

A IX. Nemzetközi Hungarológiai Kongresszus rezümékötete
Miskolc, Miskolci Egyetem Bölcsész- és Társadalomtudományi Kar, 2026.
ISBN 978-963-358-448-4

40. szimpózium

Digitális technológiák a humán tudományokban a 21. században

Szervezés: Palkó Gábor, Fellegi Zsófia
Tudományterület: digitális bölcsészet

WIKIDATA-TÓL ITIDATA-IG. A SZEMANTIKUS ADATBÁZISOK LEHETŐSÉGEI ITTHONI ÉS
NEMZETKÖZI GYAKORLATOKBAN

Dobás Kata

ELTE Humán Tudományok Kutatóközpontja, DigiPhil

Az előadás a szemantikus adatbázisok bölcsészettudományi alkalmazásának lehetőségeit mutatja be nemzetközi és hazai példákon keresztül. Kiindulópontja a Wikidata, amely napjainkra a nyílt, összekapcsolt tudás egyik legfontosabb nemzetközi infrastruktúrájává vált, és egyre meghatározóbb szerepet tölt be a digitális bölcsészeti kutatásokban. Az előadás ezt követően többek között a Wikibase technológiájára épülő, történészek számára készült FactGrid kutatási platformot ismerteti, amely számos történeti és digitális bölcsészeti projekt számára biztosít együttműködési és adatkezelési környezetet.

A nemzetközi példák után az előadás az ELTE HTK Irodalomtudományi Kutatóintézetében fejlesztett ITIdata szemantikus adatbázisra fókuszál. Az ITIdata példáján keresztül bemutatja, hogyan támogatják a szemantikus technológiák a magyar irodalomtudományi kutatásokat, a kutatási adatok strukturált kezelését, valamint az egymástól független projektek adatainak összekapcsolását. Az előadás végül arra is kitér, hogy a szemantikus adatbázisok milyen fejlődési és alkalmazási lehetőségeket kínálnak a hazai bölcsészettudományi kutatási infrastruktúrában, és milyen szerepet tölthetnek be a jövő digitális tudományos ökoszisztémájának kialakításában.

KÉZIRATTÓL A KÉPERNYŐIG: DIGITÁLIS KRITIKAI KIADÁSOK A DIGIPHIL-EN**Fellegi Zsófia**

ELTE Humán Tudományok Kutatóközpontja, DigiPhil

A digitális kritikai kiadás nem csupán a nyomtatásban megjelent kritikai kiadás képernyőre helyezett változata: létrehozása során a filológiai döntések strukturált, géppel is értelmezhető adatszerkezetben rögzülnek. Az előadás az Irodalomtudományi Intézetben működő DigiPhil kutatási infrastruktúra példáján mutatja be, hogyan jut el egy kéziratos vagy nyomtatott forrás a tudományos digitális publikációig. Röviden áttekinti a platform 2014 óta formálódó működését, valamint azt a munkamegosztást, amelyben sajtó alá rendezők, digitális filológusok és informatikusok közösen alakítják ki az adott kiadáshoz illeszkedő szerkesztési és kódolási modellt. A folyamat a források feldolgozásától és a digitális objektum meghatározásától a testreszabott virtuális kutatókörnyezetben készülő TEI XML-átíraton, az ellenőrzésen, lektoráláson és verziókezelésen át a repozitóriumi közzétételig vezet. Az előadásban bemutatom, miként teszi a DigiPhil felülete a strukturált szöveget olvashatóvá, kereshetővé és a szövegváltozatok összehasonlítására alkalmassá, továbbá hogyan kapcsolódik össze a kiadások filológiai megbízhatósága a hosszú távú megőrzés és az újrafelhasználhatóság követelményeivel. Az előadás célja, hogy a digitális kritikai kiadások létrehozását és közzétételét többféle szakértelmet, technológiai réteget és egymásra épülő munkafolyamatot összehangoló komplex vállalkozásként mutassa be, egyúttal kijelölve azokat a kapcsolódási pontokat, amelyeket a szimpózium további előadásai részletesebben tárgyalnak.

