

Bujdosóné Dani Erzsébet

Helyi testületi anyagok bibliográfiai leírásának problémái

Az 1990-es rendszerváltást követően, a demokrácia kiszélesedésével jelentősen megnőtt a közélet iránt érdeklődők száma, az emberek egyre gyakrabban kíváncsiak a szűkebb pátriájukra, városukra, községükre vonatkozó döntésekre, melyek meghatározzák azt, hogy hogyan alakul közérzetük, hogyan változnak lehetőségeik. Az önkormányzatok nyilvános helyen, a könyvtárakban tették hozzáférhetővé a települések életét alakító rendeleteiket, határozataikat, új, felelősségteljes feladatot ruházva ezzel a közkönyvtárakra. A helytörténeti gyűjtemény e rohamosan gyarapodó speciális részét a kutatók, érdeklődők csak akkor használhatják eredményesen, ha feltárása és rendszerezése ezt lehetővé teszi. A cikk a megvalósítás szerteágazó problémakörével foglalkozik, megismertet az egyik lehetséges jó megoldással.

Az önkormányzatiság szerkezetének felállítása óta, a helyi önkormányzatok testületeinek munkáit összesített formában, jegyzőkönyvként őrzik meg. A jegyzőkönyvek 5 példányban készülnek el, amelyekből egy példány a *Pest Megyei Közigazgatási Hivatalba*, négy pedig a város különböző intézményeibe kerül:

- *Polgármesteri Hivatal Irattára*,
- a címzetes főjegyző hivatala,
- a városi levéltár,
- a városi könyvtár.

A jegyzőkönyveket a könyvtárban lefűzik, és időrendi sorrendben a helytörténeti anyag szerves részét képezik. Keresni bennük ebből adódóan nagyon nehézkes, és kizárólag az adott testületi ülés dátuma szerint lehetséges. A megjelenés gyakorisága nem állandó, mivel gyakran előfordul, hogy a havi egy rendszeres testületi ülésen kívül rendkívüli üléseket is rendeznek.

Az anyag teljesen feltáratlan, nem felel meg a modern információkeresés feltételeinek, ezért valamilyen formában ezen a téren is meg kellene oldani a minőségi tájékoztatást. A dokumentumok fontos információforrást jelentenek a helytörténeti kutatás, az állampolgári tájékozódás, a közügyek helyi intézése szempontjából.

A testületi jegyzőkönyvekről

Ahhoz, hogy a helyi testületi anyagok bibliográfiai leírásával foglalkozhassunk, tisztában kell lennünk

a testületi ülés jegyzőkönyveinek részeivel. Ezek az alábbiak:

- nyitóoldal;
- ezt követi az előterjesztések felsorolása a tárgyalás sorrendjében (itt sorolják fel a napirend előtti felszólalásokat is);
- a harmadik rész az ülés leírása a megszavazott napirendi pontok sorrendjében; ez a rész tartalmazza a hozzászólásokat is;
- végül a rendeletek következnek, és azoknak a határozatoknak a számmal ellátott teljes szövege, amelyekről szavazás született. Itt található a jelenlévők aláírási íve is.

Nehéz kérdés a dokumentum típusának meghatározása, mivel ez két szempontból is megközelíthető: egyrészt a kéziratok oldaláról, másrészt az időszaki kiadványok szemszögéből.

Kézirat: minden kézzel írt feljegyzés, tanulmány, könyv és egyéb írás. Kéziratként kezeljük és gyakran így nevezzük a néhány példányban írógéppel sokszorosított dokumentumot is.¹ A kéziratoknál kétféle szabványcsoport áll rendelkezésre: a könyvtárosok által használt valamelyik MARC, vagy a kézirat- és levéltári használatra tervezett EAD (*Encoded Archival Description*).

