

Kiszl Péter – Mátyás Melinda

Digitális bölcsészet a könyvtár- és információtudományban. 2. rész Európai kutatástámogató projektek*

A digitális bölcsészet, valamint a könyvtár- és információtudomány szoros kapcsolatban álló diszciplínák. Tanulmányosorozatunk második részében a European Association for Digital Humanities (EADH) által közzétett projektek alapján digitális bölcsészeti technológiákat alkalmazó kutatástámogató programokat (#dariahTeach, The Programming Historian, CATMA – Computer Aided Textual Markup and Analysis, EVT – Edition Visualization Technology, READ – Recognition and Enrichment of Archival Documents, OES – Open Encyclopedia System) elemzünk a magyarországi könyvtárakban és könyvtárosképzésben történő elterjedésük ösztönzése, a hazai tartalomszolgáltatások színvonalának emelése érdekében.

Tárgyszavak: bölcsészettudomány; digitális technika; információmenedzsment; innováció; oktatás; kutatás; tartalomszolgáltatás; szoftver

A digitális bölcsészet európai hálózata

Az 1973-ban Association for Literary and Linguistic Computing (ALLC) néven alakult, 2012-től European Association for Digital Humanities (EADH)¹ a digitális bölcsészet legelső és legnagyobb európai szakmai szervezete, mely a formálódó új diszciplína tudományterületeken átívelő kontinensbeli képviselőjét vállalja fel. Elsősorban a kutatás, a fejlesztés, illetve a korszerű módszerek és technológiák közvetítése áll tevékenysége fókuszában.² Az EADH célja mintegy 850 egyéni tagjának tudományos hálózatba szervezése, melyet – informálisan csatlakozott partnerek (Partner Organizations) mellett – regionális tömörülések (Associate Organizations) is aktívan segítenek:

- AIUCD – Associazione per l'Informatica Umanistica e la Cultura Digitale (Olaszország),
- CZDHI – Czech Digital Humanities Initiative (Csehország),
- DHd – Digital Humanities im deutschsprachigen Raum (német nyelvek),
- DHN – Digital humaniora i Norden (északi országok),
- DH Russia – Russian Association for Digital Humanities (Oroszország).³

Az EADH az Alliance of Digital Humanities Organizations (ADHO)⁴ világszervezet tagja,⁵ az ALLC *Literary and Linguistic Computing* címmel

1986-ban útjára indított *Digital Scholarship in the Humanities (DSH)*,⁶ az *Oxford University Press (OUP)* berkeiben megjelenő nemzetközi szakfolyóirat alapítója.

Az EADH linkgyűjteményként is használható informatív honlapot üzemeltet (1. ábra). Megalakulása óta éves konferenciákat szervez, melyek 2006-tól az ADHO égisze alatt zajlanak.⁷ Továbbá támogatja a képzést,⁸ katalizálja a digitális bölcsészet ismertségét és elismertségét, ennek érdekében díjakat is adományoz,⁹ és lehetőségeihez mérten anyagi forrásokat is biztosít, valamint külsők felkutatásában is közreműködik.¹⁰

A digitális bölcsészet főbb európai kutatástámogatói projektjei

Az EADH kiemelt vállalása az európai digitális bölcsészeti projektek önkéntes alapú *regisztrálása*, az egyes kezdeményezéseknek az EADH saját honlapján történő, szabadon elérhető közzététele.¹¹

* A tanulmány az Emberi Erőforrások Minisztériuma (EMMI) támogatásával készült.



1. ábra A European Association for Digital Humanities nyitóoldalának részlete

A vállalkozások közül újszerűnek számító digitális bölcsészeti megoldásokkal, különösen a *kutatástámogató alkalmazásokkal* foglalkozunk jelen közleményünkben. Összesen hat, megítélésünk szerint *innovatív, könyvtári környezetbe adaptálható* programot vizsgálunk.

Dolgozatunkban először a portálok mindegyikére vagy egy-egy kisebb csoportjára érvényes tulajdonságairól szólnak, majd részletesebben bemutatjuk a fejlesztéseket abból a szempontból, hogy mi a profiljuk, mely szervezet(ek) támogatják, illetve melyek a partnerintézményeik (ha vannak). Végül pedig számba vesszük az adott projekt egyedi vonásait és az általuk biztosított digitális bölcsészeti szolgáltatásokat.

A programok – céljaik szerint – két nagy csoportra oszthatók:

1. Digitális bölcsészeti *oktatási anyagokat* tartalmazó kutatástámogató projektek:
 - #dariahTeach,¹²
 - The Programming Historian.¹³
2. *Szövegpublikáláshoz* kötődő digitális bölcsészeti technológiákat kínáló szoftverek:
 - CATMA – Computer Aided Textual Markup and Analysis,¹⁴
 - EVT – Edition Visualization Technology,¹⁵
 - READ – Recognition and Enrichment of Archival Documents,¹⁶
 - OES – Open Encyclopedia System.¹⁷

A digitális tananyagokat publikáló két felület a digitális bölcsészet különböző dimenzióit mutatja be. A *#dariahTeach* oldal jellemzően bevezető tananyagokat biztosít a digitális bölcsészet témakörébe és nagyobb tematikát ölel fel: például a TEI (Text Encoding Initiative) kódolásról, a multimodális műveltség területéről és a digitális szöveg-előállításról (kritikai kiadások, szótárak készítése stb.). A *The Programming Historian* portálon kisebb, specifikusabb oktatási anyagokat találunk egy-egy technológiai eszközre, technikai megvalósításra összpontosítva (pl. bevezetés a MySQL-be, hangszerkesztés az Audacity programmal, az R alkalmazás használata a szövegek elemzésére). A tananyagok közzétételét tekintve is különböznek: az első a Moodle ingyenes e-learning szoftvert használja a tananyagok publikálásához, a másodikon pedig HTML-oldalakra jelenik meg a szöveg beágyazott tartalmakkal (képek és videók). Mindkettő jól strukturált felületet biztosít a leckék közzétételéhez és mindegyikben található gyakorlati példákat és feladatokat is. Egyelőre a *The Programming Historian* portálon böngészhetünk több tananyagot.

