

Számítógép az irodalomtudományban*

BEVEZETŐ

Kiss Margit

PhD, tudományos munkatárs,
MTA Bölcsészettudományi Kutatóközpont
Irodalomtudományi Intézete

Mészáros Tamás

PhD, adjunktus,
Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Méréstechnika és Információs Rendszerek Tanszék

A MTA BTK Irodalomtudományi Intézete és a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Kar Méréstechnika és Információs Rendszerek Tanszéke közös szervezésében 2015. november 24-én *Számítógép az irodalomtudományban* címmel konferenciát szervezett az MTA Könyvtár és Információs Központban. A szakmai tanácskozáson huszonegy előadás hangzott el, tizenöt résztvevő szoftver- és poszterbemutatót tartott. Tizenhárom intézmény képviseltette magát a rendezvényen hazai és külföldi egyetemokről, kutatóintézetekből.

Az irodalomtudományban egyre nagyobb teret kap a digitális eszközök alkalmazása: a források elektronikus rögzítése és különféle korpuszok, gyűjtemények, adatbázisok, archívumok létrehozása. A konferencia megrendezésével célunk egy olyan szakmai fórum összehívása volt, amely lehetőséget teremt a hazai kutatócsoportok eredményeinek meg-

ismerésére és jövőbeli együttműködések kialakítására. A tanácskozás jelentőségét az adta, hogy az érintett tudományterületek mind-egyike képviseltette magát: az irodalomtudomány, a nyelvtudomány és az informatika.

Konferenciánk az alábbi tematikai egységekre tagolódott: lezárult vagy már számottevő eredményekkel rendelkező projektek bemutatása, amelyben kiemelt hangsúlyt kapott a bölcsészettudományi kutatásokban elért új eredmények prezentálása; a szövegkorpuszok feldolgozásához alkalmazható informatikai módszerek, eljárások ismertetése; demonstrációk, szoftver-, rendszerbemutatók; végezetül a gyakorlati megvalósítás előtt álló, de már határozott koncepcióval rendelkező tervezett kutatások, projektek, elméleti munkák felvezetése. Az előadások mellett a különféle szoftverek, eszközök megismerésére, kipróbálására is lehetőség nyílt.

Az alábbiakban ízelítőt nyújtunk a konferencián elhangzott témákból, valamint kapcsolódó irodalmakat adunk meg az érdeklődő olvasók számára. E bevezető írást követő-

en Almási Zsolt, Bodrogi Ferenc, Dömötör Adrienne és Novák Attila, Kiss Margit, Mészáros Tamás, Palkó Gábor, Péter Róbert, valamint Seláf Levente részletesebben ismereti kutatásait.

A rendezvény fővédnöke, Kecskeméti Gábor köszöntőjében hangsúlyozta az informatika térhódítását és mára már nélkülözhetetlen szerepét az irodalomtudományi kutatásokban. Az MTA Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetében zajló vállalkozások közül a kritikai kiadások korszerű feldolgozása mellett a magyar irodalomtörténet bibliográfiai adatbázisának létrehozását emelte ki, valamint az alkalmazni kívánt eszközökről is hallhattak a jelenlévők (Kecskeméti, 2014).

Prószéky Gábor előadásában arról beszélt, hogy a feladatainkat megoldó programok is emberi alkotások, nem a számítógépek szülöttei, így hibázhatnak, emellett alkotójuk szubjektív nézőpontját tükrözik. Gépi rendszereinkben megfigyelni, feldolgozni csak adatokat tudunk, amelyek azonban igen nagy utat tesznek meg a közlő szándékától a befogadó értelmezéséig. Ez a folyamat számos pontatlanságot, bizonytalanságot indukál, amely könnyen tönkretelheti mégoly tökéletes módszereink eredményességét is (Prószéky, 2013).

Köszeghy Péter *A magyar művelődéstörténeti lexikon adatbázisa* munkálatai alapján mutatta be azokat az eljárásmodokat, amelyek alkalmazásával az informatikai megoldások segítik a szerkesztési munkálatokat. Az adatbázis-struktúrába történő rendezéssel az anyagok sokkal inkább alkalmassá válnak a szerzők tematikus összefüggések láttatására, mint könyvfórmátumban (Köszeghy, 2015).

