

# ΕΞΕΛΙΞΕΙΣ ΣΤΟ ΧΩΡΟ ΤΩΝ ΣΥΝΟΛΩΝ ΧΑΡΑΚΤΗΡΩΝ ΜΕ ΑΝΑΦΟΡΑ ΣΤΟ ΕΛΛΗΝΙΚΟ ΣΥΣΤΗΜΑ ΓΡΑΦΗΣ ΚΑΙ ΣΤΗΝ ΚΑΤΑΛΟΓΟΓΡΑΦΗΣΗ

By Mr MICK RIDLEY,  
DEPARTMENT OF COMPUTING,  
UNIVERSITY OF BRADFORD, U.K

## Εισαγωγή

Θα σας παρουσιάσω για την υπάρχουσα κατάσταση σχετικά με τα σύνολα χαρακτήρων, τα οποία χρησιμοποιούνται στους υπολογιστές. Κατ' αρχάς θα παρουσιάσω μερικές βασικές έννοιες του χώρου και στη συνέχεια θα περιγράψω την τακτική που ακολουθείται, υπογραμμίζοντας όσα πιστεύω ότι προμηνύουν μελλοντικές εξελίξεις.

Περνάμε μια περίοδο πλούσια σε δραστηριότητα στο χώρο των συνόλων χαρακτήρων. Αυτή η δραστηριότητα μας έχει φέρει στα πρόθυρα της επίλυσης πολλών προβλημάτων που μας έχουν απασχολήσει μέχρι τώρα. Η λύση τους θα επιτευχθεί με τη χρήση συνόλων χαρακτήρων που αποτελούνται από πολλαπλά BYTE στη μορφή των προτύπων UNICODE και ISO 10646. Ευρισκόμενοι στις παρυφές αυτών των εξελίξεων, τα ειδικά χαρακτηριστικά και η ταχύτητα με την οποία θα εισαχθούν δεν είναι δυνατόν να καθοριστούν ακόμα.

## Τι είναι ένα σύνολο χαρακτήρων υπολογιστή (character sets)

Σ' αυτό το σημείο θα μεταχειριστώ διάφορα παραδείγματα που σχετίζονται κυρίως με προσωπικούς υπολογιστές. Παρόμοια προβλήματα παρουσιάζονται και σε άλλα περιβάλλοντα. Θα πρέπει να λάβουμε υπ' όψη μας ότι οι υπολογιστές δεν αποθηκεύουν στη μνήμη τους γράμματα ή αριθμούς, αλλά σχηματισμούς δυαδικών ψηφίων (bit-patterns) και ότι οι σχηματισμοί αυτοί στη συνέχεια ερμηνεύονται έτσι ώστε να μπορούν να χρησιμοποιηθούν ως γράμματα ή αριθμοί. Στα συμβατικά σύνολα χαρακτήρων, 8 BIT θεωρούνται ως μια μονάδα (ένα BYTE) και μια τέτοια ομάδα από δυαδικά ψηφία χρησιμοποιείται για την αναπαράσταση ενός χαρακτήρα.

## ASCII (American Standard Code for Information Interchange)

Το σύνολο χαρακτήρων που ονομάζεται ASCII είναι αναμφίβολα το πιο διαδεδομένο, που χρησιμοποιείται αυτή τη στιγμή, επιφέροντας συνοχή και ομοιογένεια στην αναπαράσταση του λατινικού αλφαβήτου. Το μοναδικό άλλο σημαντικό σύνολο χαρακτήρων είναι το

EBCDIC με ποικίλες μορφές, που συνήθως το συναντάμε σε δίκτυα και όχι σε προσωπικούς υπολογιστές. Ο όρος ASCII χρησιμοποιείται ευρέως για να περιγράψει ένα σύνολο χαρακτήρων, όπου ο κάθε χαρακτήρας αποτελείται από 1 BYTE και που τοποθετεί τους λατινικούς χαρακτήρες στις ίδιες θέσεις όπως το ASCII. Αυτό είναι μια μορφή αναπαράστασης χαρακτήρων με ομάδες 7 δυαδικών ψηφίων, όπου ένα δυαδικό ψηφίο από το κάθε BYTE δεν χρησιμοποιείται για την κωδικοποίηση χαρακτήρων. Με αυτό τον τρόπο περιγράφει 128 (δηλ. 2 στην εβδόμη) χαρακτήρες. Ο σχηματισμός 0100001 αναπαριστά τον χαρακτήρα A. Εν τούτοις, δεν είναι και οι 128 σχηματισμοί διαθέσιμοι για την αναπαράσταση χαρακτήρων, μόνο 95 από αυτούς διατίθενται για γράμματα, αριθμούς και σημεία στίξεως. Οι θέσεις μέσα σε ένα σύνολο χαρακτήρων περιγράφονται με ένα πλήθος διαφορετικών τρόπων με δυαδικούς, δεκαδικούς οκταδικούς, δεκαεξαδικούς αριθμούς, ή με αναφορά στο συνδυασμό στήλης - σειράς ενός πλέγματος σε δεκαεξαδική αρίθμηση με στήλες αριθμημένες από 0 έως 15. Έτσι το A που ανέφερα πιο πάνω είναι μεν 0100001 σε δυαδικό, όμως επίσης 65 (065) σε δεκαδικό, 101 σε οκταδικό και 41H σε δεκαεξαδικό. Στο εξής θα αναφέρομαι σε θέσεις του συνόλου χαρακτήρων με δεκαδικούς αριθμούς.