A szövegpublikálást segítő projektek közül az *EVT* és a *Transkribus*¹⁸ célja, hogy a kéziratos vagy a gép által egyelőre nem egyszerűen karakterfelismertethető szövegek (pl. régi újságok) elérhetőek legyenek a digitalizált képen kívül olyan formában is, hogy a teljes szövegben való keresés biztosított legyen. A *READ* projekt keretében létrejött *Transkribus* a kézirat vagy a régi nyomtatott oldalak átirását, akár gépi szövegfelismertetését és a végeredmény exportálását támogatja különböző formátumokba (pl. TEI XML). Az *EVT* pedig a már kész TEI XML szövegek szerkesztésére, publikálására szolgáló eszköz, amellyel kritikai apparátust is készíthetünk a művekhez. A *CATMA* és az *OES* az előzőektől eltérő funkciókat kínálnak: szövegek elemzését és közzétételét, ezeket a profiluk leírásánál és az egyedi bemutatásuknál ismertetjük.

Az 1. táblázatban áttekintjük az elemzett projektek profiljait, támogatóit, partnerintézményeit.

1. táblázat

Az EADH portálján regisztrált kutatástámogató digitális bölcsészeti projektek áttekintése

A projekt neve és profilja	Támogató(k) és partnerintézmény(ek)
<p>#dariahTeach</p> <ul style="list-style-type: none"> Nyílt forráskódú, többnyelvű, közösség által formált felület minőségbiztosított digitális bölcsészeti oktatási és tréning anyagok fejlesztésére. 2015-ben indult, eredetileg azzal a céllal, hogy oktatási anyagokat fejlesszen a digitális bölcsészet- és művészettudomány területén, továbbá erősítse az együttműködések, valamint elősegítse az innovatív tanítási és tanulási gyakorlatok elterjedését az ESFRI-roadmap Digital Research Infrastructure for the Arts and Humanities (DARIAH) tagjai között. A nyilvános portáljuk 2017-ben indult el béta tartalommal. 	<p>Támogató: ERASMUS+</p> <p>Partnerintézmények:</p> <ul style="list-style-type: none"> Maynooth University Aarhus University's DIGHUMLAB Digital Curation Unit, Research Center "Athena" Austrian Academy of Sciences Belgrade Centre for Digital Humanities Erasmus University Rotterdam University of Luxembourg University of Lausanne Swiss Institute of Bioinformatics
<p>The Programming Historian</p> <ul style="list-style-type: none"> A <i>Programming Historian</i> egy kezdő tanuló-barát, lektorált tutoriálokat tartalmazó portál, amely segíti a bölcsészeket, hogy elsajátíthassák a digitális eszközök, technikák, munkamódszerek széles spektrumát. Rendszeresen közzétesznek új leckéket és nyitottak az új témajavaslatokra. 	<p>Támogatók: Network in Canadian History & Environment (NiCHE) Roy Rosenzweig Center for New Media (RRCHNM) A projekt önkéntesek munkájával valósul meg.</p> <p>A jelenlegi közreműködők:</p> <ul style="list-style-type: none"> Adam Crymble, University of Hertfordshire Fred Gibbs, University of New Mexico Allison Hegel, University of California, Los Angeles Caleb McDaniel, Rice University Ian Milligan, University of Waterloo Evan Taparata, University of Minnesota Jeri Wieringa, George Mason University

<p>CATMA</p> <ul style="list-style-type: none"> • Ingyenes, nyílt forráskódú szövegjelölő és elemző eszköz. • Három interaktív modulból áll: <ol style="list-style-type: none"> 1. Jelölő: szövegbeli jelölések és jelölések szerkesztése. 2. Elemző: beépített keresőnyelv és előre definiált funkciók. 3. Keresőkérdés-szerkesztő. • A Voyant eszköztárat is felhasználja. • A kurrens (5.0 verzióban) teljes eszköztár áll rendelkezésre automatizált annotálási feladatokra. 	<p>Támogató:</p> <ul style="list-style-type: none"> • Federal Ministry of Education and Research (BMBF) • Google
<p>EVT</p> <ul style="list-style-type: none"> • Az EVT egy kis rendszerigényű, nyílt forráskódú rendszer arra tervezve, hogy a TEI XML-kódolású szövegekből digitális kiadásokat lehessen létrehozni. • Előnye, hogy a kutatónak nem kell web-programozónak lennie. Lehetővé teszi, hogy felhasználóbarát módon böngésszék és tanulmányozzák a digitális kiadást. • Két változata érhető el: az EVT1-et oklevélkiadások és más kéziratkiadások készítésére ajánlott használni, az EVT2-t pedig kritikai kiadásokra. A jövőben ez utóbbit fogják tovább fejleszteni, hogy az előző verzió funkcióit is támogassa. 	<p>Támogatók:</p> <ul style="list-style-type: none"> • Università degli Studi di Torino • Università degli Studi di Pisa • Progetto ALIM • Archivio e Biblioteca Capitolare di Vercelli <p>Partnerintézmények:</p> <ul style="list-style-type: none"> • Digital Vercelli Book Project • Codice Pelavicino Digitale • Laboratorio di Cultura Digitale • Digital Edition of an Ancient New Testament Arabic Manuscript
<p>READ</p> <ul style="list-style-type: none"> • A READ projekt központjában az archív anyagok közzététele áll csúcstechnológiai megoldásokkal. • A transcriptorium projektjére építve a READ új szabványokat igyekszik kialakítani a kézírásos szöveg felismerésében, a kulcsszavak azonosításában, a szövegterületek elemzésében, az írnok automatikus azonosításában és az ezekhez kapcsolódó területeken. • Szolgáltatási felületet (Transkribus) biztosít a projekt, amelyen elvégezhető az automatikus felismertetés, az átirás, és keresni is lehet a történelmi dokumentumokra. A számítógépek megtanítása a kézírások felismerésére ezen a módon úttörő lehetőséget kínál a történelmünk írásos emlékeihez való hozzáférésre. 	<p>Támogató:</p> <ul style="list-style-type: none"> • European Commission Horizon 2020 Research and Innovation Programme <p>Partnerintézmények:</p> <ul style="list-style-type: none"> • University of Innsbruck • Universitat Politècnica de Valencia • University College London • National Center for Scientific Research “Demokritos” • University of London Computer Centre • Vienna University of Technology • University of Rostock • Leipzig University • Xerox Research Centre Europe • École Polytechnique Fédérale de Lausanne • National Archives Finland • Zurich State Archives • Passau Diocesan Archives
<p>OES</p> <ul style="list-style-type: none"> • Az OES fő célja, hogy egy szabványosított, nyílt forráskódú, web-alapú felületet hozzon létre tudományos enciklopédiák publikálására és szabadon hozzáférhetővé tételére. Az első, ennek a rendszernek a felhasználásával készült portál az 1914-1918-online – International Encyclopedia of the First World War és a CeMoG-Encyclopedia on German-Greek Entanglements. 	<p>Támogatók:</p> <ul style="list-style-type: none"> • German Research Foundation (DFG) • Freie Universität Berlin <p>Partnerintézmények:</p> <ul style="list-style-type: none"> • Freie Universität Berlin • Bavarian State Library (BSB)

