

Da Unimarc a MARC21

Zeno Tajoli

CILEA, Segrate

Abstract

La conversione dei dati dell'Università di Brescia, dal formato Unimarc a MARC21. Vengono evidenziati le soluzioni software usate e i punti critici della conversione.

Keywords: Unimarc, MARC21, conversione dati bibliografici.

Come annunciato in un precedente articolo [6] il server Z39.50 dell'Università di Brescia, implementato dal CILEA, contiene dati in formato MARC21 [2]. Poiché i dati in origine sono in formato Unimarc [1] e provengono da uno scarico dell'applicativo SBN Unix C/S, è necessario convertirli. La conversione è stata effettuata partendo dall'analisi compiuta dalla Library of Congress e disponibile in rete [4]. Vengono qui illustrati i passaggi e punti ritenuti più interessanti. Chi fosse interessato al codice, può fare richiesta al responsabile della sezione, ing. Fabio Valenziano.

I set di caratteri

La prima considerazione da fare riguarda i set di caratteri usati nello scarico Unimarc da SBN Unix C/S. Si tratta di ISO 646, di ISO 5426-1983 e di ISO /DIS 6937/2. In particolare si sottolinea che ISO 5426-1983 viene usato nell'ambiente a 7-bit, cosa che agli inizi può creare qualche problema. Invece non è stato semplice documentarsi sullo standard ISO /DIS 6937/2. Questo standard attualmente è stato sostituito nell'uso comune dalla serie ISO 8859-x e da ISO 10646, pertanto la documentazione al riguardo è scarsa.

MARC21 accetta due sistemi di codifica.

Il primo si chiama MARC-8, è un sistema ad 8-bit, ed è stato definito internamente dalla Library of Congress seguendo le esigenze delle biblioteche americane. Il secondo è il sistema ad ambiente variabile 8 / 16 bit che segue lo standard ISO/IEC 10646 (UCS) e Unicode UTF-8 come regole di codifica. Nella documentazione su MARC21 questo standard viene chiamato 'UCS/Unicode UTF-8'.

Attualmente i software che lavorano sul protocollo Z39.50 gestiscono al meglio il sistema MARC-8, pertanto i caratteri sono stati traslitterati in questo sistema. La traslitterazione avviene attraverso una specifica tabella di conversione gestita dall'applicativo USMARCON ver. 1.44 [7]. Si apprezza il fatto che l'applicativo sia usabile sia su piattaforma Win 32 che Linux.

La conversione tra i due formati

La conversione avviene usando un apposito programma scritto in Basic, l'interprete usato è ScriptBasic ver. 2.0 [5]. Esso ha molte funzioni utili per questo tipo di operazioni, è distribuito con il sorgente in licenza GNU LGPL, funziona in ambiente Win32, Linux e Unix generico. Vengono ora illustrati alcuni punti di particolare interesse. I campi che vengono creati sono campi MARC21.

Va premesso che i vari MARC21 campi contengono punteggiatura interna delle aree ISBD e la punteggiatura delle intestazioni. Unimarc invece non contiene punteggiatura con significato catalografico all'interno dei dati, tranne rari casi. Questa differenza va tenuta ben presente quando si opera una simile conversione.

Indicatore 2 dei campi titolo: in MARC21 i diversi campi che contengono dati collegati al titolo hanno l'indicatore 2 per indicare quanti caratteri non prendere in considerazione nell'ordinare il titolo. Unimarc invece prevede dei caratteri specifici all'interno del campo per indicare le sezioni da non considerare nell'ordinamento. Se questa opzione fosse usata

in maniera estesa non sarebbe riproducibile in MARC21. L'indicatore infatti è composto da un solo carattere e può contenere solo cifre da 0 a 9. Tuttavia per ora non si sono trovati casi problematici sotto quest'aspetto. La Library of Congress sta lavorando per usare la metodologia di Unimarc anche in MARC21.

Creazione campo 041: viene creato leggendo Unimarc 101, se è assente come valore di default viene messo 'ita'

Creazione campo 044: viene creato leggendo Unimarc 102, se è assente come valore di default viene messo 'it'

Creazione campi 110, 111, 710, 711: nella creazione dei campi per gli autori enti temporanei e enti permanenti è visibile uno dei limiti dello scarico Unimarc di SBN Unix C/S : vengono creati solo i sottocampi ^a, ^b, ^c, ^h, ^z, ^4. Mancando gli altri, la conversione non può essere precisa.

Creazione campo 500: questo campo riceve dati da vari campi, alcuni problemi si sono avuti con i dati provenienti dai campi Unimarc 311 e 312. Lo scarico proveniente da SBN Unix C/S inserisce in questi campi dei dati non perfettamente corrispondenti a quanto detto dallo standard Unimarc.

Gestione dei dati di copia: nello scarico Unimarc di SBN Unix C/S i dati di copia vengono inseriti nel tag 950. MARC21 ha un formato apposta per le copie ('Holdings'). Si è preferito non trattare quest'aspetto ed è stata inserita solo l'indicazione di quali biblioteche possiedono una copia dell'item in questione.

Creazione del record in formato ISO 2790

Alla fine delle operazioni di conversione i dati sono codificati con il set MARC-8 e sono distribuiti nei campi e nei sottocampi MARC21. Tuttavia la loro rappresentazione fisica nel file è nel formato MicroLIF, molto più facile da gestire che il formato ISO 2709 quando si vogliono scrivere record bibliografici in un file attraverso un programma batch. La trascodifica finale da MicroLIF a ISO2709 avviene tramite uno script perl che usa la libreria MARC/Perl [3].

Bibliografia

- [1] IFLA., "Unimarc Manual: bibliographic format", Update 4 (2002), Saur
- [2] Library of Congress, "MARC 21 Concise Format for Bibliographic Data", URL: <http://www.loc.gov/marc/bibliographic/ecbdhome.html>
- [3] MARC/Perl, URL: <http://marcpm.sourceforge.net/>
- [4] Network Development and MARC Standards Office - Library of Congress, "UNIMARC to MARC21 Conversion Specifications", URL: <http://www.loc.gov/marc/unimarc21.html>
- [5] ScriptBasic, URL: <http://scriptbasic.com/html/index.html>
- [6] Tajoli Zeno, "Un server z39.50 per l'università di Brescia", Bollettino del Cilea n. 93, 2004, pp. 32-35.
- [7] USMARCON, URL: <http://www.bl.uk/services/bibliographic/usemarcon.html>