

## Salgáné Medveczki Marianna

Debreceni Egyetem Informatikai Intézet, komputergrafikai és könyvtár-informatikai tanszék

# Az XML: új perspektívák a könyvtár-informatikában

*A tanulmány ismerteti az XML (Extensible Markup Language) szabványos jelölőnyelvet, kiemelve a nyelvnek azokat a fontos jellemvonásait, amelyek a könyvtári területen történő alkalmazhatóságát erősítik. Beszámol egy új fejlesztésről, amely a Debreceni Egyetem komputergrafikai és könyvtár-informatikai tanszékén indult azzal a céllal, hogy formalizálja a bibliográfiai leírást az XML jelölőnyelv segítségével.*

### Bevezetés

Ahhoz, hogy egy könyvtár a küldetését teljesíthesse, szükséges az állományának (amelyben helyet kaphat tetszőleges típusú dokumentum), valamint a könyvtár felhasználói igényeihez szervesen kapcsolódó, a különböző hálózatokon és az interneten elérhető információforrásoknak, információknak magas színvonalú és korszerű feldolgozása.

Az elmúlt években a számítástechnika, az automatizálás és a számítógépes hálózatok, valamint a legutóbbi időkben az interneten található információforrások eredményeinek az eredendően hagyománytisztelő könyvtárakba való betörése alapvetően hat az információk feldolgozására, tárolására, keresésére. A különféle elektronikus dokumentumok és információforrások kínálta lehetőségeknek, illetve ezeknek az elektronikus dokumentumoknak a könyvtári kezelése egyre természetesebbnek tűnik mindannyiunk számára. A könyvtárak automatizálásával együtt a hagyományos bibliográfiai feltárási eszközök újabb és – reméljük – hatékonyabb eszközökkel, megoldásokkal cserélődnek le. Változik a technológia, és változnak a szellemi eszközök (szabványok, szabályzatok, metaadatok stb.) is.

A nagyobb könyvtárakban az egyes könyvtári munkafolyamatok integrált szemléletű automatizálása már végbement, ezek a számítógépes rendszerek több-kevesebb sikerrel működnek. Természetesen az iskolákban és a kisebb községek könyvtáraiban is felmerült a számítógépes könyvtári nyilvántartások kialakításának igénye, a szűkös anyagi lehetőségek azonban a megvalósításnak még sok helyen határt szabnak.

A következőkben egy olyan informatikai eszközt, eszközrendszert mutatunk be, amelynek könyvtári területen történő alkalmazása megoldást jelenthet az internet előretöréséből fakadó problémákra az integrált számítógépes rendszereket használó könyvtárakban. Az XML alkalmazása nyíltabb rendszer létrehozását teszi lehetővé a MARC alapú zártabb könyvtári rendszerekhez képest. (Nyílt rendszer az, amely olyan részekből áll, melyek egységes, gyártófüggetlen és nemzetközileg elfogadott szabványokon alapuló felületeket nyújtanak. Ennek köszönhetően az egyes részek rugalmasan összekapcsolhatók, valamint segítenek összehangolni a különböző hardver- és szoftvertermékek használatát.) Az internet globális világában egyre fontosabbá váló követelmény, hogy a könyvtári OPAC-okhoz hasonló közszolgáltatások nyíltabb, átjárható, más keresők által elérhető, mélyen linkelhető legyenek. Ma ennek a követelménynek egy átlagos könyvtári rendszer nem igazán képes megfelelni. Továbbá az XML alkalmazása segíthet a kisebb könyvtárak számára abban, hogy elinduljon könyvtárak számítógépesítése, és a számítógépek használatával ezek a könyvtárak is könnyíthessék mindennapi munkájukat.

Az XML (Extensible Markup Language = kiterjesztett jelölőnyelv) egy alig öt éves új szabvány, amely (félig) strukturált adatok, strukturált szövegek leírására és számítógépes környezetben történő kezelésére szolgál. Az XML számos olyan előnyös tulajdonsággal rendelkezik, amellyel kivívhatja a könyvtáros szakma figyelmét. Olyan valós problémákra kínál megoldást, amelyekkel a napjainkban használt könyvtári számítógépes rendszerek adatformátumai jelenleg is küzdenek. Az utolsó fejezet erről részletesebben szól, összefoglalja az XML –

könyvtáros szemszögből vizsgált – legfontosabb pozitív jellemzőit.

A Debreceni Egyetem komputergrafikai és könyvtár-informatikai tanszékén elindult egy fejlesztés, amelynek célja a bibliográfiai leírás formalizálása az XML jelölőnyelv segítségével. A fejlesztés egy olyan formális szabálygyűjtemény (bibliográfiai leírás, DTD) kialakítására irányul, amely XML nyelven megadja a szabványos bibliográfiai leírások szerkezetének sémáját. Ennek a DTD-nek az alkalmazása, reményeink szerint, komoly segítséget jelenthet minden olyan könyvtár számára, amely számítógépes környezetben könnyen és gyorsan, ezzel együtt magas szakmai színvonalon kíván a dokumentumairól bibliográfiai leírásokat készíteni. (A fejlesztés alatt álló DTD-ből egy rövid ízelítő az utolsó fejezetben található.)

### Röviden az XML-ről

„Ahogyan a HTML lehetővé tette, hogy minden számítógép-felhasználó olvashassa az internetes dokumentumokat, úgy az XML – az inkompatibilis számítógépes rendszerek bábeli zűrzavarában is – az eszperantót jelenti, melyen mindenki olvashat és írhat” (BOSAK, J.–BRAY, Tom: Az XML és a második generációs web. [http://www.scala.hu/XML\\_DAT\\_hun.htm](http://www.scala.hu/XML_DAT_hun.htm)).

