

ΥΠΟΣΤΗΡΙΞΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΧΑΡΑΚΤΗΡΩΝ ΑΠΟ ΣΥΣΤΗΜΑΤΑ ΒΙΒΛΙΟΘΗΚΩΝ ΚΑΙ ΤΑ ΔΙΕΘΝΗ ΒΙΒΛΙΟΓΡΑΦΙΚΑ ΠΡΟΤΥΠΑ

Δ. Γαροφαλλίδης,
Πληροφορική Εργασίας Α.Ε.Β.Ε

ΠΕΡΙΛΗΨΗ

Η εγκατάσταση συστημάτων αυτοματοποίησης στις ελληνικές ακαδημαϊκές βιβλιοθήκες ήταν τα τελευταία χρόνια εντυπωσιακή, αφού το σύνολο τους είτε αναβάθμισε, είτε αντικατέστησε, είτε εγκατέστησε για πρώτη φορά ειδικό λογισμικό. Οι δυνατότητες για προσφορά υπηρεσιών από τις ακαδημαϊκές βιβλιοθήκες είναι πλέον ουσιαστικές και το γεγονός αυτό είναι φυσικό να οδηγήσει στην αναβάθμιση του ρόλου των μέσα στα πλαίσια της εκπαιδευτικής διαδικασίας των ιδρυμάτων. Το επόμενο τεχνολογικό άλμα που πρέπει να επιτύχουν πλέον οι βιβλιοθήκες είναι η διασύνδεση των συστημάτων αυτών. Πόσο απλό είναι το εγχείρημα αυτό;

Επιχειρείται η επεξήγηση του θέματος της υποστήριξης των χαρακτήρων γενικότερα σε αυτοματοποιημένα συστήματα βιβλιοθηκών με περιεχόμενη πληροφορία στην Ελληνική γλώσσα, και αναλύονται οι δύο συσχετιζόμενες και ταυτόχρονες πορείες εξέλιξης: των «κωδικοσελίδων» των λειτουργικών συστημάτων ηλεκτρονικών υπολογιστών και οι προδιαγραφές που θέτουν τα βιβλιογραφικά πρότυπα UNIMARC και USMARC.

Η διπλή φύση της παρουσίασης φιλοδοξεί να οριοθετήσει το πρόβλημα της διασύνδεσης των καταλόγων βιβλιοθηκών, αλλά και να παρέχει όλο το θεωρητικό υπόβαθρο για την υποστήριξη καταλογογράφησης υλικού σε πολλαπλές γλώσσες. Η λεπτομερής επεξήγηση του τρόπου με τον οποίο τα δύο βιβλιογραφικά πρότυπα στα οποία γίνεται αναφορά προδιαγράφουν και υλοποιούν την υποστήριξη πολλαπλών αλφάβητων, θα παρέχει την αναγκαία γνώση για τη δημιουργία καταλόγων με ορθή πληροφορία σε κυριλλικό, εβραϊκό, αραβικό, αφρικανικό, κλπ. αλφάβητα.

Τα βιβλιογραφικά πρότυπα έχουν καλύψει με πληρότητα και σαφήνεια τα θέματα της υποστήριξης χαρακτήρων πολλών αλφάβητων σε μια βιβλιογραφική εγγραφή, και έχουν θέσει το γενικότερο πλαίσιο με το οποίο μπορεί να εισαχθεί ένα νέο αλφάβητο σε αυτά. Η προσεκτικότερη παρατήρηση των εξελίξεων που έχουν σημειωθεί αποκαλύπτει την προσπάθεια τους να καλύψουν το κενό που υπάρχει μέχρι σήμερα στα λειτουργικά συστήματα υπολογιστών. Το μέλλον μας επιφυλάσσει την αναστροφή του προβλήματος, αφού προωθείται στο χώρο της πληροφορικής μια λύση του προβλήματος, η υποστήριξη του Unicode.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

συστήματα αυτοματοποίησης βιβλιοθηκών, Σύνολα χαρακτήρων, Δικτύωση υπολογιστών, επικοινωνία υπολογιστών, MARC standards, codepages, UNICODE

CHARACTER SUPPORT IN LIBRARY AUTOMATION SYSTEMS AND THE INTERNATIONAL BIBLIOGRAPHIC STANDARDS

D. Garofallidis,
Ergodata S.A.

ABSTRACT

The installation of automation systems in the academic libraries was during the course of the last few years impressive, since all of them either upgraded, replaced or installed for the first time special software. Having now an enormous potential to offer new services, it is expected to improve their role in the educational process of the Institutes. The next logical step will be the development of a network among these systems. How simple this process shall be?

