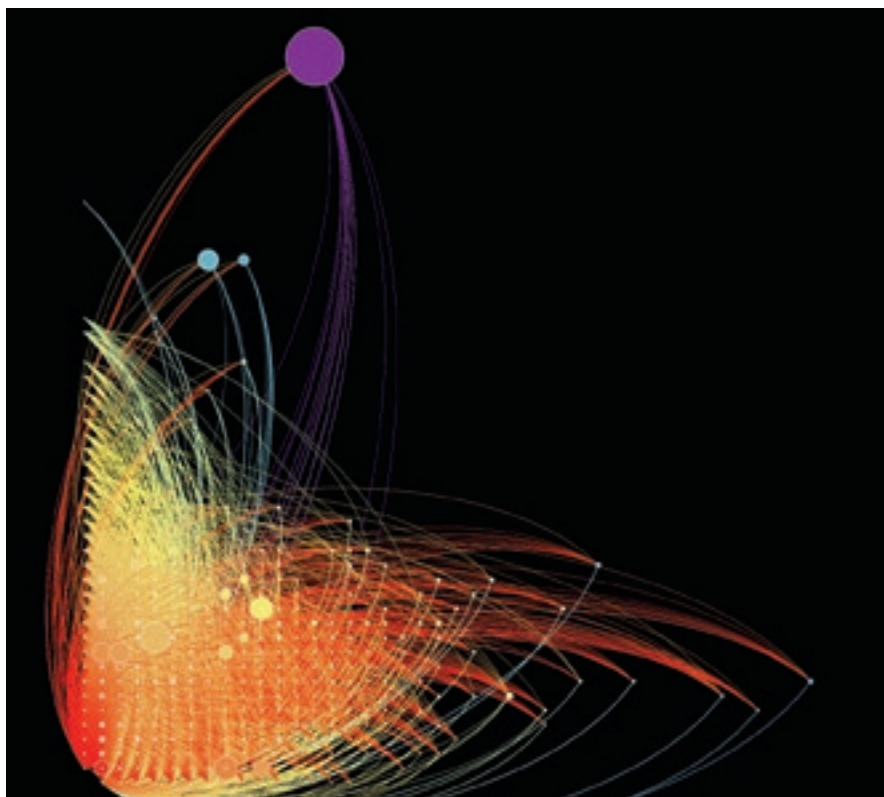


disputa



Vizualizáció a sefaria.org honlapon

BERÉNYI MARIANNA
AGGREGÁTOROK HALÁLA
ÉS FELHŐBEN ÚSZÓ ARCHÍVUMOK

IA közgyűjtemények, tudományos projektek ma már elképzelhetetlenek digitális adatbázis vagy legalább digitális adatbázis építésének szándéka nélkül. A tudományos viták és a közbeszéd egyaránt túllépett azon a dilemmán, hogy szükség van-e online elérhető, kereshető archívumokra. Így az elmúlt évtizedben világszerte naponta több ezer terrabájtnyi közgyűjteményi adat gazdagítja a virtuális világot, ugyanakkor újabb problémák vetődnek fel: hogyan őrizzük és használjuk jól az adatokat? Ezekkel a kínzó kérdésekkel négy olyan kutatóhoz fordultunk, akik teljesen más területről érkeztek, de látványos eredményeket értek el eddig az elektronikus adatbázisok építése és a különböző aggregátorokhoz való kapcsolódás területén. Fonyódi Krisztián a Szépművészeti Múzeum – Magyar Nemzeti Galériában dolgozik, országos szakfelügyelő, a MúzeumDigitár munkatársa. Horváth Iván irodalomtörténész csaknem fél évszázada foglalkozik digitalizációval. Toronyi Zsuzsannát, a Magyar Zsidó Múzeum és Levéltár igazgatóját már többször is kérdezte a MúzeumCafé a témában. Zólyomi Gábor pedig a sumerológia területén vett részt példaértékű digitális vállalkozásokban.

MC: Az információs forradalom, az informatika fejlődése régóta befolyásolja a humán tudományokat is. A magyar informatikai irodalomtudomány sokáig a világ élvonalához tartozott, többek között a te munkádnak köszönhetően. Mikor kezdtél a területtel foglalkozni?

Horváth Iván: Jelenleg az ELTE tanára vagyok, mire megjelenik a lap, már nem leszek az. Számítógépes adatbázis-építéssel – először R40-es NDK-számítógéppel, később R55-ös szovjet készülékkel – még 1976 őszén kezdtem foglalkozni a szegedi egyetemen, ahol óraadó voltam. Akkor vált világossá számomra, hogy az irodalomtudományban, az irodalomtörténetben, általában a történeti tudományokban nem lehet másként dolgozni, hiszen minden ténnyel, ami odavág, számolnunk kell. Az akkori főnökömtől azt a feladatot kaptam, hogy írjam meg a 16. századi magyar költészet történetét. Azt kérdeztem, melyek a 16. századi versek, hány darab van belőlük, hányat fordítottak idegen nyelvből? Akkor nem lehetett ezekre a kérdésekre válaszolni pontosan. Ma már tudjuk, hogy 1522 költemény van abban az adatbázisban, amelyet a tanítványaimmal építettünk. Az eredményeink

Magyarországon nem tudtak megjelenni, hosszú ideig tudományos témává sem válhattak, így 1992-ben Párizsban franciául jelent meg a régi magyar költészet adatbázisa a hozzávaló kétkötetes kézikönyvvel, amelyet életem főművének tartok. Több mint negyven éve foglalkozom a magyar kulturális örökség digitalizálásával, két könyvem is megjelent a témában, a legfrissebb is tizenkét éves, azaz nagyon elavultak mára. Ettől függetlenül nálunk, az ELTE-n készült el a világon első internetes kritikai kiadás, 1993-ban a Balassa-kódex, 1998-ban a Balassi-versek.

