

dokumentumokat ismerteti *II. András* Barca-föld elajándékozásától *Georgicus Agricola*: 12 könyv a bányászatról és kohászatról c. munkájáig, a *Könyvtár* előbbi szerző Bermannus avagy beszélgetés az ásványok világáról munkájától *Zsámboky László*: Bányászat az Árpádok korában c. művéig 46 dokumentumot ismerteti; a *Fogalomtár* az aknáktól az ércetknőig, a farbortól a lejtősaknáig, a magmától a vulkánig 62 címszót tartalmaz; a *Képtár* a 16. századi bányászatot bemutató metszettől *Zsigmond* király pecsétjéig közel 300 illusztrációt közöl; a *Filmtár* a vulkáni tevékenységtől a hodrusbányái Mindszentek-bánya tárójáig öt részletet tartalmaz; mindezt mintegy 25 percnyi középkori zenei illusztráció színesíti.

A Pentium vagy kompatibilis processzor, 64 MB RAM, 5 MB HDD, true color, 800X600 képernyő, 16XCD-ROM, Windows-kompatibilis hangkártya, Windows'98 konfiguráció igényű kiadványt a soproni Központi Bányászati Múzeum Alapítvány adta ki 2002-ben, szerzői *Bircher Erzsébet*, *Horváth József* és *Szemán Attila* voltak, a műves kiadás az Enciklopédia Humana Egyesület munkáját dicséri. A művet a múzeum ([www.kozpontibanyaszati.muzeum.hu](http://www.kozpontibanyaszati.muzeum.hu)) forgalmazza 6000 Ft-ért.

Árkos Iván  
(BME OMIKK)

## Könyvtári rendszerek és az Unicode

Az ISO/IEC 10646: Universal Multiple-Octet Coded Character Set nemzetközi szabványt az egyetemes karakterkészletre, amely a világ valamennyi nyelvének megjelenítésére alkalmas, valamint az Unicode karakterkészleteit az Unicode Konzorcium dolgozta ki a legnagyobb számítógépes cégek (IBM, Apple, Adobe, Microsoft) és néhány nagy könyvtári rendszerfejlesztő (RLG = Research Libraries Group, OCLC és mások) részvételével. Habár az első verzió már 1991-ben megjelent, csak napjainkban kezdődik el teljes körű beépítése a jelentős programnyelvekbe, operációs rendszerekbe és böngészőkbe. Előnyei jobban megmutatkoznak, amióta a Code 2000 és 2001, a James Kass és a Microsoft Arial Unicode MS [1] betűkészletek rendelkezésre állnak.

Az új karakterkészlet használata fontos a gyűjteményekben különböző nyelvű (írásrendszerű) dokumentumokat őrző könyvtárak számára, ha webkatalógusaikban eredeti írásmódban kívánják feltüntetni az adatokat (pl. olyan tételeket akarnak megjeleníteni, amelyek egyaránt tartalmaznak arab és héber írásrendszerben leírtakat). A korábban használt ASCII, EBCDIC, ISO 8859, EACC 7 vagy 8 bites karakterkészletek 256 vagy kevesebb betű megjelenítésére voltak csak alkalmasak, míg a 16 bites Unicode több mint 65 000 karakter kódolását teszi lehetővé. Az UTF-8, UTF-16 és UTF-32 kódolással már a milliót is meghaladja a megjeleníthető karakterek száma.

Kritikus kérdés a könyvtárak számára, hogy hogyan fognak bibliográfiai rekordokat cserélni az Unicode használatával. A legnagyobb bibliográfiai

rekordszolgáltató cégek mind MARC 21 karakterkészlettel, mind Unicode-dal ajánlják a rekordcserét mindaddig, amíg a szolgáltatásokat igénybe vevő könyvtárak rendszereit fel nem készítik az Unicode fogadására.