It is expected that the issue of the character support in library catalogs in its general form will be clarified, with special focus on catalogs with Greek data. Two parallel directions are analysed, the "codepages" of the computer operating systems and the character sets definition in the UNIMARC and the USMARC standards.

The dual nature of this presentation aims to set the boundaries in the problem of the interconnection among automated library systems and to provide all the theoretical background for the creation of catalogue records with information in many languages. Through a detailed review of the way the two MARC standards in question implement and handle characters in many languages, the theoretical and practical background will be offered for the generation of records with correct support of ancient and modern Greek, Cyrillic, Hebrew, Arabic, African, etc. alphabets.

All bibliographic standards address with consistency and clarity the issue of the character support of multiple alphabets within a single bibliographic record, and have defined a general frame for the adoption of any new alphabet. A closer examination of the developments in this area reveals the effort to cover the existing gap in the information science, regarding the computer operating systems. In the future we are going to witness a reversion of the situation, as a satisfactory solution becomes widely adopted in IT, namely the Unicode support.

KEYWORDS

library automation systems, Character sets, Computer networks, Computer communication, MARC standards, codepages, UNICODE

ΕΙΣΑΓΩΓΗ

Η καταγραφή του υλικού μιας βιβλιοθήκης ή ενός κέντρου τεκμηρίωσης, μιας υπηρεσίας παροχής πληροφοριών γενικότερα, περνάει από τη δημιουργία και τη συντήρηση ενός βιβλιογραφικού καταλόγου στον οποίο μπορούν να ανατρέξουν οι ερευνητές. Η καταγραφή του υλικού σε αυτόν τον κατάλογο θα πρέπει να ακολουθεί μια μεθοδολογία τέτοια που όλοι όσοι τη γνωρίζουν να μπορούν να εντοπίσουν την πληροφορία που αναζητούν. Η προσπάθεια της απλούστευσης της έρευνας όμως δεν θα πρέπει να επηρεάζει τα βασικά χαρακτηριστικά του υλικού, όπως είναι ένας τίτλος. Όσο απλή και αν είναι η περιγραφή του προβλήματος αυτού υπάρχουν τεχνικές δυσκολίες στην επίτευξη του στόχου οι οποίες ξεκινούν από τη δομή της ανθρώπινης κοινωνίας, με μεγαλύτερη αυτήν των διαφορετικών γλωσσών και αλφάβητων.

Από πολύ νωρίς οι βιβλιοθήκες ανέπτυξαν συστήματα καταγραφής και οργάνωσης του υλικού τους. Τρία είναι τα σημεία στα οποία μπορεί να σταθεί κανείς: η φυσική μορφή του καταλόγου, το σύστημα ταξινόμησης και τα χρησιμοποιούμενα αλφάβητα. Οι επιλογές για το τελευταίο σημείο δεν είναι πολλές, η καταγραφή μπορεί να ακολουθήσει το πρωτότυπο αλφάβητο, ή να χρησιμοποιηθεί μετάφραση χαρακτήρων (transliteration). Η τεχνική της μετάφρασης των χαρακτήρων για να δώσει αποτελέσματα προϋποθέτει την ηχητική αναγνώριση των όρων, αλλά σε αυτές τις περιπτώσεις είναι σύνηθες ο ερευνητής να γνωρίζει και τη γραφή του αλφάβητου στο οποίο είναι γραμμένο το υλικό που ερευνά. Η καταγραφή του υλικού σε χειρόγραφες, ή ακόμα και έντυπες, καρτέλες επέτρεπε την αποτύπωση όλων των απαραίτητων χαρακτήρων. Τα συστήματα αυτά εμφανίζουν ένα σημαντικό μειονέκτημα: τον αυξημένο χρόνο αναζήτησης για την εύρεση μιας πληροφορίας. Η πρόσφατη ανάπτυξη και διάδοση των ηλεκτρονικών υπολογιστών επέτρεψε την συγκέντρωση των πληροφοριών αυτών σε ψηφιακή μορφή, και τη χρήση των τεχνολογιών βάσεων δεδομένων που προσφέρουν ταχύτατη ανεύρεση οργανωμένων δομών πληροφορίας. Τα βιβλιογραφικά πρότυπα προσαρμόστηκαν στα δεδομένα αυτά, και σύντομα εμφανίστηκαν ειδικευμένες μορφές που μπορούν να υλοποιηθούν σε ψηφιακά συστήματα. Οι υπηρεσίες που μπορεί να προσφέρει ένας υπολογιστής αποδείχθηκαν πραγματικά πολύτιμες για τους σύγχρονους καταλόγους και μόνο μια μικρή λεπτομέρεια αμαυρώνει την εικόνα του τέλειου: η μη υποστήριξη των χαρακτήρων όλων των αλφάβητων.