MC: *Gábor, te szintén az ELTE-n tanítasz, csak másik tanszéken.*

Zólyomi Gábor: Az ELTE BTK Assziriológiai és Hebraisztikai Tanszékét vezetem, azon belül a nyelvészet és a sumerológia a fő szakterületem. 1997 és 2000 között Oxfordban dolgoztam az *Electronic Text Corpus of Sumerian Literature* projekt létrehozásán, amely az assziriológia első digitális adatbázisa volt, lényegében az óbabiloni sumer irodalmi szövegeket tette online elérhetővé átírásban és angol fordításban. Ez egy lemmatizált korpusz, amelynek jelölőnyelvi része is van. Jelentősége az Iván által említett adatbázishoz mérhető, hiszen korábban ezek a szövegek csak obskúrus szaklapokban voltak elérhetőek, és csak azok számára, akik ezen a területen dolgoztak. Az adatbázisnak köszönhetően azonban mindenki számára elérhetővé és kereshetővé váltak. Többek között olyan humán tudományok képviselői számára is, akik korábban nem férhettek hozzá ezekhez a forrásokhoz. Az azóta eltelt húsz évben már megmutatkozott az eredménye, látom a publikációs listámon, hogyan hat az idézettség. De azt is hihetetlenül megváltoztatta, hol idéznek sumerológiai szöveget. Ugyancsak fontos a *Cuneiform Digital Library Initiative* projekt, amely aggregációs adatbázisnak mondható, hiszen az az ambíciója, hogy az összes létező ékírásos szöveget feltegye a netre másolatban, fényképben, és ha lehet, átírásban. Ez az adatbázis 550 ezer rekordot tartalmaz. Egyre inkább bővül, a fényképek egyre jobb minőségűek. Robert K. Englund amerikai professzor vezeti a projektet, de a világ minden tájáról kapcsolódtak hozzá kutatók, többek között én is. Először a nagy gyűjtemények anyagát, majd a kis amerikai gyűjteményeket digitalizálták. Ma már szinte nem létezik olyan ékírásos gyűjtemény, amelynek a táblái nincsenek fent. Ez fantasztikus lehetőség a kutatóknak, és lényegi minőségi változás: több százezer egységesen átírt, korábban nem hozzáférhető szöveg vált elérhetővé. A harmadik az ORACC (*The Open Richly Annotated Cuneiform Corpus*) nevű projekt, amely szintén amerikai kezdeményezés. Ebben az egyik korpusz a tanítványaimhoz és hozzám kötődik. Az OTKA-kutatásból 2008 és 2013 között

megvalósított *Electronic Text Corpus of Sumerian Royal Inscriptions* projekt újdonsága a többihez képest, hogy nyelvtani szempontból is elemeztük az egész korpuszt. Nemcsak a szavakra, hanem a ragozott alakokra is rákereshetünk. Ebben a korpuszban körülbelül ezernégyszáz szöveg kereshető átírásban, angol és magyar fordításban.

MC: Zsuzsa, te Gábor tanszékén doktoráltál, igaz, egészen más területen, a zsidó muzeológia helyét és szerepét vizsgáltad a magyar tudományosságban.

Toronyi Zsuzsanna: Valóban. Nekem két felem van, hiszen a Magyar Zsidó Múzeum és Levéltárban dolgozom. A levéltárak és a múzeumok egészen máshogy viszonyulnak a digitalizáláshoz, úgy érzem, a levéltárak ebben előrébb járnak. Bevett gyakorlat, hogy ezen a területen mindig mindent nemcsak digitalizálnak, hanem online elérhetővé is tesznek. Ennek mi is részei vagyunk. Nagyon fontos, amire Gábor is utalt, hogy kis intézményként is látszódnunk, sőt a kánon részévé válhatunk sokszor nem várt területen. Alapvető célunk, hogy a mi örökségünket is ugyanazon a platformon tegyük elérhetővé, ahol a többi levéltár, többi ilyen típusú gyűjtemény is megtalálható. Hiszen ez azt jelenti, ott vagyunk! Mi voltunk az egyik első olyan intézmény Magyarországon, amely 2010–2012-ben feltöltötte az akkor induló *Europeanára*. Létrejött egy tíz európai intézményből álló konzorcium, hogy bemutassa a zsidó közösségek szerepét az európai városokban. Ezek nem feltétlenül zsidó intézmények, hanem inkább zsidó örökséget őrző intézmények voltak, mint a frankfurti Goethe Egyetem, amelynek hatalmas jiddis nyomtatványgyűjteménye van, vagy a görög zsidó múzeum, meg persze a mi levéltárunk, amely akkor olyan kicsi intézmény volt, hogy egyedül dolgoztam benne. A projekt több tanulsággal szolgált. Csak a végére derült ki, hogy számos nem tisztázott kérdés volt, amit mi máshogy értelmeztünk, mint a kiíró. Az utolsó pillanatok kétségbeesett megoldásai nem bizonyultak sikeresnek, be is halt az egész rendszer. Ugyanakkor fontos tanulási folyamat volt: kiderült, miként érdemes bánni a múzeumi tárgyakkal, levéltári anyagokkal, hogy azok online elérhetőek, kereshetőek legyenek. Nagyon fontos az a tény, hogy akik elég elszántak, mégiscsak tudnak ilyen típusú anyagokat találni, másrészt meg kiemelt bennünket a láthatatlanságból. Emellett nekünk is volt egy világelső projektünk: 2003-ban publikáltuk a *Deportáltakat Gondozó Országos Bizottság* jegyzőkönyveit, ami akkor a világ legnagyobb kereshető holokausztvisszaemlékezés-gyűjteménye volt a maga 3665 digitalizált, kereshető jegyzőkönyvével.

MC: Krisztián, te a múzeumokat képviseled, még hozzá rögtön többféle módon!

Fonyódi Krisztián, Zólyomi Gábor, Toronyi Zsuzsanna, Horváth Iván
Szilágyi Lenke felvétele





Fonyódi Krisztián: A Szépművészeti Múzeum – Magyar Nemzeti Galériában dolgozom a digitalizálási és fotóosztályon belül. Jelenleg a digitális fotózási részt koordinálom, ahol azzal foglalkozunk, hogyan lefz a műtárgyból digitális objektum. A két nagy intézmény mellett a Hopp Ferenc Múzeum és a Vasarely Múzeum digitalizálási munkáját is elvégezzük. Korábban, amikor még nem voltak ennyire jól strukturáltak az intézményben a digitalizálási feladatok, más területekkel is foglalkoztam: adatbázis-kezelő szoftverekkel, dolgoztam több uniós, europeanás projektben. Ezen túl az Emmi Közgyűteményi Főosztályának informatikai szakfelügyelőjeként a magyar múzeumok digitalizálási stratégiájának kidolgozásában és megvalósításában veszek részt. Néhány évvel ezelőtt a Német Múzeumkutató Intézettel és a német múzeumok szövetségével egy olyan rendszer magyarországi bevezetésével kezdtem foglalkozni, amely kis és közepes intézmények számára kínál költséghatékony gyűjteménykezelési, publikációs és aggregációs szoftvert. Ennek köszönhetően a kismúzeumok, városi múzeumok olyan korszerű technikához juthatnak, amellyel meg tudják valósítani a digitalizálást. Ez kis projektként három éve indult, most már olyan nyolcvanfős felhasználói kör áll mögötte, amelynek tagjai tájháztól kezdve országos múzeumig mindenhol megtalálhatók. Tovább lépés lehet a határon túli gyűjtemények digitalizálása, ezért két irányba mozdultunk el. Egyrészt New Yorkban a Salgó-gyűjteményt, a határon túli legnagyobb magyar képzőművészeti gyűjteményt tettük online elérhetővé, most pedig Erdélyben a Kallós-gyűjtemény is bekapcsolódik. A részt vevő intézmények közösen digitalizálnak, közös névtereket használnak.

