések, amelyek tematikus weboldalak és kiállítások számára egyaránt jól használhatók.



1. ábra



2. ábra

Összefoglalás

Azt gondoljuk, hogy a bemutatott térképes vizualizációs eszköz segíthet a kutatóknak a levelezések átfogó kutatásában, összefüggések, kapcsolatok, rejtett mintázatok feltárásában, illetve akár egyes levelek vagy háttérinformációk gyors megkeresésében is. Hasonlóan fontos szerepe lehet a szoftvernek az oktatásban, az online vagy valós kiállítások vonzóvá tételében, a műkedvelők érdeklődésének kielégítésében.

A szoftverkörnyezet folyamatos fejlesztés alatt áll, az idő szerinti szűrést a Networkshop-on elhangzott előadás óta fejlesztettük. Az aktuális állapot kipróbálható a <https://dbk.dsd.sztaki.hu/terkep/> címen.

További terveink között szerepel a TEI XML feldolgozási folyamat teljes automatizálása, a dHUplán publikált gyűjtemények és a térkép folyamatos és automatikus összehangolása, illetve a felhasználói élmény folyamatos javítása. Következő fejlesztésként pedig szeretnénk megvalósítani a napló-jellegű művekben szereplő helyszínek vizualizálását, amelynek segítségével a szövegek mellett egy életút és annak eseményei válnak követhetővé.

Irodalomjegyzék

- [1] Mihály Eszter (2022) *Mi az a dHUplá?: A Digitális Bölcsészeti Platform bemutatása*. In: Valós térben – Az online térért, Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem. Kiadja a HUNGARNET Egyesület az MTA Könyvtár és Információs Központ közreműködésével, Budapest, pp. 345–358. ISBN 978-615-82243-0-7 <https://doi.org/10.31915/NWS.2022.44>
- [2] Mihály Eszter, Micsik András (2023) *Szerkesztői környezet TEI-alapú szövegkiadásokhoz*. In: Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé (pp. 186–191). <https://doi.org/10.31915/NWS.2023.27>