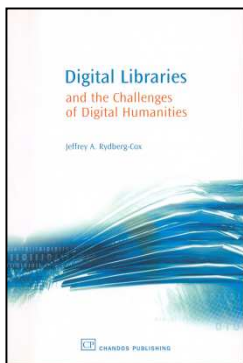


A digitális könyvtár és a digitális humaniőrak



RYDBERG-COX, Jeffrey A.
Digital libraries and the challenges of digital humanities / Jeffrey A. Rydberg-Cox. – Oxford : Chandos Publishing, 2006. – XV, 103 p. ISBN 1-84334-134-4

Jeffrey Rydberg-Cox 2000 óta a klasszika-filológia professzora a Missouri Egyetemen, s a *Persens Digital Library* projektben 1998 és 2000 között történt részvétele alatt különleges tapasztalatokat gyűjtött ahhoz, hogy érvényes és izgalmas mondanivalóval álljon elő a humaniőrak és a digitális könyvtár kapcsolatrendszeréről. Szakmai háttérét az is növelte, hogy 2001 és 2005 között az USA Nemzeti Tudományos Alapja (*National Science Foundation*) és az EU Nemzetközi Digitális Könyvtári Programja által támogatott *Cultural Heritage Language Technologies (CHLT)* projekt egyik felügyelője volt.

A szerző szerint a digitalizálás jelentősége kettős a humán tudományok szempontjából. Egyrészt a kutatók szép számmal hoznak létre digitális anyagokat azzal az elvárással, hogy a könyvtárak őrizék meg ezeket a jövő számára ugyanúgy, ahogyan most megőrzik a könyveket és a folyóiratokat. Ezek a dokumentumok a nyelvészek által létrehozott elemzett korpuszoktól a gazdag, multimédiás hipertextekig sokfélék lehetnek (akár videók, szövegek stb.). Másrészt a digitalizálás nagyon vonzó a diákok, a tudósok és az élethosszig tartó tanulásban résztvevők számára egyaránt, ugyanis ezáltal lehetővé válik az anyagok újszerű szempontok szerint történő vizsgálata, amely új utat nyit a megértés javításához. A digitalizált anyagok gyűjteménye, a digitális könyvtár a humán területen dolgozóknak kihívást jelent, és újfajta hozzáállást követel.

A pre-digitális kor vállalkozásai szinte minden esetben az eredeti szövegek reprodukálásához vezettek. Ezek különböző formákban jelentek meg: reprint, faksimile kiadás, mikrofilm stb. Szövegfordításokat adtak, a szövegek tartalmának népszerűsítő magyarázatait, és készültek jegyzetekkel ellátott, annotált, kritikai kiadások, hogy a problematikus fogalmakat, a nehezen érthető gondolatokat megmagyarázzák. Ugyanezek a kérdések motiválják a digitális könyvtárak könyvtárosait és a humaniőrak digitalizálásával foglalkozó szakembereket: hogyan tegyék hozzáférhetővé a textusokat, hogyan teremtsünk lehetőséget ezek eredményesebb olvasására és megértésére, hogyan tudjuk a szövegeket a kutatók kezébe adni.

A könyv a humán tudományokban létrehozott digitális objektumok teljes életciklusát kívánja bemutatni (létrehozás, hozzáférés, hosszú távú megőrzés). A digitalizálás számos objektuma közül (zenei partitúrák, műalkotások, történelmi tárgyak, régészeti leletek stb.) elsősorban az írott szövegekkel kapcsolatos kérdésekkel foglalkozik.

A bevezetésben a humán területek digitális könyvtári helyzetének, főbb kérdéseinek áttekintését kapjuk. Ezen belül a humán terület négy IT alkalmazásának rövid bemutatását olvashatjuk.

Valley of the Shadow projekt (USA)

<http://valley.vcdh.virginia.edu/>

Primer források digitális archívumai, amelyek a virginiai Augusta megyében és a pennsylvaniai Franklin megyében élő emberek életét dokumentálják az amerikai polgárháború korszakából. Több ezer eredeti okirat, összeírások, naplók, levelek, újságok, amelyek kereszthivatkozásokkal, egymásba ágyazott csatolókkal, korabeli és más szövegmagyarázatokkal, interaktív térképekkel, korábban soha nem látott módon mutatják be az egykori Augusta és Franklin, a polgárháború eltérő oldalán álló két közösség hétköznapijait.

Institut and Museum of the History of Science (Firenze)

<http://galileo.imss.firenze.it/>

A tudománytörténetnek szentelt intézetet és múzeumot a 20. század elején alapították a *Firenzei Egyetem* kezdeményezésére. A múzeum élen jár az új információs technológiák alkalmazásában. Látványos elektronikus szolgáltatásai a virtuális kiállítások, ahol múzeumi tárgyakat, régi objektumokat – mint például Galilei híres körzőjét – lehet virtuálisan használni.

