

**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

A magyar irodalomtörténet bibliográfiájának migrációja az ITIdata szemantikus adatbázisba: Gyakorlatok és tanulságok

Migration of the bibliography of Hungarian literary history to the ITIdata semantic database: practices and lessons learned

Dobás Kata

HUN-REN Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézet

dobas.kata@abtk.hu

ORCID: [0009-0009-7632-8276](https://orcid.org/0009-0009-7632-8276)

Tüskés Anna

HUN-REN Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézet

tuskes.anna@abtk.hu

ORCID: [0000-0003-1464-5013](https://orcid.org/0000-0003-1464-5013)

Absztrakt

Az MTA Irodalomtudományi Intézetének Bibliográfiai Osztályán készült, *A magyar irodalomtörténet bibliográfiája* című tízkötetes összefoglaló munka 1972 és 2013 között látott napvilágot. Az Irodalomtudományi Intézet a kiadott kötetek alapján online bibliográfiai szolgáltatást indított 2016 novemberében, amelyben a bibliográfia összetett keresési műveletekkel vagy szabadszavas kereséssel nyilvánosan és ingyenesen használható (<http://www.iti.mta.hu/mib/index.html>). A tíz kötet anyagából jelenleg a 2. és a 3. kötet adatbázisa használható. Az Intézet nemcsak digitalizálja a nyomtatásban megjelent köteteket, hanem folytatja is *A magyar irodalomtörténet bibliográfiáját*. A folytatás adatgyűjtő munkája 2013-ban kezdődött meg, s egyelőre a 2012-es évtől kezdve szolgáltat adatokat. 2023-ban az Irodalomtudományi Intézet bibliográfiai adatbázisának metaadatait egy átfogó mapping és több fázisra bontható névtérre fektetés után az ITIdata szemantikus adatbázisba (https://itidata.abtk.hu/wiki/Main_Page) migráltuk annak érdekében, hogy az Irodalomtudományi Intézetben található tudományos projektek adatbázisai és a különböző kutatások adatai egyetlen adatbázisban váljanak hozzáférhetővé. A tanulmány bemutatja a több mint 100 ezer rekordot tartalmazó bibliográfiai adatbázis migrációjának előkészítő munkálatait, a kidolgozott workflow-t, az általunk alkalmazott adatgazdagítás különböző formáit, a tanulságokat és eredményeket.

Kulcsszavak: bibliográfiai adatbázis, szemantikus adatbázis, adatmigráció, workflow

Abstract

The ten-volume work entitled *The bibliography of Hungarian literary history*, prepared at the Bibliographical Department of the Institute of Literary Studies of the Hungarian Academy of Sciences, was published between 1972 and 2013. Based on the published volumes, the Institute launched an online bibliographic service in November 2016, in which the bibliography can be used publicly and free of charge with complex search operations or free word searches (<http://www.iti.mta.hu/mib/index.html>). The database of the second and third volumes can currently be used from the material of the ten volumes. The Institute not only digitizes the volumes published in print but also continues the bibliography of Hungarian literary history. The follow-up data collection work began in 2013, and for the

time being, it provides data from 2012. In 2023, the metadata of the bibliographic database of the Institute was migrated to the ITidata semantic database (https://itidata.abtk.hu/wiki/Main_Page) after a comprehensive mapping and distribution into a namespace that can be divided into several phases. The process was implemented to ensure that the databases and data from various researches of scientific projects located at the Institute should be accessible in a single database. The study presents the preparatory work for the migration of the bibliographic database containing more than 100,000 records, the elaborated workflow, the different forms of data enrichment we use as well as the lessons and results.