MC: *Ez egy felhőalapú rendszer. Nagyon sok intézmény azért kritizálja a MúzeumDigitárát, mert nem érzik biztonságosnak, hogy külföldön ilyen-olyan szervereken tárolják az adatokat. Szabad-e közgyűjteményi adatokat felhőben tárolni? Jogosak-e ezek a félelmek?*

F. K: Három-négy éve kezdtünk felhőalapú technológiával dolgozni, és azóta hirdetjük az ígét. Szerintünk egy forráshiánnyal küszködő szcénában ez az egyetlen megoldás. Az az érdekes, hogy az ellenállás egyre csökken. Az elején nem is tudták, milyen felhőről beszélek, ma egyre jobban elfogadják. A tárolás kockázatával nem kell számolni. Egy központi szerveren – a MúzeumDigitár rendszer esetében egy Németországban lévő nagy szerverparkban – sokkal jobb technikát tudok biztosítani, sokkal képzetesebb szakemberekkel tudom kontrollálni az egész nagy adatbázist. A nagymúzeumokban is folyamatos probléma a szerverek működtetése, tükrözése. Egy központi szervernél ez a gond elhanyagolható. Röviden azt lehet mondani, mindenki foglalkozzon azzal, amihez ért. Egy múzeum feladata a tartalom előállítás, a tartalomfejlesztés. A technológiát bízunk

azokra, akiknek ez a szakterületük. Egy szerverközpont több helyen elhelyezve meg tudja ezeket a problémákat oldani. Egy muzeológus ne legyen szoftverfejlesztő, szerverkarbantartó, hanem felhasználó.

H. I.: Én reakciós vagyok. Azt szeretném csak megkérdezni, hány szerverparkról van szó. Felhő nincs, az csupán egy „meteorológiai képződmény”. Hány szerverparkban tartod az adataidat, Krisztián?

F. K.: Legalább három helyre tükrözik az adatokat.

H. I.: Miért pont három, miért nem hárommillió? Van egy információtörténeti szabály, hogy egyre vacakabbak az információhordozók. Az ékírásos agyagtábla, pláne amit ki is égettek, eléggé bírja még. Dolgozom pergamenkódexekkel, tökéletesek. A régi famentes, ősnymtatványpapír hófehér, rajta koromfekete a nyomás. A rotációs papír már csak pár évtizedig marad meg. Mennyi energia kell ahhoz, hogy egy merevlemez letöröljek? Szinte semmi, a merevlemez pára. A 16. századi magyar nyomtatványok esetében kétszáz példányból megmaradt kettő, 99 százalék elveszett. A winchesterekkel mi lesz ötszáz év múlva? Milliós/milliárdos példányszám alatt nem érdemes digitális másolatokkal foglalkozni. De van megoldás.

F. K.: Attól függ, mire használod a szerverparkot!

H. I.: Hogy megmaradjon az adat.

F. K.: A kulturális örökség megőrzésének illúziója az emberiség történetében mindig jelen volt. Az emlékezés és a felejtés kapcsolata bizonyára feloldható dolog: a fejlődés egyszer majd eléri azt a színvonalat, hogy mindent megőrzzünk. De a digitalizálás korszakában sem hiszek abban, hogy eljön egy olyan technológiai megoldás, hogy a felejtést kiiktathassuk. Ez most a megőrzés egy módszere, de még fontosabb a megosztás. Egy agyagba vésett táblát biztosan jobban meg tudunk őrizni, de a metaadatok könnyebben megoszthatók.

Z. G.: A kérdés az, miről szól a digitalizálás? Arról, hogy meg akarjuk örökre őrizni az adatokat, amely lehetetlen? Vagy a hozzáférésre helyezzük a hangsúlyt? Az én szakmában ez utóbbi.

F. K.: Persze fontos kérdés a hosszú távú megőrzés, folyamatosan dolgoznak rajta az informatikusok.

MC: A közgyűjtemények, a múzeumok, könyvtárak és a levéltár elsődleges célja, hogy a tárgyakat eredetiben megőrizzék. Ehhez nem férhet kétség! A digitalizáció a kutathatóságról, a bemutatásról szól.

T. ZS.: A digitalizációval ráadásul a megőrzést is segítjük, mert amivel elkészültünk, azt berakjuk egy sötét, savmentes dobozba, nem tapicskolja senki, hanem a digitalizált változatot kutatja. Az eredetit pedig a legjobb tudásunk szerint őrizzük, az első tűzvészig vagy csőtörésig... A biteket muszáj állandóan frissíteni. A digitális megőrzés költséges megoldás, jóval többbe kerül, mint papíron őrizni az adatokat. Abban hiszek, hogy felhőben tároljuk az adatokat, még hozzá nemzetközi rendszerekben. Olyan, mint a Wikipedia, vagy a rockkoncertek közönsége: ha tömegnyi ember énekel, az már nem hamis, abból csak jó jöhet ki. Ha túl sokan akarjuk, hogy valami működjön, akkor működni fog. Ha nagy nemzetközi rendszerhez csatlakozom, akkor sok ember tudása, erőfeszítése biztosítja a fejlődést. Ráadásul a munkát support group segíti.

H. I.: Nem értem az ellentétet. A Gutenberg-javaslat, a sokszorosítás mint megőrzési mód egybeesett a dolgok szétesztásával. A hozzáférés javítása a megőrzés javítása. Az emberiség egyre inkább a hozzáférés javításával teszi lehetővé a kulturális javak megőrzését, fennmaradását. Azaz a példányszámot növeli. Ennek nincs akadálya. Digitalizáljátok a nagy nemzeti kultúrákat, és adjátok oda az iskolásoknak, mindenkinek legyen saját példánya a teljes nemzeti kultúráról, ami a mi magyar esetünkben is csak száz petabájtos nagyságrendű, pedig a magyar kultúra a nagyobbak közé tartozik! Legyen meg mindenkinek, vigyék magukkal az övtáskájukban, még akkor is, amikor már senkit sem érdekel, hogy „borjú, bárány béget”!