Με εφευρετικό τρόπο τα βιβλιογραφικά πρότυπα έδωσαν λύση στο πρόβλημα αυτό και έτσι σήμερα βρισκόμαστε στην ευχάριστη θέση να διαθέτουμε οργανωμένους ψηφιακούς καταλόγους σχεδόν σε κάθε βιβλιοθήκη. Το επόμενο βήμα είναι η δικτύωση τους, ώστε από ένα και μόνο σημείο, το οποίο δεν είναι απαραίτητο να είναι ούτε καν συγκεκριμένο, να μπορεί ο ερευνητής να εντοπίζει την πληροφορία που αναζητά. Το βήμα αυτό είναι σχεδόν αυτονόητο, είναι όμως και εύκολο;

ΥΠΟΣΤΗΡΙΞΗ ΧΑΡΑΚΤΗΡΩΝ ΣΕ ΣΥΣΤΗΜΑΤΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Οι ηλεκτρονικοί υπολογιστές σχεδιάστηκαν και αναπτύχθηκαν με κύριο σκοπό την υποβοήθηση της εκτέλεσης σύνθετων αριθμητικών υπολογισμών σε μικρό χρόνο. Το θέμα της υποστήριξης χαρακτήρων διαφορετικών αλφάβητων δεν ήταν μέσα στις αρχικές προτεραιότητες που είχαν τεθεί. Η βασική υποστήριξη των χαρακτήρων του πρώτου αλφάβητου ήρθε για να καλύψει την ανάγκη απλούστευσης της διεπαφής μεταξύ προγραμματιστών και μηχανής, και κάλυπτε τα βασικά γράμματα της αγγλικής γλώσσας.

Είναι γνωστό ότι ένα ψηφιακό υπολογιστικό σύστημα μπορεί να αναγνωρίσει και να επεξεργαστεί πληροφορία μόνο αν αυτή μπορεί να αναλυθεί σε ψηφιακή μορφή, τις γνωστές σειρές μηδενικών (0) και άσπων (1). Αυτά ονομάζονται bits. Μια ομάδα από επτά τέτοια ψηφία στη σειρά σχηματίζουν μια «λέξη», ή byte. Για να επιταχυνθεί η μεταφορά πληροφορίας μέσα σε ένα υπολογιστή, η κίνηση γίνεται με τέτοιες «λέξεις», τα bytes. Το πλήθος των δυνατών διαφορετικών συνδυασμών που μπορούν να δημιουργηθούν είναι 128 (2⁷). Η συγκεκριμένη αρχιτεκτονική οδήγησε στη δημιουργία ενός πίνακα χαρακτήρων με ισάριθμες θέσεις, όπου σε κάθε μια από αυτές τοποθετήθηκε ένα συγκεκριμένο σύμβολο της αλφαβήτου. Ο χρήστης αναγνωρίζει τη μορφή του συμβόλου, το γράμμα δηλαδή, ενώ ο υπολογιστής αναγνωρίζει την αντίστοιχη θέση του πίνακα χωρίς να έχει τη δυνατότητα φυσικά να συνδέει τα γράμματα ή τις λέξεις με κάποιο νόημα, όπως συμβαίνει στον άνθρωπο. Το σύνολο των χαρακτήρων αυτό είναι πολύ γνωστό με το όνομα ASCII, και τα συστήματα αυτά είναι γνωστά ως «επτάμπιτα».

Πολύ σύντομα εμφανίστηκε η ανάγκη για αύξηση των υποστηριζόμενων συμβόλων, αφού χαρακτήρες που χρησιμοποιούνται στην Γερμανική, Γαλλική και Ισπανική γλώσσα δεν υπάρχουν στο βασικό σύνολο. Το πρόβλημα λύθηκε με την αλλαγή της αρχιτεκτονικής των υπολογιστών, αυξάνοντας κατά ένα ψηφίο το byte ώστε να διαθέτει 8 θέσεις. Τα συστήματα αυτά πήραν την ελληνική ονομασία οκτάμπιτα (8-bit) και πλέον το μέγεθος του πίνακα των χαρακτήρων είναι 256 (2⁸). Τα συστήματα που γνωρίζουμε και χρησιμοποιούμε σήμερα όλοι ανήκουν σε αυτήν την αρχιτεκτονική, η οποία επιτρέπει την τοποθέτηση πρόσθετων χαρακτήρων μετά την 128η θέση. Αρχικά τοποθετήθηκαν εδώ οι χαρακτήρες της λατινικής γλώσσας που δεν εμφανιζόταν στο σύνολο ASCII, και το υποσύνολο αυτό πήρε την ονομασία ANSEL (extended latin), ενώ σε συνδυασμό με το ASCII ονομάστηκε ANSI, και φαίνεται στο Σχήμα 1.