Αν χρησιμοποιηθούν και τα 8 BIT ενός BYTE για την αναπαράσταση χαρακτήρων τότε προκύπτουν 256 συνδυασμοί (2 στην ογδόη) αντί για 128. Συνήθως οι προσωπικοί υπολογιστές μεταχειρίζονται ASCII (7 BIT) και ότι απομένει, δηλ. οι θέσεις από 128 έως 255, διατίθενται για χαρακτήρες άλλων αλφαβήτων, γραφικούς χαρακτήρες κλπ. Αυτά είναι τα λεγόμενα εκτεταμένα ASCII σύνολα χαρακτήρων. Ο βαθμός τυποποίησης της επέκτασης του ASCII ποικίλει. Μερικές επεκτάσεις ορίζονται από εθνικά και διεθνή πρότυπα, ενώ άλλες δεν έχουν επίσημη υποστήριξη, έτσι ώστε τίποτα να μην διασφαλίζει την αποδοχή τους και ως εκ τούτου την συμβατότητά τους με άλλα περιβάλλοντα. Το ASCII αποτέλεσε με βάση του ISO 646 και ένα πλήθος επεκτάσεων που καλύπτονται από το ISO 8859-X όπου X μπορεί να είναι 1,2 κλπ. Τα τελευταία

υποστηρίζουν διάφορα λατινογενή συστήματα γραφής χρησιμοποιώντας τις επί πλέον θέσεις (128 - 255) για χαρακτήρες που μεταχειρίζεται το κάθε ένα από αυτά τα συστήματα γραφής και που δεν συγκαταλέγονται μεταξύ των βασικών λατινικών χαρακτήρων ASCII στις πρώτες 128 θέσεις.

Το σύστημα DOS προβλέπει τη χρήση ορισμένων επεκτάσεων του ASCII καθώς και τη δυνατότητα μεταπήδησης από μια επέκταση σε άλλη (Code Pages). Στις εκδόσεις του εγχειριδίου του DOS για χώρες που χρησιμοποιούν λατινογενείς αλφάβητο, το τμήμα που αναφέρεται στη χρήση αυτών ακριβώς των επεκτάσεων δεν περιλαμβάνει καμιά αναφορά στο Ελληνικό ή το Κυριλλικό αλφάβητο. Ορισμένοι από τους ελληνικούς χαρακτήρες που μπορεί να εμφανιστούν συχνά δεν είναι δυνατόν να τους αποδώσουμε με ένα συμβατικό πληκτρολόγιο. Στην αγγλική επέκταση τα ελληνικά γράμματα άλφα και βήτα είναι παρόντα και διαθέσιμα στις θέσεις ALT 224 και ALT 225 αντίστοιχα. Αυτές οι θέσεις δεν εκπροσωπούν πάντα τα γράμματα άλφα και βήτα. Φερ' ειπείν στην επέκταση που ονομάζεται Latin 1 η θέση ALT 224 αντιστοιχεί στο γράμμα όμικρον κεφαλαίο τονούμενο. Σε πολλά εκτεταμένα ASCII ένα μεγάλο τμήμα των θέσεων 128 ως 225 καταλαμβάνεται από χαρακτήρες με διακριτικά. Η πιο συνηθισμένη αναπαράσταση ενός τέτοιου χαρακτήρα συναντάται ως συνδυασμός του βασικού χαρακτήρα με το διακριτικό, που μαζί καταλαμβάνουν μια θέση στο σύνολο χαρακτήρων. Για παράδειγμα τέτοιες θέσεις στο σύνολο χαρακτήρων αποδίδονται σε όλους τους πιθανούς συνδυασμούς του χαρακτήρα e. Στο Latin 1 συγκεκριμένα υπάρχουν 4 τέτοιοι συνδυασμοί με acute, circumflex, umlaut grave και αντιστοιχούν στις θέσεις 130, 136, 137, 138 αντίστοιχα, πέρα από το e που υπάρχει μόνο του στη θέση 101 του βασικού ASCII. Μια εναλλακτική μέθοδος αναπαράστασης τονούμενων χαρακτήρων επιτυγχάνεται με τη χρήση του βασικού χαρακτήρα σε συνδυασμό με ένα ανεξάρτητο διακριτικό ή διακριτικά. Η μέθοδος παρουσιάζει το πλεονέκτημα να μας επιτρέπει να κατασκευάσουμε κατά κάποιο τρόπο χαρακτήρες με την προϋπόθεση βέβαια ότι και η βάση και τα διακριτικά θα είναι παρόντα στο σύνολο χαρακτήρων.

