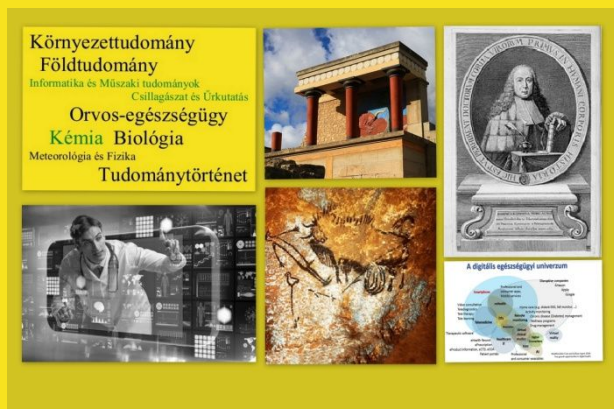


**A MAGYAR TERMÉSZETTUDOMÁNYI TÁRSULAT
TUDOMÁNYTÖRTÉNETI KÖTETEI
VI.**

**HAGYOMÁNY, ÉRTÉKMENTÉS ÉS INNOVÁCIÓ
A TUDOMÁNYOK KÖRÉBEN**



**DIGITALIZÁLT VILÁG-TÉR-KÉP A TUDOMÁNYOK,
A TECHNIKA ÉS AZ ORVOSLÁS KÖRÉBEN**



A természetvizsgálók hagyományát, immár 180 éve, 1841-től őrizve, az azóta megrendezett évi konferenciák eszmeiségét folytatta 2021-ben is a DIGITALIZÁLT VILÁG-TÉR-KÉP témával. A természettudományi és társadalomtudományi kutatások ma már elképzelhetetlenek a digitalizáció nélkül. Mégis minden tudományág más és más aspektusban, módszerrel alkalmazza ezt a lehetőséget. E különbségek közös módszerét, eszközt kívánjuk nyomon követni egyaránt az élővilág minden szerveződési szintjén, a társadalmi területeken, tudományos és intézményes formában, a természettudományban, az orvostudományban, az antropológiában, a szociológiában, az informatikában, a szabad társadalomtudományban, földrajzi, klimatológiai, mérnöki-műszaki stb. rendszerekben egyaránt. A kötet a 2021. november 18.–19. időpontban, online, Zoom rendszerrel megtartott *Digitalizált világ-tér-kép a tudományok, a technika és az orvoslás körében* című konferencia előadásait tartalmazza

**A MAGYAR TERMÉSZETTUDOMÁNYI TÁRSULAT
TUDOMÁNYTÖRTÉNETI KÖTETEI**

6.

Sorozatszerkesztő: Dr. Forrai Judit

A Hagyomány, Értékmentés és Innováció a
Tudományok történetében sorozat keretében

**DIGITALIZÁLT VILÁG-TÉR-KÉP
A TUDOMÁNYOK,
A TECHNIKA ÉS AZ ORVOSLÁS KÖRÉBEN**

Szerkesztő: Dr. Forrai Judit

Szövegszerkesztés, borítóterv és tipográfia: Pók Andrea

Budapest
2024

A MAGYAR TERMÉSZETTUDOMÁNYI TÁRSULAT TUDOMÁNYTÖRTÉNETI KÖTETEI

6.

Sorozatszerkesztő: Dr. Forrai Judit

Kiadja a Magyar Természettudományi Társulat
Budapest, 2024

Felelős kiadó: Dr. Tardy János



A kötet a 2021. november 18.–19. időpontban, online, Zoom rendszerrel megtartott *Digitalizált világ-tér-kép a tudományok, a technika és az orvoslás körében* című konferencia előadásait tartalmazza.

A kötetben másként nem jelölt webhelyek utolsó megtekintése: 2022. 12.16.

ISSN 2676-8852

ISBN ISBN 978-615-82104-6-1

ISBN 978-615-82104-7-8 [pdf]

Kötet DOI: <http://doi.org/10.23716/MTTT.6.2024>

Belovits-Print Kft.

Budapest

A kiadványra a Creative Commons – Ne add el! – Így add tovább! 3.0. licenc vonatkozik (CC BY-NC -SA 3.0). A licenc teljes szövegezése a következő linken olvasható: [http://creativecommons.org/licenses/by-nc-sa/3.0/!](http://creativecommons.org/licenses/by-nc-sa/3.0/)

TARTALOM

FORRAI JUDIT dr., DSc, egyetemi tanár, orvostörténész (S.E., WJLF): **A tudománytörténet és a digitalizáció** 5

I. TERMÉSZETTUDOMÁNYOK ÉS A DIGITALIZÁLT VILÁG TÉR-KÉPE

VERRASZTÓ ZOLTÁN DR. PHD., (a Közép-Duna-völgyi Környezetvédelmi Felügyelőség ny. igazgatója): **Több szempontú döntéstámogatás a térben**..... 8

KOMLÓSSY GYÖRGY DR.: (nyugdíjas): **Laterit bauxittelemek kimutatása távérzékelési módszerekkel** 27

TÓTH SÁNDOR LÁSZLÓ CSc, (ny. egyetemi tanár): **Digitalizáció a faiparban. Bútorgazdasági mozaikok – egy megfigyelő szemével**..... 40

II. EGÉSZSÉGÜGYI ELLÁTÓ HÁLÓZATBAN ALKALMAZOTT DIGITALIZÁCIÓ

SZALAI GY. DR., PHD, KATONA J. DR, KLENK G., SMEHÁK GY., MATESZ I., PINTÉR ZS., LEEL-ÓSSY A., SCHMIDT A. DR., HIRSCGEBERG A.: (Észak-Közép-budai Centrum Új Szent János Kórház és Szakrendelő Fül-, Orr-, Gége-, Fej-Nyak és Szájsebészeti Osztály): **Az arc-állsontsebészeti képpalkotás 3D-s rekonstrukciótól, a daganatok kezeléséhez használt molekuláris képpalkotásig**..... 45

SIMEK ÁGNES DR., PH.D., C. EGYETEMI DOCENS (SE Népegészségtani Intézet): **Digitalizáció az egészségügyben** 58

III. EMBERI MŰKÖDÉS ÉS FEJLŐDÉSTÖRTÉNETE A DIGITALIZÁCIÓS VILÁG SEGÍTSÉGÉVEL

DERGENZ-RIPPL DÓRA DSC, PhD (filozófia) (Pécsi Tudományegyetem, Filozófia Doktori Iskola KPVK): **Emberi vagy nem emberi? – A „hallgatólagos összetevő” elméleti jelentősége az MI-művészetben** 82

FORRAI JUDIT DR. DSC (S.E. Népegészségtani Intézet, WJLF egy. tanár): **Az epidemiológiai események rögzítése a papirusztekectől a digitalizációig**..... 90

IV. PEDAGÓGIA ÉS A DIGITALIZÁCIÓ

MUNKÁCSY KATALIN DR. Phd. Főisk. docens: **A MacTutor, matematikatörténet az interneten** 107

DR. KÁNTOR SÁNDORNÉ DR. VARGA TÜNDE PHD (főiskolai tanár, Apor Vilmos Katolikus Főiskola): **Geomatek – Függvények képei a monitoron** 111

HORVÁTH BALÁZS ZSIGMOND, CSORBA BOTOND (Budapesti Komplex Szakképzési Centrum Pogány Frigyes Technikum - Százhalombattai Eötvös Loránd Magyar-Angol Két Tanítási Nyelvű Tagozatos Általános Iskola): **3D-nyomtatással készült mondatkirakós játékok a nyelvoktatásban** 124

V. DIGITALIZÁLT TÁRSADALOMTUDOMÁNYI KUTATÁSOK

NAGY PÉTER TIBOR Prof. DSc, egyet. tanár, (WJLF): **Rejtett kézikönyv tartalmak feltárása** 134

SCHILLER VERA DR., bölcsészdoktorátus, nyugdíjas tanár, (SOTE) **Digitalizált világtérkép az ókortudományban** 162

VI. ELMÉLETI TÁRSADALOMTUDOMÁNYI KUTATÁSOK ÉS A DIGITALIZÁCIÓ

MOLNÁR LÁSZLÓ DR., PhD. habil, nyugalmazott egyetemi docens (BME): **Etika egy komputerezált társadalomban** 171

KISS ENDRE DR., egyetemi tanár, DSc. professzor emeritus (ELTE – OR-ZSE):
Egy bensőséges kapcsolat 183

HIDEG ÉVA DR. DSC egyetemi tanár, **Digitalizáció és informatizáció a jövő kutatásban** 196

V. DIGITALIZÁLT TÁRSADALOMTUDOMÁNYI KUTATÁSOK

REJTETT KÉZIKÖNYV TARTALMAK FELTÁRÁSA

NAGY PÉTER TIBOR Prof. Egyet. tanár, (WJLF)

E-mail:

DOI: <http://doi.org/10.23716/MTT.6.2024.11>

Absztrakt

Írásunkat, mely egy MTTT konferencia előadáson alapul azzal kezdenénk, hogy bemutatjuk, hogy a régebbi kézikönyvek digitalizálása során — melynek eredeti célja nyilvánvalóan csak a hagyományos könyvhasználatot kiváltó praktikus segítségnyújtás volt, hogyan tárultak fel a rejtett kézikönyvtartalmak általában. Ezt követően bemutatnánk, hogy a politikátörténeti források digitális használata hogyan „bányássza ki” a „rejtett tartalmat”, majd azt, hogy a valamilyen célból és logikával létrehozott kézikönyvek digitalizált formában milyen más logikájú rejtett tartalmat hordoznak. A szóképek amiket használunk nem véletlenek: az információ-tudományi szakirodalom a nem manifeszt információk megszerzésére az „adatbányászat” fogalmat használja, mely pontosan azokat a konnotációkat hozza elő, amit kell a könnyedén hozzáférhető szabad szemmel látható adatok felhasználása és a rejtett adatok felhasználása közötti különbségről. Ugyanakkor muszáj figyelmeztetni a magyar anyanyelvű olvasót, hogy a Delors jelentés az „Oktatás rejtett kincs” címen híresült el, itt kifejezett súlyos fordítási hiba Learning-et oktatásnak fordítani a within-t pedig rejtettnek, tehát kérem, ne asszociáljon arra. Kulcsszavak: kézikönyvek, 19–20.század, digitalizáció

Írásunkat, mely egy MTTT konferencia előadáson alapul azzal kezdenénk, hogy bemutatjuk, hogy a régebbi kézikönyvek digitalizálása során — melynek eredeti célja nyilvánvalóan csak a hagyományos könyvhasználatot kiváltó praktikus segítségnyújtás volt, hogyan tárultak fel a rejtett kézikönyvtartalmak általában. Ezt követően bemutatnánk, hogy a politikátörténeti források digitális használata hogyan „bányássza ki” a „rejtett tartalmat”, majd azt, hogy a valamilyen célból és logikával létrehozott kézikönyvek digitalizált formában milyen más logikájú rejtett tartalmat hordoznak. A szóképek amiket használunk nem véletlenek: az információ-tudományi szakirodalom a nem manifeszt információk megszerzésére az „adatbányászat” fogalmat használja, mely pontosan azokat a konnotációkat hozza elő, amit kell a könnyedén hozzáférhető szabad szemmel látható adatok

felhasználása és a rejtett adatok felhasználása közötti különbségről. Ugyanakkor muszáj figyelmeztetni a magyar anyanyelvű olvasót, hogy a Delors jelentés az „Oktatás rejtett kincs” címen híresült el, itt kifejezett súlyos fordítási hiba Learning et oktatásnak fordítani a within-t pedig rejtettnek, tehát kérem, ne asszociáljon arra. [BENCZÜR et al 2006, Oktatás 1997]

A papíralapon kiadott és beszkenelt, vagy levéltári forrásból bevitt adatok egyrésznél a manifeszt adatszerűség első pillanatban elnyomja a látens adattartalmat.

