

Szemle

Kiss Margit¹

A DIGITÁLIS BÖLCSÉSZET MA

– Edited by FOREST, Dominic, Diane JAKACKI, Cecily RAYNOR, Michael Eberle SINATRA and Stéfan SINCLAIR. *Special Issue: 'Digital Humanities 2017: Bridging Digital Humanities'* (Digital Scholarship in the Humanities, Volume 34, Issue Supplement_1, December, 2019, 206 lap) –

Számos kiváló összegző munkából válogathatunk, ha a digitális bölcsészet mibenlétéről, a kezdeteiről, a jelenlegi helyzetéről vagy a lehetséges távlatokról szeretnénk tájékozódni.² Ha ezekre a kérdésekre a tanulmánykötetek célirányossága, explicit jellege és didaktikussága helyett a legújabb nemzetközi kutatási eredmények megismerése révén keressük a lehetséges válaszokat, akkor a *Digital Scholarship in the Humanities*³ című folyóirat 2019-es, tizenhét tanulmányból álló különszáma⁴ erre kiváló lehetőséget teremt. Az *Alliance of Digital Humanities Organizations* (ADHO)⁵ szakmai szervezet 1990 óta minden évben nemzetközi tanácskozást szervez az újonnan elért eredményekről. A 2017 júliusában zajló esemény mottója az „Access / Accès” volt, amely a nyílt hozzáférés (*open access*) kérdése köré szervezte a terület jeles kutatóit. A konferencia anyagából készített különszám egyfelől hordozza a tanulmánykötetek gondosan megkomponált s egyben átfogó képet adó jellegét, másfelől a tanácskozás anyagára támaszkodva a legfrissebb kutatási eredményekről ad számot.

¹ A szerző a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézete XIX. századi Osztályának tudományos főmunkatársa.

² Susan SCHREIBMAN, Ray SIEMENS and John UNSWORTH, eds., *A New Companion to Digital Humanities* (New York: Blackwell Companions to Literature and Culture, John Wiley & Sons Inc., 2016); Melissa TERRAS, Julianne NYHAN and Edward VANHOUTTE, eds., *Defining Digital Humanities: A Reader* (London: Routledge, 2013); David M. BERRY and Anders FAGERJORD, eds., *Digital Humanities: Knowledge and Critique in a Digital Age* (Cambridge, Malden: Wiley, 2017); Ted UNDERWOOD, *Distant Horizons* (Chicago, London: The University of Chicago Press, 2019).

³ <https://academic.oup.com/dsh>.

⁴ Dominic FOREST, Diane JAKACKI, Cecily RAYNOR, Michael Eberle SINATRA and Stéfan SINCLAIR, eds., *Special Issue: 'Digital Humanities 2017: Bridging Digital Humanities'* (Digital Scholarship in the Humanities, Volume 34, Issue Supplement_1, December, 2019).

⁵ <https://adho.org/>.