HexCode	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0				0	@	P	.	p			.	À	Ð	à	ð	
1			!	1	A	Q	,	q			á	Ñ	á	ñ		
2			"	2	B	R	-	r			â	Ò	â	ò		
3			#	3	C	S	_	s			ã	Ó	ã	ó		
4			\$	4	D	T	`	t			ä	Ô	ä	ô		
5			%	5	E	U	~	u			å	Õ	å	õ		
6			&	6	F	V		v			æ	Ö	æ	ö		
7			'	7	G	W		w			ç	×	ç	×		
8			(8	H	X		x			è	Ø	è	ø		
9)	9	I	Y		y			é	Ù	é	ù		
A			*		J	Z		z			ê	Ú	ê	ú		
B			+		K	[k			ë	Û	ë	û		
C			,		L	\		l			ì	Ü	ì	ü		
D			-		M]		m			í	Ý	í	ý		
E			.		N	^		n			î	ÿ	î	ÿ		
F			/		O	_		o			ï	ÿ	ï	ÿ		

Σχήμα 1: Πίνακας χαρακτήρων ANSI

Το πρότυπο ANSI χρησιμοποιήθηκε με διάφορες παραλλαγές από όλους τους κατασκευαστές υπολογιστικών συστημάτων. Η δημιουργία των παραλλαγών οφείλεται σε επιμέρους συμφέροντα των εταιρειών παραγωγής λειτουργικών συστημάτων, απαραίτητο συστατικό για τη λειτουργία ενός υπολογιστή γενικής χρήσης, και οδήγησε σε μια γενικότερη αναστάτωση στο θέμα της συνεργασίας μεταξύ μηχανών διαφορετικών κατασκευαστών, που δυστυχώς συνεχίζεται ακόμη και σήμερα. Η αύξηση των διαθέσιμων θέσεων που πίνακα των χαρακτήρων είχε όμως μια πολύ θετική συνέπεια για τον ελληνικό χώρο, τη δυνατότητα να υποστηριχθεί ταυτόχρονα από ένα σύστημα το βασικό λατινικό αλφάβητο και οι νεο-ελληνικοί χαρακτήρες. Για να γίνει αυτό μεταβλήθηκαν τα περιεχόμενα του δεύτερου μισού του πίνακα των χαρακτήρων, το ANSEL δηλαδή, με τα ελληνικά σύμβολα. Η διάταξη των χαρακτήρων αυτών μάλιστα, υιοθετήθηκε και προτυποποιήθηκε από τον ΕΛΟΤ, δημιουργώντας το γνωστό ΕΛΟΤ-928. Στο πρότυπο αυτό έχουν καταλήξει πλέον και όλοι οι κατασκευαστές υπολογιστών και λειτουργικών συστημάτων που δραστηριοποιούνται στον ελληνικό χώρο. Παίρνοντας το παράδειγμα της πιο γνωστής, στο Σχήμα 2 προβάλλεται η διάταξη των χαρακτήρων της υλοποίησης της Microsoft™, η οποία έχει τον χαρακτηριστικό κωδικό "MS windows codepage 1253".

	20	30	40	50	60	70	80	90	A0	B0	C0	D0	E0	F0
0	0	@	P	ρ	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
1	1	A	Q	a	q	ι	ι	ι	ι	ι	ι	ι	ι	ι
2	2	B	R	b	r	ι	ι	ι	ι	ι	ι	ι	ι	ι
#	3	C	S	c	s	f	ι	ι	ι	ι	ι	ι	ι	ι
\$	4	D	T	d	t	ι	ι	ι	ι	ι	ι	ι	ι	ι
%	5	E	U	e	u	ι	ι	ι	ι	ι	ι	ι	ι	ι
&	6	F	V	f	v	ι	ι	ι	ι	ι	ι	ι	ι	ι
7	G	W	g	w	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
(8	H	X	h	x	ι	ι	ι	ι	ι	ι	ι	ι	ι
)	9	I	Y	i	y	ι	ι	ι	ι	ι	ι	ι	ι	ι
*	J	Z	j	z	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
+	K	[k	[ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
<	L	\	l	\	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
=	M] m] m	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
>	N	^	n	^	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι
/	O	_	o	_	ι	ι	ι	ι	ι	ι	ι	ι	ι	ι