Mielőtt ismertetnénk az egyes projektek egyedi sajátosságait és az általuk nyújtott szolgáltatásokat, hangsúlyozzuk, hogy az oktatási materiákat közzé tevő portálok a könyvtárosoknak is igen hasznosak lehetnek. Egyrészt azért, hogy elmélyíthessék a digitális bölcsészeti eszközökre vonatkozó technikai tudásukat, másrészt pedig, hogy a kutatókat is megismertethessék e lehetőségekkel, támogathassák a tanulásukat.

A szövegpublikáláshoz készült alkalmazásokat a közgyűjtemények ugyancsak használhatnák az állományukban levő különböző korú írásközvetítésére, a CATMA és az EVT modern kori, digitálisan keletkezett szövegek, szövegvariánsok elemzésére és publikálására, a READ keretében fejlesztett programot pedig kéziratos szövegek kereshetővé tételére és megosztására. Az *Eötvös Loránd Tudományegyetem Bölcsészettudományi Karának Könyvtár- és Információtudományi Intézetben (ELTE BTK KIT)* például a *Muzeális gyűjtemények számítógépesítése* című, informatikus könyvtáros alapszakos könyvtörténet specializációs hallgatók számára a 2018/2019. tanév őszi félévére meghirdetett kurzus résztvevői a READ fejlesztését, a Transkribust használva írják át egy rövidebb kéziratos az *ELTE Egyetemi Könyvtár és*

Levéltár által digitalizált¹⁹ Kaprinai gyűjteményből.²⁰

#dariahTeach²¹

A projekt során létrehozott tananyagokat a #dariahTeach's Quality Assurance Plan alapján szigorúan ellenőrzik, jelentkezni is lehet a tesztelési feladatra. A portálon kétféle oktatási anyaggal találkozunk: kurzusokkal és workshopokkal. Technikailag a Moodle e-learning keretrendszerrel használják a leckék megosztására. Ennek olyan a felépítése, hogy mindenképpen szükséges bejelentkezni a belső tartalom megtekintéséhez. Ezért vagy egy vendégbelépést használunk (regisztráció nélkül) vagy regisztrálnunk kell magunknak egy saját fiókot, és azzal böngészünk a felületen.

A kurzusok (Courses) csoportban nyolc tananyagot találunk, a nyelvi eloszlásuk szerint egy francia nyelvű, egy magyar, a többi angol. A kurzusok témája: TEI kódolás, digitális tudományos kiadások, kéziratos és szövegek TEI kódolásban, papír alapú szótárak digitalizálása, hanggal kapcsolatos kurzus, digitális eszközök használata (2. ábra).

Courses
Courses represent the equivalent level of student effort as a 5 or 10 ECTS module.

- Text Encoding and the TEI - FR
- Text Encoding and the TEI - HU
- Introduction to Digital Humanities
- Text Encoding and the TEI
- Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding

Workshops
Workshops should take anywhere from several hours to two days to complete. They represent a short, focused introduction to a topic, method, or approach.

- Spatial Image Analytics
- An introduction to Conceptual Modelling
- Spatial Image Analytics - GR

2. ábra #dariahTeach kurzusok és workshopok

A workshop menüben rövid, koncentrált bevezetést kapunk egy-egy témakörbe, módszerbe, megközelítéshez. A kurzusokhoz hasonlóan ezeket is bárholnan el lehet végezni néhány órától két napig terjedő időtartamban. Olyan tananyagokat találunk itt mint például a digitális képszerkesztés és a fogalmi modellezés, valamint az ontológiák. A tananyagok jól szerkesztettek, tartalmazznak szemléltető eszközöket, videókat, képeket is. Külön pozitívum, hogy magyar nyelven is találunk kurzust rajta: a TEI és a TEI XML (Extensible Markup Language) szövegkódolási mód alapjairól.²² TEI XML segítségével például különböző szövegeket adhatunk közre, illetve vannak olyan aggregáló gyűjtemények, amelyek kifejezetten TEI XML-ben kódolt textusokat gyűjtenek és szolgáltatnak, ilyen például a *TextGrid*²³ vagy a *University of Oxford Text Archive (OTA)*.²⁴

The Programming Historian²⁵

Minden lecke lehetőség szerint nyílt forráskódú programozási nyelveket és szoftvereket használ. 2016-tól a *Zenodo*²⁶ felületén is elhelyezték a projektet.

Jelenleg 112 leckét érhetünk el kutatói aktivitások és általános témakörök szerint csoportosítva. Ku-

tatói tevékenységek szerint: beszerzés, átalakítás, elemzés, bemutatás, fenntartás. A másik kategóriában olyan témaköröket találunk, mint az API-k, a Python programnyelv és a Distant Reading.

A leckéket rendezhetjük megjelenés dátuma és nehézségi fokozat szerint, ez utóbbi ismerv valóban hasznos a kezdő és a haladóbb programozó bölcsészek számára is. A tananyagokat két nyelven érhetjük el, angolul (74 lecke) és spanyolul (38 lecke).

