references to persons. The 9500 entries of the onomasticon contain 38,000 personal name forms including all the variants of spelling and pronunciation emerged in the sources. The personal name forms are specified by the date of year, by their original context and by the location of the quotation. The volume can be used in a digitized form owing to the CD-ROM enclosed. – In the present article one is informed about the antecedents of the onomasticon in the history of scholarship; in addition, the representatives of three branches of learning express appreciation for its utility in the fields of historical linguistics, historical onomastics, history, cultural history and philology of medieval Latin.

AZ „ÁRPÁD-KORI SZEMÉLYNÉVTÁR” SZÁMÍTÓGÉPES HÁTTERE

Minthogy a kötet CD-ROM-mellékletén eredeti formájában (is) hozzáférhető a teljes adatbázis, ezért azt a nyomtatott bevezetésben röviden ismertettük (30–2). Itt arról kívánunk szólni, hogy miért éppen azokat a számítógépes eszközöket alkalmaztuk, amelyeket, mi volt az egyes eszközök el előnye és hátránya, s ha most kezdene valaki hasonló föltételek mellett hasonló feladatba, a mi megoldásunkhoz képest milyen módosításokat javasolnánk neki.

A számítógépes eszközök kiválasztásakor a következő szempontokat igyekeztem figyelembe venni. 1. A rendszer legyen egységes, az adatbázisból a tördelt szöveg kézi beavatkozás nélkül legyen legenerálható kötegelt (batch) üzemmódban. (Kötegelt működésnek azt nevezzük, amikor a számítógépet egy feladatsor elvégzésére nem lépésenként utasítjuk, hanem a feladatokat egy parancsfájlba írjuk, és a parancsfájlt adjuk át földolgozásra a gépnek.) 2. Az alkalmazott fájlformátumok legyenek széles körben ismertek, az azokat kezelő eszközök pedig könnyen hozzáférhetők és az adatbázissal együtt továbbterjeszthetők. 3. A rendszer a munka során változó igényekhez adaptálható legyen; ez egyúttal a más, hasonló munkákhoz való fölhasználhatóságát is biztosítja. 4. Minden információ csak egy helyen szerepeljen (vagyis az adatállományokban csak egyetlen forrása legyen).

Az adatbázis-kezelésre a legkézenfekvőbb, de esetünkben legalábbis legkényelmesebb választás az Ashton-Tate dBASE program használata volt. A kéznél lévő 1992-es verzió már a munkálatok megkezdésekor sem volt fiatal, azonban teljesen megbízhatóan működött, korlátaiba (szinte) nem ütköztünk, és a további adatfeldolgozás során kötegelt módban egyszerűen használható. Minthogy a dBASE-adatbázisok tartalma adatvesztés nélkül könnyen kiírható szöveges fájlokba, kimenetét a TEX programmal lényegében közvetlenül fel lehet dolgozni (tördelni). A dBASE exportfunkciói meglehetősen kicsiszoltak, ezért csak csekély mértékű utólagos beavatkozásra volt szükség.

A munka első, adatgyűjtő szakaszában egyszerű, négymezős (névalak, év, textus, forrás) táblázatba kerültek az adatok. A névváltozatok bokrosítása, a névcikkek megszerkesztése során vált szükségessé annak rögzítése, hogy egy adott rekord mely névcikk alá sorolandó. Minthogy az azonos névváltozatok általában azonos névcikkbe kerültek, adatduplikálással járt volna, ha az erre vonatkozó információt a négymezős alaptábla

DOI: https://doi.org/10.29178/NevtErt.2005.34

bővítésével ottani újabb mezőbe helyeztük volna el. Tovább bonyolította a helyzetet, hogy a névcikkek egy részét (arab számmal jelölt) alcsoportokra kellett bontani. Létrehoztunk ezért egy második, hétmezős táblázatot (dBASE-adatbázist) is, melybe a névcikkek szerkezetére vonatkozó adatokat (címszó, címszóosztály: normál – női – genusnév, névalak, az alcímszó sorszáma, az alcímszó változatai, a főcímszó változatai, összevetendő címszavak) helyeztük el. A két táblázat közötti kapcsolatot a névalak biztosítja.