Σχήμα 2: Πίνακας χαρακτήρων Microsoft Windows™ codepage 1253

Οι πίνακες των χαρακτήρων που περιγράφηκαν στις προηγούμενες παραγράφους έχουν ένα κοινό χαρακτηριστικό: δεν μπορούν να καλύψουν ούτε όλα αλφάβητα με λατινικούς χαρακτήρες, ούτε το πλήρες ελληνικό πολυτονικό αλφάβητο. Στην τελευταία περίπτωση δε, δεν υπάρχει καθόλου υποστήριξη των συνδυασμών λατινικών γραμμάτων με τονικά σύμβολα. Κι όμως, κάποια συστήματα προβάλλουν και εκτυπώνουν τόσο λατινικούς χαρακτήρες με τονικό σύμβολο, όσο και πολυτονισμένα ελληνικά γράμματα. Για να γίνει αυτό παρεμβάλλουν δύο πρόσθετα στοιχεία στο πρόβλημα: τις γραμματοσειρές (fonts) και τις οδηγίες ελέγχου (escape sequences). Οι γραμματοσειρές δεν είναι τίποτε διαφορετικό από πίνακες γραμμάτων όπου κάθε γράφημα χαρακτήρα έχει ιδιαίτερα χαρακτηριστικά. Είναι γνωστές σε όλους όσους έχουν χρησιμοποιήσει κάποιο πρόγραμμα συγγραφής κειμένων, αφού η επιλογή της μορφής των χαρακτήρων αποτελεί βασική έννοια σε αυτά. Πρακτικά το πρόγραμμα που χρησιμοποιεί fonts παρεμβάλλει μια οδηγία στο κείμενο με την οποία αντικαθίσταται προσωρινά ο βασικός πίνακας των χαρακτήρων με κάποιον άλλο. Η τεχνική αυτή έχει ένα σοβαρό μειονέκτημα, και για το λόγο αυτό δεν χρησιμοποιείται σε επίπεδο λειτουργίας υπολογιστών, την ανομοιογένεια που παρουσιάζουν μεταξύ τους τόσο οι

μηχανές αλλά και τα λειτουργικά συστήματα και επίσης τα διαφορετικά προγράμματα που χρησιμοποιούνται από τους χρήστες.

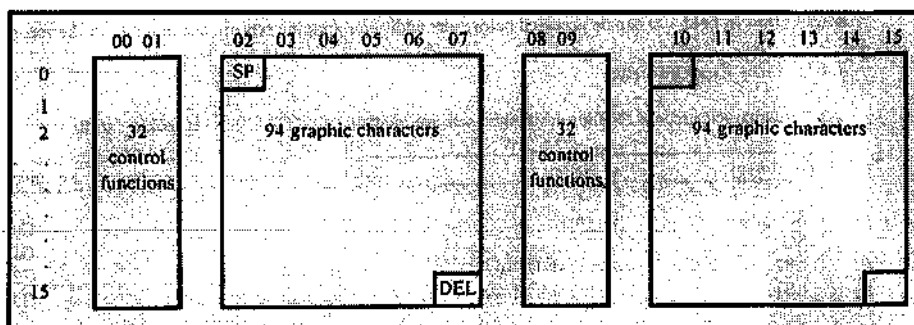
Το πρόβλημα της επιμέρους υποστήριξης των γραμμάτων των αλφάβητων, όχι μόνο της ελληνικής γλώσσας αλλά των υπολοίπων «απαιτητικών» αλφάβητων όπως το κινεζικό, έρχεται να καλύψει μια νέα επέκταση του πίνακα των χαρακτήρων και ένα νέο πρότυπο, το οποίο ονομάζεται Unicode. Ο πίνακας των χαρακτήρων από οκτάμπιτος αυξάνεται σε δεκαεξάμπιτο, και το σύνολο των χαρακτήρων ανέρχεται άμεσα σε 2^{16} , δηλαδή 65,536. Ακόμα και αυτός ο πολύ υψηλός αριθμός δεν επαρκεί δυστυχώς για να καλύψει όλους τους χαρακτήρες που χρησιμοποιούνται διεθνώς (μόνο τα κινεζικά αλφάβητα χρειάζονται πάνω από 30,000 θέσεις). Διαφορετικές μορφές του Unicode (UCS-2, UTF-8, κλπ) έχουν ήδη προταθεί χωρίς καμία να έχει επικρατήσει μέχρι σήμερα. Όσον αφορά τα λειτουργικά συστήματα των υπολογιστών, ακόμα δεν έχουν όλα τη δυνατότητα της πλήρους υποστήριξης του νέου προτύπου, αναμένεται ωστόσο αυτό να γίνει στα επόμενα ένα ή δύο χρόνια.