A digitális bölcsészet fejlődését, alakulását bemutató nagy ívű összefoglalások, válogatott szempontok szerint felfűzött narráció helyett a *The Early History of Digital Humanities: An Analysis of Computers and the Humanities (1966–2004)* and *Literary and Linguistic Computing (1986–2004)* című tanulmány szerzői a tudománytörténet feltárását is a számítógépes adatelemzés segítségével végezték el. Rangos digitális bölcsészeti folyóiratok tanulmányait elemezték informatikai eszközökkel, hogy választ kapjanak azokra a kérdésekre, milyen irányban változott e terület a publikált kutatások tükrében, s az eltér-e az eddig ismert tudománytörténeti narratívától. Rámutatnak, hogy a digitális bölcsészet „szokásos” történetírása kevésbé foglalkozik azokkal a területekkel, amelyek kívül esnek a szövegelemzésen, és ezáltal elfedi más tudományterületek érintettségét, így kimaradnak például a történet-tudomány azon korai módszertani eredményei is, amelyek statisztikai módszereket alkalmaztak olyan ismert történelmi kérdésekre, mint az amerikai polgárháború okainak a feltárása. A tanulmány szerzői szerint a számítógép és a szöveg házasítása a korai időszakban ugyan meghatározó volt, de a kezdetekben a hang és a technológiára való reflektáltság is megjelent kutatási témaként, amelyek rendszerint ugyancsak kimaradnak az áttekintésekből. Rávilágítanak, hogy a digitális bölcsészetre ma is jellemző szövegközpontú szemlélet mellett kevés szó esik a statisztikákkal, a térinformatikai rendszerekkel, a közösségi hálózatok elemzésével, a vizualizációval és a modellezéssel kapcsolatos kérdésekről, s kevés olyan munka születik, amely szisztematikusan tárná fel a diszciplína történetét a régészetre, művészettörténetre, történelemre, zenére, filmművészetre, filozófiára és vallásra vonatkozóan.⁶ A sokféleség irányába hatna, hogyha a digitális bölcsészeti kutatásokban az angol nyelv, az angol nyelvű irodalom dominanciája mellett más nyelvek és más nyelvű szövegek feldolgozásával kapcsolatos eredmények is bekerülnének a szakmai diskurzus főáramába. A módszertan és a technológia fejlődése révén a szövegre való koncentrálttság mellett ma már a digitális bölcsészet „vizuális fordulatával” kell számolni: a szövegen kívüli adattípusokkal, az informatikai támogatású képfeldolgozással komoly eredményeket értek el a kutatók.

Erre az elemzésre reflektál a különszám megkomponáltsága is: a szövegközpontúság dominanciáját megtöri a vizualitás, a hang, s helyet kap a sajtóelemzés, a történeti térinformatika, az adatbázis-struktúra. Megtudhatjuk belőle azt is, hogy a *big data* korában a különböző elemzőeljárások együttes, illetve egymásra épülő alkalmazása vezethet megfelelő eredményre, s hogy a szövegelemzések terén a szemantika gépi támogatású vizsgálata felé tolódik a hangsúly.

A képek és a szövegek kódrendszeireiben megmutatkozó alapvető különbségek ellenére a kinyert metaadatok elemezhetősége megeremti a szövegek és a képek számítógépes feldolgozása közötti átjárhatóságot. A *Distant Viewing: Analyzing Large Visual Corpora* című tanulmány szerzői a *distant reading* mintájára a nagy

⁶ Kivétel ez alól: Susan SCHREIBMAN, Ray SIEMENS and John UNSWORTH, eds., *A Companion to Digital Humanities* (Oxford: Blackwell Publishing, 2004).

terjedelmű vizuális korpuszok elemzésére alkalmas *distant viewing* módszertani és elméleti keretét vázolják fel. Az elméleti keretrendszer kialakítása mellett kísérletekben két amerikai, több száz órányi filmsorozat összehasonlító, gépi támogatású elemzésével rávilágítanak az eddigi kutatásokkal szembeni különbségekre a korabeli női szerepek ábrázolásával kapcsolatban. Másik elemzésükben 170 000 darab 1935 és 1943 között készült fénykép stílusjegyeit vizsgálták, amelynek alapján a fotósok stílusa közötti hasonlóságok és különbségek elemezhetők: így például a messter és tanítványa közötti hatás lenyomata, de kirajzolódott e csoportban a női fotóművészek (azon belül két fotós) központi szerepe is. A szövegek stilometriai vizsgálatában a statisztikai módszerek eredményes alkalmazásához a gondosan megválasztott és előkészített szövegtörzs alapvető fontosságú. A festmények hasonló módszerrel alapuló automatikus elemzésében ezek az előkészítő folyamatok még szofisztikáltabbak, amelyek révén meghatározhatóvá válik például a műalkotás alkotója, keletkezési ideje vagy akár a más munkákkal való stílusbeli hasonlósága. Az *Automatic Detection and Visualization of Garment Color in Western Portrait Paintings* című tanulmányban végigkövethetjük ezt a rendkívül összetett folyamatot, amelynek révén automatizált elemzéssel felderíthető és kimutatható, hogy a nyugati kultúrában a festményeken a férfiak és a nők ruházatának színei hogyan változtak az idő múlásával. A láttatás mellett a látás elemzése ugyancsak izgalmas kutatási terület, hiszen az emberi szemmozgás rögzítésének és mérésének a folyamata már technológiailag kivitelezhető, bár a bölcsészeti kutatásokban még nem mondható általános gyakorlatnak. Az ilyen jellegű vizsgálatok a vizuális megjelenítés és a befogadási folyamatok megismerésében és megértésében játszhatnak szerepet. Az *Aesthetic Appreciation and Spanish Art: Insights from Eye-Tracking* című tanulmány szerzőinek a szemkövetési technikák elemzésével végzett vizsgálataik jelentős mértékben hozzájárulnak ahhoz, hogy megértsük, a közönség hogyan fogadja be, hogyan látja a műalkotásokat, s mindebből a gyűjtemények összeállítása során a múzeumi gyakorlat is sokat meríthet.