Multimedia Mozart CD-ROM

Gazdag magyarázó és kiegészítő anyaggal ellátott, interaktív multimédiás CD-ROM-ok a *Voyager Company*, majd később a *Calliope Media* kiadásában. Ezek a programok lehetőséget adnak *Beethoven*, *Stravinsky*, *Dvořák* és *Mozart* klasszikus műveinek a részletes interaktív felfedezésére, a zenefolyam és a kotta párhuzamos, audiovizuális megjelentetésére.

Perseus Digital Library

<http://www.perseus.tufts.edu/>

Ez a projekt a digitális technológia egyik korai és újszerű alkalmazása a nyelvek és irodalmak kulturális környezetükben való bemutatásához. Az antik világ tanulmányozásához szükséges források gyűjteményének tartalma elsősorban a régészet, atlaszok, szövegek és fordítások, szövegeszközök és lexikonok köréből kerül ki. 7 millió görög, 5 millió latin szót tartalmaz – angol megfelelőikkel; 30 ezer kép, ezer térkép és az antik szobrászat katalógusa található benne.

A szerző ezen a négy példán keresztül vázolja fel a digitális könyvtárak és a humán területek kapcsolatát, valamint a humán tudományok számítástechnikai kihívásait, a velük kapcsolatos feladatokat. Ezután teszi fel a kérdést: mit is nevezünk digitális könyvtárnak? Megválaszolására a digitális könyvtár különböző kutatók által megfogalmazott meghatározásait veti össze. A digitális könyvtár két fő típusa, az intézményi keretet hangsúlyozó, illetve a tartalom primátusát valló. E meghatározások közül a szerző az utóbbira szavaz, és idézett példákban a digitális könyvtárat a tudós laboratóriumával hozza párhuzamba. A könyvben tárgyalt elveket, elméleteket gyakorlati példákon keresztül mutatja be. A példákhoz *Galilei* „Sidereus Nuncius” című munkájának feldolgozását, különböző kiadásait, fordításait használja fel.

Fontos kiinduló kérdés a digitális könyvtárak létrehozásánál a szövegek hozzáférhetősége. A *második fejezetben* áttekintést kapunk a szövegek géppel olvasható formába való áttételéről, a jelölések, *tagok* fogalmáról és alkalmazásáról. Ide kapcsolódnak a szövegíró módszerek (HTML, SGML, XML) és dokumentumformátumok (TEI, DocBook stb.) magyarázatai. Különösen hasznos a HTML és az XML közötti különbség leírása, mivel a HTML a tudós számára a kevésbé fontos megjelenítés eszköze, a tartalom leírásához az XML-t célszerű alkalmazni.

Példákon keresztül ad képet a kódolás kérdéseiről (ASCII és Unicode szabványok, az Unicode-ban nem megjelenő karakterek, beágyazott és nem szabványos alakzatok, tipográfia), valamint az oldalképeknek a teljes szöveges kereséssel való kombinálásáról. Részletesen ismerteti a már 1987 óta létező szövegkódolási szabványt, a TEI-t (Text Encoding Interactive), amely alkalmas a dokumentum metaadatainak tárolására is.

A *harmadik fejezet* központi témája a digitális technológiák szerepe a primer szövegek megértésében és a kutatók munkájában. Alapvetően az IT-alkalmazásokkal foglalkozik, és ezek közül három területet érint:

- Leírja, hogy a szintaktikai elemzők, elemezett korpuszok és az automatikusan generált hipertextek lehetővé teszik, hogy a felhasználók olyan szövegekkel dolgozzanak, amelyek nyelve számukra nem kellően ismert.
- Tárgyalja az információ-visszakeresési technikák kreatív alkalmazásait az irodalmi anyagok terén (kulcsszavak automatikus kigyűjtése, klaszterezés, megjelenítés).
- Bemutatja az automatikusan generált hipertexteket (nyelvi és fogalmi), amelyek kiemelik a kulcsszavakat, kifejezéseket, és ezeken az ugrópon-
tokon keresztül hozzák kapcsolatba az olvasókat még több részlettel, vagy a szövegre vonatkozó információkkal.

A görög nyelv tezaurusza, a *Thesaurus Linguae Graecae* (TLG) a homéroszi epikától 1452-ig terjedő időszak görög irodalmi szövegeit tárolja (3700 szerzőtől hozzávetőlegesen 91 millió szót

* Galilei 1610-ben, rövid értekezésben jelentette meg korai csillagászati megfigyeléseit Sidereus Nuncius (Csillag-Hírnök) c. munkájában. (Az ismertető megjegyzése.)

tartalmaz). Az 1972 óta létező TLG belső csatoló-rendszerével megváltoztatta a klasszika-filológusok kutatási szokásait.