Η τυποποίηση που έφερε το ASCII ήταν ένα μεγάλο βήμα προς τα μπρος στο χώρο της πληροφορικής, όμως παράλληλα επιβάλλει σημαντικούς περιορισμούς. Ένα κείμενο που χρησιμοποιεί τα γράμματα του λατινικού

αλφάβητου χωρίς διακριτικά μπορεί να μεταφερθεί εύκολα και να ερμηνευθεί από οποιοδήποτε υπολογιστή χωρίς πρόβλημα και κίνδυνο απώλειας πληροφοριών. Κείμενα που περιέχουν τονούμενους χαρακτήρες όμως, είναι δυνατόν να παρερμηνευθούν κατά τη μεταφορά τους από μια μηχανή σε μια άλλη.

Θεώρησα αναγκαίο τον τονισμό όλων αυτών των σημείων, διότι αντιλαμβάνομαι ότι το θέμα των συνόλων χαρακτήρων δεν βρίσκεται συνήθως στην περιοχή του άμεσου ενδιαφέροντος ενός χρήστη. Οι περισσότεροι χρήστες προσωπικών υπολογιστών δεν γνωρίζουν αν χρησιμοποιούν ASCII ή EBCDIC. Συνήθως τα προβλήματα παρουσιάζονται όταν προσπαθήσουμε να τυπώσουμε κάτι σχετικά ασυνήθιστο. Το άλλο σημείο, όπου αρχίζουν να διαφαίνονται προβλήματα, είναι όταν προσπαθήσουμε να συνδέσουμε υπολογιστές σε ένα δίκτυο και να μεταφέρουμε πληροφορίες από ένα σύστημα σε άλλο και εδώ ακριβώς είναι που το πρόβλημα γίνεται ολοένα και πιο έντονα αισθητό.

#### **Τα σύνολα χαρακτήρων και η χρήση τους στη βιβλιογραφία**

Οι βιβλιογραφικές εφαρμογές έχουν τις δικές τους απαιτήσεις, διαφορετικές και συχνά μεγαλύτερες από εκείνες του συνηθισμένου χρήστη υπολογιστών. Ακόμα και όταν η καταλογογράφηση γίνεται σε λατινική γραφή, πολλές βιβλιοθήκες χρειάζονται ένα ρεπερτόριο χαρακτήρων μεγαλύτερο από το βασικό ASCII των 128 θέσεων. Τα διάφορα εθνικά συστήματα MARC είναι συνήθως εκτεταμένα ASCII με επιπρόσθετους τονούμενους και άλλους χαρακτήρες. Ο βαθμός τυποποίησής τους ποικίλει όπως και στην περίπτωση των προσωπικών υπολογιστών που έχω ήδη αναφέρει. Με άλλα λόγια η τοπολογία των επιπρόσθετων χαρακτήρων της περιοχής θέσεων από 128 ως 255 διαφέρει στα διάφορα MARC. Σε μερικές περιπτώσεις παρατηρούνται διαφορές ακόμα και στην περιοχή από 0 ως 127. Σε αυτή την περίπτωση μιλάμε πλέον για όχι απλώς εκτεταμένο ASCII αλλά για εκτεταμένη παραλλαγή του (Extended Variant ASCII).

