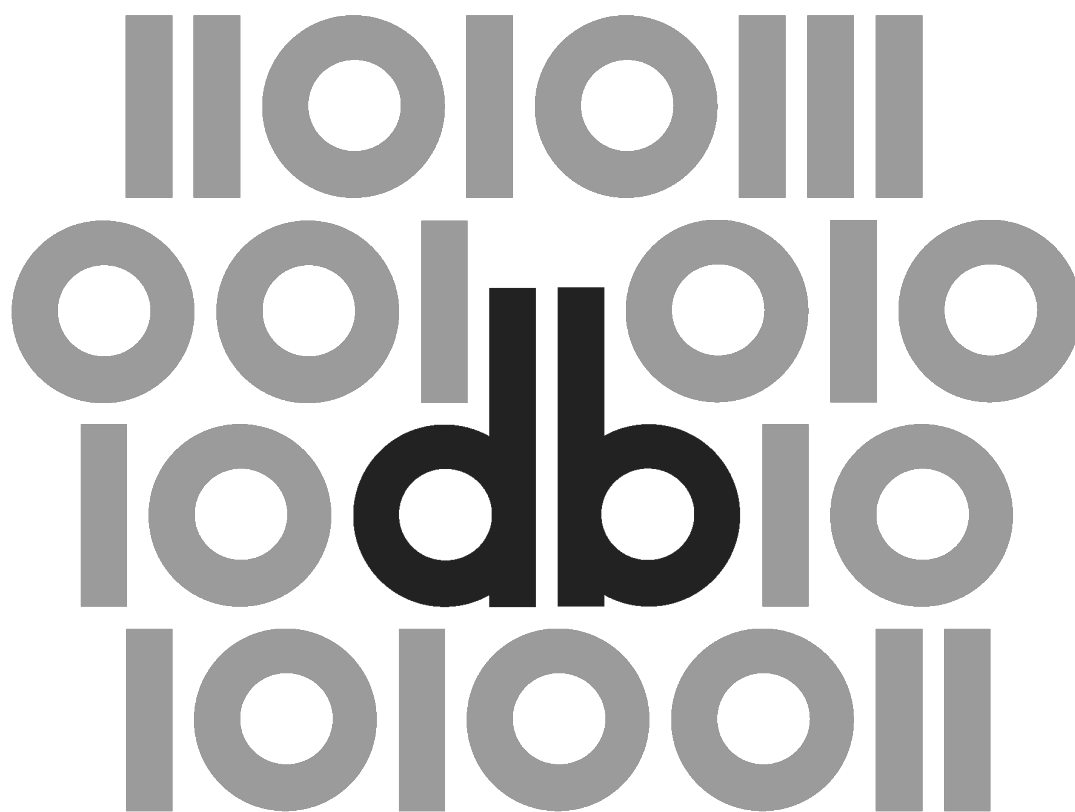


2018.01.

<DIGITÁLIS BÖLCSÉSZET>



2018.01.

</DIGITÁLIS BÖLCSÉSZET>

Digitális Bölcsészet
2018., első szám

<DIGITÁLIS BÖLCSÉSZET>



2018.01.

Felelős szerkesztő:

Maróthy Szilvia

Szerkesztőbizottság:

Bartók István, Fazekas István, Golden Dániel,
Horváth Iván (a bizottság elnöke), †Orlovszky Géza,
Palkó Gábor, Pap Balázs, Sass Bálint, Seláf Levente

Szerkesztőség:

Almási Zsolt, Fodor János, Kokas Károly, †Labádi Gergely,
Parádi Andrea

Rovatvezetők:

Tanulmányok: Kiss Margit

Műhely: Péter Róbert

Kritika: Bartók Zsófia Ágnes

ISSN 2630-9696

DOI 10.31400/dh-hun.2018.1

Kiadja az ELTE BTK Régi Magyar Irodalom Tanszéke, 1088 Budapest,
Múzeum krt. 4/A.

Felelős kiadó az ELTE BTK Régi Magyar Irodalom Tanszék vezetője.

Megjelenik az Open Journal Systems (OJS) v. 3. platformon, melynek
működtetését az ELTE Egyetemi Könyvtár- és Leváltár biztosítja.

Honlap: <http://ojs.elte.hu/index.php/digitalisbolcseszett>

Email cím: dbfolyoirat@gmail.com

Tördelés: Hegedüs Béla

Grafika: Hegyi Gábor

Tartalom

Beköszöntő	7
Prószéky Gábor előszava	9
Andrew Prescott előszava	11
Szerkesztőségi köszöntő	13
Tanulmányok	15
Labádi Gergely	
Az olvasó gép: Berzsenyi Dániel versei távolról	17
Drótos László–Kokas Károly	
Webarchiválás és a történeti kutatások	35
Markó Anita	
Hálózatok a 16–17. századi album amicorumokban: Az 1500 és 1700 közötti hungarika jellegű emlékkönyvbejegyzések hálózatelemzése az <i>Inscriptiones Alborum Amicorum</i> adatbázis alapján	55
Matthew L. Jockers	
Metaadat	83
Műhely	109
Sennyey Pongrácz	
Viták és víziók a digitális bölcsészetről	111
Horváth Iván	
Digitális bölcsészet a virtuális nemzeti könyvtárban	121
Lejtovicz Katalin–Matthias Schlögl–Bernád Ágoston Zénó–Maximilian Kaiser–Peter Alexander Rumpolt	
Digitalizáció és hálózatkutatás: Az <i>Österreichisches Biographisches Lexikon 1815–1950</i> és az APIS-projekt 139	
Cséve Anna–Fellegi Zsófia–Kómár Éva	
Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása Esettanulmány	159
Biszak Sándor–Kokas Károly	
Budapest Időgép	175
Ruttkay Zsófia	
Digitális Múzeum – a MOME TechLab projektjeinek tükrében	185
Dragon Zoltán–Sebestény Csilla	
#BREW: influencer-kísérlet az Instagram újhullámos kávéközösségében 203	

Kritika	217
Matthew James Driscoll and Elena Pierazzo, eds., <i>Digital Scholarly Editing: Theories and Practices</i> (2016) – Maróthy Szilvia	219
Laura Estill, Diane K. Jakacki and Michael Ullyot, eds., <i>Early Modern Studies after the Digital Turn</i> (2016) – Maczelka Csaba	223
Matthew K. Gold, ed., <i>Debates in the Digital Humanities</i> (2012); Matthew K. Gold and Lauren F. Klein, eds., <i>Debates in the Digital Humanities 2016</i> (2016) – Zámóné Kocic Larisa	233
George Bruseker, László Kovács and Franco Niccolucci, eds., „Digital Humanities.” <i>ERCIM News</i> 111 (2017) – Molnár Sándor Gyula	239
In memoriam	243
Szajbély Mihály: Búcsú Labádi Gergelytől	245

Lejtovicz Katalin

Österreichische Akademie der Wissenschaften (ÖAW), Austrian Centre for Digital Humanities (ACDH)

Katalin.Lejtovicz@oeaw.ac.at

Bernád Ágoston Zénó

ÖAW, Institut für Neuzeit- und Zeitgeschichtsforschung (INZ)

Agoston.Bernad@oeaw.ac.at

Matthias Schlögl

ÖAW, ACDH

Matthias.Schloegl@oeaw.ac.at

Maximilian Kaiser

ÖAW, INZ

Maximilian.Kaiser@oeaw.ac.at

Peter Alexander Rumpolt

ÖAW, Institut für Stadt- und Regionalforschung (ISR)

Peter.Rumpolt@oeaw.ac.at

Digitalizáció és hálózat kutatás: Az Österreichisches Biographisches Lexikon 1815–1950 és az APIS-projekt*

Az elmúlt években fejlődésnek induló digitalizálási módszerek lehetővé tették a digitális bölcsészet területén belül többek között az életrajzi lexikonokat feldolgozó projektek előtérbe kerülését. A biográfiai kutatások célja a már digitalizált lexikonok szövegeinek (pl. *Österreichisches Biographisches Lexikon 1815–1950* – a továbbiakban *ÖBL* – részben strukturált, részben strukturálatlan, szöveges formában tárolt adatainak) átalakítása számítógép által feldolgozható formátumra, illetve az átalakított adatokon hálózati, statisztikai és számítógépes nyelvészeti kutatások elvégzése. Ebben a kontextusban kezdődtek el 2015-ben a kutatások az Osztrák Tudományos Akadémia Új- és Jelenkortörténeti Kutatási Intézetében [Institut für Neuzeit- und Zeitgeschichtsforschung] a Mapping historical networks: Building the new Austrian Prosopographical, Biographical Information

* A tanulmány az Osztrák Tudományos Akadémia Új- és Jelenkortörténeti Kutatási Intézetében, az osztrák Nationalstiftung für Forschung, Technologie und Entwicklung által támogatott Mapping historical networks: Building the new Austrian Prosopographical, Biographical Information System (APIS) című projekt keretében készült.

System (APIS) projekt keretén belül. A projekt az *ÖBL* digitális feldolgozását és egy kutatási célokra kifejlesztett számítógépes környezet létrehozását célozza meg, mely a szintén az Osztrák Tudományos Akadémia keretében működő Austrian Centre for Digital Humanities (ACDH) [Osztrák Digitális Bölcsészeti Központ] valamint az Institut für Stadt- und Regionalforschung (ISR) [Város- és Régiókutatási Intézet] együttműködésével valósul meg. Cikkünkben bemutatjuk, hogy az APIS-projekt technikai megvalósítására fejlesztett webalkalmazás milyen technikai és nyelvtechnológiai megoldásokat használ a lexikon számítógépes feldolgozására és az adatok reprezentálására.

