

Wikidata, a többnyelvű közösségi közreműködésen alapuló tudásgráf – 2. rész

Az integráció és az alkalmazhatóság lehetséges módozatai

Molnár Bence

A kétrészes tanulmányorozat a Wikidata felépítését, szemantikus weben betöltött szerepét és lehetséges alkalmazási módozatait vizsgálja nemzetközi példák bemutatása és szakirodalmi áttekintése által. A szándék kettős: magyar nyelven íródott szakirodalom hiányában átfogó és teljes áttekintést nyújtani a Wikidatáról, összefoglalva a tudásgráffal kapcsolatos legfontosabb kutatási eredményeket, másrészt bemutatni a Wikidata integrálásának megvalósult jógyakorlatait és lehetséges alkalmazásait, különös tekintettel a közgyűteményi szereplők szempontjából. A 2. rész megvalósult együttműködések által szándékozik igazolni a Wikidata relevanciáját a természetes nyelvi feldolgozáson alapuló szövegalkotás, valamint a kapcsolt nyílt adatok elveinek megfelelő közgyűteményi adatbázisok kialakítása terén, bemutató a tudásgráf többnyelvű adatok vizualizálásában rejlő potenciákat.

besorolási adat, besorolási adatok egységesítése, gépi dokumentumleírás, SPARQL, természetes nyelvi feldolgozás, szemantikus web, tudásgráf, Wikidata

1. Bevezetés

A tanulmányorozat első részében [1] áttekintésre kerültek mindazon fejlesztési célok és lépések, amelyek a Wikidatát rövid idő alatt a szemantikus web egyik legfontosabb adatbázisává tették. A tudásgráf tartalmát felépítő RDF tripletekhez szükséges alapfogalmak mellett bemutatásra kerültek az adatok (újra)felhasználásához elengedhetetlen forráshivatkozások és azonosítók, valamint a visszakereséshez szükséges címkék és leírások. Igazolást nyert a Wikidata többnyelvű közössége által előállított adatok relevanciája a tudás megosztásának és demokratizálásának globális folyamataiban, azonban az ezt gátló, a szerkesztőközösség, valamint a címkék és leírások terén tapasztalható nyelvi egyenlőtlenségek is feltárásra kerültek. A bárki által szabadon szerkeszthető Wikidata – és a Wikipédia – fennmaradása és sikere által gyakorlati igazolást nyert közösségi közreműködésen alapuló modell működésének gyakorlati példaként történő demonstrálása által megismertetésre kerültek az adatbázis szerkesztésének manuális és (fél)automatizált módozatai is.

A történeti áttekintésben tárgyalt kezdeti célkitűzések technikai megvalósítását követően a Wikidata rövid idő alatt a Wikimedia Foundation megkerülhetetlen projektjévé vált, elsősorban azért, mert általa megvalósult a wiki- és nyelvközi hivatkozások régóta szükséges egységesítése. Ennek sikeres lezárultával kezdődhetett el a Wikidata tudásgráffá fejlődése és elszakadása a társprojektek – döntően a Wikipédia – ökoszisztémájától. Noha a Wikidata elsődleges feladata továbbra is a WMF által fenntartott társprojektek hivatkozás- és adatintegrációjának fenntartása, azonban mára tartalma és mérete okán képes külső felek – így például könyvtárak – igényeit is kielégíteni. A következőkben a tanulmány a Wikidata alkalmazhatóságát kívánja igazolni már megvalósult jógyakorlatok és lehetséges alkalmazási módozatok bemutatásával, hogy a Wikidata miképp lehet képes részlegesen kiváltani a közgyűteményi szereplők hagyományos authoritykontrollját, lehetővé téve ezáltal az intézményi adatsilókba zárt adatok nyílttá és szabadon elérhetővé tételét, illetve hogy a többnyelvű adatok milyen információ-

visszakeresési és -vizualizálási módokat kínálnak akár a természetes nyelvi feldolgozáson alapuló szövegalkotás terén.

2. Természetes nyelvi feldolgozáson alapuló szövegalkotás

A természetes nyelvi feldolgozás vagy természetesnyelv-feldolgozás (natural language processing, NLP) a számítástudomány, a mesterséges intelligencia és a nyelvészet azon területe, amely az ember és a számítógép közti interakcióval foglalkozik [2]. A tudásgráfokat és -bázisokat számos terület, így például a természetes nyelvi feldolgozás is alkalmazza, mert a természetesnyelv-feldolgozásban érdekelt közösségek kutatásai nagy adatigényűek. Azok létrehozásának idő- és anyagi vonzatai miatt azonban csak kevés nagyméretű és szakszerű módon annotált korpusz áll rendelkezésre [3]. A Wikidata mérete, komplexitása és szabad hozzáférhetősége okán egy ilyen korpusz. A természetes nyelvi feldolgozás a Wikidata és a Wikipédia között a két projekt szoros integrációja okán számos formában ölthet testet. Ennek már bevett és alkalmazott módját egy 2018-as tanulmány is vizsgálta, egy olyan strukturált és strukturálatlan adatok közötti leképezést lehetővé tévő neurális modell felállítását szorgalmazva, amely csökkenti a Wikipédiában fellelhető nyelvi egyenlőtlenségeket azáltal, hogy a szócikkkel még nem rendelkező témákról a Wikidata adatait felhasználva automatikusan generált összefoglalót kínál az olvasóknak. Továbbfejlesztve a MediaWiki szoftver ArticlePlaceholder nevezetű kiegészítőjét, amely egy entitás Wikidatában tárolt adatainak strukturált és dinamikus megjelenítését teszi lehetővé [4], a kutatók egy olyan rendszert alkottak, amely a tripletek megjelenítésén túl azokból emberi olvasásra szánt szöveges összefoglalók generálására is képes. Az így kapott eredményeket a kutatásban részt vevő Wikipédia-szerkesztők a szövegeket a minőségi elvárásoknak közel megfelelőnek ítélték, lehetővé téve azok többségének esetleges újrafelhasználhatóságát [5].

Egy 2022-es tanulmány a Wikidata-állítások mondatszintű természetes nyelvi szöveggé történő konvertálásának lehetőségeit vizsgálta az angol Wikipédia számára [6]. A projekt keretében a szócikkek számára mondatokat generáltak

a laphoz kapcsolt Wikidata-elem állításaiból, majd az így kapott kimeneti szövegkorpuszt különböző szempontok szerint értékelték, például mondat-szerkesztés-elemzés és zajsűrés segítségével. A leképezés esetleges komplikációinak csökkentése érdekében a projekt nem összefoglalók vagy bevezetők készítését, hanem önálló – tárgyias, ígét esetlegesen tartalmazó – mondatok előállítását célozta a szócikkek számára. Mivel egy RDF triplet természetéből fakadóan általában csak rövid (tő)mondattá alakítható, ezért a mindennapi írott nyelvhez jobban illeszkedő, hosszabb és teljesebb mondatok készítése érdekében a szerzők különböző RDF tripletek kombinálását javasolták. Míg a tripletek három komponensből, egy alanyból, egy állítmányból és egy tárgyból állnak (konvencionálisan ebben a sorrendben írva), addig a quadok – további kiegészítő tulajdonságokkal rendelkező tripletek – kontextussal is szolgálnak, lehetővé téve a Wikipédia-szócikkekben is használható hosszúságú mondatok előállítását. A tanulmány által bemutatott modell alkalmazhatóságának lehetőségeit mutatja – noha a szerzők nem említik –, hogy a Wikimedia Foundation jelenleg még fejlesztés alatt álló „Abstract Wikipedia” elnevezésű projektje egy strukturált adatokra épülő nyelvfüggetlen Wikipédia megteremtését célozza [7].

A modell egy sokkal egyszerűbb implementációja a Reasonator, amelynek célja a Wikidata adatainak összefogása és elérhetővé tétele az emberi felhasználók számára [8]. Az eszköz dedikált célja a Wikidata olvasási és böngészési élményének növelése akképpen, hogy száműzésre kerülnek a szerkesztést lehetővé tévő gombok és mezők, ezáltal a tartalom számára jelentős képernyőfelületet felszabadítva. A Reasonator egy önálló, független (self-contained) JavaScript osztály HTML és CSS támogatással, amely a Wikidata API-ját és SPARQL (Protocol and RDF Query Language) nyelven írt lekérdezéseket használ. A szolgáltatás a megtekinteni kívánt Wikidata-elem tulajdonságértékpárjai alapján egy automatikusan generált összefoglalót is készít az entitásról, továbbá láthatóvá teszi az elemek között meglévő szemantikus kapcsolatokat, így például a rokonsági, szerzőségi, névadói vagy egyéb kapcsolatokból fakadó hivatkozásokat. Alapvető különbség a Wikidata felhasználói felülete és a Reasonator között, hogy

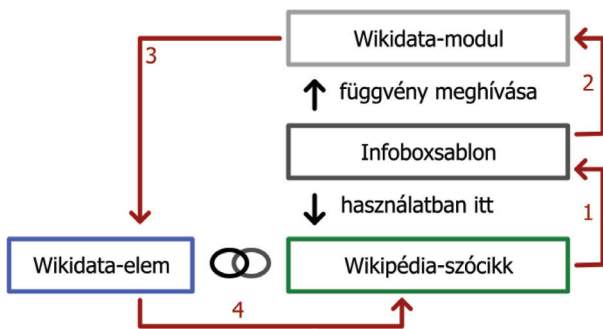
utóbbi kontextusfüggetlenül támogatja az entitástípus-specifikus információk megjelenítését, így például személyek elemeinél megjeleníti az elemek egymásra mutató hivatkozásaiból kinyert családfáját, földrajzi helyek esetében azok kontinensi osztályig tartó hierarchiáját, fajok esetében például azok rendszertanát. Az eszköz támogatja továbbá a többnyelvű adatok megjelenítését. A jobb felső sarokban lévő nyelvválasztó menüpont segítségével több száz nyelv közül választhat az olvasó, ezt követően a Reasonator által mutatott – Wikidata-elemek értékeiből és címkéiből származó – információk már a kiválasztott nyelven töltődnek be amennyiben az érintett Wikidata-elemek nyelvi címkéi szerint ez lehetséges [9, 10].

Talán nem túl meglepő, hogy a Wikidata strukturált adatainak legprominensebb és legelterjedtebb felhasználása mégis a Wikipédia-szócikkekben történik. A történeti áttekintésben említett, a szócikkek írásirányának megfelelő (jellemzően jobb) oldalán elhelyezkedő infoboxok a szócikk bevezetőjében rövid összefoglalással szolgálnak a téma legfontosabb információiról, így például földrajzi helyek elhelyezkedéséről, személyek életrajzi adatairól. A fejlesztés második fázisában megfogalmazott cél – a Wikidata képében egy központi adattér létrehozása a különböző nyelvű Wikipédia-projektek infoboxaiban tárolt információk számára – az állítások 2013-as bevezetésével valósult meg, megnyitva az utat a Wikidatán alapuló infoboxsablonok létrehozása előtt. A Wikipédia mögött húzódó MediaWiki szoftverkörnyezetben a nyílt forráskódú Lua beágyazható szkriptnyelven írt modulok segítségével lehetséges Wikidata-kompatibilis sablonok készítése. Mivel a Wikipédia különböző szerkesztőközösségeinek tagjai nem feltétlenül rendelkeznek programozói készségekkel, a Wikidata fejlesztői a kezdetektől törekedtek a folyamat leegyszerűsítésére, felhasználóbarát dokumentálására [11].

A wiki- és nyelvközi hivatkozások centralizálását követően a legtöbb Wikipédia-szócikk hozzá lett kapcsolva egy Wikidata-elemhez, közvetlen kapcsolatot teremtve az adott szócikk és az entitást reprezentáló „Q” prefixű elem között. Ez lehetővé tette, hogy a szócikkbe ágyazott sablonok közvetlenül is meghívassák a lap Wikidata-elemének értékeit a Wikidata-elemazonosító manuális meg-

adása nélkül, például a magyar Wikipédián azáltal, hogy a Wikidata-modult használó sablonok alapértelmezetten a szócikkhez kapcsolt Wikidata-elem tulajdonságainak értékét töltsék be. Amennyiben szükséges, a Wikidata-elemazonosító alapértelmezett betöltődése manuálisan felülírható, így például a magyar Wikipédia Wikidata-modulja lehetővé teszi azt is, hogy az „entityId” paraméter segítségével a modul a szócikkhez kapcsolt Wikidata-elem helyett a paraméter értékeként megadott Wikidata-elemből, a „pageTitle” által pedig a megadott lapcímhez tartozó Wikidata-elemből olvassa ki a tulajdonsághoz tartozó értéket (a modul releváns kódrészletét az *1. számú melléklet* tartalmazza). A két paraméter egyidejű jelenléte esetén a modul hibaüzenettel tér vissza [12]. A „Wikidata” elnevezésű sablon implementációjaként működő modul számos lehetőséget kínál a Wikidatában tárolt adatok kinyerésére [13]. Legegyszerűbb formájában lekérhető általa egy Wikidata-elem tulajdonságának értéke vagy értékei, így például *Douglas Adams* születési dátuma (1952. március 11.) az alábbi wikikód segítségével nyerhető ki az író szócikkébe beillesztve: „{{Wikidata|P569}}”. Amennyiben ugyanezen az oldalon lenne szükség a Galaxis útikalauz stopposoknak megjelenésének évére (1979), az az előzőekben említett paraméterek megadásával, a „{{Wikidata|P577|pageTitle=Galaxis útikalauz stopposoknak}}” vagy a „{{Wikidata|P577|entityId=Q25169}}” wikikódok használatával közvetlenül is lekérhető, annak ellenére, hogy ez az információ Douglas Adams szócikkhez kapcsolt Wikidata-adatlapon explicite nem szerepel.