A digitális bölcsészeti kutatások egyre összetettebbé válnak mind a feldolgozandó forrás, mind a feldolgozási módszerek tekintetében, ennél fogva központi szerepet kapnak a digitális infrastruktúrák tervezésének alapelvei a lekérdezhető adatszerkezetek kezelésének és feldolgozásának a szempontjából. A *Database Thinking and Deep Description: Designing a Digital Archive of the National Synchrotron Light Source* című tanulmány szerzői azt a kérdést járják körül, hogy vajon a történeti folyamatok kutatásához készített digitális archívumok mennyire determinálják, ily módon akár korlátozzák is a belőlük kiolvasható kutatási eredményeket. Egy webes platform és a felhasználói felület architektúrája szükségszerűen irányít, előre meghatároz, ezzel együtt ugyanakkor láthatatlanná tesz kutatási kérdéseket, problémákat. Ezért egy olyan adatbázis-szerkezet modelljét vázolják fel a szerzők (a rugalmas és kiterjedt osztályozás, az átláthatóság és a megváltoztathatóság pillérére támaszkodva), amely felépítése révén nem előzetesen határoz meg feltehető kérdéseket, hanem lehetővé teszi a saját kutatási irányok kialakítását. Egy olyan „rendezetlen”

struktúrát javasolnak, amelyben a mintakeresés, az okok, a következmények keresése nem eleve kódolt, hanem a kutató önállóságán és kompetenciáján alapul. A cél az, hogy a keretek olyan erősek legyenek, hogy lehetővé tegyék a feltárható történetek elmondását, de elég gyengék ahhoz, hogy mindig újak formálódhassanak. A több forrás, adat bevonása kevesebb korlátozó funkcióval, nagyobb kutatói önállóságot és gazdagabb elemzési lehetőséget teremt. Ebből arra kell következtetnünk, hogy a korszerű kutatások már nemcsak a digitális formában létrehozott források felhalmozását célozzák meg, hanem azt is feltételezik, hogy a kutató megfelelő módszertani kompetenciával rendelkezik ahhoz, hogy önálló adatbányászati, -elemzési technikákkal legyen képes eredményeket létrehozni. A nagy adatmennyiség kezelése nemcsak az elemzésnél, hanem a digitális bölcsészeti projektek lebonyolítása során is feladat. A *Using Programmed Diagnostics to Learn About, Monitor, and Successfully Complete Your DH Project* című tanulmányból egy olyan diagnosztikai eljárást ismerhetünk meg, amely azoknál a számítógépes munkafolyamatoknál bizonyul hasznos módszernek, amelyek során a nagyszámú adatok kézi ellenőrzése már nem kivitelezhető. Olyan ellenőrzési sémákról van szó, mint például a különféle duplikátumok, hibásan kódolt adatok, hivatkozások kiszűrése, de monitorozhatók a munkafolyamatok, s kiválogathatók a további ellenőrzést igénylő tételek. Az automatizált diagnosztikai eljárás jelentősen segítheti például a digitális szövegtárolással járó feladatok optimális előrehaladását.

