

Kokas Károly (Szeged)

Digitális bölcsészet 2016

A bölcsészek és az informatikai megközelítés: régen és most¹

„S akkor mindig megnyomom... F1 ... Q...”²
(Monok István, 1986)

Ez a dolgozat (talán inkább esszé?) azt próbálja körvonalazni, hogy az angol nyelvterületen honos *digital humanities* mit is foglalhat valójában magában? A manapság igen divatos téma új diszciplína lenne talán? Vagy kutatások egészen új módszertana? Esetleg a technikai fejlődés által inspirált alapos szemléletváltás? Netán mindezek keveréke?

Felülnézetből legalább kétféle szélsőséges megközelítés veszélye rajzolódik ki: az informatikusok hordják az újabbnál újabb megoldásokat, lehetőségeket és eszközöket, mi bölcsészek pedig próbálunk „képben maradni” és azokat megérteni és használni, vagy a „csillagok háborújának” bölcsész fiai álmokat szönek arról, hogy az általuk zömmel triviálisnak képzelt és informatizálható problémáikat a számítógép gyorsan megoldja, vagy legalábbis nagyban segíti.

Könnyű lenne mindezt feloldani azzal, hogy az igazság valahol középen van, esetleg hogy mindkét megközelítésnek van értelme és haszna. Azt gondolom, ahhoz, hogy legalább „madártávlatból” világosabban lássunk, érdemes áttekinteni, hogy a bölcsészet informatizálása honnan is indult, milyen folyamatok zajlottak le és most hol is tartunk. Ez a rövid áttekintés vállaltan szubjektív lesz, remélhetőleg annak néhány előnyével és – sajnos – annak minden hátrányával is (amennyiben nem ad monografikus és kimerítő képet a témáról). Ennek a megközelítésnek velejárója, hogy erős hangsúlyt kapnak a szegedi és a könyvtári jellegű projektek és megközelítések, s az is, hogy a példák – beszűkítve a témát – tipikusan történelmi, irodalomtörténelmi jellegűek lesznek.

Analógiák: ősemberek, tűzhányók és forradalmak, meg tengeralattjárók...

A *Campi Flegrei* olasz elnevezés magyarul „lángoló mezőket” jelent, hiszen ez a Nápolytól és a Vezúvtól északnyugatra fekvő nagy terület több mint kéttucatnyi vulkáni kúpnak ad otthont, s igen intenzív földtörténelmi múlttal büszkélkedhet. Ezen igen régi események sorából is kiemelkedik az a kb. 40 ezer éves kitörés, ami miatt a szakemberek „szupervulkáni” kitörésről beszélnek, s amelynek hosszan tartó hatása az akkori világra elképesztő lehetett. Ma már többé-kevésbé elfogadott a szaktudományban, hogy a sztratoszférába került nagy mennyiségű kén-dioxid és hamu minimum 2–4 °C-kal csökkentette az átlaghőmérsékletet és erőteljesen hozzájárulhatott mindez a *neander-völgyiek kihalásához és a crô-magnoni ember elterjedéséhez*. Ha a folyamat okai és összefüggései most nem is fejthetők ki, az azért rendkívül figyelemreméltó, hogy mindennek valódi bizonyítékát a grönlandi jég alatt, s a természetben máshol is föllelhető *látványos hamurétegek* biztosították, csakúgy, mint a dinoszauruszok jóval korábbi kihalását magyarázó elméletek esetében.³

Amikor 2010 tavaszán az izlandi *Eyjafjallajökull* vulkán kitört, s hatalmas korom és füstfelhője az északnyugati légteret kis híján teljesen használhatatlanná tette a repülés számára,

¹ Ez a dolgozat a Szegeden az SZTE Klebelsberg Könyvtárban 2015. október 12-én megrendezett „Digitális bölcsészet Szegeden” című minikonferencia előadásaira és tanulságaira is támaszkodik. Vö. <<http://digibolcsesz.ek.szte.hu/>>

² A közlés műfaja okán, talán idefér egy *szubjektív jegyzet* is: amikor mintegy harminc éve Monok Istvánnal, a kis Commodore 64 géppel adatbázis-kezelési kísérleteket folytattunk a Suberbase 64 nevű programmal, még álmunkban sem gondoltuk, hogy az informatikának akkora szerepe lesz egy viszonylag konzervatív tudományban, mint a könyvtár- és bölcsészettudomány. Akkor már az is boldoggá tett minket, ha ki tudtunk lépni a programból, a fent leírt módon. (Ezt azóta is sokat emlegetjük.)