Keywords: bibliographic database, semantic database, data migration, workflow

I. A magyar irodalomtörténet bibliográfiája 1957–2013, 2013–2022

A Magyar Tudományos Akadémia Irodalomtudományi Intézet Bibliográfiai Osztálya 1957-ben alakult Kemény G. Gábor vezetésével, akit e tisztségében 1962-től Gerézdi Rabán, 1965-től Vargha Kálmán, 1985-től haláláig Kókay György, 2007-től 2012-ig Csóra Karola követett. Az osztály három célt tűzött maga elé: 1. a magyar irodalomtörténet retrospektív bibliográfiai kézikönyvének létrehozása a nagy irodalomtörténeti korszakok szerint tagolva és ezen belül általános és személyi részre osztva; 2. a 20. századi irodalmi bibliográfiai katalógus elkészítése; 3. bekapcsolódás az Irodalomtudományi Intézetben folyó egyéb bibliográfiai munkákba, az Irodalomtörténeti bibliográfiák sorozat szerkesztése. A kis létszámú osztály a gyűjtést gyakornokok mellett külső munkatársak bevonásával végezte; a szakozást, a szerkesztést és a sajtó alá rendezést mindig az osztályon dolgozó munkatársak végezték. A kézikönyv-kötetek sorozatszerkesztői Vargha Kálmán, V. Windisch Éva és Kókay György voltak. A kötetek szerkesztői többnyire az adott korszakot kutató tudósok közül kerültek ki. A Bibliográfiai Osztály a mindenkori igazgatóhelyettes felügyelete alatt működött, így Klaniczay Tibor ellenőrizte a munkát, aki később igazgatóként is biztosított minden lehetőséget a bibliográfiai kézikönyvek megjelenéséhez. Az intézeti bibliográfiai kézikönyv szerkezeti kialakításához Kozocsa Sándor éves irodalomtörténeti bibliográfiáját tanulmányozták az osztály tagjai, és figyelembe vettek külföldi kézikönyveket is. Esetenként tanácsot kértek és kaptak írótól, irodalomtörténésztől és más intézmények bibliográfus munkatársaitól.

A Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetének 2013-ban Stauder Mária vezetésével újjáalakított Bibliográfiai Osztálya azt a célt fogalmazta meg, hogy a 2010-ig papír alapon létrehozott és kiadott *A magyar irodalomtörténet bibliográfiája* című, tízkötetes vállalkozás immár ne analóg módon folytatódjon, hanem a 21. század digitális online világába átemelve kereshetővé váljon, és az új bibliográfiai gyűjtés azonnal adatbázisban készüljön. A bibliográfiai gyűjtés szempontjai a következők:

1. Szépirodalmat nem gyűjtünk, kivéve, ha elhunyt magyar szerző újonnan kiadott szövege, jegyzetelt, és bizonyíthatóan új szöveggözlés: új, azaz eddig nem publikált, vagy az adott szövegváltozat nem publikált. Külföldi szerző ilyen szövegeit csak akkor vesszük föl, ha magyar irodalmi, vagy kultúratörténeti/elméleti vonatkozása van.
2. Élő vagy holt magyar író, költő, irodalomtörténész, kritikus vagy hozzátartozójának levele, naplója, notesze, jegyzete. Külföldi szerző hasonló szövegeit csak akkor vesszük föl, ha magyar irodalmi, vagy kultúratörténeti/elméleti vonatkozása van.

3. Más jeles személy holt/élő magyar íróval, költővel, irodalommal foglalkozó levele, naplója stb. Külföldi szerző hasonló szövegeit csak akkor vesszük föl, ha magyar irodalmi, vagy kultúratörténeti/elméleti vonatkozása van.
4. A magyar irodalommal, irodalomtörténettel foglalkozó cikk/tanulmány, a kezdetektől napjainkig.
5. Minden komparatistikai szemlélettel bíró, a magyar irodalommal, irodalomtörténettel, kultúrával is foglalkozó cikk, tanulmány, cikkrészlet nyelvi korlátok nélkül.
6. A tudományágak közötti összefüggésekben a magyar irodalommal, irodalomtudománnyal, kultúrával foglalkozó cikk, tanulmány, cikkrészlet.
7. A kritikai rovat feldolgozására is a fenti szempontok érvényesek. Irodalmi folyóiratok esetében minden magyar szerző regényére, verskötetére stb. vonatkozó kritikát felvesszünk, ugyanígy az antológiákat, drámaköteteket stb. bíráló/ismertető cikkeket is. Ha egy bírált műnek valamilyen tartalmi, esetleg keletkezéstörténeti kapcsolata van a magyar irodalommal, kultúrával, akkor is felvesszük a bírálatot. A bíráló nemzetisége, a bírálat nyelve nem meghatározó, tehát az bármilyen lehet, nem csak magyar.