A digitalizáció nulladik formája az, amikor bekerült kézikönyv nem karakterkereshető, olyan mint a mikrofilm, vagy mikrofisch — jelentősége tudományszociológiai, amennyiben a mikrofilmek, mikrofischek tárgyak, melyeket valakinek, valamikor át kellett hozni pl egy határon, vagy le kellett szállítania egy vidéki városba, valaki másnak be kellett raknia egy köz vagy magángyűjteménybe, ahhoz hogy kutatói döntéssel — erre szolgáló géppel megnézhesse valaki. A digitalizált kézikönyv ez esetben tartalmi értelemben úgy hat, mintha a kutató kezében lenne, nem jobban, nem másképp. Elemezni a továbbiakban azt fogjuk, amikor a nulladik fázis nyomán karakterfelismerés is történt, vagy a képet –hangot valamilyen értelemben „intelligens” módon tárolták el.

A manifeszt adattartalmú digitalizációk egyik óriási csoportja a népszámlálási kötetekre, statisztikai évkönyvekre bőséges adatpublikációt végző statisztikai szakkönyvekre terjed ki. Ezek látszólag tisztán kvantitatív természetűek, azaz maguk a táblázatok, sok-sok számmal. Egy másik tanulmányunkban rámutattunk arra, hogy a „táblázatos adatok” miért nem tisztán „kvantitatív természetűek” mégsem, azaz a kategóriák, a kategóriákhoz fűzött definíciók, az adatfelvételi módok hogyan teremtenek kvalitatív „élethelyzetet.” [NAGY,2021] Mármint ha több ezer táblázatot digitálisan publikálunk, akkor a statisztikai kézikönyvben történő szókeresés lehetősége más táblázatok felhasználását idézi elő, mint a táblázatok címét tartalmazó mutatók alapján hozott kutatói döntés arról hogy mely táblázatot nézünk meg. Tehát: más számok „látványa” fog hatni a digitalizált forrást használó kutatóra, mint a nyomtatott kézikönyvet használóra.

A digitalizáció második csoportja a szövegekből álló, de eredeti formájukban is adattárként használandó kézikönyvek digitalizációja, ilyen típusú levéltári adatok adatbázisba vitele: ide tartoznak a listák: diák névlisták, iparos névlisták, virilis listák, névmutatók, telefonkönyvek, intézménylisták, településjegyzékek, címszójegyzékek. Ennek a digitalizációnak az eredménye már kifejezetten kvalitatív értelmező munkával válik csak számsorok alapjává: profiljuk, vagy fenntartójuk alapján csoportosítjuk az intézményeket, névjelleg szerint besoroljuk a neveket, döntünk a regionális határokról, melyek segítségével a települések besorolhatók stb. De miután ezt elvégeztük másképpen keresünk, más adatokat ismerünk meg, más adatokat összesítünk papíron vagy fejben, mint papíralapon tettük volna, ahol az

egyedi adat állt rendelkezésre. Papír alapon azt látjuk, hogy Bauer István 14 éves, másodikos gimnazista református tanulónak hányasa volt kémiaiából, ezzel szemben Nagy Zoltán stb., száz, vagy ezer különböző esetre. Digitalizált formában azt látjuk, hogy német névjellegű, kortársaihoz képest kettővel alacsonyabb osztályba járó református vallású diákok kémia jegye mennyiben tér el az átlagtól, mennyiben tér el a németek átlagától, az évviztesek átlagától, vagy a reformátusok átlagától. A rejtett kézikönyv tartalom kibányászása ebben az esetben maga a kategorizálás, a számítás.

A digitalizáció harmadik csoportja olyan kézikönyvszerű anyagokra terjed ki, amelyet eredetileg folyamatos emberi olvasásra szántak, vagy legalábbis nagyobb — mondatnyi, bekezdésnyi darabok egyben-olvasását feltételezte a szerző és szerkesztő. Utóbbiakról lesz szó az alábbiakban.

A digitalizáció negyedik csoportja szintén olyan anyagokra terjed ki, amelyet eredetileg folyamatos emberi olvasásra szántak. Csakhogy míg az előző típusnál, az eredeti mű szerzője, vagy szerkesztője mindenképpen azt állítja: elsősorban az igaz/hamis dichotómia alapján sorolható be a műve, mégpedig a szerző szándéka szerint mindenképpen az igaz végponthoz sorolva, azaz a könyv „non-fiction”, addig a negyedik csoportnál az eredeti szerző/szerkesztő nem is állítja, hogy a szövegek a hétköznapi értelemben vett igaz/hamis dichotómia alapján sorolandók be, sokkal inkább a szép/csúnya, izgalmas/unalmas, hiteles/hihetetlen dichotómia alapján — de alapvetően „fiction” jelleggel. Ez a fikció (regény, dráma, eposz stb.) azonban különféle foglalkozású, nemzetiségű, nemű, életkorú stb. szereplőkkel van tele, akik „valós” társadalmat alkotnak, mely társadalmilag jelentést hordozó tárgyi és természeti környezetben létezik — az író költő drámaíró által használt szavakból, szópárokából épül fel az adatbázis. Bármely mesét, vagy fantasys, science fictiont megvizsgálunk: a történetdarabkákban, a viszonyokban, a beszédmódban vagy azok a társadalmi viszonyok tükröződnek, melyek a szerzőt körülveszik, vagy azok, amelyeket létező tendenciák extrapolálásával utópiaként megalkotott, vagy azok amelyeket olvasmányélményeiből, más művekből a múlt társadalmáról megalkotott. Elemi-történet információk ezrei kötik Scott vagy Tolkien egyes településeit a viking valósághoz, a lovagkori Anglia világához, legalábbis ahhoz a historiográfiai tényhez, hogy a 19. ill 20. század történészei milyennek látták és ábrázolták a viking falvakat vagy a lovagvárakat. Információk ezreit, művek ezreiben csak digitalizált formában lehet elemezni, egyenként elolvasva őket nem.

A digitalizáció ötödik csoportja terjed ki a hangzó és a képi, mozgóképi forrásokra, részben azokra, ahol velük kapcsolatban bevitt szöveges adatokban való keresés jelenti a használati módot, részben azonban olyanokra, ahol képkockák, vagy néhány hangra kiterjedő keresés lehetséges az adatbázis egészében. Utálnék az 1920-as évekbeli „Filmhíradó” c magyar sorozat hatására, vagy az 1960-as évekbeli „TV Híradó” sorozat ismeretterjesztő, történelmi szemlélet formálójá-

rétegspecifikusan megragadható hatására. A „megtekintés pillanatában” akár egyformán is hathat a kettő, de ha az egyik digitalizálódik és felkerül a netre, a másik pedig nem, akkor 50 év alatt a hatás erősen különbözni fog.

A digitalizáció hatodik csoportja azokra a hangzó és a képi, mozgóképi forrásokra koncentrál, ahol lényegileg fikcióról van szó, tv játékokról, játékfilmekről, mozifilmekről, rajzfilmekről, ahol mindazok a tényezők jelentik az adatbázis tartalmát, amire a negyedik csoportnál utaltunk, kiterjesztve azonban hálózat szerűen arra, hogy a szerepeket játszó színészek, rendező által kiválasztott-felépített helyszínek milyen fiktív, de összességében a kultúra valóságát reprezentáló — adatokat, műveket, szerepeket hívhatnak elő a nézőben, hallgatóban. Olyan „tömegkulturális” jelenségekről van szó, mint hogy a Simon Templár/ Angyal szerepkör matematikailag lekódolható hatásai alapján nem kultúra kutatói, hanem a filmeket egyáltalán nem ismerő empirikus szociológus eszközeivel jósolható meg, hogy évekkel a sorozat vetítésének befejeződése után a Roger Moore-t milyen szerepekben kell ahhoz felléptetni, hogy meghatározott rétegben sikeres legyen a film, hogy az „Angyal” sorozat 1970-es évekbeli magyarországi változatából néhány — KGB re utaló képkocka, párbeszéd kimarad — ezáltal a film jelentése megváltozik. Közel keleti tv csatornáknak ugyanebből a filmből az erotikus mozzanatokot vágják ki. A megfigyelt jelenség maga a vágás, ezer filmnél, bizonyító leírás csak kódolás után lehetséges.

A hatféle digitalizált tartalom megegyezik abban, hogy ha ugyanazon a módon fogyasztjuk, ahogyan a digitalizálás előtt tettük volna, semmiképpen nem tárulnak fel azok a tartalmak, amelyek digitalizálás, kódolás, számszerűsítés, stb. után igen. Van tehát rejtett tartalom. [TAKÁTS 2004, KEREKES 2015, RESS 2006, REISZ 2018]

Forrás szöveg – kézikönyvek

A preszcientikus és pozitivista korszakban kiadott óriási tömegű *eredeti forrásszövegalapú kézikönyv* digitalizálása nem egyszerűen a gyorsabb - és a forrásművek eredeti szerkezetétől független — keresést tette lehetővé, hanem olyan összefüggések feltárulását — metaadattá válását — is, amelyek az eredeti mű szerkesztőinek fejében sem lehettek meg. (vö. még: <http://www.edu-online.eu/hu/letoltes.php?fid=tartalomsor/2446>)

Több tízezer koraközépkori magyar, német, cseh, francia területeken született egyaránt latin nyelvű oklevél szófordulatait a szövegkiadások tudós szerzői nem hasonlíthatták szisztematikusan össze még egyetlen kézikönyvön belül sem, nem beszélve a különböző országokban kiadott források áttekinthetlenségéről. Furcsa mód tehát a középkori európai latin nyelvű adattömeg az eredeti - zömmel 19. századi - kiadások bázisán nemzetileg széttagolt adattömegként jelenik meg, míg a digitalizáció után *megkezdődhet az „eredeti,” azaz a nemzetállami határoktól*

független középkori kultúra kutatása. A szó és szókapcsolattömeg gépi elemzése megnyitja az utat az olyan forráselemzés felé, ahol a feldolgozónak *nem kell elolvasni* az eredeti forrásokat. A digitalizáció teszi lehetővé, hogy a történész egységnyi idő alatt többet, mélyebben ismerjen meg ebből a kultúrából, mint aki benne élt. *A digitalizáció hozza ki ismeretelméletileg győztesen a folyamatokat átélő megfigyelő és történeti elemző közül az utóbbit, legalábbis potenciálisan. Arisztotelész – bármennyire is szomorú ez a posztmoderneknek – győz: minden emberi tudás taníthatóvá és tanulhatóvá válik.*

Sok százmillió oldalnyi nemzeti nyelvű történeti forrásszöveg publikálására került sor a 19–21. században, melyek egy történelmi másodpercre aktuálisak voltak, de azután legalábbis a gazdagabb könyvtárakban kötetekbe kötettek, mintegy azt sugallva, hogy egy időszakra egy témára autentikus tudásegységek ők, magától a kronológikusságtól az adott forrásra vonatkozó teljeskörűségtől válva kvázi kézikönyvvé. Amikor őket digitalizáljuk, illetve használjuk őket digitalizáltként rejtett kézikönyv tartalmak táruznak fel.

A nemzeti nyelvű, tudósok, szakértők által írott kézikönyvekre is igaz, hogy a digitalizálás olyan metaadatokat teremt, amelyek sem a szerkesztők, sem a szerzők számára nem álltak rendelkezésre. Például a világ tudományosságában számon tartott egyik leghosszabb terjedelmű kézikönyv, amit egyetlen ember [SZINNYEI József] írt a Magyar írók Élete és Munkái 14 vaskos kötete, illetve annak még terjedelmesebb, de torzóban maradt folytatása — mely Gulyás Pál nevéhez fűződik. E könyvek szerzői — akik céduláikat természetesen a nevek betűrendjében tárolták — sem lehettek tisztában például az egyes magyarországi települések nevének könyvükben való előfordulásával, vagy egyes szavak és fogalmak felbukkanásával az idézett könyvcímekben. A digitalizálás elemi formája mely szabadszavas keresést tesz lehetővé adatok millióit „teremt meg” e kézikönyvekből, s minimális adatbázisba szervezés, mely például a címszavakat, dokumentumokat egymástól elhatárolja, vagy akár elhatárolás nélkül az egymáshoz fizikailag közeli találatokat tekinti összetartozónak már kombinatív kereséseket tesz lehetővé. Makrók segítségével sokezer szókapcsolatra kereshetünk rá.