³ Minderről számtalan cikk tudósít az interneten vagy a szaksajtóban is, például William DAVIES, Dustin WHITE, Mark LEWIS, Chris STRINGER, *Evaluating the transitional mosaic: frameworks of change from Neanderthals to Homo sapiens in eastern Europe*, *Quaternary Science Reviews*, 118(2015), 211–242.

sokaknak eszükbe juthatott egy másik, talán még vadabb és egyben híresebb izlandi vulkánkitörés, az 1783-as. A vulkánkitörés elképesztő és elnyúló hatásáról sok kortárs megemlékezik (köztük 1784-ben *Benjamin Franklin* is), de a meteorológiai, földtani vizsgálatok eredményét a tudomány csak közvetlen korunkban kapcsolta össze a nagy francia forradalom kitörésével. Ennek egyik oka az volt, hogy a történészek jelentős és kiterjedt kvantitatív kutatások után egyértelműen bizonyítottak látták, hogy nemcsak rettenetesen fenyegető élmény volt az embereknek a kitörés, hanem ténylegesen iszonyatos gazdasági károkat is okozott. S ennek egyik legexponáltabb területe éppen Észak-Franciaország volt. Így az *ancien régime* elleni elkeseredett haragnak bizonyára jelentős gyúanyagot adott a nyomor és nincstelenség, néhol éhezés, ami a vulkánkitörés utáni terméketlen év(ek)ben következett be.⁴

Ebben az esetben már világos volt, hogy nem csupán a vékony sötét színű hamuréteg megtalálásáról, azonosításáról és a humán történelmi eseményekhez való közvetett kapcsolódásáról beszélhetünk, hanem arról is, hogy a *tömeges adatfeldolgozás* a hipotézist alátámasztotta.

Ennek a módszernek persze volt már akkor előképe is, mégpedig a második világháború titkainak megfejtésében, amikor is 1959-ben nem kis szenzációt keltve megjelent Alexander L. George *Propaganda analysis: a study of inferences made from Nazi propaganda in World War II* című könyve.⁵ Ez a híres monográfia azt próbálta bemutatni, hogy a háború alatt a szövetséges hírszerzés hogyan kísérelte meg kikövetkeztetni a *háborús náci propaganda tartalom-elemzésével* a német U-Bootok elleni harc sikerességének mértékét. A háború utáni adatok és a kielemezett trendek összehasonlítása később megmutatta, hogy a nagy tömegű „adatfeldolgozás” segítségével be lehetett látni a háborús kulisszák mögé, a német propaganda tengeralattjáró-témától való elfordulása egyenes arányban volt a szövetségesek harcmodorának sikerességével.

Vagyis az adatelemzés mindkét utóbbi esetben bizonyította, hogy a geológus-történészek módszeréhez hasonlóan *a szövegvilágokból is kinyerhető a „vékony fekete csík”, ami olyan következtetésekre ad alkalmat, amit maga a szövegek egyszerű olvasása, ismerete nem biztosít.* (A „vékony fekete csík” a föld rétegződésében is ott volt, csak nem láttuk, mert nem tudtuk, hogy kell néznünk azt. Hasonlóan, mint a vulkánkitörés nyoma az iratanyagban, s a tengeralattjáró-háború sikere a goebbelsi propagandagépezet teljes termésében...)

Történeti előzmények

Az iménti gondolatmenetben már tulajdonképpen megjelent néhány a történeti előzmények közül, de mégis érdemes kitérni további, más jellegű körülményre. A *digital humanities* jelentős nemzetközi szakirodalma komoly kísérleteket tett arra, hogy tisztázza a digitális bölcsészet kezdeteit és fejlődésének főbb mérföldköveit. Nem szeretném ezeket a stációkat újrásorolni, inkább pár olyan elődöt és előképet említenék, amelyek nem igazi történelmi előzmények, de az eddig vezető gondolkodásban nagyon fontos állomások voltak. Természetesen ezek csak példalozások és illusztrációk, de talán megvilágítják, hogy *a digitális bölcsészet – paradox módon – nem a számítógépek bölcsészeti kutatásokba való bevonásával kezdődött el.* Mindkét példám szinte közhelyszerűen ismert történész és filológusi körökben, ezért nem is igényel hosszabb kifejtést.

A 19. század végi könyvtárosi, lexikográfusi eredmények egyik ékköve *Szinnyei József* (1868–1913) elképesztő teljesítménye, amivel feldolgozta az addigi magyar tudományosság anyagát. A kortársak által „gőzhangyaként” is emlegetett szakember milliós cédulaanyagot készített a teljes magyar szakirodalmi repertoárból, amely egyenként tartalmazta az egyes művek bibliográfiai adatait, legyenek azok bár könyvek vagy egyszerű sajtócikkek. Kézenfekvő volt ezt a nagy mennyiségű cédulát a szerzők betűrendjébe összerendezni, és így kialakult a majdani *Magyar írók élete és munkái* című művének nyersanyaga, ami alapján a több mint 30 ezer írói,

⁴ Vö. például <<https://en.wikipedia.org/wiki/Laki>>

⁵ Alexander L. GEORGE, *Propaganda analysis: a study of inferences made from Nazi propaganda in World War II*, Evanston, Row, Peterson, 1959.