A Wikidata-adatokat meghívó sablonok a gyakorlatban az *1. ábrán* látható módon működnek: például amikor az olvasó felkeres egy Wikidata-modult használó infoboxsablont tartalmazó Wikipédia-szócikket, akkor előbb az infobox lokális mezőiben tárolt adatok értékelődnek ki, majd csak ezt követően kerülnek lekérésre a „{{#invoke:Wikidata|...}}” varázsszóval (pontosabban parser függvényel) meghívott értékek (1). Minden érték esetében meghívásra kerül az adott projekt Wikidata-modulja (2), amely aztán szintén meghívja a szócikkhez kapcsolt Wikidata-elemet (3), amely végül megküldi a kért adatokat a Wikipédia-szócikkben szereplő infoboxsablonnak (4).



1. ábra Wikidata-modult használó Wikidata-kompatibilis sablonok működésének egyszerűsített folyamatábrája: a Wikidata-elemmel összekapcsolt Wikipédia-szócikk infoboxsablonja a Wikidata-modul segítségével meghívja a Wikidatában tárolt értékeket, amelyek végül megjelennek az olvasó számára

Szerző: Tobias1984, CC BY SA 4.0 licenc alatt, a szerző fordításával

Forrás: https://commons.wikimedia.org/wiki/File:Wikidata_Infobox_Tutorial_-_Infoboxes_using_mostly_Module-Wikidata.svg

A Wikidata indulását megelőzően a Wikipédia-projektek strukturált és félig strukturált adatai a szócikkek törzsszövegében jellemzően infoboxsablonok, táblázatok, listák és kategóriák képében voltak jelen. Ezen adatok karban- és naprakészen tartása azonban jelentős problémákkal járt a szerkesztőközösségek számára. Mint arra egy 2018-as tanulmány is rámutatott, ha egy híres – számos nyelven Wikipédia-szócikkkel rendelkező – labdarúgó gólt szerez egy nemzetközi mérkőzésen (például egy világbajnokságon), akkor a gólszerzés tényét nem csak az adott játékos szócikkének releváns infoboxában, hanem a bajnokság és a játékos csapatának gólszerzőit tartalmazó táblázatában is frissíteni szükséges [14]. Annak fényében pedig, hogy a Wikipédia több mint 330 nyelven elérhető [15] és a különböző projektek szerkesztőközösségének mérete és aktivitása széles spektrumon mozog, könnyen belátható, hogy egy szimpla gólszerzés szerkesztések százait indukálhatja. Ahol azonban ez a frissítés elmarad, mert például a közösség önkéntes humánerőforrás-kapacitása szűkös, a tartalom újdonságértéke és pontossága csorbul [16].

Mindezeket felismerve a Wikipédia számos nyelvváltozata Wikidata-értékeket meghívó sablonok által aktívan támaszkodik a Wikidatában elérhető adatokra. Mivel a közösség által meghatározott nevezetességi és formai irányelvek mentén

bármilyen témában születhetnek szócikkek a Wikipédián, könnyen belátható, hogy nem létezhet egy univerzális, minden téma leírására alkalmas infobox, szükséges azok differenciálása. A magyar Wikipédián a generális „Személy infobox” mellett létezik „Vezető infobox” és „Zenész infobox” is, de például a „Település infobox” is rendelkezik országspecifikus – magyar, francia vagy épp ugandai – változattal is. A Wikidatában elérhető információkat használó sablonok – a magyar Wikipédia esetében több mint kétszáz ilyen létezik [17] – legnagyobb előnye, hogy azok Wikidata-kompatibilissé tett paramétereit nem szükséges lokálisan kitölteni, hogy a kívánt érték (például a település irányítószáma) láthatóvá váljon az olvasó számára: amennyiben ugyanis ez a laphoz kapcsolt Wikidata-elemből is kiolvasható, a sablon emberi közreműködés nélkül képes megjeleníteni azt. Hogy mindez a gyakorlatban milyen haszonnal is jár a szerkesztőközösség és az olvasók szempontjából, jól példázza a francia Wikipédia személyekhez használt „Infobox Biographie2” elnevezésű sablonja és modulja [18]. Szentágothai János magyar anatómus és egyetemi tanárnak a tanulmány leadásáig még nem készült szócikke a francia Wikipédián: amennyiben bárki belefogna ennek megírásába, a szócikk bevezetőjén túl egy infoboxot is kellene hozzá készítenie Szentágothai életútjáról és kiterjedt munkásságáról. Ahelyett azonban, hogy a dokumentáció áttanulmányozása után kezel kellene kitölteni a sablon paramétereit, a sablon wikikódjának („{{Infobox Biographie2}}”) szimpla beillesztésével a 2. számú mellékletben látható komplett infobox kerülne automatikusan legenerálásra. A sablon Szentágothai Q662084 Wikidata-azonosítószámú elemének legtöbb releváns adatát megjeleníti, így például – az infobox strukturális sorrendjét követve – a viselt tisztségeket és minősítőiket („Fonctions” szakasz), illetve bibliográfiai és egyéb adatokat („Biographie” és „Autres informations” szakasz). Ezen utolsó szakasz tartalmazza többek között a Szentágothai tanulmányaira és szakintézeti tagságára vonatkozó információkat, valamint kitüntetései és díjainak listáját is. A sablon mögött húzódó modul az olvasók és a szerkesztők szempontjából több kényelmi funkcióval is rendelkezik, így például képes megjeleníteni a francia nyelvű címekkel nem rendelkező

Wikidata-elemeket: ahol nem érhető el a projekt nyelvének megfelelő címke, ott alapértelmezetten előbb az angol nyelvű címke (például a harmadik magyar köztársaság első, 1990 és 1994 közötti parlamenti ciklusa esetében), ennek hiányában a rendelkezésre álló címkék egyike kerül megjelenítésre (például a Tényeket Tisztelők Társasága esetében). A más szócikkre mutató – belső hivatkozással – nem rendelkező entitások esetében a sablon vagy az adott Wikidata-elemre vezető „(d)”, vagy az angol Wikipédia szócikkére mutató „(en)”, félkövéren szedett kattintható hivatkozással teremti meg a hipertextualitást. Az infoboxban megjelenő, Wikidatából származó – és ezért lokálisan közvetlenül nem módosítható – információk szerkesztését az állítások mellett egy közvetlen ugrópontként funkcionáló kis ceruza ikon teszi lehetővé, ami közvetlenül a Wikidata-elem megfelelő tulajdonságára viszi a rákattintó felhasználót, és még a Wikidata felhasználói felületének franciára állításáról is gondoskodik [19]. Bizonyos esetekben a modul a Wikidatából meghívott adatokat helytakarékossági megfontolásból „dobozolja”, egy alapértelmezetten összecsukott, kattintásra kinyíló listába rendezi. Példánk esetében csak a legfontosabbnak rangsorolt Kossuth-díj jelenik meg a „Distinctions” sorban. Végezetül az infobox a szócikk vizuális és forrásszövegű szerkesztésére mutató hivatkozások („modifier” és „modifier le code”) mellett a Wikidata-elemre és a modul dokumentációjára is kínál elérési linket.

Léteznek azonban olyan, a Wikimedia Foundation által üzemeltetett ökoszisztémán kívül található infoboxok is, amelyek a Wikidata adataira támaszkodnak. A kategória legprominensebb szereplője az Alphabet és az általa birtokolt Google-keresőmotor. Keresés során a Google bizonyos könnyen értelmezhető és megválaszolható kérdésekre – például, hogy mikor született egy bizonyos híresség, vagy hogy mikor alapítottak egy adott intézményt – azonnal válasszal is szolgál az oldal tetején, még a találati lista első eleme előtt. A Google továbbá – a Wikipédia-szócikkek felépítéséhez hasonlóan a felhasználói felület nyelvének írásrendszeréhez igazodva jobb vagy bal oldalt – egy tudáspanelnek (Knowledge Graph) nevezett infoboxban foglalja össze az entitás legfontosabb információt. Az utóbbira vonatkozó legfris-

sebb, 2020 májusában publikált beszámoló szerint a Google a szolgáltatás 2012-es elindítása óta 5 milliárd entitásra vonatkozóan 500 milliárd állítással rendelkezik [20]. Mindkét funkció közös jellemzője, hogy információik egy részét a Wikidata strukturált adataiból nyerik, a felhasználói felület nyelvvel megegyező adatok természetes nyelvi szövegalkotás által kerülnek generálásra [21]. *Bibor Máté* Q113145838-as azonosítószámú Wikidata-eleme például 2022-ben egy félautomata eszközzel került létrehozásra, aminek információit a Google tudáspanelje is felhasználja [22, 23]. A természetes nyelvi feldolgozás és -szövegalkotás kapcsán említett érdemelnek még a virtuális asszisztensek, amelyek szintén támaszkodnak a Wikidatára, azonban érdekességüket az adja, hogy mind az input, mind az output élőbeszédű szöveg. Az Amazon virtuális asszisztense, az Alexa esetében a fejlesztőknek a japán zenei katalógusok kapcsán azzal a problémával kellett megküzdeniük, hogy a különböző írásrendszereket használó nyelvek esetében a transzkripció több esetben téves eredményhez vezethet, mivel egy japán katalógusban a latin mellett hiragana, kandzsi és katakana írásrendszerű dalcímek is előfordulhatnak. A probléma megoldására fejlesztett gépi tanuláson alapuló többnyelvű transzkripció és -literáció modelljükhöz a fejlesztők például a Wikidata adatbázisát vették alapul [24]. De az Apple *Siri* elnevezésű virtuális asszisztense is nagyban támaszkodik a Wikidatára – néha túlzottan is. 2018-ban bejárta a világot a hír, hogy a Siri tévesen állította felhasználóknak, hogy elhunyt *Stan Lee*, a Marvel-univerzum híres képregényírója. Mint utóbb kiderült, a hiba abból eredt, hogy egy szerkesztő beszúrt egy P570 (halálozási idő) tulajdonság-értékpárt *Stan Lee* Wikidata-elemébe, amit aztán a Siri igaznak feltételezett. Ez az eset is rávilágított arra, milyen fontos az adatok forráshivatkozásokkal történő alátámasztása, az adatok validitásának kontrollja által a tudásgráf struktúrájának és hitelességének megőrzése, az applikációk oldaláról pedig az adatok kritikus és kontrollált átvétele [25]. Mint minden ekkora adatbázis, a Wikidata információi sem minden esetben teljeskörűek és pontosak, azonban a fejlődés íve kimutatható és töretlen, ami lehetővé teszi az adatok felhasználását a természetes nyelvi feldolgozásban és -szövegalkotásban – a Wikipédián és azon túl.

3. Azonosító- és adatcserén alapú authoritykontroll

A 2010-es éveket megelőzően a szemantikus web jelentette technológiai lehetőségek csak korlátozottan jelentek meg a könyvtári szabványokban és integrált könyvtári rendszerekben. A szemantikus web adaptációja a könyvtárak technikai lehetőségei okán kezdetben csak rendkívül lassú ütemben valósult meg – állapították meg 2018-ban kanadai szerzők [26]. A magyarországi könyvtárak általános (számítás)technikai felszereltségi szintjének és anyagi lehetőségeinek ismeretében feltételezhető, hogy a hazai helyzet a nemzetközi környezethez hasonlóan, ha nem rosszabbul alakult. A Wikidata elsősorban mint közösségi közreműködésen alapuló, adatainak (újra)felhasználását szabad licenc által garantáló tudásgráfként kerül a könyvtárak és közgyűjtemények fókuszába, mint a metaadatok megosztásának és cseréjének egy lehetséges platformja. A közgyűjtemények számára a Wikidata különösképp authoritykontroll-tevékenységük terén rejt kiaknázható lehetőségeket, ugyanis képes az intézmények authority adatainak központi csomópontjaként is szolgálni, lehetővé téve az egységesített besorolási adatok egyértelműsítését és összekapcsolását, linkelését, kitorvve ezáltal a tradicionális diszciplináris adatsilók korlátaiból, megteremtve annak lehetőségét, hogy a könyvtárak és hasonló intézmények a jövőben áttérjenek a Wikidata-azonosítók használatára forrásdokumentumaik és authority rekordjaik alapértelmezett azonosítóiként [27, 28].