A szöveges „adatbázisok” — alapvető forrásként szolgálnak a gyakoriságokat és számszerű összefüggéseket kutató kvantitatív és az egyetlen életpálya vagy jelenség potenciális kapcsolatait felgöngyölítő, egyidejűségeket, párhuzamosságokat mérlegelő klasszikus kvalitatív elemző számára is.

Az interneten használható — s különösen az ingyenesen használható — kézikönyvek idézési valószínűsége lényegesen megnövekszik az interneten nem használható újabb könyvekkel szemben. A kiadói érdekek alapján számos modern kézikönyv nem, vagy csak igen korlátozott hozzáférhetőséggel kerül a netre (pl. a könyvet papíralapon is megvásárlók, vagy előfizetők számára). Ez növelte annak valószínűségét, hogy a régebbi, szabadon hozzáférhető kézikönyvek, ill. az ingyenes

site-okra felhordott kivonatok szolgáljanak adatforrássul. A torrentezés világa azonban erre is kiterjed. A torrentezés világa a szólásszabadság és az információszabadság világa: tudományos individuumok számára hasznos, könyvkiadó érdekek, intézményes érdekek szempontjából káros. A torrentezés ugyanannak a háborúnak a partizánalakulata, mint amelynek a open access elv a reguláris hadserege. 2030-ból dönthető majd el, hogy torrentezés súlya Nógrádi vagy Tito alakulatainak szerepéhez lesz hasonlítható mármint a Szovjet Hadsereg 1944-es szerepvállalásához képest.

A mesterséges intelligencia (azaz a fordítógép a 2020-as évek elején) e tudástömeg nemzeti széttagoltságának megszüntetésére is képes.

A 19–21 század politikátörténései számára parlamenti naplón és a sajtón keresztül mutatjuk be a gépi fordítás jelenségének hatását. Később kifejtjük azt is, hogy nemcsak a források, hanem a szakkönyvek adatainak átcsoportosítása, más logika alá rendelése is új adatokat teremt. Három saját példát mutatunk be, az egyik amikor egy tematikus lexikont kézimunkával tematikusan indexelünk, a másik amikor egy személylexikonból gépi úton kronológikus, intézményes és topografikus fókuszú adtabázist készítünk, s a harmadikat, amikor kronológikus kézikönyvből személyes, intézményes és topografikus fókuszú adtabázist készítünk. [SIPOS 2000, KÓTA 2015, VITA 2000]

A parlamenti napló, mint forrás-kézikönyv

A parlamenti jegyzőkönyveket, naplókat nagyjából azóta használják az úgynevezett „nagypolitika” elemzésére, amióta ilyenek készülnek. Ez főleg kvalitatív munka — „nagy emberek” „látványos” összecsapásainak — szövegelemzése, sok retorikai minősítő-elemet megmozgatva. Sokkal kevésbé népszerű forrás a parlamenti jegyzőkönyv, hogyha nemzetközi összehasonlítást teszünk, vagy ha valamilyen ágazattörténeti kérdést — tehát pl. oktatás-történetet, egészségügy-történetet, közlekedés-történetet stb. elemzünk. Ez érthető — vannak specifikusabb források erre, az az előfeltevés, hogy amikor egy-egy ágazati törvény születik, a szakfolyóiratok jobb, részletesebb tanulmányokat tartalmaznak. Ami pedig kronológiailag, vagy „komparatistikailag” érdekes-releváns kérdés, tehát pl. „hogyan változik a magyar mezőgazdaságpolitika 1870 és 1880 között” vagy „miben és miért különbözik a porosz, dán, magyar, osztrák és román mezőgazdaságpolitika az 1870-es években” az éves költségvetési vitákból lenne tanulmányozható, ami egy nem elektronikus szövegben gigantikus feladat, hiszen a költségvetés általános vitájához hozzászóló képviselő a kétórás beszédében vagy szán tíz percet a mezőgazdaságpolitikának vagy nem, s ha van valami konkrét lobbizó javaslata, azt a fejezeti vitában vagy megismétli, vagy nem. (Magyarán: Túl sok mindent kell elolvasni a várható haszonhoz képest).

Csakhogy a parlamenti jegyzőkönyveknek van valamiféle olyan tulajdonsága, amittől egyrészt a történettudomány analitikus, algoritmikus, szociológikus, tudásszociológikus irányú kitágulása miatt e jegyzőkönyvek lehetnek éppolyan fontosak, mint az ágazattörténeti források, másrészt a digitalizáció — s különösen a legújabb, mesterséges intelligenciával kapcsolatos fejlemények a hagyományos ellenmotivációkat eltávolították — erről lesz szó a továbbiakban. [BARTOS 2020, MIHÁLYI 2016, SIEGRIST 2009, NOIRIEL 2002, ERŐS 2010, RÁCZ 2017, WELKER 2013, ROMSICS 2010]

A parlamenti viták — ha törvénygyárnak tekintjük a parlamenteket, és csak a gyár termékei iránt érdeklődünk — nem feltétlenül fontosak, hiszen pl. a Gömbös kormány törvény előkészítő munkája, vagy a második Orbán-kormány törvényelőkészítő munkája és/vagy a nagyobb kormánypárti frakciók erős kézben tartottsága szinte kizárja, hogy a törvényszövegekben érdemi változások történjenek. Ez azonban — és ez a hozzászólástömegben való digitális tájékozódással, szógyakoriség elemzéssel végezhető munka — egyáltalán nem zárja ki, hogy a parlamenti naplót, mint ágazat politikailag releváns viták *legjobb tartalmi forrását használjuk*

Olyan forrást, melyet tárgyi tartalmában *nem torzít* az egyes szereplők eltérő hozzáférése a médiához, magyarárn, hogy kinek a szövegét mennyire torzítja, szelektálja a média.

A 19–21. század történetében a többpártrendszerű politikai rendszerekben, a parlamentben mindig *megjelennek* az eltérő társadalmpolitikai és oktatáspolitikai, egészségpolitikai stb. álláspontok, mivel a többpártrendszer természeténél fogva (szavazatokért való versengés, különbözőség hangsúlyozásának igénye) megjeleníti ezeket valamilyen mértékig. Még akkor is, ha e többpártrendszerű politikai rendszerek a liberális jogállam kritériumai szerint továbbra is erősen különböznek egymástól. (A különböző választási rendszer vagy a médiaszabályozás a vélemények elterjedésének, kormány rákerülési esélyeinek befolyásolására képes, a vélemények mögött álló társadalmi csoportok méretére vonatkozó látszatot eltorzíthatja, a véleményeltérések *tartalmát* elemző kutatót viszont nem akadályozza *maguknak a véleményeknek* a megismerésében.)

Aki nemcsak azt ambicionálja, hogy a 19–21 századra nézve a politikai álláspontok *tartalmát* standard — úgy nemzetközileg, mint idősorosan összehasonlítható — formában megismerhesse, hanem azt is, hogy az egyes álláspontok *súlyát* is elemezhesse, az joggal fogalmazza meg azt a fenntartást, hogy a társadalomban fellelhető vélemények súlyának *arányos* tükröződése a parlamenti nyilvánosságban sem egyenletes, sem időben, sem a nemzetközi térben. A választójog korlátozottsága, a tisztán választókerületi többségi választójog következtében nyilvánvalóan „látható” méretű pártálláspontok – pl. az 1918 előtti

magyarországi szociáldemokratáké, vagy az 1930-as évekbeli angol nációké — sem általános sem szakkérdésben nem jelenhettek meg a parlamentben.

Olyan esetekben, amikor választójogi okai ennek nem voltak a fennálló kormány a választások előtt és alatt jogszabályalkotással, közigazgatási, ügyészi aktivitással akadályozhatta meg, hogy egy nyilvánvalóan létező vélemény képviselői a parlamentbe bekerüljenek. Az 1918–19-es magyarországi városi tömegbefolyásuk ismeretében, illetve 1922-es eredményeik ismeretében nem hihető például, hogy a szociáldemokraták legalább néhány képviselővel ne kerültek volna be az 1920-ban megválasztott parlamentbe, hogyha választási kampányukat a közigazgatás-bíráskodás nem akadályozza olyan mértékben, hogy a párt végül is a visszalépés mellett kénytelen dönteni, (ami azért, , tehetjük hozzá off-topic megjegyzésként 2023 januárban- napjainkig figyelemreméltó példája a „diktatórikus körülmények között született választási eredményeket nem legitimálunk” politikai moráljának).

A konszolidált Horthy korszakban valamekkora befolyással azért mindenképpen rendelkező és ezt pl. a szakszervezeti mozgalom és a szociáldemokrata ifjúsági szervezetek választásain megjelenítő kommunista párt az 1921:3 tc. miatt nem juthatott mandátumhoz a parlamentben.

Az 1939-es választási eredmények és a 1944 márciusa után regnáló rendszer társadalmi támogatottságának jelzőszámai (pl. a szélsőjobboldali sajtó népszerűsége, a nyilas párttagok száma, az elkobzott zsidó boltokra benyújtott igények száma) valószínűtlenné teszik, hogy a Szövetséges Ellenőrző Bizottság permisszívebb álláspontja esetén ne kaptak volna mandátumot az 1945-ös választáson szélsőjobboldaliak. (Az 1945-ös választáson minden korábbi és későbbi meghaladó részvétel mellett a szélsőjobboldali polgárok — mai tudásunk szerint — a kommunisták, parasztpártiak és kiscgazdák között osztották meg szavazatukat, azaz csak a szociáldemokrata párt és polgári demokrata párt nem lehetett vonzó számukra, de még ezügyben is várhatók talán emlékezetpolitikai „csalódások”.)

A digitalizáció azonban ezen a torzításon némiképp segít: ugyanis, hogyha kvantitatív módon az *egykori* parlamenti párt hozzászólásaiból, a *leendő* parlamenti párt megfigyelési időpontunkban még nem létező hozzászólásaiból összeállítunk egy témalistát, vagy kifejezéscsoportot, ami parlamentbe nem került párt „ügye”, (vagy ha muszáj kvalitatív tudással — sok szakkönyv elolvasásával — parlamenten kívüli forrásokból tesszük ezt) mégiscsak szerezhethünk arról képet, hogy *más* pártok képviselőin „keresztül”, hogyan politizál a parlamentbe nem jutott párt. Esetleg a parlamentben benn lévő, de pártként még *nem létező* véleménycsoport is megismerhető így: a későbbi MIÉP a későbbi MDNP, későbbi „kis-MDF” képviselők hozzászólásait elemezhetjük az 1990–1994-es magyar parlament „nagy-MDF” frakció tagjai hozzászólásainak elemzésekor, a későbbi (2010–2014-es)

Demokratikus Koalíció és „kis - MSZP” hozzászólásait elemezhetjük a 2006–2010-es magyar parlament „nagy-MSZP” frakciótagjai hozzászólásai elemzésekor – hiszen azok az emberek, akik X időpontra – a parlamentből való kihullást is kockáztatva eljutnak az egymással való szakításig, nyilván már x-1 vagy x-2 időpontban is a legtöbb dologról szövegszerűen, szó és szókapcsolat gyakoriságon keresztül megragadhatóan eltérő dolgokat gondoltak.