tudósi portrét és munkásságot feldolgozó sokkötetes kézikönyv elkészülhetett (1891). Kevésbé ismert, hogy a cédulaanyag korábban már egyszer életre kelt, hiszen 1874-ben Szinnyei kiadta a *Hazai és külföldi folyóiratok magyar tudományos repertóriumát* is. Ha valaki ismeri az anyagot, s kicsit könyvtáros, lexikográfus szemmel tekint rá, világossá válik, hogy nagyrészt az történt, történhetett, hogy Szinnyei „kifordította” a meglévő hatalmas cédulaanyagát, s amit tematikusan már egyszer elrendezett, azt most – sok adattal kiegészítve persze – új sorrendbe, szerzői betűrend szerint „újrahasznosította”. Véleményem szerint az óriási adatmennyiség és a módszer együtt nyugodtan tekinthető a hazai digitális bölcsészet egyik, számítógép előtti előzményének. (Más hasonló példák is vannak a gondolkodásmód kialakulására a pozitívizmus korában, de ez talán eklatánsan mutatja, mire gondolok.)

Egy másik, a tartalmi vonatkozáshoz még közelebbi példa a számítógép előtti korból. 1929-ben francia történészek egy csoportja új szaklapot hozott létre. Lucien Febvre (1878–1956), Henri Hauser (1866–1946) és Marc Bloch (1886–1944) alig volt fiatalabb, mint a szeretve tisztelt és most mégis meghaladni kívánt Henri Pirenne mester nemzedéke, de aztán a másodiknak induló csoport, Fernand Braudel (1902–1985), Georges Duby (1919–1996), Pierre Goubert (1915–2012), Robert Mandrou (1921–1984), Pierre Chaunu (1923–2009), Jacques Le Goff (1924–2014) és Ernest Labrousse (1895–1988) munkássága már a 20. század derekára esett. Történész berkekben ez a névsor hatalmas tisztelettel övezett, mint ahogy a lap is, amiről szó van, az *Annales d'histoire économique et sociale*. A szakfolyóirat és köre iskolaként is ismert, a híres *Annales Iskoláról* van szó.⁶

Az *Annales* nemzedéke, nemzedékei sok tekintetben forradalmasították a történettudomány módszertanát, szemléletét, témaválasztását, sőt előadásmódját is. Itt most az első, a módszertani változás érdekes a számunkra. Lényegében arra szeretnék utalni, hogy sajátos, a gazdaság és a társadalom mozgásait megragadó és leíró szempontjaikhoz az út a korábban ismeretlen mértékű alapadatgyűjtésen és feldolgozáson keresztül vezetett. Ez értelemszerűen azzal járt, hogy történészként az elsők közt szembesültek a nagy adattömegek elemzésének problémáival, a statisztikai módszerek, korrelációk alkalmazásával, illetve mindezek eredményeinek grafikus ábrázolásaival. Nem hiszem, hogy bővebb magyarázatra szorul, hogy miért jönnek számításba a digitális bölcsészet kapcsán már 1950 előtt írott műveik is. (Érdeemes megnézni például a talán legnagyobb alak, Fernand Braudel fő művét a *Mediterrániumot*, amelynek már első kiadása is ebbe az irányba mutatott, 1949-ben.⁷)

Különösen az *Annales* utóbbi példája mutatja, hogy a gondolkodás, a szakmai munka kihívásai vezettek el oda, hogy olyan módszerek alkalmazása kezdődött meg, amelyek aztán magától értetődően megtalálták a számítógépet, a modern adatfeldolgozás legújabb útjait. Ebben az összefüggésben kimondhatjuk, itt a tojás volt előbb és nem a tyúk. Ezzel semmiképpen nem tagadnám, hogy a későbbiekben az informatikai lehetőségek ne implikáltak volna olyan bölcsész metodikusokat, amelyekre korábban gondolni sem igen lehetett.

Magára a módszer kifejlődésére aztán már az informatika használatának korszakában került sor, amelyre olyan széles hatású, szakmai körökön is túlmutató könyv lehet a kitűnő példa, mint például John Naisbitt első jelentős műve (*Megatrends: Ten New Directions Transforming Our Lives*, 1982).⁸ A szerző ebben a művében tíz olyan folyamatot és változást (*megatrend*) tárt fel, amelyek nem az adott társadalmak vagy a kormányok akaratából keletkeztek föltétlenül, és nem is a látható felszínen, hanem a mélyben bontakoztak ki, de később mégis mindennél erősebben meghatározták a gazdaság és a társadalom jövőjét. Mai szemmel is lenyűgöző olvasni, hogy az akkor feltárt megatrendek szinte mind meghatározták a 20. század végének változásait. Naisbitt jelentősége tehát a mostani kontextusban az, hogy a széles körű sajtóelemzés jelentőségét ilyen

⁶ Vö. TÓZSA RIGÓ Attila cikke a *Klió*-ban Peter Burke könyve kapcsán. <<http://www.c3.hu/~klio/klio991/klio006.htm>>, illetve az *Annales* lap 1929 és 2001 között online: <<http://www.persee.fr/collection/ahess>>. [Letöltés: 2016. március 1.]

⁷ Magyar kiadása: *A Földközi-tenger és a mediterrán világ II. Fülöp korában*, I–III, Bp., Akadémiai–Osiris, 1996.

