

olyan stresszhelyzetet, amelyben a függő személy inkább több anyagot használ fel. A nyomozó úgy jelenítette meg, hogy a racionalitás világából ez nem érthető meg, hogy a szerhasználó számára a tudatmódosítás, a misztikum mindent felülír. Bajzáth érdekesnek találta a találkozást. Úgy vélte, kikerülve a szerhasználók világából, az egykori nyomozó már nem volt számára félelmetes.

Bajzáth szerhasználó múltjával kapcsolatban számos fotót és zárójelentést töltött fel a Facebook-oldalára. A fényképek sajátos emlékezési mechanizmusokat indítanak be a társaság egykori tagjainál. Az emlékező közösség tagjai megpróbálják beazonosítani a képek szereplőit, illetve ironikusan viszonyulnak önmagukhoz. Mint fentebb említettem, azok közül, akik a szerhasználó társaság tagjai voltak, sokan meghaltak. Az egykori barátokra emlékezés is megjelenik a kommentekben, akár a "R.I.P." kommenttel, de konkrét történetek felidézésével is, a hiány kifejezésével vagy emlékező zeneszám feltöltésével.⁴⁰

Összegezve: tanulmányomban rövid nemzetközi kitekintés után, a Magyar Nemzeti Levéltár Mark Pittaway Oral History Gyűjteménye alapján két konkrét példát választva elemeztem, milyen hatással van a hely, illetve a társasági szituáció az emlékezés mechanizmusaira. Ahogy Horváth Sándor megállapította fentebb idézett tanulmányában, az oral history a másodlagos szóbeliség korában, a tömegmédiával és a nyilvános történelem térhódítását kihasználva hozzájárul ahhoz, hogy a társadalom jóval szélesebb rétegei szólhassanak hozzá a saját múltjukról szóló diskurzusokhoz.⁴¹ Indián esetében a Nagyfa, mint hely kiemelt szerepet játszott az életút elbeszéléseiben, mivel az ellene indított koncepciók per is ide köthető. Radics Bélával kapcsolatos emlékeinek elbeszélése szintén hely- és szituációfüggő: nagyközönsége előtt tartott beszédében élt a nagyítás és pátosz retorikai eszközeivel. A szocialista időszak szerhasználója és az egykori nyomozó, amikor kutatói szorgalmazásra 25 év után újra találkozott, emlékeiket más-képpen kontextualizálták, mint a külön-külön készített interjúkban. A nyomozó visszaemlékezéseiben hangsúlyosabbá vált a korabeli rendőrség és társadalom igazoló jellegű ábrázolásmódja. Az egykori szerhasználót normatív alapon kivételként jellemezte, mivel nem bukott el, hanem a szerhasználatot abbahagyva ma szociális munkásként dolgozik és rendezett családi életet él. A szerhasználó számára ez egykori traumatikus rendőrségi élmények feloldásában is segítséget nyújtott a találkozó. Míg a Nagyfa-galeri kommunikatív emlékezetének kollektívba fordulásánál feltételezhetjük az egykori galeri tagok társadalmi rehabilitációját, addig a szocialista időszak drogpolitikája és a szerhasználói szubkultúrák nem normatív, hanem interpretatív alapon lehetnek részei az emlékezetnek.

⁴⁰ Például: „nagyon hiányzol szívem”, „jaaj de rohadt dolog hogy ez lett a vége....”

⁴¹ Horváth i.m. 38.

DIGITÁLIS FORRÁSOK – DIGITÁLIS FILOLÓGIA

KALCSÓ GYULA

A számítástechnikai eszközök felhasználása ma már a humán tudományokban is mindennapos gyakorlatnak számít. A számítógép, valamint a világháló által kínált lehetőségek azonban hozzáértő, felkészült, gondos előkészítő munka nélkül a humán tudományokban (sem) volnának kiaknázhatók. Az informatikai eszközök bölcsészettudományi implementálásának ma már külön módszertana van, sőt: egyesek szerint önállósulóban lévő tudományterület. Az angolszász világban *digital humanities*nek, azaz magyarul kb. *digitális bölcsészet*nek nevezett terület hazánkban is elterjedt: az ide kapcsolódó tevékenységek közül jó néhányat már évtizedek óta művelnek (nyelvtechnológia, számítógépes nyelvészet), viszonylag önálló kutatási irányként azonban csak az utóbbi években jelent meg (l. a Pázmány Péter Katolikus Egyetem, valamint a Debreceni Egyetem digitális bölcsészet c. MA-képzését, vagy a Szegedi Tudományegyetemen megalakult Digitális Kultúra és Elméletek Kutatócsoportot).