Mi több *a képviselők listájából nyerhető ki* azok a nevek is, akiknek nyilatkozataira a következő vagy előző ciklus idején, amikor a párt éppen nem parlamenti párt érdemes rákeresni *nem parlamenti nyilvánosságban* — ezt teszteltük azzal, hogy az 1922–1926-os ciklus szociáldemokrata képviselőinek nevére kerestünk rá az 1920–1922-es Népszavában. A frakció létszám csökkenése illetve a frakció összetételének átalakulása teremt egy csomó kereshető nevet — azokét akik az előző ciklusban voltak képviselők, az ő nevükkel kapcsolatba hozható találatok arányai mutathatják a „még fontos” de a parlamenti frakció szintjét el nem érő dolgokat. (Nyilvánvalóan: szavak, szópárok felülreprezentációjáról van szó e személyekhez kötődő Népszava szövegekben, a teljes Népszava korpuszhoz viszonyítva.) Másik tesztünkben az 1998–2002-es MIÉP képviselők listájából nyertük ki a neveket, melyekre a 1994 és 1998 közötti, illetve 2002–2006 közötti Magyar Nemzetben kereshettünk rá. (Olyan szövegtömegekről van szó, mely kizárólag gépi úton elemezhető. Az alapanyag a <https://adt.arcanum.com/hu/> -része, digitalizálási produktum, digitálizáltság nélkül lehetetlen lenne az eljárás.)

A parlamenti naplók véleménymegjelenítő tulajdonságát — pontosabban azok összehasonlíthatóságát — akadályozza a parlamenti küszöb eltérő volta, illetve az arányosságot biztosító „listás” jelleg jelenléte vagy hiánya az egyes országokban ill. időszakokban. Az alacsony parlamenti küszöb és a listás jelleg erőteljessége a viszonylag kis véleménycsoportok önálló megszerveződését, mandátumszerzését és a vélemény felbukkanását valószínűsíti a parlamenti naplóban. (Ennek közismerten markáns példája a tisztán listás izraeli parlament, ahol egyetlen mandátumos pártok is vannak, a másik véglet a választókerületen belüli többség kizárólagos megjelenését lehetővé tevő brit parlament, ahol elvileg előfordulhatna, hogy a szavazatok 50,001%-t megszerző párt a mandátumok 100%-át birtokolja).

A — német és magyar esetből ismert — bejutási küszöbös ill. teljesen lista nélküli (brit) rendszerekben a kisebb csoportok a szembenálló nagy pártokhoz csapódva juttatják be elittagjaikat a nemzeti parlamentbe — s ennek megvannak a következményei arra nézve, hogy mely társadalompolitikai, oktatáspolitikai stb. álláspont jelenhet meg a parlamentben. (Nyilvánvaló például, hogy a brit vagy amerikai politikai erőtér többszínű, mint ami a parlamenti jegyzőkönyvekben tükröződik, mert nagy pártok frakciótagjaként, széles választói csoportokra építő egyéni képviselőként nem képviselhetőek olyan markáns kisebbségi vélemények,

mint pl az olasz radikális párt színeiben...) De mégis: a kétpártrendszerű parlamentekben a felszólalások szövegelemzésével éppúgy kiállíthatunk mondjuk 5 csoportot, ahogyan egy formálisan is 5 párti parlamentben. Kizárólag azért tehetjük ezt, mert a teljes szöveg a rendelkezésre áll.

A parlamenti küszöb közelében lévő pártok elitje nem csapódik hozzá az esélyesebb „nagytestvérhez”, tehát — legalábbis arra a ciklusra, amit az jellemez, hogy *éppen* kimaradtak belőle — hangjuk nincs jelen : az 1990–2010 közötti magyar történelemből jó példa erre a Munkáspárt, a Szociáldemokrata Párt, a Magyar Igazság és Élet Pártja. (Más kérdés, ahogy erre már utaltunk, hogy e pártok elitje, a következő parlamenti választásnál már egy nagyobb párt listáján bukkan fel, így az eredetileg MIÉP-es FIDESZ-es képviselők, eredetileg Szociáldemokrata Párthoz tartozó MSZP-s képviselők álláspontja a következő parlament jegyzőkönyvéből már elemezhető.)

A *legfontosabb társadalompolitikai, oktatáspolitikai* kérdések — ha másutt nem az éves költségvetés vitájakor — *minden* parlamentben, minden politikai párt képviselői felszólalásaiban felbukkannak. Az egészségről, az oktatásról „mindenkinek” van véleménye, álláspontja — ha hallgat róla az önmagában jelzés értékű, szinte csak olyankor fordul elő, hogyha egy koalíciós megállapodás a *másik* kormánypárt hatáskörébe utalta azt. A kisebb kérdések, napi aktualitások, „botrányok” tárgyalási lehetőségét nagyban befolyásolja viszont az egyes országok parlamentjének különböző házszabálya, mely szintén az összehasonlíthatóság bizonyos korlátját jelenti. (praktikusan: hol-mekkora lehetőség van egyedi ügyeket interpelláció formában vitatni.) De: az „elmondani kívánt” történetek száma és iránya mégis összehasonlítható, hiszen a költségvetés tárgyalásakor, a tárca általános bírálatakor az interpellációt akadályozó parlament képviselője is sort keríthet erre. A „visszásságok számbavétele” még az egypárti parlamenteket is jellemezte, tehát — bármilyen hihetetlen is ez — eldönthető, hogy az 1958-ban, 1963-ban, 1967-ben, 1971-ben, 1975-ben „választott” parlamentben hogyan alakult a képviselők véleménynyilvánítási pluralizmusa. Ha a szöveg digitalizált.

A 21. században (az EU bővülése következtében is, meg az európai parlament jelentőségének növekedésével is) a tisztán listás európai választások parlamenti jegyzőkönyv-szinten láthatóvá tehetnek olyan véleményeket, amelyek a nemzeti parlamentekből kimaradtak. Közismert e vonatkozásban a strasbourgi magyar és a budapesti parlamenti képviselők összetételének különbözősége 2019 után, legalább két vonatkozásban, azaz az ellenzéki-FIDESZ-es képviselők arányát tekintve, másrészt, a budapesti parlamentből teljesen hiányzó liberális szín intézményes megjelenése révén (Momentum-frakció). A strasbourgi brit és londoni parlamenti képviselők pártpolitikai összetételének különbözősége a Brexitet megelőző évtizedekben, láthatóvá tette, hogy a brit politikai élet két-három klasszikus pártja mellett más erők is vannak. E forrás nélkül érthetetlen lenne a 21 századi brit

történelem eddig legfontosabb eseménye a Brexit. Az európai parlament tisztán listás jellege tehát *javítja* az európai parlamenti jegyzőkönyv használhatóságát az egyes országok véleményeskálájának megismerése szempontjából, bár kétségtelen, hogy az európai hatáskörbe *nem* tartozó témák felbukkanási valószínűsége kisebb. (A kutatók szerencséjére az erőteljesebb európai föderalizmus iránt elkötelezett politikusok folyamatosan keresik annak módját, hogy miképpen lehet egyes európai hatóságok hatáskörét úgy értelmezni, hogy az európai értékekkel szembenálló nemzeti szakpolitikákkal szemben használhatóak legyenek. Így például az európai biztosok személyével kapcsolatos vitákban olyan kérdések is szóba kerülnek, melyek nem „európai hatáskörök”. Némiképp hasonló ez ahhoz, ahogy az Amerikai Egyesült Államok nagyobb központi befolyást kívánó oktatáspolitikusai a szigorúan tagállami hatáskörnek minősülő oktatáspolitikát a 20-21 században nemzetbiztonsági üggyé, vallásszabadsági kérdéssé, faji egyenlőségi kérdéssé stb. fogalmazzák át, e kérdésekben ugyanis már a szövetségi hatóságok, bíróságok hatásköre is vitathatatlan.)

A parlamenti jegyzőkönyvek minden más forrásnál cenzúrázatalanabban tükrözik a helyben elhangzott álláspontot, nem kell külön forráskritikával élni, (ami minden más forrásnál — pl. nem szó szerinti jegyzőkönyvvel rendelkező tüntetés szónoklatainál, vagy tudományos ülés jegyzőkönyvével fennáll), hogy a cikk, a tudósítás stb. mennyire az eredeti véleményt tükrözi, s mennyire az újságíróét, a szerkesztőségét, a jegyzőkönyvvezetőét stb.

Ellenérvként fogalmazható meg, hogy a parlament honlapján vagy a Hungariananában hozzáférhető jegyzőkönyv nem tökéletesen tükrözi az ülésen elhangzottakat, mert a képviselőknek jogukban áll a gyorsírói feljegyzés és a publikálás között belenyúlni a szövegbe. Az ellenérv azonban cáfolható, ugyanis annak is digitálisan nyoma marad, hogyha egy képviselő szóban teszteli bizonyos szavak kimondhatóságát, majd a botrányt érzékelve rákényszeríti a jegyzőkönyvvezetőt a szavak módosítására. Erre példa a „Hordót a zsidónak” — „Hordót a szónoknak” történet, a szélsőjobboldal nevezetes kísérlete az antiszemita fegyver használhatóságának tesztelésére 1990 szeptemberében. (Fiatalabb olvasók kedvéért: egy populista politikus nyilvánvalóan azt akarta tesztelni, hogy egy liberális politikus felszólalása közben bekiabált „Hordót a zsidónak” szavak, mekkora részét csábíthatják át a szélsőjobboldali választókörnek. A többség és az elitcsoportok körében pedig a jegyzőkönyvmódosítás jogával élve — miszerint a „Hordót a szónoknak” szavak hangzottak volna el — kívánta a pártjával szembeni elidegenítő hatást csökkenteni. Szempontunkból azonban a „botrány” részletes utólagos parlamenti pertraktálása következtében az egyébként nem túl gyakori „zsidó” szó 1000 szóra eső gyakorisága megnőtt, tehát ha nem emlékeznénk az esetre, akkor is — a zsidó szót indikátornak tekintő kutatás megtalálta volna a

botrányt.) Tehát: a digitalizálatlan parlamenti szöveg nem lenne alkalmas az antiszemitizmussal kapcsolatos jelenségek kutatására, a digitalizált pedig alkalmas.

A parlamentben elhangzó vélemény „beszélője” legalábbis egy kormánypárti-ellenzéki tengelyen a megfigyelő számára triviálisan beazonosítható — míg az oktatással, egészségüggyel, közlekedéssel stb. kapcsolatban a nyilvánosságban megjelenő vélemények többi beszélőjével külön kutatást kell végezni, hogy az illető elhelyezkedését a politikai térben az elemző megállapítsa — s e megállapítással való olvasói egyet nem értés az egész interpretáció elfogadhatóságát relativizálja. A „parlamenti képviselők” csoportja mindenütt zárt számú, nem kell külön mérlegelést végezni, mint az oktatásügyi, egészségügyi, közlekedési „stakeholderek”, érdekcsoport álláspont képviselők esetében, hogy mekkora véleményritkaságig kell „lemenni” a kisebbségi álláspontok feltérképezésében. A nemzeti ki kicsodák, parlamenti almanachok nagy biztonsággal tartalmaznak adatot a megszólaló képviselők születési helyéről, iskolázottságáról, pályájáról, multipozicionális elit tulajdonságairól (pl. mely egyesületek vezetőségi tagja stb.) — melyek ismerete egy diskurzuselemzés számára alapvető, de a többi (mármint nem parlamenti) oktatásügyi, egészségügyi, közlekedési megszólaló esetében a rendelkezésre álló adatok nem szisztematikusak. A kényelmes többséggel átvitt 1924-es és 1934-es magyar középiskolai törvények, az 1993-es oktatási törvény vitáiban nemcsak az ellenzéki képviselők, de a kormánypártiak is konkrét intézmények, szervezetek lobbistáiként viselkednek — a nagybetűs szavakból készített lista pl. tökéletesen igazolja ezt.

Természetesen, ahhoz hogy a „digitális parlamenti napló” forrásértékét most bemutassuk, valamiképp standardizálni kellett, hogy mit tekintünk *egyébként is* ismertnek, *doxikus* információnak. A régebbi magyar történeti eseményekkel kapcsolatos példákat Benda Kálmán Magyarország történeti kronológiáját használva ellenőriztük, egyetemes történethez Pók Attila Világtörténet évszámokban-ját használtuk. Az utóbbi évtizedek magyar és egyetemes történeti kronológiájaként egy szerkesztőségi munkatárs szívességéből használatra megkapott HVG CD szolgált.