⁸ Akkor meglepően friss hazai kiadása az OMIKK-ban készült el, szintén 1982-ben. A könyv igen széles hatása miatt is fontos, mert 57 országban és 14 millió példányban jelent meg.

sikeresen – és tulajdonképpen a mai úgynevezett *big data* módszereket is használva – ő tudatosította sokakban. Ahogy Naisbitt a jövőt szerette volna kiolvasni *adatbányászati módszereivel*, úgy magától értetődő volt, hogyha a múlt minél inkább számítógépen volna, úgy a történeti trendek (vagy a „vékony fekete csíkok”) abból is kiolvashatók.

A módszer, a gondolkodásmód alkalmazása

E rövid dolgozat messze nem alkalmas arra, hogy bemutassa, hogy ma hol és milyen széles spektrumban található fel a „bölcész” területeken a digitális bölcsészetnek nevezett módszer, gondolkodás. Itt csak néhány példával érzékeltetni szeretném a tipikus és új területeket, amelyek már felbukkantak, esetleg bevettnek számítanak, vagy éppen napjainkban kerülnek elénk mint lehetőségek.⁹ A példák nem a konkrét projektek kataszterét, hiszen azok száma még Magyarországon is nagyon nagy, hanem a *lehetséges szakterületek* körét mutatják, mégpedig úgy, hogy a hozzám közelebb álló könyvtárosi és könyvtári szolgáltatási körben mozgunk leginkább. Ezeket a témákat az alapozástól, a kezdetektől a legújabbak felé haladva szeretném megemlíteni, röviden tipizálni:

a) a virtuális kiállítás

A virtuális kiállítások ötlete spontán jelent meg könyvtáros és muzeológus körökben, amikor az internet már multimediális lehetőségeket is kínált. Így viszonylag egyszerűen lehetővé vált a szöveges és képes, hipertexttel egymásba ágyazott oldalak egész rendszerét publikálni a nyilvánosság felé, ráadásul úgy, hogy azok a nap 24 órájában elérhetők, egymással és az egész világháló tartalmával összekapcsolhatók, s nem utolsósorban részben vagy egészben letölthetők, anyagaik felhasználhatók. (Most tekintsünk el a szerzői jogi aspektustól.) Nyilvánvaló, hogy ez a digitális bölcsészet szempontjából anyagközlő, prezentációs, ismeretterjesztő lehetőségekkel egyként rendelkező megoldás volt. Tipikus példája ennek a 2001-ben készült és a Magyar Elektronikus Könyvtárban megjelent „*Arany János: Őszikék*” kiállítás,¹⁰ amely az illető Aranykorpusz feldolgozásához szükséges „kellékeket” gyűjtötte össze, úgymint hasonló kiadás, teljes szövegű művek és kommentárok, tanulmányok, adatközlések, képanyagok stb.

b) szövegtörzsek létrehozása és használata

A másodikként megjelölt téma tulajdonképpen logikailag meg is előzhette volna az előzőt, részben így is lett, ha például a *Magyar Elektronikus Könyvtár* úttörő munkájára gondolunk, amely az internet mulandóságával harcolva a világháló magyar vagy magyar vonatkozású (és értékesebbnek tűnő) szövegeit kezdte begyűjteni még a kilencvenes évek elején. A *Neumann-ház* már nem csupán könyvtárosi szempontból gyűjteményezett, de ki is adott szövegeket, mégpedig minőségileg a tudós társadalom által is respektált módon. Még ezeket is megelőzve, vagy e projექtekkel egy időben (s időnként össze is dolgozva) kezdte birtokba venni az irodalmi, történelmi szövegek, források és kiadványok teljes szövegű közlését az *Arcanum Kft.*, ekkor még persze CD-ROM, majd DVD kiadványok formájában, hogy aztán ők is felköltözzenek a világhálózatra.

Mindez természetesen felvetette és szakmai körökben vitatottá tette, hogy filológiai, sőt textológiai szempontból ezek a könyvtárosi és vállalkozói törekvések hasznosak-e és megnyire?¹¹ Ennek ellenére a szövegtörzsek nőttön nőttek, s a felhasználók száma is, amatőr és professzionális oldalon egyaránt. Mindezen folyamat aztán sokban professzionizálódott is,

⁹ Csak az EBSCO szakirodalmi adatbázisában több mint 7000 (!) könyv található, amelyek címében vagy tárgyszavában megtalálható a kifejezés. Ezen művek közt nagyon sok a konferenciaanyag és tanulmánykötet, amelyekben ezrével találni a legkülönbözőbb területekről alkalmazási példákat.

¹⁰ KOKAS Károly és TÓTH Margit: <<http://www.bibl.u-szeged.hu/oszikek/>>

¹¹ Vö. például KOKAS Károly, *Az elektronikus könyvtáros textológiai gondjai* (1997 Networkshop) <<https://www.niif.hu/rendezvenyek/networkshop/97/tartalom/NWS/3/10/index.htm>>

gondoljunk csak a modern magyar irodalom digitális forrásaira a *Digitális Irodalmi Akadémia* projektben.