Időszaki kiadvány, periodikum: olyan előre meg nem határozott időtartamra tervezett kiadvány, amely egymást követő részegységekből (számokból, füzetekből, kötetekből, évfolyamokból stb.) áll; ezeket rendszerint számozásuk, keltezésük, kronologikus vagy egyéb megjelölésük különbözteti meg egymástól.²

Az ISBD/CR szabvány szerint új meghatározások születtek a dokumentumok különféle kategóriákba sorolásához, ennek megfelelően lett az időszaki kiadványok leírási szabványából, az ISBD/S-ből ISBD/CR, vagyis folytatódó források (continuing recourses) leírási szabványa (1. ábra).³

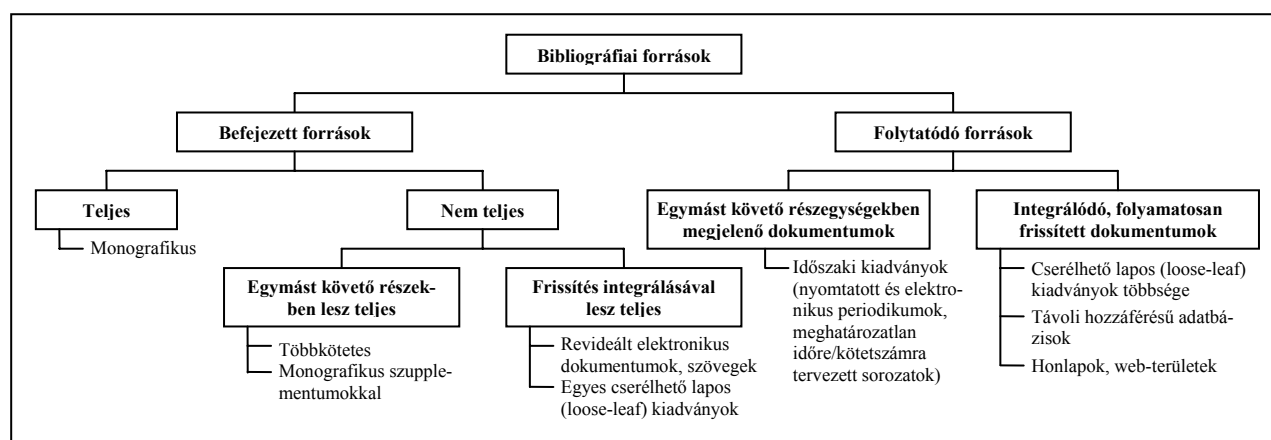
A definíciók szerint a példányszám vonatkozásában a testületi anyag definiálható kéziratként, de a megjelenés, szerkezeti összetétel alapján időszaki kiadványként is (kivéve a címdoldal, nyitóoldal megéléte).

Testületi jegyzőkönyvek feldolgozása

Megnéztem különböző könyvtárakban, hogy hogyan oldják meg ezt a problémát. Kiderült, hogy

sehogyan, mivel az esetek nagy százalékában nem találtam ilyen jellegű munkát. Helyi rendeletekkel, jegyzőkönyvekkel, előterjesztésekkel találkoztam, de ezek rendszerint az adott település honlapján található meg, a teljesség igénye nélkül. Általános gyakorlat, hogy a testületi üléseken született döntésekről valamilyen nyomtatvány jelenik meg, és az abban szereplő rendeletek, határozatok címei jelennek meg a honlapon. Egy rendhagyó esettel találkoztam Hévíz város honlapján, ahol a városi rendeletek és határozatok letölthetők pdf vagy doc formátumban.

A Katona József Megyei Könyvtárban az Olib 7 integrált könyvtári szoftvert használják, és ott megtalálhatók az információhordozó szerinti szűkítésben a helyi jogalkotás adatai (2–5. ábra).



1. ábra A bibliográfiai források kategóriái

Kecskemét. Közgyűlés	
05068-039/1998. Előterjesztés : Kecskemét Megyei Jogú Város Önkormányzata 1998. évi költségvetésének módosítása. – 1998 jún. 3-i Közgyűlés. – 1998-JÚN-03	
Kecskemét	
Költségvetés	
Testületi tevékenység	

2. ábra Testületi jegyzőkönyvben szereplő előterjesztés cédulaformátumú leírása

Cím	05068-039/1998. Előterjesztés: Kecskemét Megyei Jogú Város Önkormányzata 1998. évi költségvetésének módosítása. – 1998 jún. 3-i Közgyűlés	Helyi jogalkotás adatai
Nevek	Közgyűlés Kecskemét. (közread.)	
Kiadás	1998-JÚN-03	
Tárgyszavak	Kecskemét Költségvetés Testületi tevékenység	

3. ábra Testületi jegyzőkönyvben szereplő előterjesztés Olib 7-ben megjelenő leírása