2016. november 23-án került sor az Irodalomtudományi Intézet bibliográfiai portáljának nyilvános elindítására a kurrens gyűjtésből, ami a 2013-as évtől tartalmazza a magyar irodalom történetére vonatkozó, mintegy 730 magyar és idegennyelvű időszaki kiadványban megjelent szak- és ismeretterjesztő írásokat. Az elmúlt évek során a tízkötetes irodalomtörténeti bibliográfia köteteiből az első három kötet ugyancsak kereshetővé vált a BBL interaktív adatbázisban, a többi hét továbbra is csak kereshető pdf-ként letölthető. A 2. kötet mintegy 26400 rekordot, adatbázis tételt tartalmaz 3000 elemű szakrendi besorolásban, tehát ez egy nagyon erősen cizellált, túlszakozott bibliográfiai vállalkozás, amelynek a szakrendjét megmenti az adatbázis, tehát a szabadszavas keresésen túl az eredeti szakrendi kategóriák rendjében kereshető.

Munkafázisok	Sárga cédulák korszaka 1958–2013	Adatbázis 2013–2022
Tervezés	<ul style="list-style-type: none"> •Osztály- és egyéni munkaterv, tervtárgyalások. •Általános alapelvek, tematikai tervek, korszakolás 	<ul style="list-style-type: none"> •A gyűjtendő folyóiratok listájának összeállítása, a szakrend összeállítása.
Gyűjtés	<ul style="list-style-type: none"> •Kézírt anyaggyűjtés nagy közgyűjteményekben cédulákra. •Kézírt cédulák gépelése sárga kartonokra. 	<ul style="list-style-type: none"> •Laptonon tételfelvétel (címfelvétel, annotáció és szakozás - szakszámmal)
Ellenőrzés	<ul style="list-style-type: none"> •Adatellenőrzés, összeolvasás. •A gépelt cédulák formai egységesítése, esetleg újragépeléssel. 	<ul style="list-style-type: none"> •Az adatbázis hajlékonyságának megfelelően azonnali hibajavítás lehetőségei.
Rendezés	<ul style="list-style-type: none"> •Szakozás, a cédulák szakszámmal való ellátása, azok alapján rendezése a szakrendnek megfelelően feliratokkal ellátott rózsaszín cédulák rendjében. •Egy-egy szakon belüli idő- és betűrendbe rakás. •Selejtezés, szerkesztői átvizsgálás. 	<ul style="list-style-type: none"> •Gép automatikusan rendezi. •Azonnali lekérdezés lehetősége.

Munkafázisok	Sárga cédulák korszaka 1958–2013	Adatbázis 2013–2022
Karbantartás, mentés, tárolás	<ul style="list-style-type: none"> •A cédulák tárolása katalógusfiókokban és katalógusszekrényekben történt. •A cédulaformában lévő, megszerkesztett anyag lektorálása. •A lektori jelentésnek megfelelő javítások elvégzése. •A lektorált anyag A4-es papírokra gépeltetése 4 példányban. •A legépeltetett anyag összeolvasása a cédulákkal. •Folyamatos könyvtári kontrolok végzése. •Nyomdai előkészítés. 	<ul style="list-style-type: none"> •Konzisztencia ellenőrzése •Helyi mentés •MySQL adattáblába mentés •A program automatikusan jelzi a véletlenül elkövetett, de formalizálható hibákat.
Közzététel	<ul style="list-style-type: none"> •Nyomdába adás. •Korrekktúrák: egy hasábkorrekktúra, két tördelt korrekktúra, és egy átnéző korrekktúra elkészítése. •Mutatók: névmutató készítése, rövidítésjegyzékek készítése és más technikai feladatok ellátása. •Imprimálás, kiadás, könyvbemutató. 	<ul style="list-style-type: none"> •Interneten. •Nincs nyomtatott változat, de aki akar, nyomtathat magának.