### Irodalmi áttekintés

Számos cikk foglalkozik az Unicode szabvány magyarozatával, többek között *Ericksoné* [2] részletezi történetét és céljait, ugyanakkor kevés az olyan, amely a bevezetésével foglalkozik.

*Zhang* és *Zeng* [3] a könyvtárak előtt álló nehézségeket tárgyalják, különös tekintettel Kelet-Ázsiára, nevezetesen: a MARC szabványok nem támogatták az Unicode-ot, a jelentős rekordszolgáltatók (OCLC, RLG) nem tervezték a bevezetését, és csak kevés könyvtári rendszerfejlesztő cég kísérletezett vele. Hangsúlyozzák, hogy még az Unicode sem támogatja néhány speciális, személynévben, teljes szövegű dokumentumokban és ritka könyvekben használt kínai, japán és koreai karakterek megjelenítését.

*Aliprand* [4] azokról a kompromisszumokról ír, amelyeket a könyvtárosoknak meg kell kötniük egy-egy olyan mű leírásánál, amely egyedi jeleket, matematikai szimbólumokat tartalmaz. A karakterek kódolása mindenesetre egyszerűbb, mint glyphként (ugyanazon karakter több megjelenítési formában, pl. félkövéren, dölten) való rögzítésük és továbbításuk, hatékonyabb keresést biztosít, és kevesebb szoftverfejlesztést igényel.

*Chacra* [5] a karakterkezelés kérdésének összetettségével foglalkozik, és arra mutat rá, hogy az Unicode csak egy szempontból jelentős a nyelvek könyvtári kezelésében, más megközelítésben a nyelvek fordítására alkalmas eszközök, rendezési mechanizmusok, a speciális nyelvek szótani elemzése, a régi karakterkészletek megfeleltetésére használatos megoldások fejlesztéséről is gondoskodni kell.

Zhang és Zeng [6] számos kifogása, amely főként a dél-ázsiai nyelvek karaktereinek kezelésére vonatkozott, megoldódni látszik. A MARBI Character Set Subcommittee és az East Asian Character Set Task Force a megfeleltetések során számba vették a hiányzó és az összetett karaktereket, aminek eredményeképpen ezeket a változtatásokat az Unicode-ba be lehet vezetni.

Az OCLC és az RLG is komolyan foglalkoztak az Unicode-dal. Az RLG 2000-ben kezdte használni az Eureka-ban, adatbázisainak webalapú interfésznél. Ugyanakkor kifejlesztették a rekordok UTF-8-ba való konvertáló programjait, s amint igény van rá, a MARC 21 rekordokat Unicode-dal is tudják szolgáltatni, illetve a weben arra alkalmas böngészővel megjeleníteni. 2001 nyarán az OCLC bejelentette, hogy Oracle adatbázis-technológiára cserélte bibliográfiai adatbázisát, a WorldCat elavult rendszerét, s ez lehetővé teszi az Unicode-ra váltást. Természetesen a rendszer egy ideig párhuzamosan működik, de a rekordokat importáló könyvtáraknak adott időre fel kell készülniük az Unicode fogadására.

## Módszer

A szerző áttekintést kívánt nyújtani arról, hogyan állnak a piac rendszerfejlesztő cégei az Unicode bevezetése terén. Ehhez az American Library Association kiadványában, a *Library Systems Newsletter*-ben [7] évenként megjelenő összeállítást, valamint a Vanderbilt University munkatársának, *Marshall Breeding*-nek [8] az egyetem honlapján közölt adatait tekintette át, s ha szükségesnek mutatkozott, felkereste a cégek weblapjait is. 15 cégről gyűjtött így adatokat, és e-mailben kapcsolatot keresett a cégek Unicode-ban jártas szakértőivel is.