Ένα σημαντικό ζήτημα στις βιβλιογραφικές εφαρμογές είναι η δυνατότητα/ή μη, διαφορετικών συσκευών να αναπαράγουν ολόκληρο το σύνολο χαρακτήρων που βρίσκεται σε χρήση. Επί πλέον οι δυνατότητες ενός εκτυπωτή μπορεί να είναι διαφορετικές από εκείνες ενός τερματικού OPAC με το οποίο είναι συνδεδεμένος. Αυτό έχει ιδιαίτερη σημασία για το ελληνικό υλικό που έχει αποθηκευτεί σε λατινοποιημένα αρχεία. Διάφορα πρότυ-

πα μεταγραμματισμού χρησιμοποιούν το χαρακτήρα  $\mu\alpha\sigma\rho\eta$  μαζί με το χαρακτήρα  $\omicron$  για να επιτύχουν τη διάκριση μεταξύ των γραμμάτων  $\omicron$  μικρον και  $\omega$  μέγα. Σπάνια όμως μπορεί να παραχθεί το  $\mu\alpha\kappa\rho\eta$  και συχνά αυτό το τμήμα πληροφορίας χάνεται προκαλώντας σύγχυση.

### Τα σύνολα χαρακτήρων και το ελληνικό αλφάβητο

Είναι δύσκολο να διευρινίσουμε τι αποτελεί ένα ελληνικό σύνολο χαρακτήρων ή ποιά από αυτά υποστηρίζουν ελληνικά. Πολλά σύνολα χαρακτήρων για λατινική γραφή περιλαμβάνουν μερικούς ελληνικούς χαρακτήρες. Αυτό συνήθως γίνεται για να επιτρέψουν την αναγραφή μαθηματικών κυρίως συμβόλων. Αυτά τα σύνολα χαρακτήρων δεν μας παρέχουν τη δυνατότητα να γράψουμε ελληνικό κείμενο. Επί πλέον το ρεπερτόριο ενός συνόλου, που επιτρέπει μονοτονική ελληνική γραφή, μπορεί να μην ικανοποιεί τις ανάγκες της βιβλιογραφικής χρήσης. Μπορεί να είναι αρκετό για την καθημερινή επικοινωνία, αλλά ένας κατάλογος βιβλιοθήκης μπορεί να απαιτεί τη χρήση συμβόλων παλιότερων μορφών της ελληνικής γραφής.

Τα σύνολα χαρακτήρων που θα περιγράψω στη συνέχεια έχουν αναπτυχθεί με σκοπό την υποστήριξη του μονοτονικού συστήματος γραφής. Δεν καλύπτουν το πλήθος τονούμενων χαρακτήρων που χρησιμοποιούνται σε παλαιότερες μορφές ελληνικής γραφής και δεν υποστηρίζουν τη σύνθεση πιο πολύπλοκων συνδυασμών χαρακτήρων με διακριτικά.

Τα παρακάτω παραδείγματα από τον Διεθνή Καταχωρητή του ISO (ISO International Register of Coded Character Sets), επιδεικνύουν μερικές από τις παραλλαγές των συνόλων χαρακτήρων που υποστηρίζουν ελληνικά. Θεωρητικά οι τοπολογικές διαφορές δεν θα έπρεπε να δημιουργούν προβλήματα αφού όλα αυτά τα σύνολα χαρακτήρων χρησιμοποιούνται σε συνδυασμό με το ISO 2022 και ως εκ τούτου το υπό χρήση σύνολο μπορεί να αλλάξει με τη βοήθεια ενός χαρακτήρα ESCAPE. Εν τούτοις στην πράξη, όπως παρατηρεί ο Adams, ελάχιστα συστήματα υποστηρίζουν τη δομή κωδικοποίησης που χρησιμοποιείται το ISO 2022, «... very few systems support the general ISO 2022 encoding architecture». Τα παραδείγματα επιδεικνύουν τη μεταβλητότητα της τοπολογίας των ελληνικών χαρακτήρων σε αντιπαράθεση με τη γενική σταθερότητα και ομοιογένεια του ASCII. Οι χαρακτήρες περιγράφονται με το συνδυασμό στήλης - σειράς που χρησιμοποιεί ο Καταχωρητής ISO, όπου οι θέσεις μεταξύ 128 και 255 είναι αριθμημένες ως στήλες από 0 έως 7, με 16 διαστήματα ανά σειρά. Έτσι 6/1 αντιστοιχεί στο δεκαδικό αριθμό 193.