Kecskemét. Közgyűlés

1/1996. (I.10.) KH. sz. határozat Az ülés napirendje. - 1996 jan 10-i Közgyűlés jegyzőkönyve. – 1996-JAN-10

Napirend

4. ábra Testületi jegyzőkönyvben szereplő határozat cédula formátumú leírása

Cím	1/1996. (I.10.) KH. sz. határozat Az ülés napirendje. - 1996 jan 10-i Közgyűlés jegyzőkönyve	Helyi jogalkotás adatai
Nevek	Közgyűlés Kecskemét. (közread.)	
Kiadás	1996-JAN-10	
Tárgyszavak	Napirend	

5. ábra Testületi jegyzőkönyvben szereplő határozat Olib 7-ben megjelenő leírása

A tervezés kivitelezhetőségét nehezíti az is, hogy a kézi feldolgozást követően célszerű lenne gépi adatfeldolgozásban is gondolkodni, hiszen a retrospektív feldolgozás elkészülte után az lenne kívánatos, hogy az állandó feldolgozottság állapota a lehető legnagyobb legyen. Könyvtárunkban a Szikla 21 integrált könyvtári szoftverrel dolgozunk, a retrospektív konverzió régen megtörtént (a könyvállomány esetében), a gépi kölcsönzés már több mint nyolc éve élő gyakorlat. A helytörténeti gyűjtemény feldolgozása az adatbeviteli modul cikkúrlapján történik. A fentieket figyelembe véve, a Szikla21 integrált könyvtári szoftver alkalmatlan az ilyen jellegű feldolgozásra.

A legcélravezetőbb megoldás a teljes szövegű adatbázis létrehozása lenne. Ennek érdekében az XML⁴ alapokon nyugvó szövegfeldolgozást és a Dublin Core⁵ (DC) metaadat-elemeit választottam, és ennek a feladatnak a fényében próbálom bemutatni a lehetséges megoldásokat.

Egy XML-dokumentum lehetséges felépítése testületi jegyzőkönyvnél

A testületi jegyzőkönyvet leíró XML dokumentum két fő részből áll: fejrészből és dokumentumelemekből. A dokumentumelemet gyökér elemnek is szokás nevezni.

A fejrész:

```
<?xml version='1.0' encoding='ISO-8859-2'?>
```

```
<!--Fájlnév: Testületi jegyzőkönyvek.xml-->
```

Az első sor az XML deklaráció, ami azt fejezi ki, hogy ez egy XML dokumentum, és itt kell megadni a verziószámot, valamint azt a karakterkészletet, amit használunk, annak érdekében, hogy használhassuk a magyar ékezetes betűket. Az XML deklaráció nem kötelező, bár a specifikáció előírja, hogy a dokumentumnak tartalmaznia kell.

A fejrész második sora jelentés nélküli térköz. Akármennyi jelentés nélküli térköz előfordulhat.

A fejrész harmadik sora egy megjegyzés, amelynek használata növelheti a dokumentum olvashatóságát. Bármilyen szöveget írhatunk a karakterek közé. Az XML feldolgozó figyelmen kívül hagyhatja a megjegyzést, de átadhatja az alkalmazásnak is, átadhatja a weblapszkripteknek. (Az XML feldolgozó az a szoftvermodul, ami beolvassa az XML dokumentumot, és hozzáférést biztosít a dokumentum tartalmához és szerkezetéhez.)

A fejrész tartalmazhatja még a következő (elhagyható) összetevőket is:

- a DTD: Dokumentumtípus-deklaráció (Document Type Definition), amely definiálja a dokumentum típusát, tartalmát és szerkezetét. Ha ez szerepel, akkor az XML deklaráció után kell következnie. A DTD formális szabályokat definiál, amelyek egyebek között lehetőséget nyújtanak egy adott dokumentumosztályban (dokumentumtípusban) felhasználható elemek előzetes megadására. A DTD meghatározza a megengedett elemek egymáshoz fűződő viszonyát, meghatározza az elemek jellemzőit, az ún. attribútumokat, és az általuk felvehető értékeket stb., elkészítve ezzel az XML dokumentumok logikai szerkezetének a

sémáját. Ezzel gyakorlatilag kialakítja a dokumentumelem hierarchiáját és tagoltságát. Tagoltságnak nevezzük annak a mértékét, hogy az elemek tartalma mennyire felosztott gyermek-elemekre. Ahol sok a gyermekelem vagy más néven leszármazott, finom tagoltságról, ahol pedig kevés a leszármazott, ott durva tagoltságról beszélünk.⁶ A fejrész külső DTD, vagy séma, pl. xsd állományra való hivatkozást is tartalmazhat.