Az XML szabvány és a hozzá kapcsolódó szabványos eljárások új lehetőséget kínálnak a szöveg és más médiák együttes kezeléséhez. Az XML a World Wide Web Consortium (W3C) által kifejlesztett új, hatékony dokumentumleírási és megjelenítési formátum. Két meglévő nyelv, a HTML (HyperText Markup Language = hipertext jelölőnyelv) és az SGML (Standard Generalized Markup Language = szabványosított általános jelölőnyelv) elvei és szabályai alapján egyszerű, de mégis hatékony információátviteli, -feldolgozási és -megjelenítési rendszert hoz létre. Az XML elsődlegesen az SGML-re épül, de átvette a HTML néhány jellemzőjét is, valamint kiegészül olyan további lehetőségekkel, amelyek az interneten való felhasználását könnyítik meg.

Az ajánlás 1998-ban jelent meg, amelyet az informatikus szakma nagy tetszéssel fogadott, és azóta széles körben elterjedt. Az XML sikerében használhatósága mellett további jelentős szerepet játszott az is, hogy az XML nyelvnek nincs tulajdonosa, és egyetlen gazdasági érdek sincs rá kizárólagos befolyással.

Az XML metanyelv, amelynek segítségével újabb jelölőnyelvek készíthetők. Kiterjeszhetősége lehetőséget ad a különböző adatstruktúrák tartalmi és nem formai szempontok alapján történő leírására, amelyeket elektronikus úton tárolunk olyan formában, hogy egyaránt érthető legyen a számítógépes programok és az emberek számára is. Követelmény, hogy az XML-dokumentumok emberi olvasásra is alkalmasak legyenek, ami azt jelenti, hogy az XML fájlok nem binárisak. Egy XML formátumú adatállomány az XML-dokumentum szövegének és az XML-dokumentum egyes részeit összerendező és azonosító XML jelöléseknek a keveréke.

Az XML ideális adatformátum a különböző médiákon terjesztett, közvetlenül megjelenített strukturált vagy félig strukturált szövegek számára. Az XML formában megjelenő információ „önleíró” – a megcélzott közönség és a használt média követelményeihez igazodóan használható, alakítható és formázható. A megjelenítésen kívül a különböző programok és számítógépes rendszerek közti információcsere egyre gazdagabbá és bonyolultabbá válik az új, hálózatalapú alkalmazások fejlődésével. Az ilyen információkkal szemben követelmény: önleíróak legyenek annak érdekében, hogy a kliensprogram tudja értelmezni a kapott adatokat, és végrehajthassa a felhasználtól kapott feladatokat anélkül, hogy újra a kiszolgálóhoz kellene kapcsolódni.

Egy XML alapú dokumentumnak van logikai és fizikai szerkezete is. A logikai szerkezet lehetővé teszi, hogy a dokumentumot nevet viselő egységekre, alegységekre bontsuk, amelyeket elemnek neveznek. A fizikai felépítés során pedig lehetőség van arra, hogy a dokumentum részeit, ún. egyedeit elnevezzük, és külön tároljuk általában olyan adatállományokban, amelyekből az információ újra kinyerhető; és hivatkozásokkal, nem XML típusú adatokkal is (pl. képekkel) bővíthetjük.

Az XML segítségével definiált nyelvek bővíthetők, ugyanis ezekben nincs előre definiált elemlista (mint pl. a HTML-ben), hanem saját elemeket lehet tetszőlegesen deklarálni. A deklarált elemek szabadon egymásba ágyazhatóak, s a legmagasabb szinten valamennyi elemet egyetlen dokumentum-elembe kell zárni. A dokumentumelem egy konkrét előfordulását nevezzük XML-dokumentumnak. Az egy osztályba tartozó XML-dokumentumok elemei alkotják a definiált nyelv szókincsét, és az elemek tartalmát, valamint egymáshoz való kapcsolatát rögzítő szabályok adják a definiált nyelv nyelvtanát. Az XML specifikáció megad egy szintaxist

mind az XML-dokumentumokra – vagyis az elemek jelölésére –, mind pedig a szabályok leírására (DTD). A megadott szintaktikai szabályok betartásával bárki saját nyelvet (dokumentumtípust) készíthet, s azt a megfelelő XML-konform eszközzel ellenőrizheti, feldolgozhatja.

A DTD (*Document Type Definition*) formális szabályokat definiál, amely egyebek közt lehetőséget nyújt 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 gyermek elem vagy más néven leszármazott, finom tagoltságról, ahol pedig kevés a leszármazott, ott durva tagoltságról beszélünk.

Az XML-ben rejlő lehetőségek néhány jól megválasztott szabályból erednek. Ezek egyike az, hogy a jelölőelemek szinte mindig párosával szerepelnek, mintegy zárójelként veszik körül a szöveget, amelyre vonatkoznak. A jelölőelemek általában közvetlenül nem határozzák meg a megjelenítés stílusát, hanem csak azonosítják a különböző szövegrészek jellegét. A jelölőelem nyitó része és a jelölőelem záró része a benne lévő adattal együtt jelent egy elemi egységet. Az XML elemek tartalmazhatnak további, beágyazott elemeket tetszőleges mélységben, és végül az egész dokumentumot egyetlen dokumentumelembe kell zárni. Az elkészült XML-dokumentum szerkezetének felépítése sokféle grafikus módszerrel ábrázolható, az egyik lehetőség például a fastruktúra alkalmazása.