A címszavaknak ez a rendszere akkor működik helyesen, ha azonos (fő)címszavú rekordokban azonosak a főcímszóváltozatok, és ezeken belül azonos alcímszavú rekordokban azonosak az alcímszóváltozatok. Ez adatduplikálást jelentett volna és a szerkesztést is nehezítette volna. Az egyik megoldásnak az kínálkozott, ha újabb táblázatokat hozunk létre, és ezekbe áthelyezzük az ötödik és a hatodik mező tartalmát. Az egyszerűbb áttekinthetőség és szerkeszthetőség érdekében azonban inkább az adatbázis-kezelőn kívüli utat választottunk: a cleancimszok7.sh burokprogram által meghívott segédprogramok egyetlen (a sorrendben első) rekord adatai alapján kitöltik az ötödik és a hatodik mezőt, és bizonyos mértékű konzisztencia-ellenőrzést is végeznek.

Néhány esetben a névváltozatok előbb leírt szigorúan hierarchikus elrendezése nem volt lehetséges. Például a *Boc* általában a *Bok* változatának tekintendő, azonban egy előfordulása *Boch* alá volt sorolandó. Ilyenkor az ellentmondást a négymezős táblázat módosításával oldottuk föl: az utóbbi rekordban névalakként nem a tényleges *Boc*, hanem *Boch* szerepel. Ilyen esetekben automatikusan létrejön a gépi feldolgozás során egy *Boc* v. *Bok* utaló, mely a kettős utaló mellett fölöslegessé vált; törléséről a hackcimszok.flx program gondoskodik (vö. a kötet bevezetőjében: 32).

Mindkét táblázatot elsődlegesen szöveges (csv = comma separated values) formátumban szerkesztettük és javítottuk. Ennek előnye az volt, hogy ki-ki a kedvelt és jól ismert szövegszerkesztőjét használhatta a munkához, és hogy nem kellett tekintettel lenni az adatbázis-kezelőnek a szövegmezők terjedelmére vonatkozó korlátjára; hátránya pedig, hogy a mezőket jelölő idézőjelek és vesszők a szerkesztés során rossz helyre kerülhettek. Az ilyen hibák kiszűrésére egy-egy kis segédprogramot használtunk (chckutxt.pas, illetve chckutxt7.pas), melyek segítségével a javítás elég gyorsan elvégezhető volt. A segédprogramok nem kerültek a kötet mellékletére, azonban hozzáférhetők a http://www.cs.elte.hu/gam/ címen keresztül.

A munkának az adatgyűjtést követő szakaszaiban az adatbázis-kezelőt tehát már csupán kötegelt üzemmódban használtuk. A kommerciális programnak szabad és közvetlenül linux alatt futó helyettesítővel való kiváltása is fölmerült, mivel már akkor is több hasonló irányú fejlesztés zajlott (pl. http://www.harbour-project.org), azonban – legalábbis 2003-ban – egyik sem volt teljesen kompatibilis a dBASE-zel. (Problémát elsősorban a „set relation to” és a „_pform” utasítás okozott). A dbf-fájlok olvasására illetve konvertálására unix/linux alatt rendelkezésre áll például a dbf nevű program, l. http://www.anubisnet.de/products/dbf/.