## Eredmények

A megkeresett 15 cég közül 11 adott választ a kérdésekre, ezek közül néhányan nevük elhallgat-

tását kérték. A 11-ből hat azt közölte, hogy már bevezették az Unicode-ot. E cégek java része több országban is rendelkezik székhellyel. A másik öt cégből négy 2003 végére tervezi a bevezetést. Általában kétféle megoldást alkalmaznak a könyvtári szoftverek terén. Az egyik esetben az eredeti (az adott nyelvben előforduló) nyelv speciális karaktereit építik be az Unicode-ba, a másikon már létező karakterkészletet, mint amilyen az EACC (East Asian Character Code) feleltetnek meg az Unicode-nak. Öt cég jelentette, hogy az eredeti nyelv karaktereinek beépítését választotta, illetve hogy az adatok egy részét Unicode-ban tárolja. A konkrét rákérdezésnél, hogy mely adatokat, már azt közölték, hogy valamennyit.

Hatból öt könyvtári rendszer használ Unicode-ot rekordszerkesztésre. Számos rendszer fejlesztett ki az Unicode előtt különböző módszereket az eredeti írásrendszer karaktereinek rögzítésére és megjelenítésére, de erre többé nem lesz szükség.

Hat rendszer és a legtöbb webböngésző támogatja az UTF-8 kódolást, s minthogy az online katalógusok tételeinek megjelenítéséhez is webböngészőket használnak, ésszerű, hogy a könyvtári rendszerek erre álljanak át. Ahogyan a szoftverfejlesztés világszerte átveszi az Unicode-ot, a jövőben az egyszerű rekordcserében már az UTF-8-on is túl kell lépni. Az UTF-16-os forrású rekordok importálásához ugyanis nem elegendő az UTF-8. Két rendszer az UTF-32-t támogatja más kódolási formák mellett. Az UTF-16 több mint egymillió karakter kódolását teszi lehetővé.

A rendezés és indexelés ugyancsak problémákat jelenthet a különböző nyelvek besorolási szabályainak eltéréséből adódóan. A programozásnál figyelembe kell venni a felhasználói elvárásokat. A vizsgált rendszerek közül három válaszolt igennel arra a kérdésre, hogy Unicode-ot használnak-e valamennyi írásrendszer rendezésére; öt igen válasz volt arra, hogy indexelésre is ezt használják; hat igen válasz érkezett, amikor a kérdés az volt, hogy az Unicode alapján lehet-e keresni. A kritikusként ítélt írásrendszerek közül legfontosabb a kínai, a japán és a koreai, illetve ezek változatai.

A vizsgálat kiterjedt arra is, hogy a cégek mely írásrendszereket támogatják. Az Unicode-ot bevezető cégek mindegyike jegyzékben sorolta fel ezeket (pl. latin 1, arab), valamint hogy melyek vannak fejlesztés alatt, és melyek szerepelnek tervükben. A legtöbb esetben nem tervezik a létező összes írásrendszer támogatását, csak azokra koncentrálnak, amelyeket a felhasználók igényelnek.

A két utolsó kérdés arra vonatkozott, hogy mi készítette a cégeket az Unicode bevezetésére vagy mellőzésére. Legtöbben a piaci kihívások miatt döntenek a bevezetés mellett, vagy azért, hogy megfeleljenek már meglévő üzletfeleik elvárásainak, vagy azért, mert a felhasználói kört soknyelvű országokban is bővíteni kívánják. Az egyik válaszadó arra hivatkozott, hogy az Unicode bevezetésére Ausztráliában a kínai népesség növekedése készítette. Ugyanez a helyzet az Egyesült Államokban, ahol a spanyol és az ázsiai nyelveket beszélők száma növekszik jelentősen. Amint a webböngészőknél elterjed az Unicode, szükségtelemmé válik a speciális adatrögzítő eszközök és szoftverek használata. Minél inkább megfelelnek a piaci termékek (idegen nyelvű kiadványok SGML-ben, XML-ben) az Unicode szabványnak, annál inkább felgyorsul az ezekkel kompatibilis rekordok cseréje, és az Unicode bevezetésének mértéke.