Az XML ilyen módon az egyes szakterületek specialistái számára lehetővé teszi új és egyedi nyelvek tervezését, azonban valóban jó nyelvet létrehozni korántsem lebecsülendő feladat. Egy-egy új XML alapú jelölőnyelv kialakítása előtt lényeges végiggondolni: milyen jelölőelemeket használjunk, a kódolt elemek hogyan ágyazódhatnak egymásba, és hogyan történjen a kódok végrehajtása. Az első kettőt (a nyelv „szókincsét” és „nyelvtanát”) általában a DTD szabályozza, a harmadikról az XML-hez kapcsolódó, kiegészítő szabványok és a megfelelő alkalmazás gondoskodik. Az XML szabvány nem kényszeríti egy új nyelv tervezőit a

DTD kötelező használatára, azonban a legtöbb nyelv feltehetően mégis tartalmazni fogja, mert segíti a későbbiekben a programozókat a helyesen működő szoftverek elkészítésében.

Az XML nem elszigetelt szabvány. Az alapvető szintaxisspecifikáción kívül az XML-ben megtalálhatók a hipertextkapcsolatokat létrehozó (XLink és XPointer) és dokumentumformázó (XSL és XSLT) szabványok is. Ezen túlmenően létező, kidolgozott és fejlesztés alatt álló szabványok sora egészíti és terjeszti ki az XML hatókörét (pl. ASCII, UNICODE, ISO/IEC 10464, HTTP, CSS, SPDL).

### **Az XML felhasználási területei és lehetőségei**

Az XML megjelenése óta komoly sikereket ért el az elektronikus kereskedelem, a tudományban, a számítástudományban és valamennyi iparágban, a gyártástól kezdve a gyógyításig. A legjelentősebb programgyártó cégek behódoltak a szabvány előtt, és számos XML-t támogató termékkel jelentek meg.

Ha bonyolult adatformátumot kell két program között mozgatni, az XML minden esetben megfelelő formátum lehet. Az XML már bizonyított bizonyos területeken, mint pl. az információszolgáltató technológiák (push technology), EDI (Electronic Data Interchange) és köztes adatok generálása (MCF, XML adatok) és RDF (Resource Description Framework).

Sok ún. ipari DTD létezik, amelyet sokan használnak, így egyfajta szabvánnyá válhat egy adott felhasználói körben. Az ipari DTD esetenként kifejezetten megengedő szabályokat tartalmaz annak érdekében, hogy sokféle projekthez alkalmazható legyen. Leggyakrabban ezért ezek az ipari DTD-k testre szabhatók, bővíthetők. Ilyen DTD például a TEI (Text Encoding Initiative) DTD, amelyet elsősorban szépirodalmi szövegek digitalizálásánál használnak.

Ma már sokféle adat leírásához létezik DTD, pl. XHTML, amely a HTML webes formátumleíró nyelv XML változata; DocBook a dokumentációk, könyvek leírásához; XFRML (XML Financial and Reporting Markup Language), amely pénzügyi jelentések leírására szolgál; az XMLNews a hírek tartalmának és metaadatainak leírását teszi lehetővé; a WML a WAP mobiltelefonok nyelve.

Az XML alkalmazható relációs adatbázisok mozgására is, ebben az esetben az XML elemek csupán határolóként használatosak a rendszerek közti rekordok, mezők és kapcsolatok átvitelekor. Az XML szabványok fejlődésük során elérték a könyvtári világot is. Több projekt is elindult a MARC formátum XML-ben történő leírására (pl. XMLMARC). Ennek jövőbeli bevezetésével és elterjedésével lehetővé válhat az így feldolgozott rekordok kezelése, visszakeresése XML vagy XML-re épülő eszközökkel. Teljes szövegű elektronikus dokumentumok esetén pedig közvetlen kapcsolat teremthető a bibliográfiai leírás és a teljes szöveg között. Az XML-t alkalmazzák már a teauruszépítésben és -szerkesztésben is (a Virtual HyperGlossary pl. XML formátumú teauruszszoftver).

Hazánkban a Szegedi Egyetem Egyetemi Könyvtárában *Bakonyi Géza* vezetésével indult egy program, amelyben digitalizált szöveget XML eszközökkel alakítanak át strukturált szöveggé. Az általuk kidolgozott módszer lényege, hogy egy digitális formában létező folyóirat esetén milyen módon lehet a strukturált szövegfeldolgozás eszközeit hasznosítani úgy, hogy az adott folyóirat szövegéből mintegy „automatikusan” előállítható legyen egy virtuális repertórium (lehetőleg a nyomtatott változatnál nagyobb lehetőségekkel).

A Neumann-ház gondozásában futó Digitális Irodalmi Akadémia névre keresztelt projekt a TEI kódolást alkalmazza az irodalmi szövegek rögzítésére, amelyet a hazai szélesebb körű elterjedés végett magyarítottak.

Ne feledkezzünk meg a Magyar Elektronikus Könyvtár új rendszeréről sem. Míg az előző két példa a dokumentumok szintjén alkalmazza az XML/SGML technológiát, addig a MEK 2.0 a bibliográfiai leírások szintjén biztosítja az XML formátumban való letöltést. Az alkalmazott DTD-t az Empolis cég közreműködésével fejlesztették ki a TEI alapján, azt szűkítve. A DTD elérhető és megtekinthető a következő címen:

<http://mektukor.oszk.hu/iroda/mekdtd/empolis/dtd/>

### **Az XML alkalmazásának lehetősége a könyvtárak állományának formai feldolgozásában**

Az XML nyelvnek számos olyan, a könyvtári munka szempontjából előnyös tulajdonsága van, amely lehetővé teszi az XML könyvtári területen történő sikeres felhasználását. Az XML érdekes lehet

mindazok számára, akik dokumentumaik kialakításában és kezelésében nagyobb rugalmasságot kívánnak elérni, akik sokoldalúan támogatott rugalmas tárolási vagy általános adatformátumot, valamint fejlett, ám ezzel együtt nem túl drága alkalmazásokat keresnek. A számítógép számára megkönnyíti az adatok generálását, olvasását, és biztosítja, hogy az adatszerkezet egyértelmű legyen.