Az elektronikus szövegtárolások készítése ma már alig képzelhető el valamilyen jelölőnyelv alkalmazása nélkül. Ebben a tárgykörben a TEI vagy nem TEI kérdése hegemoniája révén jó ideje megosztja a kutatókat. Vannak, akik zászlóshajóként tekintenek a TEI-re,⁷ mások inkább a rajta kívüli lehetőségekkel kecsegtetnek.⁸ Az ambivalencia ihlette az *A World of Difference: Myths and Misconceptions about the TEI* című tanulmány szerzőjét, aki a TEI-vel kapcsolatos általános tévképzetektől táplálkozó mítoszokat veszi górcső alá, s egyben konkrét megoldásokat ad a felmerülő problémákra. Ilyenek például: a TEI egyenlő az XML-lel *versus* az XML már halott; a TEI túl nagy és bonyolult *versus* a TEI túl egyszerű és általános; a TEI-ben nem lehet változtatni *versus* TEI-gurunak kell lenni ahhoz, hogy testre szabható legyen; az interoperabilitás nem megoldható; csak digitális szövegtárolásokhoz alkalmazható; csak nyugat-európai nyelveken íródott, leginkább angol szövegekhez használható stb. A TEI-iránymutatások és a szövegtárolással kapcsolatos aggodalmak általában nem a formátumot érintik, hanem inkább a jelölendő információk gazdag részletességéről és a megfelelő technológiáknak az egyes feladatokhoz való

⁷ Georg VOGELER, ed., *What is text, really? TEI and beyond* (Graz: Centre for Information-Modelling Austrian Centre for Digital Humanities, 2019).

⁸ Peter ROBINSON, „Creating and Implementing an Ontology of Documents and Texts”, in Jonathan Girón PALAU and Isabel Galina RUSSEL, eds., *Digital Humanities 2018: Puentes-Bridges* (Mexico City: El Colegio de México, Universidad Nacional Autónoma de México, Red de Humanidades Digitales, 2018), 266–268.

felhasználásáról szólnak. A tanulmány ugyan nem hoz áttörő eredményt a szövegkódolás terén, mégis a kutatóközösség nagy részét érintő problémákra ad konkrét megoldási javaslatokat.

Az adatvizualizáció az elemzések értelmezésében kulcsfontosságú, mondhatni természetes velejáró. Ennek a magától értetődő jellegnek a természetrajzát elemzik az *In Defense of Sandcastles: Research Thinking Through Visualization in Digital Humanities* című tanulmány szerzői. Az adatvizualizáció mind a mennyiségre, mind a tartalomra vonatkozó információt mutat be egyszerre, amellyel az elemző mintázatok, trendeket, állandóságot vagy variációt lát: észrevehetővé teszi azt, ami egyébként észrevehetetlen lenne. Maga a vizualizáció esztétikája – ahogy nevezik – segít abban, hogy a vizualizáció hogyan emelheti ki, és teheti fókuszba az adatok kategorizálásának különböző módjait. A homokvár-metaforával a szerzők kiemelik az átmeneti, instabil, gyakran befejezetlen jeleget, mivel a vizualizáció felvet olyan további problémákat, amelyekből kiindulva újabb vizualizációk készíthetők, s aztán megint további kérdések adódnak. A homokvár-metaforán alapuló elméleti keretrendszer a részletekbe menő kidolgozottsága miatt széttartónak hat, amelyhez mértén a kísérlet tárgyalása részletkérdésnek tűnik, ennek ellenére ez utóbbi a lényegi mondanivalót mégis szemléletesebben megvilágítja. Az 1840 és 1990 közötti időszak több mint 10 000 tudományos-fantasztikus írás nyelvi és metaadatait elemezték a szerzők, hogy a műfaji változásokat feltárják. A kapott eredmények egyfelől azt mutatták, hogy a természetfeletti témák sokkal jobban befolyásolták a sci-fi alakulását, mint az a korábbi kutatásokból kiviláglik, de ezzel együtt az adatvizualizációk elemzése során jóval több, további kérdés vetődött fel, mint ami szorosan a műfaji változásokra vonatkozott volna.