Kulcsszavak:

életrajzi lexikon, információ kinyerés, entitás kapcsolat, történeti hálózatkutatás



1. Bevezetés

Az Osztrák Tudományos Akadémia Új- és Jelenkortörténeti Kutatási Intézetében 2015 óta zajlanak a Mapping historical networks: Building the new Austrian Prosopographical, Biographical Information System (APIS) című projekt keretében végzett kutatások. A projekt célja az *Österreichisches Biographisches Lexikon 1815–1950 (ÖBL)*¹ című életrajzi lexikon új, digitális rendszerben való feldolgozása, mely a szintén az Osztrák Tudományos Akadémia keretében működő Osztrák Digitális Bölcsészeti Központ valamint a Város- és Régiókutatási Intézet együttműködésével valósul meg. A modern információtechnológiai környezetnek és a számítógépes nyelvészeti adatfeldolgozásnak köszönhetően az *ÖBL* egyrészt a szemantikus web Giant Global Graph csomópontjává válhat,² ugyanakkor pedig a kulturális ismeretanyag tárolójaként is funkcionálhat, amely új kérdésselvetéseket és megoldási javaslatokat tehet lehetővé a szellem-, társadalom- és kultúratudományok területén, különösen pedig a migráció- és a történeti hálózatkutatások körében. Jelen tanulmány elsőként az *ÖBL* (digitalizálási) történetét vázolja fel, majd bemutatja az APIS-adatbázis jelenlegi változatát, és rávilágít a számítógépes nyelvészeti eljárásokra, végezetül pedig betekintést enged néhány történeti jellegű kutatásba is.

2. Az *ÖBL* története és digitalizálási fázisai

Az *ÖBL* kezdetei a 20. század elejéig nyúlnak vissza. 1916-ban kezdeményezte az irodalomtörténész Anton Bettelheim (1851–1930) és a történész Oswald Redlich (1858–

¹ *Österreichisches Biographisches Lexikon 1815–1950*. Lfg. 68, Bd. 14 (Wien: Böhlau Verlag, Verlag der Österreichischen Akademie der Wissenschaften, 1954ff.). *Österreichisches Biographisches Lexikon 1815–1950 Online Edition*, 2003–2017, hozzáférés: 2017.12.14, <http://www.biographien.ac.at/>.

² Tim Berners-Lee, James Hendler and Ora Lassila, „The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities,” *Scientific American*, 2001. máj. 17., 34–43; Dave Raggett, „W3C Data Activity: Building the Web of Data,” hozzáférés: 2017.12.14, <http://www.w3.org/2013/data/>.

1944) a Constant von Wurzbach által szerkesztett, 1856 és 1891 között megjelenő *Biographisches Lexikon des Kaiserthums Österreich*³ című életrajzi lexikon folytatását és átdolgozását. A tervet a történelmi helyzet miatt csak részben lehetett megvalósítani,⁴ az összegyűjtött forrásanyag viszont lehetővé tette, hogy a második világháborút követően elkezdődhessen egy újabb életrajzi szakmunka összeállítása: az Osztrák Állami Levéltár igazgatója, Leo Santifaller (1890–1974) sürgetésének köszönhetően 1946-ban az Osztrák Tudományos Akadémia határozatot hozott az *ÖBL* kidolgozására. A lexikon első füzeté 1954-ben jelent meg.

A lexikonba azon személyek adatai kerülnek be, akik 1815 és 1950 között az egykori osztrák államszerkezetben éltek vagy ott hunytak el, és a közélet valamelyik területén jelentőset alkottak. Ezáltal az *ÖBL* a határokon átívelő nemzeti biográfiai munkák paradigmaticus példája lesz, hiszen adatait nem szűkíti a mai Ausztriára, hanem összefogja az egykori császári birodalom területeit, illetve az Osztrák–Magyar Monarchiában tevékenykedők adatait. A lexikon ily módon a középeurópai kultúra leképezőjeként is funkcionál, mely átfogóan mutathatja be e régió kultúrájának 19. és 20. századi változásait. A nyomtatott változat jelenleg 14 kötetből áll (68 füzet), A és Vo betűk között több mint 18 000 életrajzot közöl. A mű magyar vonatkozásaival kapcsolatban kiemelendő, hogy a kötetek több mint 2500 olyan személy adatait tartalmazzák, akik az egykori Magyar Királyság vagy Erdély területén születtek. Ezáltal az *ÖBL* a magyar és erdélyi történelmi kutatások egyik legfontosabb német nyelvű biográfiai forrásmunkájaként konstituálódik.

A lexikon, illetve a szerkesztőség által gyűjtött nyilvántartólapok a digitalizálás több fázisán vannak túl. Az 1990-es években az addig kartotékokon rögzített adatokat bevezették egy Microsoft Access adatbázisba. Ezt a 2000-es évek elején egy OracleTM alapú relációs adatbázis váltotta fel, amely 450 000 rekordot tartalmaz, amelyből több mint 80 000 életrajzi adatnak tekinthető.⁵ Ennek a továbbra is rendelkezésre álló, platformfüggetlen integrációt megvalósító digitális rendszernek köszönhetően jött létre az *ÖBL* online kiadása 2009-ben azáltal, hogy mind a strukturált, mind pedig a nem strukturált személyi adatok átvételét és XML-formátumba való átmentését lehetővé tette. Az XML-formátum rögzített formában tartalmazza a nevet, születési és elhalálozási adatokat (a születési és elhalálozási helyet is), foglalkozást, műveket

³ Constant von Wurzbach, Hg., *Biographisches Lexikon des Kaiserthums Oesterreich, enthaltend die Lebensskizzen der denkwürdigen Personen, welche seit 1750 in den österreichischen Kronländern geboren wurden oder darin gelebt und gewirkt haben*, Bd. I–LX (Wien: L. C. Zamarski, k. k. Hof- und Staatsdruckerei, 1856–1891).

⁴ Anton Bettelheim, Begr., *Neue Österreichische Biographie 1815–1918, Abteilung 1: Biographien*, Bd. I–VIII (Wien–Zürich: Wiener Drucke–Amalthea Verlag, 1923–1935); Anton Bettelheim, Begr., *Neue Österreichische Biographie ab 1815: Große Österreicher*, Bd. IX–XXII (Wien–München: Amalthea Verlag, 1957–1987); Hans Bohatta, Jaroslav Sutnar und Julius Stockinger, Hg., *Neue Österreichische Biographie 1815–1918, Abteilung 2: Bibliographie zur Neuen Österreichischen Biographie*, Bd. I (Zürich–Wien–Leipzig: Amalthea Verlag, 1925).

⁵ Christine Gruber und Roland Feigl, „Von der Karteikarte zum biografischen Informationsmanagementsystem: Neue Wege am Institut Österreichisches Biographisches Lexikon und biographische Dokumentation,” in *Biografische Lexika im Internet: Internationale Tagung der „Sächsischen Biografie” in Dresden (30. und 31. Mai 2008)*, Hg. Martina Schattkowsky und Frank Metasch, Bausteine aus dem Institut für Sächsische Geschichte und Volkskunde 14 (Dresden: Thelem Universitätsverlag, 2009), 55–75.

és vonatkozó irodalmat. A legtöbb információ azonban az életrajzi szócikkek nem strukturált folyó szövegében található.

3. Az APIS-projekt technikai aspektusai

Az új digitalizálási lehetőségeknek és módszereknek köszönhetően az elmúlt években az életrajzi lexikonokat feldolgozó projektek előtérbe kerültek. Ezeket a munkálatokat egy közös cél vezérli: a már digitalizált biográfiai kézikönyvek szövegeinek (melyek az adatokat félig strukturált formában tárolják) átalakítása a gépek által feldolgozható adatokká (pl. Deutsche Biographie,⁶ BiographyNet),⁷ illetve a hálózati kutatások módszereinek alkalmazása ezen adatok értelmezésében.⁸ Ebben a kontextusban kezdődtek el az ÖBL-ben található biográfiák számítógépes nyelvészeti feldolgozásai is az Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart (Stuttgarteri Egyetem Gépi Nyelvfeldolgozás Intézet) falai között.⁹ A következőkben bemutatjuk, hogy az APIS-projekt¹⁰ technikai megvalósítására fejlesztett webalkalmazás milyen technikai és nyelvtechnológiai megoldásokat használ a lexikon számítógépes feldolgozására és az adatok reprezentálására.

3.1. Az adatmodell

Az ÖBL-ben található adatok leírására egy saját adatmodellt fejlesztettünk ki. Már létező modellek felhasználását, mint például a CIDOC-CRM¹¹ elvetettük, mert ezek jellemzően generikusak, a digitális bölcsészet témakörébe tartozó sokféle adat leírására lettek kifejlesztve. Ha egy ilyen, már létező, generikus adatmodellt használtunk volna az ÖBL adatainak jellemzésére, túl komplex reprezentációhoz jutunk, ahol az olyan egyszerű események, mint például születés, házasságkötés, elhalálozás többszörösen elágazó struktúrákat eredményez (legalább négy csomópont CIDOC-CRM-ben). Az

⁶ Deutsche Biographie, hozzáférés: 2017.12.14, <https://www.deutsche-biographie.de/>.

⁷ BiographyNet, hozzáférés: 2017.12.14, <http://www.biographynet.nl/>.

⁸ Sophia Stotz, Valentina Stuß, Matthias Reinert and Maximilian Schrott, „Interpersonal Relations in Biographical Dictionaries: A Case Study,” in *BD2015: Biographical Data in a Digital World 2015, CEUR Workshop Proceedings*, vol. 1399, eds. Serge ter Braake, Antske Fokkens, Ronald Sluijter, Thierry Declerck and Eveline Wandl-Vogt (Amsterdam: CEUR-WS.org, 2015), 74–80, <http://ceur-ws.org/Vol-1399/paper12.pdf>; Antske Fokkens, Serge ter Braake, Niels Ockeloen, Piek Vossen, Susan Legêne, Guus Schreiber and Victor de Boer, „BiographyNet: Extracting Relations Between People and Events,” in *Europa baut auf Biographien: Aspekte, Bausteine, Normen und Standards für eine europäische Biographik*, Hg., Ágoston Zénó Bernád, Christine Gruber und Maximilian Kaiser (Wien: New Academic Press, 2018), 193–224.

⁹ André Blessing, Andrea Glaser and Jonas Kuhn, „Biographical Data Exploration As a Test-bed for a Multi-view, Multi-method Approach in the Digital Humanities,” in *BD2015*, 53–60, <http://ceur-ws.org/Vol-1399/paper9.pdf>; Jonas Kuhn und André Blessing, „Die Exploration biographischer Textsammlungen mit computerlinguistischen Werkzeugen: Methodische Überlegungen zur Übertragung komplexer Analyseketten in den Digital Humanities,” in *Europa baut auf Biographien*, 225–257.