Αρ. Καταχώρησης Κεφαλαίο αλφα Κεφαλαίο ωμεγα Μικρό αλφα

18		6/1	7/6
	4/1		
19 (κεφαλαία)	6/1	7/6	
	(4/1=κεφαλαίο Α)		
27 (κεφαλαία)	4/1	5/11	
	(6/1= μικρό α)		
31		4/1	5/8
	6/1		
55		4/1	5/13
	6/1		
126 (ΕΛΟΤ 928)	4/1	5/9	6/1

### IBM 437 και ΕΛΟΤ 928

Τα δύο σύνολα χαρακτήρων γνωστά ως IBM 437 και ΕΛΟΤ 928 είναι τα πιο διαδεδομένα σε χρήση με προσωπικούς υπολογιστές στην Ελλάδα. Άλλα σύνολα χρησιμοποιούνται, επίσης σε πολύ μικρότερη κλίμακα όμως. Το IBM 437 ήταν το πρώτο διαθέσιμο σύνολο που χρησιμοποιήθηκε στην Ελλάδα. Το ΕΛΟΤ 928 κατασκευάστηκε από τον οργανισμό ΕΛΟΤ (Ελληνικός Οργανισμός Τυποποίησης) και αντιστοιχεί στο ISO 8859-7. Επειδή και τα δύο αυτά χρησιμοποιούνται ευρέως, αποφασίσαμε να παραγάγουμε, στα πλαίσια του ερευνητικού προγράμματος HELEN, μια μορφή του λογισμικού για το καθένα. Και το δύο είναι εκτεταμένα ASCII, δηλ. σύνολα χαρακτήρων που αποτελούνται από ένα BYTE: οι θέσεις από 0 - 127 καταλαμβάνονται από το βασικό σύνολο ASCII ενώ οι ελληνικοί χαρακτήρες και όλα τα σχετικά σύμβολα βρίσκονται στις θέσεις από 128 έως 255.

Αμφότερα τα παραπάνω σύνολα υποστηρίζουν ένα ρεπερτόριο ελληνικών χαρακτήρων που επιτρέπει τις περισσότερες σύγχρονες εφαρμογές ελληνικής γραφής. Το ΕΛΟΤ 928 είναι λίγο λεπτομερέστερο. Και τα δύο σύνολα περιέχουν κεφαλαίους και μικρούς αλφαβητικούς χαρακτήρες και τούμενα φωνήεντα καθώς και μικρά «ι» και «υ» με διαλυτικά. Μόνο το ΕΛΟΤ όμως διαθέτει κεφαλαία «ι» και «υ» με διαλυτικά καθώς και τα μικρά «ι» και «υ» με διαλυτικά συν τόνο. Επί πλέον το ΕΛΟΤ 928 έχει τρεις θέσεις για διακριτικά μόνο, διαλυτικά, τόνο και διαλυτικά συν τόνο και μερικούς άλλους χαρακτήρες όπως ελληνικά εισαγωγικά.

Μερικές φορές το ΕΛΟΤ 928 παρουσιάζεται σαν να περιλαμβάνει ορισμένους χαρακτήρες στις ίδιες θέσεις που ορίζει το IBM 437, εκτός από τις σωστές θέσεις που προβλέπει το ΕΛΟΤ 928. Αυτό δεν είναι το σύνολο που έχει καταχωρηθεί επισήμως ως ΕΛΟΤ 928. Επίσης έχω βρει μια ανατύπωση αυτού του συνόλου χαρακτήρων όπου το σύμβολο επί τοις εκατόν ήταν στη θέση 181, όπου κανονικά έπρεπε να βρίσκεται ο συνδυασμός των χαρακτήρων διαλυτικά συν τόνο. Δεν ξέρω αν αυτό ήταν σκόπιμο ή έγινε κατά λάθος, π.χ. από ανεπάρκεια του

εκτυπωτή. Τέτοιου είδους αποκλίσεις από την επίσημη μορφή του προτύπου είναι συγκρίσεις με την κατάσταση που επικρατεί στην εφαρμογή του ASCII στην Αγγλία, όπου το σύμβολο της λίρας αντικαθίσταται από το σύμβολο του δολαρίου.