A konferencián meghívásunkra tutoriált tartott Maciej Eder, a Lengyel Tudományos Akadémia kutatója, a terület nemzetközi szín-

tű elismert szakembere. Az élvonalbeli kutatásokról szóló áttekintésében – a modern kor módszertani újításaként – a számítógépes stilometriai, stilisztikai módszerek irodalomtudományban történő alkalmazásáról beszélt. Latin irodalmi korpuszokat vizsgálva rávilágított a hagyományos módszerrel és az informatikai eszköztárral végzett kutatások különbségeire és a különböző forrásszövegek közti feltáratlan összefüggésekre. Előadásában bemutatta az emberi szem számára láthatatlan szerzői „ujjlenyomat” számítógépes felismerésének módszereit, a klaszterezés és más statisztikai módszerek kínálta lehetőségeket (Eder, 2014).

Pallinger Péter előadásában az elosztott tudásbázisok építése témakörben az MTA Sztaki Elosztott Rendszerek Osztálya által létrehozott és karbantartott lod.sztaki.hu szolgáltatást ismertette. Ez a rendszer könyvek és más művek adatait tartalmazza, valamint elérhetővé teszi kapcsolataikat külső szerverekhez (például a DBpediához). A rendszer bemutatása során ismertette a Linked Open Data technológiában rejlő lehetőségeket, és kitért az adatok böngészésére szolgáló Lodmilla rendszerre is (Micsik et al., 2014).

Simon Zsolt egy 17. századi verseket tartalmazó adatbázis technikai megoldásait ismertette meg a hallgatósággal. Bemutatta az alkalmazott jelölőnyelveket és alternatíváikat, megmutatta, hogy nemcsak emberek, hanem külső gépi felhasználók számára is készíthetünk szolgáltatásokat. Utalt a programozói felületek verziójának fontosságára is. Az általuk kifejlesztett eszközök a Github szoftvertárban érhetőek el, így más munkák számára is hasznosakká válhatnak (Horváth, 2010).

Mártonfi Attila a Mikes- és a József Attila-szótár informatikai munkálatai kapcsán tartott előadást az írói szótárak készítése során

* Az összeállítás a 2015-ös Tudományünnepen tartott konferencia előadásainak alapján készült.

alkalmazott informatikai eljárásokról. Prezentációjában kitért a szótárkészítés folyamatában egyre nagyobb térhez jutó számítógépes módszerek bemutatására, de felhívta a figyelmet egyben arra is, hogy mindezek új, további kérdések megválaszolását is szükségessé teszik (Mártonfi, 2014).

Sass Bálint a *Korpuszok, lekérdezők, Nemzeti Korpuszportál* című előadásában áttekintette a magyar nyelvű szövegek korpuszait abból a szempontból, hogy mit adhatnak az irodalmi és a nyelvészeti kutatások számára. A meglévő korpuszok és a hozzájuk tartozó lekérdezők módszereinek értékelése mellett felvázolta a Nemzeti Korpuszportál kialakítását is, amelyben az eddig létező összes hazai korpusz lekérdezhetővé válna, és bármikor újabb korpuszokkal bővíthetne ez a szolgáltatás (Sass, 2016).

Ruttkay Zsófia *Krúdy nyomában – Szindbád-szövegvizualizációk* című előadásában Krúdy Gyula nyelvének zeneiségét kívánta láthatóvá, kutathatóvá tenni interaktív szövegvizualizációs eszközökkel. A Szindbád-szöveg vizualizációs elemzésével és az eredmény képi formában történő megjelenítésével a szövegek hangtani, ritmikai mintái és azok szövegbeli szerepének a meghatározása válik kutathatóvá. Eredményeit látványos demonstráció keretében is bemutatta.

Rákai Orsolya az új irodalomtörténeti szintézis készítése kapcsán arra a kérdésre kereste a választ, hogy az irodalomtudományban számot vetettünk-e azzal, hogy mit jelent, ha hagyományos, papíralapú irodalomtörténeti összefoglalás helyett digitális, hálózati változatot készítünk. A digitalizált online munkák fontosabb jellemvonásait áttekintve a kutató felhívta a figyelmet arra, hogyan befolyásolja munkánk médiuma az olyan alapvető, nélkülözhetetlen irodalomtörténe-

ti fogalmainkat, mint a folyamat, korszak, kánon, korpusz, szerző, hivatkozás, forrás, hitelesség.

Kiss Dániel a klasszikus latin irodalom eddig megjelent digitális kritikai és nem kritikai kiadásait mutatta be az informatika és az új média kínálta lehetőségekkel együtt. Az internet térhódításával a klasszikus latin irodalom nagy része is elérhetővé vált a világhálón. Napjainkban e művek igényes, kritikai kiadásának a létrehozása kezdődött el. Saját munkáját prezentálva ismertette ennek az időigényes, még gépi támogatással is hatalmas erőfeszítésekkel járó feladatnak a részleteit (Kiss, 2015).