A programszerűen vezérelhető tördelés legismertebb (és igen kifinomult) eszköze a TEX, illetve LATEX programcsalád. (Ezekről a világhálón pl. a www.tug.org honlapról kiindulva hatalmas mennyiségű információ érhető el. Könyv formában magyar nyelven a BUJDOSÓ GYÖNGYI – FAZEKAS ATTILA: *TEX kezdőlépések*. Bp., 1997, illetve a WETTL FERENC – MAYER GYULA – SZABÓ PÉTER: *LATEX kézikönyv*. Bp., 2004. áll rendelkezés-

re.) A kettős elnevezés kettős felhasználói felületet jelöl. A „plain TEX” változat felülete technikailag egyszerűbb, általa a „tördelő motor” közvetlenebbül vezérelhet ő és így gyorsabb; ezeknek az előnyöknek elsősorban a hasábok kiegyenlítésénél és az élőfej programozásánál vettük hasznát. A többhasábos szedésre és a szótári élőfejre természetesen már sok-sok éve közkézen forog több megoldás is, azonban PRŐHLE PÉTER programja kiküszöböli a korábbiak hibáit, ezért a „Magyarországi Középkori Latinság Szótár”-a szedésének és tördelésének az Akadémiai Kiadótól való átvétele (1992) óta az MKLSz. készítése során is ezt használjuk. Az algoritmus viszonylag könnyen volt módosítható a célból, hogy a hasábok tetején a fattyú (azaz kimenő) sorok félig automatikusan elkerülhet ők legyenek.

A teljesen automatikus megoldásnak lennének bizonyos elméleti buktatói (a kérdés hosszabb ideje kutatások tárgya, l. pl. KAREL SKOUPÝ: NTS: A new typesetting system. TUGboat 1998: 321; = http://www.tug.org/TUGboat/Articles/tb19-3/tb60nts.pdf.), a hagyományos szedői megoldások pedig a kötet formátuma és rögzített adattartalma miatt kevéssé voltak alkalmazhatók.

A névtárban idézett (közép)latin szövegek szavainak elválasztásánál az általános latin elválasztási szabályokat alkalmaztuk, azonban ezek érthető módon nem ölelik föl a névtárban szereplő, igen változatos ortográfiájú neveket. Az elválasztási hibák esetenkénti javítása helyett itt is általános megoldást alkalmaztunk: a szabályrendszert bővítettük oly módon, hogy ezeket a neveket is helyesen válassza el a program, a bizonytalan esetekben pedig inkább letiltottuk az elválasztást. (A TEX elválasztási algoritmusának ismertetése D. E. KNUTH: The TEXbook. Reading/Ma. 1984, számos további kiadás, H. függelék = ftp://ftp.dante.de/tex-archive/systems/knuth/tex/texbook.tex.) A szabályrendszer kiépítését és javítását is érdemes már a munkálatok során minél korábban megkezdeni, hogy a kinyomtatott levonatokban az elválasztást az egyéb hibákkal együtt folyamatosan javítani lehessen. – A kötet bevezetője és mutatói (1–37, 863–96.) tipográfiailag egyszerűbbek voltak, ezért ezeket a LATEX programváltozattal tördeltük.

A három mutató anyaga közvetlenül az eredeti (csv) adatbázisfájlokból készült, a lépéseket a makeatergo.sh burokprogram foglalja össze. (A pascal nyelvű segédprogramokkal együtt megtalálható a CD-mellékleten.) Az automatikus generálás nemcsak egyszerűbb, hanem a főszöveg legutolsó változtatásainak a mutatókba való pontos bevezetését is garantálta.

A címszavak rekordjait karbantartó, a mutatókat generáló és a tördelt névtári részt előállító burokprogram példái a unixosnak nevezett szellemiség alkalmazásának: az egyes részfeladatokat olyan eszközökre (awk, flex, gpc, sh) bízzuk, amelyek valamilyen okból éppen a legalkalmasabbak vagy a legkézenfekvőbbek, és ezek összefűzése révén összetettebb feladatok is viszonylag egyszerűen és csekély programozói ráfordítással megoldhatók.