Όσον αφορά στη σχέση μεταξύ των δύο συνόλων χαρακτήρων IBM 437 και ΕΛΟΤ 928, το αρχικά πιο δημοφιλές IBM 437 φαίνεται να εκτοπίζεται από το ΕΛΟΤ 928 που χρησιμοποιείται στην ελληνική έκδοση του MS WINDOWS και στο λογισμικό δικτύων προσωπικών υπολογιστών. Ως αποτέλεσμα της εξάπλωσης του WINDOWS και των δικτύων, το ΕΛΟΤ 928 κερδίζει έδαφος. Η υιοθέτηση του WINDOWS από χρήστες στο διεθνή χώρο έγινε πολύ γρήγορα, λέγεται πως είναι το ταχύτερο λογισμικό σε πωλήσεις. Το ίδιο είναι πολύ πιθανό να συμβεί και στην ελληνική αγορά. Παράλληλα το IBM 437 μπορεί να συνεχίσει να χρησιμοποιείται από μεγάλους ελληνικούς οργανισμούς που μηχανογραφήθηκαν νωρίς και κατά συνέπεια έχουν επενδύσει κεφάλαια στη χρήση του IBM 437.

Από την άποψη της καταλογογράφησης υπάρχει και ένα ακόμα σημαντικό σύνολο χαρακτήρων και αυτό είναι το σύνολο που αναπτύχθηκε για την ελληνική εθνική βιβλιοθήκη από το ερευνητικό πρόγραμμα CEC LIBACT1/GR. Το βασικό αλφάβητο στο σύνολο χαρακτήρων της Εθνικής Βιβλιοθήκης της Ελλάδος καταλαμβάνει τις ίδιες θέσεις όπως το IBM 437. Περιλαμβάνει όμως μεγαλύτερο αριθμό χαρακτήρων επιχειρώντας έτσι να λύσει τα προβλήματα που προέκυψαν από την υποστήριξη αποκλειστικά του μονοτονικού συστήματος γραφής από τα άλλα σύνολα χαρακτήρων. Όντας ακόμα ένα σύνολο χαρακτήρων που αποτελείται από ένα BYTE, δεν παύει να περιορίζεται στις 256 θέσεις.

#### **Το Apple Mectintosh και οι ελληνικοί χαρακτήρες**

Το APPLE MACINTOSH είναι δημοφιλές στην Ελλάδα ειδικά για εφαρμογές στην επεξεργασία κειμένου. Έτσι στο ερευνητικό πρόγραμμα HELEN αναπτύσσεται παράλληλα και μια μορφή του λογισμικού προς χρήση με το APPLE MACINTOSH.

Για το σκοπό αυτό χρειάστηκε να γίνει έρευνα σχετικά με την υποστήριξη ελληνικών χαρακτήρων στο MACINTOSH. Ένα πρόβλημα που παρουσιάζεται σε σχέση με αυτό το είδος υπολογιστή είναι η σύγκριση ανάμεσα στις έννοιες του συνόλου χαρακτήρων και της γραφικής αναπαράστασης. Κανονικά θα έπρεπε να είναι ανεξάρτητες και η γραφική αναπαράσταση του συνόλου θα έπρεπε να αφορά μόνο στην εμφάνιση ενός χαρακτήρα. Οι χρήστες του MACINTOSH επιλέγουν μια γραφική αναπαράσταση για ένα κείμενο όπου μια συγκεκριμένη θέση

χαρακτήρα μπορεί να αντιστοιχεί σε ένα πλήθος διαφορετικών συμβόλων αντί να αντιστοιχεί σε διαφορετικές αναπαραστάσεις του ίδιου συμβόλου (π.χ. έντονο, πλαγιαστό, κλπ.). Οι γραφικές αναπαραστάσεις κατά TIMES και GR TIMES, φαίνονται ίδιες όσον αφορά το λατινικό αλφάβητο από άποψη εμφάνισης και τοπολογίας. Διαφορετικά σύμβολα παρουσιάζονται στην περιοχή από 128 ως 255 αλλά με την ίδια εμφάνιση. Το MACINDOSH διαθέτει μια γραφική αναπαράσταση χαρακτήρων συστήματος (System font), που είναι ουσιαστικά εκτεταμένο ASCII. Ελληνικά κείμενα μπορούν να παραχθούν στο MACINTOSH χρησιμοποιώντας γραφικές αναπαραστάσεις διαφορετικές από αυτήν και υπάρχουν διάφορες για αυτό το σκοπό, όπως Gr Times. Σε μια τέτοια γραφική αναπαράσταση μπορεί κανείς να συνδυάσει αγγλική και ελληνική μονοτονική γραφή. Ο συνδυασμός ελληνικής μονοτονικής με κάποια άλλη λατινογενή γραφή μπορεί να αποβεί προβληματικός. Φερ' ειπείν, για το συνδυασμό ελληνικών και ισπανικών το Gr Times δεν περιέχει όλα τα ισπανικά διακριτικά, ενώ βρίσκονται όλα στο Times, που όμως δεν περιλαμβάνει ελληνικά.