¹⁰ APIS (GitHub), hozzáférés: 2018.01.09, <https://github.com/acdh-oeaw/apis-core>.

¹¹ CIDOC-CRM: Conceptual Reference Model, hozzáférés: 2018.01.09, <http://www.cidoc-crm.org/>.

ehhez hasonló, komplex modellek megnehezítik egy webalkalmazás fejlesztését, karbantartását és bővítését. A mi esetünkben elegendő az osztályoknak, az osztályokat jellemző tulajdonságoknak és az osztályokat összekötő relációknak egy részhalmaza az életrajzok leírására. Az APIS-projektben öt entitást definiáltunk: Személy, Hely, Intézmény, Esemény, Mű. Ezek az osztályok tovább bonthatók, tipizálhatók a felhasználó által definiált ontológiákkal, továbbá az osztályok közötti kapcsolatok típusa is leírható hierarchikus kontrollált szótárakkal.

Művek esetében például a felhasználó megadhatja az adott mű típusát: irodalmi, képzőművészeti vagy tudományos stb. mű, ezeken belül az ontológia által definiált további kategóriákat rendelhet az entitásokhoz, például: festmény, szobor vagy fénykép stb. A tipizálás pontosságát a definiált szótár leíróképessége határozza meg. Az entitások mellett a relációk is elláthatóak típussal, melyek szintén a felhasználó által definiált struktúrából választhatóak ki. A személy–személy osztályok közötti kapcsolatok elláthatóak *házastárs*, *szülő*, *tanár* stb. típussal, a személy–intézmény kapcsolatok a *tanult*, *dolgozott* stb. típusokkal. Az APIS-projektben az osztályok közötti kapcsolatok leírására használt típusok mindig kétirányúak, ezt tükrözik a relációk elnevezései is: *parent of*, *child of* („szülő”, „gyerek”), *works in*, *has employee* („dolgozik valahol”, „munkavállaló”).

3.2. Az APIS-webalkalmazás

Az APIS-alkalmazás¹² segítséget nyújt a kutatóknak nagy mennyiségű adathalmaz szisztematikus és félautomatikus feldolgozásában, az adatok elemzésében, és az ennek során megtalált összefüggések megjelenítésében. Az adatok megjelenítése lehetővé teszi, hogy az egyébként strukturálatlan adatokban nehezen átlátható összefüggések felszínre kerüljenek, továbbá, hogy a kutatók könnyebben és intuitívabban dolgozhasanak nagy mennyiségű adatokkal a felhasználóbarát megjelenítésnek köszönhetően.

Az APIS-projekt keretében egy széles körben alkalmazható VRE-t (Virtual Research Environment, Virtuális Kutatói Környezet) fejlesztettünk ki, amely az indulás óta számos kutatási területen és több intézetben is felhasználásra került. Az APIS egy *Django*¹³ és *MySQL*¹⁴ frameworkre épülő webalkalmazás, amely nemcsak felhasználóbarát felhasználói felülettel, hanem egy RESTful API-val is rendelkezik, lehetővé téve, hogy a felhasználók ne csak a felhasználói felületen keresztül, hanem programkódból is hozzáférhessenek az adatokhoz. Az adatmodell által leírt információ a webalkalmazásban kerül megjelenítésre. Az alkalmazás lehetővé teszi bejegyzések egyszerű létrehozását, módosítását és törlését. A felület három fő komponensből áll: az első, az adatok különböző kritériumok szerint szűrhető listája, a második az entitások tulajdonságait megjelenítő nézet, a harmadik pedig a vizualizációs és fájlletöltés funkciók weboldala. A felhasználók az adatokat a részletes leírást tartalmazó nézetben tudják szerkeszteni; az oldalt úgy terveztük meg, hogy a felhasználót

¹² Matthias Schlögl und Katalin Lejtovicz, „APIS – Eine Linked Open Data basierte Datamining-Webapplikation für das Auswerten biographischer Daten,” in *DHd 2017 Bern: Digitale Nachhaltigkeit: Konferenzabstracts*, eds. Michael Stolz et al. (Bern: Universität Bern, 2017), 220–222.

¹³ Django Software Foundation, *Django*, hozzáférés: 2018.01.04, <https://www.djangoproject.com/>.

¹⁴ *MySQL*, hozzáférés: 2018.01.04, <https://www.mysql.com/>.

támogassa az adatok szerkesztésében. A bal oldalon található a kiválasztott bejegyzés metaadatai (pl. születési hely, idő), jobb oldalon pedig az aktuális entitás összeköthető más bejegyzésekkel (pl. egy személy összeköthető az intézménnyel, ahol tanult).

Teleki von Szék, László d. J. Gf.

Texts

Metadata Person

Name

Start

Please enter a date (DD).(MM).YYYY

End

Please enter a date (DD).(MM).YYYY

Status

First name

The persons's forename. In case of more then one name...

Profession
 ✖ Politik ✖

Title

Gender

MetaInfo

related to

Event

Institution

Delete	Start Date	End Date	relation type	Related Institution	Edit
✖	01.01.1836	01.01.1844	war Mitglied >> war k. Mitglied	Ungarische Akademie der Wissenschaften	Edit
✖	01.01.1844	—	war Mitglied >> war Ehrenmitglied	Ungarische Akademie der Wissenschaften	Edit
✖	—	—	war Mitarbeiter >> war Praktikant	Ungarische Statthaltere	Edit
✖	01.01.1828	01.01.1830	war in Ausbildung >> war Student	Reformiertes Kollegium Sárospatak	Edit
✖	—	—	war Mitarbeiter >> war Praktikant	Ungarische Hofkanzlei	Edit
✖	—	—	war Mitglied >> war Vizepräsident	Ungarischer Schutzverein (Védegylet)	Edit
✖	—	—	war Mitglied >> war Präsident	Pester Kreis (Pesti Kör)	Edit
✖	—	—	war Mitglied >> war Präsident	Oppositionskreis (Ellenzéki Kör)	Edit
✖	—	—	war Mitglied >> war Abgeordneter	Ungarischer Landtag	Edit
✖	—	—	war Gründer	Radicalkreis (Radical Kör)	Edit

[Weiter](#) Seite 1 von 2

Relation type*

Institution*

1. ábra. Az életrajzban bemutatott személy metaadatait és kapcsolatait megjelenítő webes felület az APIS-alkalmazásban

Az ilyen entitás–entitás kapcsolatok egyszerű létrehozását a webes felület *auto-complete* (automatikus kiegészítés) funkcióval támogatja, továbbá a földrajzi helyek (falvak, városok, országok) egy térképen jelennek meg, ezzel segítve a felhasználót a megfelelő hely kiválasztásában.

Az APIS-alkalmazásban nemcsak webes felületet lehet használni az adatok bevitelére, hanem lehetőség van annotációk megadására a szabad szövegben is. A felhasználó az egér segítségével kijelöli az annotálni kívánt szövegrészt, majd a felugró contextmenüből kiválasztja a reláció típusát (pl. *utazott valahova*). A kiválasztott reláció mindig az életrajzban szereplő személyt köti össze egy másik entitással. A reláció kiválasztásakor a másik entitás típusának megfelelő webes felület automatikusan felugrik, a felhasználó ebben adja meg – automatikus kiegészítés segítségével – a kijelölt szövegrészhez tartozó adatbázis bejegyzést (ebben az esetben pl. egy város vagy ország nevét). A megjelölt reláció az adatbázisban elmentésre kerül, továbbá a szövegben

szinkóddal emeljük ki az annotációt a könnyebb kereshetőség és módosíthatóság érdekében.