1–2. ábra: A munkafázisok összehasonlítása. Forrás: saját szerkesztés

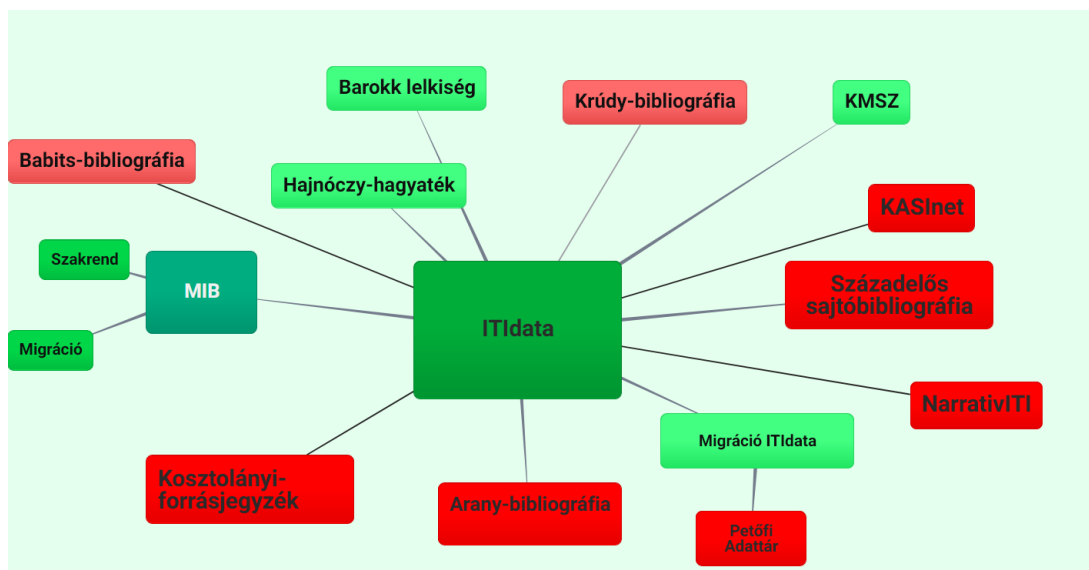
Összehasonlítás szempontjai	Sárga cédulák korszaka 1958-2013	Adatbázis 2013-2022
Személyi állomány	Intézeti munkatársak: 15 Külső felvevők: 20	3 és ½ intézeti munkatárs
Munkaeszköz	Cédulák, írógép, szekrény	Laptop
Gyűjtés helyei	Nagy közgyűjtemények (főként OSZK)	MTA BTK ITI Eötvös Könyvtár, OSZK, MTA KIK, FSZEK stb.
Gyűjtőkör: irodalomtörténet és irodalomtudomány, szépirodalom nem	Könyvek és időszaki kiadványok, 10 kötetben: 150.000 tétel	Egyelőre csak a 2013. évi időszaki kiadványok (hetilapok, folyóiratok, évkönyvek) Eddig felgyűjtve: Jelenleg 5.000 tétel
Szakozás	I. Általános rész II. Személyi rész Aprólékos beosztás	Szakok összevonása Írói jegyzék folyamatos bővülése Új szakok, pl. irodalomtörténészek
Forma	Saját szabvány	ItK-szabvány
Analitika	Kevés tartalmi feltárás []-ben	Annotáció: 3-5 mondat, említett szerzők és művek

3. ábra: A munkafolyamat összehasonlító elemzése. Forrás: saját szerkesztés

A keresési feltételek kombinálhatók: lehet szabadszavas keresés mellett a megjelenés éve szerint, a szerzők nevére, az írás címében előforduló szavakra, a periodikumoknak a címe szerint, amelyekben megjelentek, és mindezek kombinálható keresési lehetőségek. Továbbá lehet keresni a szakrendi beosztás szerint, amely szakozást érvényesítették a bibliográfia elkészítése során: az MTMT (Magyar Tudományos Művek Tára) kategóriarendszerét némiképpen meghaladva és cizellálva műfaji kategóriák szerepelnek. Nem változtathatatlán a szerkezet és a gyűjtőkör, s nem egy teljesen hibátlan adathalmazt tartalmaz, hanem egy folyamatosan változó, frissülő, javítható adatbázis az irodalomtudomány iránt érdeklődő, a tudományokat művelő, vagy csupán kíváncsi kollégáinknak, diákoknak, egy szóval felhasználóknak.

II. ITIdata 2023–

Az Irodalomtudományi Intézet a DigiPhil szakmai vezetésével 2021-ben kezdte meg Intézeti szemantikus web-alapú adatbázisának a fejlesztését, az ITIdata-t.¹ Mivel az Intézet célja, hogy minden, az Intézetben készülő adatbázist egy rendszerben képezzük le, így a döntés *A magyar irodalomtörténet bibliográfiáját* is érintette. A migráció előkészítését 2022-ben kezdtük meg. Az ITIdata egy wikibase-alapú szemantikus adatbázis, melynek struktúráját a WikiData mintájára alakítottuk ki. Az adatbázisban tulajdonságok (property) struktúrálják a bibliográfiai leíró rekordokat. Ezen tulajdonságok egy részéhez tartozó állítások stringként, mások entitásként kerülnek fel az adott rekordhoz – ez utóbbi nagyon szerteágazó kapcsolati rendszerek, hálózatok kiépítésére is lehetőséget ad. Ahogy azt már említettem, az ITIdata-ban kapnak helyet az intézeti tudományos projektek, ezeket a következő buborékábra mutatja.