Bátori Anna és Labádi Gergely az 1730 és az 1840 között megjelent magyar regények jellemzőinek a sztenderdizálódási folyamatát kutatják. Arra keresik a választ, hogyan írható le valami (például egy műfaj átalakulása) változásai közepette. Ennek tanulmányozásához létrehoznak egy olyan adatbázist, amely tetszőleges szempontú lekérdezésekkel képes hozzájárulni a kérdéses irodalomtörténeti folyamatnak az elemzéséhez (Labádi, 2014).

A konferencia előadásai rávilágítottak arra, hogy bár az egyre növekvő méretű szövegek korpuszok számítógéppel támogatott létrehozásával gyorsabbá és hatékonyabbá vált a kutatás, egyszerűbben tárolhatók és kevesebb nehézséggel érhetők el a szövegek, de a digitális eszközökben rejlő lehetőségek jobb kihasználása új feldolgozási módszertanok kialakítását is igényli, amelyre az utóbbi időig kevesebb figyelem irányult. Azokon a területeken, amelyekben már eredménnyel alkalmazták a modern informatika kínálta lehetőségeket, az eljárások adaptálása az irodalomtudományi kutatásokba olyan módszertani újítást hozott, amely a korábbiakhoz képest teljesen más megvilágításba helyezi az eddig

elért eredményeket, és immár lehetőség nyílik azoknak a területeknek is a kutatására, amelyeket hosszasan, sokszor pontatlanul, manuális eszközökkel végeztek a kutatók, vagy lehetőségük sem adódott rá a terjedelem nagysága és az erőforrások hiánya miatt. Az informatikai eszközök alkalmazása tehát nemcsak könnyebbé teheti a bölcsezők feladatainak az elvégzését, hanem módszertani innovációt is ösztönözhet, ami jelentős mér-

tékben megújíthatja az irodalomtudományi kutatásokat. Mindehhez azonban az is szükséges, hogy a kutatócsoportok jobban megismerjék egymás eredményeit, újításait, és rendszeres eszmecsere alakuljon ki közöttük a modern informatikai eszközök sikeres alkalmazása érdekében.

Kulcsszavak: *irodalomtudomány, informatika, szövegfeldolgozás*

IRODALOM

- Maciej, Eder (2014): Does Size Matter? Authorship Attribution, Small Samples, Big Problem. *Digital Scholarship in the Humanities*. DOI:10.1093/lc/fqt066 Advance Access published December 2, 2014. • <http://tinyurl.com/jchr18j>
- Horváth Iván (2010): Számítógép és irodalomtudomány. *Alföld irodalmi, művészeti és kritikai folyóirat*, 9–28.
- Kecskeméti Gábor (2014): Electronic Textual Criticism: A Challenge to the Editor and to the Publisher. In: Dávidházi Péter (szerk.): *New Publication Cultures in the Humanities: Exploring the Paradigm Shift*. Amsterdam University Press, Amsterdam, 91–98. • <http://real.mtak.hu/19684/>
- Kiss Dániel (szerk.) (2015): *Catullus Online*. digitális szövegtárhely www.catullusononline.org
- Kőszeghy Péter (2015): A Magyar Művelődéstörténeti Lexikon – középkor és kora újkor (MAMÜL) múltja és jövője. In: Monok István (szerk.): *Mérföldkövek a magyar művelődéstörténet-írásban*. Kossuth, Buda-

- pest, 143–152. • <http://tinyurl.com/hr69max>
- Labádi Gergely (2014): Könyvek távolról: A magyar regény 1807-ben. *Irodalomtörténet*. 95, 3, 311–332. • <http://tinyurl.com/hsz2fbx>
- Mártonfi Attila (2014): Számítógép és írói szótár – különös tekintettel a készülő József Attila-szótárra. *Magyar Nyelv*, 110, 30–46. • <http://tinyurl.com/zu6drtw>
- Micsik András – Turbucz S. – Tóth Z. – Sziládi Z. (2014): *Linked Open Data: konverzió és vizualizáció*. Workshop 2014 konferencia. Video. • <http://tinyurl.com/jugmk3d>
- Prószéky Gábor (2013): How “Truly Electronic Dictionaries” of the 21st Century Should Look Like? In: Stickel, Gerhard – Várad, Tamás (eds.): *Lexical Challenges in a Multilingual Europe*. Peter Lang Academic Publishers, Frankfurt am Main, 51–60. • <http://real.mtak.hu/34421/>
- Sass Bálint (2016): Nyelvészeti szövegkeresők, Nemzeti Korpuszportál. *Magyar Tudomány*. 177, 7, 798–808. • <http://www.matud.iif.hu/2016/07/05.htm>