T. ZS.: Szerintem meg tegyék fel a netre, ahonnan bárki letöltheti, akinek szüksége van a bárányra a Toldiból. Ez egy szelekciós, új kánont kreáló folyamat is. Látni fogjuk, hogy a Mona Lisát naponta kétszázán töltik le, egy másik művet pedig senki sem. Utóbbi akkor nem frissül, az összes algoritmus hátradobja.

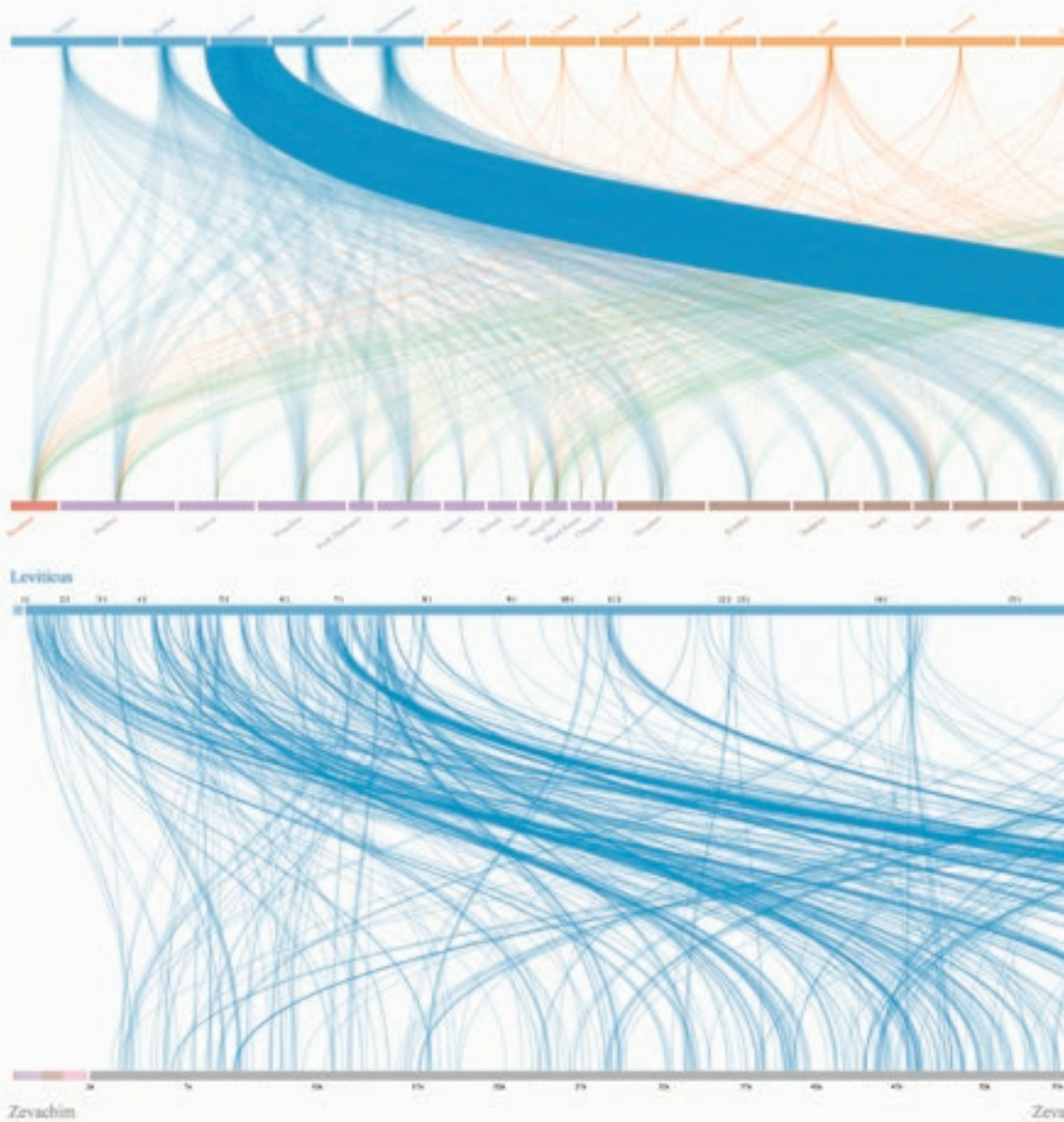
MC: Ennek viszont nagyon nagy a veszélye. Másrészt ott a felhasználó személye. Arról beszéltek, hogy a legfontosabb cél: „bárki” felhasználhassa ezeket az adatbázisokat. Egy sumerológiai adatbázis kinek szól? Ki nézheti meg? Tényleg bárki? Nincs veszélye, hogy újabb konteók szövődnék, kontrollálhatatlanná válnak az áltudományos nézetek?