Default
All selected (17)
All selected (20)
Update

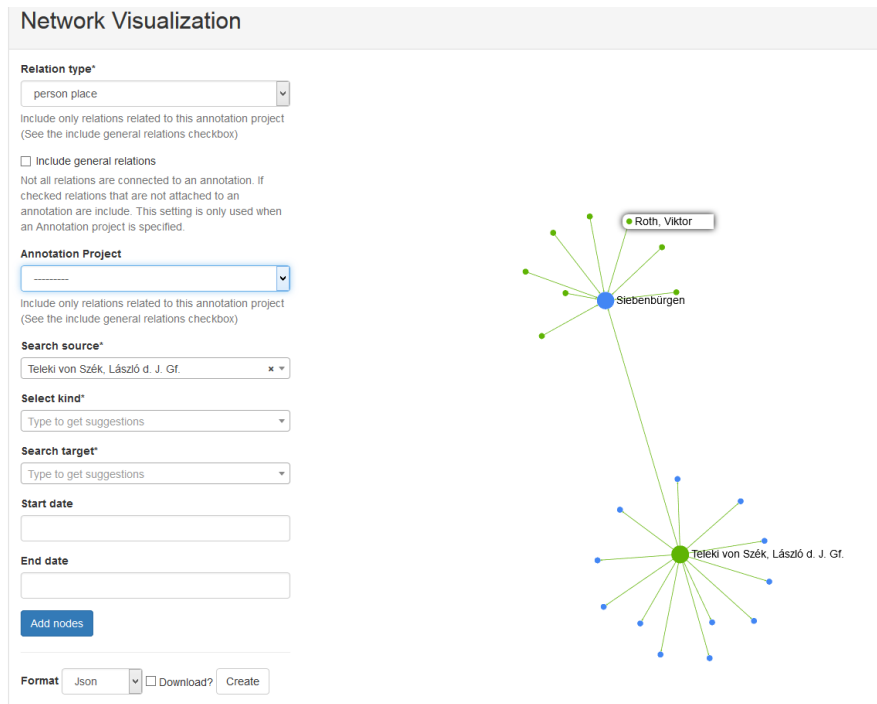
Sohn von → [László Gf. T. v. S. d. Á.](#), Halbbruder von → [József Gf. T. v. S. d. J.](#) – Nach privatem Unterricht absolv. T. 1828–30 Jus am [ref. Kollegium von Sárospatak](#). Darauf folgte ein kurzes Praktikum bei der [Statthaltereirei in Ofen \(Buda\)](#), dann in [Wien](#) bei der [Ung. Hofkanzlei](#). Stud.reisen führten ihn 1833–36 nach [Dtd.](#), in die [Niederlande](#), nach [England](#) und [Frankreich](#). Nach seiner Rückkehr wandte sich T. der Politik zu und wurde eines der aktivsten Mitgl. der liberalen Reformopposition in [Siebenbürgen](#) und [Ungarn](#). Sein literar. Erstlingswerk war das Drama „Kegyenc“, 1841. In den 1840er-Jahren war T. eine der führenden und gleichzeitig radikalsten Figuren der hochadeligen Opposition. 1844 wurde er zum Vizepräs. des [Schutzver.](#), 1845 zum Präs. des [Pester Kreises](#) gewählt, der ersten Organisation der liberalen Opposition. 1847 wurde er Präs. des [Oppositionskreises \(Elenzékí Kör\)](#). Auf dem letzten ständ. [LT](#) verlangte T. an der Seite von → [Ludwig Gf. Batthyány](#) die sofortige Inkraftsetzung des Reformprogramms. Während der Revolution schuf er aus dem Oppositionskreis den „[Radicalkreis](#)“ und wurde in [Abony](#) als Abg. in den LT gewählt. Im Repräsentantenhaus galt er als geistiger Anführer einer Opposition, die die [Batthyány-Regierung](#) von links kritisierte und entschiedenes Auftreten gegenüber Wien forderte. Obwohl T. als ung. Gesandter in [Paris](#) für sein Land keine diplomat. Anerkennung erreichen konnte, blieb er dennoch dort, um die ung. Interessen in Westeuropa weiterhin entscheidend zu vertreten. T. verstand es, die ausländ. Medien geschickt für die Zwecke der ung. Revolution einzusetzen. Nach Besprechungen mit den Anführern der poln. und rumän. Emigration entwarf er den Plan für ein föderatives Ungarn als Zentrum einer neuen Donau-Konföderation, worin die Nationalitätenkonflikte gelöst werden sollten. Nach der Niederlage der Revolution hielt sich T. ab 1851 größtenteils in der [Schweiz](#) auf. Während des Krieges in Italien gründete er Anfang Mai 1859 gem. mit → [Lajos Kossuth v. Udvard](#) u. [Kossut](#) und → [Georg Klapka](#) die Emigrantenregierung „[Ungarisches Nationaldirektorium](#)“. Im November 1860 reiste er – aus privaten Gründen – illegal nach [Dresden](#). Dabei wurde er von der [sächs. Polizei](#) verhaftet und an [Osterr.](#) ausgeliefert. Nach zehntägiger [Haft](#) wurde er unter der Bedingung freigelassen, sich von polit. Tätigkeiten fern zu halten. Doch schon im April 1861 nahm er erneut als Abg. von Abony am parlamentar. Leben teil, wobei er jedes Zugeständnis gegenüber Osterr. zurückwies und erklärte, die Zustände von 1848 als Ausgangspunkt für weitere Verh. betrachten zu wollen. Im Unterschied zur „[Adresspartei](#)“ → [Franz v. Deáks](#) forderte T., dass dieser Vorschlag dem [Kg.](#) per Beschluss – daher „[Beschlusspartei](#)“ – überbracht würde. T. wurde Anführer der Letzteren, doch schreckten die Mitgl. vor notwendigen Zugeständnissen gegenüber den nichtmagyar. Nationalitäten zurück und so blieb T. mit seiner vergleichsweise konzilianten Haltung innerhalb der Partei allein. Ab 1836 war er [K. M.](#), ab 1844 [Ehrenmitgl.](#) der MTA.

Teleki von Szék László Gf. d. J., Politiker und Schriftsteller. Geb. Pest (Budapest, H), 11. 2. 1811; gest. ebd., 7. (8.) 5. 1861 (Selbstmord); evang.

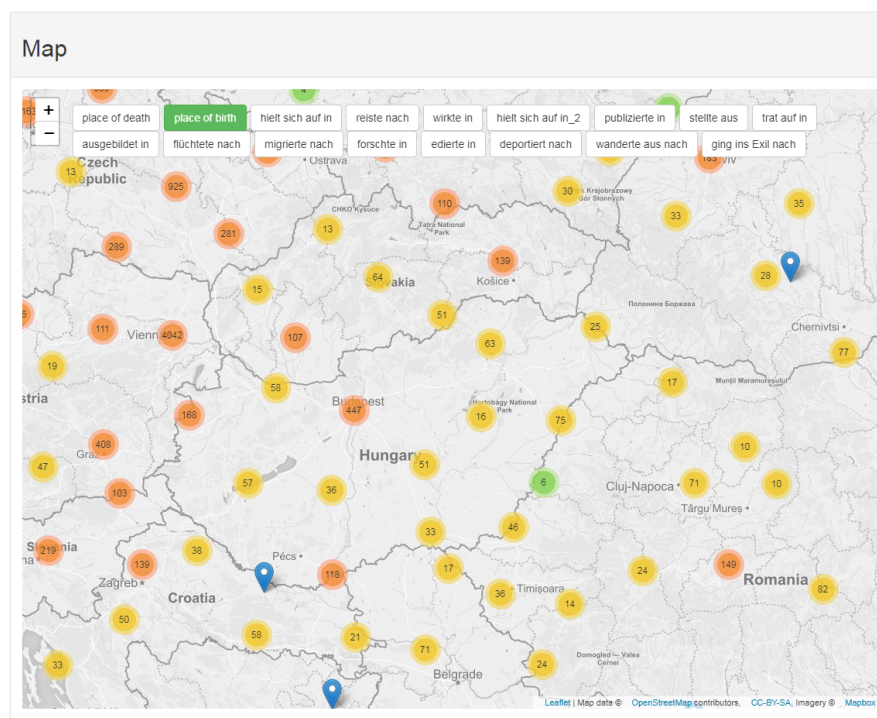
2. ábra. Az életrajzok annotációját támogató webes felület az APIS-alkalmazásban

Az APIS-webalkalmazás fejlesztésekor nemcsak arra törekedtünk, hogy az adatok beírása és szerkesztése egyszerű legyen, hanem hogy a nagy mennyiségű adathalmazban kódolt információt a kutatók gyorsan és könnyen megtalálják, és egy átfogó képet kapjanak róla. Ennek érdekében több vizualizáció is a felhasználó rendelkezésére áll. Az egyik az entitások közötti relációkat bemutató hálózat, melyben a személyek, helyek, intézmények, művek, események a hálózat csomópontjai, ezeket kötik össze a relációkat reprezentáló élek, melyeken minden esetben szerepel a reláció típusa. A gráf személyre szabható, lehetőség van a megjelenítendő csomópontok és élek típusának kiválasztására. A kutatók ezáltal megjeleníthetik a számukra releváns információkat, például az összes személyt, aki egy bizonyos intézményben dolgozott. Ez lehetővé teszi, hogy egy átfogó képet kapjanak az adatokról, anélkül, hogy át kellene tekinteniük egy nagyméretű szövegtörzset. Továbbá az így megjelenített adatokat exportálni lehet, és le lehet tölteni JSON, valamint GraphML fájlformátumokban, így lehetőség nyílik az információ megosztására, továbbá nyílt forráskódú vizualizációs szoftverek használatára.

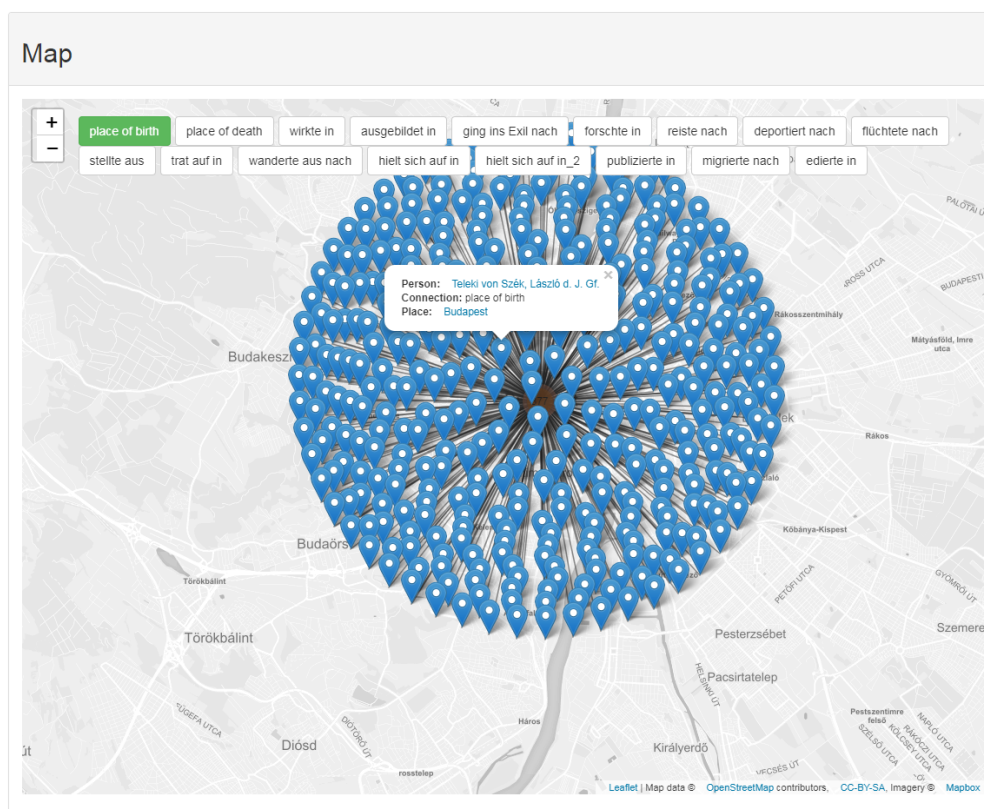
Egy másik vizualizációs lehetőség az APIS-ban az életrajzokhoz köthető helyek térképen történő megjelenítése. A térkép nézete aszerint változtatható, hogy a felhasználó milyen helyek megjelenítését választja ki a rendelkezésre álló típusok közül (pl. születési hely, tartózkodási hely stb.). Az egyes helységekre ráközelítve a felhasználó az áttekintést nyújtó nézetből (amelyen körökbe írt számok mutatják, hogy hány életrajz kapcsolódik egy adott városhoz vagy területhez) átválthat a személyeket egyesével bemutató nézetre.