Természetesen itt (és az egész problémakör szempontjából) alapvető változást hozott az SGML szövegjelölő „filozófia” megjelenése, s az, hogy az eddigi *HTML megjelenítési szabvány* mellett megjelent a jóval precízebb *XML*, majd a kifejezetten irodalmi szövegek elektronikus publikálására kifejlesztett *TEI szabvány*. Nincs arra most tér, hogy ezt sorban kifejtsük, az érdeklődő rengeteg helyen talál erről információt.¹² Legyen szabad csak arra utalni, hogy ez azt jelentette, hogy a pusztán elektronikus szöveg „öntudatlan” állapotából öntudatra ébredt, vagyis a szöveg nem csupán a megjelenítés (lásd *HTML*), hanem a feldolgozás és keresés szempontjából is kapott egy erőteljes struktúrát, így a tudományos munka szempontjából óriásit nőtt, nőhetett annak értéke.

c) közös keresőrendszerek

Trivialitásnak tűnik, de egyáltalán nem az, hogy az informatikai fejlesztések és a hálózat révén lehetővé vált a szabványos interfészekkel ellátott tartalmak összekapcsolása, együttes keresése, így olyan szövegtörzsek, adatbázisok kerülhettek egy „virtuális” térbe, ami önmagában véve is jelentős hozzá adott értékeket tudott produkálni. A szöveges adatbázisok például *OAI-PMH protokoll*-al való „leiratása” és máshova való betöltése és annak közös indexelése rendkívüli lehetőségeket adott a virtuális térben kutakodóknak, így például a „*Bodza*” projekt keretében történő magyar levéltári és szakirodalmi alkalmazások,¹³ vagy az *Arcanum* hatalmas indexelt állományai.¹⁴

d) képi adatbázisok

A képpel való illusztrálás mellett megjelent a pusztán vagy jelentősen képi információt tartalmazó adatbázisok tömege is, amelyek már nem pusztán kataszterei voltak például a művészeti alkotásoknak, hanem lehetővé tették azt is, hogy azokban descriptorok mentén kifinomultan keresni is lehessen. Ez azt jelentette, hogy a képi adatbázisok metaadatai kiterjedtek a hagyományos formai leíráson túl a képek tartalmára is, lehetővé téve például a motívum, tárgy és személy keresését.¹⁵

e) származtatott valós idejű adatok

Az intelligens szerver oldali megoldások lehetővé tették, együttműködve a fejlettebb *HTML 5.0* szabvánnyal és más grafikai alapú XML szabványokkal, hogy az úgynevezett „big data” adatbázisokból a lekérdező valós idejű származtatott adatot is megkapjon, vagyis összekapcsolja azokat, tulajdonképpen tetszés szerinti kombinációkban, s *facettákkal* megszűrve. Nagyon érdekes megoldásokat mutat ezen a téren például a Monok Istvánhoz kötődő hazai *Eruditio projekt*, amely az adatok vizualizációjában is hazai úttörő volt a 2000-es évek elején a bölcsész területen.

f) adatábrázolások és a földrajzi tér

A nagyon nehezen tanulmányozható és elérhető régi térképek digitalizálása a laikus közönséget is lázba hozta, de a történelmi adatok térképekre, sokszor akár történelmi térképekre vetítve történő ábrázolása szakmai berkekben is elismerést okozott. Hiszen például egy százötven éves, vagy még régebbi részletes térkép rétegátmenetes ábrázolása, keresési szempontokkal össze-

¹² Pl. BÍRÓ Szabolcs, *A szövegfeldolgozás modern eszközei – az SGML és XML nyelvek*, Tudományos és Műszaki Tájékoztató, 2004/10: <http://tmt.omikk.bme.hu/show_news.html?id=3733&issue_id=455>

¹³ Zawiasa Róbert fejlesztői munkái, lásd például: <<http://bodza.bibl.u-szeged.hu>>

¹⁴ <<http://arcanum.hu/hu/>>

¹⁵ Az ezer példából csak egy-egy, a San Franciscó-i szépművészeti kollekción hatalmas kereshető adatbázisa: <<https://art.famsf.org>> vagy a magyar készítésű, s világhírű világképtár és annak hazai képtári párja a hálózaton: <<http://www.wga.hu/index1.html>> és <<http://www.hung-art.hu>>.