Υπάρχουν επίσης γραφικές αναπαραστάσεις χαρακτήρων, όπως το Dg Times Poly που υποστηρίζουν ένα μεγαλύτερο ρεπερτόριο ελληνικών χαρακτήρων, συνδυασμούς γράμματος με διακριτικά και αρχαϊκά σύμβολα όπως το δίγαμα, όμως τότε δεν υπάρχει αρκετό περιθώριο για να χωρέσει και το λατινικό αλφάβητο. Επομένως δεν είναι δυνατόν να συνδυάσουμε αγγλική με ελληνική πολυτονική γραφή.

#### **Σύνολα χαρακτήρων πολλαπλών BYTE**

Όπως το περιέγραψα παραπάνω, το ASCII είναι ένα σύνολο χαρακτήρων που αποτελείται από ένα BYTE, (8 δυαδικά ψηφία). Ένας άλλος τρόπος αναπαράστασης χαρακτήρων δημιουργείται με τη χρήση ενός συνόλου χαρακτήρων που αποτελείται από περισσότερα BYTE. Υπάρχουν μερικά τέτοια σύνολα που χρησιμοποιούνται για την υποστήριξη συστημάτων γραφής των χωρών της ανατολικής Ασίας (βλ. αρ. 2 της βιβλιογραφίας για παραδείγματα). Ο λόγος είναι ότι γι' αυτά τα συστήματα γραφής το σύνολο χαρακτήρων χρειάζεται να περιλάβει περισσότερους από 256 χαρακτήρες. Τα σύνολα αυτά δεν είναι συμβατά μεταξύ τους και υπάρχει ένα πλήθος από αλληλοσυγκρουόμενα κορεατικά, κινεζικά και ιαπωνικά.

Μια επιφύλαξη που εκφράζεται ως προς τη χρήση συνόλων χαρακτήρων που αποτελούνται από περισσότερα BYTE, είναι ότι καταλαμβάνουν πολύ μεγαλύτερο χώρο στη μνήμη ενός υπολογιστή. Αυτό είναι μεν αλήθεια, όμως τα πλεονεκτήματα που προκύπτουν από τη

χρήση ενός τέτοιου συνόλου είναι αρκετά για να ξεπεράσουν το πρόβλημα του κόστους. Άλλωστε τώρα το κόστος της μνήμης μειώνεται, ενώ εκείνο που αυξάνεται είναι το κόστος του λογισμικού. Η διαδικασία της τυποποίησης των συνόλων χαρακτήρων πολλών BYTE έχει ξεκινήσει και υποστηρίζεται από επιχειρήσεις στη μορφή του UNICODE και από εθνικούς οργανισμούς τυποποίησης στη μορφή των προτάσεων (1.5) για το ISO 10646. Τα σύνολα αυτά αποβλέπουν όχι μόνο στην ικανοποίηση των αναγκών της κάθε γλώσσας, αλλά και στην ενοποίηση, δηλ. στο να ικανοποιούν με ένα μόνο σύνολο τις ανάγκες όλων των συστημάτων γραφής. Έτσι αυτό το σύνολο χαρακτήρων προσφέρει μια ολοκληρωμένη λύση που κανένα σύνολο χαρακτήρων του ενός BYTE δεν μπορεί να προσφέρει. Επί πλέον, με την ολοκλήρωση της διαδικασίας της τυποποίησης, θα παρέχεται η δυνατότητα επικοινωνίας και μεταφοράς δεδομένων χωρίς απώλεια πληροφοριών και σύγχυση.

### Μελλοντικές εξελίξεις και συμπεράσματα

Όπως είδαμε μέχρι τώρα, υπάρχει ένα πλήθος λύσεων στο πρόβλημα της αναπαράστασης ελληνικών χαρακτήρων σε υπολογιστές. Πολλές από αυτές δεν είναι συμβατές μεταξύ τους. Είναι εύκολο να μεταπηδήσουμε από ένα σύνολο χαρακτήρων σε άλλο π.χ. μεταξύ των IBM 437 και ΕΛΟΤ 928, αυτό μπορεί να επιτευχθεί με τη βοήθεια απλών προγραμμάτων. Πρέπει όμως πάντα να ξέρουμε ποιο σύνολο δημιούργησε ένα κείμενο, διότι μπορούμε να μεταπηδήσουμε μόνο όταν γνωρίζουμε τόσο το σημείο εκκίνησης όσο και τον προορισμό. Άλλωστε το γεγονός ότι ένας χαρακτήρας μπορεί να μην είναι παρών και στα δύο σύνολα, μπορεί να οδηγήσει σε απώλεια πληροφοριών.