4. ábra: Az ITIdata irodalomtudományos projektjei. Forrás: saját szerkesztés

¹ https://itidata.abtk.hu/wiki/Main_Page

Ezeken túlmenően a tavalyi év során egy nagy migrációs folyamat is elindult, melynek eredményeként nagy mennyiségű bibliográfiai adat került fel az adatbázisba. A rekordszámokat tekintve ez az összrekordszám majdnem a duplája a tavalyi évhez képest.²

1. A magyar irodalomtörténet bibliográfiájának migrációja

A korábbi SQL adatbázisból kinyert dump adatstruktúrája alapján készítettük el a mappinget. A korábbi BBL-ben található mezőneveket megfeleltettük az ITIdata tulajdonságainak. Két nagyon fontos és hosszadalmas előkészítő munkafolyamat zajlott ezt követően: mivel az ITIdata-ban bizonyos tulajdonságokhoz entitás típusú adatok tartoznak, korábban ezek stringként kerültek rögzítésre a BBL felépítéséből következően, ezeket létre kellett hozni, és a létrehozott rekord számát a dump-ból kinyert xml-fájlban az adott helyére vissza kellett illeszteni, hogy a betöltés problémamentesen megtörténhessen. Az eddig felsorolt munkafázisokban a Qulto Kft. munkatársai voltak segítségünkre, a velük való együttműködés elengedhetetlen része a migrációnak.

Entitástípusú rekordok előkészítése vette a legtöbb időt igénybe: ilyenek voltak a periodikák, a könyvek, összefoglalóan a bennfoglaló művek, a sorozatok, a műfajok, valamint a szakrendi nevek. A periodikák címéről írnék részletesebben: a BBL-ben ezek nagy hányada rövidített névvel került fel a korábbi adatbázisba (pl. ItK – Irodalomtörténeti Közlemények) – a lapok neveit a Bibliográfiai osztály munkatársai azonosították, a rövidítéseket feloldották, valamint az adatgazdagítást elősegítendő, rögzítették Excel-táblázatban a lapok metaadatait (megjelenés helye, ideje, különböző azonosítók: EPA, WikiData, MTMT-folyóiratazonosító). Ez közel háromezer (3000) periodikát jelentett. Ugyanilyen módon történt a sorozatok, valamint a szakrendi nevek azonosítása is.

A szakrendi nevek a korábbi adatbázis struktúrájában nagyon fontos szerepet játszottak, az adott szakcikkhez kapcsolódó személyneveket tartalmazták. Ezek egy része már bent volt az ITIdata-ban, a többit létrehozták a munkatársak, az ITIdata személynév-létrehozása szabályai szerint. Itt is az xml-dumpból kinyert listából dolgoztak Excel-táblázatokban, a létrejött személynév rekordja pedig a fenti módon visszakerült az xml megfelelő mezőjéhez. Közel kétezer (2000) személynév azonosítása, majd rekordjuk létrehozása történt meg.

A korábbi adatbázisban a munkatársak egy adott műfajlista alapján dolgoztak, de ez nem volt teljes átfedésben az ITIdata-ban használatos műfajokkal, ezért ezek harmonizációját is végre kellett hajtanunk. Az új gyűjtés és adatrögzítés során már egy új műfajlista alapján dolgozhat a Bibliográfiai osztály, amely teljes mértékben az ITIdata műfaji felosztását követi.

Az előkészítő munkák másik része ezt követően zajlott, szintén kilistázott adatok alapján: a kiadás helye, a kiadói és szerzői nevek ellenőrzése és feloldása. A rövidített névváltozatok feloldása sok esetben nem volt egyszerű, néha az eredeti dokumentumig kellett visszanyúlni, mert a nyomtatott változatban sem volt egyértelmű a metaadat. A szerzői neveknél még betöltés előtt a listák alapján lehetett javítani az elgépeléseket, elírásokat. Itt utómun-

