

Hozzászólás a
Magyar Könyvtárosok Egyesületének 40. Vándorgyűlésén 2008. július 25-én tartott
MOKKA-ODR nyílt fórumon

Thék György (InfoKer Számítástechnika-alkalmazási Szövetkezet)

Bár a felhatalmazásuk nélkül, főként a TextLibet használó intézmények képviselőit figyelembe véve állítottam össze hozzászólásom anyagát.

A MOKKA a textlibes könyvtárak számára egyelőre csak rekordforrás, mivel a közös adatbázis építéséhez nem járulhatnak hozzá. A rekordok letöltésének célja a feldolgozó munka egyszerűsítése, haszna pedig akkor van, ha a MOKKA-ból nyerhető rekordok minősége e célnak megfelel. Hozzászólásommal azt próbálom igazolni, hogy a MOKKA rekordjai jelenleg rengeteg hibát tartalmaznak, így a katalogizáló munka egyszerűsítésére, esetleg helyettesítésére csak szigorú ellenőrzés és javítás után alkalmasak.

A TextLib adatbázis szerkezete nem Hunmarc alapú. Mivel a KSZ/4.1 címe és bevezetőjének első mondata szerint is a Hunmarc a bibliográfiai rekordok adatsere formátuma, a KSZ/4.1-ben leírt szabályok alkalmazásával az eltérő szerkezetek közötti konverzió megoldható. A megvalósítás során szükség van a Hunmarc rekordok alapos elemzésére, melynek eredményeként kiderül a bibliográfiai szint, a szerzőség típusa, a gyűjtemény típusa, stb. Az elemzés 50-nél több szempontot vizsgál, kiterjed a rekordfejre, a mutatóra, a technikai és a bibliográfiai adatmezőkre, ismételhetőségre, kötelezőségekre, stb.

A következő anyag összeállításának elindítója a MOKKA adatbázisából a mindennapi munkához szükséges rekordok letöltésekor talált hibák magas száma volt. Egy jelentős számú rekordból álló minta részletes elemzése után mód van a MOKKA minőségének általános megítélésére.

Az elemzéshez 516 119 MOKKA rekordot használtunk fel. Ezekben a sokféle hiba összes előfordulásának száma több millió, 10 olyan hiba van, amelyek egyenként 100 000-nél többször szerepelnek.

Következzen egy táblázat a legjellemzőbbekből:

A hiba jellege	száma
a KSZ/4.1 szerint ismeretlen indikátorok száma	1 140 421
egyéb indikátorhiba (főtétel, melléktétel, részletező melléktétel tévesztés)	498 061
ismeretlen mezőhívójel	303 744
ismeretlen almező	117 980
mezőhívójel ismételhetőségének megsértése	1047
almező ismételhetőségének megsértése	5396
megengedhetetlen jel a technikai (005, 007, 008) mezőkben	235 062
meglévő mezők összefüggései alapján kötelező mező (245) hiánya	26 277
többkötetesek rekordjai közötti kapcsolati (787) hiba	1448
további egységesített cím (730) további cím (740) nélkül	8688
kapcsolatjelölő hiánya, hibás értéke	11 334
ismeretlen nyelvkód (41)	1144

A hibák súlyossága természetesen nem azonos. A konverzió végrehajtása szempontjából súlyosnak ítélt hibát tartalmazó rekordok aránya a teljes mintában 17,36%, a mintából kiválasztott 2007-es és 2008-as megjelenési évű tételekben pedig 11,96%, ez javuló tendenciát mutat.

A hibák létéért három tényező felelős:

- a könyvtári rendszer, ami megengedi a hibás rekord elkészítését
- a könyvtáros, aki elkészíti a hibás rekordot
- a MOKKA, ami befogadja a hibás rekordot

A helyzet javítása érdekében a Hunmarc rekordok elemzésére és ellenőrzésére képes programunk kétféle változatát ajánlom fel. Az első egy önálló program, amely tetszőleges méretű, Hunmarc rekordokat tartalmazó (MSZ ISO 2709) állományt listáz, a rekordok végéhez szöveges megjegyzésekben hozzáfűzve az elemzés eredményét. A második a <http://www.textlib.hu/hmchk.htm> cím, ahol on-line módon lehet Hunmarc rekordokat ellenőrizni az előzőhöz hasonló végeredménnyel.

Végül két megjegyzés a fórumon korábban elhangzottakhoz:

- Nem értek egyet Koltay Klára egy rövid mondatával, mely szerint a lelőhely adatok számának növelése érdekében kellett lazítani a bibliográfiai rekordok MOKKA-ba kerülését engedélyező szabályokon. Véleményem szerint nem szabad hibás rekordokat beengedni a MOKKA-ba.
- Támogatom Bariczné Rózsa Mária ötletét, hogy a Kelló teljes bibliográfia rekordokat küldhessen a MOKKA-ba.