A digitális bölcsészet rendkívül szerteágazó irányzatai és problémái közül jelen tanulmány az egyik legalapvetőbbet: a szöveg- és dokumentumdigitalizálást, valamint az ezzel kapcsolatban felmerülő digitális filológiai kérdéseket érinti. A *digitális filológia* kifejezés ma már egyértelműen az elektronikus gépek által kettős számrendszerben kódolt szövegekkel kapcsolatos tudományos feladatok megoldását jelenti. A *digitális* jelző azonban eredetileg a latin *digitus* ('ujj') szóból származik, és a változó valóságot diszkrét jelekkel (pl. számokkal) leképező (kódoló) dolgot jelöl. Ilyen értelemben az emberiség első nagy digitalizálási hulláma a hangjelölő írások feltalálása volt, hiszen akkor rögzítettek először egy folytonosan változó entitást (hangok sorozatát) diszkrét jelekkel (csak ezek nem számok, hanem másfajta vizuális jelek voltak). A hagyományos *filológia* fő feladata éppen az írásban rögzített szövegek tudományos vizsgálata: a szöveg eredetiségének, a hiteles szövegváltozatoknak a megállapítása, valamint a szövegek tudományos igényességű rögzítése és publikálása. A digitális filológia azonban nemcsak abban több ennél, hogy a szövegek tárolása és közzététele már nem papíron történik, hanem elektronikus számítógépeken. A számítástechnikai eszközök új lehetőségeket jelentenek a tudományos igényű szövegkezelésben.

A digitális filológia egyik legfontosabb feladata a hagyományos filológia által már kidolgozott forrásközlési metódusok számítógépes adaptálása. A tudományos igényességű forrásközlési eljárások közül kiemelten fontos a *betűhív kiadás*, a *kritikai kiadás*, valamint a *fakszimile*. A betűhív kiadás az eredeti szö-

veg grafematikailag pontos nyomtatott betűs átíratát jelenti, a kritikai kiadás a textust részletes kiegészítő magyarázatokkal kísérv, a forrás minden jellemzőjét részletesen feltárv és kommentálv, tudományos célokra szánt közlési forma, míg a faksimile az eredeti dokumentumról készült fotómásolat. Ezen forrásközlési módok átültetése digitális (és ma már sokszor főként online) közegbe jórészt megoldott feladat, azonban a rendelkezésre álló eszközök segítségével az egyszerű közlés által kínált lehetőségeknél jóval több is megvalósítható volna.

A digitális forrásközlésben alapvetően két eset különíthető el: a *szövegközlés* és a *dokumentumközlés*. Az első esetben valamely szöveg történetileg kialakult változatainak figyelembevételével (vagy sajnos sokszor anélkül) kialakítanak egy szövegváltozatot, amelyet közlétesznek.¹ Ez lényegében megegyezik a hagyományos szövegkiadással, a digitális közeg ez esetben semmilyen problémát nem jelent. Ennek fejlettebb változata a digitális kritikai kiadás, amely a hagyományoshoz hasonlóan szövegkritikai apparátust használ, tudományos igényességgel annotálja a szöveget, eligazítást nyújt a különböző szövegváltozatok között.² A hagyományos kritikai kiadással szemben (amely a szövegkritikai megjegyzéseket jegyzetek formájában tudta csak közölni) a digitális kritikai kiadásnak nagy előnye a digitális hipertextualitás lehetősége (a szövegkritikai megjegyzéseknek nem kell feltétlenül megbontaniuk a szöveg linearitását, akár az is megoldható, hogy a jegyzetek csak gombnyomásra jelennek meg, vagy a különböző szövegváltozatok között is lehet kattintással váltogatni). A digitális kritikai kiadások létrehozásakor már valamilyen *jelölőnyelvet* kell használni (l. lentebb).

A történeti források publikálásakor természetesen jóval gyakoribb a dokumentumközlés, amelynek Tószegi Zsuzsanna szerint háromféle szintjét lehet elkülöníteni. „Ha a digitális változat tulajdonságait az eredeti műhöz viszonyítjuk, három szintet különböztethetünk meg:

A reprodukív szint a forrásmű formai és tartalmi jegyeit egyaránt tükrözteti (az esetleges hibákkal, eltérésekkel együtt). A digitalizált változat az eredeti művel gyakorlatilag egyező hatást vált ki, azzal szinte egyenértékű. Ebbe a csoportba elsősorban a faksimile állományok (képfájlok) tartoznak.