Számtalan eszköz, vizuális megjelenítési technika létezik, de ezek közül keveset integrálnak a digitális könyvtárakba. A szerző felhívja erre a könyvtárosok figyelmét, és ösztönzi őket ezeknek a könyvtári környezetbe történő beépítésére. Ebben a fejezetben is számos gyakorlati példát kapunk Galilei művének feldolgozása kapcsán.

A *negyedik fejezet* a humán számítástechnikai tudományok lehetőségeit kutatja, megvizsgál néhány projektet, amelyek a digitálisan támogatott humán tudományok vállalkozásaiból jöttek létre (pl. *Oxford English Dictionary* = *OED*). Olyan területekről esik szó, mint az irodalom és a szókincs mennyiségi vizsgálatai, a szótárak készítése és más hagyományos referenciamunkák. Az IT szerepét kutatja a lexikográfiában és a szövegelemzés mennyiségi megközelítésében. Ezek a területek többségükben szintén ki vannak zárva a digitális könyvtárak infrastruktúrájából. Feladatuk nem a zárt kereskedelmi rendszerekhez való hozzáférés, hanem inkább megállapodás a tartalomszolgáltatókkal szabványos, szöveges adatbázisokhoz való hozzáférésre, amelyek olyan rendszerekkel szolgálnak, ahol a kutatóknak lehetőségük van az előzőekben említett vizsgálatok, kutatások megvalósítására.

Az *utolsó fejezet* a digitális források hosszú távú megőrzésének módjait tárgyalja. Foglalkozik az adaptáció problémájával, amivel számos jól ismert intézményi tárház, repozitórium szembekerült. Bizonyítja, hogy azok az intézményi tárházak, amelyek nem integrálják megfelelő mértékben a digitális eszközöket, kevés motivációt adnak a humán területek számítógépes kutatóinak, akik szeretnének közreműködni a digitális tárházak munkájában.

Végül összefoglaló gondolataiban a szerző rávilágít arra, hogy a számítástechnikai eszközökkel támogatott humán tudományok magukban rejtik a primer források segítségével történő tanulást, kutatás változásait, megújuló formáit. Az ismertetett projektek a források új fajtáit szolgáltatják a széles közönség számára. Megadják a lehetőséget olyan kérdések feltevésére a különböző szövegekkel kapcsolatban, amelyekre a hagyományos nyomtatott források nem adnak választ.

A *tagelt* szövegek létrehozása a hozzáférés feltétele. A humán területek kutatói többségükben átvették azokat a szabványokat, amelyeket a TEI alkotott az irodalmi szövegek XML-ben történő kódolására. Ezek a szövegek olyan archív verzióként szolgálnak, amelyek a megjelenítéstől függően transzformálhatók különböző formátumokba, például HTML-be. A rendszer, hasznát tekintve megéri a ráfordítás költségét. Lehetővé teszi, hogy a *tagek* félreérthetetlenül leírják az egyes elemeket, míg a tipográfiai kódokon keresztül közvetített információ alkalmazása körülményesebb, nehezebb.

A kódolásnak ez a módja hasznos a hosszú távú megőrzés szempontjából is. A TEI-konform XML ugyanis nem épít saját szabványra, vagy nem igényel speciális hardver-szoftver kombinációt, különböző típusú számítógépeken használható. Ezen kívül a számítógép-technológiák fejlődésével a szövegek könnyen konvertálhatók egyik megjelenítési formátumból a másikba. Jelenleg létező megjelenítő-technológiákon alapuló példa erre, hogy a TEI-konform XML szöveg könnyen átalakítható PDF vagy WAP dokumentumformátumba, könnyebben, mint a HTML. Ezen kívül, ha egyszer már kódolták ezeket a szövegeket, akkor a számítástechnikai háttérrel támogatott új, digitális humán tudományok bázisául szolgálnak, amelyek működése papíralapú, nyomtatott szövegekkel nehezen megoldható, vagy éppen lehetetlen lenne.

Nem túl sok azoknak a digitális könyvtári rendszereknek a száma, amelyeket azért terveznek, hogy megkönnyítse a kutatásoknak ezt a fajtáját. A *JSTOR* folyóirat-archívum adatbázis például kiválóan működik, aminek egyik oka, hogy központjában a nyomtatott folyóiratok funkciójának lemásolása áll. A tárgyalt gyűjtemények használatának azonban nem ez a feladata. Mivel az intézményi tárházak és digitális könyvtárak többsége nem rendelkezik ezekkel a feltételekkel, kevés érdeklődést váltottak ki azok között a kutatók között, akiknek leginkább használni kellene őket. A digitális könyvtárak feladata megérteni, hogy ezek a lehetőségek segítik a „nehéz anyagok” feldolgozását, tanulását, ezért módszereiket, eszközeiket integrálniuk kell eszköztárukba.

Czegledi László
(Eszterházy Károly Főiskola
Médiainformatika Intézet)