A számítógépes szövegelemzések terén az utóbbi időkben a *word embedding* (szóbeágyazás) számos látványos eredményt hozott a bölcsészettudományokhoz köthetően is.⁹ A *word embedding* a természetesnyelv-feldolgozás egyik alapvető eszköze, amely egyre elterjedtebbé válik, mivel a gépi elemzés már a szemantikai tartalmak feldolgozásában nyújt támogatást. Ezt a módszert alkalmazták a *Text Analysis of Government Reports on the Irish Industrial School System with Word Embedding* című tanulmány szerzői az úgynevezett Ryan-jelentés vizsgálatára. A jelentés az 1936–1990 közötti katolikus vallási gyülekezetek által vezetett ír ipari iskolákban történő gyermekekkel szembeni visszaélések jegyzőkönyvszerű leírását tartalmaz-

⁹ Néhány példát említve: Siobhan GRAYSON, Maria MULVANY, Karen WADE, Gerardine MEANEY and Derek GREENE, „Novel2Vec: Characterising 19th Century Fiction via Word Embeddings”, in Derek GREENE, Brian MAC NAMEE and Robert ROSS, eds., *Artificial Intelligence and Cognitive Science 2016* (Dublin: University College Dublin, Dublin Institute of Technology, 2016), 68–79; Nina TAHMASEBI, „A Study on Word2Vec on a Historical Swedish Newspaper Corpus”, in Eetu MÄKELÄ, Mikko TOLONEN and Jouni TUOMINEN, eds., *CEUR Workshop Proceedings. Vol. 2084: Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference, Helsinki Finland, March 7–9, 2018*. (Helsinki: University of Helsinki, 2018), 1–12, hozzáférés: 2020.07.10, <http://ceur-ws.org/Vol-2084/paper2.pdf>.

za. Sajátos írásmódja révén több mindent elfed, mint amennyit feltár: az egyéni felelősség hangsúlyozása mellett elhomályosulnak a rendszerszintű feltételek, amelyek lehetővé tették a visszaélések kialakulását, így inkább mélyreható eseti tanulmányt tükröz az írásmód. A cikk megmutatja, hogy a szövegelemzés miként használható fel olyan új mintázatok, összefüggések felvázolására, amelyek a *close reading* során nem voltak felismerhetők. A szisztematikus elemzés érdekében a *word embedding* eljárást a kutatók kiegészítették más szövegosztályozó technikákkal, a tartalomelemzést egy domain-specifikus kulcsszókészlet létrehozásával. A tanúvallomások gépi támogatású elemzéséhez az élőbeszéd jellegű szövegek automatikus kiszűrését végezték el. A szemantikai tartalmak feldolgozásához a WordNet felhasználásával szinonimacsoportokat hoztak létre. A *word embedding*-gel a szövegtörzset fogalomkészletet elemezték. A Ryan-jelentés gépi támogatású tartalomelemzése lehetővé tette a meglévő szerkezet felbontását, a releváns megállapítások feltárását és új módon történő olvasását. Módszertani szempontból a tanulmány legfőbb értéke annak a szemléltetése, hogy a *big data* korában a többféle, együttesen alkalmazott automatizált elemzőeljárások milyen hatékony módon működnek együtt az elemzések eredményességének érdekében.

A szemantikai tartalmak gépi feldolgozása kerül más kutatások középpontjába is. Az *An Ontology and a Memory Island to Give Access to Digital Literature Works* című tanulmány szerzői a digitális irodalom fogalmát és az idetartozó művek ismerőit tekintik át. Kutatásukban a szemantikus web egyik eszköze, az ontológia a digitális irodalom műfaji kereteinek és kritériumainak a kialakításában nyújt informatikai támogatást.