Minden írásnál feltüntetik a szerzőt, a lektoráló nevét, a publikálás és az utolsó módosítás idejét is. A tananyagok egy weboldalon jelennek meg, az általános adatok után legfőképpen találunk egy tartalomjegyzéket, amely segítségével a szöveg megjelölt egységeire ugorhatunk (3. ábra). Azt is jelzi a rendszer, ha egy több részből álló leckét választottunk ki, és lehetőségünk van a többi egységet is megtekinteni. A tananyagok általában képekkel, képernyőképekkel illusztráltak a jobb követhetőségért. Sokszor „lépésről lépésre” vezető útmutatót nyújtanak a megadott cél érdekében a program vagy bizonyos technológia használatára. Legtöbbször tartalmazznak gyakorló feladatokat is.

The Programming Historian

ABOUT ▾ CONTRIBUTE ▾ LESSONS BLOG [es]

Introduction to the Principles of Linked Open Data

Jonathan Blaney

Introduces core concepts of Linked Open Data, including URIs, ontologies, RDF formats, and a gentle intro to the graph query language SPARQL.

Peer-reviewed CC-BY 4.0

EDITED BY Adam Crymble	REVIEWED BY Terhi Nurmi-Ko-Fuller Matthew Lincoln	
PUBLISHED 2017-05-07	MODIFIED 2018-05-16	DIFFICULTY Low

This lesson has been translated into Spanish: [Introducción a los Datos abiertos enlazados](#)

Contents

- Introduction and Lesson Scope
- Linked open data: what is it?
- The role of the Uniform Resource Identifier (URI)
- How LOD organises knowledge: ontologies
- RDF and data formats
 - Serialisation
 - Turtle
 - RDF/XML
- Querying RDF with SPARQL

3. ábra A Linked Open Data alapjaiba bevezető lecke a The Programming Historian honlapján

Leckék készítésére is van lehetőségünk, ehhez a Contribute link alatt részletes leírásokat találunk.

A Programming Historian jól felépített tananyagokat tartalmaz, amelyek segítségével a kutatók vagy akár a könyvtárosok az élethosszig tartó tanulás megvalósításaként modern technológiákat sajátíthatnak el a munkájuk megkönnyítéséhez. Egyelőre a kurzusok érhetőek el a portálról, véleményünk szerint azonban hasznos lenne egy fórum is, amelyben a felhasználók kérdezhetnek egymástól vagy akár a tananyagkészítőktől is, ha elakadnak. Az angol nyelvű oktatási anyagok között például találunk a könyvtári vonatkozásban is sokat emlegetett Linked Open Data-ról is két leckét.²⁷

CATMA²⁸

Az analízisek készítését biztosító felület kulcsfontosságú eleme, hogy felhasználói kombinálhatják a hermeneutikus/undogmatikus és a digitális/taxonómiai megközelítést a szövegekhez és a szövegkorpuszokhoz. Az undogmatikus megközelítés azt jelenti, hogy a rendszer nem követeli meg a felhasználótól egy előre meghatározott, *annotációs* modell használatát, valamint azt sem írja elő, hogy egy merev „igen/nem, helyes/helytelen” taxonómiát alkalmazzon a szövegekhez az elemző. Ehelyett a CATMA biztosítja, hogy a használók saját jelölési

készleteket (Tagsets) dolgozzanak ki, és akár egymásnak ellentmondó értelmezésekkel is annotálhatják a korpuszokat.²⁹

A CATMA funkciói:

- *Elemzés és keresések* – szólisták előállítás. Gyakoriság szerint rendezi a szövegben előforduló szavakat, egy-egy szót ki lehet jelölni és elemezni a kontextusában (KWIC). A 4. ábra egy tanulmányból készült mintaelemzést mutat. Kereséseket is indíthatunk szavakra, kifejezésekre, tag-ekre, tulajdonságokra, valamint összetett kereséseket is össze lehet állítani.
- *Kézi annotálás* – szövegek annotálása az elemzéshez. Szövegrészt magunk jelölhetünk ki és láthatunk el tag-ekkel. A meglévő (akár általunk létrehozott) tag-készletekből is lehet választani.
- *Automata annotálás* – a heureCLÉA³⁰ segítségével Natural Language Processing (NLP) eszközökkel kínál lehetőséget az annotálásra.
- *Kollaboráció* – betöltött szövegek megosztása, ezzel a közös munka támogatása.
- *Vizualizáció* – az elemzések kiválasztott eredményeiből készíthető vizualizált megjelenítés. Grafikus felületen látszanak a szavak, a kifejezések, a vizsgált tag-ek. A kulcsszavak kontextusban való bemutatása is ilyen eszköz, valamint a rendszer a 3DH projekt³¹ új vizualizációs eszközeit is integrálja.

Phrase	Frequency	Visible in Kwic
Pray	9	<input type="checkbox"/>
szerint	9	<input type="checkbox"/>
1	9	<input type="checkbox"/>
archivo	9	<input type="checkbox"/>
Continet	8	<input type="checkbox"/>
Hóman	8	<input type="checkbox"/>
Dedek	8	<input checked="" type="checkbox"/>
15	8	<input type="checkbox"/>
%	8	<input type="checkbox"/>
például	8	<input type="checkbox"/>
11	7	<input type="checkbox"/>
ELTE	7	<input type="checkbox"/>
van	7	<input type="checkbox"/>
ha	7	<input type="checkbox"/>
of	7	<input type="checkbox"/>
10	7	<input type="checkbox"/>
Tóth	7	<input type="checkbox"/>
Total count: 1.432	Total frequency: 3.572	

Document/Annotations	Left Context	Keyword	Right Context
HKP projekt	vannak hiányzó kötetek, amelyeket	Dedek	is említi a katalógusában:
HKP projekt	jezsuita tudósok káziatgyűjteményében. Ehhez	Dedek	Crescens Lajos katalógusát(9)
HKP projekt) esetben jobban szükséges támaszkodni	Dedek	leírására, és a káziatolvasási
HKP projekt	A Pray-gyűjtemény 9. kötetében (Dedek	szerinti 22. tétel)
HKP projekt	, akkor az bejegyzésre került	Dedek	írását átvéve az egyik megjegyzés
HKP projekt	ezt az adatot többnyire	Dedek	is azonosítja, és mivel
HKP projekt	. A források és helynevek	Dedek	katalógusa alapján Az ötödik oszlopban
HKP projekt	Gyűjtemény Rész Kötet Tétel	Dedek	szövegének részlete Forrás helye Hevenesi

4. ábra Szógyakoriság- és kulcsszó-elemzés a CATMA rendszerében

A Try it out link alatt regisztrációval vagy Google-bejelentkezéssel kipróbálhatjuk az elemző eszközt. Részletes dokumentáció is elérhető a funkciókhoz a projekt főoldalának Documentation menüjében. A könyvtári szakértőoktatásban történő alkalmazási lehetőségekre cikksorozatunk előző részében, a *Voyantnál* utaltunk, az egyes elektronikusan meglévő művek tartalomelemzésénél.