A parlamenti jegyzőkönyvek — legalábbis a többpártrendszerű országok vonatkozásában — európai nagy könyvtárakban — s ide tartozik budapesti parlamenti könyvtár is — már a 2010-es éveket megelőzően is rendelkezésre álltak. A komoly hazai elemzés külföldi parlamenti jegyzőkönyvek bázisán, mégis ritka volt, mint a fehér holló.

Az e ritkaságot érthetővé tevő három legfontosabb akadályozó körülmény azonban 2022-re tovatűnt.

Az egyik akadályozó körülmény az volt, hogy bár *maguk* a parlamenti naplók rendelkezésre álltak a megszólalók „pozíciójának” értelmezéséhez rengeteg olyan, a képviselők személyére vonatkozó, információra volt szükség, melyet sem a

parlamentari könyvtár, sem más hazai könyvtárak nem biztosítottak, hiszen természetesen nem voltak meg olyan „helyi lapok” melyek olasz, német tartományi nyilvánosságokat tükrözve akár országnyi területek helyi politikáit, oktatáspolitikáit, közlekedéspolitikáit tükrözik, és segítenek elhelyezni egy-egy politikust különféle tengelyeken.

A helyi nyilvánosság még nem demokratikus rendszerekben is képes a véleménykülönbségek mögötti folyamatok megértésére: az egykori neves kelet-európai államszocializmus szakértő Daniel Chirot, aki 1989 után Kína szakértővé képezte át magát, árulta el egyszer azt a módszertani trükköt, hogy amit a Népszabadság szövegeinek mélyelemzésével, vagy a Népszabadság/Magyar Nemzet/Népszava összehasonlításával az 1980-as években sikeresen gyakoroltunk, azaz a pártállam belső csoportjainak megértését, az Kína kapcsán manapság úgy gyakorolható, hogy a tartományi pártlapokat kell olvasni, mert a pártközpont hatalmi harcaiban kissé hátrább szoruló elvtársak egy „kevésbé fontos tartományt” — amelynek méretét, lakosságát, vagy a világ GDP-jéhez való hozzájárulását bármely a német és az orosz zóna között fekvő országozskáéval bizvást összemérhetjük - kapnak kárpótlásul, annak pártsajtója pedig ezt tükrözni fogja. De sokkal inkább igaz ez valódi többpártrendszerekben. A kora 1960-as években az amerikai kongresszusban egy „déli demokrata” és egy északnyugati demokrata között a „faji kérdésben” akár nagyobb különbség is lehetett, mint a demokraták és republikánusok közötti átlagos különbség. Fontos jelentéssel rendelkezik a Dél és Észak közötti különbség Olaszországban is, ahol a nagyobb központi forráselosztásban érdekelt déliek akkor sem támogathatták igazán lelkesen az adócsökkentést, ha pártpolitikailag egyébként azt tették volna. Aki rátekint az egyesek szerint nacionalista közhangulatú Románia 2014-es elnökválasztási térképre a német kisebbségi vezető területi értelemben „egyhangú” többséggel megszavazó „kék” Észak Nyugatot és „Tiszta piros” regátot lát, ez azért a képviselők összetételén is meglátszik.

A 2010-es és különösen a 2020-as évek alapvető változása, hogy a helyi könyvtár — óhatatlanul a külföldi *országos* lapokra korlátozódó gyűjtő aktivitása — e tekintetben már nem akadályozza a kutatót, hiszen nemcsak a helyi lapok váltak elektronikusan is elérhetővé, hanem mindenütt megkezdődött a helyi lapok évtizedekre visszamenő digitális feldolgozása is. (Ha bárki azt gondolná, hogy e tekintetben a „történelem” iránti nagyobb állami érdeklődés a különbség indikátora, annak felhívjuk a figyelmét arra, hogy a 19. századi USA nyomtatott anyagait digitalizáló projekt már elkészült, míg pl. német nyelvterületen roppant hiányosságok vannak- miközben a közvélekedés szerint az európai középrétegek tudatában a „történelem” fontosabb, mint a szimbolikus univerzumot inkább a „jog” köré szervező amerikaiban.)

A másik korlát eltűnése azzal áll összefüggésben, hogy az újságok nemcsak képként, nemcsak tárgyszavazva, hanem egyre jobban felismert szöveggént, „Full-textként” is rendelkezésre állnak a neten. (Biszak Sándor tájékoztatása szerint jelenleg az ADT-Plusz olyan irányú fejlesztése folyik, mely a napilap szövegtömeg strukturálását teszi lehetővé. Magyarán: lehet majd tudni, hogy címben van-e egy kifejezés.) Az egyes politikusok nevének múltjának kapcsolatainak internetes kereshetősége azt a korlátot is kiemelte, mely — pl a huszadik század utolsó harmadára emlékezők tanúsága szerint — rendkívüli versenyelőnyhöz juttatta azokat a kollégákat akik – főleg persze személyes kapcsolatok révén — a nagy világlapok archívumaiban működtetett kartotékrendszerekhez hozzáfértek, amihez viszont a Párizsban, Londonban, Frankfurtban élés nagy előnyt jelentett.

A harmadik korlát az volt, hogy a nemzeti parlamenti naplók szövege – ellentétben a több nyelven rendelkezésre álló jogszabályanyaggal – nemzeti nyelvű, viszont összehasonlításra leginkább a kisméretű európai országok parlamentjei vonzóak. Nagyon formalizáltan fogalmazva : ha az egymillió választópolgárra jutó érdemi parlamenti felszólalást, vagy vitamozzanatot tekintenénk megismerendő egységnek, akkor a 2000–2005-ös cseh, a szlovák és a magyar parlament mindenképpen több elemezhető vitát produkál, mint mondjuk a lengyel egyedül, noha az utóbbi döntései több embert érintenek, mint az előző háromé együttvéve. Ezt a természetes korlátot is a 2010-es évek vége kezdte elmozdítani, mivel az ingyenes fordítóprogramok már lehetővé teszik e szövegek tanulmányozását.

Az újságévfolyam mint „kézikönyv”– illetve webes utódjai

A véleményeltérések forrásaként szokták megnevezni a többpártrendszerű rendszerekben *valamennyire mindenképpen plurális* országos sajtót.

Lehetnek olyan időszakok, amikor annak is forrásértéke van, hogy a sajtó mekkora véleménypluralizmust tükröz és a parlament mekkora véleménypluralizmust tükröz. Pl az 1940-es évek elején a magyar hadsereg délvidéki viselt dolgait, vagy a magyar-német kapcsolatok egyenlőtlenségét taglaló parlamenti üléseken a kisgazda , szociáldemokrata vagy liberális hozzászólásokat természetesen nem korlátozta senki, s gyorsírói feljegyzések nyomán be is kerültek a naplóba, onnét pedig – napjainkra a netre, az ellenzéki lapok parlamenti tudósítói viszont nem mindent mertek beleírni a cikkekbe, s ha bele is írták, az óvatosabb szerkesztők kihúztak mondatokat, s ha még ezen a szűrőn is átment a cenzura vagy saját belátásból kihúzott mondatokat, vagy megérdeklődte a külügyminisztériumtól, a vezérkari főnökségtől, a kormányzói kabinetirodától mi mehet át és mi nem. Tehát egyetlen parlamenti ülés jegyzőkönyvének és a róla szóló tudósításnak a különbségeiből — ez digitalizált szöveg esetén könnyebben állapítható meg, mint papíralapú olvasással — érdemi információ szerezhető arról, hogy milyen témák,

történetek, történetelemek, jelzők azok, amik a pártpluralizmus ellenére, a parlamentben való elmondhatóság ellenére a sajtónyilvánosságot már nem tűrik el.

A sajtó azonban nemzetközi ill. idősoros összehasonlíthatósági szempontból mindenképpen gyengébb forrás, mint pl. a parlamenti beszédek, hiszen a napi és hetilapok politikai orientációját néha csak tartalomelemzés állapíthatja meg, néha a lap vállalja az orientációt „baloldalinak,” „nemzetinek” vagy „polgárinak” nevezve magát, a szerkesztő számára fenntartva annak megállapítását, hogy mi „baloldali”, mi „nemzeti.” Ritkább esetekben a lap nyíltan kötődik valamelyik politikai párthoz, még ritkábban annak egyik frakciójához, tehát még cenzurális hatás nélkül is van szerkesztői döntés.

Minden sajtótermékre igaz azonban, hogy a szerkesztőségekbe beérkező cikkek, vagy közlésre szánt olvasói levelek, tehát a *potenciális írások száma meghaladja a leközlött cikkekét*. Nem lehetünk biztosak tehát abban, hogy a véleménycsoporthoz tartozók valamennyi véleményére rábukkanhatunk az adott lapban. Ezt a problémát a történeti kutatások esetében a 2010-es években sem szüntette meg semmi, bár kétségtelen, hogy a személynevekre, intézménynevekre, szókapcsolatokra keresés lehetősége a világháló olyan pontjairól is a kutató szeme elé hoz forrásokat, ahova egyébként nem jutott volna el. (Formalizálva: valamely tárggyal, keresőszóval kapcsolatban a valamire reprezentatívnak minősített újságból kiemelve a 100 leggyakrabban előforduló nevet, már van valamiféle eszközünk arra, hogy a keresőszóval létrehozott világháló találatok millióit minek alapján szűkítsük. De ez már majdnem maga a big-data kutatás.)

Rejtett kézikönyv tartalom viszont, hogy az évek múlásával változik a szavak, kifejezések, személynevek, intézménynevek, településnevek aránya. Nyilvánvalóan vannak szókapcsolatok, pl a „klerikális reakció”, vagy tulajdonnevek „Mindszenty” melyeknek felbukkanási gyakorisága mögött nyilvánvalóan az a szerkesztői tudás áll, hogy az olvasó is, a lapengedélyező is, a lapkiadó is, a szerkesztő munkáját minősítő politikai munkatárs is és — nem utolsósorban — a lapnak megnyerendő szerzők is jelentőséget tulajdonítanak a szavak felbukkanási gyakoriságának, kontextusának. Az elemzések azonban azt mutatják, hogy számítástechnika előtti világban e kiemelt szavakon kívül legfeljebb a cikkek címének a szavai tudatos üzenetek, a több százezer szavas korpuszokban néhány száz előre ki nem található szó vagy szókapcsolat arányának ezrelékes nagyságrendű elmozdulása fejtette ki az objektív hatást az olvasók százezreire. [GYÖRFFY 2017, SZÉCHENYI 2004, LENGYEL 2011, RUFF 2010, GYÁNI 2006]

Az eleve internetre készülő újságok, archívumát is lehet kézikönyvnek tekinteni, bár kétségtelen sosem léteztek papíralapon. Viszont valószínűleg csak a szerkesztőkkel, újságírókkal készített mélyinterjúk alapján dönthetnénk el, mi a rejtett tartalom és mi a szándékolt üzenet. Azért az valószínűsíthető hogy amennyiben a, b és c közszereplő x-1 és x időpont között ismertségét folyamatosan

javítva elérte pl. a holtversenyben 3.. legismertebb közszereplő pozícióját , akkor az, hogy x mínusz 1 év időpontban a lap a-t a 10.b-t viszont csak 20. pozíciónak megfelelő gyakorisággal említette, valószínűsíti, hogy a lap rejtett üzenete volt, hogy a ből inkább „nagy ember” lesz, mint b ből. Ha ez több különböző portálnál alakul ugyanígy, akkor az újságírók témaválasztása, *a hírpiac láthatatlan keze hozta létre a különbséget, ha csak egyiknél, akkor vélhetőleg szerkesztői megrendelés, azaz nagyon is látható kéz.* Utóbbi esetben beszélhetünk ugyan az olvasó elől elrejtett tartalmi üzenetről de naivitás lenne azt feltételezni, hogy a szerkesztők keresésekkel, vagy akár erre rendszeresített makroval, programocskával ne ellenőriznék folyamatosan, mennyire alakul kívánatos irányba ez az üzenet, s ne tennének lépéseket, ha elégtelennek érzik az eredményt.

