

BIBLIOGRÁFIAI ADATOK ELLENŐRZÉSE ÉS EGYSÉGESÍTÉSE A CDS/ISIS NAGYGÉPES VÁLTOZATÁVAL

Varga Sándor

Számítástechnika-alkalmazási Vállalat

A bibliográfiai adatbázisok építése sokféle tevékenységet foglal magában. A bibliográfiai tételek adatait tartalmazó rekordok elkészítése, a keresőtár kialakítása és a keresőállomány naprakészen tartása (az új indexelemek invertálása), a folyamatosan növekedő állományok védelme és növekedő helyigényük biztosítása a mágneses adathordozókon csupán néhány alaptevékenység a legfontosabbak közül. Valójában ezek a tevékenységek is több résztevékenységből állnak, amelyeket elsősorban az adott programsomag használatában jártas könyvtári és dokumentációs szakemberek, illetve az adatrögzítők és a számítástechnikusok látnak el. Ez utóbbiakra jól működő számítógépes rendszer esetében csak kevés feladat hárulhat.

E cikk keretében nem célom az összes tevékenység áttekintése. A közelmúltban egyébként e folyóirat olvasói *Válasz György* jóvoltából részletesen ismerkedhettek meg egy klasszikus, nagy adatbázis-építő cég, az INSPEC tevékenységével, az adatbázis-építés munkájának ottani szervezettségével [1]. Az olvasó számára világosan kitűnhetett a cikkből, hogy egy bibliográfiai adatbázis jól szervezett, olajozott építése a munkában részt vevő emberek és számítógépprogramok kifinomult munkamegosztásán alapszik. Az adatbázis-építés minőségének és automatizáltságának foka nagymértékben függ attól, hogy milyen mértékben képes az adott programrendszer mentesíteni az embert a mechanikus és ismétlődő feladatok alól, és milyen mértékben vállalja magára az ember által végzendő, de ellenőrizendő és ellenőrizhető tevékenységek gépi ellenőrzését.

E meglehetősen általános bevezető után tehát eljutottunk az ellenőrzés problémaköréhez. Ezen belül e cikk keretében csupán a *bibliográfiai adatok ellenőrzésével* és a bibliográfiai rekordban tárolt nem bibliográfiai jellegű, de az *adatbázis használhatósága szempontjából fontos*, ún. vezérlőinformációk helyességének ellenőrzésével foglalkozunk. Az egységesítés, pontosabban az egységesített adatokat tartalmazó állományok létrehozása és fejlesztése (authority control) szintén vizsgálódásom keretébe tartozik.

Azt az olvasót, akit néhány számítástechnikai terminus technicusom már most elriasztott, kérem, hogy legyen még egy kicsit türelmes. Az a célom, hogy a vázolt problémákat a gyakorló könyvtáros vagy dokumentációs szakember szemszögéből világítsam meg, s ennek során — remélem — az eddig leírtak is tisztázódnak.

Gyakorlati példán mutatom be az ellenőrzés szintjeit és konkrét formáit, feltételezve, hogy olyan könyvtári célú bibliográfiai adatbázist építünk, amely elsősorban a könyvek (monográfiák) rekordjait tartalmazza, de helyet kapnak benne a konferenciakiadványok egyes előadásainak analitikus szintű rekordjai is.

Philip Bryant említi "A katalógus" című, jelentős, a hetvenes évek angol–amerikai szakirodalmát áttekintő tanulmányában, hogy sokan összetévesztik a *bibliográfiai tétel* és az ettől világosan elkülönítendő *bibliográfiai rekord* fogalmát [2]. Ezt a különbségtételt rendkívül fontosnak tartom, elsősorban gyakorlati szempontból. Fontos szem előtt tartanunk, hogy az ún. "új címléirési szabványok" (MSZ 3423, MSZ 3424, MSZ 3440 stb.) nem a rekordra, hanem a *tételre*, illetve a tételekből felépített jegyzékekre, kártyasorozatokra (pl. kártyakatalógus) vonatkoznak. Jól szemlélteti a különbséget azoknak az — egyelőre sajnos nem magyar — könyvtáraknak a gyakorlata, akik a könyv beszerzésekor annak bibliográfiai rekordját is beszerzik és betöltik saját házi adatbázisukba, s *ebből* állítják elő a bibliográfiai tételeket saját online vagy nyomtatott (kötet- vagy kártya-) katalógusuk számára. A *tétel* tehát ilyen értelemben a rekordban lappangó (csak számítógéppel olvasható) adatoknak vagy azok egy részének adott formában történő *megjelenítése képernyőn vagy papíron* stb., emberi szemmel olvasható formában. Természetesen *ellenőrzési* célból gyakran a rekord egész adattartalmát meg kell jeleníteni. Ilyenkor nemcsak a rekordban szereplő adatok, hanem a mezőazonosítók (hívójelek vagy "tag"-ek), almezőjelek stb. is megjelennek. Ez az úgynevezett *ellenőrző* vagy *korrektúra* formájú megjelenítés.

Az ellenőrzés automatizálásának lehetőségei természetesen nagymértékben függenek az adott programrendszer lehetőségeitől. Szemléltetésre a CDS/ISIS programcsomag nagygépes változatának 4.5 számú verzióját választottam, mivel egyre több helyen alkalmazzák Magyarországon, és ennek alkalmazásában van némi jártasságom. Az UNESCO által tesztelésre megküldött 4.6-os verzió néhány újdonságára (pl. azonnali ellenőrzés online bevitel esetén) szintén utalni fogok. A CDS/ISIS általános ismertetésére nem térek ki, átfogó programozói szemléletű ismertetést talál az olvasó *Huba Zoltán* cikkében [3], de felhasználói szempontból mikrogépes változatot bemutató alkalmazási példa is sok fontos tudnivalót tartalmaz [4].

Miért van szükség ellenőrzésre?

A válasz e kérdésre kézenfekvőnek tűnik: hogy ne kerüljenek hibás adatok az adatbázisba. Ez a válasz viszont egy újabb kérdést vet fel: mit tekintünk hibának egy bibliográfiai adatbázis esetében?