EVT³²

Az EVT eredetileg a *Digital Vercelli Book*³³ keretében fejlesztett alkalmazás kéziratok átírására és digitális kiadására. Később más célokra is felhasználták, például a *Pelavicino kódex*³⁴ elektronikus változatának elkészítéséhez. A szükség, hogy adaptálják a különböző típusú dokumentumokhoz és a TEI-kódolású szövegekhez, megváltoztatta a fejlesztés irányát egy általánosan használható eszköz megalkotására. Az EVT2 főbb funkciói:

- a kritikai kiadás támogatása: megnövelt kritikai apparátus, különböző heatmap (vizualizáló eszköz), szövegváltozatok kapcsolása, különböző szűrések;
- könyvjelzők: a szöveg aktuális nézetére közvetlen hivatkozások hozhatók létre;
- személyre szabhatóság: a szerkesztő be tudja állítani mind a felhasználói felület, mind a grafikus összetevők kinézetét.

Várható fejlesztések: hagyományos kritikai apparátus megjelenítők, a szöveg képének linkelése egy kijelölt szövegváltozathoz, nagyon nagy felbontású képnézegető; GIS-hez (Geographic Information System) hasonló támogatás a névelemek megjelenítéséhez a térképen.

A szoftver mindkét változata letölthető. Az EVT2-nek egy online demóverziója is megtekinthető egy már létrehozott kritikai kiadással. Az eredeti latin nyelvű szöveghez (*Avicenna: Logica* című műve) az egyik fülön elérhetjük a kritikai apparátust kattintható linkekkel a különböző szövegvariánsokra. Az egyes kritikai megjegyzéseket kiemelhetjük magunknak (Toggle Pin), amelyek a képernyőn egy külön oszlopban fognak megjelenni (5. ábra). A források fülön ez esetben főleg a filozófus szerzők szöveghelyeire való utalást követhetjük nyomon. Egy gombnyomásra (Align Quotes) kiemeli a hivatkozás szövegét és odaugrik a szövegnézetben. Az Analogues a párhuzamos szöveghelyeket emeli ki. A nézetek között is válthatunk a jobb felső sarokban:

1. szövegolvasás (kritikai apparátussal),
2. több kritikai kiadás összevetése,

3. a szöveg digitalizált kép változatának az átírt szöveg mellé helyezése,
4. több szöveg összevetése,
5. a vizsgált mű megjelenítése a forrásszöveggel.

A főszövegnél (a bal oldalon) lehetőségünk van ki-be kapcsolni a Heatmap megjelenítését és a különböző szűrőket, valamint növelhetjük vagy csökkenthetjük a betűméretet.

Ha a szövegkiadás-készítő programot feltöltjük az általunk elemezni kívánt adatokkal (képek, XML-fájlok), akkor megfelelő eszköz lehet korszerű digitális kiadások készítéséhez. Még tradicionális mongol nyelvű szövegeket is átírtak és megjelenítettek az EVT segítségével.³⁵