Az XML nem programozási nyelv, használatához vagy megtanulásához nem szükséges programozónak lenni (bár egy „kis” informatikai ismeret sosem árt). A legtöbb számítógépes adatformátumtól eltérően az XML jelölések az ember számára is érthetők, mivel semmi mást nem tartalmaznak, mint közönséges szöveget. Az XML biztosítja az információ tervezését, autonómiáját, többszöri felhasználását, kereshetőségét, számítógépes feldolgozhatóságát, információk összekötését, az információ karbantartását és a minőséget, azaz a tudatos bánásmódot.

Az XML elkerüli a programozási nyelvekben gyakran előforduló csapdákat, ugyanis bővíthető, platformfüggetlen, támogatja a nemzetköziesítést. Jelentőségének egyik záloga, hogy az új szabványon, a UNICODE nevű karakterkódolási rendszeren alapul, amely lehetővé teszi a világ főbb nyelveiben készült szövegek kevert használatát. A UNICODE megengedi a diakritikus és egyéb speciális jelek, valamint a nem latin betűk alkalmazását, normál szöveggént való kezelését. Az XML nyelvet értelmezni képes szoftver e karakterkészletek bármely kombinációját kezelni tudja, így az XML az adatcserét nemcsak a különböző számítógépes rendszerek, hanem az országok és kultúrák között is lehetővé teszi.

XML-dokumentumokat tetszőleges szövegszerkesztő programmal létrehozhatunk, de léteznek kifejezetten XML-érzékeny szerkesztők is. Ez utóbbiak képesek „olvasni” a DTD-t, ezzel segítve a szerkesztő munkáját, így elkerülhetők az XML-dokumentum létrehozása során a szintaktikai és a logikai hibák.

Az XML jelölőnyelv használata lehetővé teszi a bibliográfiai leírásban rejlő lehetőségek jobb kihasználását a leírás azonosító funkciójának csorbítása nélkül.

A katalogizálási szabvány a bibliográfiai leírás fogalmát a következőképpen határozza meg: a bibliográfiai leírás meghatározott szabályok szerinti

egységes szerkezetben, formában és sorrendben leírt adatok összessége, amelyek a dokumentumok azonosítására, nyilvántartásba vételére és a róluk szóló tájékoztatásra szolgálnak. Ebből következően a bibliográfiai leírás formálisan értelmezhető, az adatelemek formálisan felismerhetők. Ha a könyvtáros szakember rápillant a leírás adott pontjára, máris tudja, hogy ott például egy kiadó megnevezéséről van szó, hiszen az adatelem környezetéből, illetve a megelőző egyezményes jelből ez egyértelműen kiderül. A számítógépek nem ilyen „okosak”: pontosan meg kell mondanunk a számukra, mi mit jelent, hol található, és mi a teendő vele. Az XML olyan új nyelv, amelynek kifejezett célja az, hogy az adatok önmagukért beszéljenek.

A DTD-ben tudunk rendelkezni az elemek sorrendjéről, amely a bibliográfiai adatformátumoknál kritikus szempont. A DTD-ben megadható az elemekhez szöveggörnyezettől függően sorrendi szabályozás (az elemek XML-ben két logikai művelettel kapcsolhatók össze, a sorozat-, illetve a választás-összekapcsolóval), és megadhatók (az elemek kötelezőségére és ismételhetőségére vonatkozó) előfordulási szabályok az ún. mennyiségi jelzők alkalmazásával. Mivel XML-ben lehetőség van az elemek tetszőleges mélységű egymásba ágyazására, nagyfokú rugalmasság érhető el. Ennek köszönhetően a bibliográfiai leírás adatelemeinek, vagy néhány összetartozó adatelemének ismételhetősége, opcionálitása, feltételektől függő kötelezősége (például a bibliográfiai leírás 1. adatsorozatjában a különböző párhuzamosan meglévő adatelemek kezelése, akár – könyvek esetében – szerzői vagy egyes gyűjtemények esetén is) XML formátumban nem jelent problémát, rugalmasan kezelhetők a különböző előfordulás-variációk.

Az XML kódok önmagukban nem tartalmaznak utasítást arra vonatkozóan, hogyan jelenjenek meg az adatok a képernyőn vagy nyomtatásban, azaz a kódok a tartalom jellegére koncentrálnak a megjelenítő közegtől függetlenül. A kódolt tartalomhoz tetszőleges számú ún. stíluslap készíthető, s az ezekben a stíluslapokban szereplő szabályokat alkalmazva az anyag a különböző eszközöknek megfelelően automatikusan formázható újra és újra.

A bibliográfiai leírások formalizálásánál jól kihasználható az XML legfőbb jellemzője: a tartalom, a megjelenítés és a kapcsolások különválasztása, amelyek így külön-külön optimálisan kezelhetők. A felhasználást tekintve ez azt jelenti, hogy a bibliográfiai leírások különböző megjelenítési (output)

formátumai külön-külön egy-egy stíluslap elkészítésével kényelmesen megoldhatók (figyelembe véve azt is, hogy az elemek, pl. a leírás adatelemeinek, adacsoportjainak sorrendje XML-ben egyszerűen előírható). A gyengén látó emberek számára is előnyös ez a megközelítés. A stíluslap, illetve megfelelő szoftver segítségével átalakíthatják az XML-t Braille-írássá vagy beszéddé.