- egy vagy több feldolgozási utasítást, amelyben olyan információk találhatók, amelyeket az XML feldolgozó továbbít az alkalmazásnak.

A dokumentumelemek:

<TestületiJegyzőkönyvek>

<Kézirat>

<Cím>1644/2005.(XI.24.)Testületi ülés jegyzőkönyve</Cím>

<Alcím>Nagykőrös Város Önkormányzat Képviselő-testületének 2006. február 10-i rendkívüli üléséről</Alcím>

<Dátum>2006. február 10.</Dátum>

<Szám>1644/2005</Szám>

<Közread>Nagykőrös Város Önkormányzat Polgármesteri Hivatal Aljegyzője</Közread>

</Kézirat>

<Kézirat>

<Cím>1645/2005.(XII.12.) Testületi ülés jegyzőkönyve</Cím>

<Alcím>Nagykőrös Város Önkormányzat Képviselő-testületének 2006. március 12-i üléséről</Alcím>

<Dátum>2006. március 18.</Dátum>

<Szám>1645/2005</Szám>

<Közread>Nagykőrös Város Önkormányzat Polgármesteri Hivatal Aljegyzője</Közread>

</Kézirat>

</Testületi jegyzőkönyvek>

Az XML dokumentum másik fő része egyetlen elem, a dokumentum-, vagy gyökérem, amely további beágyazott elemeket tartalmazhat. Ezek az elemek építik fel a dokumentum logikai szerkezetét és hordozzák az információtartalmát. Egy tipikus elem nyitórészből, elemtartalomból és zárórészből áll.⁷ Tehát a gyökérelemen belül az egyes jegyzőkönyvek jelentik a gyökérelembe beágyazott elemeket.

Az XML dokumentum jelen esetben felállított szabályai:

- A testületi jegyzőkönyvek csak jegyzőkönyveket tartalmazhatnak, semmi más.
- A jegyzőkönyvnek van címe és alcíme, dátuma, száma és közreadója.
- A jegyzőkönyvnek vannak napirendi pontjai, azon belül sorok. Ezek a sorok csakis napirendi ponton belül állhatnak.
- Egy napirendi pontot egy másik napirendi pont követ.
- A jegyzőkönyvnek vannak határozatai, amelyek sorokból állnak. Ezek a sorok csak határozaton belül állhatnak.
- Egy határozatot egy másik határozat követ, vagy egy újabb jegyzőkönyv.

Ezek a szabályok szabadon felállíthatók, megváltoztathatók, csak arra kell figyelni, hogy mindig következetesen kell használni őket, és a munka elkezdése előtt kell megfogalmazni, elfogadni, és a véglegesen elfogadott változathoz kell kiindulni, következetesen ragaszkodni. Ezek a szabályok megfogalmazhatók egy DTD-ben, vagy sémában is, amelynek használatával ellenőrizhetővé válik az XML dokumentumok érvényessége.

A metaadatok jelentősége az információkeresésben

A könyvtári gyakorlatból is ismert, hogy az információforrásokat a másodlagos adatokkal írjuk le (azonosítjuk), majd e tulajdonságok alapján tudjuk visszakeresni az elsődleges dokumentumokat.

A keresés után a találatokról is a megjelenített metaadatok alapján szerezhetünk információt. Az egyes metaadatokat nagyobb egységbe rendezzük – ez a metaadatrekord –, ezek a rekordok azokat a tulajdonságokat/jellemzőket (formai és tartalmi) tartalmazzák, amelyek feltétlenül szükségesek a források leírásához. Hálózati dokumentumok esetében a metaadatok alkalmazásának célja a keresési lehetőségek finomítása, várt eredménye pedig a releváns források megtalálhatóságának biztosítása lesz.

A metaadatoknak (ill. a metaadat-rekordoknak) két fő típusa létezik: a beágyazott és a csatolt metaadat.