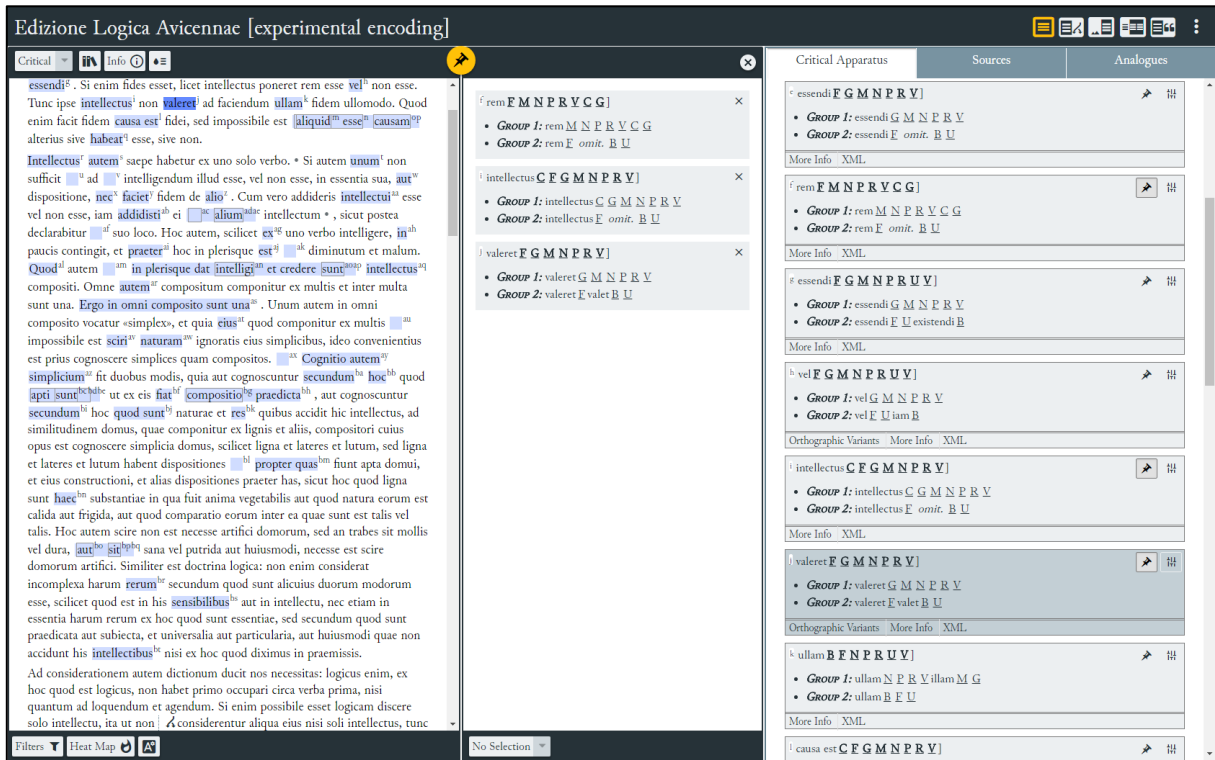
READ³⁶

A READ projekt keretében, a tranScriptorium³⁷ eredményeire is építve hozták létre a Transkribus nevű programot, amellyel automata szövegfelismerést végezhetünk kéziratok szövegeken, átírhatjuk a kézírásos textusokat, valamint kereshetünk bennük. Az alkalmazást letölthetjük a Transkribus főoldaláról, és a linkek mellett rögtön el is érünk egy hasznos felhasználói segédletet az elinduláshoz: *How To use Transkribus – in 10 steps (or less)*.³⁸ Az útmutatóból megtudhatjuk, hogy ingyenes regisztráció szükséges az alkalmazás használatához, mivel ekkor létrejön egy tárhelyünk a szerveren. A Transkribus oldalán is regisztrálhatunk új fiókot, illetve a Google fiókunkkal is bejelentkezhetünk és a szükséges engedélyek megadása után eképp is használhatjuk a programot.

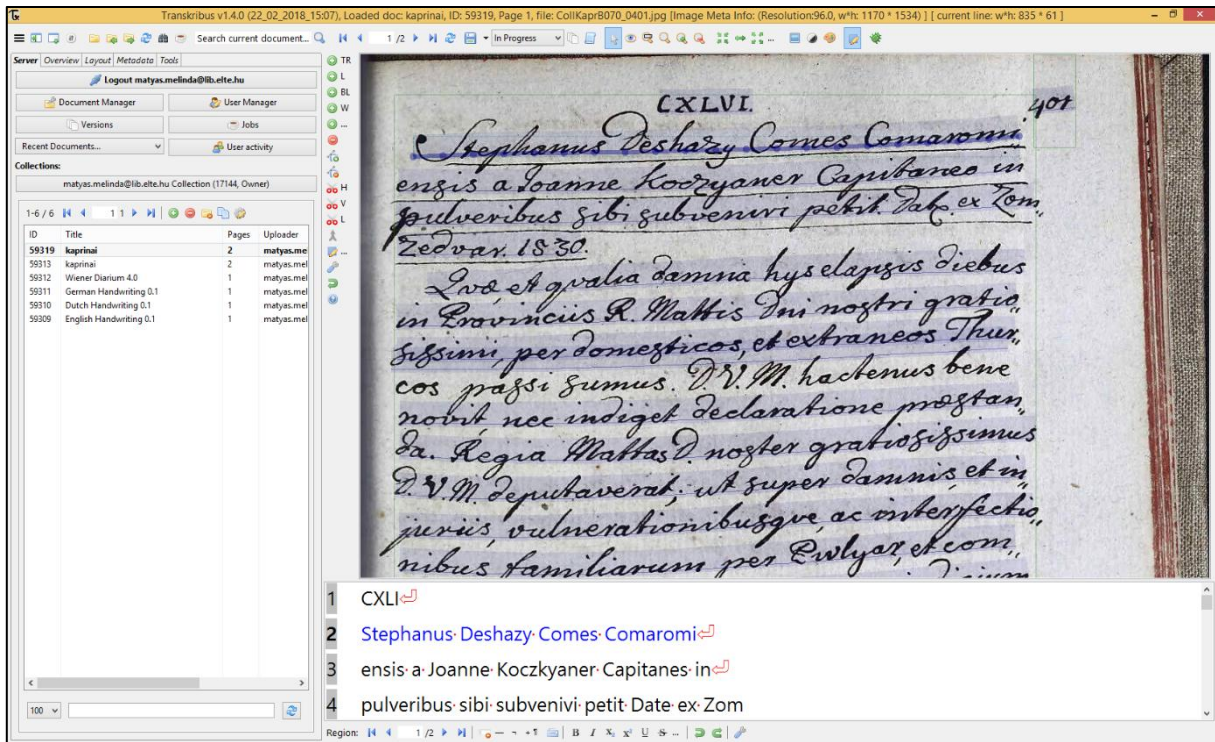
Alapértelmezésben négy, már átírt mintaszöveg áll rendelkezésünkre a lehetőségek tanulmányozásához. Az egyik egy 18. századi folyóirat egy oldala, a másik három pedig különböző kézírás: német, holland és angol nyelven. Természetesen saját dokumentumokat is tölthetünk fel, a rendszer automatikusan elvégzi a feladat kiválasztása után (Find Text Regions, Find Lines in Text Regions) a sorokra való tagolást, ezután kezdhetjük el sorról sorra átírni a szövegünket (a különleges karakterek bevitelét virtuális billentyűzet segíti). Az átírás funkció kinézetére a 6. ábrán mutatunk példát. Az automatikus kézirat-felismerést akkor használhatjuk, ha nyomtatott szövegben (pl. régi folyóiratok) legalább 5000 szót, a kézirat esetében legalább 15 000 szót átírtunk. Ha már rendelkezünk ennyivel valamelyik típusból, akkor szükséges egy emailt küldeni a fejlesztőknek ahhoz, hogy aktiválják az automatikus felismerés tanuló modulját. Az

átírást az online felületen web-böngészőből is éle-
síthetjük, viszont a számítógépről indított szoftver-
ben több funkciót érünk el, a webes felületen lé-

nyegében a kollekciónkat tekinthetjük meg és
folytathatjuk vagy elkezdhetjük a transzkripciót.



5. ábra Avicenna: Logica című műve az EVT2 kritikai apparátusában



6. ábra Példa az átírást menetére a READ szoftverrel

A program lehetővé teszi, hogy többen együtt dolgozzunk egy szövegen, ehhez minden felhasználónak kell rendelkeznie Transkribus fiókkal (vagy összekapcsolt gmail-es hozzáféréssel).