Ha a kérdést közelebbről megvizsgáljuk, kitűnik, hogy a válasz nem éppen egyszerű, hiszen az adathiákat súlyozni kell abból a szempontból, hogy milyen mértékben zavarják vagy teszik kellemetlenné az adatbázisban tárolt adatok kikeresését, illetve mennyire lassítják, nehezítik vagy teszik kellemetlenné a keresést végző ember – a felhasználó – tevékenységét.

Induljunk ki egy konkrét példából! Egy könyvet kell katalogizálnunk, s mivel – sajnos – még nem áll rendelkezésre a könyv adatait tartalmazó eredeti bibliográfiai rekord, minden adatot nekünk kell bevinnünk, amely akár a bibliográfiai leíráshoz, akár a tartalmi kereséshez (tárgyszavak, szakjelzetek) szükséges.

Az 1. ábrán látható a könyv címdala és annak verziója. Az itt látható adatok és az oldalszámzás alapján készítjük elő a bevitelre szánt adatokat.

A kész rekord korrekció formában kiírva a 2. ábrán látható. Ezzel kapcsolatban megjegyzem:

- ◆ A rekordban valójában nem szerepelnek a mezők nevei, csupán numerikus azonosítói (hívójeljei). Ezek az azonosítók sem részei az ún. adatmezőnek, hanem csak a rekord úgynevezett directory-jában (tartalomjegyzékében) szerepelnek, ellentétben az *almezőjellel*, amely része az adatmezőnek. Például a 02-es hívójelű "ISBN és ár" mezőben a $\neg a$ jelkettős az ISBN almezőjének jele. Látható, hogy az almezőjel mindig a vonatkozó almező előtt áll, azt vezeti be. A 24-es "Cím" mezőben a főcím almezőjele $\neg a$, az alcímé $\neg b$. Az egyszerűség kedvéért más mezőkben mellőz-

tük az almezőjeleket. Egyszerűbb szerkezetű adatbázisok ilyen formában is létrehozhatók a CDS/ISIS-szel.

- ◆ A magyar nyelvű címfordításban és a tárgyszavakban megfigyelhető hárombetűs kódokkal, az ékezetes betűk belső kódjaival részletesen foglalkozik *Huba Zoltán* említett cikke [3].
- ◆ A 69-es "Tárgyszavak" mezőben látható $\langle \rangle$ jeleket arra a célra használjuk, hogy a közjük zárt szövegrészeket a későbbiekben mint ismérveket használhassuk fel a dokumentumrekord kikeresésére. A szótárban ugyanis az "operációs" és "rendszer" szavak nem külön-külön, hanem "operációs rendszer" deskriptorként szerepelnek, s ezt valahonnan a feldolgozó (invertáló) programnak is tudnia kell.

Térjünk vissza ezek után fő témakörünkhöz, az ellenőrzés kérdéséhez. Tegyük fel, hogy a 2. ábrán bemutatott tételt a munkaidő végén már ugyancsak fáradt kolléga a 3. ábrán látható módon, számos hibával rögzítette. Melyek ezek a hibák, és mi a következményük az adatbázis minősége, használhatósága szempontjából?

Az első csoportba azok a hibák sorolhatók, amelyek nem zavarják az adatbázis rendeltetésszerű használatát, csupán sérthetik a felhasználó "nyelv-érzékét". A 21-es "Címfordítás" mezőben a hosszú magyar ó betű helyett csupán "o" szerepel, s ez bizony szemet szúrhat egy felhasználónak, ha valamely keresés eredményét ékezetes nyomtatóval készített listán kapja meg. Attól függően, hogy milyen nyomtató áll rendelkezésre, ugyanebből a rekordból készíthető ékezetes, kis- és nagybetűs, ékezet nélküli kis- és nagybetűs vagy csupán ékezet nélküli nagybetűs tétel. Megfelelő eszközök hiányában Magyarországon sajnos az utóbbi gyakran előfordul.

Hasonló jellegű *szépséghiba*, hogy a "Megjelenési adatok" mezőben a kiadó neve, Addison-Wesley helyett Adison-Wisley szerepel. Ezt talán nem is veszi észre minden felhasználó, és a hiba – eltekintve az esetleges kiadó szerinti keresés, ill. rendezés torzulásaitól – csak "szépséghiba". Az ebbe a csoportba tartozó hibák tehát lényegében azonos jellegűek a *nem értelemzavaró sajtóhibákkal*, helyesírási hibákkal. Nem okoznak nagyobb bajt, de azért még hibák maradnak.

A második csoportba – az enyhébbtől a súlyosabb esetek felé haladva – az *értelemzavaró*, de az adatbázis rendeltetésszerű használatát még nem akadályozó hibákat sorolom. Ilyenek adódhatnak a szöveg egyszerű elírásából is, de almező-szerkezetű mezők esetén "ravaszabb" módon is jelentkezhetnek, mint például a 24-es "Cím" mező "alcím" almezőjében. A leíró vagy az adatrögzítő kihagyta az almezőt azonosító *b* betűt. Ha az almezőkódokat

ESSENTIAL CONCEPTS OF OPERATING SYSTEMS USING IBM MAINFRAME EXAMPLES

C. K. Yuen

National University of Singapore

© 1986 Addison-Wesley Publishers Limited
© 1986 Addison-Wesley Publishing Company, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the publisher.

The programs presented in this book have been included for their instructional value. They have been tested with care but are not guaranteed for any particular purpose. The publisher does not offer any warranties or representations, nor does it accept any liabilities with respect to the programs.

Cover design by Ken Arlotte.
Typeset by Columns, Reading, Berks.
Printed in Singapore by Toppan.

British Library Cataloguing in Publication Data

Yuen, C. K.
Essential concepts of operating systems : using IBM
mainframe examples.
1. IBM Computer 2. Operating systems (Computers)
I. Title
005.4'3 QA76.8.1102

ISBN 0-201-12917-5

Library of Congress Cataloguing-in-Publication Data

Yuen, C. K.
Essential concepts of operating systems.