Már nem is emlékeztet kézikönyvre, de aki a 21. század elejét tanulmányozza történészként, annak fontos a „web2”. Twitter és facebook bejegyzések, intézményes és magán site-ok tömege, melynek révén gyakorlatilag minden vélemény nyilvánosságra kerül. A nemzetközi és idősoros összehasonlíthatóság korlátja e vonatkozásban módszertani természetű: azaz az igen különböző rendszerekben, különböző időszakokban hogyan határozható meg *elvileg* a véleményformálók és véleménynyilvánítók *azon köre*, akiket az egyes országokban ill. egyes időszakokban figyelembe kell vennünk ahhoz, hogy az összehasonlítás ne vezessen torzulásokhoz, ne vonódjanak be az elemzésbe egyes országok és egyes időszakok teljesen partikuláris (létükkel tehát semmit sem bizonyító) véleményei. A like-ok és más emotikonok, követések, ismeretségi hálók számszerűsítésén, illetve ezen számok változásán keresztül komolyabb számítástechnikai felkészültség nélkül is csoportosíthatóak az emberek és a vélemények is. az idősorossághoz azonban jelen állás szerint úgy tűnik arra van szükség, hogy az ember valós időben rögzítse az adatokat, amelyeket 10-20 év múlva elemez, ez nem igazán történésznek való helyzet... [Koltay 2010, Koltay 2011]

A források hierarchiája...

Az álláspontkülönbségek elemzésének – torzulásmentesség szempontjából — tehát van egy hierarchiája:

- legarányosabban tükröző, s legaggálymentesebb álláspont-összehasonlítást lehetővé tevő forrás az európai parlament felszólalástömege — mármint azokra az országokra és időszakokra nézve, amikor ez releváns
- ezt követi a nemzeti parlamentek éves költségvetési viták során szükségképpen elhangzó szakpolitikai véleménytömege
- ezt követi az időben és térben változó házsabályok szerint erősen befolyásolt egyéb parlamenti műfajok során elhangzó beszédek tömege — e harmadik helyről még az oktatási törvények, vagy szociális törvények

vitáit sem vesszük ki, hiszen önálló oktatási vagy szociális törvényjavaslat beterjesztésének elkerülése, salátatörvények használata kedvelt eszköze az autokratikus kormányoknak

- ezt követi az a sajtómegnyilatkozás-tömeg, melyet az újságok példányszáma szerint minősíthetünk valamennyire reprezentatívnak
- ezt követi a web1-hez kötődő adattömeg, melyet az elektronikus újságok kattintásszáma szerint minősíthetünk valamennyire reprezentatívnak
- ezt követi a — létrehozók pozíciója, illetve like-ok száma szerint standardizálhatóan fontos web 2 nyilvánosság

Az álláspontok jó megismerhetőségével ellentétben a „reálfolyamatokat” tükröző szöveges adatokkal némiképp más a helyzet.

A „jogszabályi valóságtól” eltérő, „devianciákat”, (semlegesebben fogalmazva a valóság sokszínűségét) dokumentáló lokális folyamatok vagy újságírók tucatjainak-százainak munkája, vagy helyi oktatáspolitikai-oktatásügyi aktorok, egészségpolitikai-egészségügyi aktorok aktív kommunikációja, vagy szisztematikus független társadalomtudományi tevékenység, vagy szisztematikus állami adatgyűjtés révén válnak megismerhetővé. A nagyszámú „történet” feltárása a cikkekből ill a stakeholderek kommunikációjából maga is egy nagyobb kutatás, ami erőforrásokat igényel, vagy fizetett munkaerőt, vagy nagyobb időráfordítást a kutató részéről. Az oktatásügyi, egészségügyi esettanulmányok természetesen összehasonlíthatók Európa bármely országa vonatkozásában. Az „esettanulmányok” összehasonlíthatóságával szemben ugyanakkor módszertani aggályok fogalmazhatóak meg, hiszen semmiképpen nem standardizáltak az újságírói megfigyelés körülményei, arról pedig csak nehezen empirizálható előfetevéseink lehetnek, hogy az egyes országokban és egyes időszakokban minek van „hírértéke”, azaz mikor, hol milyen, mekkora „deviancia” vagy milyen „jógyakorlat” szolgálhat motivációul az újságcikkek megírásához. Ennek a módszertani aggálynak a kiküszöbölése csak szisztematikus terepkutatással lehetséges, melynek össz-európai költségvetése, illetve a szükséges engedélyek megszerzése akkora kihívás, melyre csak egy igen komoly nemzetközi kutatói konzorcium vállalkozhat.

A komparatív kutatások kapcsán fentebb felsorolt kihívások és lehetőségek egy része automatikusan igaz a nemzeti történeti kutatásokra is, hiszen pl a digitalizációból eredő összes jelenség itt is kiszélesítette a forrásbázist, lehetővé tette egészen távoli asszociációk bekapcsolását.

Közép Európa

A Közép-Európával kapcsolatos történeti kutatásokkal kapcsolatban külön is megjegyzendő, hogy az oktatásügyi, közigazgatástörténeti stb mainstream források a 1844 előtt alapvetően latin, azt követően magyar nyelvűek, így az oktatástörténet közigazgatástörténet valóságának legalább három olyan „szintje” van, melyben a mesterséges intelligenciával működő fordítógépek alapvetően átalakítják a kutató 2020 utáni helyzetét a korábbiakhoz képest.

Az egyik ilyen „szint” a történeti-helyi szint. 1918 előtt az oktatáspolitiká, közigazgatáspolitiká stakeholderei mielőtt az országos oktatáspolitikába, jogpolitikába — tehát a Parlamentbe vagy az Országos Közoktatási Tanácsba, közigazgatási bíróságra stb. — vitték volna álláspontjukat — ahol tehát magyarul olvasó számára is elégséges mértékben tanulmányozható —, helyi nyilvánosságban vitatták meg. A helyi nyilvánosság pedig elsősorban a helyi, részben nem magyar nyelvű lapokat jelenti, részben különféle egyházi testületek üléseit. Az egyházi testületekben pedig — noha voltak nyelvpolitikai harcok az egyes egyházak *országos* nyelvét illetően — végképp az adott közösség nyelvét használták. Az osztrák magyar monarchiában német, magyar, cseh, szlovák, szerb, horvát, szlovén, olasz, bosnyák, román, rutén, jiddis, lengyel nyelvű forrásokkal kell szembenéznie annak, aki pl az államérvényes és zugiskolákkal kapcsolatos írott örökséggel akar megismerkedni.

A másik szint az 1918 illetve 1945 után keletkezett közigazgatási források tömege. A magyar köztudatban szereplő 1918 utáni ill 1945 utáni konfliktusoknak és együttműködési kísérleteknek, (melyeket általában csak a magyar kisebbségiek újságcikkeiből ismerünk) megvan a teljes „tükröződése” román, szerb, szlovák stb nyelvű közigazgatási iratokban, amelyek részben áttelepülésekhez, részben magyar nemzetiségi oktatási intézményszervezésekhez, részben „oktatáspótló” nemzetiségi identitás szervezésekhez, részben pedig magyar nemzetiségű, magyar háttérű, magyar származású személyek teljes vagy részleges nyelváltásához is vezettek román, szlovák, szerb stb társadalomba való integrációjuk ill asszimilációjuk során.

A helyi iratok természetesen, a magyar nemzetiséggel való viszony pertraktálásán túl még arról — a magyar forrásokban szinte egyáltalán nem tükrözött — társadalmi jelenségről is szólnak, hogy a 1918 után más országokhoz került területek elitpozícióira nemcsak a többségi nyelvhez tartozó helybéliek hanem a Cseh és Morvaországból, az 1918 előtti román királyságból, illetve szerb királyságból érkezettek is pályáztak, s 1945 után is elit-önjelöltként bukkantak fel dél-erdélyiek, közülük is különösen azok, akik román hazafiságukat azzal bizonyították, hogy 1940 ben a Partiumba, Észak Erdélybe érkező magyar csapatok elől a határ román oldalára húzódtak, illetve azok, akik a Bácskában nem túlzottan

jelentős, de a szerb hegyvidéken nemzetközi mértékben is számontartott partizánmozgalomban szereztek ahhoz elegendő kapcsolati tőkét, hogy az újjászerveződő Jugoszláviában pozíciókra számíthassanak, vagy azok a dél-szlovákok, akiknek 1938 és 1939 után a szlovák bábállam, még mindig jobban tetszett, mint Horthy Miklós Magyarországa, s ezért áttelepültek oda. A szlovák, magyar, román digitalizáció nem hala párhuzamosan, ha párhuzamosan haladna egészen konkértan megvizsgálható lenne, hogy egy ügytípus — „pl. boltnyitási engedélyek intézése” hogyan folyik a helyi hivatalokban az utolsó olyan héten, amikor Csehszlovákiához tartoznak a Máramaros megyei települések, s hogyan az első olyan héten, amikor már Magyarországhoz, s hogy a 50. hétre hogyan változik ez. Egységes rendszerbe látszik viszont szerveződni a romániai román és a romániai magyar helyi sajtó az adtplus nál.

A harmadik szint az utódállamok történetírása (illetve a német szakirodalom), amely nagyon sok információt illetve figyelemreméltó narratívát tartalmaz az 1918 előtti és utáni oktatástörténeti, közigazgatástörténeti viszonyokról egyaránt, s a fordítógépek természetesen ezekkel is boldogulnak.

A pályakezdő történész nemzedék számára — válassza kutatási területét akár a szélesebb nemzetközi összehasonlítást, akár közép-európai történeket rendkívüli lehetőségeket nyitott a politikai nyilvánosság megváltozása, a források digitalizálása és internetesedése ill. a fordítógépek megjelenése. Legalább 2030-ig kell azonban várunk, hogy egy akkori historiográfiai elemzés kiértékelhesse, élt-e ezzel a lehetőséggel. [TÓTH 2014, SZAB, 2013, JANEK 2021, TÖMÖSKÖZI 2020, SIMON 2013, MISKOLCZY 2014]

A színművészeti lexikon projekt

A színművészeti lexikon feldolgozása az első konkrét példa a digitalizálásra. úgy ennél mint a következő példáknál kissé didaktikusabb lesz a leírás.

Az 1929-31-es magyar színművészeti lexikont feldolgozó projekt kapcsolódott mind a reputációs eliteket vizsgáló történetiszociológiai, mind általánosabb érdekű digitalizációs projektjeinkhez, melyeket a Wesley Egyház- és Vallásszociológiai Kutatóintézetben folytatunk.

Reputációs elitvizsgálataink során — melyeket Karády Viktorral közösen építettünk fel az elmúlt bő két évtizedben alapvető meggyőződésünké vált, hogy az elitek meghatározásának legjobb történeti módszere a lexikonok vizsgálata. A lexikoncímszavak kiválasztása (általánosságban, kevés lexikon kivételével) sokkal alaposabban tükrözi egy-egy kor (társadalmi, vagy szakmai) közvéleményében „fontosnak tartott” személyek (az un. reputációs elit) névsorát, mint a bizonyos pozíciókat elérték listájából megalkotható elitlista. A pozíciók betöltése, kitüntetések elnyerése elé a magyar történelem szinte bármely korszakában számos esetben

felekezeti, nemzetiségi, politikai akadályok tornyosultak. A magyar társadalom valamennyi elitesoportját - így a színművészeti elitet is - jellemzi, hogy felekezeti, nemzetiségi és politikai összetétele - nemcsak a hazai átlagnak nem felelt meg, hanem annak a képnek sem, amit a kiténtetések és pozíciókat adományozók megfelelőnek láttak volna. A „kisebbségi” és „ellenzéki” pozícióban lévő — de saját szakmai köreikben elismert, vagy a történelmi mérlegen megmaradó — személyek megfelelő súlyát általános és szaklexikonok segítségével mutathatjuk be. Ezek egyike az Országos Színészegyesület által 1929–31-ben kiadott Schöpflin Aladár szerkesztette négykötetes Magyar Színművészeti Lexikon.

Hasonlóképpen igaz ez nemcsak az elitekre, a személyi kérdésekre, hanem a „dolgokra” is, a színművészet-szignifikáns helyekre, intézményekre, művekre, fogalmakra is.