A könyvtárak által legszélesebb körben használt adatcsere-formátum, a Machine Readable Cataloguing (MARC) hátránya ugyanis, hogy noha a könyvek leírására kétségtelenül kipróbált és széles körben elfogadott formátum, nem alkalmas a GLAM közgyűjtemények (galleries, libraries, archives, and museums – galériák, könyvtárak, archívumok és múzeumok) eltérő és egyedi dokumentumai jelentette speciális igények teljeskörű kielégítésére, mindezt pedig fokozza még az a tény is, hogy a könyvtári katalogizálás folyamata átalakulóban van: a szemantikus webre történő katalogizálás hatalmas lehetőségeket és előnyöket rejt magában. A közgyűjtemények zárt ökoszisztémájában létező adatsilók – adattárolásra szolgáló elszigetelt pontok vagy elszigetelt rendszerek egy tágabb rendszeren

belül [29] – ideje leáldozóban van. Az intézményeknek már most a közös szabványok lefektetését előrevetítő jövőbe szükséges tekinteniük arra a kérdésre keresve a választ, hogy miképp fogják (meta) adataikat nyílttá és szabadon elérhetővé tenni. [30] Az Online Számítógépes Könyvtári Központ (Online Computer Library Center, OCLC) szerint a „Wikidata/Wikibase a hagyományos authoritykontroll alternatívájának tekinthető, amely olyan lehetséges előnyökkel is rendelkezik, mint például a tartalom többnyelvűségének támogatása vagy épp az intézmény erőforrásait leíró adatsilók áthidalása” [31]. A Wikidata szorosabb integrálása a könyvtári – mindenekelőtt a katalogizálási – munkafolyamatokba, az integrált könyvtári rendszerekbe és nyilvános könyvtári katalógusokba azonban jelenleg még a jövő ígérete [32].

A való világ objektumainak leírására a szemantikus weben az egységes forrásazonosítók szolgálnak. Könyvtári környezetben az URI-k segítik az authoritykontrollt és az adatok géppel olvasható RDF-alapú formátumba való konvertálását, a külső adatok megjelenítésének lehetőségével pedig fokozhatják a felhasználók böngészési élményét. Utóbbi lehetővé teszi a gyorsabb katalogizálást és a részletesebb authorityrekordok készítését – például számos amerikai könyvtár a Kongresszusi Könyvtár leírásait veszi át részben vagy egészben. A személyazonosítók legnagyobb nemzetközi adatbázisai a már említett VIAF-on túl a Nemzetközi Szabványos Névezőazonosító (International Standard Name Identifier, ISNI) vagy épp az ORCID (Open Researcher and Contributor ID). Mivel az URI-k alkotják az RDF-modell fő nyelvét, az URI-k hozzáadása a MARC-rekordokhoz megkönnyíti a más adatmodellekre (pl. BIBFRAME) való későbbi átállást, ami azért is fontos, mert noha a MARC 21 a könyvtári katalogizálás bevett formátuma, nem kompatibilis a kapcsolt nyílt adatok (Linked Open Data) elveivel, csupán néhány almezőjében megengedett a rekordok URI-kkal való bővítése. Az URI-k MARC 21-be történő integrálása révén azonban az intézmények és a felhasználók már akkor is élvezhetik a nyilvános könyvtári katalógus külső forrásokkal történő összekapcsolásának előnyeit, amikor a könyvtár még nem áll készen a kapcsolt adatok környezetére való átállásra [33].

A jövőt illetően az amerikai Kongresszusi Könyvtár 2012-ben mutatta be a Bibliographic Framework (BIBFRAME) modellt, mint a bibliográfiai leírási szabványok kapcsoltadat-modellé történő fejlesztését célzó kezdeményezést. A projekt célja, hogy a könyvtárak különleges igényeit változatlanul kiszolgálva integrálja a bővebben vett információs közösséget. A modell a Functional Requirements for Bibliographic Records (FRBR) fogalmi modelljéből származik. 2012 és 2016 között világszerte számos könyvtár vett részt a BIBFRAME tesztelésének és véleményezésének procedúrájában, az első változatot 2016 áprilisában követte a 2.0-ás, tapasztalatok alapján átdolgozott verzió. A BIBFRAME 2.0 egy lehetséges megoldás arra, hogy a bibliográfiai metaadatok kijussanak az online katalógusokból a keresőmotorokig, növelve ezáltal a digitális gyűjtemények webes hozzáférhetőségét. A BIBFRAME adatcsere-keretrendszerként készen áll a MARC formátum leváltására a kapcsolt adatok világában – lehetséges kooperációkat teremtve olyan adatbázisokkal, mint például a Wikidata [34].

A Wikidata alkalmazása az authoritykontroll terén számos tanulmány elméleti és gyakorlati vizsgálatának tárgya volt. A tengerentúlon a Utahi Egyetem J. Willard Marriott Könyvtára például 2016-ban fogott bele a Western Name Authority File (WNAF) fejlesztésébe, amely az Egyesült Államok nyugati államaiban található intézmények digitális gyűjteményeinek metaadataiban előforduló személy- és testületi neveit gyűjti. A projektről készült tanulmányukban a szerzők a WNAF regionális névtér nemzeti és nemzetközi adatbázisokhoz való kapcsolásának lehetőségeit vizsgálták. Mivel a WNAF – általánosságban egy regionális adatbázis – esetében felülreprezentáltak a helyi relevanciával bíró rekordok, így nem meglepő, hogy kevés elem rendelkezett saját azonosítóval a szövetségi Kongresszusi Könyvtár Library of Congress Name Authority File (LCNAF) névadatbázisában. A projekt keretében először összevetették a WNAF rekordjait a LCNAF elemeivel, az így kapott közös halmaz entitásaival aztán kétféleképp dolgoztak tovább. Elsőként olyan feltételezhetően már elhunyt – 1918-nál korábban született, vagy 70 és 100 közötti életkorú – személyek LCNAF-rekordjaira szűrtek, ahol nem volt megadva halálozási dátum, majd megpróbálták beazonosítani azon, időközben elhunyt

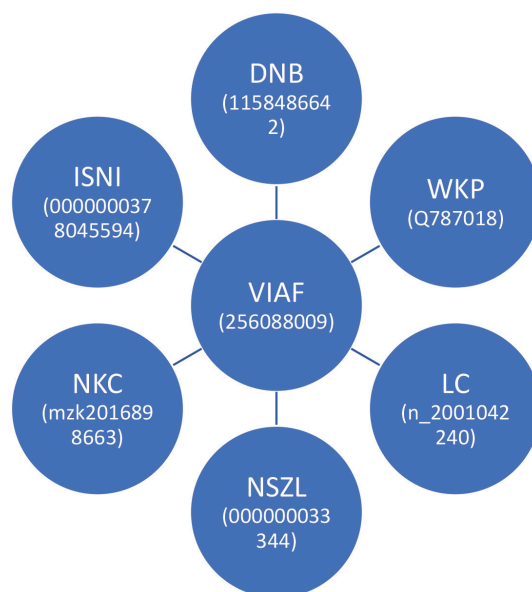
személyeket, akiknek katalógusrekordja frissítésre szorult. A projekt másik ága keretében olyan legalább három gyűjteményben szereplő WNAF-rekordokat kerestek, amelyek még nem rendelkeztek LCNAF-azonosítóval, hogy aztán azokat kiegészítve felvételre javasolhassák a nemzeti névtér számára. A Wikidata-integráció lehetőségeit vizsgálva hasonló módon jártak el, olyan WNAF-személyneveket hasonlítottak össze a Wikidata adatbázisával, amelyeket legalább három intézmény használt. Ennek folyamán egyrésztől megfigyelték, hogy a LCNAF-azonosítók megléte nagyban segítette a pontos egyezés biztos megállapítását, másrésztől az így kapott Wikidata-elemek elemzésekor arra jutottak, hogy a Wikidata-elemek döntő többsége az egységesített besorolási adatok minőségének javítása szempontjából hasznos bibliográfiai információkat tartalmaztak. A különféle adatbázis-azonosítók (LCNAF, VIAF, ISNI) megléte és a WNAF-azonosítók importálása lehetővé tenné még továbbá a Wikidata által már lefedett adatbázisokhoz való kapcsolódás lehetőségét is – zárták a tanulmányt [35]. A bibliográfiai adatok Wikidatából történő importálásának kérdése azért is releváns, mert visszautal a Wikidata-elemek állításainak ellenőrizhetőségének kérdéskörére. Noha a Wikidata a Wikipédiához hasonlóan másodlagos és harmadlagos forrásokra támaszkodik, a forrásmegjelölés nélküli állítások esetleges átvétele élő személyek esetében különösen problematikus tud lenni. Azon jelenleg, miszerint élő személyek kéri bizonyos információk kihagyását authority rekordjaikból nem újkeletű: egy 2020-ban készült, katalógizáló könyvtárosok kortárs szerzőkkel történő név-authority tevékenységre irányuló kommunikációját vizsgáló tanulmány amerikai válaszadóinak 60%-a válaszolt igennel arra a kérdésre, hogy a szerzők kérték-e már bizonyos információk kihagyását rekordjaikból (pl. születési idő, névváltozatok, lakhely, nem, elérhetőség). A tapasztalatok szerint álnéven publikáló szerző esetében előfordult például a valódi név közlésének megtagadása, de sokan születési idejük közléséhez sem járultak hozzá [36].

Világviszonylatban az előzőekben már említett Nemzetközi Virtuális Katalógustár a legjelentősebb példa a sikeres kétoldalú integrációra. Az Online Számítógépes Könyvtári Központ (OCLC) által kezelt authority adatbázis bővüléséhez közel 60 GLAM-

intézmény, adatbázis és szervezet járul hozzá. A 2003-ban indult projekt – amely 2012-től az OCLC szolgáltatásaként működik – célja a partnerek által beszolgáltatott, egymástól szeparáltan létező authority fájlok összekapcsolása egy egységes névrekordba, ezáltal a különböző entitásokról nemzeti szinten születő egységesített besorolási adatok duplikációjának felszámolása. A VIAF bővülése elsősorban a projektben részt vevő partnerek – a nemzeti könyvtárak mellett például a Wikidata – önkéntes adatszolgáltatásán alapszik [37].

A Wikidata és a VIAF közötti integráció előfutára volt, amikor 2012 nyarán az angol Wikipédia szócikkeibe közel negyedmillió VIAF-azonosító került automatizált módon importálásra. A folyamat során felmerült nehézségekről és a projekt tapasztalatairól a résztvevők esettanulmányban is beszámoltak [38]. Mivel azonban ezek az azonosítók egyedül az angol Wikipédia szócikkeihez kerültek hozzáadásra, és ezek szinkronizálása a különböző nyelvváltozatok között – hasonlóan a nyelv- és wiki-közi hivatkozások karbantartásához – akkor még megoldhatatlan kihívást jelentett, a projekt sikere csak korlátozott számú olvasót volt képes elérni. A Wikidata azonban megoldást kínált a külső azonosítók centralizált tárolására és disszeminálására a projektek között, így hamar kialakult a VIAF-Wikidata-Wikipédia-integráció. 2023. szeptember 16-án a Wikidata „VIAF-azonosító” (P214-azonosítójú) tulajdonsága több mint 7,54 millió alkalommal, 3,29 millió esetben állításként (a személyeket leíró Wikidata-elemek esetében 10. leggyakoribb tulajdonságként – lásd 2. számú táblázat), míg 4,24 millió esetben minősítőként volt használatban [39].

A VIAF-azonosítók felvitele a Wikidatába a közösségi közreműködésen alapuló modellt taglaló fejezetben ismertetett módszerek által részben automatizált módon történik. Azonban minőségügyi szempontból különösen fontosak a humán felhasználók által végzett szerkesztések, ugyanis az integráció egyik következménye, hogy a VIAF problémás – duplikált vagy hasonló nevű entások összekeveredett – elemei a számos importáláson szüntűgy átvételre kerülnek. A tévesen egy VIAF-azonosító alá keveredett azonos vagy hasonló nevű entások jelzésére egy dedikált Wikidata-lap szolgál, lehetővé téve a szerkesztők közösség és az OCLC munkatársainak a problémák



2. ábra Barátné Hajdu Ágnes VIAF-adatlapján szereplő intézményi rekordok és a hozzá tartozó azonosítósorszámok 2023. szeptember 16-án. A rövidítések feloldása: DNB=Német Nemzeti Könyvtár, WKP=Wikidata, LC=Kongresszusi Könyvtár, NSZL=Országos Széchényi Könyvtár, NKC=A Cseh Köztársaság Nemzeti Könyvtára, ISNI=Nemzetközi Szabványos Névazonosító
A szerző saját gyűjtése és szerkesztése
Forrás: <https://viaf.org/viaf/256088009/>

nyomon követését [40]. Jellemzőbb azonban, hogy a partnerintézményektől kapott rekordok betöltésekor az egységesített besorolási tételekben előforduló – elírásból, transliterációból vagy -kripcióból fakadó – különbözőségek okán duplikált VIAF-azonosítók születnek. A Wikidata közösségi közreműködésen alapuló modellje ilyenkor különösen hasznossá válik, ugyanis mivel egy entáshoz ideális esetben csak egy azonosító tartozik, a P214-es (VIAF-azonosító) tulajdonság egyetlen-érték-kikötése megkönnyíti a duplikált azonosítók megtalálását: amennyiben egy szerkesztő egy entáshoz több VIAF-azonosítót is talál, és ezeket hozzáadja az entás Wikidata-eleméhez, akkor az az elem előbb megjelenik a kikötésmegszegéseket tartalmazó adatbázis-jelentésben [41], majd idővel a VIAF-azonosítók összevonásra kerülnek az alacsonyabb azonosítósorszámú (korábban létrehozott) rekordba. Az így megszűnt azonosítók átirányításként szolgálnak tovább. Konkrét példával szemléltetve a folyamatot, Barátné Hajdu Ágnes első „256088009” azonosítósorszámú VIAF-eleme (lásd a 2. ábrát) 2012-ben került létreho-

zásra az amerikai Kongresszusi Könyvtár rekordja alapján [42]. Ennek duplikátumaként további két VIAF-azonosító („177145541830096600119” és „775152636066320050980”) született még a Német Nemzeti Könyvtárnak és A Cseh Köztársaság Nemzeti Könyvtárának rekordjai alapján [43, 44]. Miután ezek 2021 novemberében beszúrára kerültek Barátné Wikidata-elemébe, a VIAF-azonosítók még abban a hónapban összevonásra kerültek: a két azonosító azóta átirányítja a felhasználókat az elsőként létrehozott VIAF-elemre, amely immáron a két említett nemzeti könyvtár rekordjait is tartalmazza [45].