kötve mondjuk a mai Google-alapú térképekkel valóságos történeti térképtani felderítő tevékenységnek adott, adhat alapot. Hazai és mégis úttörő példák erre az Arcanum térképtárai.¹⁶

g) korrelációs módszerek és vektorgrafikus ábrázolások

Ez az irány viszonylag újabb, s a legmisztikusabb a bölcsész közönség számára, hiszen itt már a szövegekből, képekből és adatokból olyan új összefüggések merülhetnek fel, amelyek jelentősen új ötleteket adhatnak. Lehetetlen összefoglalni ezeket a törekvéseket, s összegezni pár sorban működési modelljüket, de annyit talán le lehet szögezni, hogy itt lényegében arról van szó, hogy a pusztán statisztikai és visszakeresési eljárásokon jelentősen túlmutatva igen korszerű *matematikai eljárásokat alkalmaznak* a háttérben, amelyek a megfelelő paradigmasorokkal leírt adatok egymáshoz való viszonyát gráfokkal ábrázolják, mégpedig akár térbeli szerkezetben. A legizgalmasabb ebben a metódusban az, hogy új összefüggések jelenhetnek meg: például az MTA SZTAKI egyik kísérleti modelljében a Wikipédia szócikkeinek szerkezetét feldolgozva a *LODmilla projekt* keretében bármely szócikk bármelyik másikkal való összefüggéseit feltárhatjuk, vizualizálhatjuk. Természetesen ez még önmagában nem tudományos eredmény, hanem egy felkiáltójel lehet abban a vonatkozásban, hogy az adott összefüggés az valós, érdekes és kutató-e?¹⁷

h) virtualizációs módszerek

A legújabb terület divatos beszédtema, s a digitális bölcsészet területén mint kutatási módszer ingoványos is, ugyanakkor a közönség számára határtalan lehetőségeket rejt. Van ennek egy része, ami például *3D modellalkotás*, s amelynek segítségével csak történetileg létező terek teremthetők újra a virtuális térben, lehetővé téve azt, hogy a bevitt adatok tömegének vizualizálásával látványos élményt, de kutatási felületet is hozunk létre. Hiszen a kutató számára átélhetővé, „láthatóvá” válik például a római Fórum császárkori rekonstrukciója úgy, hogy az térben bejárható, a valóságos viszonyok plauzibilisen érzékelhetők. Ez még akkor is inspiráló lehet, ha tulajdonképpen egyetlen új adatot sem tartalmaz, de maga a virtualizáció új összefüggéseket sugallhat.¹⁸ Természetesen absztrakt „terek”, például információs galaxisok is létrehozhatók, ahol az információk sűrűsödési pontjai, vagy egymáshoz való térbeli elhelyezkedésük, méretük és távolságuk önmagában összefüggéseket rajzol fel a kutató előtt a virtuális térben.

i) big data és internetarchiválás itthon

A szakirodalomban a „Big Data” a cégek, a tudományos intézmények és eszközeik, az intelligens hálózatok, a magánszektor és az egyéni felhasználók által világszerte és napi szinten előállított óriási adatmennyiséget jelentik. Az igény pedig úgy fogalmazódik meg, hogy strukturáltan tárolva és rendezve, aztán kielemezve ez a rengeteg információ könnyen egészen új eredményekhez vezet, a következtetések általa meghatározottabbak lesznek. Ennek összefüggései a fentiekkel és a digitális bölcsészettel azonnal átláthatók, ha a digitalizált, kereshetővé tett, állandóan elérhető és soha nem látott mennyiségű bölcsész jellegű adattömegre gondolunk.

Persze az, hogy a történelemtudomány, irodalomtudomány stb. számára nagyon hasznos, ha a múltból nagy tömegű adatot tud gyorsan és sokféleképpen kielemezni, szinte magától értetődő. De azt se felejtsük el, hogy a jelenünk minden pillanatban múlttá válik. *Kosáry Domokos* így fogalmazott, sok évvel az internet hazai megjelenése előtt: „A történelem totális [...] mindazt, ami a társadalom és az egész emberiség életében ma végbemegy, holnapra vagy már ma estére mindenestül, minden részével és részletével együtt elnyeli a nagy Moloch: a történelem.”¹⁹ Félő, hogyha nem gondoskodunk ennek (az egész keletkező és részben eltűnő kultúrának) egyik meg-

¹⁶ <<http://mapire.eu/hu/>>

¹⁷ <<http://lodmilla.sztaki.hu/lodmilla/>>

¹⁸ Hasonlóra volt nagyon korai hazai kísérlet 2001-ben az Eruditio program keretében a ma már meghaladott VRML technikával elkészített Zrínyi-könyvtár térbeli rekonstrukciója, ahol az elképzelt könyvtárhelyiségben az eredeti sorrendben váltak láthatóvá Zrínyi Miklós 17. század közepi könyvespolcai, lásd: <<http://www.eruditio.hu/zrinyi3d/>>.

¹⁹ KOSÁRY Domokos, *Művelődés a XVIII. századi Magyarországon*, Bp., Akadémiai, 1983, 14.

határozó adatmennyiségéről, magáról az internetről, akkor a jövő bölcsészeinek is nagy hiányérzetei lesznek. A hazai internet, a mi internetünk jelenleg nem archiválódik sehol. Lassan könnyebb lesz egy 20. század elei mikrotörténeti mozzanat rekonstrukciója, mint azt kinyomozni, hogy például 15 évvel ezelőtt mi volt a Nemzeti Színház honlapján, ami akkor óriási sajtó- és médiavisszhangot keltett. Bizony, a jelenlegi *magyar internetet archiválni kellene* (még hozzá szemantikusan, vagyis felcímkézve mindenféle automatikusan/félautomatikusan előállítható metaadattal). Ennek metódusának kidolgozása és a munka irányítása biztosan közös digitális bölcsészeti és közgyűjteményi feladat lenne.²⁰

A digitális bölcsészet „Szent Gráljai”

Sokféle módon lehet meghatározni, hogy a digitális bölcsészet rövid történetének vágyálmai, kutatási *leading edge* projektjei éppen miben álltak. Az mindenesetre látszódik, hogy vannak olyan sarokkövek, amelyekre építve hatalmas lehetőségek bontakoztak ki.