A Transkribus szoftverrel egy használatra kész alkalmazást kapunk a régi folyóiratoldalak (amelyekkel nem birkóznak meg az OCR-programok) és kéziratok feldolgozására, átírására. Az elkészült átírást több formátumban exportálhatjuk, például PDF-ben (kétrétegű eredménnyel) vagy TEI XML-ben is. A gépi kézirat-felismerés eddig nem ismert lehetőségeket kínál az egyébként igen munkaigényes kézírásos szövegek kereshetővé tételére könyvtári környezetben is.

OES³⁹

Az OES ötletgazdái úgy látják, hogy nincs megfelelő *tudományos enciklopédia* szerkesztésére alkalmas, ingyenesen elérhető modern keretrendszer. Ezért a céljuk, hogy egy nyílt forráskódú, testre szabható felületet hozzanak létre az enciklopédia szerkesztési munkafolyamatának koordinálására, a tudományos referenzmunkák editálására és folyamatos fejlesztésére. A víziójuk, hogy egy web-alapú, kollaborációs lehetőségeket kihasználó, könnyen megtanulható keretrendszert alakítsanak ki.

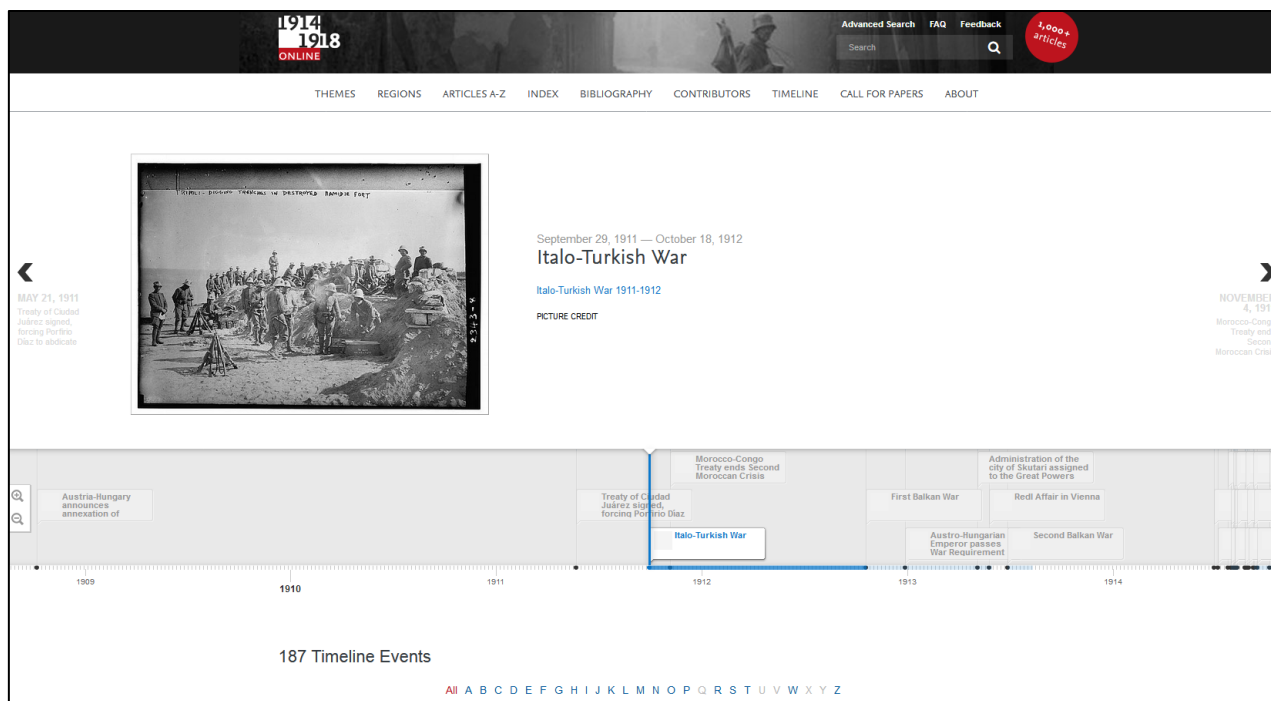
Az OES a következőket kínálja:

- tartalomkezelő eszközök és szövegszerkesztő;
- a publikált cikkek verziókövetése;
- beállítható feldolgozási eszközök a lektoroknak és külső multimédia-források;
- online közösség a szerzők, szerkesztők és más kutatók számára;
- crowd-sourcing lehetőségek úm. kommentelés;
- médiatartalom, linkek, bibliográfiák hozzáadása;
- egyedi belépés és személyre szabható környezet.

A *Freie Universität Berlin* által koordinált innovációt egyelőre három német enciklopédián teszteljük:⁴⁰

1. Online Encyclopedia on German-Greek Entanglements,⁴¹
2. 1914–1918-online. International Encyclopedia of the First World War (7. ábra),⁴²
3. Compendium heroicum.⁴³

Az általánosan használható keretrendszer debütálása 2019-re várható. A fejlesztések eredményeképpen akár egy tudományos igényű, a *Wikipédiához* hasonlóan működő, de lektorált tartalmak elhelyezését biztosító alkalmazás is létrejöhet, amely támogatja a különböző tematikájú enciklopédiák szerkesztési folyamatait és az online közzétételt.



7. ábra 1914–1918-online. International Encyclopedia of the First World War. Timeline nézet

Összefoglalás

A magas szintű *információs kompetenciák mélyre szabott oktatása, továbbá a szövegfeldolgozás és -publikáció* előrehaladott informatikai támogatása létkérdés a közgyűteményi rendszer hatékony működtetéséhez. A digitális bölcsészet ernyője alatt, rendszerint széles körű nemzetközi intézményi összefogással fejlesztett, konzorciális kutatástámogatási innovációk bevezetése így még *eredményesebbé* teheti a *hazai könyvtári hálózat* szolgáltatásait, ezáltal javíthatja *társadalmi elismertségét*.