A Wikidata integrálásának és alkalmazásának hazai úttörője a Petőfi Irodalmi Múzeum (PIM). Az intézmény évtizedes kitaró munkával egy kiterjedt és átfogó személynévteret épített. A PIM nyilvános katalógusa „Gyűjtemények”, „Névtér”, „Díjak” és „Adattárak” részekre oszlik fel, a következőkben azonban csak PIM személynévteréről lesz szó. A múzeum munkatársai által végzett kutatások és adatfeldolgozások során született gyűjtések 2008 óta érhetőek el egyetlen integrált névtéradatbázis részeként. Az adatbázis 2020-ban több mint félmillió személy adatait tartalmazta. A PIM egy 2017 óta zajló szakmai projektje a lehető legtöbb Wikidata- és PIM-rekorddal rendelkező személy beazonosítását, valamint az adatlapok közötti kölcsönös hivatkozás kiépítését tűzte ki céljául, lehetővé téve a felhasználóknak, hogy egyetlen kattintással átugorhassanak az egyik oldalról a másikra [46]. Ennek eléréséhez a PIM személynévtér-azonosítóinak nagyszámú importálása történt a Wikidatába, ami aztán lehetővé tette az azonosítók közvetlen automatikus megjelenítését a magyar Wikipédia szócikkeiben a „Nemzetközi katalógusok” [47] elnevezésű sablon segítségével, valamint a PIM nyilvános katalógusának adatlapjain – az online hivatkozások tárolására szolgáló, MARC 856-os hívójelű mező által – is létrejött egy hivatkozás az adott entitás Wikidata-elemére és az esetlegesen hozzá kapcsolt magyar nyelvű Wikipédia-szócikkére [48]. A projekt másik célja a PIM gyűjteményének láthatóbbá tétele az interneten és a különböző keresőszolgáltatások számára. Korábban már szó volt arról, hogy a közgyűjteményi intézmények technikailag még nem állnak készen a kapcsolt adatok kör-

nyezetére való átállásra, aminek következtében a közgyűjteményi adatbázisok jelentős része még mindig a mély (vagy más fordításban láthatatlan) web része. Definícióját tekintve ez azt jelenti, hogy a tartalmak – esetünkben az adatbázisrekordok – a keresők számára a legtöbb esetben elérhetetlenek, nehezen feltérképezhetők [49]. Ahhoz, hogy például a PIM személynévtér-rekordjai megjelenhessenek a Google-keresések találati listáiban, tartalmuknak értelmezhetőnek kell lennie a keresőszolgáltatás számára. Amennyiben ez teljesül, a találati listán való előrébb kerülés feltétele a hipertextualitás, más oldalról származó hivatkozások. A személynévtér-azonosítók Wikidatába importálása tehát keresőoptimalizálási célokat is szolgált [50]. Egy, a PIM személynévterének forgalmi adatait elemző 2021-es kutatás arra kereste a választ, hogy a projekt megfigyelhető látogatószám-növekedést eredményezett-e. A 2019 áprilisa és 2021 februárja közötti időszakban a Wikidatában szereplő PIM-névtér-azonosítók száma 9986-ról 27 097-re nőtt: a vizsgált periódusban a magyar Wikipédiáról érkező látogatók a Google és a közvetlen (nem más weboldaltól származó) elérések után a harmadik legjelentősebb forrásként a névtér forgalmának nagyjából ötödét tették ki. A nyers adatok anomáliájának tűnhet, hogy a magyar Wikipédia asztali változatáról majdnem hatszoros forgalom érkezik, mint a mobil felületről, ami ellentmondani látszik a mobilkészülékek számának 2010-es évek óta tartó exponenciális növekedésének és napjaink hegemoniájának. Ennek megoldása abban keresendő, hogy a PIM személynévtér-azonosítóinak szócikkbeli megjelenését biztosító „Nemzetközi katalógusok” sablon mobil nézetben nem jelenik meg az olvasók számára, így a magyar Wikipédián szereplő PIM-személynévtér-linkek csupán 17,3%-a volt elérhető a mobil nézetet használó olvasók számára [21].

A Wikidata könyvtári alkalmazásai elsősorban vagy az authority adatbázisok integrálása és átjárhatósága, vagy a gyűjteményi metaadatok minőségének növelése kapcsán kerülnek említésre a szakirodalomban: utóbbi történhet az intézmények azonosítóinak elérhetővé tétele, valamint a lokális rekordok Wikidata-adatokkal történő bővítése által is [51]. Kijelenthető tehát, hogy a Wikidata elméleti és gyakorlati alapon egyaránt

partnere lehet a közgyűjtemények azonosító- és adatcserén alapú authoritykontrolljának, segítve kitörésüket a tradicionális diszciplináris adatsilók jelentette korlátokból, támogatva átállásukat a kapcsolt (nyílt) adatok környezetére és láthatóvá téve a közgyűjteményi adatbázisokat a keresőszolgáltatások számára, végső soron növelve ezáltal az intézmény által elért felhasználók számát.

4. Információ-visszakeresés és -vizualizálás

A visszakereshetőség biztosításának egyik lehetséges módja az információk strukturált formában – például adatbázisok által – történő tárolása. A Wikidata egyedisége abban rejlik a legtöbb szemantikus adatbázishoz képest, hogy tartalma teljes egészében szabadon hozzáférhető, a Creative Commons 0 (Public Domain) licenc ugyanis lehetővé teszi bárkinek az adatok bármilyen – akár üzleti – célú (újra) felhasználását.

A Wikidatában tárolt információk elérésének legegyszerűbb módja az alapértelmezett webes felhasználói felület használata. A visszakeresés legkézenfekvőbb eszköze a honlap jobb felső sarkában található kereső, amennyiben ismert a megtalálni kívánt elem címkéje. Mivel a Wikidata fejlesztése a kezdetektől fogva a többnyelvű adatok tárolásának egyedi követelményeinek figyelembevételével zajlott, a projekt keresőeszköze is akképp került megalkotásra, hogy natív módon képes legyen kezelni a többnyelvű adatokat. Ennek eredményeképp a kereső rugalmas módon képes kezelni a felhasználó által – bármilyen beviteli nyelven – megadott keresési kifejezéseket, a kereső a találatokkal minden esetben a felhasználói felület alapértelmezett megjelenítési nyelvén tér vissza (ahol ez a rendelkezésre álló nyelvi címkék alapján lehetséges). A gyakorlatban ez azt jelenti, hogy a felhasználói felületet magyar nyelvre állítva a fordított névsorrendű (keresztnev, vezetéknév) kifejezésekkel keresve a találatok a 3. ábrán látható módon az elemek magyar nyelvű címkéivel és leírásaival jelennek meg. A bemeneti keresőkifejezés nyelvének megadási kötelezettségének hiánya és az a tény, hogy keresés a tartalmi elemeket tartalmazó fő névtér halmazán kerül végrehajtásra, a felhasználói élmény szempontjából egy rendkívül hasznos funkció, ugyanis a nagy internetes keresőmotoroknál megszokott azonnal megjelenő

tamas horva	🔍
Horváth Tamás (<i>Tamás Horváth</i>) (1977–) magyar mérnök-orvos, kutató	
Horváth Tamás (<i>Tamás Horváth</i>) magyar labdarúgó (1987–)	
Horváth Tamás (<i>Tamás Horváth</i>) (1951–) sakkjátékos, a sakkválogatott szövetségi ...	
Horváth Tamás (<i>Tamás Horváth</i>) labdarúgó (sz. 1991)	
Horváth Tamás (<i>Tamás Horváth</i>) (1963) magyar várospolitikus	
Tamas L Horvath (<i>Tamas Horvath</i>) researcher	
Horváth Tamás (<i>Tamás Horváth</i>) egyértelműsítő lap	
	több
Kifejezést tartalmazó oldalak keresése <i>tamas horva</i>	

3. ábra A Wikidata keresőjének találati eredményei a „tamas horva” kifejezésre magyar nyelvű felhasználói felület esetén.

A szöveg CC0 licenc alatt közzététel

Forrás: <https://www.wikidata.org/>

releváns találatok intuitív keresési élményt eredményeznek. Amennyiben szükséges, haladó keresési beállítások segítségével lehetséges a keresésbe bevont névterek halmazának szűkítése és személyre szabása, ami szerkesztői szempontból például a tulajdonságok és a különböző útmutatók keresésekor lehet hasznos funkció. A felhasználók számára is általában ismert alapvető Boole-operátorok (AND, OR és NOT) és a pontos egyezést eredményező idézőjelek használata mellett állítások és címkék kombinálásának visszakeresése is lehetséges a „haswbstatement” kulcsszóval. A megtalálni kívánt (ezáltal szűkítő paraméterként funkcionáló) tulajdonság-értékpárok a kulcsszót követően adhatók meg „haswbstatement:P...=Q... alakban [52]. Ezen keresési lehetőség akkor bizonyulhat hasznosnak, ha a kulcsszó a sok azonos vagy hasonló címkéjű elem miatt túl sok találati eredménnyel járna. A viszonylag gyakori névnek számító Horváth Péter kifejezésre keresve 2023. szeptember 16-án a Wikidata 1091 darab találatot tért vissza, ami nyilvánvalóan túl nagy halmaz ahhoz, hogy abból a felhasználó képes legyen kikeresni az ELTE-n végzett személyeket [53]. A „Hor-

váth Péter haswbstatement:P31=Q5 haswbstatement:P69=Q390287” keresőkifejezésével azonban már könnyen megkapható az a három fő, akinek Wikidata-elemére igaz az az állítás, hogy osztály, amelynek példánya (P31) = ember (Q5), valamint alma mater (P69) = Eötvös Loránd Tudományegyetem (Q390287) [54].

A korábbi fejezetekben a segédeszközök és a Reasonator kapcsán már szó esett arról, hogy a Wikidata alapértelmezett felhasználói felülete elsősorban az alapvető szerkesztési műveletek elvégzésére van optimalizálva, annak áttekinthetősége nagyban függ az aktuálisan megtekintett Wikidata-elem „zsúfoltságától”. A reprezentált entitás tulajdonság-értékpárjainak túlzott száma könnyen emberi befogadásra alkalmatlanná teheti az elemet, különösen akkor, ha az információmentesség nem fér el a felhasználó képernyőjén. Gyakran előfordul ugyanis, hogy egy elem tulajdonságához több érték is társulhat. Országok – például Magyarország Q28-as azonosítószámú eleme [55] – esetében nyilvánvaló, hogy a népesség (P1082), a születéskor várható élettartam (P2250) vagy a teljes termékenységi mutató (P4841) tulajdonságok attól függően, hogy milyen időpontban kerülnek megállapításra, többféle értéket is felvehetnek. Hazánk fogyatkozó népességének tendenciája ekképpen a tulajdonság időpont minősítővel ellátott évenkénti adataiban jelenik meg. A történeti adatok retrospektív jellegű reprezentáltsága az adatbázisban azonban azt is eredményezi, hogy Magyarország Wikidata-eleme a rengeteg tulajdonság-értékpár okán rendkívül hosszúvá nyúlik, még a Reasonator által generált olvasásra szánt összefoglaló is túl sok információt tartalmaz egy humán felhasználó számára [56]. Egy szemantikus adatbázisnak azonban nem is a felhasználói igényekhez igazodó információreprezentáció az elsődleges célja, hanem az adatok strukturált és géppel olvasható formátumú tárolása és szolgáltatása többek között a felhasználók böngészési élményét fokozó és az adatok vizualizálására szolgáló eszközök számára. A természetes nyelvi feldolgozáson alapuló szövegalkotást vizsgáló fejezetben már bemutatott Reasonator és Wikidata-kompatibilis infoboxsablonok jó példái annak, hogy miképp hasznosíthatóak a Wikidata strukturált adatai közérthető formába rendezve.