A reprezentatív szint a forrásmű tartalmát helyezi előtérbe, de alapvetően nem változtatja meg a szöveg lineáris olvasatát. Ezen a szinten az analóg szövegből digitalizált szöveget állítunk elő, amelynek információtartalma a számítógép nyújtotta szokásos eszközökkel könnyebben kereshető.

Az interpretatív szinten az eredeti forrás tartalmához hozzáadódik a feldolgozást végző szakemberek tudása és tapasztalata, melynek eredményeként új mi-

¹ Ily módon teszik közzé pl. a Himnusz egy (filológiaiilag több ponton is kifogásolható) szövegváltozatát a nemzeti jelképeinket bemutató internetes oldalon: <http://www.nemzetijelkepek.hu/himnusz-szovegek.shtml>.

² Ilyen pl. az ELTE Magyar Irodalomtörténeti Intézet Reneszanszkutatások Posztgraduális Központja (CHER) és a Bölcsészeti Informatika Önálló Program (BIÖP) műhelyében készült hálózati kritikai kiadás Balassi verseiről: <http://magyar-irodalom.elte.hu/gepesk/bbom/cimlap.htm>.

nőség jön létre. Az eredeti művet kiegészítő elemek (amelyek lehetnek magyarázatok, mutatók, hipertext hivatkozások, vagy a szövegtől eltérő műfajú elemek: hang-, videofájlok stb.) megbontják az eredeti szöveg lineáris egységét.”³

Felosztását azzal kell kiegészítenünk, hogy a digitális dokumentumokra akár egyszerre több szint is jellemző lehet: mind a reprodukzív, mind a reprezentatív szinten lehetséges az „interpretáció”, azaz a *hozzáadott információk (annotáció)* rögzítése. Ily módon tehát a dokumentumdigitalizálás esetében valójában két eset különíthető el: *a dokumentum szövegének a digitalizálása* (ezt nevezi Tószegi reprezentatív szintnek), amely tartalmazhat akár hozzáadott információkat (annotációt) is;⁴ valamint *a dokumentum képének vagy képének és szövegének a digitalizálása (digitális fakszimile)*, mindkét esetben az annotáció lehetőségével.⁵ Amennyiben a dokumentum képét és szövegét is digitalizálják, további két eset különíthető el: a két digitális objektum egymástól függetlenül érhető el, vagy összekapcsolják őket. A dokumentum képét és szövegét összekapcsoló, azokat hozzáadott információval ellátó, valamint a szövegben és az annotációban egyaránt keresést biztosító digitális objektumokat tekinthetjük a digitalizálás legmagasabb rendű formájának.

Bármelyik módot választjuk is, a digitalizálás első lépése a számítógépes szöveg létrehozása. Ez kétféleképpen történhet: szkenneléssel és kézi bevitellel. A tudományos igényességű, betűhív digitális szöveg létrehozása jelenleg kizárólag manuálisan lehetséges, és képzett szakember munkáját igényli. A nehézséget az optikai karakterfelismerés jelenti. A szkennelés során a lapolvasó sorról-sorra, a sorokon belül pedig pontról-pontra haladva letapogatja a másolandó képet, és minden egyes képpontra vonatkozóan rögzít bizonyos információkat (fedettség, szín stb.). A képpontokra vonatkozó információkat a számítógép digitális formában dolgozza fel. A szkennelés során létrejövő képfájlból karakterfelismerő programmal (Optical Character Recognition – OCR)⁶ lehet digitális szöveget előállítani. Jó eredménnyel azonban csak a jól olvasható, mai helyesírással íródott, nyomtatott szöveget lehet szoftverrel felismertetni. A képként beolvasott állományokat be kell tölteni a programba, majd ki kell jelölni azokat a zónákat, amelyeket szöveggént szeretnénk felismertetni. Ha a forrásdokumentum nem tökéletes állapotú, akkor a karakterfelismertetést megelőzően képfeldolgozó programot kell alkalmaznunk pl. a kontraszt növelése érdekében. Az optikai karakterfelismerő programok hatékonysága megfelelő előkészítés esetén, mai helyesírással íródott, jó minőségű nyomtatványok esetén is csak kb. 90% (emiatt

³ Tószegi Zsuzsanna: A szövegdigitalizálás döntési folyamata. Könyvtári figyelő, 2006/2. 245–260. Interneten: <http://epa.oszk.hu/00100/00143/00059/toszegi.html>

⁴ Erre példa a Magyar Antikvakorpusz: <http://korpusz.ektf.hu>.

