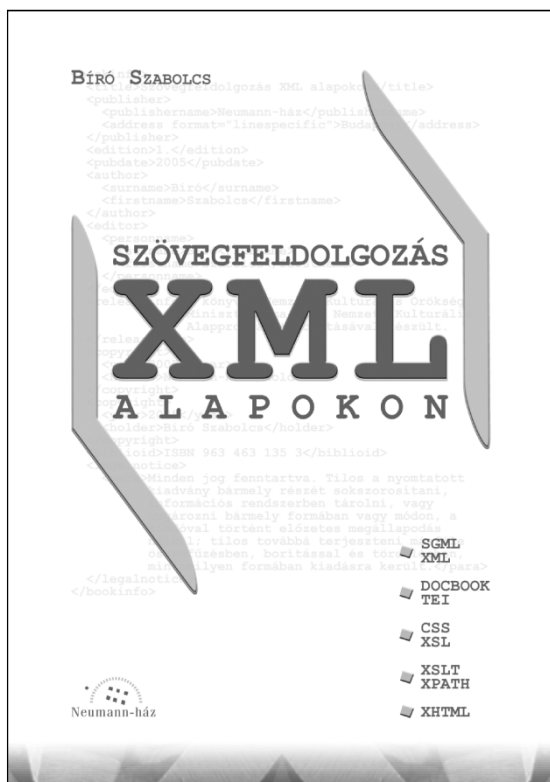




XML és szövegfeldolgozás



Szövegfeldolgozás XML alapokon / Bíró Szabolcs ; [közreadja a Neumann János Digitális Könyvtár és Multimédia Központ]. – Budapest : Neumann-ház, cop. 2005. – 208 p. : ill. ; 24 cm
Bibliogr. p. 194-201.
ISBN 963-218-740-7

Napjainkban az internet az információgyűjtés, az ismeretszerzés, az élethosszig tartó tanulás egyik elsődleges forrása. Egyre több dokumentum érhető el elektronikus formában, de messze elmarad a kínálat az igények mögött. Még több tartalomra – szakirodalomra, adattárakra, kulturális forrásokra lenne szükség ahhoz, hogy kielégítsük a netezők információéhségét. A régi, pusztulóban lévő értékek megőrzése, átörökítése és a legfrissebb információk elérhetővé tétele egyformán fontos feladat. A könyvtárak mint a nagy tömegű információ gyűjtői és szolgáltatói, szintén bekapcsol-

lódnak a tartalomszolgáltatásba, és ezen a területen is szükség van jól képzett szakemberekre.

Az újonnan megjelentetett könyvvel nemcsak a könyvtári informatikusoknak szeretnénk segíteni, hanem bárkinek, aki most ismerkedik a szövegdigitalizálással, annak eszközeivel, módszereivel. Számos lehetőség kínálkozik, ha elektronikus dokumentumokat kívánunk készíteni, és a különböző műhelyek ténylegesen is eltérő megoldásokat alkalmaznak. Való igaz, hogy nincs egyetlen üdvöztető technológia minden feladatra, hiszen a forrásdokumentumok állapota, mennyisége, milyensége (és még sok más tényező, pl. az anyagi háttér) jelentősen befolyásolja az eszközök és módszerek megválasztását. Mégis szükségesnek látszik, hogy az eddiginél nagyobb figyelmet fordítsunk arra, hogy korszerű, szabványos megoldást alkalmazzunk, mert így érhető el, hogy a jelentős ráfordítással előállított digitális dokumentum sokoldalúan felhasználható, többféle céllal feldolgozható legyen, más környezetbe könnyen áthelyezhető, időtálló formátumban készüljön el.

A ma használatos megoldások közül talán az XML technológia alkalmazása a legigéretesebb, hiszen a fentebb leírt feltételeknek szinte teljes körűen megfelel. Az sem elhanyagolható szempont, hogy egy szabványos jelölőnyelvről van szó, amely átjárást biztosít többféle formátumhoz, és támogatja a dokumentumok hordozhatóságát is.

A Neumann-ház digitális gyűjteményének megalapozásakor az SGML szabvány mellett döntött, ebben a formátumban készültek el a szépirodalmi és a szakirodalmi művek elektronikus kiadásai. A több éve folyó digitalizálási tevékenység során összegyűlt tapasztalatokat korábbi munkatársunk, *Kora András* egy kéziratot tankönyvben foglalta össze. Ez a könyv szolgált később – az elsősorban könyvtári szakembereknek szánt – akkreditált tanfolyamunk alapjául. A szervezett oktatás céljainak megfelelően átdolgoztuk a tananyagot, belekerültek az időközben megjelent frissítések, és a ko-

rábbinál nagyobb hangsúllyal tettük be a tárgyalt témakörök közé az XML-t.

Az első tanfolyam sikere után döntöttünk úgy, hogy könyv alakban is megjelentetjük a tananyagot, természetesen bővített formában, számos újabb forrás felhasználásával. A szerző, *Bíró Szabolcs* hasznosította a gyakorlati oktatás során szerzett tapasztalatokat is. A Nemzeti Kulturális Alapprogram támogatásának köszönhetően 2005-ben jelent meg a „Szövegfeldolgozás XML alapokon” című nyomtatott kiadvány, amely reményeink szerint betekintést nyújt a korszerű technológiákba, megismerteti az olvasót a választható lehetőségekkel, és gyakorlati útmutatót nyújt mindazoknak, akik most ismerkednek az internetes szolgáltatások világával.

A könyv nem egyetlen programra vagy egyetlen technológiára koncentrál, és nem is vállalkozik arra, hogy professzionális digitalizáló szakembert képezzen az olvasóból, hanem elsősorban arra törekszik, hogy a választható lehetőségek tárházát és a korszerű technikákat mutassa meg, valamint felhívja a figyelmet az interneten megtalálható számos eszközre, ajánlásra. A történeti fejlődést követve könnyebben érthetővé válik, hogy az idők folyamán hogyan és miért változtak a szövegfeldolgozási módszerek. Miért került előtérbe a műveleti jelöléssel szemben a leíró jelölés? Miért van jelentősége a tartalom és a forma szétválásának?

Hogyan oldható meg, hogy a lehető legtöbb információt rögzítsük a dokumentumról, ugyanakkor sokoldalú feldolgozásra tegyük alkalmassá? A közérthető nyelven íródott könyv mindezekre a kérdésekre választ ad, és közben bevezet minket a kódolási technikák részleteibe. Számos gyakorlati példán keresztül elsajátíthatjuk a DTD (Document Type Definition = dokumentumtípus-definíció) készítést, valamint a leíró jelölést, a formázó és megjelenítő programok szerkesztésének alapjait. Az XML-hez vezető úton megismerkedünk a szövegjelölési eljárások közti különbségekkel, a HTML helyébe lépő XML alapú XHTML-lel, a TEI ajánlással, a DocBook jelölérendszerével, megtanulhatjuk a lépcsős és bővíthető stíluslapok használatát.

A könyv erénye, hogy a módszereket párhuzamosan mutatja be, felhívja a figyelmet az azonosságokra és az eltérésekre. Így mindenki könnyen eldöntheti, hogy melyik ajánlást célszerű az adott feladathoz alkalmaznia. Reményeink szerint gyakorlati segítséget nyújt a technológia kiválasztásához és sikeres alkalmazásához. Mellőzi a lexikális ismeretek halmozását, inkább számos kapaszkodót ad ahhoz, hogy az alapok megértése után önállóan is eligazodjunk a digitalizálási szabványok és ajánlások között.

Bánkeszi Katalin
(Neumann-ház)