XML-TŐL A NYOMDÁIG: KRITIKAI KIADÁS ELŐÁLLÍTÁSA PYTHON ÉS LATEX SEGÍTSÉGÉVEL**Hernády Judit**

ELTE Humán Tudományok Kutatóközpontja, DigiPhil

A digitális technológiák megjelenésével a kritikai kiadások elkészítése területén is új lehetőségek nyíltak meg, a szövegek strukturált tárolása, automatikus feldolgozása és professzionális tipográfiai megjelenítése ma már programozási eszközökkel is támogatható. Az XML-ben a szöveg a jelölőnyelvi kódok révén összetett adatszerkezetként jelenik meg. A Python nyelv segítségével automatizálhatók azok a feladatok, amelyek kézi elvégzése nagy mennyiségű szöveg esetén lassú és hibalehetőségekkel teli lenne. A LaTeX különösen alkalmas tudományos és tipográfiai összetett dokumentumok létrehozására. Míg a hagyományos szövegszerkesztőkben a szerző gyakran közvetlenül a dokumentum kinézetét alakítja, a LaTeX elválasztja a tartalmat és a megjelenítést. A kutató meghatározza a dokumentum szerkezetét, a tipográfiai szabályokat pedig a rendszer kezeli. A három technológia együttes alkalmazása

zása lehetővé teszi, hogy a born digital kiadásokból tudományos igényű, nyomdakész kritikai kiadást hozzunk létre. Előadásomban a Babits Mihály összes versei kritikai kiadássorozat 3. és 4. köteteivel kapcsolatos munkafolyamatokat fogom bemutatni.

KÉZIRATOK A DIGITÁLIS TÉRBEN – HTR TECHNOLOGIÁK ÉS ESZKÖZÖK

Kovács-Bobák Barbara

ELTE Humán Tudományok Kutatóközpontja, DigiPhil

A kéziratos források kutatásában, az utóbbi években egyre nagyobb szerepet kapnak a mesterséges intelligenciára épülő dokumentumfeldolgozó technológiák. Ezek a kézírásfelismerő (Handwritten Text Recognition, HTR) technológiák lehetővé teszik a kézzel írott források automatikus átírását, kereshetővé tételét és hatékony kutatását, ezzel új perspektívákat nyitva a levéltári, könyvtári, bölcsészeti és filológiai munkában. Az előadásban bemutatásra kerülnek a HTR-technológiák működésének alapjai, különös tekintettel a Transkribus platformra, amely jelenleg az egyik legelterjedtebb megoldás a történeti kéziratok automatikus feldolgozásában. Részletes ismertetést kap egy saját fejlesztésű, TrOCR-re épülő alkalmazás is, amelynek kialakítását és továbbfejlesztését gyakorlati kutatási igények motiválták, és amely különböző kéziratos források feldolgozásában szerzett tapasztalatok alapján kerül bemutatásra. Röviden kitér az eScriptorium nyílt forráskódú platformra, valamint a Vision Language Model (VLM) megközelítésre is, amely a dokumentumok vizuális és szöveges információinak együttes értelmezésével a jövő egyik ígéretes kutatási irányát képviseli. A bemutatott technológiák saját fejlesztésű projektek példáin keresztül kerülnek összehasonlításra, rávilágítva a gyakorlati alkalmazásuk során szerzett tapasztalatokra.

FILOLÓGIA A NAGY NYELVMODELLEK KORÁBAN

Palkó Gábor

ELTE Humán Tudományok Kutatóközpontja, DigiPhil

A generatív mesterséges intelligenciáról, vagy, általánosabban, a nagy nyelvi modellekről többnyire úgy beszélünk, mint különös módon működő, nehezen átlátható eszközökről: szövegeket fogadnak be, szövegeket hoznak létre, miközben működésük lényege rejtve marad előttünk. De miben áll valójában egy nyelvtechnológiai apparátus „megbízhatósága”? Ki „hibáztatható”, ha a bemenet, a feldolgozás és a kimenet egymásba fonódó műveletsorában már nem választható el élesen az emberi és a gépi közreműködés?

Az előadás különböző nyelvtechnológiai eljárásokon keresztül azt vizsgálja, miként lényegül át az irodalmi szöveg adattá, numerikus reprezentációvá, majd ismét ér-

telmezhető szöveggé. Az egyszerű szógyakorisági elemzéstől a stilometriai reprezentáción át a nagy nyelvmodellekig minden nyelvtechnológia sajátos módon tesz láthatóvá bizonyos mintázatokat, miközben másokat elfed. Ezen fekete dobozok (részleges) felnyitása ezért nem pusztán technikai feladat, hanem filológiai és hermeneutikai kérdés is.