Egy DTD megalkotásával és használatával számos előnyhöz jutunk. A programozók készíthetnek adatkezelő és feldolgozó szűrőket anélkül, hogy a programokat nem várt adatok kezelésére fel kellene készíteni. Ugyanilyen nagy biztonsággal készíthetők stíluslapok is, és a DTD minden eleméhez stíluszabályok rendelhetők. XML-kompatibilis szerkesztő használatával a feldolgozó könyvtárosok irányíthatók, és egységes szabályoknak megfelelő XML-dokumentumok (bibliográfiai leírások) készíthetők (bár az XML megengedi a felhasználó által definiált újabb elemek létrehozását, a szabályozott környezetben használt szerkesztőprogramokban ez a lehetőség letiltható).

Befejezésként következzen egy rövid ízelítő a készülő bibliográfiai leírás DTD-ből, amely a könyvek monografikus szintű leírásának első adatsorozatját ülteti át XML nyelvre. A DTD-ben *félkövér* betűvel kiemelt elemek ebben a rövid kivonatban nem lesznek kifejtve.

Röviden az alkalmazott jelölésekről. Ha egy modelcsoportban egynél több elem található, akkor a gyermek elemek különböző módon szervezhetők. Az elemek két logikai művelettel kapcsolhatók össze: *'* (sorozat-összekapcsoló, amely rögzíti egyben az elemek sorrendjét is) és *|* (választás-összekapcsoló).

A DTD azt is meghatározhatja, hogy egy elem milyen gyakran fordulhasson elő az egyes helyeken. A mennyiségi jelzők a következők. Ha egy elem szükséges, és nem ismételhető, akkor nincs szükség további jelölésre. Ha az elem nevét a *'?* követi, akkor az elem csak választható, de nem ismételhető. Ha egy elem szükséges és ismételhető, akkor az elem nevét egy *'+'* követi. És végül, ha egy elem választható is és ismételhető is, akkor a nevét egy *'\*\*'* követi.

A DTD-részlet után a legfontosabbnak érzett DTD-részeket fastruktúrát alkalmazva grafikusán is ábrázoltam (1. és 2. ábra). Bízom benne, hogy ezek a hierarchikus ábrák segítik majd a DTD könnyebb megértését.

Végül pedig a DTD-ben leírt 1. adatszoportnak egy egyszerű előfordulása következik, vagyis egy konkrét XML alapú bibliográfiai leírás 1. adatszoportja (3. és 4. ábra). Az XML-jelölőelemek általában körbefogják az azonosított adategységet az adatfolyamban. A jelölőelem nyitó és záró része a bezárt adattal együtt alkot egy elemet. A nyitóelemet a '<' és a '>', a záróelemet a '</' és a '>' karakterek határolják. Ezeket az ábrákon *félkövér* betűvel emeltem ki. A bibliográfiai leírás egyezményes jeleit jelölő elemek üres elemekként lettek definiálva, azaz ezek olyan elemek, amelyeknek nincs tartalmuk. Az ilyen tartalom nélküli elemek esetén létezik a jelölésnek egy másik formája is. Ekkor a zárórész jelölőelem elmarad, és a nyitó (illetve

üres) jelölőelemnek kicsit eltérő a szintaxisa, '>' helyett '/>' zárja. Az ábrákon az elemek nevét *dőlttel* szedtem, a tényleges adattartalmat a *félkövér* betű jelzi.