ΥΠΟΣΤΗΡΙΞΗ ΧΑΡΑΚΤΗΡΩΝ ΣΕ ΒΙΒΛΙΟΓΡΑΦΙΚΑ ΠΡΟΤΥΠΑ

Τα βιβλιογραφικά πρότυπα που αναφέρονται σε μηχανογραφημένα συστήματα καταλόγων (MARC) δεν ήταν δυνατό να μην βρίσκονται σε άμεση συνάρτηση με τις τεχνολογικές εξελίξεις των υπολογιστικών συστημάτων στο θέμα των χαρακτήρων. Άσχετα με το βαθμό υποστήριξης των χαρακτήρων στα υπολογιστικά συστήματα, οι εγγραφές των καταλόγων πρέπει να έχουν τη δυνατότητα υποστήριξης όλων των χαρακτήρων όλων των αλφάβητων. Η λύση που δόθηκε από την πρώτη στιγμή περιέχει τεχνικές ευθέως ανάλογες με αυτές που χρησιμοποιήθηκαν στα υπολογιστικά συστήματα, αλλά και καθιερώσεις σελίδων με τους χαρακτήρες των υποστηριζόμενων αλφάβητων. Ανάλογα με την τεχνολογία που υλοποιεί ο υπολογιστής που φιλοξενεί τον εκάστοτε κατάλογο (7-bit ή 8-bit) διαφοροποιούνται και οι οδηγίες ελέγχου. Στις επόμενες παραγράφους δεν θα παρουσιαστεί η υποστήριξη των 7-μπιτων συστημάτων, αφού είναι αμφίβολο αν υπάρχουν πλέον τέτοια συστήματα.

Το πρότυπο UNIMARC

Το UNIMARC επιμερίζει τον πίνακα χαρακτήρων των 256 θέσεων σε δύο κύρια μέρη, από 128 θέσεις το ένα, τα οποία ονομάζονται σελίδες. Κάθε μια από αυτές ξεκινάει με 32 θέσεις που ορίζονται ως χαρακτήρες ελέγχου, ενώ οι υπόλοιπες 96 θέσεις είναι διαθέσιμες για γράμματα. Οι δύο σελίδες αυτές ονομάζονται «χαμηλή» και «υψηλή» σελίδα, από τις τιμές που έχουν οι θέσεις σε αυτές (0-127 και 128-255). Κάθε σελίδα διατάσσεται σε στήλες των 16 θέσεων, με 8 στήλες, ακολουθείται δηλαδή επακριβώς η δομή των προτύπων των υπολογιστικών συστημάτων. Οι δύο πρώτες στήλες κάθε σελίδας των χαρακτήρων ελέγχου έχουν τους χαρακτηριστικούς κωδικούς G0 και C1 (Control functions 0 και 1). Οι στήλες των χαρακτήρων έχουν τους κωδικούς G0 και G1 (Graphic characters 0 και 1). Στο Σχήμα 3 φαίνεται με ευκρίνεια η διάταξη αυτή.



Σχήμα 3: Γενικός σχεδιασμός πινάκων χαρακτήρων σε 8-bit UNIMARC συστήματα

Το πρότυπο περιλαμβάνει μια σειρά από σελίδες που ορίζουν τους υποστηριζόμενους χαρακτήρες. Κάθε μια χαρακτηρίζεται από ένα μοναδικό κωδικό αριθμό αναγνώρισης. Έτσι, η βασική λατινική έχει τον κωδικό «01», η εκτεταμένη λατινική τον κωδικό «03», η ελληνική «05», η κυριλλική «04» κλπ. Στην περίπτωση της 8-μπιτης τεχνολογίας, δύο από αυτές τις σελίδες μπορούν να είναι ταυτόχρονα ενεργές στα δύο μέρη της κωδικοσελίδας του υπολογιστή. Το πρότυπο επιτρέπει την τοποθέτηση των σελίδων και στη χαμηλή και στην υψηλή θέση, ωστόσο μέχρι δύο σελίδες μόνο μπορούν να είναι ενεργές κάθε στιγμή. Το ερώτημα το οποίο προκύπτει άμεσα είναι ποια σελίδα είναι σε ποια θέση. Απάντηση έρχεται να δώσει το πεδίο 100, Γενικά Δεδομένα Επεξεργασίας, το οποίο περιλαμβάνει τους κωδικούς μέχρι τεσσάρων σελίδων που μετέχουν στην εγγραφή. Οι δύο πρώτες σελίδες ορίζονται ως το «Σύνολο Χαρακτήρων», θέσεις 26-29, και οι δύο επόμενες ως «Πρόσθετο Σύνολο Χαρακτήρων» και είναι στις θέσεις 30,-33. Αν χρησιμοποιείται μόνο μια σελίδα (π.χ. η λατινική) τότε μόνο ο κωδικός «01» εισάγεται, και όλες οι άλλες θέσεις μένουν κενές.