Az alkalmazott rendszer leggyöngébb pontja természetesen az adatbázis-kezelő volt; képességeinek a határait több ponton súroltuk. Ma már könnyen hozzáférhetők és széles körben elterjedtek olyan szabad forráskódú adatbázis-kezelők, melyek kellőképpen megbízhatók és a miénkénél nagyságrendekkel nagyobb adathalmaz kezelésére is alkalmazhatók; elsősorban a MySQL programra (http://www.mysql.com/) gondolunk. Ez egyebek között a magyar karakterkészletet is kezelni képes (beleértve a rendezést; az

angol nyelvű dokumentáció a „collation” terminust alkalmazza a rendezési típusok jelölésére), s őt az utf8 és ucs2 unicode-reprezentációkat is. A mezőméretre vonatkozó felső korlát problémája is elegánsabban oldható meg MySQL használata esetén.

Amennyiben a rekordokban tárolandó szövegek sokszínűsége megkívánja a unicode használatát, akkor a tördeléshez a TEX program erre adaptált változata (omega, illetőleg aleph) használható. (L. pl. WETTL és társai i. m. 520–8. Az ott nem említett aleph nevű változat levelezőlistája: http://www.ntg.nl/pipermail/aleph/. Az omegából leszármaztatott aleph vállalkozás célja az, hogy a radikális változásoknak kitett omegával szemben stabil platform álljon a felhasználók rendelkezésére.)

MAYER GYULA

A HOLLAND NÉVTUDOMÁNY NAPJAINKBAN

1. A Névtani Értesítő Szemle és Figyelő rovatából, a konferenciakötetekből, valamint idegen nyelvű folyóiratokból (Onoma, Names, Namenkundliche Informationen stb.) rendszeresen értesülhetünk a külföldön folyó kutatásokról, így alkalomadtán a hollandiai és belgiumi tudományos munkáról is. A nemzetközi konferenciák szintén fórumai a tapasztalatcserének, ám kétségkívül nehéz lehetőséget találni arra, hogy tüzetesebben szemügyre vegyük egy-egy tőlünk sok tekintetben távoli nép névtudományának jelenlegi helyzetét. Úgy gondolom, hogy Németalföld (Hollandia, valamint Belgium északi része, azaz Flandria) is meglehetősen ismeretlen terület a hazai névtanosok számára. Bár HAJDÚ MIHÁLY új összefoglaló tankönyvében mind a tudománytörténeti, mind a személynevekkel foglalkozó fejezetekben foglalkozik a holland nyelvterülettel, célja magától értetődően nem lehetett a teljesség (HAJDÚ 2003: 17−36, 236−59).

Minthogy e nyelvterület a legkorábbi időkig visszatekintő tudománytörténetének és kutatási céljainak leírása jóval több oldalt venne igénybe, ez alkalommal azt ismertetem, mivel foglalkozik ma a névtudomány a Németalföldön. Az utóbbi bő húsz évről szólok csupán, az intézményi háttérről, a periodikákról, és egy figyelemre méltó, bár kevésbé tudományos igényű névtani munkáról. Az áttekintésben ANN MARYNISSEN és HANS BENNIS összefoglaló cikkét vettem elsősorban figyelembe (MARYNISSEN 1999: 77−87, BENNIS 2002: 129−35), amelyet más, elsősorban friss internetes források felhasználásával egészítek ki.

2. Intézményi háttér. – A névtant mint tanegységet egészen a közelmúltig három holland nyelvű egyetemen vehették fel a hallgatók. A Leuveni Katolikus Egyetemen Belgiumban, ahol a tizennyolc egyetemből nyolc kínál holland nyelvű képzést. Hollandiában, ahol összesen 16 egyetem található, az Amszterdami Egyetemen és a Groningeni Egyetemen szerepelt eddig tanegységként a névtudomány. A három egyetem mindegyikén hozzávetőleg 3500–4000 bölcsészhallgató tanul, akiknek 2001-ig a névtan is a választható tantárgyaik közé tartozhatott.