Az első esetben (3. ábra) jól látszik a leírás tagolódása a soremeléseknek és behúzásoknak köszönhetően. Az elemek között lévő szóközök és sorvége jelek azonban általában nem tartoznak az XML-dokumentumokhoz, így az elrendezés nincs hatással az XML-dokumentum jelentésére. Gyakorlatilag a 3. és a 4. ábrán szereplő XML-dokumentum egy XML-értelmező számára egy és ugyanazt jelenti.

```
<!-- XML BIBLIOGRÁFIAI LEÍRÁS DTD
      Bibliographic Description Markup Language (BDML)
              SZERZŐ: Medveczki M. (marianna@inf.unideb.hu)
              DOKUMENTUMTÍPUS: könyvek (MSZ 3424/1-78 alapján)
              VERZIÓ: 1.1 (2003.10.01.) -->
```

```
<!-- EGYEDEKLARÁCIÓK -->
```

```
<!-- Külső egyedek nevének és helyének meghatározása -->
```

```
<!ENTITY % ISOnum PUBLIC "ISO 8879:1986//ENTITIES Numeric and
      Special Characters//EN" SYSTEM "/ENT/ISOnum.ent">
<!ENTITY % ISolat1 PUBLIC "ISO 8879:1986//ENTITIES Added Latin
      1//EN" SYSTEM "/ENT/ISolat1.ent">
<!ENTITY % ISolat2 PUBLIC "ISO/IEC 8859/2//ENTITIES Added Latin
      2//EN" SYSTEM "/ENT/ISolat2.ent">
<!ENTITY % ISOgrkl PUBLIC "ISO 8879:1986//ENTITIES Greek
      Letters//EN" SYSTEM "/ENT/ISOgrkl.ent">
<!ENTITY % ISOpub PUBLIC "ISO 8879:1986//ENTITIES Publishing//EN"
      SYSTEM "/EGYE/ISOpub.egy">
<!ENTITY % ISOdia PUBLIC "ISO 8879:1986//ENTITIES Diacritical
      Markss//EN" SYSTEM "/ENT/ISOdia.ent">
```

```
<!-- Külső egyedek bevonása -->
```

```
%ISOnum;
%ISolat1;
%ISolat2;
%ISOgrkl;
%ISOpub;
%ISOdia;
```

```
<!-- Belső egyedek definiálása -->
```

```
<!ENTITY % ism_elso_adatcsop "(pont, elso_adatcsop)*">
<!ENTITY % ism_eadatcsop "(ejel, parh_eadatcsop)*">
<!ENTITY % ism_cimadatok "(pvesszo, cimadatok)*">
<!ENTITY % opc_szerzosegi_kozles "(per, szerzosegi_kozles)?">
```

```

<!ENTITY % pegyebcimek "((kpont, egyebcim), (ejel, pegyebcim)*)">
<!ENTITY % palcimek "((kpont, alcim), (ejel, palcim)*)">
<!ENTITY % pfocimalcim "(ejel, pfocimek, (kpont, palcim)+)">
<!ENTITY % focimalcimegyebcim "focimek, (kpont, alcim)+, (kpont, egyebcim)+">
<!ENTITY % pfocimalcimegyebcim "(ejel, pfocimek, (kpont, palcim)+,
(kpont, pegyebcim)+)">

<!-- A párhuzamos adatelemek nem teszik lehetővé a csoportképzést -->
<!ENTITY % cimadatok_valt1 "(focimek, (ejel, pfocimek)*, (kpont,
alcim)*, (kpont, egyebcim)*)">

<!ENTITY % cimadatok_valt2 "(focimek, (ejel, pfocimek)*, (kpont,
alcim)+, %pegyebcimek;)">

<!ENTITY % cimadatok_valt3 "(focimek, %palcimek;, (kpont, egyebcim)*)">

<!-- A párhuzamos adatelemek lehetővé teszik a különböző szintű
csoportképzéseket -->
<!ENTITY % cimadatok_valt4 "(focimek, (kpont, alcim)+, %pfocimalcim;,
(kpont, egyebcim)*)">

<!ENTITY % cimadatok_valt5 "%focimalcimegyebcim;, %pfocimalcimegyebcim;">

<!ENTITY % ism_szerzosegi_adat "(pvesszo, szerzosegi_adat)*)">

<!-- A FŐSZERKEZET -->

<!-- A bibliográfiai leírás dokumentumelem -->
<!ELEMENT bibl_leiras (monografikus_szintu | osszefoglalo_szintu |
analitikus_szintu)>
<!ATTLIST bibl_leiras id ID #REQUIRED
autopszia (i | n) "i">

<!-- A monografikus szintű leírás dokumentumelem -->
<!ELEMENT monografikus_szintu (monografikus | monografikus_kotet)>

<!-- Egyetlen fizikai egységből álló könyv leírása -->
<!ELEMENT monografikus (adatcsopl, adatcsop2?, adatcsop3, adatcsop4,
adatcsop5?, adatcsop6*, adatcsop7?)>

<!-- Közös főcím nélküli vegyes gyűjteményekre is gondolva -->
<!ELEMENT adatcsopl (ujbekezdes, betuhely3., elso_adatcsop,
%ism_elso_adatcsop;)>

<!ELEMENT elso_adatcsop (eadatcsop, %ism_eadatcsop;)>

<!-- Közös főcím nélküli szerzői gyűjteményekre is gondolva -->
<!ELEMENT eadatcsop (cimadatok, %ism_cimadatok;, %opc_szerzosegi_kozles;)>

<!-- A cimadatok dokumentumelem definiálása, különböző változatai a párhuzamos
címek, alcímek és egyéb cimadatok kezelésére szolgálnak -->
<!ELEMENT cimadatok (%cimadatok_valt1; | %cimadatok_valt2; | %cimadatok_valt3;
| %cimadatok_valt4; | %cimadatok_valt5;)>