Többféle lehetséges magyarázat született már arra vonatkozóan, hogy Pablo Picasso miért szentelt jelentős időt 1935 körül az írásnak. A *Semantic Domains in Picasso's Poetry* című tanulmány szerzői ugyancsak ezt az időszakot vizsgálják abból a szempontból, hogy mi a kép és a szöveg viszonya ezekben a művekben (ahol a kép szavakból áll, s az egyes szavak képként hatnak); továbbá hogy a francia és spanyol nyelvű versekben használt fogalmak milyen szemantikai kategóriák köré rendeződnek, s ezek a kategóriák hogyan viszonyulnak egymáshoz. A különféle taxonómiák és a WordNet alkalmazásával végzett számítógépes szemantikai elemzés lehetővé tette, hogy eltéréseket és összefüggéseket találjanak bizonyos lexikális kifejezésekkel jelölt konkrét témák (pl. a 'megöl' fogalom lexikális kifejezéseinek az aránytalansága a francia és spanyol versekben) és a különböző nyelveken írt versek közti fogalmi kapcsolatokra vonatkozóan. Megfigyelték bizonyos témák nyelvspecifikus jellegét is, például a francia nyelvű művekben a politika, a vallás, a szexualitás témája dominánsabb a spanyol nyelvű költeményekkel szemben. Megközelítésük segít választ adni azokra a kérdésekre, hogy mi állhat Picasso költészetének spanyol, illetve francia nyelvi jellege mögött, s hogy a feltárt tartalmi kategóriák miként alkotnak egy komplex s összefüggő szemantikai hálózatot.

A jelentéshez, értelmezéshez kapcsolódó gépi szövegelemzések terén a korpuszalapú metaforakutatás új, izgalmas terület. A *'Wind of Change'—Shaping Public*

Opinion of the Arab Spring Using Metaphors című tanulmányban az *arab tavasz* mint metafora közvéleményt formáló szerepét vizsgálják a szerzők német folyóiratokban. A metaforakutatás elméleti háttérét és a szövegbányászati módszereket a kognitív nyelvészettel ötvöző módszertant is részletesen bemutató tanulmány kiterít a kézzel gyűjtött metaforák jellegére, a lexikai adatbázis, a szövegbányászati módszerek alkalmazására; a további metaforák gépi kinyerésére, csoportosításukra és részletes szemantikai feldolgozásukra.

Hogy a digitális bölcsészet ma már vagy ma sem csak elektronikus szövegkiadás, nyelvfeldolgozás, stilometria, szövegbányászat, hanem ennél jóval gazdagabb, azt nemcsak a részletesebben tárgyalt tanulmányok, hanem a különszám további írásai is tükrözik. Ezekben az írásokban szó esik a kulturális örökség megőrzéséről (*Remembrance of Contemporary Events: On Setting Up the Sunflower Movement Archive*); a digitális rendszerek projektalapúságának, így élettartamának a következményeiről (*Shelf Life: Identifying the Abandonment of Online Digital Humanities Projects*); a tudományos, nyílt hozzáférésű elektronikus publikációk kérdéséről (*Informing Library-Based Digital Publishing: Selected Findings from a Survey of Scholars' Needs in a Contemporary Publishing Environment*); a 16. és 17. századi tájak térinformatikai feldolgozásának lehetőségeiről (*A 3D GIS and 'The Royaltie of Sight': Recreating 'Prospects' and 'Perspectives' within an English Designed Landscape c. 1550–1660*); valamint a kopt dialektusok OCR-feldolgozásának eredményeiről (*Optical Character Recognition of Typeset Coptic Text with Neural Networks*). Tágabb értelemben a sokfélesége mellett a korábbi bölcsészeti módszerekkel való elégedetlenség, a mindennapjaink részét képező egyre komolyabb technológiai fejlettség, az ebből adódó elképzelések és az erre alapuló kutatói kíváncsiság jellemzi ma a digitális bölcsészetet. S hogy évtizedek múlva majd azonos lesz-e magával a bölcsészettel, vagy önálló diszciplínává válik, esetleg a bölcsészeti kutatások informatikai támogatása lesz a szerepe, vagy akár teljesen elhalványul a jelentősége, az nagyban függ attól, hogy milyen eredményeket érnek el segítségével a kutatók, s az milyen mértékben járul hozzá a humán tudományok műveléséhez. Ha az első jövőkép túlzásnak is tűnik, a *Digital Scholarship in the Humanities* különszámát tekintve borúlátásra nincs okunk.