- **Beágyazott metaadat:** olyan leíró (másodlagos) adat, amelyet magában a leírt forrásban helyeznek el. Általában HTML-formátumú dokumentumoknál szokták alkalmazni, de lehetséges más

típusú forrásnál is. A beágyazott metaadatok előállítására különböző eszközök állnak rendelkezésre, amelyek segítségével (pl. egy űrlapot kitöltve) előállíthatók a szerkesztett és beilleszthető metaadatok. Ilyen metaadat pl. a CIP (Catalog in publication) leírás a könyvekben, valamint a TEI fejléc az elektronikus szövegekben.

- **Csatolt metaadat (stand-alone metadata):** olyan leíró (másodlagos) adat, amelyet a leírt forrástól függetlenül tárolnak (pl. egy adatbázisban). A leírt forrásban egy csatolón keresztül oldják meg a metaadatrekord és a forrás közötti kapcsolatot. A csatolt metaadatokat általában nem szöveges forrásoknál alkalmazzák, de a szöveges dokumentumoknál is egyre inkább elterjedtek, elsősorban azért, mert lehetőséget adnak a leírások megosztására (pl. a közös katalógizálás analógiája alapján).

Az alkalmazott metaadat-szabványok nem írják elő kötelezően egyik típusú metaadat alkalmazását sem.

A metaadatok előállításának két lehetséges módja: intellektuális és automatikus. Intellektuális előállításnál segédprogramok megfelelő mezőit az adott forrás adataival kitöltve megszerkeszthetők a metaadatrekordok, amelyeket a dokumentumba ágyaznak be, vagy külön tárolnak. A metaadatok automatikus előállítása az eredeti dokumentumból akkor oldható meg, ha a forrás szövege funkcionálisan strukturált, például XML formátumú. A metaadatok ez esetben automatikusan kiolvashatók a szövegből a már említett DC, vagy qDC formátumban.

Dublin Core kódolás XML-be

A leírtakból egyértelműen következik, hogy a feldolgozott dokumentumok számbavétele, szolgáltatása és archiválása szempontjából kiemelkedően fontosak a formai, illetve tartalmi jellemzőket leíró metaadatok. A legnagyobb probléma sokáig az volt, hogy a formai feltárás hagyományos eszközeivel nem lehetett megoldani a digitális dokumentumok feldolgozását. Ez vezetett a Dublin Core kidolgozásához. A DC rendszer nagy előnye, hogy számos nyelvbe – XHTML, XML, RDF stb. – kiválóan beépíthető, és a böngészőprogramok is kezelik. Következésképpen nem kell mást tennünk, minthogy XML állományaink forrásában DC elemek felhasználásával leírjuk jegyzőkönyveink tartalmi és formai jellemzőit. Ennek eredményeként XML forrásunk jól strukturált lesz, és bibliográfiai információkat magában foglaló fejléceket is tartal-

mazni fog, tehát a metaadatok teljes egésze automatikusan kiolvashatóvá válik.

Testületi anyagok jegyzőkönyveinek metaadatszintű leírása cédulaformátumban:

Nagykőrös, Testületi ülés jegyzőkönyve

1644/2005. (XI.24.) Testületi ülés jegyzőkönyve : Nagykőrös Város Önkormányzat Képviselő Testületének 2005. nov. 24-i üléséről / [közread.] Nagykőrös Város Önkormányzat Polgármesteri Hivatal Aljegyzője.– Nagykőrös, 2005. december 7.

Nagykőrös

Testületi ülés jegyzőkönyv

Önkormányzat

Napirend

Előterjesztés

Határozat

Testületi anyagok jegyzőkönyveinek metaadatszintű leírása Dublin Core XML-ben:

```
<dc_title>1644/2005. (XI.24.) Testületi ülés jegyzőkönyve : Nagykőrös Város Önkormányzat Képviselő Testületének 2005. nov. 24-i üléséről</dc_title>
<dc_subject>
  <topicgroup>
    <keyword>Nagykőrös</keyword>
    <keyword>Testületi ülés jegyzőkönyv</keyword>
    <keyword>Önkormányzat</keyword>
    <keyword>Napirend</keyword>
    <keyword>Előterjesztés</keyword>
    <keyword>Határozat</keyword>
  </topicgroup>
</dc_subject>
<dc_description>Nagykőrös Város Önkormányzat Képviselő Testületének 2005. november 24-i üléséről készült jegyzőkönyv, a napirendi pontokkal, az előterjesztésekkel, a hozzászólásokkal, és az ülés végén meghozott határozatokkal.</dc_description>
<dc_type>szöveg</dc_type>
<dc_creator>Nagykőrös Város Önkormányzat Képviselő Testület</dc_creator>
<dc_publisher>Szabó Károly Városi Könyvtár, Nagykőrös</dc_publisher>
<dc_date>2006-05-05</dc_date>
<dc_format>text</dc_format>
```

```
<dc_identifer>URL:
www.nagykoros.konyvtar.c3.hu</dc_identifer>
<dc_language>hu</dc_language>
```