⁵ A történeti forrásokat publikáló levéltári digitális objektumok többsége valamilyen képfájl, amely esetleg PDF formátumban jelenik meg. Ezek esetében a dokumentum szövegének hiányában egyszerűen tanulmányozhatjuk az eredeti dokumentum képét.

⁶ Ilyenek pl.: Omnipage, Abby FineReader, TextBridge, Adobe Capture, Recognita stb.

van szükség a kézi korrektúrára). A történeti források esetében tehát ez a módszer egyáltalán nem alkalmazható, a kézi bevétel fáradságos és idő- valamint szaktudásigényes munkafázisa sajnos jelenleg nem mellőzhető.

A digitalizált szöveg már önmagában is jelentős könnyebbséget jelenthet a tudományos kutatómunkában, ám a valódi előnyét akkor élvezhetjük, ha hozzáadott információval (annotációval) látják el, valamint egy megfelelő (lehetőleg online) felületen a szöveg bármely elemére és az annotációra is rá tudunk keresni. A digitalizált szövegben elhelyezett annotációt valamilyen módon el kell különíteni magától a szövegtől. A legelterjedtebb megoldás az olyan kódrendszerek használata, amelyek általában valamilyen speciális jelölésmóddal jelelik a hozzáadott információt. Ilyen kódrendszert akár magunk is kidolgozhatunk, ám mégis célszerű valamilyen szabványos, többféle eszközzel is feldolgozható, szabadon hordozható formátumot választani. A legelterjedtebb megoldás az XML, illetőleg valamilyen speciális, XML-alapú kódrendszer használata.

Az XML (eXtensible Markup Language) az 1980-as évek nagyszabású vállalkozásának, az SGML-nek az egyszerűsített változata. Az SGML (Standard Generalized Markup Language) szabványos jelölőnyelv dokumentumok belső szerkezetének leírására, beleértve az egyes elemeket jelölő címkék (szakszóval tagek) definiálásának módját is. A Nemzetközi Szabványügyi Szervezet (ISO) által elfogadott nemzetközi szabvány (ISO 8879:1986). Segítségével elvben bármilyen dokumentum leírható, függetlenül az azt tároló és megjelenítő számítógépes környezettől. Az SGML valójában metanyelv, vagyis formálisan, megadott szabályok alapján leírhatunk vele egy másik nyelvet. Az információt annak tartalma, illetve szerkezete alapján jelöli meg, innen származik a jelölőnyelv elnevezés. Tervezésekor az egyik alapvető célkitűzés az volt, hogy az SGML szabályait követő dokumentumok információvesztés nélkül hordozhatók legyenek az eltérő hardver- és szoftverkörnyezetek között. Manapság az elektronikus formában tárolt dokumentumokban nagyon nehéz – esetenként lehetetlen – megtalálni a számunkra fontos információkat, mivel az azokat kezelő szoftverek nem képesek értelmezni az ember számára értelmes szöveget. Amíg nincsenek nagy teljesítményű, majdnem emberi intelligenciával rendelkező számítógépeink, addig kénytelenek leszünk valamilyen módon megjelölni a szoftverek számára az információkat. Erre szolgál – többek között – az SGML.

Bíró Szabolcs az alábbi szemléletes példával mutat rá a kódolás lényegére: „A World Wide Web egyik legnépszerűbb keresőgépét, mondjuk a Google-t szeretnénk használni. Keresőkérdésünk igen egyszerű: meg szeretnénk tudni, hogy mi is az a Jáva. Ennek a szónak napjainkban már két értelmezése is ismert: Jáva – mint sziget, és Jáva – mint programozási nyelv. Milyen eredményeket kapunk, ha csupán az egyszerű »Jáva« szót adjuk meg? A jelöletlen szövegekben a keresőrobot mindkét értelmezést ugyanolyan »értékűnek« tekinti, tehát a találati listában szigetként és nyelvként is előfordul majd a Jáva szó. Jelölve viszont különbséget tud tenni közöttük:

```
<sziget>Jáva</sziget>  
<programozasi_nyelv>Jáva</programozasi_nyelv>”7
```

Látható, hogy a szövegben szereplő relációjelek az annotációt különítik el magától a szövegtesttől. A címke (tag) nyitóeleme a „<” jellel kezdődik, és a „>” jellel zárul. A záróelemben a „/” jelöli, hogy itt ér véget az a szövegrész, amelynek jelölését az aktuálisan megnyitott címke végzi.