Digitalizációs projektjeink két tényből indultak ki. Az egyik, hogy napjaink lexikonhasználati gyakorlata az internetes információforrások javára szinte teljesen háttérbe szorította a papíralapú lexikonok használatát. A régi lexikonokat háttérbe szorító — s már interneten található — lexikonok, adattárak néha egyszerűen gyengébbek, mint évtizedekkel ezelőtti tematikus társaik. Gyakran persze jobbak és korszerűbbek, ahogyan jelen esetben is igaz, hogy a 90-es évek színháztudományi lexikonja korszerűbb, mint a projektünkbe bevont kézikönyv. De még ezekben az esetekben is igaz, hogy a minden kézikönyvnél jelentkező terjedelmi korlátok szükségszerűen csökkentik a régebbi időkről szóló címszavak arányát, illetve a rájuk eső terjedelmet. A digitalizálás tehát annak a munkának hasznosítása — potenciális köztudatban tartása — amit egy-egy értelmiségi csoport évtizedekkel ezelőtt befektetett.

A másik tény, hogy a régebbi papíralapú kézikönyvek — a leggondosabb tárgymutatókészítési gyakorlat mellett is — számtalan olyan információt hordoznak, mely nem tárul fel a kézikönyv használója előtt, hiszen — noha releváns egy személy vagy jelenség elemzéséhez — nem abban a szócikkben, nem abban a fejezetben van, ahol a felhasználó keresi. Az egyszerű — szkenneléses, karakterfelismeréses digitalizáció is segít némiképpen ebben, hiszen karaktorsorokra minden dokumentumban lehet keresni, de Boolean operátorokhoz, azaz kombinatív keresési lehetőségekhez szoktatott korunkban ennél differenciáltabb igények mutatkoznak. E keresési lehetőségek olyan összefüggéseket is feltárhatnak, melyeket a kézikönyvek egykori szerkesztői sem láthattak.

Ezért az 1929-31-es színművészeti lexikont egy minden korábbinál mélyebb analízisnek vetettük alá, hogy kialakítsunk egy jól kereshető adatbázist. Ennek fázisai következnek az alábbiakban. Ezek nem technikai részletek, hanem a keresési lehetőségeket, s a korlátokat is belőlük értheti meg az olvasó.

1 A projekt első fázisában a négykötetes lexikon beszkenelése és a karakterfelismertetés történt meg. A pdf file-ok — egy oldal egy file — azt tették

lehetővé, hogy az azonos oldalon előforduló karaktersorozatokra, szavakra kombinatív kereséssel rá lehessen találni a <https://mek.oszk.hu/08700/08756/> oldalról elindulva E keresés kiszolgáltató volt a téves karakterfelismerésnek, s annak, hogy az abc véletlene szerint egymástól teljesen független címszavakban előforduló információk között hamis összefüggést sugallt, míg a több oldalra széttagolódó hosszabb szócikkek belső összefüggéseit rejtve is hagyhatta. Érvényesült viszont az a fontos — előzetesen megállapított — elvünk, hogy a lexikonban található szavakra akkor is rá lehessen találni, ha valaki ismeri a lexikont, s akkor is, ha minden előzetes ismeret nélkül egyszerűen beír egy nevet az internetes keresőbe.

2. A következő fázisban azzal könnyítettük a felhasználó dolgát, hogy elkészítettük a címszavak listáját, a születési évszámok segítségével lehetővé tettük az azonos nevű személyek elkülönítését. A címszavakat kategóriákra osztottuk, férfinevekre, női nevekre, település nevekre, idézetekre, műcímekre, színháznevekre, intézményekre és színházi fogalmakra. Ez már 2010-ben használhatóvá vált a MEK-ben.

3. Ebből keletkezett a segédadatbázis, melyben minden egyes címszó külön sort kapott, sőt a hosszabb címszavak, melyek a könyvben több oldalra terjedtek ki, annyi sort ahány oldalra kiterjedt. A segédadatbázis 12327 sorból áll, ebből 11972 sor új címszó kezdetét jelzi. A címszavak végleges listája — 2017 elejére kialakult formájában, azaz az azonos címszavak elkülönítését első szavaik segítségével lehetővé téve abc-ben mutatóként kereshető, de természetesen szövegrészeket formákában is.

4. A következő fázis volt a munka leginkább időigényes része. B. Kádár Zsuzsanna végigolvasta a teljes szöveget, szegmentálta a címszavakat, kijavította a karakterfelismerési hibákat és a szövegben megjelölte az előzetesen meghatározott tárgyszó-kategóriákat: a településeket, a személyneveket, a nemzetiségeket, a nyelveket, a szerepeket, a műveket, az intézményeket. Az első kötettel már 2013-ban végzett, az utolsóval csak 2017-ben. Az első kötet már 2013-tól kereshetővé vált a MEK-ben, s 2017. 04. 18-ig 7070 érdeklődő kereste fel.

5. 2017 márciusában a szövegbankból egy olyan, több mint egymillió soros full text SPSS adatbázist alkottunk, melyben egy-egy szó jelentett egy-egy esetet. Az adatbázis természetesen tartalmazta azt a 20000-es és 30000-es számkörben alkotott fiktív számot, amit az egyes címszavak elkülönítésére használtunk, valamint az egyes sorok szócikken belüli pozícióját, egy 1–4 jegyű szám formájában. Ebben az adatbázisban került sor az előzetesen elhelyezett kódok ellenőrzésére, javítására. A kódokkal felszerelt szavak illetve kifejezések önálló változóba kerültek, s a szövegből olyan változat is keletkezett amelyet nem terheltek kódok.

6. A kódokkal felszerelt szavakból, kifejezésekből képzett változókon kívül kialakítottuk a Budapesttel kapcsolatos változót, mely mindazokat a településneveket egyben és külön is kezeli, melyek a jelenlegi Budapesthez

tartoznak. A településváltozóból az 1910-es állapotnak megfelelő megyeváltozót hoztunk létre. A Kárpát-medencén kívüli településnevekhez — vállalt következetlenséggel — hozzárendeltük azon országok nevét, melyek területén 2000-ben fekszenek. (Ha tehát valaki nem tudja, hogy egy konkrét megyében, hol történt olyan színházreleváns esemény amelyre rákereshet, a megyenév megadásával ajánlatokat kaphat: mire keressen.)

7. Kialakítottuk az évszámváltozót, úgyelve arra, hogy a műcímeken belüli évszámok ebbe ne kerüljenek bele. Az évszámváltozóból évtizedváltozót alakítottunk ki és minden egyes tárgyszó mellé odarendeltük azt az évtizedet, melyre feltehetően vonatkozik, azaz ami a szócikkben a szó előtt utoljára szerepelt. (Természetesen a halál évét, mely a szócikk elején szerepelt ebből kizártuk.) Ha a következő évszám illetve a halál éve alapján nem kellett ezt kizárnunk, akkor a „valószínű évtized” mellé még „bizonytalan évtizedeket” (maximum kettőt) is rendeltünk. Minthogy az évtizedváltozó kialakítása gépi úton történt többféle hiba is lehet benne: elképzelhető, hogy az adott szerep, vagy mű már másik évtizedhez kötődik, de a lexikonszócikk szerzője ezt nem évszámmal, hanem stílárius eszközökkel jelzi. Néhány esetben az „1001.” és „1931.” közé eső jelzések nem évszámot, hanem például honoráriumot jeleznek, de ez csak szövegösszefüggésből világlik ki. Kivétel nélkül kimaradnak az évszámváltozóból az ókori és koraközépkori események, hiszen az 1–3 jegyű számok és az évszámok elválasztására gépi úton nem találtunk módot. A születési év változó azonban ezekben az esetekben eligazít.

8. Ezt követően a segédadatbázis segítségével egy „elosztó adatbázist” hoztunk létre, melyben az azonos szóval kezdődő címszavak közös kódot kaptak. Ebből az adatbázisból leválogatás, szöveges mentés, majd htm-mé alakítás segítségével mutató file-okat hoztunk létre. Tehát például az „abonyi” mutató file hatjegyű kódot kapott s az összes olyan címszót tartalmazta, mely az abonyi szóval kezdődött.

9. A tárgyszóként megjelölt szavakból listákat hoztunk létre, a listákba bekerült az évtizedváltozó és természetesen a címszó is. A következő listák állnak rendelkezésre: megye, Budapest, ország, személynév, szerep, műcím, intézmény, nemzetiség, nyelv, születési év, időpont. A megyei és ország listák a településnevek abc-ben keresését is támogatják.

10. Ezt követően az elosztó-adatbázisból képzett keresőszavak segítségével a fő adatbázis minden egyes tárgyszóként megjelölt szava mellé odahelyeztük a megfelelő hatjegyű számot, így minden tárgyszóról látni lehet: van-e az adott szóval kezdődő címszó.

11. Leválogattuk az egyes szócikkekhez tartozó adatbázisdarabokat, s azokból három-három dat-mentést hoztunk létre.

12. Az első mentésbe azok a változók kerültek, melyek az eredetileg megjelölt szavak mellé olyan „fórróponokat” iktattak be, melyek az egyes listák felé

mutattak, illetve információt kínáltak fel. A listák a tárgyszavakból készültek, lehetővé téve tehát a felhasználó számára annak elemzését, hogy meghatározott szóhoz (meghatározott feltételezett évkörben) milyen más címszavakban tartozik hasonló találat. (lásd 9. pont) Az információk pedig az elosztó-file-ok felé vezetnek, azt mutatják, hogy olyan kezdettel milyen címszavak találhatóak még a lexikonban. (Lásd 8. pont) Ennél biztosabb módszert nem találtunk ahhoz, hogy az olvasót el is vezessük az őt érdeklő címszavakhoz, de egyben figyelmeztessük is: mérlegelni kell, hogy valóban arról a személyről, vagy műről van-e szó az adott szócikkben, vagy csak szóazonosság esete forog fenn.

13. A második mentésbe a főszöveg mellett a kódokkal felszerelt szavak olyan formában kerültek, hogy a Google keresőmotorja megtalálja őket, pl. yszemelynevy aradi vagy yintezmenynevy aradi formában. Ez az emberi olvasás számára haszontalan, apróbetűs anyag.

14. A harmadik mentésbe viszont csak az eredeti szöveg került. Ennek funkciója az, hogy összefüggő szabadszavas kereséseket a kódok és linkek zavaró hatása nélkül lehetővé tegyen, illetve, hogy aki a szöveget vagy annak egy részét kódok vagy linkek nélkül szeretné magának kimásolni, ezt könnyedén megtehesse.

15. A három-három szöveget egyetlen „.htm” file-ba kapcsoltuk össze, a szövegbe beírt „.htm”-eket valódi linkekké változtattuk.

16. Index file-okat hoztunk létre melyek az elosztó-file-okra, címszó-file-okra és a lista-file-okra mutattak. Így a Google keresőmotorja az index file-ok segítségével éri el az egyes file-okat.

17. Az adatbázist előre meghatározott struktúrában helyeztük el a MEK-ben. Az index file-ok kerültek legközelebb a gyökérhez, az elosztó-file-ok (lásd 8. pont) az e alkönyvtárba, a lista-file-ok (lásd 9. pont) az a alkönyvtárba, míg a szócikk file-ok (15. pont) a w alkönyvtárba kerültek.