Testületi anyagok feldolgozása XML-ben:

```
<?xml version="1.0" encoding="ISO-8859-2"?>
```

```
<!--Fájlnev: Testületi jegyzőkönyvek.xml-->
```

```
<TestületiJegyzőkönyvek>
```

```
<Kézirat>
```

```
<Cím>1644/2005.(XI.24.)Testületi ülés jegyzőköny-
ve</Cím>
```

```
<Alcím>Nagykőrös Város Önkormányzat Képviselő-
testületének 2006. február 10-i rendkívüli üléséről</Alcím>
```

```
<Dátum>2006. február 10.</Dátum>
```

```
<Szám>1644/2005</Szám>
```

```
<Közread>Nagykőrös Város Önkormányzat
Polgármestei Hivatal Aljegyzője</Közread>
```

```
<Napirendi pont>
```

```
< sor>1. Városi tejbegyűjtő és kimérő megvalósításra
javaslat. Előadó: Révay György települési
képviselő</sor>
```

```
< sor>2. KÖVA Rt. Pénzügyi ellenőrzése. Előadó: dr.
Koltay Zoltán települési képviselő</sor>
```

```
< sor>3. Lakások és helyiségek bérletéről szóló
13/2004. (V.28.) ÖT. Sz. rendelet módosítása.
Előadó: de. Karay Kornélia címzetes főjegyző</sor>
```

```
</Napirendi pont>
```

```
<Határozat>
```

```
< sor>204/2005. (XI.24.) ÖT. Sz Határozat</sor>
```

```
< sor>1. Nagykőrös Város Önkormányzat Képviselő-
testülete a 154/2005. (IX.29) ÖT. Sz. határozat alapján, a megvalósítandó helyi tejbegyűjtő-kimérő érdekében készített tájékoztatót elfogadta.</sor>
```

```
...
```

```
</Határozat>
```

```
</Kézirat>
```

```
...
```

```
</Testületi Jegyzőkönyvek>
```

Az elkészült XML dokumentumok megjelenítése

Alapvetően háromféleképpen tudjuk jelölni a böngészőnek azt, hogy hogyan kezelje és jelenítse meg az XML elemeinket.

- Stíluskapcsolás: ezzel a technikával az XML dokumentumhoz stíluslapot kapcsolunk. A stíluslap egy külön állomány, amelyben az egyes XML elemek formázására vonatkozó utasítások szerepelnek.
- Adatkötés: ezzel a módszerrel először létre kell hozni egy HTML lapot, és hozzá kell csatolni az XML dokumentumot, majd szabványos HTML elemeket kell kötni az oldal XML elemeihez.
- XML DOM szkriptek: először egy HTML oldalt készítünk el, majd hozzákötjük az XML dokumentumot, és az egyedi XML elemekhez a hozzáférést és a megjelenítést szkriptkóddal tesszük lehetővé.⁹

Az XML filozófia a formázás helyett a tartalomra helyezi a hangsúlyt. Amikor viszont a felhasználók számára szolgáltatható állományokról beszélünk, akkor fontos kérdés lesz a formázási információ. Rendszerint sokféle technikát lehet alkalmazni a szövegek olvasmányosabbá tételére és a dokumentumok fontos összetevőinek kiemelésére. Ilyenek:

- betűtípus és szövegméret módosítása,
- színek és hatások használata,
- betűstílusok beállítása,
- margók, keretek, hivatkozások alkalmazása.

Ezeket a technikákat eltérő mértékben használhatjuk a közzététel jellegétől és a feldolgozott jegyzőkönyv papíralapú kiadásainak megjelenésétől függően.¹⁰

Összegzés

A téma, vagyis a testületi anyagok bibliográfiai leírásának problémái, nagyon szerteágazó, és a hagyományos könyvtári munkán kívül mélyreható informatikai ismereteket igényel. Mivel a munka célja egy teljes szövegű adatbázis létrehozása, és az adatbázis interneten való hozzáféréseinek megoldása, közös munkát igényel a könyvtár dolgozói és az informatikusok között.