Bibliography: p.
Includes index.
1. Operating systems (Computers) 2. IBM computers -
Programming. I. Title.
QA76.76.063Y84 1986 005.4'3 85-28766
ISBN 0-201-12917-5 (U.S.)

ABCDEF 89876



ADDISON-WESLEY PUBLISHING COMPANY

Singapore • Wokingham, England • Reading, Massachusetts • Menlo Park,
California • Don Mills, Ontario • Amsterdam • Sydney •
Tokyo • Mexico City • Bogotá • Santiago • San Juan

1. ábra A katalogizálendő könyv címdala és verzója

nem ellenőrizzük, a kiíróprogram "átugorja" a jel utáni *u* betűt (hiszen az almezőt mindig egy jel és az azt követő betű vagy számjegy azonosítja, és az almezőjel a tétel formában nem íródik ki, csak az ellenőrző formában), s így a következő cím jelenik meg a képernyőn vagy a papíron:

Essential concepts of operating systems : sing
IBM mainframe examples

A harmadik csoportba olyan hibákat sorolok, amelyek már a rekord kikereshetőségét is érintik. Ilyen hiba például esetünkben a 10-es mező elhagyása, hiszen ebben szerepelt a szerző neve, mégpedig keresésre alkalmas besorolási adat formájában. Ha a rekordot nem javítjuk ki, hanem véglegesen ebben a formában kerül az adatbázisba, Yuen professzor könyve elérhetetlen lesz mind a szerzői név szerinti online kereső, mind a főtétel elsődleges besorolási adata alapján rendezett nyomtatott bibliográfiát böngésző könyvtárlátogató számára.

Végezetül a legveszélyesebb hibák azok, amelyek a rekordot szinte eltüntetik. Az ilyen jellegű hibák azt eredményezik, hogy a rekord ugyan benne van az adatbázisban, de mivel egyetlen ismérve szerint sem található meg, gyakorlatilag hozzáférhetetlenné

válak a kereső számára. Ilyen helyzet adódhat elő például, ha adatbázisunkat csak a szerzői nevek és a tárgyszavak szerinti keresésre terveztük, de egy szerencsétlen rekord esetében nemcsak a szerzői nevet felejtjük el leírni adatbevitelkor, hanem még a tárgyszavakat is rossz helyre, mondjuk a "Megjegyzés" mezőbe írjuk. Ez utóbbi mezőt nyilván nem "invertáljuk", hiszen a benne tárolt ismérveket nem szándékozunk beépíteni a keresőállományba. Így a tárgyszavak benne vannak ugyan a rekordban, de keresési szempontból teljesen használhatatlanok.

E meglehetősen ritkán előforduló komplex hibánál talán még veszélyesebb, ha a rekord valamely kulcsfontosságú, kiírást vezérlő adatának hibás beírása idézi elő a rekord, illetve helyesebben a megjelenítendő tétel eltűnését. Példánkban a 00-ás mező 7-edik pozícióján áll a bibliográfiai szint kódja (M monografikus, A analitikus stb.). A kiíróprogramot vezérlő nyelv első utasításai a bibliográfiai szint kódja szerint más és más kiírási formát írnak elő a monografikus tételek és mást az analitikus tételek, esetünkben konferencia-előadások számára (a konferencia-előadást tartalmazó kötet bibliográfiai adatait ugyanis nem az előadást leíró rekord, hanem

A mező neve	Mezőazonosító	A mező tartalma
Dokumentumtípus*	00/6	K
Bibliográfiai szint*	00/7	M
ISBN és ár	02	7a0-201-12917-5
A főtétel elsődleges besorolási adata: személynév	10	Yuen, C. K.
Címfordítás	21	Az oper@7aci@7os rendszerek alapfogalmai : IBM nagy@7epes p@7eld@7ak alapj@7an
Cím és szerzőségi közlés	24	7aEssential concepts of operating systems 7using IBM mainframe examples 7 cC. K. Yuen
Megjelenési adatok	26	Singapore : Addison-Wesley, cop. 1986
Terjedelem	30	VIII, 199 p.
Raktári jelzet	37	C-19782
Tárgyszavak	69	<oper@7aci@7os rendszer> <IBM> <multiprogramoz@7as> <virtu@7alis t@7arol@70> <lapszervezet@7-es>
Nyelvi kód	98	eng
Megjelenési év	99	1986

* A 00 azonosítójú mező fix hosszúságú elemeire a mező kezdetétől számított relatív helyükkel hivatkozunk. Pl.: 00/7, azaz a 7. pozíciót követő elem.

2. ábra Az 1. ábrán bemutatott könyv bibliográfiai rekordja

A mező neve	Mezőazonosító	A mező tartalma
Dokumentumtípus	00/6	K
Bibliográfiai szint	00/7	N
ISBN és ár	02	7a0-201-12917-5
Címfordítás	21	Az oper@7aci@7os rendszerek alapfogalmai : IBM nagy@7epes p@7eld@7ak alapj@7an
Cím és szerzőségi közlés	24	7aEssential concepts of operating systems 7using IBM mainframe examples 7 cC. K. Yuen
Megjelenési adatok	26	Singapore : Addison-Wesley, cop. 1986
Terjedelem	30	VIII, 199 p.
Raktári jelzet	37	C-19872
Megjegyzések	50	<oper@7aci@7os rendszer> <IBM> <multiprogramoz@7as> <virtu@7alis t@7arol@70> <laptchnika>
Nyelvi kód	98	hun
Megjelenési év	99	1986

3. ábra A hibásan rögzített bibliográfiai rekord (A hibás részeket bekereteztük.)

a könyvet monografikus szinten leíró rekord tartalmazza, s ezért kiírni is abból kell). Tegyük fel, hogy nem ellenőrizzük a bibliográfiai szint kódjának érvényességét, s a program a következőképpen működik: **M** esetén kiírja a tételt a monografikus leírás előírásai szerint, **A** esetén az analitikus leírás szerint stb. Mivel az **M** helyett tévesen leírt **N**-re nincs felkészítve, ilyenkor nem ír ki semmit, a tétel tehát *látszólag* elveszett, a kereső számára pedig ténylegesen elveszett.