Az interneten található weblapok legtöbbször az SGML-ből fejlesztett egyszerűbb jelölőnyelv, a HTML (Hypertext Markup Language) kódolja. Ennek nagy hátránya azonban, hogy kizárólag a megjelenítendő tartalom formai jellemzőit tudjuk vele megadni (milyen betűtípussal, milyen színben, milyen elrendezésben jelenjen meg stb.), tartalmi kódolásra nem alkalmas.

Az SGML bonyolultsága, továbbá a HTML megjelenítés-orientáltsága miatt 1996-ra a szövegjelölés területének több szakértője úgy gondolta, elérkezett az idő az SGML egyszerűsített verziójának létrehozására, amely a nagyközönség számára vonzóvá tenné az általánosított jelölés alkalmazását. Így jött létre az XML, amely az SGML és a HTML konvencióira és elveire épül, hogy ezáltal egyszerű, ugyanakkor mégis hatékony mechanizmust hozzon létre az információk feldolgozására, tárolására, illetve szolgáltatására. Bíró Szabolcs így foglalja össze a kódolás lényegét: „Az SGML és az XML leíró jelölést alkalmazó jelölőrendszer, vagyis olyan jelölőkódokat használ, amelyek nevekkal azonosítják (kategorizálják) a dokumentumok bizonyos részeit. Az olyan jelölőkódok, mint például <cim> vagy <bekezdés> csupán a dokumentum bizonyos részeinek azonosítására szolgálnak, és mindössze annyit jelentenek, hogy »a következő elem egy cím«, vagy »most egy bekezdés következik«. Az SGML/XML nyelvben a dokumentumok bizonyos célú feldolgozásához – pl. formázott megjelenítéséhez – szükséges utasítások élesen elválnak a dokumentumban található leíró jelölésektől, rendszerint a dokumentumon kívül, külön eljárásokban vagy programokban találhatóak...”⁸

Egy rendkívül egyszerű példa SGML/XML-kódokkal ellátott szövegre:

```
<!-- SGML/XML -->  
<antologia>  
<vers>  
<szerzo>Petőfi Sándor</szerzo>  
<cim>Méz és csók</cim>  
<versszak>  
<sor>Kis méh! te a füvet, fát,</sor>
```

⁷ Bíró Szabolcs: Szövegfeldolgozás XML alapokon. Budapest, 2005. 14.

⁸ Bíró i. m. 20.

```

< sor>S virágokat leped,</ sor>
< sor>Hogy édes kelyheikből</ sor>
< sor>Gyűjthessed mézedet.</ sor>
</ versszak>
< versszak>
< sor>Kis méh! Lidim füvet, fát</ sor>
< sor>S virágokat nem lep,</ sor>
< sor>Mézednél csókja mégis</ sor>
< sor>Mi sokkal édesebb.</ sor>
</ versszak>
</ vers>
<!-- Antológiáról lévén szó, itt további versek kö-
vetkeznek. -->
</ antologia>

```

Az ily módon kódolt szöveggel többféle műveletet is végezhetünk. Egy nyomdai tördelőprogram pl. megfelelő módon tudja kezelni a versszakokat és a sorokat, hiszen a kódok alapján fel tudja ismerni. Egy böngészőprogram ugyancsak megfelelő módon jelenítheti meg a képernyőn. Egy erre a célra fejlesztett keresőprogram pedig megkeresheti és kiírhatja számunkra az általunk keresett elemeket (pl. egy korpuszban szereplő összes vers címét). Az XML az alkalmazható kódok tekintetében semmilyen megkötést nem tartalmaz, „csupán” egy szabványos kódolási szintaxist biztosít. Ezért jöttek létre belőle speciális rendszerek, amelyek egy-egy felhasználási területre ajánlanak nemzetközileg elfogadott és egységes kódrendszert.