### Következtetés

A könyvtári rendszereket fejlesztő cégek egyre nagyobb érdeklődést mutatnak az Unicode iránt, s a felmérésre adott válaszaikból kiderül, hogy vagy bevezették már, vagy 2003 végére tervbe vették használatát. Nyilvánvaló előny mutatkozik mind a fejlesztők, mind a könyvtárak részére. A MARC 21 specifikációk már magukba foglalják az ISO/IEC 10646 használatának támogatását MARC 21 rekordok cseréjére; a karakterek megfeleltetése megtörtént, csakúgy, mint a webböngészők Unicode támogatása. Az újabb operációs rendszerek lehetővé teszik, hogy a használó saját anyanyelvének karaktereivel keressen. A fejlesztő cégek eltérő felkészültséget mutatnak, a vásárlók számára kulcskérdés, hogy mely írásrendszerek kezelésére van szükségük. Előfordul, hogy az üzleti vállalkozó támogatja az arab nyelvet, de ez nem jelenti szükségszerűen azt, hogy a kiterjesztett arab karakterkészletet, beleértve a perzsa nyelvet is, képes kezelni. Ugyanilyen fontos az eltérő betűrendezés kérdésének megoldása.

Több mint tíz év telt el az Unicode szabvány első megjelenése óta, de tényleges bevezetése csak az utóbbi években kapott lendületet. Számos tényező egybeesése segíti az Unicode előnyeinek kihasználását. Ebben a kis iparágban a versenyhelyzet diktálja, hogy a könyvtári rendszereket minél több országban piacképesse tegyék, azaz egyre több nyelvet tudjanak kezelni.

### Irodalom

- [1] KASS, James: Does your browser support multi-language? URL: <http://home.att.net/~jameskass>
  - [2] ERICKSON, Janet C.: Options for presentation of multilingual text: use of the Unicode standard. = *Library Hi Tech*, 15. köt. 3–4. sz. 1997. p. 172–188.
  - [3] ZHANG, Foster J.–ZENG, Marcia Lei: Multiscript information processing on crossroads: demands for shifting from diverse character code sets to the Unicode standard in library application. = *IFLA Journal*, 25. köt. 3. sz. 1999. p. 162–167.
  - [4] ALIPRAND, Joan: The Unicode standard: its scope, design principles, and prospects for international cataloging. = *Library Resources and Technical Services*, 44. köt. 3. sz. 2000. p. 160–167.
  - [5] CHACRA, Vinod: Unicode and the world's languages. Paper presented at the ALA Annual Meeting, San Francisco, June 16, 2001.
  - [6] i.m.
  - [7] *Library Systems Newsletter* (Chicago: Library Technology Reports, American Library Association). Az éves áttekintés általában a márciusi számban jelenik meg.
  - [8] BREEDING, Marshall: Library Technology Guides: key resources and content related to library automation. URL: <http://staffweb.library.vanderbilt.edu/breeding/ltg.htm>
- TULL, Laura: Library systems and Unicode: A review of the current state of development. = *Information Technology and Libraries*, 21. köt. 4. sz. 2002. p. 181–185./

(Berke Barnabásné)

## A freiburgi virtuális orvosi könyvtár

### Virtuális könyvtárak

Napjainkra a virtuális könyvtárak a könyvtárosi munka magától értetődő részeivé váltak. A Global Info, a német Digital Library Projekt már 1998-ban leírta a világon fellelhető valamennyi forrás felhasználásával történő informálódás jövőjét. A Szövetségi Oktatási és Kutatási Minisztérium 2003-ig

terjedő programja az információforrások munkahelyekről való hozzáféréseinek optimalizálását tűzte ki célul. A Deutsche Forschungsgemeinschaft 1998-as memoranduma a régióközi irodalomellátásban *virtuális szakkönyvtárak* létesítését kezdeményezte; időközben 14 ilyen virtuális könyvtár jött létre, közös portállal.