A fentiekkel — remélem — meggyőzően érzékeltetem, hogy a számítógépes bibliográfiai adatbázisok rekordjaiba igen sokféle hiba kerülhet az emberi munka jellegéből adódó feledékenység, fáradtság, váratlan munkamegszakítás stb. következtében. Ezek a hibák — további következményeiket tekintve — nem egyformák, ezért az adatbázis tervezőinek és üzemeltetőinek elsősorban a keresés eredményét torzító és a súlyosan értelemzavaró hibák kiszűrésére kell felkészülniük, s nem szabad beérniük a szépséghibák kozmetikázásával.

Milyen hibák szűrhetők ki automatikus ellenőrzéssel?

Példaként térjünk vissza a 3. ábrán látható tétel néhány hibás adatához. Nem nehéz belátni, hogy elméletileg az összes rögzített szót, kifejezést egybevetethetnénk valamilyen óriási ellenőrző szótár szavával, kifejezéssel, s ez hibaként jelezhetné az eltéréseket. Ebben az esetben bizonyára hibaüzenetet kapnánk az "operáció" vagy az "Adison-Wisley" szavak, ill. szókapcsolatok ellenőrzésekor. Eddig rendszerben is lenne a dolog.

De mi történik akkor, ha — tételezzük fel — az adatrögzítő hibásan "rendszerek" helyett "rendszer"-t ír, vagy pl. kihagyja a "mainframe" szót. Ilyen hibát még az angol és/vagy a magyar nyelv szintaktikai és szemantikai szabályait ismerő "szuperintelligens" program sem tudna kideríteni.

Ellenőrző programunk ugyancsak hatástalannak bizonyulna, ha pl. nemcsak egy "Addison-Wesley", hanem egy "Adison-Wisley" nevű kiadó is létezne. Programunknak fogalma sem lenne róla, hogy mikor melyik a jó. Én viszont kézbe veszem a könyvet, ránézek a címdalra, utána a kinyomtatott rekordra, és máris észlelem a hibát (tükrözött adatok hibás írása esetén). Összefoglalva: a szöveghibák — esetünkben ezek a dokumentumleírás adataiban előforduló hibák — csak igen rossz hatásfokkal és nagy költséggel ellenőrizhetők gépi úton. Kiszűrésük, csökkentésük sokkal hatékonyabb módja a hagyományos korrektúraolvasás, illetve a bevitt adatok egybevetése a leírt dokumentummal.

Azok az adatok, amelyeknek csak meghatározott értékei lehetnek — pl. egy tezausz deszkriptorai, szabványosított nyelvi kódok, országcódok, megjelenési évszámok, szakjelzetek stb. —, s ezért érvényességi (validitási) szempontból jól ellenőrizhetők, semmiképpen sem mondhatók megbízhatónak a kiegészítő *intelligens emberi tevékenység*, a hozzáértő korrektúraolvasás nélkül. Hiszen egy 1986-ban megjelent könyv megjelenési évét tévesen írhatjuk 1989-nek, de írhatjuk 1968-nak is. Ha az évszám érvényességét mondjuk az 1960–1987 intervallumban fogadja el az automatikus ellenőrzés, akkor *csak az 1989-es évszámot fogja hibaként jelezni!*

Mire jó akkor a számítógépes ellenőrzés? A jelzett korlátokon belül sok mindenre. Főleg akkor, ha nemcsak egy adat önmagában vett érvényességét képes vizsgálni, hanem összefüggését is más adatokkal. Lehet például, hogy egy nyelvi kód önmagában helyes, pl. "hun", azaz a magyar nyelv kódja, de az adatbevitt végző személy tévesen egy angol nyelvű könyvhöz rendelte hozzá (előtte ugyanis vagy egy tucat magyar nyelvű könyv adatait rögzítette).

Ha ellenőrző programunknak meg tudjuk mondani, hogy vizsgálja meg a nyelvi kódot, és ha az a magyar nyelv kódja (hun), akkor ellenőrizze, hogy a 21-es mező (a főcím magyar nyelvű fordítása) nincs-e kitöltve (ami tévedés lenne), és ha igen, akkor a rekordot hibásnak jelölje meg, nyomtasson ki hibaüzenetet, miszerint "TÉVES CÍMFORDÍTÁS VAGY HIBÁS NYELVI KÓD", akkor programunk alkalmas az adatok *konzisztenciájának* ellenőrzésére, s ez ugrásszerűen megnöveli a gépi ellenőrzés eredményességét. Nos, a CDS/ISIS 4.5-ös verziója *képes* az ilyen típusú ellenőrzésre.

Néhány ellenőrzési módszer szemléltetése a CDS/ISIS programeszközeivel

Néhány évvel ezelőtt a *Nemzetközi Tudományos és Műszaki Információs Központ* (Moszkva) munkatársai cikket jelentettek meg a CDS/ISIS ott kidolgozott első ESZR-változatáról, amely az eredeti CDS/ISIS 3.2 verzióján alapul [5]. Az ott röviden vázolt ellenőrzési és hibajavítási lehetőségek az újabb verziókban jelentős mértékben kibővültek. Különösen jelentősnek tartom a következőket:

- ◆ a HA... AKKOR... EGYÉBKÉNT (IF... THEN... ELSE...) típusú konzisztencia-ellenőrzési lehetőséget,
- ◆ a teljes képernyős módosítási lehetőséget (Full Screen Editor) online javításkor,
- ◆ a 4.6-os verzió már megfelelően működő új lehetőségét: a beépített *ellenőrzési funkciók* azonnali online működését.

Az alábbi példák ellenőrzési megoldásai ma már online bevétel, ill. módosítás esetén is azonnal működnek.