Az adatbázist használók száma 2023 január 31-n 11301-en állt

Két általános célú adatbázis

Az egyik adatbázis alapja a Magyar Életrajzi Lexikon, melyet ez alkalommal nem elitadattárnak, hanem eseményadattárnak használunk. Az spss-be behívott teljes életrajzi lexikon — mely a Magyar Elektronikus Könyvtárból tölthető le ingyenesen — teljes szövegének minden egyes szava új esetet képez. A szöveg oszlopot átmásoltuk két szomszédos oszlopba; négy, illetve öt karakter hosszúságúra csonkoltuk őket, majd numerikussá változtattuk. Ahol az első oszlop (EVSZAM) értéke 1990-nél (a lexikon készülésének événél) nagyobb, vagy nincs, ott az értékét nullává változtatjuk, ha a második oszlopban (OTJEGYU_NUM) van egy 9999-nél nagyobb érték, az első oszlop hasonlóképpen nulla lesz, ugyanis ez az érték azt mutatja, hogy az első oszlopban szereplő évszámnak látszó szám valójában egy

nagyobb szám első négy karaktere. A születés és a halál időpontját egy-egy oszlopba kivisszük, ez könnyű feladat, hiszen a személyi címszó elején, a félkövérral szedett mezőben, minden esetben csak egyetlen nyitózároljel van, s a nyitózároljel után következő első négyjegyű szám mindig a születés éve, a következő évszám (ami egyébként a záró-zároljelet megelőző évszám is egyben) a halál éve. A félkövér mezőt megterheli néha még egy évszám, a névváltoztatásé, ezt is kivisszük.

A félkövérral szedett mező soraiban a születési évszám kivételével minden számot nullára változtattunk.

Ezt követően a magyarországi településnevek teljes listájából és az európai városok listájából alkotott parancssorral egy TELEPULES nevű változóba lekódoltuk a teljes adattár minden szavát, azokat az eseteket, melyek nem kaptak numerikus kódot nullákkal töltve fel.

Majd az EVSZAM-ban és a TELEPULES-ben minden nullát egyenlővé tettünk a felette lévő értékkel, ezzel a szövegbank minden egyes szava mellé egy rá vélhetőleg jellemző évszám került, s egy rá vélhetőleg jellemző településnév is került.

Ezt követően létrehoztunk egy-egy olyan változót, mely két illetve három szót vont össze, kidobva azokat a kifejezéseket ahol a pont, vessző, pontosvessző volt a szavak között.

Így keletkezett egy olyan adattár, mely *virtuálisan azaz a konkrét történeti igazság értéke nélkül* intézmények, jelenségek tömegéről tájékoztatott, mármint arról, hogy *valaki* az egyesület élére került, valaki publikált a folyóiratban stb., s minden egyes jelenség mellé egy tájékoztató évszám és egy tájékoztató helynév is került.

A jelenségek átlagos (mean) időpontja, középidőpontja (medián), első és utolsó időpontja egyaránt fontos adat (A medián és a mean viszonya tájékoztat arról, hogy a jelenség a történeti időben előrehaladva sűrűbben vagy ritkábban fordul elő). Filológiai értelemben nem állíthatjuk, hogy az egyesület abban az évben jött létre, amely legkisebb dátum így mellette állt, de az mégiscsak jellemző, hogy sokezer elittag egyikével sem merült fel korábban az egyesülettel kapcsolatos életrajzi mozzanat.

A másik segédadatbázist a szintén Arcanum által managed „szaktars” adattárban megtalálható Benda Kálmán féle négykötetes történeti kronológiából készítettük, mely a kezdetektől 1970-ig tekintette át a magyar történelmet. A letöltött és karakterfelismerővel feldolgozott adattárat először az egy bekezdés egy eset logikával nyitottuk meg spss-ben, azt mondva, hogy nincsenek delimiter, majd egy másikban ugyanígy azt mondva, hogy az üres karakter delimiter. A második filera a első három oszlopot megőrizve, a következőket törölve rákapcsoltuk az első filet, az egykori első oszlop elé beszúrva egy oszlopot melynek adattartalma az volt, hogy xxevszamxx, a második elé beszúrva egy oszlopot

melynek adattartalma az volt, hogy xxhónapxx, a harmadik elé beszúrva azt szót, hogy xxpontosdatumxx. Az így — txt-ben — elmentett adattárat spss-ben ismét megnyitottuk, egy szó – egy eset logikával. Ezt követően külön változóba kihoztuk az évszámra és a hónapra és a napra vonatkozó adatokat, a maradék helyett nullákkal töltve fel, majd mindent ami nullás értékű a fölötte lévő szóval tettünk egyenlő értékűvé. A hónapneveket számokra cserélve és az oszlopokat összevonva, minden egyes szó elé egy nyolcjegyű szám került, mely a szóra jellemző dátumot tartalmazta eeehhnn formában.

Ebből az adattárból is készült egy olyan aggregáció, mely a szavak, szópárok, köztük köznevek, tulajdonnevek időbeni elhelyezésére alkalmas volt.

E két adattárból készült metaadatbázis egyrészt arra való rejtett kézikönyvtartalom, hogy személyek, személycsoportok, események, helyszínek közötti rejtett összefüggések feltárhatóak legyenek — bár ez is igen érdekes eredményeket hoz. Kísérletképpen ellenőrizve: másféle az 1890-1895 ös időszak másféle kronológiája bontakozik ki, a személyi címszavakból, hogyha közigazgatásreleváns, mintha irodalomreleváns, vagy iparreleváns személyekből alkotjuk meg azt.

Másrészt viszont ezek a metaadatbázisok olyan indikátorforrásokká válnak, melyek *minden* szöveg minden tulajdonnév jellegű szavának minősítésére alkalmasak, az újabb és újabb szövegekkel megerősítve a metaadatbázist, gyakorlatilag végtelen iterációban kapcsolva össze a legkülönbözőbb történeti és kortárs forrásokat.

A fenti írásban azt kívántuk bemutatni, hogy a digitalizáció révén út nyílt arra, hogy a történettudomány bölcsészettudományból mindinkább társadalomtudományá váljék.

Irodalom

BARTOS-ELEKES Imre: A posztmodern játéka a történetírással: a történettudomány helyzete a globalizmus korszakában. in: *Korunk*, 2020 / 7. szám

BENCZÜR András – BÍRÓ István – CSALOGÁNY Károly – RÁCZ Balázs – SÁRLOS Tamás: PageRank és azon túl: hiperhivatkozások szerepe a keresésben. In: *Magyar Tudomány* 2006 / 11. sz

ERŐS Vilmos: A historiográfia fogalma. In: *Századok* – 2017 / 1. szám

ERŐS Vilmos: Az amerikai történetírás. In: *Valóság*, 2010/ 2. szám

- FEITL István: *Vita a parlamentarizmusról 1944 és 1949 között* (PALASIK Mária: Parlamentarizmustól a diktatúráig (1944–1949). Országgyűlés Hivatala, Budapest, 2017). In: *Múltunk* 2018 / 2. szám
- GYÁNI Gábor: Sajtótörténet a társadalomtörténész szempontjából. In: *Médiakutató* 2006 / 1. szám
- GYÓRFFY Gábor: Újabb magyar sajtótörténet. In: *Korunk* 2017 / 11. szám
- GYULAI Attila: Partikularitás, univerzalitás és alternativitás a parlamentáris kormányzatban. In: *Politikatudományi Szemle*, 2013 / 2. szám,
- JANEK István: Popély Gyula: Felvidék 1929–1939. A második évtized csehszlovák uralom alatt. In: *Századok* 2021 / 1. szám
- KEREKES Pál: Változó könyvészeti előképek a digitalizáció szemszögéből: Gutenberg-átértelmezések, avantgard könyvmegújító törekvések e-könyves nézőpontból. In: *Iskolakultúra*, 2015 / 7–8. szám ,
- KOLTAY Tibor: A Wikipédia, a Web2.0 és az amatőrök. In: *Magyar Tudomány* –2011 / 11. sz.
- KOLTAY Tibor: Web 2.0. In: *Iskolakultúra*, 2010 / 7-8. szám
- KÓTA Péter: Digitális forráskiadás: gondolatok alapvető kérdésekről. In: *Turul* 2015 / 4. füzet,
- LENGYEL András: A magyar sajtótörténet válogatott bibliográfiája 1705–1944. I. A-J. Összeáll. Lakatos Éva. Bp. 2010. OSzK, 372 l.. In: *Magyar Könyvszemle*, 2011 / 1. szám
- MIHÁLYI Dorottya: *Történet, történelem, történetírás* (ERŐS Vilmos: Modern historiográfia: Az újkori történetírás egy története. Ráció Kiadó, Budapest, 2015) In: *Aetas*, 2016 / 4. szám,
- MISKOLCZY Ambrus: A Vasgárda az 1937 évi romániai választásokon. In: *Levéltári Közlemények*, (2014)
- NAGY Péter Tibor: Kvantitatív és kvalitatív felsőoktatás- és tudománytörténetírás. In: *Iskolakultura* 2021.10 sz
- NOIRIEL, Gérard: A történetírás „válsága” (Lakatos Ágnes). In: *Múltunk* 2002/2. szám
- Oktatás – rejtett kincs*. A Jacques Delors vezette Nemzetközi Bizottság jelentése. Budapest, 1997, Osiris Kiadó Magyar UNESCO Bizottság.
- PÜSKI Levente: A parlamenti politizálás nyilvánossága a Horthy-korszakban in: *Múltunk* 2011 / 3. szám
- RÁCZ György: Történetírás és levéltár in: *Történeti Tanulmányok* 18. (2010)

- REISZ T. Csaba: Történelem egy kattintásra? Klasszikus forráskiadás és/vagy levéltári adatbázisok. In: *Századok*, 2018 3. 685–694
- RESS Imre: Digitális forráskiadás vagy elektronikus mutatózás? A kora újkortól az államszocializmus végéig keletkezett tömeges forrásanyag közzétételének magyar gyakorlata. In: *Levéltári Közlemények*, 2006 1. 13–31.
- ROMSICS Ignác: A magyar történetírás intézményrendszere és fórumai. In: *Aetas*, 2010 / 4. szám
- RUFF Borbála: Sajtótörténet másképp • Hírharang, vezércikk, szenzációs riport *Élet és Irodalom*, 2010-06-25 / 25. szám (20. oldal)
- SIEGRIST, Hannes: A transznacionális történetírás mint kihívás a tudományos történetírás számára (Fordította: Cora Zoltán) In: *Aetas*, 2009 2. szám
- SIMON Attila: A magyar szociáldemokrácia útkeresése a húszas évek Csehszlovákiájában. In: *Múltunk* 2013 / 1. szám
- SIPOS Péter: A XX. századi forráskiadás problémái és ajánlás a forráskiadás szabályzatára Fons (*Forráskutatás és Történeti Segédtudományok*) VII. évfolyam 2000.1. szám
- SZABÓ Lilla: A felvidéki magyarság képzőművészete 1920–1937. In: *Art Limes* 2013/2. szám
- SZÉCHENYI Ágnes: A huszadik század hiányzó magyar sajtótörténete – adósságlista és javaslat. In: *Magyar Tudomány* 2004 / 10. sz
- SZENTE Zoltán: Kormányzás a dualizmus korában. A XIX. századi európai parlamentarizmus és Magyarország kormányformája a kiegyezés után 1867–1918. Budapest, Atlantisz, 2011
- SZENTPÉTERI-NAGY Richárd: Az ország értéke? *Élet és Irodalom* 2011-02-18 / 7. szám 5.o
- TAKÁTS Béla (szerk.): Helyismereti könyvtárosok XI. országos tanácskozása : *Helyismereti – digitalizálás – Európai Unió* : Szolnok, 2004. július 14–16.
- TOTH, Andrej: Magyar kisebbségi politikai pártok Csehszlovákiában és az 1935. évi elnökválasztás. In: *Századok* 2014 VI/1517p.
- TÖMÖSKÖZI Ferenc: A református egyház önszerveződésének folyamata Csehszlovákiában 1918–1923 között In: *Múltunk* 2020 / 2. szám
- VILLÁM Judit: Múlt és jelen országgyűléseinek dokumentumai – az Országgyűlési Könyvtár Magyar Parlamenti Gyűjteménye. In: *Aetas*, 2016 / 1.

- Vita a középkori forráskiadás módszertanáról. In: Fons (*Forráskutatás és Történeti Segédtudományok*) VII. évfolyam 2000.1. szám
- Vita az újkori forráskiadás módszertanáról. In: Fons (*Forráskutatás és Történeti Segédtudományok*) VII. évfolyam 2000.1. szám
- WELKER Árpád: Konzolidáció, historiográfia és a populizmus kihívása. In: 2000, 2013 október