Hivatkozások

- 1 European Association for Digital Humanities (EADH): <https://eadh.org>
- 2 About. EADH: <https://eadh.org/about>
- 3 EADH Flyer: <https://eadh.org/sites/eadh.org/files/eadh-flyer.pdf>
- 4 Alliance of Digital Humanities Organizations (ADHO): <https://adho.org>
- 5 További tagok: Australasian Association for Digital Humanities (aaDH), European Association for Digital Humanities, Association for Computers and the Humanities (ACH), Canadian Society for Digital Humanities / Société canadienne des humanités numériques (CSDH/SCHN), centerNet, L'association francophone des humanités numériques/digitales (Humanistica) and Japanese Association for Digital Humanities (JADH).
- 6 Digital Scholarship in the Humanities (DSH): <https://academic.oup.com/dsh>
- 7 Annual Conferences. EADH: <https://eadh.org/conferences> A negyvenedik eseményre 2019 márciusában Koppenhágában kerül sor: <http://dig-hum-nord.eu/dhn-2019>
- 8 Education. EADH: <https://eadh.org/education> Itt fontos megemlíteni a Digital Humanities Course Registry adatbázist, melyben Magyarországról az ELTE digitális bölcsészeti kurzusai jelennek meg: <https://registries.clarin-dariah.eu/courses> A digitális bölcsészet művelésére létrejött európai – jellemzően egyetemeken létrehozott – kutatóközpontok is rendszeresen indítanak képzéseket: <https://eadh.org/education/digital-humanities-centres>
- 9 Awards. EADH: <https://eadh.org/awards> 1998-ban a Debreceni Egyetem részesült Roberto Busa díjban: <http://eadh.org/awards/adho-roberto-busa-award>
- 10 Support. EADH: <https://eadh.org/support>
- 11 Projects. EADH: <https://eadh.org/projects>
- 12 #dariahTeach: <http://eadh.org/projects/dariah-teach>
- 13 The Programming Historian: <http://eadh.org/projects/programming-historian>
- 14 Computer Aided Textual Markup and Analysis (CATMA): <http://eadh.org/projects/catma>
- 15 Edition Visualization Technology (EVT): <http://eadh.org/projects/evt-edition-visualization-technology>
- 16 Recognition and Enrichment of Archival Documents (READ): <http://eadh.org/projects/read>
- 17 Open Encyclopedia System (OES): <http://eadh.org/projects/open-encyclopedia-system-oes>
- 18 Transkribus: <https://transkribus.eu/Transkribus>
- 19 MÁTYÁS Melinda: Jezsuita tudósok digitalizált kéziratgyűjteményei az ELTE Egyetemi Könyvtárban és lehetséges kutatási témáik. = Kiszl Péter – Boda Gáborné Köntös Nelli (szerk.): Valóságos könyvtár – könyvtári valóság. Könyvtár- és információtudományi tanulmányok 2016. Budapest, ELTE BTK Könyvtár- és Információtudományi Intézet. 2017. 203-209. p. <https://edit.elte.hu/xmlui/handle/10831/34637>
- 20 Hevenes Kaprinai Pray. ELTE EDIT: <https://edit.elte.hu/xmlui/handle/10831/124>
- 21 #dariahTeach: <https://teach-blog.dariah.eu>
- 22 Szöveggódolás és a Text Encoding Initiative. #dariahTeach: <https://teach.dariah.eu/course/view.php?id=40>
- 23 TextGrid: <https://textgrid.de>
- 24 University of Oxford Archive (OTA): <http://ota.ox.ac.uk>
- 25 The Programming Historian: <https://programminghistorian.org>
- 26 Zenodo: <https://zenodo.org>
- 27 Linked Open Data. The Programming Historian: <https://programminghistorian.org/en/lessons/?topic=lo>
- 28 CATMA: <http://catma.de>
- 29 MEISTER, Jan Christoph et al.: CATMA 5.0 Tutorial. = Rhian LEWIS et al. (Eds.): Digital Humanities 2017. Conference abstracts. Montréal, McGill University & Université de Montréal. 2017. 822-823. p. <https://dh2017.adho.org/abstracts/DH2017-abstracts.pdf#page=822>
- 30 heureCLÉA: <http://heureclea.de>
- 31 Three-Dimensional Dynamic Data Visualisation and Exploration for Digital Humanities Research (3DH): <http://threedh.net>

- ³² Edition Visualization Technology (EVT):
<http://evt.labcd.unipi.it>
- ³³ Digital Vercelli Book: <http://vbd.humnet.unipi.it/beta2>
- ³⁴ Codice Pelavicino:
<https://dig-ed-cat.acdh.oeaw.ac.at/editions/detail/212>
- ³⁵ BATJARGAL, Biligsaikhan – KHALTARKHUU, Garmaabazar – MAEDA, Akira: Creating a Digital Edition of Ancient Mongolian Historical Documents. = Digital Humanities 2018. Book of Abstracts. Mexico City 26-29 June 2018. Alliance of Digital Humanities Organizations, Red de Humanidades Digitales A. C. 534-536. p.
https://dh2018.adho.org/wp-content/uploads/2018/06/dh2018_abstracts.pdf
- ³⁶ READ: <https://read.transkribus.eu>
- ³⁷ tranScriptorium: <http://transcriptorium.eu>
- ³⁸ How To use Transkribus – in 10 steps (or less). RE-AD:
https://transkribus.eu/wiki/images/7/77/How_to_use_TRANSKRIBUS_-_10_steps.pdf
- ³⁹ Open Encyclopedia System (OES):
<http://www.open-encyclopedia-system.org>
- ⁴⁰ Use Cases. OES:
<http://www.open-encyclopedia-system.org/oes/use-cases/index.html>
- ⁴¹ Online Encyclopedia on German-Greek Entanglements. Center for Modern Greece (CeMoG) – Freie Universität Berlin:
<http://www.cemog.fu-berlin.de/en/index.html>
- ⁴² 1914-1918-online. International Encyclopedia of the First World War: <http://www.1914-1918-online.net>
- ⁴³ Compendium heroicum:
<https://www.compendium-heroicum.de>

Beérkezett: 2018. X. 20-án.



Kiszl Péter

habilitált egyetemi docens,
az ELTE BTK Könyvtár- és
Információtudományi Intézetének
igazgatója, Információtudományi
Tanszékének és Könyvtártudományi
doktori programjának vezetője.
E-mail: kiszl.peter@btk.elte.hu



Mátyás Melinda

informatikus könyvtáros, MTMT admi-
nisztrátor,
ELTE Egyetemi Könyvtár és Levéltár,
Egyetemi Könyvtár,
Gyűjteményszervezési Osztály.
E-mail: matyas.melinda@lib.elte.hu