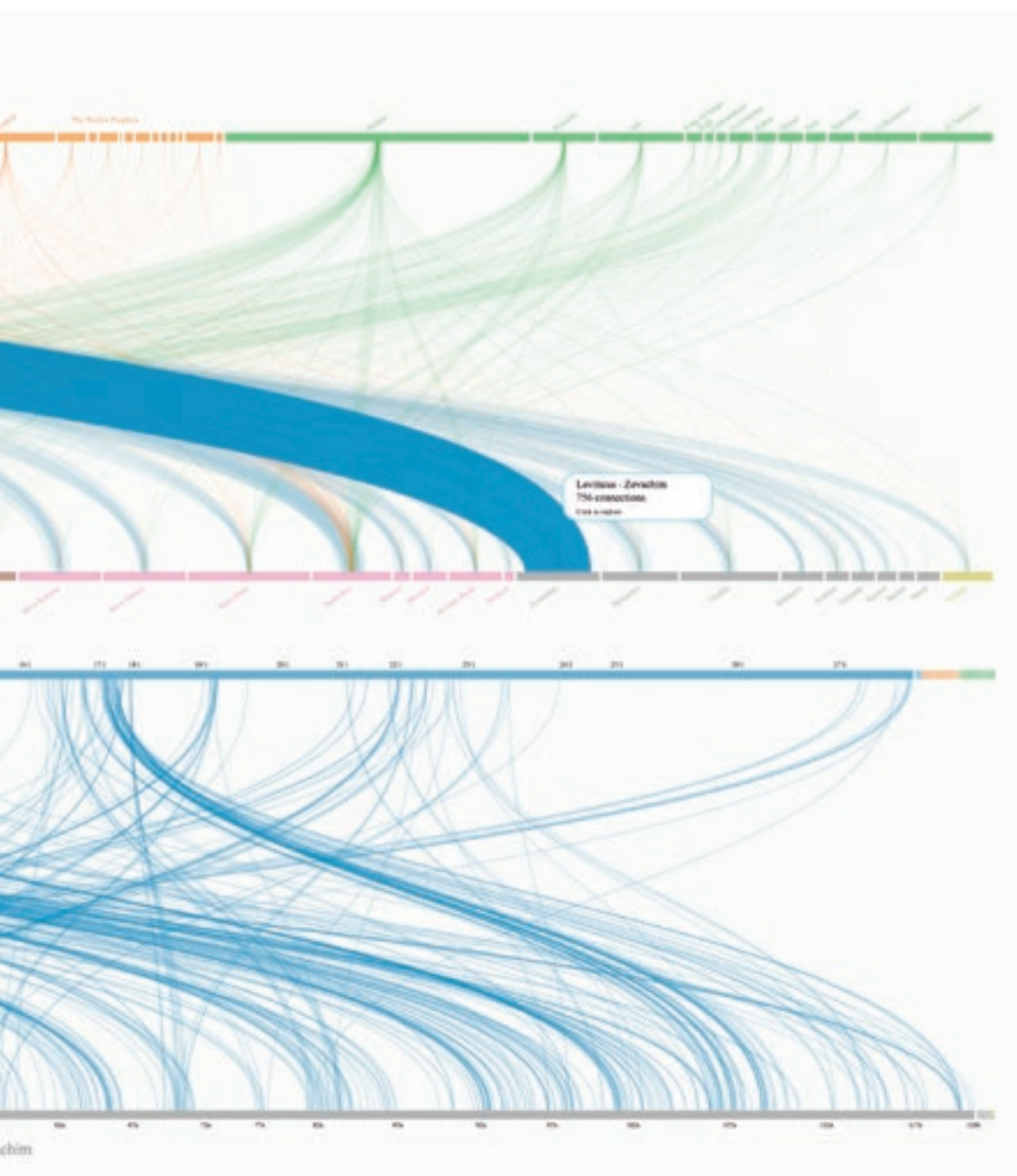
Z. G.: A mi tudományágunk korábban olyan volt, mint a beavatottak közössége. Mostanra ez teljesen megváltozott. Bárki számára elérhető ismeretek halmaza jött létre, ebből semmi rossz nem származhat. Azt azonnal lehet látni, hogy valaki ért hozzá vagy sem, ettől nem félek. Ráadásul a digitalizáció előtt írták azt a sok marhaságot, amelyre Komoróczy Géza is reagált a *Sumer és magyar?* című könyvében. Ehhez nem szükségesek a digitalizált táblák, sőt most azonnal lehet mondani, hogy ez a tábla nem így néz ki, nem az van rajta. Egy kattintásnyira vagyunk az igazságtól!

F. K.: Ugyanígy működik a Google víziója is, hogy az egész világot feltérképezi kívülről belülről. Természetesen nem tudnak mindenhol bejutni az autójukkal, de gyűjthetik a fotókat geokoordinátákkal. Ha nagyon sok turistafotó készült a múzeumról, egyre objektívebb kép áll össze. Crowdsourcinggal, a nagyközönség bevonásával a mennyiségből minőség is létrejöhét.

T. ZS.: Ismeritek a Zooniverse projekteket? Nagy kedvencem és álmom, hogy ilyeneket csináljak a saját tartalmaink alapján. A Zooniverse olyan civil tudományos platform, amely tömegek számára teszi lehetővé, hogy feltöltött digitális tartalmak feldolgozásában, azaz tudományos munkában vegyenek részt, sőt az önkéntesek aktív részvétele szükséges a kutatások elvégzéséhez. Ebben a weboldal játékosításra épülő megoldásai segítenek, amelyek élményszerűvé teszik az egyébként monoton feladatok elvégzését. Mindenféle digitális tartalmakat lehet ide feltölteni. Az én kedvencem az a platform, ahol ókori kopt szövegek átírását végezték el közösen a felhasználók. Ez úgy működik, hogy ha belépsz, kapsz egy ókori kopt szöveget görög jelekkel. Nem könnyű őket elolvasni, és rettenetesen sok vár átírássra. (Nyilván az ékírásos táblákkal is működik, én héber szövegek átírására szeretném használni.) Kapsz mellé egy görög ábécét, és elkezdhetsz klikkelgetni, hogy szerinted melyikre hasonlít legjobban. Ha három felhasználó is lambdát vél felfedezni, akkor elfogadja a rendszer, hogy valóban egy lambda. Ugyanígy crowdsourcinggal tömegek írják át a forrásokat, persze szakmai ellenőrzéssel. Egy másik projektben az Atlanti-óceánon a 19–20. század fordulóján áthaladó hajók naplóját tették fel digitálisan, hogy írják át a felhasználók őket közösen. Egy játék formájában ebből kell kiszedni adatokat a világ különböző pontján tartózkodó játékosoknak, amely során, szórakozás közben, hosszú távú meteorológiai kutatást segítenek elő. Most már több mint nyolcvan ilyen projekt van. A Tate Gallery például feltette a gyűjteményében őrzött művész-naplókat. Akinek az élete érdekel, azt segíthetsz átírni.

Connections between Talmud and Tanakh





F. K.: A természettudományi múzeumok is gyakran élnek ezzel a módszerrel. Komplet bogár- vagy lepkegyűjteményt tesznek fel az internetre, hogy a felhasználók határozzák meg őket, hiszen az intézmény néhány fős csapatának erre nincs lehetősége.