```

```

<!-- Vagylagos címet a főcím részeként kell leírni -->
<!ELEMENT focimek (Focim, Vagylagos_cim*)>
<!ELEMENT pfocimek (Pfocim, Pvagylagos_cim*)>

<!ELEMENT Focim (#PCDATA)>
<!ELEMENT Pfocim (#PCDATA)>
<!ELEMENT Vagylagos_cim (#PCDATA)>
<!ELEMENT Pvagylagos_cim (#PCDATA)>
<!ELEMENT alcim (#PCDATA)>
<!ELEMENT palcim (#PCDATA)>
<!ELEMENT egyebcim (#PCDATA | (nyszogljzel, egyebcim.mf, zszogljzel))>
<!ELEMENT egyebcim.mf (#PCDATA)>
<!ELEMENT pegyebcim (#PCDATA | (nyszogljzel, pegyebcim.mf, zszogljzel))>
<!ELEMENT pegyebcim.mf (#PCDATA)>

<!-- Szerzőségi közlés definiálása -->
<!ELEMENT szerzosegi_kozles (szerzosegi_adat, %ism_szerzosegi_adat;)>
<!ELEMENT szerzosegi_adat (szadat, (ejel, pszadat)*)>

<!ELEMENT szadat (#PCDATA | (nyszogljzel, szadat.mf, zszogljzel))>
<!ELEMENT szadat.mf (#PCDATA)>
<!ELEMENT pszadat (#PCDATA | (nyszogljzel, pszadat.mf, zszogljzel))>
<!ELEMENT pszadat.mf (#PCDATA)>

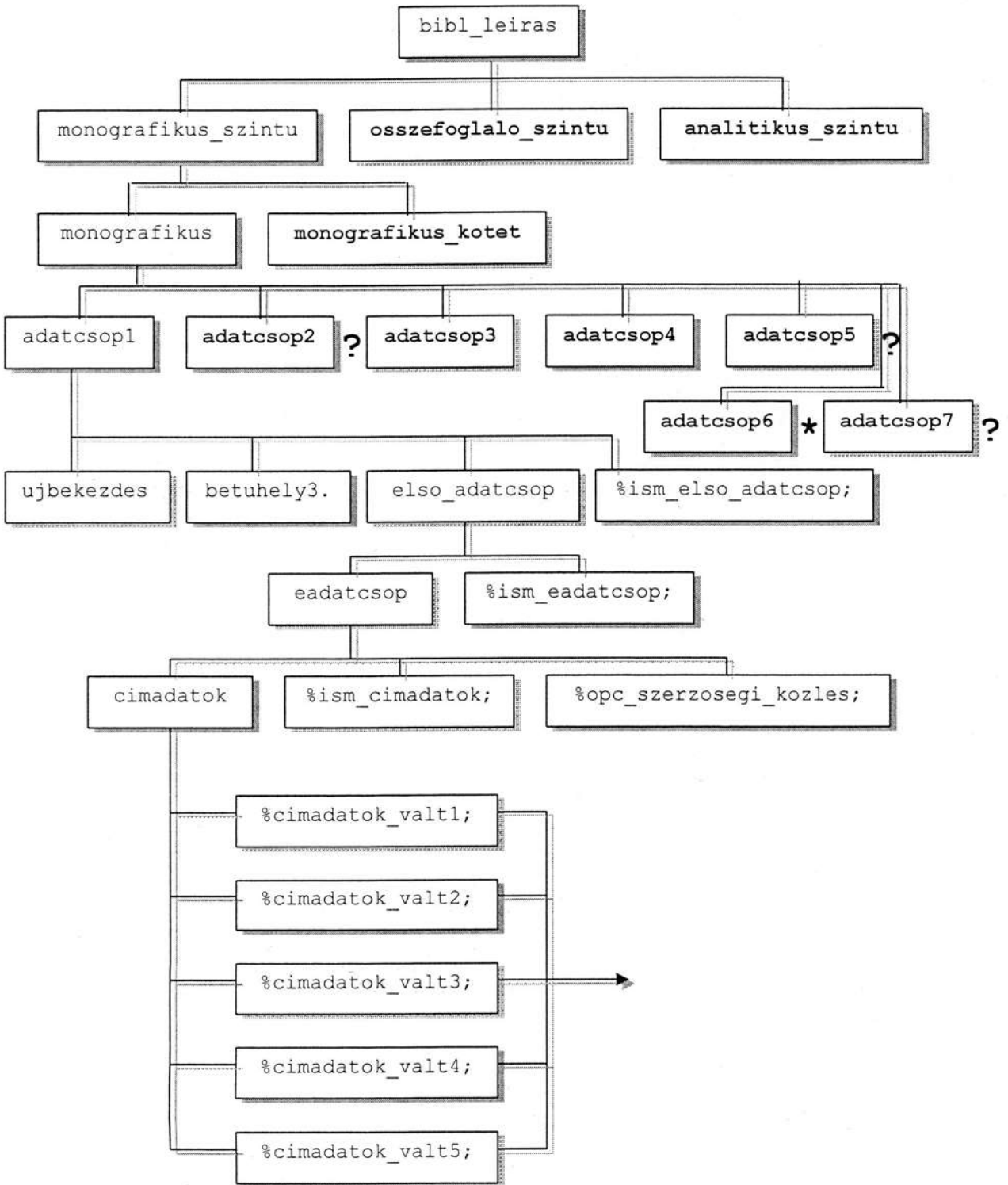
<!-- Egyezményes jelek definiálása -->

<!ELEMENT betuhely3. EMPTY>
<!ELEMENT acselvjel EMPTY>
<!ELEMENT ejel EMPTY>
<!ELEMENT felkjel EMPTY>
<!ELEMENT kerdojel EMPTY>
<!ELEMENT kpont EMPTY>
<!ELEMENT nykerekzjel EMPTY>
<!ELEMENT nyszogljzel EMPTY>
<!ELEMENT per EMPTY>
<!ELEMENT plusz EMPTY>
<!ELEMENT pont EMPTY>
<!ELEMENT pvesszo EMPTY>
<!ELEMENT ujbekezdes EMPTY>
<!ELEMENT vesszo EMPTY>
<!ELEMENT zkerekzjel EMPTY>
<!ELEMENT zszogljzel EMPTY>

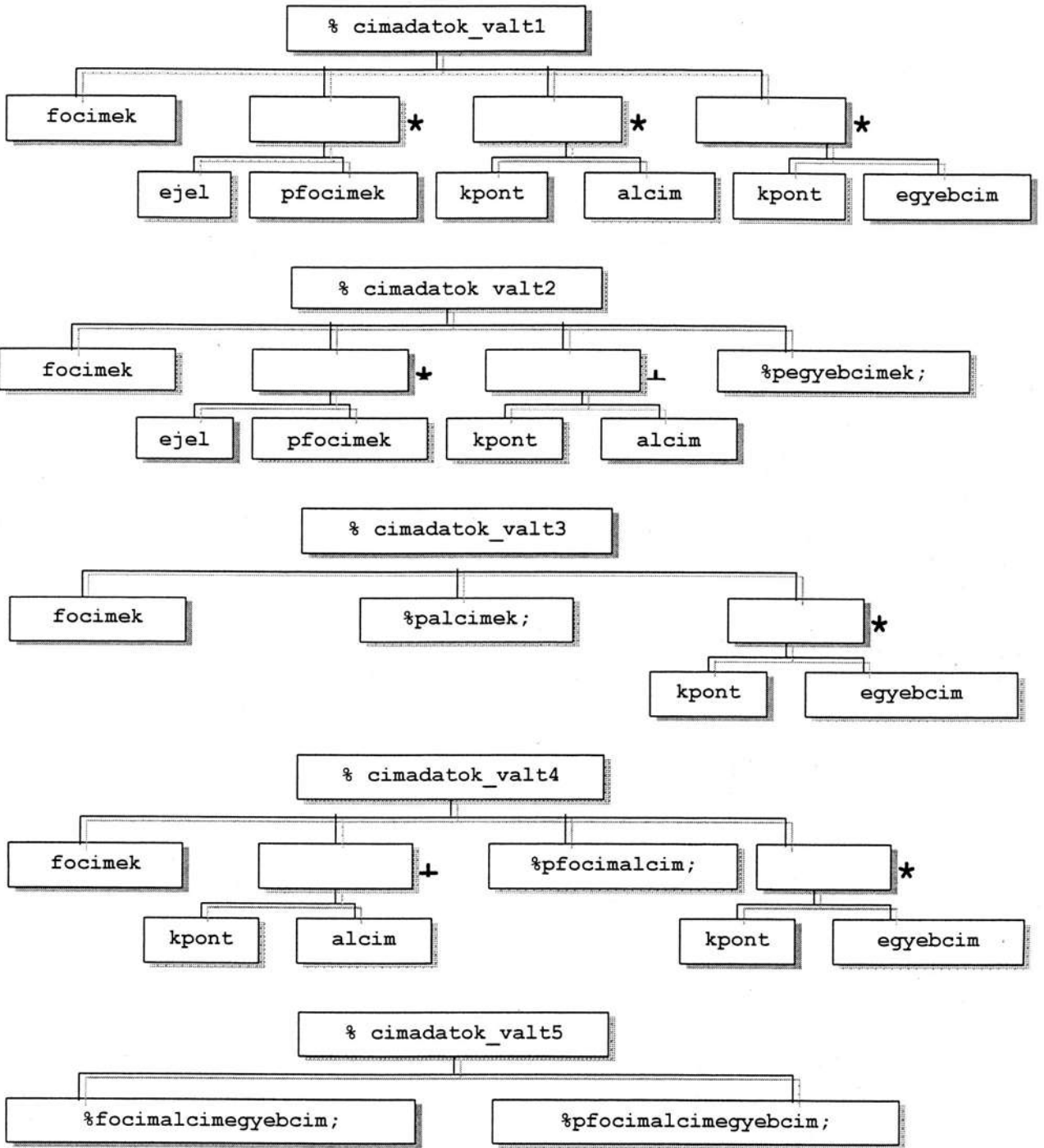
<!-- MÉG NINCS VÉGE! FOLYT. KÖV. -->