Csökkenő súlyossági sorrendben nézzük meg sorra a 3. ábrán bemutatott adathibákat, és vizsgáljuk meg, hogy milyen ellenőrzési módszerrel küszöbölhető ki. Felhívom az olvasók figyelmét, hogy az itt leírt példák nem okvetlenül a probléma egyedüli üdvözítő megoldásai. A CDS/ISIS – hosszú fejlesztési múltja következtében – az ellenőrzési eszközök széles választékát kínálja, s ugyanaz a hibatípus gyakran többféle ellenőrzési módszerrel is elhárítható.

1. példa: a megjegyzés mezőbe írt tárgyszavak esete

Feladat: Ne kerülhessen rekord az adatbázisba kitöltött tárgyszómező nélkül, mert tárgyszavak híján adatbázisunk a művek tartalma alapján egyáltalán nem kereshető. A tárgyszómezőben csak a tezauszban meghatározott deskriptorok szerepelhetnek.

Megoldás: Az adatbázis ún. mezőleíró táblájába (Field Definition Table) ellenőrző (érvényességi, validáló) utasítások írhatók. A következő utasítást írhatjuk le: "Ha a 69-es mező hiányzik, a rekord hibás, ird ki a HIÁNYZIK A TÁRGYSZÓMEZŐ üzenetet".

Ennek pontos formája:

```
IF 69A THEN 'HIANYZIK A TARGYSZOMEZO' FI
```

Ugyancsak a mezőleíró táblában jelezzük a feldolgozó program számára, hogy a mezőben szereplő < > közötti szavakat, ill. szókapcsolatokat melyik szótár alapján kell ellenőrizni. Ez a szótár lehet csak ellenőrzésre használt állomány (Look up file), de lehet pl. egy adatbázis invertált állományának szótárrésze is.* Ha a mezőleíró táblánkban a 69-es mező ilyen típusú ellenőrzését írjuk elő, a "laptechnika" szót mint nem deskriptort vissza fogja utasítani, azaz hibaüzenetet küld, és a rekordot hibásnak jelöli meg.

2. példa: a bibliográfiai szint hibás kódja miatt "eltűnő" tételek esete

Feladat: Csak érvényes kódok kerülhessenek a mezőbe. Azt is célszerű ellenőrizni, hogy a mező mindig ki legyen töltve.

Megoldás: Mivel csak néhány kódról van szó, nem érdemes szótár alapján ellenőrizni. Egyszerűbb az ilyen utasítás a mezőleíró táblában:

```
IF 00.7='A' OR 00.7='M' THEN ELSE  
'HIBAS BIBLIOGRAFIAI SZINT' FI
```

vagyis: "Ha a 00 azonosítójú mező 7. pozíciója utáni jel A vagy M, akkor ne csinálj semmit, de ha nem így van, akkor jelezd a hibát."

* Például a tezausz adatbázisé.

3. példa: az elveszett szerző esete

Feladat: Ne fordulhasson elő, hogy valamely kulcsfontosságú besorolásiadat-mezőt elfelejtik kitölteni.

Megoldás: Talán már az olvasó is megfogalmazta a mezőleíró tábla újabb ellenőrző utasítását. Mondjuk:

```
IF 10A THEN 'HIANYZIK AZ ELSODLEGES SZERZO' FI
```

vagy valami hasonló.

Álljunk meg egy pillanatra! Nem minden főtételek szerzői személynév az első besorolási adata. Lehet testületi név, egységesített konferencianév, egységesített cím, sőt maga a főcím is a főtétel besorolási adata. Esete és dokumentumleírás precizitása válogatja.

Feltételezve, hogy a felsorolt eseteknek megfelelő adatok rendre 11-es, 12-es és 13-as mezőben fordulnak elő, továbbá a 24-es mezőben szereplő főcím elsődleges besorolási adat esetén nem az *a* almezőjelet, hanem az *e* almezőjelet kapja, az alábbi utasítást javaslom:

```
IF 10A AND 11A AND 12A AND 13A AND 24.1-7e  
THEN 'NINCS FOTETEL BESOROLASI ADAT' FI
```

vagyis: "Ha a felsorolt mezők hiányoznak, ráadásul a 24-es mező 2. karaktere nem *e* (vagyis nem *e* az első almező jele), akkor a rekord hibás, mert hiányzik a főtétel elsődleges besorolási adatának feltüntetése."

4. példa: a "dalolj IBM nagygépes példát!" esete

Feladat: Ne fordulhasson elő, hogy a kiíróprogram az almezőkód elhagyása miatt az almező tartalmának első betűjét (karakterét) tekinti almezőkódnak, s ily módon "módosítja" a szöveget (mint esetünkben "using" helyett "sing").

Megoldás: CDS/ISIS eszközökkel sajnos csak félmegoldást tudok javasolni.

A mezőleíró táblán ugyan meghatározhatjuk a mezőben előforduló almezőket (esetünkben: *e*, *a*, *b*, *c*), előírhatjuk sorrendjüket és kötelező vagy opcionális jellegüket, s mindezt a rendszer automatikusan ellenőrizni is fogja, de semmiféle védelmet nem nyújt arra az esetre, ha a mezőben tárolt szöveg első betűje történetesen megegyezik az elhagyott almező-azonosítóval. Példánkban tehát az *u* almezőkód hibajelzést fog adni, és a bemutatott hiba kivédhető, de a *c* almezőben, ha az almezőkód tévesen elmarad, a program nem ad hibajelzést, mivel a név *C* betűjét fogja almezőkódnak tekinteni. (Almezőkódként a kis- és nagybetű egyenértékű.)

A sajnálatos eredmény a kiírásakor:

```
Essential concepts of operating systems :  
using IBM mainframe examples / . K. Yuen
```

Ne feledjük, hogy az almezőkód az adat része, s ezért rögzítése is az adatbevitelt végző személy feladata.

Ilyen hibák kiküszöbölése céljából tehát a korrek-túraolvasónak is fokozottan kell figyelnie az almezők pontos jelölését.