Ασφαλώς δεν μπορούμε να προβλέψουμε το μέλλον, όμως ένας αριθμός παραγόντων οδηγεί στην άποψη ότι τα σύνολα χαρακτήρων πολλών BYTE θα βρουν ευρεία εφαρμογή, χωρίς αυτό να σημαίνει ότι η μετάβαση από την υπάρχουσα κατάσταση θα είναι σύντομη ή εύκολη. Ένας παράγων είναι η συνεργασία ανάμεσα στα UNICODE και ISO 10646, αν και τα προβλήματα όσον αφορά στις γλώσσες της ανατολικής Ασίας παραμένουν. Ένας άλλος παράγων είναι το επιχειρηματικό ενδιαφέρον που έχει επενδυθεί στο UNICODE εμπλέκοντας κατασκευαστές όπως IBM, MICROSOFT, APPLE, SUN, DIGITAL, NOVELL. Θα πρέπει να αναφερθεί ότι σε όλη την πορεία της εξέλιξης υπήρξε και βιβλιογραφική συμμετοχή. Στην ανάπτυξη του UNICODE, η ομάδα American Research Libraries Group έχει συμμετάσχει από τα πρώτα βήματα. Το UNICODE χρησιμοποιείται ήδη και είναι δυνατόν να χρησιμοποιηθεί σε διάφορα προϊόντα, όπως τα

πιο πρόσφατα X (για τις οθόνες γραφικής απεικόνισης σε συστήματα UNIX), και στα WINDOWS NT. Υποστήριξη του UNICODE από το WINDOWS NT και την APPLE μπορεί να οδηγήσουν στην υιοθέτηση αυτής της λύσης για την επίλυση των προβλημάτων που αντιμετωπίζουν οι χρήστες προσωπικών υπολογιστών. Αν και πολλοί χρήστες ίσως να μη θελήσουν ή να μη χρειαστούν να το υιοθετήσουν, κάτι τέτοιο θα πρέπει να αποτελέσει την πλατφόρμα εκκίνησης για σημαντικές εξελίξεις. Ένα γεγονός που συνηγορεί με αυτή την άποψη είναι η δημοσίευση δύο άρθρων σε περιοδικά που απευθύνονται σε χρήστες προσωπικών υπολογιστών. Συγκεκριμένα στο BYTE και στο PC MAGAZINE. Ένα άρθρο στο BYTE του Ιουλίου του 1991 με τον τίτλο «ASCII Goes Global» εξέθεσε στο κοινό τα ζητήματα και τις πιθανότητες να γίνουν τα δύο χωριστά (τότε) σύνολα χαρακτήρων αποδεκτά. Το άρθρο στο PC MAGAZINE, δύο χρόνια αργότερα, τον Οκτώβριο του 1993, ήταν πιο θετικό υποστηρίζοντας την υιοθέτηση των συνόλων χαρακτήρων πολλαπλών BYTE, με τον τίτλο «Move over ASCII, Unicode is Here». Αν και αυτή ίσως θεωρηθεί υπερβολικά ενθουσιώδης αντίδραση, δεν είναι υπερβολή να πούμε ότι το UNICODE έχει σχεδόν καταφθάσει και ότι οποιαδήποτε ανάπτυξη σ' αυτό το χώρο που δεν θα το λάβει υπόψη της είναι πιθανόν να μετατραπεί σε ιστορικό αδιέξοδο. Κάθε τέτοιο εγχείρημα θα οδηγούσε στην εξειδίκευση και τον τοπικισμό, ενώ μόνο ένα σύνολο χαρακτήρων που αποτελείται από πολλαπλά BYTE είναι σε θέση να υποστηρίξει όλα τα υπαρκτά συστήματα γραφής συμβάλλοντας έτσι στην ενότητα.

### Βιβλιογραφία

1. Adams, G. Internationalization and Character Set Standards. ACM Standard View 1,1 Sept. 1993, 31-39.
2. Glews, J. Language Automation Worldwide: The Development of Character Set Standards, Sesame Vomputer Projects, 1988.
3. Petzold, G. Move over, ASCII. Unicode Is Here, PC Magazine, 26 Oct. 1993, 374 - 376.
4. Sheldon, K.M. ASCII Goes Global, Byte July 1991, 108 - 116.
5. The Unicode Consortium. The Unicode Standard, Version 1.0 vulumew 1 & 2, Addison Wesley, 199.