```





1. ábra A könyvek monografikus szintű leírása 1. adatcsoportjának grafikus ábrázolása a BDML DTD alapján



2. ábra A címadatok dokumentumelem grafikus ábrázolása

```

<adatcsop1>
<ujbekezes/>
<betuhely3./>
<elso_adatcsop>
  <eadatcsop>
    <cimadatok>
      <focimek>
        <Focim>Matematika</Focim>
      </focimek>
      <kpont/>
      <alcim>összefoglaló feladatgyűjtemény 10-14
        éveseknek</alcim>
    </cimadatok>
    <per/>
    <szerezosegi_kozles>
      <szerezosegi_adat>
        <szadat>Kosztolányi József [et al.]</szadat>
      </szerezosegi_adat>
      <pveszo>
        <szerezosegi_adat>
          <nyszogljzel/>
          <szadat.mf>az ábrákat rajzolta Vassné Németh
            Katalin</szadat.mf>
          <zszogljzel/>
        </szerezosegi_adat>
      </szerezosegi_kozles>
    </eadatcsop>
  </elso_adatcsop>
</adatcsop1>

```

3. ábra BDML alapú bibliográfiai leírás 1. adatcsoportjának egy előfordulása soremelésekkel és behúzásokkal tagolva

```

<adatcsop1><ujbekezes/><betuhely3./><elso_adatcsop><eadatcsop>
  <cimadatok><focimek><Focim>Matematika</Focim></focimek><kpont/>
  <alcim>összefoglaló feladatgyűjtemény 10-14 éveseknek</alcim>
</cimadatok><per/><szerezosegi_kozles><szerezosegi_adat><szadat>Ko
  sztolányi József [et al.]</szadat></szerezosegi_adat><pveszo>
  <szerezosegi_adat><nyszogljzel/><szadat.mf>az ábrákat rajzolta
  Vassné Németh Katalin</szadat.mf><zszogljzel/></szerezosegi_adat>
</szerezosegi_kozles></eadatcsop></elso_adatcsop></adatcsop1>

```

4. ábra BDML alapú bibliográfiai leírás 1. adatcsoportjának egy előfordulása tagolatlan elrendezéssel

## Irodalom

BAKONYI Géza: Tartalomszolgáltatás – egy folyóirat digitális feldolgozása. <http://nws.iif.hu/ncd2000/docs/eloadas/29/index.htm>

BRADLEY, Neil: Az XML-kézikönyv. Bicske: Szak, 2000. 564 p.

Extensible Markup Language (XML) 1.0., 2. ed. <http://www.w3.org/TR/REC-xml>

MILLER, Dick R.: XML and MARC: a choice or replacement? Chicago: ALA, 2000. 12 p. <http://elane.stanford.edu/lanauth/ALACHicago2000.html>

TENNANT, Roy: MARC must die. = Library Journal, 2002. 10. sz. <http://libraryjournal.reviewsnews.com>

Beérkezett: 2003. X. 31-én.