3. ábra. A személy–hely kapcsolatokat bemutató hálózati ábra ifjabb gróf Teleki László (1811–1861) életrajza alapján



4. ábra. Az ŐBL-be felvett személyek születési helyei – térképvizualizáció az APIS-adatbázisban (részlet)



5. ábra. Budapesten született személyek az APIS-adatbázis térképén

3.3. Információkinyerés

Az APIS-projekt egyik célja, hogy a kutatók munkáját megkönnyítő automatikus feldolgozási lehetőségeket kínáljon, amelyek biztosítják, hogy az életrajzok szövegében található információ automatikusan és hatékonyan a felhasználó rendelkezésére álljon. A szövegek feldolgozását és értelmezését számítógépes nyelvészeti módszerekkel végeztük el, amelyek magukban foglalják a biográfiákban előforduló entitások (személyek, helyek, intézmények stb.) azonosítását, ezeknek az automatikus hozzákapcsolását a Linked Open Data Cloud ¹⁵ adatforrásaihoz, a kapott URI-k automatikus egyértelműsítését, az entitások közötti relációk felismerését és az eredmények manuális kiértékelését. A következő pontokban ezeket a lépéseket fogjuk részletesen bemutatni.

3.4. Névelem-felismerés és entitás-összekapcsolás

Bár az ÖBL-ben az életrajzok XML-formátumban is rendelkezésre állnak, ezek a fájlok néhány kulcsfontosságú eseményt (pl. születés és elhalálozás) kivéve nem tartalmaznak strukturált adatokat a leírt személyek életéről. A projekt egyik fő célja, hogy a természetes nyelvű szövegben kódolt információkat, például a személyek, helyek, in-

¹⁵ Linked Data – Connect Distributed Data Across the Web, hozzáférés: 2018.01.04, <http://linkeddata.org/>.

tézmények, események stb. neveit és az életrajzban bemutatott személynek az ezekhez fűződő viszonyát automatikusan feltárja. Ahhoz, hogy ezt hatékonyan és mindemellett jó minőségben végezhessük el, automatikus és a manuális információkinyerési technikákat ötvöztünk.

Az APIS-ban egy olyan információ kinyerésére alkalmas automatikus rendszert hoztunk létre, amely alkalmazható más digitális bölcsészeti projektek természetes nyelvű szövegeinek annotálására. Az általunk használt rendszer a következő komponensekből áll: a névelem-felismerő (Named Entity Recognizer), az entitás-összekötő (Entity Linker) és az egyértelműsítő komponens. Az automatikus információkinyerésre az Apache Stanbol¹⁶ nevű nyílt forráskódú szoftvert használtuk, amely a természetes nyelvű szövegben megtalált entitásokat kapcsolja össze a referenciaként szolgáló ontológiákkal és tudástárakkal, mint például a GND (Gemeinsame Normdatei),¹⁷ a GeoNames¹⁸ vagy a DBpedia.¹⁹ Az összeköttetések, amelyek az entitások és az életrajzok között így létrejönnek, nemcsak a szövegek szemantikai információkkal történő kiegészítését, hanem a hiányzó vagy hibás adatok kijavítását, illetve pótlását is lehetővé teszik.

A projektben használt Apache Stanbol előnye, hogy bármilyen RDF/XML-formátumban rendelkezésre álló ontológiát fel lehet használni a szövegek szemantikus kiegészítésére, továbbá biztosítja a dokumentumok NLP (Natural Language Processing, természetesnyelv-feldolgozás) szoftverekkel történő feldolgozását. A szemantikus annotálás elvégzéséhez az RDF/XML-formátumban rendelkezésre álló adatokból úgynevezett Referenced Site-okat²⁰ hozunk létre, ezek egy Solr-indexben²¹ tárolják az ontológiák adatait, ami *cache*-elten vagy a szerveren lokálisan érhető el.

Az APIS-ban létrehozott folyamat, amely a strukturálatlan, szöveges életrajzokból strukturált, szemantikai információkkal kiegészített és számítógépesen feldolgozott dokumentumokat hoz létre, két lépésből áll. Először feloldjuk a rövidítéseket, köztük a személynevek, intézménynevek, az akadémiai címek, a helynevek és a gyakori igék rövidítéseit. Két változatát is kifejlesztettük a rövidítések feloldásának. Az egyik, egy reguláris kifejezéseken alapuló Java program, ami az ÖBL-ben használt rövidítések listáját használja. A másik megoldás egy Python szkript, amely reguláris kifejezéseket és egy nagy méretű német nyelvű korpuszt használ a többjelentésű rövidítések feloldására és a helyes megoldás kiválasztására. A program lekérdezi a rövidítéseket és azok környezetét (előtte és utána álló szavakat) a német nyelvű adathalmazban, és amelyik változatra a legtöbb találatot kapja, azt a rövidítésfeloldást választja.

A szemantikus annotálási folyamat második lépése a Solr-indexek létrehozása az ontológiákból. Az entitás-összekapcsolás (Entity Linking) során ezekben az indexekben keres az Apache Stanbol a szövegben megtalált entitásoknak (személyek, helyek, intézmény nevei stb.) megfelelő fogalmakat az ontológiákban. Az APIS-projektben

¹⁶ Apache Stanbol, hozzáférés: 2018.01.04, <https://stanbol.apache.org/index.html>.

¹⁷ Deutsche Nationalbibliothek, *Gemeinsame Normdatei*, hozzáférés: 2018.01.04, http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html.

¹⁸ GeoNames, hozzáférés: 2018.01.04, <http://www.geonames.org/>.

¹⁹ DBpedia, hozzáférés: 2018.01.04, <http://wiki.dbpedia.org/>.

²⁰ „Working with Custom Vocabularies,” hozzáférés: 2018.01.04, <https://stanbol.apache.org/docs/trunk/customvocabulary.html>.

²¹ Apache Solr, hozzáférés: 2018.01.04, <http://lucene.apache.org/solr/>.

a GeoNames-ből és a GND-ből készítettünk indexeket, hogy a szövegben előforduló helyneveket, személyneveket és intézmények neveit hozzá tudjuk kapcsolni a Linked Open Data Cloud-hoz. Az indexeket a következőképpen hoztuk létre: letöltöttük a fent említett források RDF/XML-formátumú *dump*jait, amelyeket egy saját fejlesztésű szkripttel felváltunk részfájlokkra. A fájlok feldarabolásának két oka, hogy kezelhető méretű bemenetet kapjunk, és hogy könnyen tudjunk külön indexeket készíteni a különböző típusú entitásoknak. Az így kapott kisebb méretű fájlokból az Apache Stanbol indexelésre létrehozott Java package-ével készítettük el az Apache Solr-indexeket és az OSGi-keretrendszerben²² futó OSGi *bundle*-t (OSGi-köteget).

A Solr-index és az OSGi-köteg létrehozása és installálása után történik az Entity Linking komponens konfigurálása, amely a következő, kötelezően végrehajtandó lépésekből áll: az Entity Linking komponens azonosítására egy egyedi név megadása, a használni kívánt forrás (pl. GeoNames vagy GND) és a forrás Solr-indexében a kereséshez használni kívánt mező kiválasztása. Ezeken felül további beállítások megadására is van lehetőség, ami pontosabb és hatékonyabb entitás-összekapcsolást (*entity linkinget*) tesz lehetővé. Szűkíthetjük például csak a tulajdonnevekre a keresést, ekkor az algoritmus csak a tulajdonneveket választja ki a szövegből, és ezeket kérdezi le a Solr-indexben, így optimalizálva a találati precizitást és a futási időt. Lehetőség van továbbá entitástípusok leképezésére is, amennyiben az index tartalmaz erre vonatkozó információt. A leképezés eredményeként a felhasználó az indexben különböző típusokkal ellátott entitásokat átfogó kategóriákba sorolva kapja meg az alkalmazástól, így például összevonva, a *Személy* kategória alatt lehet megmutatni az összes, az indexben férfiként, illetve nőként megjelölt nevet.

Az Entity Linking komponens konfigurációját követően a bemeneti adatok feldolgozására vonatkozó nyelvtechnológiai információkat adjuk meg. Az alkalmazás számára ez írja le, hogy milyen NLP-lépéseket végezzen el, és azokat milyen sorrendben hajtsa végre a természetes nyelvű szövegeken. Az APIS-ban az Apache OpenNLP²³ nyílt forráskódú szoftvert használjuk az életrajzok számítógépes nyelvészeti feldolgozására. A szövegfeldolgozás folyamata a következő lépésekből áll:

- A bemeneti szöveg nyelvének megállapítása.
- A szöveg mondatokra tagolása (opennlp-sentence).
- A mondatok szavakra bontása (opennlp-token).
- A szavak szófajának megállapítása (opennlp-pos).
- A főnévi kifejezések megkeresése (opennlp-chunker).
- Entitás-összekapcsolás. (Entity Linking.)

Utolsó lépésként az NLP-feldolgozás során kapott főnevek és főnévi kifejezések a Solr-indexszel kerülnek összehasonlításra (*entity linking*). Amennyiben egy főnévi kifejezés megegyezik az indexben tárolt entitások valamelyikével, az entitást és a hozzá tartozó információkat az alkalmazás a kért formátumban szerializálva adja vissza. Ha több találat is van, 0-tól 1-ig terjedő pontszám jelzi, hogy melyik a legvalószínűbb helyes eredmény. Az általunk használt Entity Linking-szoftver előnye,

²² OSGi Alliance, hozzáférés: 2018.01.04, <https://www.osgi.org/>.

²³ Apache OpenNLP, hozzáférés: 2018.01.04, <https://opennlp.apache.org/>.

hogy tetszőleges RDF/XML-formátumban rendelkezésre álló ontológiát hatékonyan tud indexálni, így a felhasználó kiválaszthatja, hogy az alkalmazás melyik adatforrással annotálja meg a bemeneti szövegtörzset.