² Jelenlegi rekordszám: 483 086 (2024. október 1.) Vö. Dobás Kata – Fellegi Zsófia – Palkó Gábor, *A kis gömböc meséje – az ITIdata irodalomtudományos adatbázis fejlesztése 2022–2023-ban*, In: Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé: 32. Workshop országos konferencia: 2023. április 12-14. Pannon Egyetem, Veszprém, HUNGARNET Egyesület, Budapest, 2022, 192–198.

kálat is zajlik: a betöltést követően is javítják a Bibliográfiai osztály munkatársai az elírt neveket, az egyéni szerzői névhasználat egy része is feloldhatóvá válik így (pl. Csokonai). Az előkészítő munka közel egy évig tartott, és nagy tanulsága volt a munkafolyamatnak, hogy az egyes mezők kilitázása számos olyan hibát is láthatóvá tett, amivel valószínűleg egyébként soha nem találkoztunk volna (pl. nem létező folyóiratok, amelyekhez mégis rendeltek bibliográfiai metaadatokat).

A betöltés Python-nyelven írt betöltő kóddal történt, a Wikidata Integrator Python-csomag testre szabásával.

A dokumentáció fontosságát külön ki szeretném emelni. A mapping és a betöltő kód készítése során számos olyan információ cserélt gazdát, aminek a rögzítése egy későbbi migráció esetében nagyon hasznos lesz. Minden munkafázisról írtunk részletes leírást, a felmerülő kérdések és válaszok szintén rögzítésre kerültek, ahogy a különböző munkafázisok dokumentumai is. Az egyplatformos kommunikációt a DigiPhil csapaton belül és az Irodalomtudományi Intézet munkatársaival is a Teams biztosítja, itt tároljuk a migráció különböző dokumentumait is, ami így a csapat egészének hozzáférhető.

2. Ami kimaradt...

14 246 rekordot hibásnak észlelt a betöltő kód, így azok nem kerültek át az adatbázisba. Ezek részletes vizsgálata folyamatban van, egyelőre két nagy hibalehetőség rajzolódott ki. Számos olyan cikk került gyűjtésre, amelynél a cím az adott lap rovatcímének felel meg, vagyis számról számra ugyanaz. Az ITIdata Description/Leírás részében a következő adatok kerülnek rögzítésre: szerző, műfaj, bennfoglaló mű, évszám. Ebben az esetben a rekord előbb említett Leírás részénél kell majd felvegyünk olyan adatokat is (lapszámok, oldalszámok megjelölése), amelyek elkülönítik egymástól a cikkeket.

A nyomtatott kötet struktúrájából fakadó hiányos rekordok: a kiadványok olyan módon épültek fel, hogy egy-egy tárgyalt szerző adott művénel nem kerültek rögzítésre az ismeretések (kritikák, recenziók) címei, ami elvileg ugyanaz lett volna, mint a műcím, amiről a recenzió, kritika íródott. Így egy látszólag hiányos rekord jött létre. Ezek utólagos korrekcióját a szakrendi besorolás teszi majd lehetővé.

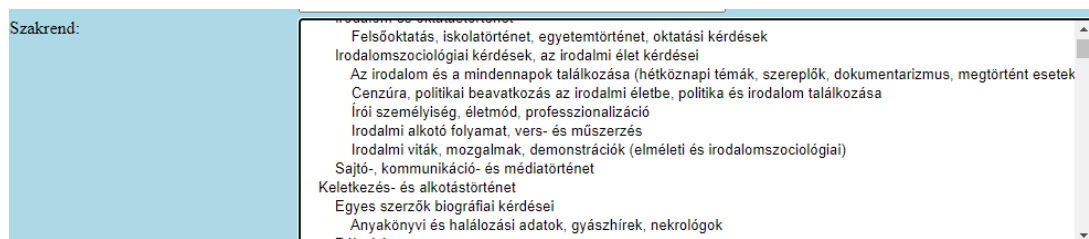
A betöltést követően többféle javításra is szükség lesz (hibásan felismert karakterek, lemaradt oldalszámok, kötetek esetében hiányos bibliográfiai adatok stb.), ezek szűrését listák alapján végezzük, és Python-kóddal javítjuk. Ezen felül folyamatosan javítjuk a korábbi adatbázis apróbb hibáit (elgépelések, indokolatlan duplumok), hogy az ITIdata-ban már helyesen szerepeljenek.