MC: *Hasonló módszerrel dolgoztak a brit képzőművészeti gyűjtemények, amikor arra kérték a felhasználókat, segítsék megkeresni az ismeretlen képek alkotóit. De a Fortepan is indít ilyen kampányokat. Megváltozik a cél: az intézmény, az archívum a kutatásba vonja be a felhasználót. Az adatok megosztásakor nemcsak az a fontos, hogy láthatóvá, kereshetővé tegyenek gyűjteményeket, hanem különböző aggregátorokon keresztül összehasonlíthatóságuk, összekapcsolásuk is lehetővé válik. Korábban is szóba kerültek már ezek a szervezetek, hiszen legalább három évtizede törekedik rá a világ, hogy szülessenek olyan univerzális adatbankok, ahová bárki feltöltheti hasonló szempontok szerint az adatait. Ezek szükségességéhez nem fér kétség, mégis úgy tűnik, a kezdeményezések rendre kudarcba fulladnak. Milyenek a ti tapasztalataitok?*

T. ZS.: Én az Europeanát láttam bedőlni technikai problémák és félreértések miatt. Ilyen projekt akkor tud működni, ha irgalmatlanul pontosan meghatározzák, mit kell csinálni, és ha egy metaadatsort a világon mindenki ugyanúgy értelmez. Nálunk a legnagyobb félreértés a már említett projektben, ahol múzeumok, levéltárak és könyvtárak dolgoztunk közösen, abból a kérdésből fakadt: hány oldalt – page-et – fogunk digitalizálni. Én négy kötetet kívántam digitalizálni, amelyben volt körülbelül kétezer oldal, és beírtam, hogy kétezer page-re számíthatnak. Aztán kiderült, hogy az Europeana valójában nem oldalra gondolt, hanem metaadatszetre. Én feltöltöttem négy metaadatszettet, és megígértem kétezeret. Ez az egész projektünket meggyilkolta. Volt olyan partner, aki ezek után minden oldalhoz rendelt egy rekordot, amitől pedig kereshetlenné vált a rendszer. Ha rákeresek valamire, ahol ötmillió találatot kapok, nem nyertem semmit, főleg egy olyan rendszerben, ahol az elnevezésekkel is komoly probléma van. Amíg nincsenek egységes szótáraink, hogy mindenki ugyanazt értse egy kifejezésen, buta rendszereket hozunk létre, aminél összehasonlíthatatlanul jobb a Google.

MC: *Gábor, ti speciális adatbázist hoztatok létre speciális célcsoportnak. Nálatok hogyan működik ez?*

Z. G.: Tényleg kell egy központ, amely szigorú szabályokat fektet le, meghatározza a jelölések módját. Az ORACC-ban van harminc-negyven projekt, amelynek munkatársai mind másként viszonyulnak az anyaghoz. Viszont van egy alapszintje

a jelölésnek, amelyet mindenki elfogad. De ettől működik. Ugyanakkor van egy nagyon idegesítő része ennek a szoftvernek, a feltöltött szöveget validálni kell, és ha formailag hibásan töltünk fel valamit, nem működik. A validálással viszont megelőzzük, hogy rossz rekordok kerüljenek fel.

MC: Tágabb területen ez miként valósítható meg? Megvalósítható egyáltalán?

T. ZS.: Ha nem mindig a világot szeretnék megváltani, hanem egy minimum programot akarunk megvalósítani, lenne rá esély. Számomra sokkoló, hogy mindig túl sok adatot akarnak felvinni, amikor egy átlagos kutató számára hat adat is elég lenne. Meg kell határozni, mennyi az a minimum adat, amivel használható a rendszer, és nem szabad tovább bonyolítani.

MC: Pedig az Europeana az ókori Alexandria híres könyvtárának mintájára kívánta felhalmozni és digitálisan elérhetővé tenni Európa kulturális kincseit.

H.I.: Ott voltam az Europeana egyik ünnepén, ahol ketten csak azzal foglalkoztak, hogy pénzt hajtsanak fel a projekt számára. Ez a hágai királyi könyvtár egyik osztálya. Lángelme volt, aki kitalálta az Europeana elnevezést, holott nem igazán európai intézményről van szó. Mégis minden ország mögójük állt, egy európai uniós biztos is jelen volt. Kár, hogy sosem volt jó. Mégis védeném az aggregáció gondolatát, mert elhiszem a Gutenberg-receptet: egyre kisebb energiával kódoljuk a jelhordozókat, egyre több példányban, és egyre inkább a példányszám révén őrizzük meg a tartalmat. Az aggregátor célja a központi mentés és sokszorosítás. Az internet megsérti az információ-történet szabályát. A kulturális adatokat anélkül osztja szét, hogy megtöbbszöröznék őket. A kultúra megőrzésére ezt nekünk magunknak kell megtennünk. A nemzeti kultúrákat szét kell osztani, hogy mindenkinek legyen példánya belőle. Hamarosan gyártható lesz egy olcsó és kisméretű eszköz, rajta az összes múzeum, könyvtár, levéltár, filmek. Ilyenből kell sok millió, és akkor egyik-másik fennmaradhat. Nem szabad megfeledeknünk arról, hogy a közgyűjteménynek csak másodlagos feladata, hogy megmutassa a rá bízott kincseket. Első kötelessége: hogy megőrizze őket.

F.K.: Tiltakozom! A megőrzés és a nyilvánossá tétel nem zárja ki egymást. Ha megőrzöm, de nem mutatom meg, csak egy múzeumi raktárt tartok fenn. Viszont egy aggregáció sikeréhez három dolgot érdemes végiggondolni. Ki az aggregátor célcsoportja? Hogyan biztosíthatom a hozzáférhetőséget az aggregátorhoz? Miként határozhatom meg a minőség és a mennyiség viszonyát? Az Europeana első

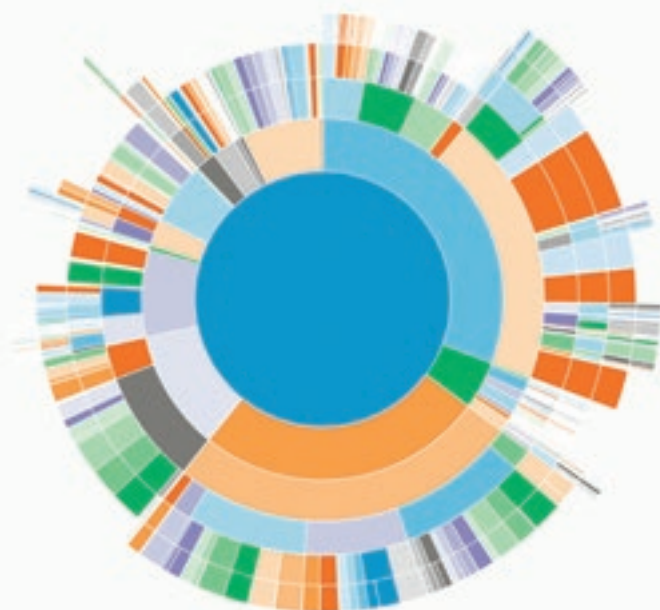
Sefaria English Library Mishnah

The Mishnah with Chofadiah Berenson by Rabbi
Shraga Silverstein (ms)
6,217 words



Sefaria English Library

Sefaria Community Translation (en)
479,086 words



fázisában adattömeget akartak létrehozni, összegereblyézték teljesen különböző szektorokból, nem volt rendes metaadat-struktúra, nem lehetett összehozni az adatokat. Hatalmas összegekkel támogatták meg az adatfeltöltést, miközben a többnyelvűséget sem oldották meg.