Azt hiszem, hogy a viszonylag részletesen leírt fenti négy példa elegendő annak bemutatására, hogy milyen körültekintően kell megtervezni egy-egy hibatípus ellenőrzésének módszereit. Nemcsak a CDS/ISIS ellenőrző funkciót kell alaposan ismernünk, hanem megfelelő képünk kell, hogy legyen az előforduló hibák fajtáiról és súlyosságáról. A leggondosabb tervezésnél sem tekinthetünk el egy viszonylag hosszú, nyugodt próbaüzem hiba-kereső időszakától, az "éles" adatokkal tesztelt adatbázis és a belőle származó termékek gondos böngészésétől.

A képernyőn megjelenő egy-egy furcsa tétel, a "tudom, hogy megvan nekünk, mégsem találok az online katalógusban" stb. ismerős esetei mind a fentiek igazát bizonyítják.

A teljesség kedvéért még röviden a 3. ábrán szereplő, de a példákban nem elemzett hibák ellenőrzéséről. A hibás nyelvi kód kiszűrését sajnos csak valószínűsíteni tudjuk a "hun" nyelvi kód és a "Címfordítás" mező együttes előfordulásán alapuló hibajelzéssel. A döntő ellenőrzési eszköz itt is elsősorban a korrek-túraolvasás, akárcsak a címfordításban elkövetett ékezetes betű hiba ("operációs") és a "Megjelenési adatok"-ban hibásan írt kiadói név esetében.

Néhány szót érdemel még a megjelenési év ellenőrzése. Már említettem, hogy a gyakorlatban itt is csak az évszám adott tartományon belüli érvényességét szoktuk ellenőrizni, s ez csak a hibák egy része ellen nyújt védelmet. Az egyszerű szám-cserés elírás ellen viszont elég eredményesen védekezhetünk, ha tudjuk, hogy az adott időszakban beviendő dokumentumok megjelenési éve kellően szűk tartományba, mondjuk 1985 és 1987 közé esik. Az ellenőrző utasítás formája a Mezőleíró táblában a következő lesz:

```
IF 99.0 <'1985' OR 99.0 >'1987' THEN 'ERVENYTELEN
EVSZAM' FI
```

Persze ne felejtjük el a jövő év elején módosítani az évszámot a mezőleíró táblának ebben az utasításában.

A rekordok közötti hivatkozások mint az egységesítés eszközei

Úgy vélem, hogy felhasználói, könyvtárosi szempontból az ellenőrzés problémaköréhez tartozik az is, hogy az analitikusan feltárt részegységet tartalmazó kötet adatait helyesen és — természetesen — minden analitikus tétel esetében *egységesen*, azonos formában írjuk le. Hasonló igény a többkötetes művek közös adatainak és a sorozatba tartozó művek sorozati adatainak leírásával kapcsolatban is felvetődhet (noha a dokumentumleírási szabványok "tűkröztetési elvei" bizonyos esetekben ellentmondanak az egységesítésnek).

Nézzünk egy példát! A 4. ábrán látható kötet tizenkét előadás anyagát tartalmazza, tehát analitikus feltárás esetén tizenkét 'A' szintű rekord kerül adatbázisunkba. Ha minden egyes alkalommal leírnánk az előadások szövegét tartalmazó kötet monografikus szintű adatait is, a tizenkét eset közül óhatatlanul előfordul néhányban — talán a szabványok szempontjából még megengedhető — eltérés egy-egy adat írásmódjában.

E "változatosság" kiküszöbölésére és a felesleges, ismétlődő munka elkerülésére szolgál a CDS/ISIS-ben a rekordok közötti logikai kapcsolatok kezelése, mégpedig a rekord sorszáma szerinti "hivatkozás" formájában.

A hivatkozáshoz tehát az adott rekord valamelyik mezőjében szerepelnie kell egy *másik* (ún. "alternate") rekord sorszámának. Ez akár másik adatbázisban is lehet (ha például az analitikus tételek rekordjait külön adatbázisban tartjuk).

Példánkban az analitikus tételek hivatkoznak a monografikus tételre. Az 5. ábrán látható két előadás rekordjának és a hivatkozott könyvrekordnak a sematikus rajza. Az egyszerűség kedvéért csak a 24-es mezőt (cím és szerzőségi közlés) és a rekord sorszámát tartalmazó 39-es mezőt tüntetem fel, s ugyancsak az egyszerűség kedvéért a címet és a szerzőségi közlést megjelenítési formában írom le. Látható, hogy mind *Vanautgaerden*, mind *Husby* előadása a 081723-as számú CDS/ISIS rekordra hivatkozik, mivel ez a rekord tartalmazza a 4. ábrán látható kötet adatait.

Ez a hivatkozás háromféle módon használható:

- ◆ A rekordok megjelenítésekor a kimeneti formátumot vezérlő kódsor (print formatting language) hivatkozhat olyan adatokra, amelyek nem az aktuális, hanem a hivatkozott rekordban szerepelnek. Esetünkben például egy DOBIS/LIBIS-re vonatkozó keresés eredményeként megjelenítjük *Vanautgaerden* cikkének bibliográfiai adatait, s a 39-es mezőben található rekordsorszám alapján a *081723-as* rekordból írjuk ki az előadást tartalmazó *könyvfőcímét, megjelenési helyét* stb.