A fenti példalózást más aspektusból nézve láthatjuk, hogy a *digitalizálás* magában óriási lehetőségekhez juttatott minket. Aztán maga a *hálózatba kerülés*, az, hogy amit digitalizáltunk vagy digitalizáltak, az folyamatosan és bárhol elérhető.²¹ A digitalizálás egyik külön aspektusa volt, és meghatározó eredmény, a nyomtatott szövegek analógból digitálissá való visszaalakítása, az *optikai karakterfelismerés (OCR)* révén. Aztán újabb fordulópont, az említett szövegkódolási iniciatívák, az SGML nyomában: az XML és TEI lehetőségei. A következő állomás a szintén vázolt „élek és csúcsok” korszak, a *gráfok, korrelációk* és ábrázolásaik megjelenése. Ez már a ma problémája. Mi hát a digitális bölcsészet legújabb Szent Grálja? Ami ma még talán hihetetlennek tűnik, de mégis lehet, hogy a küszöbön áll? Véleményem szerint az, amikor a számítógép nem csupán olvasni tanul meg, hanem megtanítjuk a szöveget is megértetni vele. Mindezt úgy foglalthatnánk össze, hogy az *AI- (mesterséges intelligencia) alapú szövegértés*, ami magában foglalja a szövegek klaszterálását és tárgyszavazását is. Ez a digitalizálás révén létrejött óriási és vegyes szöveg-galaxis értelmezését és feltárását tenné lehetővé, hiszen emberi munkával ez ebben a méretben aligha képzelhető el. Úgy is mondhatnánk, azt, amit a digitalizálási technikák segítségével könnyedén elő tudunk állítani, akár sok millió oldalas szövegvilágokat, az is egy jelentős feladat volt, de azokat metainformációkkal ellátni, s így jelentős hozzáadott értéket termelni, a digitalizálás tempójában biztosan nem tudunk.

Ebben az aspektusban is felmerül a digitális bölcsésszel partner szakmák és intézményi rendszerek szerepe. Azt gondolom, ahogy eddig is, az említett „nyersanyagok” feldolgozásában és szolgáltatásában a digitális könyvtáraknak és múzeumoknak, valamint levéltáraknak (amelyek nem új dolgok, az eddigiek hálózati és digitális másai) igen jelentős tennivalói vannak és lesznek. Úgy tűnik, a digitális bölcsész még a korábbiaknál is erősebb partnerkapcsolatba kerül az öt kiszolgáló, anyaggal és információval ellátó könyvtárossal, levéltárossal és muzeológussal. Arról nem beszélve, hogy talán megnyílik a lehetőség arra, hogy például *a könyvtárak három-ezer év után újradefiniálhatják magukat*, alapvető feladataikat.²²

Diszciplína, módszer vagy szemléletváltás?

A dolgozat elején azt a kérdést tettem fel, vajon a digitális bölcsészet hogyan írható le leginkább? Reményeim szerint a dolgozat elárulja azt is, összességében mit érdemes minderről vélekedni ma.

²⁰ A hazai internetarchiválásnak egyik ötletadója, Drótos László és sokan mások is leírták, leírtuk, hogy mekkora hátrány már ma is, hogy nincs hazai internetarchiválás. Ezzel kapcsolatban lásd például: DRÓTOS László, *Mi a MIA? – Javaslat egy Magyar Internet Archivum létrehozására*, Tudományos és Műszaki Tájékoztatás, 2006/6: <http://tmt.omikk.bme.hu/show_news.html?id=4431&issue_id=473>.

²¹ Ennek könyvtári aspektusairól, de az egész digitális bölcsészetre is kiterjeszhető értelmezéséről lásd: SENNYEY Pongrácz, KOKAS Károly, *Könyvtárak a hálózatban*, Tudományos és Műszaki Tájékoztatás, 2011/10: <http://tmt.omikk.bme.hu/show_news.html?id=5548&issue_id=531>.

²² KOKAS Károly, *Könyvtárak a Rubiconnál: A tudományos szakirodalmi információellátás új könyvtári paradigmái*, Educatio, 2013/ 3: <<http://folyoiratok.ofi.hu/educatio/konyvtarak-a-rubiconnal>>.