T. ZS.: Mi 2010-ben olyan policy mentén kezdtünk el dolgozni, hogy mindenkinek a saját anyanyelvén kell leírni. Azt ígérték, hogy lesz beépített egységes szótárrendszer, de ez azóta se valósult meg. Úgy próbálják megmenteni az eddigi munkát, hogy olyan tematikus anyagokat állítanak fel, mint például az *Europeana Fashion*.

F. K.: Narratívákat hoznak létre, mert narratívákon keresztül tanul az ember, nem úgy, hogy különböző objektumokra kattintgat. A célcsoportok meghatározása ennél is nagyobb gond, méghozzá valamennyi múzeumi aggregátor számára. A múzeumlátogató nem alkot webes közösséget, nem tudunk velük olyan portált felépíteni, mint a zsebragyűjtőkkel vagy a hobbi barkácsolókkal, esetleg egy tudományos közösséggel. Ugyanakkor az aggregáció elve már szakmai portálok esetén is rendkívül jól működhet. Illúzió, hogy egy átlag európai polgár beüsse először, hogy *europa.eu*, majd egy keresőfelületen elkezdjen keresgélni. Csak az fog működni, amit a Google-kereső az első oldalon kidob. Egy tudós tudja, hova menjen, a nagyközönség nem. Ezt mutatta több aggregációs projekt is. Sok pénzből, jó minőségben digitalizálták a nyugat-európai múzeumok anyagait, és kiderült, a kutya nem nézi. Nem lehetett odacsábítani senkit. A hollandok jobban kihegyezték a célcsoportok számára az aggregátorokat, például az állami adminisztráció felé. A *MúzeumDigitár* német fejlesztői lemondtak arról, hogy ők lesznek a mindentudó aggregátor, és elkezdtek együttműködni az olyan, átlagemberek által is használt weboldalakkal, mint a Google. Azaz úgy kell együttműködni velük, hogy első oldalon dobja ki a tartalmainkat. Németországban a Deutsche Digitale Bibliothekot 27 millió euróból hozták létre, de jelenleg nem igazán működik, pedig nyolc-kilenc millió eurót költenek rá évente.

MC: *Gábor, a ti adatbázisaitok elérhetők a Google-ről? Ingyenesen?*

Z. G.: Természetesen. A Google a digitalizált tartalmakat és szövegkiadásokat is megtalálja.

MC: *Iván, a hetvenes évektől hogyan változott azoknak a felhasználóknak a köre, akiknek a digitalizált tartalmak készültek? Eleinte nem volt kérdés: a tudományos kutatókon kívül nem vetődött fel más célcsoport.*

H. I.: Egész életemben a kényszerolvasókkal törődtem. Ez jó közönség, nem is változik. Ne becsüljük le őket. Attól, hogy sokan megtanulnak írni-olvasni, érdeklődniük is kell-e irodalmi dolgok iránt? A finanszírozás ügyében 2001-ben írtam egy cikket *Művelődés közpénzen* címen az internetes könyvtárakról. Abból indultam ki, hogy közpénzből csak olyasmit szabad vásárolni, aminek a hasznát az adófizetők élvezhetik, miközben az adófizetőket sem szabad ama jogukban korlátozni, hogy hozzáférhessenek a közgyűjtemények anyagához, még ha annak egy része nem feltétlenül az ő befizetések nyomán jött is létre. A kutatásaimért fizetést kapok, esetleg más támogatást, aztán a kiadáshoz is. Megjelenik a könyv, a könyvtár is kap támogatást, hogy megvehesse. Úgy gondolom, hogy az én anyagaimat ingyenesen elérhetővé kell tenni. Minden, a közösség költségén létrejött adat felhasználási jogának tulajdonosa maga a közösség, s nem az az intézmény, amelyben az adatok létrejöttek.

MC: *Mennyire ismeritek a célcsoportjaitokat, a közösségeiteket?*

Z. G.: Szinte név szerint. Az ORACC rajta van a Google Analyticson. Látom, melyik városból használják az adatbázist.

T. ZS.: Két nagy csoportunk van. Egyrészt a külföldi látogatók, családkutatók, másrészt van egy szűk közösség, akiket mi is szinte név szerint ismerünk. Fontos, amikor valaki megosztja, nézegeti a nálunk őrzött családi hagyatékról szóló anyagokat. A magyar és az angol nyelvű honlap között látjuk a különbséget, annak ellenére, hogy több a külföldi látogató, a magyarok szívesebben böngésznek az adatbázisban.

F. K.: A *MúzeumDigitár* oldallal kapcsolatban sokat kutattunk. A külföldi kollégáknak érdekes tapasztalatai vannak a célcsoportok váratlan felbukkanásával kapcsolatban. Észrevették például, hogy vasárnap este a portál látogatottsága megugrott. Mi lehet a magyarázata? Vasárnap délután a kávé és a süti után odaülnek a német polgárok a gép elé, hogy böngésszenek? Utánanézték, hogy kik is a felhasználók. Kiderült, hogy az eBay régiségei közül vásárlókat vonzották a digitális gyűjtemények. Eredeti forrásokat, hiteles tárgyakat kerestek, hogy össze tudják hasonlítani a kínálattal. A képzőművészeti adatbázisok között az *Artbuy* fizetős szolgáltatása is érdekes. Ötszáz euróért regisztrálhatnak a módosabb gyűjtők. Megnézhetik, hogy ha Rembrandtot akarnak venni, mit lehet kapni, mennyi volt az aukciós ára az elmúlt években, de ez nem kapcsolódik múzeum adatbázishoz. Az én víziómban nem csak véghasználói csoportok vannak, hanem két nagy adatbázis innovatív összekapcsolása révén jöhet létre újdonság.