Gesamthochschulbibliothek Essen
Essen University Library

Hierarchical Relationships in Bibliographic Descriptions

INTERMARC
Software-Subgroup Seminar 4

Library Systems Seminar Essen
25 March - 27 March 1981

Contributors

E. Bonness · B. Delbrück · P. Goossens
O. Husby · B. Jedwabski · M. Lodder
H. Meulengracht-Madsen · P. Noerr
A. Regent · J. Vanautgaerden

Edited by

Ahmed H. Helal
Joachim W. Weiss

Gesamthochschulbibliothek Essen

Essen 1982

4. ábra Több részegységet tartalmazó kötet

- ◆ Hivatkozhatunk egy másik rekordra a keresőállomány bővítésének céljából is. Ezzel elérhetjük, hogy például az analitikus tétel adatait tartalmazó rekord nemcsak az előadás specifikus deskriptorai, hanem a könyv általánosabb deskriptorai (pl. "INTERMARC") szerint is kereshető legyen.
- ◆ A hivatkozás módszere igen alkalmas rendezett bibliográfiai listák készítésére. Ha például olyan listát akarunk készíteni, amelyben az analitikus tételek az őket tartalmazó kötetek címe szerint, ezen belül pedig a saját címük szerint rendezetten szerepelnek, az itt vázolt hivatkozások és a megfelelően elkészített rendezési mezőkiválasztó

táblák segítségével eredményesen megoldhatjuk feladatunkat.

Ennél többet e cikk keretében nem lehet és — talán — nem is kell elmondani a bibliográfiai rekordok közötti kapcsolatok, kölcsönös hivatkozások CDS/ISIS-lehetőségeinek izgalmas, ámde meglehetősen bonyolult problémaköréről. Ellenőrzési-egységesítési szempontból azonban fontos tudnunk, hogy a rekordok közötti hivatkozás eszköze is rendelkezésünkre áll a többszörös leírásból eredő hibalehetőségek számának csökkentésére.

Egy további probléma azért itt is jelentkezik. Bevitelkor adatként visszük be a hivatkozott rekord sorszámát, tehát adathiba itt is lehetséges. Ha például 081723 helyett 081273-at írunk, a hivatkozás tárgya egy egészen más rekord lesz. Ezért az adatként bevitt rekordsorszámot is ellenőrizni kell szótár, kódtábla vagy valamilyen felhasználó által írt programrutin segítségével.

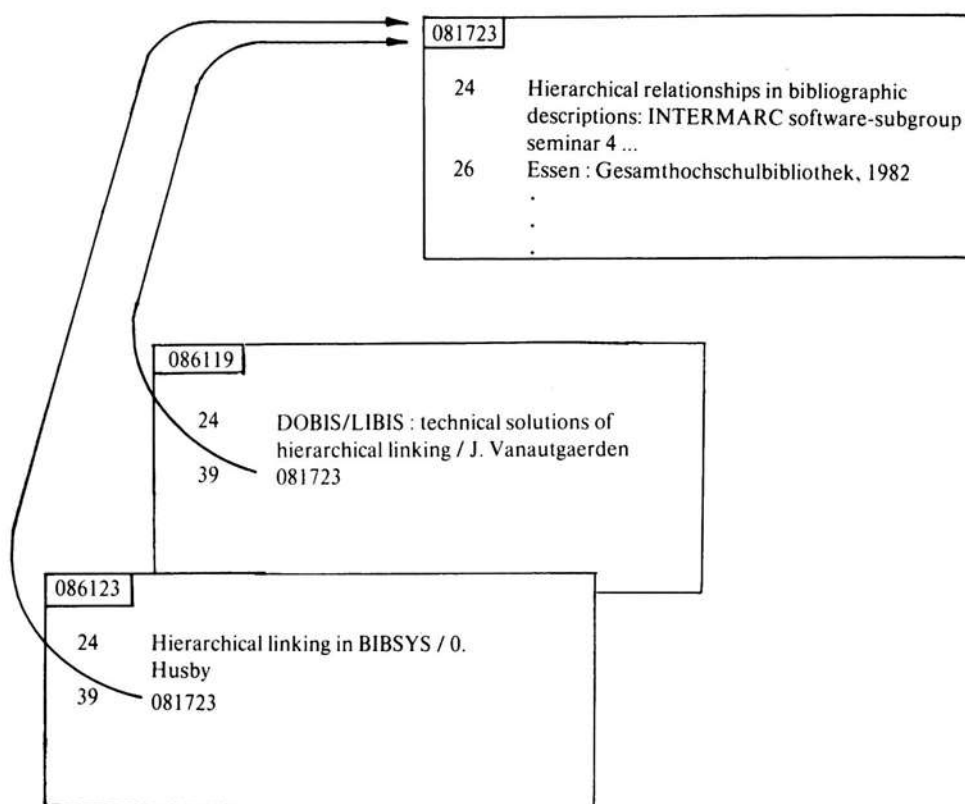
Ha a hivatkozás egy adatbázison belüli rekordokat kapcsol, és a hivatkozott rekordok száma viszonylag kevés az adatbázis összes rekordjához képest, valamint megfelelően szóródnak az adatbázisban, a szótár szerinti ellenőrzés viszonylag megbízható. Nagyon ritkán fordul elő, hogy a hivatkozott rekord sorszámát bevitel során egy másik, ugyancsak érvényes hivatkozás számával cserélik fel.

Utószó

Remélem, hogy a bibliográfiai adatbázisok építése során jelentkező ellenőrzési gondok felvetése és azok automatikus vagy félautomatikus megoldási lehetőségeinek felvázolása nemcsak a CDS/ISIS alkalmazása iránt érdeklődő olvasó számára jelentett esetleg megfontolandó szempontokat, hanem mindazoknak, akik érdeklődnek a nemnumerikus (szöveges) adatbázisok építésének problémái iránt.

Noha ilyen tevékenység nemcsak a CDS/ISIS eszközeivel történhet, fontosnak tartom hangsúlyozni, hogy a cikkemben leírt lehetőségek korántsem merítik ki a CDS/ISIS ellenőrzési lehetőségeit. Nem szóltam a mezők karakterkészletének szinte hagyományos ellenőrzéséről (numerikus, felhasználó által definiált stb.), a különböző mezőtípusok ellenőrzéséről (pl. a < > jelek ellenőrzéséről az ún. D-típusú deskriptormezőben), és a mezőhossz ellenőrzését sem említettem (pl. fix hosszúságú kódok esetében).

Mindezt az olvasó megtudhatja a rendszer referenciakönyvéből. A cikk megírásával elsődleges célom az volt, hogy azt a szemléletmódot próbálja erősíteni, amely az adatbázis-építés felelősségteljes munkájához véleményem szerint szükséges.