A bölcsészettudományok művelésének háttérében ezred évekig az *emlékezet és az eruditio* állt. A „nagy tudós” egyik legfontosabb attribútuma volt mindig is az anyagismeret, a széleskörű olvasottság, amely lehetővé tette, hogy az ötletek, kérdések megfogalmazódjanak. Nagy a kísértés, hogy a digitális bölcsész mindezt „tárolt memóriával” és visszakereső rendszerrel pótolja. Tévedés, nem a pótlásra, s a régi típusú felkészültség elvetésére van szükség, hanem annak jelentős *megettoldására*. De az, aki nem rendelkezik a kellő szakmai előképzettséggel, szakirodalmi eruditióval és forrásismerettel, nem fogja tudni sem a jó kérdéseket feltenni, sem a kapott válaszokat helyesen értelmezni. S pláne nem tud digitális bölcsészeti háttéradatbázisokat tervezni, létrehozni és azokat hatékonyan használni.

A modern természet- és műszaki tudományok a 20. században civilizációnk fejlődésének/fejlesztésének meghatározó eszközei lettek. Valószínűleg a 19. és a 20. század fordulója körül történt meg velük a váltás, amikor igazán azokká lehettek: biztos, hogy ebben nagy szerepet játszott az is, hogy a gondolati konstrukciókon túl az adatok gyűjtése és teljes körű feldolgozása erre alkalmasabbá tette őket. Az úgynevezett kemény társadalomtudományok már jócskán megindultak ezen az úton, ezt persze jellegük (közgazdaság, szociológia stb.) is indokolta. Nem mellesleg az itt használt csoportképző nevet („kemény”) is ezért kapták. A „lágý” társadalomtudományokkal kapcsolatban az volt a vélekedés, hogy a művészethez állnak közelebb (*arts and humanities*), illetve hogy nem kvantitatív a megközelítésük. Ha ez lényegében nem is változik meg, az jól látható, hogy a bölcsészettudomány – az informatika segítségével – most megy át ezen a változáson, aminek következtében például a múlt kérdéseivel kapcsolatban (például társadalom-, eszme- és gazdaságtörténeti területek makró szintjén) objektívebb, tényekkel sokkal jobban alátámasztott válaszaink szülehetnek.²³

Hangsúlyozni szeretném azonban, hogy az *alapszinten* nem föltétlen és elsődleges tudományos eredményekről van szó, amelyek a módszerek nyomán csak úgy adódnak. Amelyeket a számítógép felfedez vagy elénk tár. Régen is így volt ez. Gondoljunk arra, hogy egy monumentális oklevélkiadás, sok ezer regesztával vagy egy áhított és alapos kritikai kiadás mennyivel tette kényelmesebbé és innovatívabbá a kutatást. Az alapok tekintetében itt sincs szó többről: a számítógép és informatika nyújtotta lehetőségek nyomán új és mélyebb ismeretek szerezhetők. De természetesen az, aki ezeket használja, aki a következtetéseket levonja, az továbbra is a bölcsész kutató. A digitális bölcsészet ehhez rengeteg és a régmúltnál sokkal kényelmesebben elérhető matériát kínál, új keresési és megjelenítési lehetőségeket, amelyek új vagy biztosabban megalapozott eredményekhez vezetnek.

S ne feledkezzünk meg arról sem, azért a kapu másfelé is nyitva van. Innen *van még tovább is*. A digitális bölcsészeti módszertan az egyes tudományok esetében olyan kérdésfeltevésekhez is elvezethet, amelyeket eddig nem is tehattünk fel. Könnyen lehetséges, hogy új összefüggések merülnek fel a hatalmas adattengerből. S az sem elképzelhetetlen, hogy maga a kutatási segédlet, módszer vezet megkérdőjelezhetetlen eredményhez, ott, ahol ez korábban elképzelhetetlen volt.²⁴ Elő fog fordulni velünk, hogy *a vékony fekete csík*, lerakódva láthatóvá lesz és előbukkan az eddig feldolgozatlan/feldolgozhatatlan adattömegből.

De ezt meglátni, és még előtte az adatokat összegyűjteni, strukturálni, s nem utolsósorban a metódusokat kitalálni, mindehhez új típusú ismeretek bevonására van szükség. Mivel a módszertan egyre komplexebb és nem fejlődhet maguk nélkül a bölcsészek nélkül a kívánatos irányba, a „diszciplínát” magát is kutatni és fejleszteni, és persze tanítani is kívánatos.

²³ Ezt a gondolatmenetet Drótos László barátom vetette fel nekem. Köszönöm.

²⁴ Hogy ez nem merő fantáziálás, mutatják például az olyan, most itt csak éppen említhető vizsgálatok, amelyek például részletes – számítógép számára feldolgozható – paradigmákkal ellátott ismeretlen vagy vitatott szépirodalmi szövegekről „döntötték el”, hogy azok ma még ismeretlen szerző művei vagy egyértelműen valamely ismert szerzőhöz tartoznak. Vagy gondoljunk bele az ókori kétoldalas papirusztöredékek azonosításába, ahol a számítógépes alkalmazás, például az összes ókori szöveg ismeretében, egy-egy verzón-rektón meglévő szó alapján a töredéket egyértelműen egy bizonyos mű részének állítja. Nyilván ez újra felveti a tudomány mint emberi emlékezet, illetve olvasottság – dolgozatban említett – problematikáját is.