Το Σχήμα 4 περιέχει τη σελίδα των 96 θέσεων των λατινικών χαρακτήρων, και το Σχήμα 5 τη σελίδα των ελληνικών. Οι στήλες όλων των σελίδων έχουν αρίθμηση για τοποθέτηση στην χαμηλή σελίδα, στήλες δηλαδή 0-7. Αν κάποια σελίδα τοποθετηθεί στην υψηλή θέση η αρίθμηση που θα πρέπει να χρησιμοποιηθεί ξεκινάει από το 8 και εφεξής.

67	68	69	70	71	72	73	74
0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	1
0	0	1	0	0	1	0	1
0	0	1	1	0	0	1	1
0	0	1	1	1	0	0	1
0	1	0	0	0	0	0	0
0	1	0	0	1	0	0	0
0	1	0	1	0	0	0	0
0	1	0	1	1	0	0	0
0	1	1	0	0	0	0	0
0	1	1	1	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0
1	0	1	0	0	0	0	0
1	0	1	1	0	0	0	0
1	1	0	0	0	0	0	0
1	1	0	1	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	1	0	0	0	0

2	3	4	5	6	7
0	α	β	γ	δ	ε
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π	ρ	σ	τ	υ
φ	χ	ψ	ω	α	β
γ	δ	ε	ζ	η	θ
ι	κ	λ	μ	ν	ξ
ο	π				

προβληθεί σε ένα μηχανογραφικό σύστημα που δεν υποστηρίζει στην κωδικοσελίδα του ελληνικούς χαρακτήρες (π.χ. βλέπε Σχήμα 1) να υποστηρίζεται πλήρως. Αντίστοιχη περίπτωση είναι για την ελληνική πραγματικότητα οι εκτεταμένοι λατινικοί χαρακτήρες, που ανήκουν στη σελίδα των Εκτεταμένων λατινικών χαρακτήρων, ISO 5426-1983.

b	0	0	0	0	1	1	0	1				
b	0	0	3	1	1	0	0	1				
b	0	1	0	1	0	1	0	1				
			0	1	2	3	4	5	6	7		
b	b	b	b	0				<	N			
b	b	b	b	1				>	A	E	a	ξ
b	b	b	b	2				..	B	O	β	ο
b	b	b	b	3				..		Π	β	π
b	1	0	0	4				-	Γ	Φ	γ	φ
b	1	0	1	5				.	Δ	P	δ	ρ
b	1	1	0	6				.	E	Σ	ε	σ
b	1	1	1	7				.	ς		ς	ς
1	0	0	0	8					F	T	f	t
1	0	0	1	9					Z	Y	z	y
1	0	1	0	10					H	Φ	h	φ
1	0	1	1	11					Θ	X	θ	χ
1	1	0	0	12					I	Ψ	i	ψ
1	1	0	1	13					K	Ω	k	ω
1	1	1	0	14					Λ	λ	λ	λ
1	1	1	1	15					M		μ	

Σχήμα 5: Σελίδα ελληνικών – ISO 5428-1984

Όπως αναφέρθηκε, οι σελίδες που μπορούν να μετέχουν σε μια UNIMARC εγγραφή είναι καταρχήν τέσσερις, ωστόσο μόνο δύο μπορούν να είναι ενεργές ταυτόχρονα. Επομένως, είναι απαραίτητη η ύπαρξη μιας διαδικασίας με την οποία μια από τις σελίδες που βρίσκονται στην τρίτη και τέταρτη θέση να μετακινούνται προσωρινά στην πρώτη ή δεύτερη. Αυτό επιτυγχάνεται μέσω κάποιων οδηγιών ελέγχου, που ονομάζονται μετακινήσεις κλειδώματος (locking shift). Υπάρχουν καθιερωμένες οδηγίες ελέγχου για όλες τις πιθανές μετακινήσεις σελίδων, π.χ. της τρίτης στη δεύτερη, της τέταρτης στην πρώτη, κλπ. Αν υποθέσει κανείς ότι οι δύο πρώτες σελίδες είναι η «01» και «03», βασική και εκτεταμένη λατινική, η τρίτη η ελληνική «05», και θελήσει να εισάγει ελληνικούς χαρακτήρες, τότε θα πρέπει να εισαχθεί μια οδηγία μετακίνησης η οποία να λέει «τοποθέτησε την τρίτη σελίδα στην δεύτερη» πριν την εισαγωγή τους. Το πεδίο 100 θα περιέχει στα συγκεκριμένα σημεία τα ψηφία «010305bb» (όπου «b» διάστημα). Η συγκεκριμένη οδηγία είναι η «LS2R», και σε ορολογία υπολογιστή «ESC 7/13». Ενεργοποιείται η σελίδα «05», και ο καταλογογράφος θα πρέπει να εισάγει τους χαρακτήρες με τη θέση που έχουν στον πίνακα του UNIMARC, «C1» για το κεφαλαίο ελληνικό άλφα, «C2» για το βήτα, κλπ.