Ilyen kódrendszer a Text Encoding Initiative (TEI), amely amelyet 1987-ben három amerikai számítógépes nyelvészeti és irodalmi kutatásokkal foglalkozó tudományos társaság, az Association for Computers and the Humanities (ACH), az Association for Computational Linguistics (ACL), és az Association for Literary and Linguistic Computing (ALLC) indított el. Egy tervező konferenciából nőtt ki, amelyet 1987-ben tartottak a New York melletti Vassar College-ban ezen szervezetek támogatásával. A találkozón harminc reprezentatív szövegarchívum, tudományos társaság és kutatási program képviselői vettek részt, hogy megvitassák egy irányadó kódolási tervezet lehetőségeit, és ajánlásokat készítsenek annak szerkezetére és tartalmára.

Az induló projekt feladata irányvonalak kifejlesztése, terjesztése volt a géppel olvasható szövegek kódolására, közvetíthetőségére, és cserélhetőségére, valamint javaslatok tétele új szövegek kódolására az SGML szabvány alapján. A TEI-t elsősorban általános tartalmú szövegek, szépirodalmi művek, kritikai kiadások, történeti források, illetve előszöveg-átiratok elektronikus feldolgozására alkalmazzák.

Az ajánlások kidolgozása 1988 januárjában kezdődött a már korábban említett szervezetek támogatásával. Első verzióját (1.0) 1990 júliusában mutatták be, amely tartalmazza a TEI P1 nevű dokumentumot. Eredeti címe: Guidelines for the Encoding and Interchange of Machine-Readable Texts (Ajánlások géppel olvasható szövegek kódolására és átalakítására). Megjelenése óta 5 verzióját adták ki, ezek közül a legutóbbi a TEI P5, mely 2007-ben (nyomtatásban 2008-ban) jelent meg (TEI Consortium (szerk.) 2008). A TEI P4 már tartalmazta az XML-támogatást is, tehát a DTD-nek (a kódolási szabályokat, a nyelv jelölő-elemeit és egymáshoz való viszonyukat leíró fájlnak, a dokumentumtípus-deklarációnak) az SGML mellett egyaránt létezik XML és XML Schema változata. Azért ajánlják az XML-t, mert az jóval egyszerűbb, rugalmasabb, valamint számos szoftvereszköz támogatja.

A TEI 1999-től konzorciális keretek között működik, fejlesztésében mára számos tudományos társaság és tanszék vállal szerepet, évente konferenciákat tartanak, az egyes részterületeket – például a kéziratok kritikai kiadását vagy a karakterkódolást – munkabizottságok vizsgálják. Bár formálisan csak egyetlen magyar tagja van a konzorciumnak (a Szegedi Egyetem Informatikai Tanszék-csoportja), számos projekt használja (pl. a MEK, a BIÖP „gépeskönyvei”, hálózati kritikai kiadásai stb.).

A TEI nem szabvány, semmit nem ír elő kötelezően, „csupán” egy ajánlás. Hogy ki mit használ fel belőle, az elsősorban a konkrét feladat függvénye. A TEI kidolgozottsága azonban garantálja, hogy a legszigorúbb filológiai követelményeknek is megfelelhessen az, aki ezt az utat választja. A tudományos célú szövegkutatásban ma aligha van olyan kódolási szempont, amelyre ne találnánk megoldást a TEI-ben. (A TEI történetére nézve l. Bíró 2005, valamint a konzorcium honlapját.⁹)

A TEI-XML nagy előnye, hogy az ideális elektronikus dokumentummal szemben támasztott követelményeknek maximálisan megfelel. Olyan szabványos formátumot biztosít, amelyet egységesen megjeleníthetünk böngészőben, könnyedén konvertálhatunk belőle hordozható szövegformátumot (PDF-et), valamint a felhasználás céljainak megfelelő annotációval láthatjuk el a szövegeket (fontosnak tartott tartalmi elemek, pl. ábrák, táblázatok, képek stb. kódolása és még számos egyéb). A TEI-XML feldolgozására kifejlesztett eszközökkel bárki képes információkat nyerni a szövegekből, nem kell hozzá semmilyen sajátos kódrendszert ismernie.