Még nem tudok pontos példát mondani, de például az *Artbuy* is gazdagodhatna autentikus közgyűjteményi tartalmakkal. A magángyűjtők, művészeti galériák mindenre azt mondják, hogy Rembrandt, mert el akarják adni, de egy közgyűjtemény az eredetiségre épít. Ha a kettőt összekötnénk, létrejönne egy olyan felület, amely során a közgyűjteményi adatbázis is újabb szerephez jut. Az új, 2017-os kormányhatározatban elfogadott *Közgyűjteményi Digitalizálási Stratégia* sem az aggregátorokra koncentrál, hanem a felhasználhatóságból indul ki. Az oktatási célú felhasználásban látják a közgyűjteményi adatbázisok legegyszerűbb, legfontosabb szerepét. A közgyűjteményi adatbázisokat kössük be a közoktatásba, a jövő tananyagába. Ha azt mondom, hogy honfoglalás, be lehessen linkelni tárgyakat, lehessen hozzá illusztrációkat keresni. A jövőben az illusztráció helyén kérdés állhat, és a gyerekek kereshetik meg a képeket. Ennek az alapja viszont a megfelelő közgyűjteményi adatbázis, ami még nem áll rendelkezésre.

H. I.: Azt szeretem, ha a narratívakészítést nem hagyják rá a muzeológusra vagy a könyvtárosra, hanem külső szakembernek adják át. A Széchényi Könyvtártól kaptunk ilyen feladatot a magyar irodalomtörténet területéről, a felhasználók (gimnazisták és egyetemisták) narratíván keresztül jutnak hozzá a tartalmakhoz. Párizsban magyar nyelvtörténet órán egy piciny, magyar társalgási zsebkönyvet vettem elő a Zsigmond-korból, amelyben főleg szexuális közeledésre, étkezésre vonatkozó mondatok voltak. Sikeresebb volt, mint a *Halotti beszéd*. Az emberek érdeklődését könnyű kihasználni! Az *eBay* népében is komoly szakértelmek alakulnak ki.

Z. G.: Az ORACC ismeretterjesztő szövegei is kifejezetten népszerűek, többen olvassák, mint a szakmai tartalmakat.

F. K.: Mondhatjuk azt, hogy aggregátoroknak nem maga az aggregáció a céljuk. Ha szabványok szerint összehozunk minőségi tartalmakat, az tovább gyúrható, az aggregátoron keresztül pedig a társadalom különböző szegmensébe szétfolyhat a tartalom. Például a Szépművészeti Múzeum egyiptológusai a *Minecraft*-tal fejlesztenek egy olyan alkalmazást, amely során a Nílus áradásával kell szembenézni. Lehet hidat építeni, járványokról beszélni, kiderül, mi történik a piramissal, ha víz alá kerül. Sok múlik azon, ha az adatokat játékokba, online piacterekre, a turizmus területére folytatjuk. Itt a lehetőség a széles körű társadalmi hasznosításra.

T. ZS.: Nem beszéltünk még a tömeges adatvizualizációról, amely különösen levéltári területen lehet fontos. Csaták, jegyzőkönyvek adatai önmagukban unalmasak,

viszont adatok tömegéből új tudások születnek meg, juthatnak el széles körbe. Ilyen forradalmi volt a *sefaria.org* alkalmazása, amelyen az összes klasszikus héber szöveget beírták, és így megnézhetjük, hány szóból áll a *Talmud*, összekapcsolhatunk különböző fogalmakat, bővíthetjük. A *Sefaria* a „digital humanities” legszebb gyakorlatát követve „vizualizálja”, ábrázolja is a roppant mennyiségű zsidó szakrális szöveget. Vizualizációik egyrészt megmutatják a szövegek közti kapcsolatokat, azaz tanulást könnyítő eszközként is értelmezhetők, de talán ennél is többet jelentenek. A több ezer éves szövegek közti kapcsolatok – amelyek eleinknek is ismertek voltak, vizuális megjelenésük azonban nem – gyönyörű, szimmetrikus és/vagy repetitív mintát adnak. Szekuláris-művészetelméleti olvasatban ez nem más, mint az igazi, a zsidó szellemet megmutató, transzcendens, metafizikus zsidó művészet.

MC: Kijelenthetjük, hogy a játékosítás és a széles körű, asszociatív együttműködések irányába mozdult el az online adatbázisok világa?

H. I.: Azon dolgozom, hogy a magyar irodalomtörténet egy képernyőoldalon tökéletesen érthető formában elmondhassuk. Következő kattintásra nagyobb korszakokat, majd alkotókat látunk, utána pedig a műveket, aztán az első kiadások, kéziratok, a tudomány lényege tárul fel. Egészen kicsi dolgokból kiindulva az egészet meg lehet közelíteni.

F. K.: Hierarchizálsz a tudást? Mi kerül az első oldalra?

H. I.: Az utak anarchikusak. De az ezer nagyobb szám, mint az egy, és vannak dobozok, Arisztotelész óta ilyen módon dolgozik a tudomány.

Z. G.: A kérdés inkább az, hogyan hat mindez a tudományra? Hogyan hatnak a nem hierarchikusan elérhető ismeretek? Kisantal Tamás *Az élet tanítómesterei* című könyvében van egy remek tanulmány *TörtéNet – elmélet* címmel, ahol leírja, hogy a nem hierarchikus adatbázisok miként hatnak a történelemszemléletünkre.

T. ZS.: Sokkal könnyebb így levéltárban dolgozni. Régen a proveniencia elve alapján helyeztük el a dolgokat a rendszerben. Sok esetben véletlenszerűvé vált, milyen anyagok kerülnek egy-egy kutató kezébe. Forradalmasította a kutatást, hogy a digitalizációnak köszönhetően számtalan elérési útvonal, szempontrendszer élhet egymás mellett. A levéltár szempontjai már nem határozzák meg kétszáz évre a kutatók felhasználási lehetőségeit. Végtelenné váltak a lehetőségek.

„A kulturális örökség megőrzésének illúziója az emberiség történetében mindig jelen volt. Az emlékezés és a felejtés kapcsolata bizonyára feloldható dolog: a fejlődés egyszer majd eléri azt a színvonalat, hogy mindent megőrzzünk. De a digitalizálás korszakában sem hiszek abban, hogy eljön egy olyan technológiai megoldás, hogy a felejtést kiiktathassuk.” (Fonyódi Krisztián)