A vizsgálat hátterét egy több mint ezer magyar regényt tartalmazó korpusz építése, valamint a RACKA magyar nyelvmodell fejlesztésének tapasztalatai adják. A cél nem annak eldöntése, hogy a gépi olvasás helyes vagy hibás, hanem annak megértése, milyen döntések, korlátok és előfeltevések alakítják azt, amit a modell egy szövegből érzékelni és megmutatni képes.

FELTÁRATLAN ÖRÖKSÉG NYOMÁBAN: A MAGYAR BÁNYÁSZLAP (1913–1962) ÉS AZ APPALACHE-I MAGYAR KÖZÖSSÉG DIGITÁLIS FELTÉRKÉPEZÉSE

Péter Róbert

Szegedi Tudományegyetem

Az 1870-es és az 1920-as évek között az Osztrák–Magyar Monarchia területéről jelentős kivándorlási hullám indult az Egyesült Államokba. Ezek a bevándorlók többnyire az északkeleti, a közép-atlanti és a délkeleti államok gyáraiban, szénbányáiban és acélüzemeiben találtak munkát. Közülük kiemelkedő szerepet játszott Himler Márton (1888–1961), a Magyar Bányászlap című hetilap alapítója és szerkesztője. A lap létfontosságú támasz volt a magyar bevándorló közösségek számára. Az újság a kapcsolattartás és az identitásmegőrzés egyik legfontosabb fóruma volt: nemcsak az amerikai magyarsággal és az óhazával kapcsolatos híreket közölte, hanem lehetőségeket is közvetített, például állás- és lakáshirdetéseket, közösségi felhívásokat, leveleket, személyes történeteket, sőt eddig ismeretlen irodalmi műveket is. Így tehát nem pusztán a bányászoknak szóló lapként működött. Jelentősége ellenére a Magyar Bányászlap mindaddig javarészt feltáratlan maradt az amerikai–magyar kapcsolatok kutatásában. Az Országos Széchényi Könyvtár fontos előrelépést tett azzal, hogy digitalizálta a lapot. A legkorszerűbb technológia alkalmazásával a teljes folyóiratot újra OCR-eztük, majd megkezdtuk a kinyert szövegek tisztítását és előfeldolgozását. A digitális bölcsészet módszereire támaszkodva elemzéseket készítettünk a csaknem teljes Magyar Bányászlap anyagán. Az AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) kutatási eszköz segítségével többek között n-gram elemzést, témamodellezést és névelem-felismerést végeztünk. Ezek a vizsgálatok új kutatási irányokat nyitnak meg az appalache-i magyar örökség, valamint a közösségi élet és a nyelvhasználat feltárásában. A digitális elemzés láthatóvá teszi a közösség legjelentősebb személyeit, helyszíneit, intézményeit és a lap visszatérő témáit, emellett új távlatokat nyit az amerikai magyar kölcsönszavak kutatásában is.

A MAGYARORSZÁGI KATASZTERI FELMÉRÉS ÉS LEVÉLTÁRI FORRÁSANYAGA, 1850–1918

Reisz T. Csaba

Magyar Nemzeti Levéltár

A Habsburg Birodalom örökös tartományaiban 1817-ben indult meg az országok részletes, a földadó igazságos kivetését megalapozó kataszteri felmérése. A Magyar Korona országaiban csak a levert szabadságharc után, 1850-ben vezették be a földadót, és az ehhez szükséges kataszteri felmérési munkálatok is megindultak. Az igen részletes térképek és az ahhoz készült egyéb kataszteri iratok (kataszteri telekkönyv, birtokosok névjegyzéke, helynevek jegyzéke stb.) a történettudomány számos területének (gazdaság-, társadalom-, család- és helytörténet stb.) biztosíthat kiemelkedő forrásanyagot. A kataszteri térkép- és iratanyag fennmaradása és megőrzése napjainkban igen sok – és nem csak magyarországi – (köz)gyűjteményhez kapcsolódik, azok fellelése és kutatása sokszor az információhiány miatt nem lehetséges. A forrásegyüttes virtuális egyesítését és településszintű adatbázisát a Magyar Nemzeti Levéltár tudományos projektjei, 2019 óta a „Négyzetkataszter” című NKFIH-támogatású projekt keretében végezzük. Az előadás a projekt célkitűzéseit és eredményeit ismerteti a kataszteri felmérések forrásértékének kontextusába ágyazva.