A TEI-XML lehetőséget biztosít az eredeti dokumentumról készült képfájlok és a digitalizált és annotált szöveg összekapcsolására.¹⁰ Ez úgy lehetséges, hogy

⁹ <http://www.tei-c.org/About/history.xml>

¹⁰ A szöveg és a kép összekapcsolására alkalmazott eljárás az ún. kétrétegű PDF technológia is. Azonban a PDF-fájlok megjelenítéséhez speciális szoftver szükséges, továbbá a PDF nem támogatja az annotáció kezelését.

a képfájl megfelelő területeit a szövegben elhelyezett címkék segítségével a szöveghez kapcsolják. A területek megjelölését és kóddal a szöveghez kötését általában erre a célra kifejlesztett programokkal lehet elvégezni. Ily módon lehetőség van arra, hogy a keresőfelületek ne csak a digitalizált szöveg szöveget, hanem a dokumentumról készült kép megfelelő részét jelenítsék meg. A dokumentumdigitalizálásnak ez a módja kétségtelenül a legfejlettebbnek tekinthető, azonban speciális humánerőforrás- és időigénye miatt sajnos jelenleg csak viszonylag kevés projektben alkalmazzák. A jövő azonban mindenképpen az ilyen jellegű forrásközlésre: a hagyományos kritikai kiadások előállítására sem volt kevésbé időigényes, ráadásul a digitális változatok állandóan módosíthatók, javíthatók, bővíthetők.

Irodalom

- Bartók István–Golden Dániel–Horváth Iván et al.: Digitalizálás. Magyar tudomány, 2006/7. 831–836. Interneten:
<http://www.matud.iif.hu/06jul/09.html>
(A letöltés ideje: 2012. november 2.)
- Bates, Chris: XML: elmélet és gyakorlat. Budapest, 2004.
- Bíró Szabolcs–Kora András: Kulcs az SGML-hez. Budapest, 2004.
- Bíró Szabolcs: A szövegfeldolgozás modern eszközei – az SGML és XML nyelvek. Tudományos és Műszaki Tájékoztatás, 2004/10. 453–459. Interneten:
http://tmt.omikk.bme.hu/show_news.html?id=3733&issue_id=455
(A letöltés ideje: 2012. november 2.)
- Bíró Szabolcs: Szövegfeldolgozás XML alapokon. Budapest, 2005. Csak interneten elérhető:
<http://www.tankonyvtar.hu/informatika/szovegfeldolgozas-xml-080906-159> (A letöltés ideje: 2012. november 2.)
- Bradley, Neil: Az XML-kézikönyv. Budapest, 2005.
- Chapman, Stephen: Managing text digitisation. Online Information Review, 2003/1. 17–27. Magyarul Lepp Tünde fordításában: A szövegdigitalizálás menedzselése. Tudományos és Műszaki Tájékoztatás, 2003/8. 325–330. Interneten:
http://tmt.omikk.bme.hu/show_news.html?id=2271&issue_id=78
(A letöltés ideje: 2012. november 2.)
- Golden Dániel–Tóth Tünde–Turi László: Virtuális örökkévalóság: objektumok a digitális könyvtárban. Tudományos és Műszaki Tájékoztatás, 1998/8–9. 299–314. Interneten:
http://tmt.omikk.bme.hu/show_news.html?id=2017&issue_id=3
(A letöltés ideje: 2012. november 2.)

- Hász-Fehér Katalin: A filológia diszciplináris helyzete. *Helikon*, 2000/4., 469–480.
- Hockey, Susan: *Electronic Text in the Humanities: Principles and Practice*. Oxford, 2001.
- János István: A digitális filológia problémái. In: *A Herman Ottó Múzeum Évkönyve XLIII*. Miskolc, 2004. 549–555.
- Koltay Tibor: Virtuális, elektronikus, digitális. Elméleti ismeretek a 21. század könyvtárához. Budapest, 2007. Interneten:
<http://www.tankonyvtar.hu/konyvtartudomany/virtualis-elektronikus-080906-82> (A letöltés ideje: 2012. november 2.)
- Kora András: Szemben a változással: időtálló fájlformátumok, szövegfeldolgozás, digitalizálás. In: *Negyvenéves a szombathelyi könyvtárosképzés*. Szombathely, 2002. 116–121.
- TEI Consortium (szerk.): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Oxford – Providence – Chralottesville – Nancy, 2008.
Interneten: <http://www.tei-c.org/Guidelines/P5> (A letöltés ideje: 2012. november 2.)
- Tószegi Zsuzsanna: A szövegdigitalizálás döntési folyamata. *Könyvtári figyelő*, 2006/2. 245–260. Interneten:
<http://epa.oszk.hu/00100/00143/00059/toszegi.html> (A letöltés ideje: 2012. november 2.)