Összesen 108 593 MIB-rekord került betöltésre a migráció során.³

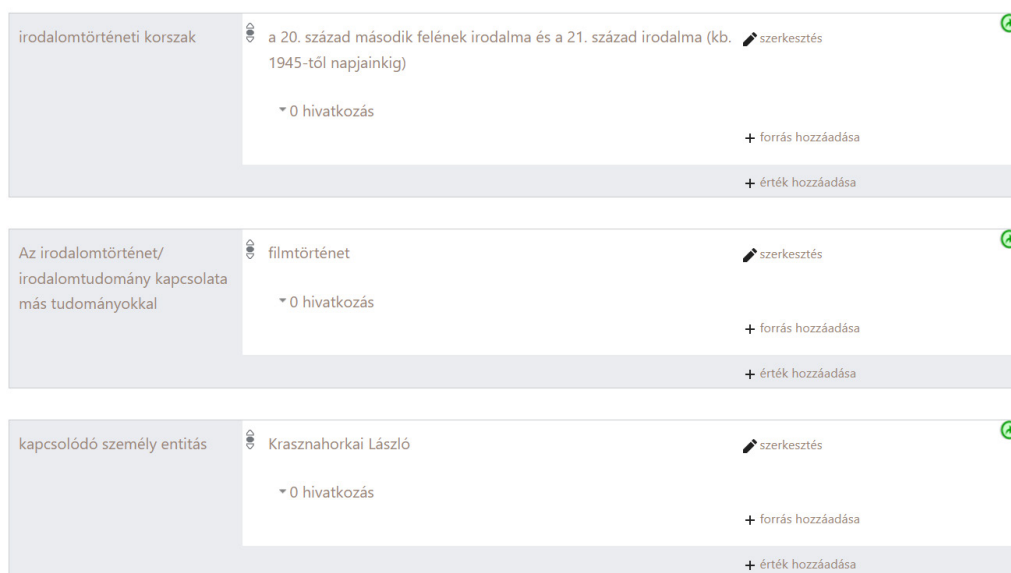
3 A projekthez fejlesztett kereső élesítése folyamatban van, a MIB-rekordok az alábbi linken található SPARQL-lekérdezés után hozzáférhetőek: <https://tinyurl.com/296qtgq4>

3. A szakrend

A korábbi szakrendi besorolások természetesen átkerültek az ITIdata-ba is, a szemantikus struktúra azonban új, másfajta szakrendi besorolást tett lehetővé. Ennek kialakítása során figyelembe vettük a korábbi szakrendet, de igyekeztünk olyan szerkezetet kialakítani, ami már az ITIdata-ban, vagyis egy szemantikus struktúrában való kereshetőséget tartja szem előtt.



5. ábra: Részlet a régi szakrendből
Forrás: <http://www.iti.mta.hu/mib/>



6. ábra: Részlet az új szakrendből
Forrás: ITIdata, https://itidata.abtk.hu/wiki/Main_Page

Az ITIdata-n belül lehetőség van arra, hogy az adatrögzítést segítsük ún. megkötésekkel (a WikiData-hoz hasonló módon). A gyakorlatban ez azt jelenti, hogy a nem megfelelő tulajdonság kiválasztásánál a rendszer felkiáltójelet tesz az állítás mellé. Vagyis születési helyet nem lehet felvinni pl. egy szakcikk metaadataihoz. Ugyanezt az eljárást alkalmazzuk a szakrendnél is, illetve ahol nincs sok tétel a hierarchiában, vagyis még emberi számára átfogható mennyiség, ott legördülő listából lehet kiválasztani az adott szakrendi kulcsszót.

4. Használhatóság, kereső

Minden ilyen projekt felhasználói környezetbe érkezik, ez talán kevésbé szokott hangsúlyt kapni, pedig nagyon fontos tényező, ezért az új gyűjtéshez tartozó próbarekordokat a sandboxunkban⁴ próbálták ki a munkatársak az új szakrendi besorolással együtt.

Az ITIdata-ban a keresés SPARQL lekérdező nyelven történik, ami nagyon összetett keresésekre is lehetőséget biztosít. Ennek elsajátítását azonban – és itt megint visszautalnék az emberi, kutatói, felhasználói környezetre – nem kértük a kutatóktól, lévén nem a legegyszerűbb, legmegszokottabb módja a keresésnek. Külön keresőt fejlesztettünk a projekthez, ami jelenleg tesztfázisban van, de egy minden felhasználó által könnyen kezelhető keresőplatformot eredményez majd.

4 https://sandbox.itidata.dh-lab.hu/wiki/Main_Page