3.5. Relációfelismerés

Az entitás-összekapcsolás az első lépés a természetes nyelvű szövegek jelentésének automatikus értelmezésében. Segítségével a szövegben szereplő személyneveket, helyneveket, intézmények neveit össze tudjuk kapcsolni a Linked Open Data Cloud szótáraival. Ennek az előnye, hogy a szövegben karaktersorozatokként szereplő neveket egyedi azonosítóval (Uniform Resource Identifier, URI) rendelkező entitások váltják fel. Az ontológiák URI-val rendelkező fogalmai nemcsak egyértelműen azonosíthatók és hivatkozhatók, megoszthatók a különböző alkalmazások között, hanem használatukkal lehetőség nyílik a strukturálatlan szövegek információval történő kiegészítésére, illetve a helytelen adatok kijavítására.

A második lépés a korpuszok szemantikus feldolgozásában a megtalált entitások közötti kapcsolatok és azok típusainak megállapítása, más néven az automatikus relációfelismerés.

A relációfelismerés során az NLP-modul a szövegben detektált személyek, helyek, események között keres szemantikai kapcsolatokat, mint például *szülő-gyerek*, *utazott valahova*, *tanult valahol*, *részt vett egy eseményen* kapcsolat. Az automatikus relációfelismerésre több módszert is kipróbáltunk, melyek közül tesztelést követően fogjuk kiválasztani az alkalmazásba véglegesen integrált megoldást. Az első változat egy szabályalapú algoritmus, melyet a GATE-keretrendszer²⁴ használva implementáltunk. Az implementáció a GATE-annotációkon is működő reguláris kifejezési nyelvet, a JAPE-et²⁵ használja a szemantikai kapcsolatok szövegből történő automatikus kivonatolására. Első lépésként az Entity Linking modul kimenetét XML-formátumba konvertáltuk, ahol minden entitás külön elemként jelenik meg, az elemek attribútumai pedig az entitást tartalmazó Solr-index neve (pl. GeoNames_Capital_Cities_Index, GND_Persons_Index) és az entitáshoz köthető (pl. GeoNames vagy GND) URI.

Az XML-fájlokat ezután betöltöttük GATE-be, majd a beépített NLP-modulokkal tokenizáltuk és azonosítottuk a szövegekben a névelemeket. Az importált fájlok entitásai (az entity linking kimenete) és a GATE-ben végrehajtott NLP-feldolgozás eredménye (tokenek, névelemek) annotációkként tárolódnak GATE-ben. A JAPE reguláris kifejezések az annotációkon végeznek műveleteket, a különböző relációkra jellemző mintázatokat keresve a szövegekben. Amennyiben az alkalmazás talál egy, az adott relációra jellemző mintázatnak megfelelő szövegrészt, azt automatikusan ellátja egy újabb annotációval, amely a reláció típusát írja le. A relációazonosítás kimenetét XML-fájlokba – az NLP alkalmazások által széles körben használt formátumba – exportáljuk. A második relációfelismerési megoldás, amelyet teszteltünk, a nyílt for-

²⁴ GATE: General Architecture for Text Engineering, hozzáférés: 2018.01.04, <https://gate.ac.uk/>.

²⁵ „JAPE: Regular Expressions over Annotations,” hozzáférés: 2018.01.04, <https://gate.ac.uk/sale/tao/splitch8.html>.

ráskódú, Pythonban implementált *IEPY*-szoftver,²⁶ amely gépi tanuláson alapuló relációfelismerést végez. Az alkalmazás webes felületén a felhasználó annotálja az általa definiált relációkat (pl. *utazott valahova, megházasodott valakivel*), majd a szoftver az így létrehozott korpuszból megtanul egy modellt, amely az entitások között fennálló relációk felismerésére alkalmas. Az *ÖBL* esetében az *IEPY* nem bizonyult megfelelő szoftvernek, mert a relációk mindkét tagjának bejelölését megköveteli (pl. a *tanult valahol* esetében a személy- és helynevet egyaránt), az *ÖBL*-ben azonban az ismétlések elkerülése érdekében sokszor csak egyszer, az életrajz elején szerepel a személynév.

A harmadik megközelítés, amelyet megvizsgáltunk, a mondatok szintaktikai elemzésével (*parszolásával*) kapott fastruktúrából mélytanulás (Deep Learning) módszerrel történő relációfelismerés,²⁷ ahol a bemenetet néhány száz annotált biográfia adja. Ennek a megoldásnak a hátránya, hogy nagyon nagy mennyiségű annotált adatot kíván meg a tanulási folyamat, amely jelenleg nem áll rendelkezésünkre. A későbbiekben a reguláris kifejezésen és a Deep Learning módszerrel alapuló megoldásokat fogjuk összehasonlítani egy erre a célra létrehozott tesztkorpuszon, hogy a pontos eredményességi mutatók alapján (precision, recall, F1-score) döntsük el, melyik megoldást választjuk.

3.6. Egyértelműsítés és manuális szöveggondozás

A fent bemutatott automatikus szövegfeldolgozási módszerek jellemzője, hogy bár könnyen és hatékonyan képesek természetes nyelvű szövegeket feldolgozni, az eredmények sok esetben helytelenek. A hibás eredmények két formája az APIS-ban, amikor egy entitást vagy relációt nem talál meg az alkalmazás (pl. nem ismer fel a szövegben egy személynevet), illetve amikor a megtalált entitás vagy reláció azonosítása helytelen (pl. *tanult valahol* helyett a *dolgozott valahol* relációt adja vissza a program). A szoftver pontosságának növelése érdekében a két típushibát egyrészt automatikus, másrészt manuális módszerekkel próbáljuk kiküszöbölni. A helytelenül azonosított városok, országok és egyéb helynevek automatikus javítására a következő heurisztikákat alkalmazzuk:

Amennyiben több találat is érkezik az Entity Linking modultól, azt választjuk, amelyik:

- magasabb adminisztratív egységnek felel meg, azaz egy fővárost előnyben részesítünk például egy kisvárossal, faluval szemben. (Pl. a *Bécsben járt egyetemre* mondat esetében *Wient*, Ausztria fővárosát választjuk az USA-ban található *Wien* kisvárossal szemben.)
- A nagyobb lélekszámú település (azonos adminisztratív szintű helységek esetében nagyobb a valószínűsége, hogy a népesebb a helyes megoldás.)
- Közelebb van az *ÖBL*-ben releváns földrajzi területhez, az Osztrák-Magyar Monarchiához (nagyobb valószínűséggel éltek, dolgoztak az életrajzokban szereplő

²⁶ *IEPY: Information Extraction Tool*, hozzáférés: 2018.01.04, <https://github.com/machinalis/iepy>.

²⁷ Matthias Schlögl, Katalin Lejtovicz, Ágoston Zénó Bernád, Maximilian Kaiser and Peter Rumpolt, „Using Deep Learning to Explore Movement of People in a Large Corpus of Biographies,” *Zenodo*, 2018. jan. 16., <http://doi.org/10.5281/zenodo.1149023>.

The image shows two panels from the APIS interface. The left panel is a form for 'Záluží, kr. Středočeský'. It includes sections for 'Texts', 'Metadata Place', and 'Metalinfo'. The 'Metadata Place' section has fields for Name (filled with 'Záluží, kr. Středočeský'), Start, End, Status (filled with 'ambigue'), Kind, Latitude, and Longitude. At the bottom, there are buttons for 'delete', 'modify', and 'Revisions', and a 'Linked Open Data' link.

The right panel is titled 'related to' and contains a 'Map' section. It features a search box with the text 'Search for...' and a 'Search' button. Below the search box is a map showing a location in the Czech Republic near Prague. A popup window displays the following information: 'Place', 'ÖBL name: Záluží, kr. Středočeský', 'Geonames: Záluží', 'Select this URI', 'Confidence: 37.65585', and 'Feature: http://sws.geonames.org/3061899/'. Below the map, there are several tabs: 'Institution', 'Person', 'Event', 'Place', and 'Work'.

6. ábra. Az egyértelműsítést támogató felhasználói felület az APIS-ban. A többértelmű helységnevek térképen jelennek meg a döntést segítő plusz információkkal (név, GeoNames URI)

emberek a Monarchia területén, vagy az ahhoz közeli országokban, mint távolabbi területeken.)

A fenti heurisztikák sem tudják minden esetben a helyes megoldást kiválasztani, ezért az APIS-webalkalmazás lehetővé teszi a találatok kézi javítását, és a többértelmű találatok egyértelműsítését. Az alábbi ábra az egyértelműsítésre kifejlesztett felhasználói felületet mutatja be.

Az automatikus és kézi hibajavítás ötvözése az APIS-ban megfelelő megoldásnak bizonyult az adatok pontos és hatékony feldolgozására, de hogy minél gyorsabban és minél kevesebb emberi erőforrással tudjunk nagy mennyiségű szöveges adatot feldolgozni, a jövőben tervezzük a kézi hibajavítás automatizálható részeinek implementálását.

4. Kutatási témák

Az APIS-projekt egyik célkitűzése az ÖBL-ben publikált biográfiai adatok alapján végzett történeti népesség- és migrációkutatás. A több mint 18000 életrajzot tartalmazó

korpusz szellem- és társadalomtudományi szempontból történő vizsgálatát – beleértve a humángeográfia tárgykörébe tartozó, főként népességföldrajzi és mobilitás-, illetve migrációkutatási kérdéscsoportokat – egyrészt egy transzdiszciplináris kutatócsoport együttműködése, másrészt a digitalizált életrajzi szócikkekre épülő virtuális kutatási környezet (Virtual Research Environment, VRE) teszi lehetővé.