A jelenlegi könyvtári számítógépes infrastruktúra hardveres része elegendő a munka elvégzéséhez, viszont a szükséges szoftverek még nem állnak rendelkezésre. Gondolok itt a Microsoft XML Core Services (msxml) 4.0 és a Microsoft XML SDK 4.0 programokra, valamint a Recognita Omni Page Pro 11-re.

A város jelenlegi honlapja, valamint a könyvtár saját honlapja HTML alapú, így erre a célra is megfelelne, mivel az XML alkalmazása a feldolgozás során nem kötelezi az XHTML használatát a megjelenítésben.

A munkafolyamat terve

A munkálatokra kidolgozott technológiai folyamat elemei (6. ábra) – a szkennelés, az optikai karakterfelismerés (OCR), a korrektúrázás, az XML kódolás, annak ellenőrzése – egymásra épülő, ugyanakkor jól elkülöníthető részekből állnak.

Szkennelés

A hagyományos úton előállított input dokumentumok szkennelése után az OCR program segítségével a papírra vetett szöveg elektronikus formátumúvá alakítható, így azután formázható, szerkeszthető.

OCR előkészítés

A beolvasott oldalakat mindenképpen célszerű valamilyen képszerkesztő programmal az optikai karakterfelismeréshez előkészíteni. A nemkívánatos fekete foltokig bezárólag eltüntethetjük mindazt, amit egész biztosan nem akarunk felismertetni, illetve a nem szigorú rendben feliratozott olda-

latat a jobb eredmény érdekében rendbe akarjuk szedni.

Optikai karakterfelismerés/felismertetés – OCR

Az optikai karakterfelismertetéshez a Recognita OmniPage Pro 11-et használjuk. Az OCR-ezés során nyert szöveget tárolása előtt korrektúrázhatjuk, továbbá igénybe vehetjük az OmniPage szoftver IntelliTrain szolgáltatását is.

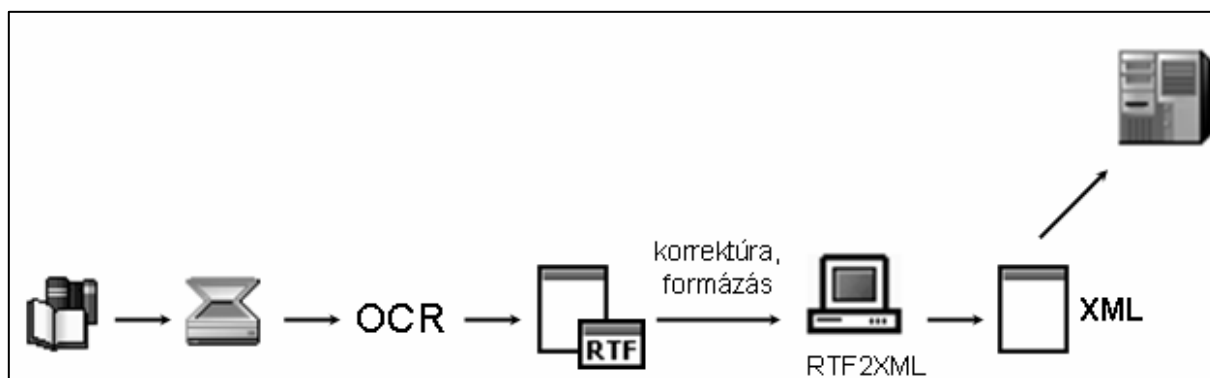
Korrektúrázás

A korrektor szerepe hasonlít azokhoz a feladatkörhöz, melyeket például kiadóknál, szerkesztőségeknél vagy nyomdáknál végeznek. Célja, hogy a digitalizálandó eredeti szöveg teljes mértékben megegyezzen a majdan az interneten megjelenő szöveggel.

XML kódolás

A már korrektúrázott forrásanyagot a PSPad szövegszerkesztőbe töltjük be, mert ez ingyenesen rendelkezésre áll. Itt kell lecserélni az ékezetes betűket entitásokra, valamint itt kell az XML címkéket (tag-eket) elhelyezni a kódban. (A PSPad helyett vannak más, ingyenes programok is, sőt, az XML szerkesztésére talán nem is a kifejezetten karakteres szerkesztők a legalkalmasabbak, de a célnak jelen esetben ez is megfelel.)