Külső alkalmazások és platformok esetében – amikor a felhasználó nem közvetlenül a projekt weboldalán keresztül használja az adatbázist – a Wikidata strukturált adatainak lekérése egy közvetítő lekérdezőnyelv, a SPARQL segítségével történik, ami lehetővé teszi az RDF-formátumban tárolt adatok kinyerését [57]. A szemantikus adatbázisok egyik legfontosabb technológiájaként számontartott szabvány 1.0-ás verziója 2008-ban, míg 1.1.-es verziója 2013-ban vált a W3C hivatalos ajánlásává. A SPARQL-lekérdezések triplet mintákból, konjunkciókból, diszjunkciókból és egyéb opcionális mintákból állhatnak [58]. Az SPARQL-nyelv a lekérdezések célját tekintve négy típust specifikál:

- SELECT: a SPARQL-végpontok nyers adatainak kinyerésére szolgál, az eredményeket táblázatos formában adja vissza,
- CONSTRUCT: a SPARQL-végpontok nyers adatainak kinyerésére és a kapott adatok érvényes RDF-formátumba történő átalakítására szolgál,
- ASK: egy SPARQL-végponton végrehajtott lekérdezés igaz/hamis eredményét adja meg,
- DESCRIBE: egy SPARQL-végponthoz tartozó RDF-diagramot hoz létre, amely az egyes végpontok hasznosságát adja meg a tartalmazzott információ tekintetében.

A Wikidata Query Service (WDQS) a Wikidata saját SPARQL-végpontszolgáltatása, amely a SPARQL-nyelven íródott lekérdezések eredményeképp a projekt elemeit adja vissza [59]. A WDQS használata elsősorban akkor célszerű, ha egyedül a kívánt adatok karakterisztikája ismert. Hátránya, hogy feltételezhetően nagyszámú adat (az adatbázis jelentős hányadának) visszakeresésére nem használható, valamint a szabadszavas és fuzzy – rugalmas szabályú, akár reguláris kifejezést is tartalmazó – keresési mód sem alkalmazható [60]. A Wikidata-szerkesztőközössége a jelölőnyelv alapjainak elsajátításához számos interaktív, statikus és lépésről-lépésre útmutató áll rendelkezésre [61].

Felhasználói szemszögből azonban a SPARQL lehet bármennyire adatorientált és kezdőbarát nyelv, létezhet akárhány útmutató és leírás hozzá, ha annak használata nem intuitív a programozási készségekkel és képességekkel nem, vagy csak korlátozottan rendelkező többség számára. Ezen

helyzetet kívánja orvosolni a Wikimedia Foundation által üzemeltetett projektekkel kompatibilis „PetScan” eszköz, amelynek többnyelvű felülete egy egyszerű, de robotsztus információ-visszakereső módszert rejt [62]. Az eszközzel elsősorban SELECT és CONSTRUCT típusú lekérdezések végezhetők, lehetővé téve a különböző nyelvű Wikipédia-projektek és a Wikidata tartalmi kombinálását. Amellett, hogy natívan kezeli a WDQS lekérdezéseit, a felhasználó számára nagyfokú szabadságot kínál a különböző szűrők és generátorok definiálása terén [63]. Ez különösen a Wikidatában kevésbé reprezentált nyelv- és országspecifikus információk visszakeresésénél bizonyulhat hasznosnak. Például, ha a felhasználó az Örmény Tudományos Akadémia (Հայաստանի Հանրապետության Գիտությունների ազգային ակադեմիա, Hajasztáni Hanrapetutján Gitutjunneri Azgajin Akademia) [64] tagjainak Wikidata-elemeit szeretné megkapni, akkor 2023. szeptember 16-án a 3. számú *mellékletben* látható egyszerű SPARQL-lekérdezést használva 167 rekordot kapna eredményül. [65] Ezzel szemben a PetScan az örmény Wikipédia – az akadémia rendes [66], levelező [67] és külső [68] tagjait tartalmazó – kategóriáit betöltve már 384 személy Wikidata-elemével térne vissza [69]. A két keresést OR művelettel ötvöző PetScan-lekérdezés visszaadja a Wikidatában szereplő 389 akadémiai tag rekordjának listáját [70], köztük például Armenak (Armen) Nazarov Gevorgi (Արմենակ [Արմեն] Գևորգի Նազարով) örmény szeizmológus Q62104559-es elemazonosítójú, tagság (P463) tulajdonság nélkül álló rekordját is [71].

A Wikidata tartalmának SPARQL általi lekérdezése elsősorban a fejlesztők számára tartogat szerteágazó lehetőségeket, amiknek egyedüli korlátozója az adatbázis tartalmi lefedettsége. Ennek egyik kiemelkedő példája a 2017-ben fejleszteni kezdett Scholia projekt, amely a Wikidatát használja a nemzetközi tudományos élet bibliográfiai adatainak kezelésére. Az eszköz egyedi tudományos profilokat hoz létre a kutatók, szervezetek, folyóiratok, kiadók, tudományos produktumok és kutatási témák számára. Az ehhez szükséges információkat a webszolgáltatás SPARQL-lekérdezések által a WDQS segítségével nyeri ki a Wikidatából. A webes szolgáltatást implementáló Python programnyelvben íródott csomag képes a Wikidata

bibliográfiai adatokat tartalmazó elemeit LaTeX és BIBTeX formátumúvá alakítani. A Scholia projekt a Wikidata kiterjedt bibliográfiai adataira támaszkodva – 2020 februárjában az akkor létező 71,6 millió elem 31,5%-a tudományos cikkről szólt [72] – kívánja bemutatni a tudományos élet szereplőit és eredményeit. A Scholia fejlesztése a WikiCite kezdeményezés keretében zajlik, aminek célja olyan bibliográfiai (meta)adatok indexelése a Wikidatában, amelyek később felhasználhatók például Wikipédia-szócikkekben vagy máshol tett állítások alátámasztására [73]. A nyilvánosan és ingyenesen elérhető adatok mellett alternatíváját jelenthetik az üzleti modellben működő bibliográfiai szakadatbázisok tudományometriai mutatóinak is. Noha a Wikidatában fellelhető bibliográfiai adatok mennyisége limitált az erre szakosodott citációs adatbázisokhoz képest, számuk folyamatosan növekszik. Erre jó példa, hogy *Kiszl Péter* [74], *Bibor Máté* [22] és *Németh Katalin* [75] Wikidata-eleme még – az ELTE Egyetemi Könyvtárban található itáliai kódexekről szóló – közös tanulmányuk megjelenésének napján létrehozásra került a tanulmány rekordjával együtt [76]. A folytonos növekedés azonban buktatókat is rejt magában: a túlságosan sok elemet érintő és ezért hamar erőforrásigényessé váló SPARQL-lekérdezések idővel fenntarthatatlanná válnak, azt eredményezve, hogy a WDQS megszakítja azokat, ugyanis a szerverek kapacitása – a Wikimedia Foundation non-profit és adományokra épülő működési modellje miatt – limitált. A WDQS-lekérdezések által a Scholia képes a Wikidata különböző típusú entitásokat leíró elemeinek adatait oly módon kombinálni, ami más (tudomány)bibliográfiai webszolgáltatás esetében általában nem lehetséges, például lehetséges vele gráfként ábrázolni a társszerzők kapcsolati hálózatát. A grafikonok készítésénél felhasználhatók továbbá a Wikidata-elemek nemi identitásra vonatkozó információi is, láthatóbbá téve ezáltal a tudományos élet nemi egyenlőtlenségeit [77].

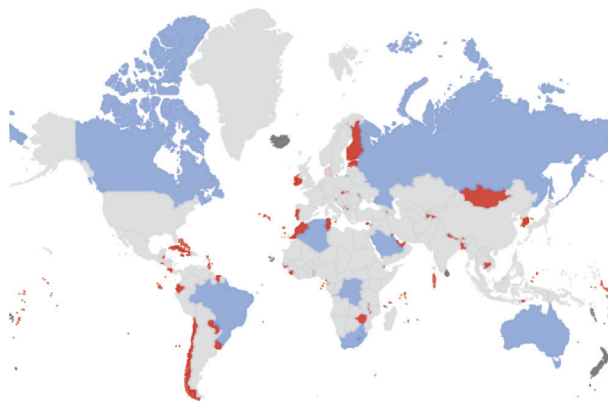
A Wikidata adatait tekintve azonban a nemi reprezentáción túlmutató szisztematikus egyenlőtlenségekkel küszködik, amelyek egyrésztől torzuláshoz vezethetnek az információk visszakeresése során, másrésztől viszont lehetővé is teszik ezek látványos vizualizálását [78]. A Wikimedia Foundation által fenntartott projektek nemi egyenlőtlen-

seégeinek – 2023. szeptember 16-án az életrajzok csupán 18,6%-a szólt nőkről, szemben a férfiak 81,3 százalékaival (a fennmaradó 0,076 százalék a további nemi identitásoké) – vizualizálására például egy dedikált, az adatok különböző megjelenítési lehetőségeit kínáló weboldal készült [79]. A Wikidata többnyelvűségének vizsgálatakor már szó esett arról, hogy a nyelvek beszélőinek száma nem korrelál a Wikidata nyelvi adatainak megoszlása tekintetében elfoglalt helyével, felülreprezentáltak az európai nyelvek. Hasonló jelenség figyelhető meg akkor is, ha a Wikidata koordinátákat tartalmazó elemeit pontként ábrázoljuk: a 4. számú mellékleten a kontinensek néhol halvány, de markáns kirajzolódása mellett az is szembetűnő, hogy Európa és Észak-Amerika mennyivel lefedettebb (és ezért fényesebb) a világ többi részéhez képest, látványosan illusztrálva ezen területek periférikus mivoltát az adatbázisban.

A SPARQL-nyelv alapjainak elsajátítása azonban egy egyszerű felhasználó számára is előnyös lehet, elsősorban azért, mert a Wikidata elemei világunk entitásainak széles spektrumát lefedik, általuk lekérdezhetővé és vizualizálhatóvá válnak elsősorban nem függő, komplex kérdések is. Szerkesztői oldalról azért is hasznos lehet ez a tudás, mert segítségével kinyerhetők a felhasználók érdeklődésének megfelelő – például magyar Wikipédia-cikkkel még nem rendelkező – témák, de a lekérdezések által akár bizonyos szerkesztői tevékenységek elvégzése is megkönnyíthető. Több mint két tucat Wikipédia-projekt használ például SPARQL-lekérdezésen alapuló automatikusan frissülő listaoldalt (vagy adatbázisjelentést), ami kigyűjti helyi hivatkozással rendelkező személyek friss haláloseit, megkönnyítve a szerkesztők számára azok aktualizálását [80]. Egy programnyelv autodidakta módon történő elsajátításakor azonban fontos a belső motiváció fenntartása: a WDQS által kínált sokrétű adatvizualizációs megoldások (fatérkép, pont- és vonaldiagram, gráf, idővonal stb.) látványos lekérdezések készítését teszik lehetővé, amelyekhez a rendszer a könnyebb és karaktertakarékosabb megosztás érdekében egyedi és rövidített URL-eket is képes generálni. Ez különösen akkor tud hasznos lenni, ha a lekérdezések eredményeinek vizualizálása közönség számára készül, például oktatási célból. A Wikidata adatai ugyanis

akár a közoktatás keretei között is felhasználhatók: a vizualizációk segítségével a tanárok például bemutathatják a Moby Dick szereplőiről elnevezett gleccsereket [81], de akár interaktív formában is taníthatják a világ országainak zászlóit egy olyan lekérdezéssel, ami összegyűjti a piros (vörös), fehér, zöld színeket tartalmazó zászlókat [82].

Annak illusztrálására, hogy lehetséges oktatási célból hasznos lekérdezéseket készíteni, az 5. számú melléklet SPARQL-lekérdezésének segítségével arra a földrajtanításhoz kapcsolódó kérdésre kerestem a választ, hogy melyek azok az országok, amelyek területe nagyobb, illetve kisebb valamilyeni határos szomszédjánál. Noha elsősorban könnyűnek tűnhet a kérdés, hiszen kézenfekvő a válasz, hogy Oroszország és Ausztrália minden szomszédjánál nagyobb, míg a különböző városállamok kisebbek. A feladat célja rávilágítani az olyan komplex esetekre, mint például Franciaország, amely – noha az Európai Unió legnagyobb országa – tengerentúli területei miatt határos a nála nagyobb Brazíliával, ezért nem nagyobb minden szomszédjánál. A lekérdezés eredményeképp kapott, a 4. ábrán látható térképen kék színnel vannak jelölve a szomszédjaiknál nagyobb, valamint piros színnel a kisebb országok. A világosszürkén jelölt országok kisebb és nagyobb szomszédokkal egyaránt rendelkeznek, míg a sötétszürke a szárazföldi határral nem rendelkező, és ezért kiszűrendő országok egy részét jelöli.



4. ábra Az összes szomszédjuknál kisebb (piros) vagy nagyobb (kék) területű ENSZ-tagállamok. A kisebb és nagyobb szomszédokkal egyaránt rendelkező országok világosszürkén, a szárazföldi határral nem rendelkező országok – egy része – sötétszürkén vannak jelölve.

Az adatok és a vizualizáció CC0 licenc alatt közzétéve

Forrás: <https://query-chest.toolforge.org/redirect/WnHa7LML1kmWm4GuualWE0oIEMyQaqAIAcOEmCwYWqj>

A térkép forrását felkeresve a lekérdezés eredménye egy interaktív felületen jelenik meg – szemben a 4. ábra statikus formájával –, amit szabadon lehet nagyítani, kicsinyíteni és mozgatni. A jobb felső sarokban emellett hozzáadhatók és eltávolíthatók a már bemutatott színnel jelölt rétegek, valamint egy adott országra kattintva annak területe négyzetméterben és határos országainak listája is rögtön elérhető. Mindezek lehetővé teszik a tanár számára az átadni kívánt tudás modern és hatékonyabb szemléltetését, hiszen az iskolai papíralapú atlaszokkal csak nehézkesen lehetséges az országok területi adatainak gyors visszakeresése.