A lexikonba felvett személyek migrációs útvonalainak és mintáinak rekonstrukciójára összpontosító szociodemográfiai elemzés egyúttal egy készülőben lévő disszertációs dolgozat témája. Az *ÖBL* életrajzi szócikkei ugyanis térvonatkozású információkat is tartalmaznak (születési és halálozási hely, tartózkodási helyek stb.), melyek segítségével rekonstruálhatók a térbeli életutak. Ebben a kontextusban mind a térbeli mobilitás mérete, illetve a migráció, a migrációs útvonalak és vándorlási minták, mind pedig az érintett népességcsoportok demográfiai struktúrája figyelmet érdemel, és a kutatás tárgyát képezi. Emellett vizsgáljuk a városok mint migrációs csomópontok szerepét, valamint a települések és az intézmények hatókörét. Az életrajzokban említett, de nem egyértelműen identifikálható helyneveket történeti kartográfiai források segítségével azonosítjuk. Az elemzés összességében a 19. század elejétől a 20. század közepéig tartó korszak politikai és politikai-adminisztratív változásainak keretei között helyezhető el. Ebben az összefüggésben értékelhetők a projekt eddig publikált részeredményei is, melyek az egyes személyek migrációjának példáján keresztül a földrajzi nevek közötti összefüggésekre világítanak rá a közösségkeresési módszerek (*community detection*) segítségével.²⁸

Az életrajzi szócikkekben előforduló helynevek jelentőségét a migrációs folyamatok vizsgálata szempontjából idősebb gróf Andrassy Gyula (1823–1890) és ifjabb gróf Teleki László (1811–1861) életrajza alapján szemléltetjük.²⁹ Andrassy 1867–1871 között magyar miniszterelnök, 1871–1879-ben pedig a Monarchia külügyminisztere volt. Az *ÖBL* Andrassy-szócikke számtalan térvonatkozású adatot tartalmaz: többek között Kassát mint születési helyet, Pestet, illetve a pesti egyetemet mint tanulmányai színhelyét, Isztambul mint egyik munkahelyét, Londont és Párizst, ahol száműzetésben élt, valamint az isztriai Voloskót, ahol elhunyt.

²⁸ Ágoston Zénó Bernád, Maximilian Kaiser, Sebastian M. Mair und Alexander Rind, „Communities in biographischen Netzwerken,” in *FMT 2017: Proceedings of the 10th Forum Media Technology and 3rd All Around Symposium, CEUR Workshop Proceedings*, vol. 2009, eds. Wolfgang Aigner, Thomas Moser, Kerstin Blumenstein, Matthias Zeppelzauer, Michael Iber, Grischa Schmiedl (St. Pölten: CEUR-WS.org, 2017), 83–87, <http://ceur-ws.org/Vol-2009/fmt-proceedings-2017-paper12.pdf>; Ágoston Zénó Bernád, Maximilian Kaiser, Matthias Schlögl and Katalin Lejtovicz, „APIS Dataset Humanities Scholars,” *Zenodo*, 2017. szept. 19., <https://doi.org/10.5281/zenodo.896281>.

²⁹ „Andrassy Julius (d. Ältere) Graf,” in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. I., Hg. Leo Santifaller und Eva Obermayer-Marnach (Graz/Köln: H. Böhlau, 1957), 20–21; Zoltán Fónagy, „Teleki von Szék László Gf. d. J.,” *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 14, Hg. Ernst Bruckmüller und Christine Gruber (Wien: Verlag der Österreichischen Akademie der Wissenschaften, 2015), 135–136.

Andrássy, Julius d. Ä. Gf.

Texts

Metadata Person

Name

Start

Please enter a date (DD).(MM).YYYY

End

Please enter a date (DD).(MM).YYYY

Status

First name

The persons's forename. In case of more then one name...

Profession

Title

Gender

related to

Event

Institution

Person

Place

Delete	Start Date	End Date	relation type	Related Place	Edit
<input type="checkbox"/>	—	—	place of birth	Košice	<input type="button" value="Edit"/>
<input type="checkbox"/>	—	—	place of death	Volosko	<input type="button" value="Edit"/>
<input type="checkbox"/>	—	—	ausgebildet in	Pest	<input type="button" value="Edit"/>
<input type="checkbox"/>	01.01.1847	01.01.1847	wirkte in	Pressburg	<input type="button" value="Edit"/>
<input type="checkbox"/>	01.01.1848	01.01.1848	wirkte in	Schwechat	<input type="button" value="Edit"/>
<input type="checkbox"/>	01.01.1849	01.01.1849	wirkte in	Istanbul	<input type="button" value="Edit"/>
<input type="checkbox"/>	01.01.1849	—	reiste nach >> ging ins Exil nach	London	<input type="button" value="Edit"/>
<input type="checkbox"/>	—	01.01.1857	reiste nach >> ging ins Exil nach	Paris	<input type="button" value="Edit"/>
<input type="checkbox"/>	08.06.1867	08.06.1867	wirkte in	Buda	<input type="button" value="Edit"/>

Relation type*

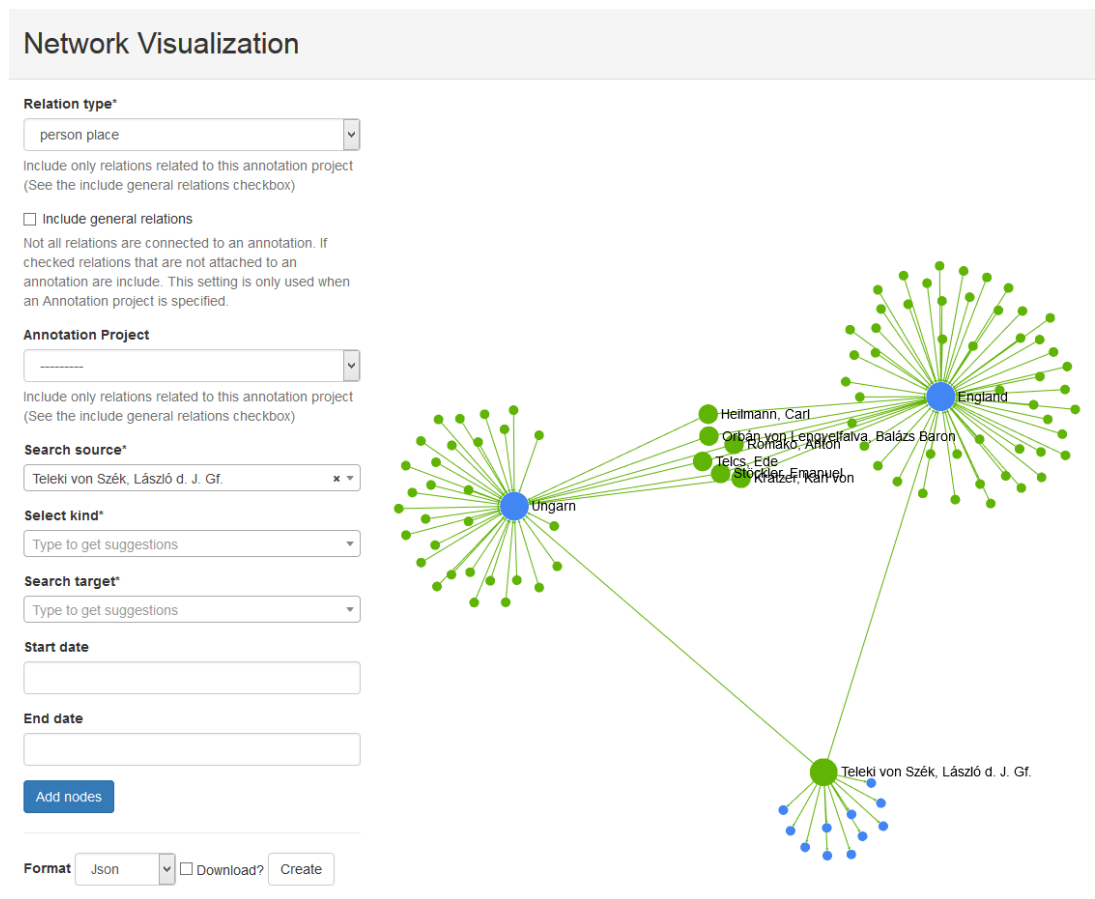
Place*

7. ábra. Andrássy személy–hely kapcsolatait megjelenítő webes felület az APIS-alkalmazásban

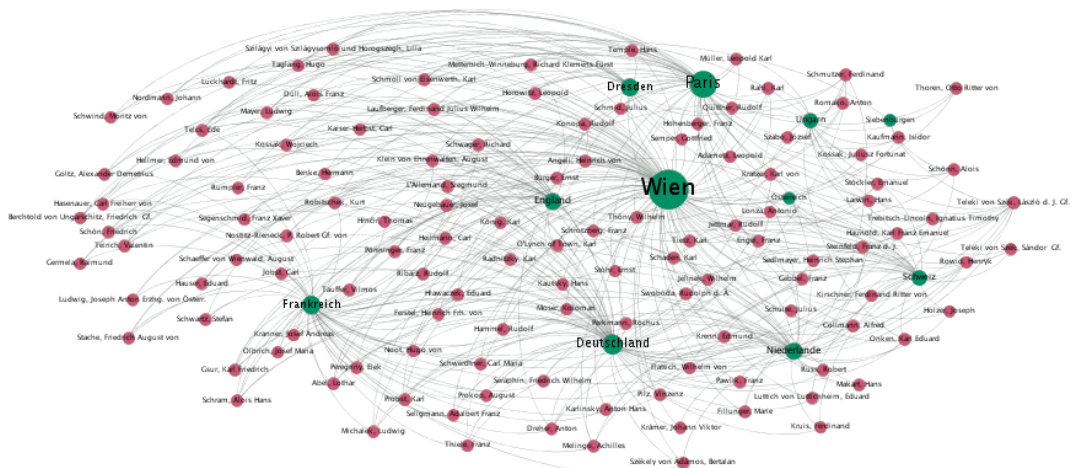
Teleki, az 1848–49-es forradalom és szabadságharc egyik vezető alakja, a Határozati Párt politikusa, a Magyar Tudományos Akadémia tiszteleti tagja volt. Pesten született és halt meg, Sárospatakon, Budán és Bécsben tanult. Ezt követően európai tanulmányútja során bejárta Németországot, Hollandiát, Angliát és Franciaországot. Politikai pályafutásával kapcsolatosan a szócikkben említett földrajzi nevek Erdély, Magyarország, Svájc és Ausztria, illetve Abony, Párizs és Drezda.