A helyes szintaktikai kódolás ellenőrizhető a felhasználott DTD-vel való összevetéskor, de csak akkor, ha például egy validálásra is képes XML szerkesztőt használunk – a PSPad erre nem alkalmas. A kódolás befejeztével az XML kódot át kell konvertálni HTML formátumúvá.



6. ábra Tervezett munkafolyamat

Összegezve a közös munkát

- Első lépés a meglévő jegyzőkönyvek gépre vitele szkenneléssel, rtf formátumban. Ez a könyvtárban megoldható.
- Az OCR előkészítés és az optikai karakterfelismerés is megoldható a könyvtárosok szaktudásával.
- A korrektúrázás szintén a könyvtárosokra váró feladat.
- A DTD-k kidolgozásában, alapjainak lefektetésében nagy szerepe van a könyvtárosi munkának, ez a folyamat közös munkát igényel az informatikusokkal.
- Az XML alapú szövegfeldolgozás informatikus feladat lesz.
- A Dublin Core metaadatok előállítása a könyvtáros feladata, a fejlécbé történő behelyezés az informatikusé.
- A kész XML kód lefordítása HTML formátumba a számítástechnikus feladata lesz.

Az elkészült terv elméleti, a gyakorlati kivitelezésben számos probléma fog felmerülni, az elkészült elméleti alapok is valószínűleg néhány tekintetben változni fognak, de bízom abban, hogy a könyvtárosok és az informatikus szakemberek következetes és gyümölcsöző munkájának eredményeképpen meg fog valósulni ez a terv, elérve mindazon célokat, amelyeket a munka elkészülte előtt kitűztünk.

Irodalom és jegyzetek

- 1 VÉRTESY Miklós: Könyvtárosok kislexikona, Budapest, Múzsák, 1987, 110. p.
- 2 KSZ/3-2001 Bibliográfiai leírás. Időszaki kiadványok. p. 8.
- 3 GAZDAG Tiborné: Az időszaki kiadványok számbavétele és feltárása – nemzetközi tendenciák – és a hazai környezet, <http://epa.oszk.hu/00100/00143/00043/gazdag.html>; [2006. 04. 25.]
- 4 Extensible Markup Language (XML) 1.0 (Fourth Edition), <http://www.w3.org/TR/REC-xml/>; [2007. 05. 22.]
- 5 Dublin Core Metadata Initiative (DCMI), <http://www.dublincore.org/>; [2007. 05. 22.]
- 6 SALGÁNÉ MEDVECZKY Marianna: Az XML : új perspektívák a könyvtár-informatikában, http://mt.omikk.bme.hu/show_news.html?id=3541&issue_id=448; [2006. 03. 18.]
- 7 YOUNG, Michael J.: XML lépésről lépésre. Budapest, Szak Kiadó, 2002, p. 22–24.
- 8 BÍRÓ Szabolcs: Szövegfeldolgozás XML alapokon. Budapest, Neumann-ház, 2005. p. 128–132.
- 9 YOUNG, Michael i. m. p. 9.
- 10 BÍRÓ Szabolcs i. m. p. 109–112.

Beérkezett: 2007. IV. 2-án.



Bujdosóné Dani Erzsébet

a nagykőrösi Szabó Károly Városi
Könyvtár igazgatóhelyettese.
E-mail: budani@vipmail.hu

HIRDESSZEN A **tmt** -BEN!

2008-ban megrendelt hirdetéséhez
ingyenes bemutatkozási lehetőséget (pr cikk) adunk
a lap elektronikus változatában.

A hirdetés ára:

- beszerkesztett (fekete-fehér): 60 000 Ft/oldal + 20% áfa
- szórólap elhelyezése (25 g-ig): 45 000 Ft/alkalom + 20% áfa

A kiadvány formátuma: A/4 • Tükörméret: 168 x 225 mm

Megrendelhető a szerkesztőségben:

1111 Budapest, Budafoki út 4-6. • Levélcím: 1510 Bp., Pf. 111
Tel.: 463-2446; 463-1111/5652 • E-mail: tmt@omikk.bme.hu