A Wikidata adatainak vizualizálása kapcsán említést érdemel még a Databánya, a Digitális Bölcsészeti című szakfolyóirat és a Petőfi Irodalmi Múzeum által 2019. március 2-án a PIM-ben megrendezett Irodalmi hálózatok I. – Nyugathon esemény, ami egy edit-a-thonnal (szerkesztőmaratonnal) egybekötött adatvizualizációs workshop volt. Ennek keretében a résztvevő kutatók, gyakorlott Wikipédia-szerkesztők és egyszerű érdeklődők közösen dolgoztak a 20. század első felének magyarországi sajtójával kapcsolatos Wikidata-elemek és Wikipédia-szócikkek létrehozásán és szerkesztésén. A széles körben ismert Nyugat folyóiraton kívül számos egyéb folyóirat is működött a korban, az esemény szervezői pedig arra keresték a választ, hogy hogyan kapcsolódtak össze a korszak szerzői és folyóiratai, ki kivel állt kapcsolatban. A résztvevők egy előre összeállított, folyóiratokat és szerzőket tartalmazó adatlistával dolgoztak, amelyek entitásainak aztán Wikipédia-szócikkeket és Wikidata-adatlapokat hoztak létre a PIM által kínált nyomtatott és online forrásokra támaszkodva. Az esemény végén a szervezők a résztvevők közreműködésével kiegészült Wikidata adataiból vizualizálták a kor folyóiratainak és szerzőinek hálózatát, valamint a folyóiratok szerzőgárdájának heterogenitását [21]. Az esemény szervezői számára a workshop hasznos tapasztalatokkal szolgált, amelyeket később fel is tudtak használni az ELKH Bölcsészettudományi Kutatóközpont hálózatvizualizációs projektje során [83].

A strukturált adatok visszakeresése a felhasználói igényeknek megfelelően tehát többféle módon is megvalósulhat. Látható, hogy a Wikidata keresőmotorjának címkealapú keresési módja szí-

gorúbb szűrést lehetővé tevő állításokkal is kombinálható. Mivel az alapértelmezett felhasználói felület egyben a szerkesztőközösség közreműködéseinek elsődleges színtere is, annak böngészési és adatvizualizációs hiányosságainak orvoslására kínál megoldást a belső fejlesztésként született Reasonator, ami természetes nyelvi szövegeket generáló Wikipédia-szerű felépítésével helyezi kontextusba és teszi emberi befogadásra alkalmasabb formájúvá a nyers adatokat. A Wikidata adatainak géppel olvasható formátumú kinyerését elsősorban a SPARQL lekérdezőnyelv és az erre épülő szolgáltatások teremtik meg. A szemantikus adatbázisok egyik legfontosabb technológiaként számotartott szabvány a Wikidata saját SPARQL-végpontszolgáltatása által teszi lehetővé az RDF-formátumban tárolt adatok visszakeresését. Számos webes alkalmazás és szolgáltatás támaszkodik ezáltal a Wikidatára, köztük például a tudományos élet bibliográfiai adatainak kezelésére és vizualizálására indított Scholia kezdeményezés is. Az Irodalmi hálózatok I. – Nyugathon szerkesztőmaratonnal egybekötött adatvizualizációs workshop példáján keresztül és a fejezetben ismertetett konkrét lekérdezések képében kerültek bemutatásra jógyakorlatok az információvizualizálás oktatási és tudományos célú felhasználására: a szaktárgyi tananyagba beilleszthető lekérdezéseken túl a Wikidata strukturált adatai akár hálózatvizualizációs projektek alapjaként is szolgálhatnak bizonyos esetekben.

5. Összegzés

A Wikidata 2012-es indulásakor megfogalmazott kezdeti célkitűzések elérését követően a projekt hamar meghaladta a Wikimedia Foundation nyelv- és wikiközi hivatkozásait gyűjtő repozitórium szerepét, és rövid idő alatt a szemantikus web egyik legfontosabb adatbázisává nőtte ki magát. A tudásgráf adatainak gyorsléptékű növekedése és alkalmazásának széleskörű elterjedése ellenére azonban a Wikidatával kapcsolatos kutatások alacsony száma és egyenlőtlen eloszlása a kutatási téma alulreprezentáltságát és kiforratlanságát jelzi. A téma Magyarországon szintén alig publikált, az ezzel foglalkozó kutatások, konferencia-előadások és workshopok szinte kivétel nélkül a – korábban a Petőfi Irodalmi Múzeumhoz, 2021 decembe-

rétől az Országos Széchényi Könyvtárhoz tartozó – Digitális Bölcsészeti Központ munkatársaihoz kötődnek. A Wikidata átfogó, tudományos igényű és magyar nyelvű feldolgozását célozva a dolgozat részletesen ismertette a tudásgráf strukturális felépítését és legfontosabb ismérveit. Az információkat felépítő RDF tripletekhez szükséges – elem, állítás, tulajdonság és érték – alapfogalmakon kívül az adatok (újra)felhasználásához elengedhetetlen forráshivatkozások és azonosítók, valamint a visszakereséshez szükséges címkék és leírások is részletesen bemutatásra kerültek.

A Wikidata és társprojektjei által támogatott nyitott tudás mozgalom terjesztésének egy fontos aspektusa a többnyelvűség: az összesen több mint 300 nyelven elérhető 60 millió Wikipédia-szócikk adataival kiegészült Wikidata kiváló adatforrásként szolgálhat a humán interakcióra tervezett alkalmazások, így például a mesterséges intelligenciával támogatott csevegőrobotok (pl. ChatGPT) és virtuális asszisztensek (pl. Amazon Alexa, Apple Siri) számára. A magyar nyelv szerencsés helyzetben van e tekintetben, ugyanis a Wikidata Európa-centrikussága, a magyar nyelvű közreműködők felülreprezentáltsága a szerkesztőközösségben és a magyar Wikipédia–Wikidata szoros kooperációja előnyös kiindulási helyzetet teremt az integrációban gondolkodó hazai intézmények számára. Az adatbázis strukturált adatainak Wikipédia-szócikkek információiból, partnerségi adatcserékből, közgyűjteményi (adat)hozzájárulásokból, és nem utolsósorban a közösségi közreműködésen alapuló modellből fakadó töretlen növekedése egyben a projekt népszerűségét indukálta. A Wikidata adatainak ingyenes és megkötések nélküli felhasználhatóságában rejlő lehetőségeket a legnagyobb szilícium-völgyi cégek mellett lassan a közgyűjtemények is kezdik felismerni, hogy a Wikidata a hagyományos authority-kontroll alternatívájának tekinthető, amely segít

het az intézményi adatsilókba zárt adatok nyílttá és szabadon elérhetővé tételében. A könyvtárak számára a Wikidata különösen az authority adatok egyértelműsítése, karbantartása és linkelése terén bizonyulhat hasznosnak. Noha a világ könyvtáraiban töretlenül a MARC a legelterjedtebb adatcsere-formátum, a kapcsolt adatok környezetére való átállás szemléletében fejlesztett BIBFRAME adatcsere-keretrendszer készen áll a MARC formátum jövőbeli leváltására, megteremtve a mélyebb integrációk lehetőségét a szemantikus web adatbázisaival, többek között a Wikidataival is. Ennek megtörténteig például a bemutatott nemzetközi és hazai kezdeményezések szolgálhatnak mintaként a kapcsolt (nyílt) adatokra történő átállás eredményeire.

A Wikidata információinak visszakeresésére számos mód áll rendelkezésre, azonban a strukturált adatok géppel történő lekérdezésére a SPARQL nyelv a legalkalmasabb. A szemantikus adatbázisok egyik legfontosabb technológiájaként számontartott szabvány a Wikidata esetében is kulcsszerepet játszik az adatok gyors és hatékony lekérdezésében, megnyitva ezzel az utat olyan külső alkalmazások előtt, mint például a nemzetközi tudományos bibliográfiai adatok rendszerezését és vizualizálását célzó Scholia projekt. Adatorientált és kezdőbarát nyelvként a programozói készségekkel nem, vagy csak alig rendelkező felhasználók is rövid idő alatt elsajátíthatják az alapokat például a szerkesztőközösség által készített útmutatók és segédletek által, megteremtve ezáltal akár a tudásgráf adatainak és az azt használó eszközök tantermi környezetben történő alkalmazását is. Végül soron pedig a Wikidata tartalma is tudományos kutatások témájául szolgálhat, ugyanis az adatokban megjelenő nyelvi, nemi és szocioökonómiai egyenlőtlenségek társadalmunk kihívásainak leképeződéseként is felfoghatóak, lehetővé téve ezek adataalapú elemzését az adatbázisban.

Irodalom és jegyzetek

- [1] Molnár, B. *Wikidata, a többnyelvű közösségi közreműködésen alapuló tudásgráf – 1. rész: Egy szakirodalmi és strukturális áttekin-tés, Tudományos és Műszaki Tájékoztató*, 70(3), p. 316–333, 2023.
<https://doi.org/10.3311/tmt.13260>
- [2] Interactive Terminology for Europe, *natural language processing*, Elérhető: <https://iate.europa.eu/entry/result/130121/en-fr> (Utolsó elérés: 2023. 09. 16.)
- [3] Mora-Cantalops, M., Sánchez-Alonso, S., García-Barriocanal, E. *A systematic literature review on Wikidata*, Data Technologies and Applications, 53(3), p. 250–268, 2019.
<https://doi.org/10.1108/DTA-12-2018-0110>
- [4] MediaWiki, *Extension:ArticlePlaceholder*, Elérhető: <https://www.mediawiki.org/wiki/Extension:ArticlePlaceholder> (Utolsó elérés: 2023. 09. 16.)
- [5] Kaffee, L. A. et al. *Mind the (language) gap: Generation of multilingual wikipedia summaries from wikidata for articleplaceholders*, In: The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, p. 319–334.
https://doi.org/10.1007/978-3-319-93417-4_21
- [6] Ta, H. T., Gelbukha, A., Sidorov, G. *Mapping Process for the Task: Wikidata Statements to Text as Wikipedia Sentences*, arXiv, 2022.
<https://doi.org/10.48550/arXiv.2210.12659>
- [7] Vrandečić, D. *Building a Multilingual Wikipedia*, Communications of the ACM, 64(4), p. 38–41, 2021.
<https://doi.org/10.1145%2F3425778>
- [8] Toolforge, *Reasonator*, Elérhető: <https://reasonator.toolforge.org/> (Utolsó elérés: 2023. 09. 16.)
- [9] Manske, M. *The Reason For It All*, The Whelming, Elérhető: <http://magnusmanske.de/wordpress/?p=138> (Utolsó elérés: 2023. 09. 16.)
- [10] Odell, J. D., Lemus-Rojas, M., Brys, L. *Wikidata for Scholarly Communication Librarianship*, IUPUI University Library, 2022,
<https://doi.org/10.7912/9Z4E-9M13>
- [11] Wikidata, *Wikidata:Infobox Tutorial*, Elérhető: https://www.wikidata.org/wiki/Wikidata:Infobox_Tutorial (Utolsó elérés: 2023. 09. 16.)
- [12] Wikipédia, *Modul:Wikidata*, Elérhető: <https://hu.wikipedia.org/wiki/Modul:Wikidata> (Utolsó elérés: 2023. 09. 16.)
- [13] Wikipédia, *Sablon:Wikidata*, Elérhető: <https://hu.wikipedia.org/wiki/Sablon:Wikidata> (Utolsó elérés: 2023. 09. 16.)
- [14] Sáez, T., Hogan, A. *Automatically generating Wikipedia info-boxes from Wikidata*, In: WWW '18: Companion Proceedings of the The Web Conference 2018, International World Wide Web Conferences Steering Committee, Genf, 2018, p. 1823–1830.
<https://doi.org/10.1145/3184558.3191647>
- [15] Wikipedia, *List of Wikipedias*, Elérhető: https://en.wikipedia.org/wiki/List_of_Wikipedias (Utolsó elérés: 2023. 09. 16.)
- [16] Lewoniewski, W. *Completeness and Reliability of Wikipedia Infoboxes in Various Languages*, In: Abramowicz, W. (szerk.) Business Information Systems Workshops: BIS 2017 International Workshops, Revised Papers, Poznań, Poland, June 28–30, 2017, p. 295–305.
https://doi.org/10.1007/978-3-319-69023-0_25
- [17] Wikipédia, *Kategória: Wikidatát is használó sablonok*, Elérhető: https://hu.wikipedia.org/wiki/Kateg%C3%B3ria:Wikidat%C3%A1t_is_haszn%C3%A1l%C3%B3_sablonok (Utolsó elérés: 2023. 09. 16.)
- [18] Wikipédia, *Modèle:Infobox Biographie2*, Elérhető: https://fr.wikipedia.org/wiki/Mod%C3%A8le:Infobox_Biographie2 (Utolsó elérés: 2023. 09. 16.)
- [19] Például a 2. számú mellékletben látható infobox tanulmányokra vonatkozó szakaszának ceruzája a következő hivatkozást rejti: <https://www.wikidata.org/wiki/Q662084?uselang=fr#P106>. Az URL Szentágothai János Q662084 Wikidata-azonosítószámú elemére mutat, kiegészülve két query stringgel: az egyik az elem foglalkozás tulajdonságának (lásd 2. számú táblázat) szakaszára navigálja a látogatót („#P106”), a másik a Wikidata felhasználó felületének nyelvét módosítja („uselang=fr”).
- [20] Sullivan, D. *A reintroduction to our Knowledge Graph and knowledge panels*, The Keyword, Elérhető: <https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/> (Utolsó elérés: 2023. 09. 16.)
- [21] Molnár, B. *A Wikidata és a Nemzeti Névtér kapcsolódási lehetőségei*, Könyvtári Figyelő, 67(1), p. 46–55, 2021.
- [22] Wikidata, *Bibor Máté János*, Elérhető: <https://www.wikidata.org/wiki/Q113145838> (Utolsó elérés: 2023. 09. 16.)
- [23] Google, *Máté János Bibor*, Elérhető: <https://www.google.com/search?kgmid=/g/11c5s62dym&hl=en> (Utolsó elérés: 2023. 09. 16.)
- [24] Merhav, Y., Ash, S. *Design challenges in named entity transliteration*, In: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fae, 2018, p. 630–640. ISBN: 978-1-948087-50-6
- [25] Curotto, P., Hogan, A. *Suggesting Citations for Wikidata Claims based on Wikipedia's External References*, In: Proceedings of the 1st Wikidata Workshop (Wikidata 2020) co-located with 19th International Semantic Web Conference(OPub 2020), Virtual Conference, November 2-6, 2020., CEUR-WS.org, 2020.
- [26] McKenna, L., Debruyne, C., O'Sullivan, D. *Understanding the position of information professionals with regards to linked data: a survey of libraries, archives and museums*, In: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, Association for Computing Machinery, New York, 2018, p. 7–16.
<https://doi.org/10.1145/3197026.3197041>