További kutatások témáját képezik a művészek és gyűjtők életrajzai, sajtóhálózatok feltérképezése és más tudománytörténeti kérdésfelvetések. A művészi életrajzok esetében mintegy 506 személyből, a bécsi Künstlerhaus tagjaiból álló csoportról van szó, az összetartozásukat éppen ez a tagság biztosítja.³⁰ Két fő szempontot érvényesítünk a kutatáskor: egyrészt különös figyelmet fordítunk a művészek életútjának és karriertörténetének összehasonlító elemzésére, másrészt a lexikonban található, ám ehhez a csoporthoz nem tartozó további személyeket is műgyűjtőként határozhatunk meg. Ez utóbbi esetben egy még feltárandó csoportról van szó, amelynek tagjai között

³⁰ Maximilian Kaiser, „Was uns Biographien über Künstlernetzwerke sagen. Konzepte für eine historische Netzwerkanalyse auf Basis biographischer Texte aus dem *Österreichischen Biographischen Lexikon (ÖBL)*,” in *Europa baut auf Biographien*, 383–403.



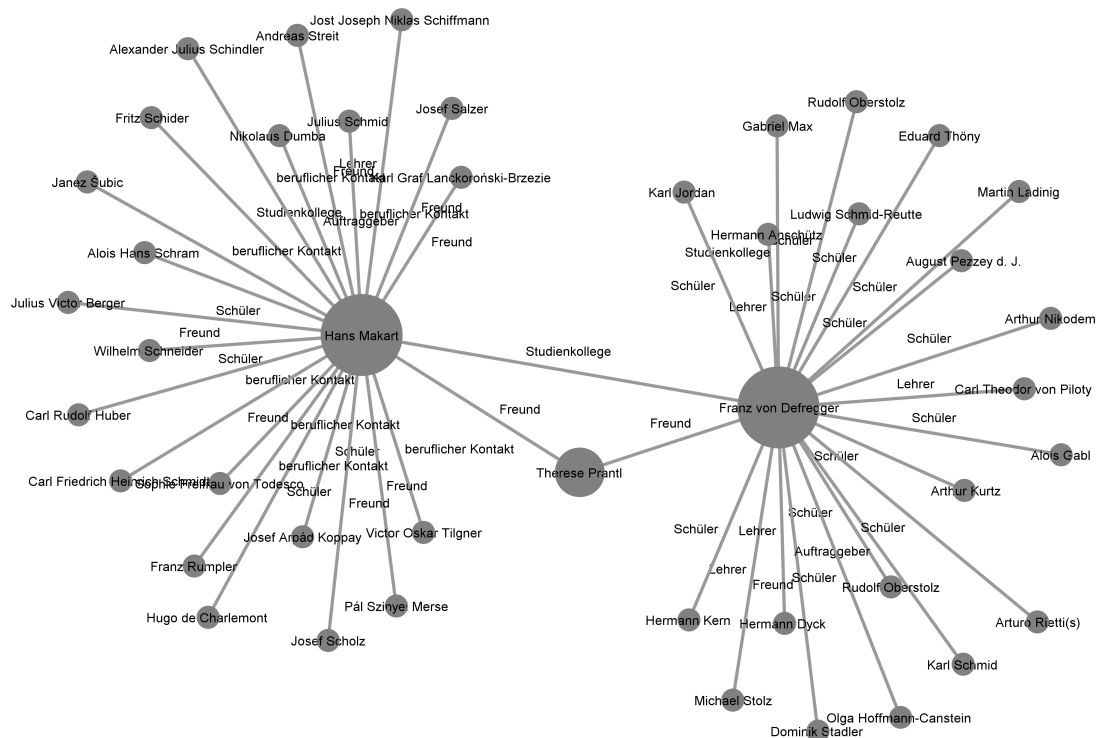
8. ábra. A Teeki-életrajzban előforduló földrajzi nevek hálójának vizualizációja az APIS-adatbázisban (részlet)



9. ábra. A Teeki-életrajzban előforduló földrajzi nevek és a hozzájuk kapcsolódó személyek hálójának megjelenítése a Gephi segítségével³¹

³¹ Gephi: The Open Graph Viz Platform, hozzáférés: 2017.12.14, <https://gephi.org/>.

orvosokat, ügyvédeket, kereskedőket és nagyiparosokat találunk. Ők megbízóként vagy a műalkotások felvásárlójaként nem elhanyagolható mértékben befolyásolták az osztrák művészettörténet fejlődési folyamatait.



10. ábra. Hans Makart (1840–1884) és Franz von Defregger (1835–1921) festőművészek – a *Visone*-alkalmazás segítségével³² megjelenített – kapcsolati hálója, az *ÖBL*-életrajzokban említett személynevek alapján (© Maximilian Kaiser)

Választásunk azért esett a bécsi Künstlerhaus közösségének vizsgálatára, mert ez jelentette a Habsburg Monarchia legfontosabb reprezentációs és műkereskedelmi platformját a 19. századi művészek számára. A Künstlerhaus kiállításain való részvétel egy fiatal művész számára nem csupán a tagfelvétel és az azzal összefüggésben álló integrálódás miatt volt döntő jelentőségű, hanem a további anyagi sikert is meghatározta. Sok esetben ez biztosította számukra a művészetekre fogékony nyilvánosság előtti megjelenést, és így kerültek kapcsolatba a műgyűjtőkkel is. Az intézménynek kiemelt szerepe van a művészi biográfiák szempontjából igen jelentős műgyűjtői életrajzok esetében is, melyeknek forrásai a Künstlerhaus tagjegyzékei. A csoportok közötti hasonlóságok az életrajzi adatok összehasonlításából és értékeléséből derülhetnek ki. A szövegekben gyakran előforduló helyek, intézmények és személyek nevei olyan hálózatokat rajzolnak ki, amelyek a csoportok közötti kapcsolatokat is kontúrozni tudják.

A 19. és 20. század sajtóhálózatainak vizsgálata a tágabb értelemben vett közép-európai szellem- és kultúrtörténeti kutatások, valamint a történeti hálózatkutatás

³² *Visone: Visual Social Networks*, hozzáférés: 2017.12.14, <http://visone.info/>.

kereteibe illeszkedik. Ebben az esetben nem csupán a kézi annotálással ellátott *ÖBL*-beli életrajzokat, valamint az *ÖBL*-hez tartozó GIDEON-adatbázis kiegészítő adatait használjuk, hanem külső sajtótörténeti forrásokat is bevonunk a kutatásba. Nem törekszünk az intézmények és szereplők összefonódásainak és kapcsolatainak teljes körű feltárására, hanem inkább egy reprezentatív minta feldolgozása a célunk. A kutatás a személyközi viszonyok mellett a személy–intézmény és az intézmények közötti relációkra koncentrál, és azokra a hasonlóságokra és különbségekre kérdez rá, amelyek a Monarchia két államának területén a sajtó és a politikai szféra kölcsönös függőségét és ezen kapcsolat megjelenési formáit jellemezte. Az életrajzok feldolgozásánál feltűnt, hogy a kenyérkereseti foglalkozástól függetlenül igen sok személy jelentős sajtótevékenységet is kifejtett, ennek következtében a továbbiakban fokozottan figyelünk az egyes foglalkozási csoportok közötti többszörös átfedésekre.

Projektünk egyik további aspektusát képezi az *ÖBL* szerzői hálózatának feltérképezése. Ebben az esetben a lexikon létrejöttének körülményei kerülnek előtérbe, ami az *ÖBL* adatainak kritikai feldolgozásához és értelmezéséhez járul hozzá. Mindemellett további adatok bevonása és levéltári kutatások is szükségesek ahhoz, hogy feldolgozhassuk és rekonstruálhassuk azokat a tanár–diák, illetve a személyek és intézmények közötti kapcsolatokat, valamint azoknak a különböző kül- és belföldi szerzői nemzedékeknek a tudományos, politikai és ideológiai szocializációját, amelyek az *ÖBL* általános narratíváját meghatározták. Ezen a téren első kutatási eredményeink időben a hidegháború szakaszára koncentrálnak, és ezen periódusban az *ÖBL*, valamint a magyarországi,³³ illetve romániai³⁴ tudományos intézmények közötti együttműködéseket vizsgálják.

Digitisation and Network Research: The Austrian Biographical Dictionary and the APIS project

Digitisation methods have significantly evolved in recent years and made it possible for all types of digital humanities projects, among them biographical projects to gain ground. One of the aims of a biographical research is to transform the texts of already digitized encyclopaedias (e.g. the *Österreichisches Biographisches Lexikon 1815–1950*, *ÖBL*, which consists of structured and unstructured data) to a machine readable format, and to perform network-, statistical- and computational linguistic analysis on the transformed data. The biographical research of *ÖBL* started in 2015 in the Institute for Modern and Contemporary Historical Research [Institut für Neuzeit- und Zeitgeschichtsforschung (INZ)] of the Austrian Academy of Sciences within the project Mapping historical networks:

³³ Ágoston Zénó Bernád, „Europäische Biographik während des Kalten Krieges. Zur Geschichte der Kooperation zwischen dem Österreichischen Biographischen Lexikon und wissenschaftlichen Institutionen in Ungarn 1956–1989,” in *Europa baut auf Biographien*, 81–105.

³⁴ Ágoston Zénó Bernád, „Das Österreichische Biographische Lexikon 1815–1950. Ein Nachschlagewerk im Wandel der Zeit,” in *Hortus amicorum. Köszöntőkötet Egyed Emese tiszteletére*, szerk. Bartha Katalin Ágnes, Biró Annamária, Demeter Zsuzsa és Tar Gabriella-Nóra (Kolozsvár: Erdélyi Múzeum-Egyesület, 2017), 457–467.

Building the new Austrian Prosopographical, Biographical Information System (APIS). The project is aimed at the digital processing of the *ÖBL* and the creation of a virtual research environment which is implemented in cooperation with the Austrian Center for Digital Humanities (ACDH) and the Institute for Urban and Regional Research [Institut für Stadt- und Regionalforschung (ISR)]. This paper discusses the technical and natural language processing solutions used in the APIS web application to open up biographical data for the researchers by processing and representing them in a structured format.

Keywords:

entity linking, biographical lexicon, information retrieval, historical network research