5. ábra Rekordok közötti kapcsolatok egy CDS/ISIS adatbázisban

(Hivatkozás a konferencia-előadásokat leíró analitikus rekordokban az előadást tartalmazó kötet monografikus rekordjára.)

Irodalom

- [1] VÁLAS Gy.: Adatbázis-építés: például az INSPEC. = Tudományos és Műszaki Tájékoztatás, 33. köt. 10. sz. 1986. p. 518–534.
- [2] BRYANT, Ph.: The catalogue. = Journal of Documentation, 36. köt. 2. sz. 1980. p. 133–163.
- [3] HUBA Z.: A CDS/ISIS szöveges információkezelő rendszer egy programozó szemével. = Tudományos és Műszaki Tájékoztatás, 31. köt. 1–2. sz. 1984. p. 14–18.
- [4] SZÜCS A.–VARGA S.: Online könyvkatalógus a MICRO-ISIS segítségével. Alkalmazási példái = Tudományos és Műszaki Tájékoztatás, 33. köt. 5–7. sz. 1986. p. 268–280.
- [5] FLORENCEV, S. N.–OŽOGIN, B. A.–VASIL'EV, S. M.: A CDS/ISIS/ESZ alkalmazói programcsomag és alkalmazási lehetőségei. = Tudományos és Műszaki Tájékoztatás, 31. köt. 10. sz. 1984. p. 402–409.
- [6] CDS/ISIS Reference Manual: Rev. 9, Release 4.5, November 1984 / Unesco Computerized Documentation System (CDS); Division of the Unesco Library, Archives and Documentation Services (LAD). – [Paris] : Unesco, cop. 1984. – 223 p.

VARGA Sándor: *Bibliográfiai adatok ellenőrzése és egységesítése a CDS/ISIS nagygépes változatával*

A bibliográfiai rekordok készítése során többféle hibát lehet elkövetni. Egyes hibák csak a szöveg olvashatóságát csökkentik, mások viszont a leírt bibliográfiai tétel eredményes kikeresését is akadályozhatják. Az adatbeviteli, leírási hibák egy része csak figyelmes emberi munkával, korrektúraol-

VARGA, S.: *Checking and standardization of bibliographic data with the mainframe version of CDS/ISIS*

During the production of bibliographic records, several types of errors can be committed. Some of them cause only the readability of a text to be difficult but others can even prevent the retrieval of a bibliographic record. A part of the recording and input errors can only be eliminated by human efforts

vasással szűrhető ki, más részük viszont programmal automatikusan is felderíthető. Az ellenőrzés e típusait mutatjuk be a CDS/ISIS programcsomag ellenőrzési lehetőségeivel szemléltetve azokat.

* * *

ВАРГА, Ш.: Проверка и унификация библиографических данных с использованием ППП CDS/ISIS на большой ЭВМ

При подготовке библиографических записей можно допустить различные ошибки. Некоторые ошибки только уменьшают читаемость текста, другие же могут препятствовать успешному поиску библиографической записи. Часть ошибок ввода или описания может быть найдена только при корректуре, т. е. при внимательном чтении. Другая же часть может быть определена автоматически, при помощи программы. Статья описывает типы проверки текста, демонстрируя примерами применения ППП CDS/ISIS на большой ЭВМ.

like proof reading, but others can be detected and/or corrected automatically using suitable computer programs. The author discusses this latter type of error checking illustrated by the checking routines of the CDS/ISIS program system.

* * *

VARGA, S.: Kontrolle und Vereinheitlichung der bibliographischen Daten mit der CDS/ISIS Grossmaschine

Man kann während der Fertigung bibliographischer Rekorde zahlreiche Fehler begehen. Manche Fehler vermindern nur die Lesbarkeit des Textes, die anderen können aber auch die erfolgreiche Ausschichtung des abgeschriebenen bibliographischen Satzes verhindern. Ein Teil der Fehler der Dateneingabe und -abschreibung kann nur mit aufmerksamer menschlicher Arbeit mit Korrekturlesen ausgesondert werden, ein anderer Teil ist aber mit dem Programm automatisch aufdeckbar. Der Artikel stellt die Typen dieser Kontrolle mit der Möglichkeiten der CDS/ISIS Programmpackung dar.

A SZELLEMI MUNKÁT IS TANULNI KELL

Egy szakma, mesterség vagy akár csak valamely köznapi tevékenység gyakorlására hosszú évek fáradságos munkájával, de legalábbis néhány kisebb-nagyobb fogás elsajátításával készülünk fel – ez természetes. De ha meggondoljuk, mennyire kézenfekvőnek tartjuk, hogy a termelésben vagy más gyakorlati tevékenység során meg kell tanulni és be kell gyakorolni az egyes műveletek valamennyi fortélyát, önkéntelenül felvetődik a kérdés, miért nem tartjuk ugyanilyen nyilvánvalónak, hogy ezt tegyük a szellemi munka terén is, vagyis megtanuljunk – megtanítsunk – tanulni, hatékonyan olvasni, figyelmesen hallgatni, információkat, gondolatokat rendszerezni, vagy ezeket világosan, érthetően közreadni. Pedig ahogy még a legegyszerűbb eszközök használatának is megvannak a maga szabályai, fogásai, ugyanígy akadnak a szellemi munkát, az ismeretek kezelését megkönnyítő szabályok, módszerek is.

Ezek ismertetésére, illetve az egyes módszerek felhasználásának, alkalmazásának előmozdítása érdekében íródott *dr. Dezső Zsigmondné: A szellemi munka technikája* (Ismeretszerzés/gyorsolvasás/ -rögzítés, -közlés) c. könyve, amely értékes segítséget nyújthat mindenkinek, aki szellemi munkával foglalkozik.

A könyv – korlátozott példányszámban – még kapható az OMIKK értékesítési osztályán (Bp., Múzeum u. 17. félemelet 53.).

Ára: 120,- Ft.