- [27] Neubert, J. *Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings*, In: Proceedings of the 17th European Networked Knowledge Organization Systems Workshop co-located with the 21st International Conference on Theory and Practice of Digital Libraries 2017 {(TPDL} 2017), Thessaloniki, Greece, September 21st, 2017, CEUR-WS.org, 2017, p. 14–25.
- [28] Van Veen, T. *Wikidata: From “an” Identifier to “the” Identifier*, Information Technology and Libraries, 38(2), p. 72–81, 2019. <https://doi.org/10.6017/ital.v38i2.10886>
- [29] Interactive Terminology for Europe, *data silo*, Elérhető: <https://iate.europa.eu/entry/result/3572020/en-fr> (Utolsó elérés: 2023. 09. 16.)
- [30] O’Neill, B., Stapleton, L. *Digital cultural heritage standards: from silo to semantic web*, AI & Soc, 37, p. 891–903, 2022. <https://doi.org/10.1007/s00146-021-01371-1>
- [31] Smith-Yoshimura, K. *Transitioning to the Next Generation of Metadata*, OCLC Research, Dublin, 2020, ISBN: 978-1-55653-167-5
- [32] Sardo, L., Bianchini, C. *Wikidata: a new perspective towards universal bibliographic control*, J LIS.it, 13(1), p. 291–311, 2022. <https://doi.org/10.4403/JLIS.IT-12725>
- [33] Serra, L. G., Schneider, J. A., Segundo, J. E. S. *Person Identifiers in MARC 21 Records in a Semantic Environment*, Cataloging & Classification Quarterly, 58(5), p. 505–519, 2020.
- [34] Jin, Q. *Enhanced discovery with linked open data for library digital collections*, Technical Services Quarterly, 38(1), p. 17–32, 2021. <https://doi.org/10.1080/07317131.2020.1854575>
- [35] Myntti, J. et al. *Regional connections to national authority files*, Cataloging & Classification Quarterly, 58(1), p. 76–89, 2020. <https://doi.org/10.1080/01639374.2019.1690087>
- [36] Stanishevskaya, I. *Authority Work, Catalogers, and Contemporary Authors: Results of an Online Survey*, Technical Services Quarterly, 37(3), p. 272–290, 2020. <https://doi.org/10.1080/07317131.2020.1768702>
- [37] OCLC, *VIAF: Convenient acces to name authority files*, Elérhető: <https://www.oclc.org/en/viaf.html> (Utolsó elérés: 2023. 09. 16.)
- [38] Klein, M., Kyrios, A. *VIAFbot and the integration of library data on Wikipedia*, Code4lib Journal, 22, 2013.
- [39] Wikidata, *Property talk: P214*, Elérhető: https://www.wikidata.org/wiki/Property_talk:P214 (Utolsó elérés: 2023. 09. 16.)
- [40] Wikidata, *Wikidata:VIAF/cluster/conflating entities*, Elérhető: https://www.wikidata.org/wiki/Wikidata:VIAF/cluster/conflating_entities (Utolsó elérés: 2023. 09. 16.)
- [41] Wikidata, *Wikidata:Database reports/Constraint violations/P214*, Elérhető: https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/P214 (Utolsó elérés: 2023. 09. 16.)
- [42] VIAF, *256088009*, Elérhető: <https://viaf.org/viaf/256088009/> (Utolsó elérés: 2023. 09. 16.)
- [43] VIAF, *177145541830096600119*, Elérhető: <https://viaf.org/viaf/177145541830096600119/> (Utolsó elérés: 2023. 09. 16.)
- [44] VIAF, *775152636066320050980*, Elérhető: <https://viaf.org/viaf/775152636066320050980/> (Utolsó elérés: 2023. 09. 16.)
- [45] Wikidata, *Barátné Hajdu Ágnes*, Elérhető: <https://www.wikidata.org/wiki/Q787018> (Utolsó elérés: 2023. 09. 16.)
- [46] Hubay, M. *A tudás és a részek összege*, A Petőfi Irodalmi Múzeum szakmai blogja, Elérhető: https://pimblog.blog.hu/2020/06/10/wikidata_pim (Utolsó elérés: 2023. 09. 16.)
- [47] Wikipédia, *Sablon:Nemzetközi katalógusok*, Elérhető: https://hu.wikipedia.org/wiki/Sablon:Nemzetk%C3%B6zi_katal%C3%B3gusok (Utolsó elérés: 2023. 09. 16.)
- [48] Hubay, M. *Neked is jó, nekem is jó: A Wikidata és a PIM személynév-entitásainak összekapcsolása és adatgazdagítása*, In: Networkshop 2020, online konferencia, 2020. szeptember 2–4.
- [49] Herendy, Cs. *A kereső, a dokumentumok és a user*, Médiakutató, 11(1), p. 41–55, 2010.
- [50] Hubay, M. *Rekordkapcsolattól a szemantikus kapcsolatig*, In: Hagyományok és kihívások VIII., p. 115–123, ELTE Egyetemi Könyvtár és Levéltár, 2020. ISBN: 9789634892540
- [51] Zhao, F. *A systematic review of Wikidata in Digital Humanities projects*, Digital Scholarship in the Humanities, 00, p. 1–23, 2022. <https://doi.org/10.1093/llc/fqac083>
- [52] Wikidata, *Help:Navigating Wikidata*, Elérhető: https://www.wikidata.org/wiki/Help:Navigating_Wikidata (Utolsó elérés: 2023. 09. 16.)
- [53] Wikidata, *A keresés eredménye [keresőkifejezés: Horvath Peter]*, Elérhető: <https://w.wiki/6Zy6> (Utolsó elérés: 2023. 09. 16.)
- [54] Wikidata, *A keresés eredménye [keresőkifejezés: Horvath Peter haswbstatement:P31=Q5 haswbstatement:P69=Q390287]*, Elérhető: <https://w.wiki/6Zy5> (Utolsó elérés: 2023. 09. 16.)
- [55] Wikidata, *Magyarország*, Elérhető: <https://www.wikidata.org/wiki/Q28> (Utolsó elérés: 2023. 09. 16.)
- [56] Toolforge, *Magyarország*, Elérhető: <https://reasonator.toolforge.org/?q=Q28&lang=hu> (Utolsó elérés: 2023. 09. 16.)
- [57] W3C, *SPARQL 1.1 Query Language: W3C Recommendation 21 March 2013*, Elérhető: <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/> (Utolsó elérés: 2023. 09. 16.)
- [58] XML.org, *XML and Web Services In The News - 6 October 2006: SPARQL Query Language for RDF*, Elérhető: <http://www.xml.org/xml/news/archives/archive.10062006.shtml#5> (Utolsó elérés: 2023. 09. 16.)
- [59] Wikidata, *Wikidata Query Service*, Elérhető: <https://query.wikidata.org/> (Utolsó elérés: 2023. 09. 16.)

- [60] Wikidata, *Wikidata:Data access*, Elérhető: https://www.wikidata.org/wiki/Wikidata:Data_access (Utolsó elérés: 2023. 09. 16.)
- [61] Wikidata, *Wikidata:SPARQL query service/Wikidata Query Help*, Elérhető: https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/Wikidata_Query_Help (Utolsó elérés: 2023. 09. 16.)
- [62] Meta, *PetScan*, Elérhető: <https://meta.wikimedia.org/wiki/PetScan/> (Utolsó elérés: 2023. 09. 16.)
- [63] Manske, *Of cats and pets*, The Whelming, Elérhető: <http://magnusmanske.de/wordpress/?p=385> (Utolsó elérés: 2023. 09. 16.)
- [64] Az örmény nevek a dolgozat egészében az akadémiai magyar átírásnak megfelelő alakban szerepelnek. Ligeti, L., Terjék, J. *Keleti nevek magyar helyesírása*, Akadémiai Kiadó, 1981, ISBN: 963-05-2080-X
- [65] Wikidata, *Wikidata Query Service*, Elérhető: <https://w.wiki/6Yg2> (Utolsó elérés: 2023. 09. 16.)
- [66] Վիքիպեդիա, *Գտնագրիչ: ՀՀ ԳԱԱ սկսող նիկուներ*, Elérhető: https://hy.wikipedia.org/wiki/%D4%BF%D5%A1%D5%BF%D5%A5%D5%A3%D5%B8%D6%80%D5%AB%D5%A1:%D5%80%D5%80_%D4%B3%D4%B1%D4%B1_%D5%A1%D5%AF%D5%A1%D5%A4%D5%A5%D5%B4%D5%AB%D5%AF%D5%B8%D5%BD%D5%B6%D5%A5%D6%80 (Utolsó elérés: 2023. 09. 16.)
- [67] Վիքիպեդիա, *Գտնագրիչ: ՀՀ ԳԱԱ թղթակից անդամներ*, Elérhető: https://hy.wikipedia.org/wiki/%D4%BF%D5%A1%D5%BF%D5%A5%D5%A3%D5%B8%D6%80%D5%AB%D5%A1:%D5%80%D5%80_%D4%B3%D4%B1%D4%B1_%D5%A9%D5%B2%D5%A9%D5%A1%D5%AF%D5%AB%D6%81_%D5%A1%D5%B6%D5%A4%D5%A1%D5%B4%D5%B6%D5%A5%D6%80 (Utolsó elérés: 2023. 09. 16.)
- [68] Վիքիպեդիա, *Գտնագրիչ: ՀՀ ԳԱԱ արտասահմանյան անդամներ*, Elérhető: https://hy.wikipedia.org/wiki/%D4%BF%D5%A1%D5%BF%D5%A5%D5%A3%D5%B8%D6%80%D5%AB%D5%A1:%D5%80%D5%80_%D4%B3%D4%B1%D4%B1_%D5%A1%D6%80%D5%BF%D5%A1%D5%BD%D5%A1%D5%B0%D5%B4%D5%A1%D5%B6%D5%B5%D5%A1%D5%B6_%D5%A1%D5%B6%D5%A4%D5%A1%D5%B4%D5%B6%D5%A5%D6%80 (Utolsó elérés: 2023. 09. 16.)
- [69] PetScan, *24676957. azonosítószámú lekérdezés*, Elérhető: <https://petscan.wmflabs.org/?psid=24676957> (Utolsó elérés: 2023. 09. 16.)
- [70] PetScan, *24676974. azonosítószámú lekérdezés*, Elérhető: <https://petscan.wmflabs.org/?psid=24676974> (Utolsó elérés: 2023. 09. 16.)
- [71] Wikidata, *Armenak Nazarov*, Elérhető: <https://www.wikidata.org/wiki/Q62104559> (Utolsó elérés: 2023. 09. 16.)
- [72] Wikidata, *Wikidata:Statistics*, Elérhető: <https://www.wikidata.org/wiki/Wikidata:Statistics> (Utolsó elérés: 2023. 09. 16.)
- [73] Wikidata, *Wikidata:Scholia*, Elérhető: <https://www.wikidata.org/wiki/Wikidata:Scholia> (Utolsó elérés: 2023. 09. 16.)
- [74] Wikidata, *Kiszl Péter*, Elérhető: <https://www.wikidata.org/wiki/Q113145828> (Utolsó elérés: 2023. 09. 16.)
- [75] Wikidata, *Németh Katalin*, Elérhető: <https://www.wikidata.org/wiki/Q113145894> (Utolsó elérés: 2023. 09. 16.)
- [76] Wikidata, *Italian codices in Eötvös Loránd University Library*, Budapest, Elérhető: <https://www.wikidata.org/wiki/Q113148888> (Utolsó elérés: 2023. 09. 16.)
- [77] Nielsen, F. Å., Mietchen, D., Willighagen, E. Scholia, *Scientometrics and Wikidata*, In: The Semantic Web: ESWC 2017 Satellite Events: ESWC 2017 Satellite Events, Revised Selected Papers, Portorož, Slovenia, May 28–June 1, 2017, p. 237–259. <https://doi.org/10.1007/978-3-319-70407-4>
- [78] Odell, J. D., Lemus-Rojas, M., Brys, L. *Wikidata for Scholarly Communication Librarianship*, IUPUI University Library, 2022, <https://doi.org/10.7912/9Z4E-9M13>
- [79] Wikimedia Cloud Virtual Private Server, *Humaniki*, Elérhető: <https://humaniki.wmcloud.org/> (Utolsó elérés: 2023. 09. 16.)
- [80] Wikipédia, *Wikipédia:Wikidata/Friss halálesetek*, Elérhető: https://hu.wikipedia.org/wiki/Wikip%C3%A9dia:Wikidata/Friss_hal%C3%A1lesetek (Utolsó elérés: 2023. 09. 16.)
- [81] Wikidata, *Wikidata Query Service*, Elérhető: <https://w.wiki/6EJ5> (Utolsó elérés: 2023. 09. 16.)
- [82] Wikidata, *Wikidata Query Service*, Elérhető: <https://w.wiki/6YsC> (Utolsó elérés: 2023. 09. 16.)
- [83] Mihály, M., Maróthy, S. *Nyugat és Napkelet között: folyóiratok és írók hálózata*, Bölcsészettudományi Kutatóközpont, Elérhető: <https://btk.mta.hu/ismerettar/ismeretterjesztes/1598-nyugat-es-napkelet-kozott-folyoiratok-es-irok-halozata> (Utolsó elérés: 2023. 09. 16.)