Ένα σημαντικό σημείο είναι το θέμα των τονικών συμβόλων. Αν συγκρίνει κανείς τις κωδικοσελίδες των υπολογιστικών συστημάτων με τις σελίδες του UNIMARC, θα εντοπίσει εύκολα ότι οι δεύτερες δεν περιέχουν συνδυασμούς γραμμάτων με τονικά σύμβολα. Τα τονικά σύμβολα, ή γενικότερα τα σύμβολα τα οποία δεν υπάρχουν παρά μόνο

συγκεκριμένη μορφή του Unicode που θα υποστηρίζεται (UCS2, UCS4, UTF-16, UTF-8, κλπ), ακόμα και αν θεωρείται ως επικρατέστερη η επιλογή του UTF-8.

ΔΙΑΔΙΚΤΥΩΣΗ ΜΗΧΑΝΟΓΡΑΦΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΒΙΒΛΙΟΘΗΚΩΝ

Δεν είναι πολλά τα διαφορετικά συστήματα τα οποία έχουν εγκατασταθεί σήμερα στον ελληνικό χώρο. Ωστόσο υπάρχει μια ανομοιογένεια στις επιλογές που έχουν υιοθετηθεί, η οποία προφανώς δεν επιτρέπει την ανταλλαγή βιβλιογραφικών πληροφοριών μεταξύ των διάφορων συστημάτων. Το παράδοξο είναι ότι από το έτος 1984 υπάρχει διεθνές πρότυπο για την ανταλλαγή βιβλιογραφικών πληροφοριών με ελληνικούς χαρακτήρες, το οποίο έχει ακολουθηθεί από ισχυρή μειονότητα κατασκευαστών λογισμικού μόνο. Έτσι, ενώ το UNIMARC αναφέρεται σαφώς στο πρότυπο αυτό, τα συστήματα που λειτουργούν περιέχουν υλοποιήσεις του προτύπου ΕΛΟΤ-928 ή του συγγενικού Microsoft Windows 1253, του παλαιότερου IBM 437, κλπ. Ακόμα και αν οι επιλογές αυτές παραβιάζουν τα βιβλιογραφικά πρότυπα, τουλάχιστον βασίζονται σε ένα διεθνές πρότυπο, ενώ σε μια περίπτωση για να ξεπεραστεί το πρόβλημα των χαρακτήρων που δεν υποστηρίζονται υλοποιούνται ιδιότυπες επιλογές, οι οποίες προφανώς οδηγούν σε καταλόγους που δεν έχουν τη δυνατότητα να ανταλλάξουν με αυτόματο τρόπο οποιαδήποτε βιβλιογραφική πληροφορία που περιέχει τους χαρακτήρες αυτούς.

Είναι προφανές ότι πριν από την όποια προσπάθεια διασύνδεσης μηχανογραφικών συστημάτων βιβλιοθηκών θα πρέπει να υπάρξει μια ολική αναγνώριση του προβλήματος και παροχής συγκεκριμένων λύσεων οι οποίες θα πρέπει να μην παραβιάζουν τα διεθνή καθιερωμένα πρότυπα, ενώ θα ήταν θετική εξέλιξη η δραστηριοποίηση των βιβλιοθηκών του ελληνικού χώρου προς την κατεύθυνση της συμμετοχής στις διεθνείς επιτροπές που δημιουργούν τα πρότυπα αυτά, και την παρακολούθηση των εξελίξεων τους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Unimarc manual: bibliographic format / International Federation of Library Associations and Institutions, IFLA Universal Bibliographic Control and International MARC Programme. Die Deutsche Bibliothek, Frankfurt am Main. - 2nd ed. - Munchen ; New Providence ; London ; Paris : Saur, c1994. - (UBCIM Publications - New Series ; Vol. 14). — ISBN 3-598-11211-4
2. Unimarc manual: bibliographic format / International Federation of Library Associations and Institutions, IFLA Universal Bibliographic Control and International MARC Programme. Die Deutsche Bibliothek, Frankfurt am Main..- 2nd ed. ; Update 3. - Munchen : Saur, c2000. - (UBCIM Publications - New Series ; Vol. 14). - ISBN 3-598-11214-9
3. USMARC specifications for record structure, character sets, and exchange media / prepared by Network Development and MARC Standards Office. - 1994 ed. - Cataloging Distribution Service Library of Congress : Washington, c1994. - ISBN 0-8444-0746-1 (pbk.)
4. ISO 5428 - 1984 (E) Greek alphabet coded character set for bibliographic information interchange
5. The Unicode Standard : version 3.0 / The Unicode Consortium. - Reading, Mass. (et al): Addison - Wesley, c1991-2000. - ISBN 0-201-616333-5