Beérkezett: 2023. szeptember 18.



Molnár Bencea

PTE BTK Könyvtár- és Információtudományi Tanszékének egyetemi tanársegéde, az Európai Unió Kiadóhivatalának munkatársa

ORCID: <https://orcid.org/0009-0002-0274-7784>

E-mail: molnar.bence@hotmail.com

Melléletek

1. számú melléklet

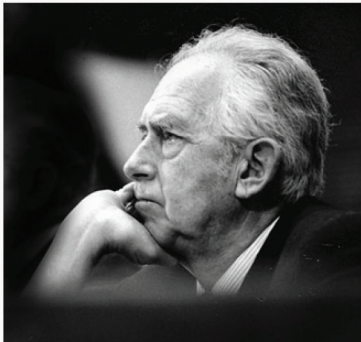
```
1. local entity = item
2. local statements = {}
3. if entity then
4.     statements = entity:getAllStatements(property)
5. else
6.     if options.entityId and options.pageTitle then
7.         return formatError('excluding-arguments', 'entityId', 'pageTitle')
8.     end
9.     if options.pageTitle then
10.        options.entityId = mw.wikibase.getEntityIdForTitle(options.pageTitle)
11.    elseif not options.entityId then
12.        options.entityId = mw.wikibase.getEntityIdForCurrentPage()
13.    end
14.    if not options.entityId then
15.        return options['felsorolás'] == 'table' and {} or nil
16.    end
17.    if mw.wikibase.isValidEntityId(options.entityId) then
18.        statements = mw.wikibase.getAllStatements(options.entityId, property)
19.    end
20. end
```

M1 ábra A magyar Wikipédia Wikidata-moduljának a Wikidata-elemazonosítók meghívásáért felelős Lua-kódrészlete

Szerzők: a Wikipédia szerkesztői, CC BY SA 4.0 licenc alatt (a szerző kiemelése)

Forrás: <https://hu.wikipedia.org/wiki/Modul:Wikidata>

2. számú melléklet

János Szentágotthai		Autres informations	
		A travaillé pour Université Semmelweis (1963-1986) Medical University of Pécs (d) (1946-1963) Université Loránd-Eötvös (1932-1946)	
Fonctions		Parti politique Forum démocrate hongrois	
Président Tényeket Tisztelők Társasága (d) 1992-1994		Membre de Académie hongroise des sciences (1967) Académie américaine des sciences (1972) Presidential Council of the Hungarian People's Republic (en) (1985-1989) Tényeket Tisztelők Társasága (d) (1992) Académie norvégienne des sciences et des lettres Académie slovène des sciences et des arts Académie royale des sciences de Suède Académie serbe des sciences et des arts Académie des sciences et des lettres de Mayence Académie finlandaise des sciences Académie des sciences de l'URSS (en) Académie pontificale des sciences Royal Society Académie royale de médecine de Belgique Académie américaine des arts et des sciences Académie Léopoldine Académie polonaise des sciences	
Député de l'Assemblée nationale de Hongrie 1990–1994 legislative term (d) 2 mai 1990 - 27 juin 1994		Conflit Seconde Guerre mondiale	
Député de l'Assemblée nationale de Hongrie 1985–1990 legislative term (d) 28 juin 1985 - 1 ^{er} mai 1990		Distinctions Prix Kossuth (1950)	
Président de l'Académie hongroise des sciences 26 octobre 1976 - 10 mai 1985 ◀ Tibor Erdey-Grúz (en) Iván T. Berend ▶		Liste détaillée [becsuk] Prix Kossuth (1950) Prix d'État de la République populaire hongroise (d) (1970) Prix Karl-Spencer-Lashley (1973) Ferrier Lecture (en) (1977) Membre étranger de la Royal Society (1978) Prix János Csere Apáczai (d) (1980) Honorary doctor of the Medical University of Pécs (d) (1980) Médaille académique d'or (d) (1985) Commandeur avec étoile de l'ordre du Mérite hongrois (1992) Prix de l'héritage hongrois (en) (2000) Docteur honoris causa de l'université de Turku Ordre de l'Amitié des peuples Docteur honoris causa de l'université d'Oxford	
Vice-président Académie hongroise des sciences 10 mai 1973 - 5 mai 1977		modifier - modifier le code - modifier Wikidata 1	
Biographie			
Naissance	31 octobre 1912 Pestújhely 		
Décès	8 septembre 1994 (à 81 ans) Budapest 		
Sépulture	Cimetière de Farkasrét 		
Nom de naissance	Schimert János 		
Nationalité	hongroise 		
Formation	Université Loránd-Eötvös (1930-1936) Université de Bâle (1939-1940) 		
Activités	Neurobiologiste , anatomiste , homme politique , professeur d'université 		
Parentèle	Géza Antal (d) (grand-père maternel) Sándor Lumniczer (d) (arrière-grand-père) 		

M2 ábra A francia Wikipédia „Infobox Biographie2” elnevezésű sablonja által Szentágotthai János Wikidata-elemének információiból generált infobox. A kép méretének csökkentése érdekében a sablonban letiltásra került a nyughely képe (P1442) és az emléktábla képe (P1801) tulajdonságok meghívása.

A felhasznált wikikód: „{{Infobox Biographie2|wikidata=Q662084|tombe=-|plaque=-}}”
 A szöveg és a portré CC-BY-SA 4.0 licenc alatt közzétéve, szerzője Eifert János, forrása:
<https://commons.wikimedia.org/entity/M76838001>

3. számú melléklet

```
1. SELECT DISTINCT ?item WHERE {
2.   ?item p:P31 ?statement0.
3.   ?statement0 ps:P31 wd:Q5. #osztály, amelynek példánya=ember
4.   ?item p:P463 ?statement1.
5.   ?statement1 ps:P463 wd:Q2379496. #tagság=Örmény Tudományos Akadémia
6. }
```

M3 ábra Az Örmény Tudományos Akadémia tagjait kigyűjtő SPARQL-lekérdezés. Azon személyek listája, ahol osztály, amelynek példánya (P31) = ember (Q5) és tagság (P463) = Örmény Tudományos Akadémia (Q2379496)

A szerző saját SPARQL-lekérdezése

Link a lekérdezéshez: <https://w.wiki/6Yg2>

4. számú melléklet



M4 ábra A Wikidata koordinátákat tartalmazó elemeinek pontszerű (százszoros elemintenzitású) hengervetületi ábrázolása a Wikidata 2023. június 26-ai adatainak megfelelően.

Az adatok és a vizualizáció CC0 licenc alatt közzétéve

Forrás: <https://wmde.github.io/wikidata-map/dist/index.html>

5. számú melléklet

```

1. #defaultView:Map
2. SELECT ?item ?itemLabel ?area ?geoshape ?layer ?rgb (CONCAT("Határos: ",
GROUP_CONCAT(DISTINCT(?neighbour_label); separator=', ')) AS ?neighbours)
3.
4. WITH {
5.   SELECT ?item ?itemLabel ?area WHERE {
6.     ?item p:P463 ?UN_stmt .
7.     ?UN_stmt ps:P463 wd:Q1065 .
8.     ?UN_stmt a wikibase:BestRank .
9.     MINUS {?UN_stmt pq:P582 []} .
10.    MINUS {?item wdt:P576 []} .
11.    ?item wdt:P2046 ?area .
12.    ?item rdfs:label ?itemLabel FILTER(lang(?itemLabel) = 'hu') .
13.  }
14. } AS %countries
15.
16. WHERE {
17.   INCLUDE %countries .
18.   OPTIONAL {
19.     FILTER NOT EXISTS {
20.       ?item wdt:P527?/p:P47 ?stmt .
21.       ?stmt ps:P47/wdt:P17? ?neighbour0 .
22.       MINUS {?stmt pq:P5102 wd:Q3089219} .
23.       FILTER(?item != ?neighbour0) .
24.     }
25.     BIND("n/a" AS ?layer) .
26.     BIND("777777" AS ?rgb) .
27.   }
28.   OPTIONAL {
29.     ?item wdt:P47 [] .
30.     FILTER NOT EXISTS {
31.       ?item wdt:P527?/p:P47 ?stmt .
32.       ?stmt ps:P47/wdt:P17? ?neighbour1 .
33.       MINUS {?stmt pq:P5102 wd:Q3089219} .
34.       ?item wdt:P2046 ?area1 .
35.       {
36.         SELECT (?item AS ?neighbour1) (?area AS ?neighbour1_area) WHERE {
37.           INCLUDE %countries .
38.         }
39.       }
40.       FILTER (?area1 > ?neighbour1_area) .
41.     }
42.     BIND("kisebb" AS ?layer) .
43.     BIND("CC3333" AS ?rgb) .
44.   }
45.   OPTIONAL {
46.     FILTER NOT EXISTS {
47.       ?item wdt:P527?/p:P47 ?stmt .
48.       ?stmt ps:P47/wdt:P17? ?neighbour2 .
49.       MINUS {?stmt pq:P5102 wd:Q3089219} .
50.       ?item wdt:P2046 ?area2 .
51.       {
52.         SELECT (?item AS ?neighbour2) (?area AS ?neighbour2_area) WHERE {
53.           INCLUDE %countries .
54.         }
55.       }
56.       FILTER (?area2 < ?neighbour2_area) .
57.     }

```

```

58.   BIND("nagyobb" AS ?layer) .
59.   BIND("88AAFF" AS ?rgb) .
60.
61.   }
62.   OPTIONAL {?item wdt:P3896 ?geoshape} .
63.   OPTIONAL {?item wdt:P527/wdt:P3896 ?geoshape} .
64.   OPTIONAL {?item ^wdt:P361/wdt:P3896 ?geoshape} .
65.   OPTIONAL {
66.     BIND("egyik sem" AS ?layer) .
67.     BIND("DDDDDD" AS ?rgb) .
68.
69.   }
70.   OPTIONAL {
71.     ?item wdt:P527?/p:P47 ?stmt .
72.     ?stmt ps:P47/wdt:P17? ?neighbour .
73.     MINUS {?stmt pq:P5102 wd:Q3089219} .
74.     {
75.       SELECT (?item AS ?neighbour) (?itemLabel AS ?neighbour_label )WHERE {
76.         INCLUDE %countries .
77.       }
78.     }
79.   }
80. } GROUP BY ?item ?itemLabel ?area ?geoshape ?layer ?rgb
81. ORDER BY ?itemLabel

```

M5 ábra A világ országainak (tagság [P463] = Egyesült Nemzetek Szervezete [Q1065]) hengervetületi térképét négy felhasználó által definiált réteggel leképező SPARQL-lekérdezés. Az összes szomszédjuknál kisebb és nagyobb (terület [P2046]) országok piros és kék, a vegyesen mindkettővel rendelkező országok világosszürke színnel vannak jelölve. A sötétszürke a határos (P47) tulajdonsággal nem, vagy állítás természete (P5102) = tengeri határ (Q3089219) minősítővel kiszűrt országokat jelöli.

A szerző saját SPARQL-lekérdezése

Link a lekérdezéshez: <https://query-chest.toolforge.org/redirect/WnHa7LML1kmWm4GuualWE0oIEMyQaqAIaC0EmCwYWqj>