Τμήμα Μηχανικών Πληροφοριακών
και Επικοινωνιακών Συστημάτων

«ΣΧΕΔΙΑΣΗ ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ ΕΛΛΗΝΙΚΩΝ ΧΕΙΡΟΓΡΑΦΩΝ ΧΑΡΑΚΤΗΡΩΝ»

Η Διπλωματική Εργασία
παρουσιάστηκε ενώπιον
του Διδακτικού Προσωπικού του
Πανεπιστημίου Αιγαίου

Σε Μερική Εκπλήρωση
των Απαιτήσεων για το Δίπλωμα του
Μηχανικού Πληροφοριακών και Επικοινωνιακών Συστημάτων

των:

Μαργαρώνη Ιωάννη	321/2001038
Χρήστου Μηνά	321/2001074

ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2007

ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή

Η συγκεκριμένη διπλωματική εργασία στοχεύει στη δημιουργία μίας βάσης δεδομένων ελληνικών χειρόγραφων χαρακτήρων κατάλληλους για εκπαίδευση και αξιολόγηση εφαρμογών οπτικής αναγνώρισης χαρακτήρων (OCR – Optical Character Recognition).

Σήμερα η χρήση εφαρμογών οπτικής αναγνώρισης είναι ιδιαίτερα διαδεδομένη ακόμα και στον ελληνικό χώρο της πληροφορικής και είναι εμφανής η ανάγκη μίας βάσης δεδομένων με κατάλληλο σχεδιασμό που να ικανοποιεί τις απαιτήσεις αυτών των εφαρμογών τόσο για την αξιολόγηση τους, όσο και για την τελική εκπαίδευση του συστήματος.

Η διπλωματική εργασία αυτή, αποτελείται από μία βάση δεδομένων, υλοποιημένη και σε SQL Server και MS Access, με την ονομασία GCDB (Greek Characters DataBase – Βάση Δεδομένων Ελληνικών Χαρακτήρων), μία φόρμα εισαγωγής δεδομένων, και δύο προγραμμάτων λογισμικού, ένα για την εισαγωγή των δεδομένων από τη φόρμα στη βάση και ένα για την εξαγωγή δεδο μένων από τη βάση για εκπαίδευση ή αξιολόγηση εφαρμογών οπτικής αναγνώρισης.

Το παρών κείμενο χωρίζεται σε 4 τμήματα, όπου περιγράφονται τα επιμέρους τμήματα της διπλωματικής εργασίας. Το πρώτο τμήμα είναι εισαγωγικό και περιγράφει την υπάρχουσα κατάσταση στις ελεύθερες βάσεις χειρόγραφων χαρακτήρων καθώς και την φόρμα εισαγωγής δεδομένων που έχει δημιουργηθεί για τη συλλογή δεδομένων της βάσης. Στο δεύτερο τμήμα γίνεται περιγραφή της βάσης δεδομένων αυτής κάθε αυτής και παρατίθενται τα απαραίτητα διαγράμματα και πληροφορίες. Στο τρίτο τμήμα γίνεται εκτενής περιγραφή του λογισμικού που δημιουργήσαμε τόσο για την εισαγωγή, όσο και για την εξαγωγή δεδομένων από τη βάση, το διαγράμματα ροής δεδομένων και όποιες σχετικές πληροφορίες με το λογισμικό. Τέλος στο τεχνικό παράρτημα της διπλωματικής δίνονται οδηγίες επέκτασης της βάσης καθώς και οι αλλαγές που πρέπει να γίνουν σε περίπτωση που αλλάξει η φόρμα εισαγωγής δεδομένων.

1.1 Περιγραφή της υπάρχουσας κατάστασης

Στις μέρες μας κρίνεται απαραίτητη η ύπαρξη Βάσεων Δεδομένων (ΒΔ – Βάση Δεδομένων) με χειρόγραφους χαρακτήρες οι οποίοι μπορούν να χρησιμοποιηθούν είτε για εκπαίδευση προγραμμάτων OCR εφαρμογών ή για σχετική εκτίμηση απόδοσης (Benchmark) τέτοιου είδους λογισμικού. Στο παρόν κεφάλαιο θα αναλύσουμε την υπάρχουσα κατάσταση στον τομέα αυτό σήμερα και θα αναφέρουμε τις κυριότερες ΒΔ χειρόγραφων χαρακτήρων.

Τέτοιες βάσεις συνήθως χρησιμοποιούνται κυρίως σε on-line και off-line αναγνώριση όμως μπορούμε να τις διαχωρίσουμε με βάση τον τρόπο αποθήκευσης τέτοιων χαρακτήρων (π.χ. σύμβολα, λέξεις, αποσπάσματα κειμένων, ψηφίων, κ.ά.). Στην πραγματικότητα όμως κάθε τέτοια βάση εξυπηρετεί το δικό της σκοπό, άλλες για εκπαίδευση συστημάτων και άλλες για benchmarking.

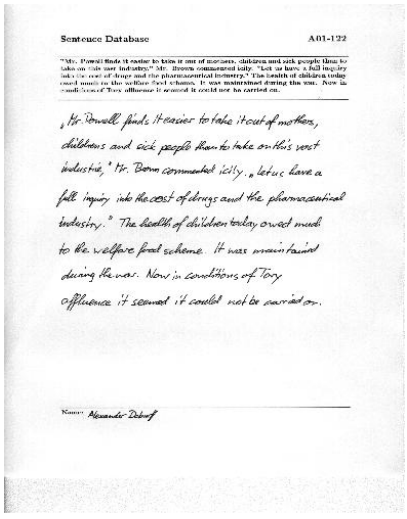
On-line αναγνώριση χαρακτήρων είναι η διαδικασία μετατροπής συμβόλων από την χωρική τους μορφή στην συμβολική τους αναπαράσταση την ίδια στιγμή της διαδικασίας γραφής. Σε αντίθεση η off-line αναγνώριση χαρακτήρων χρησιμοποιείται κυρίως για την ψηφιοποίηση είδη υπαρχόντων εικόνων εγγράφων.

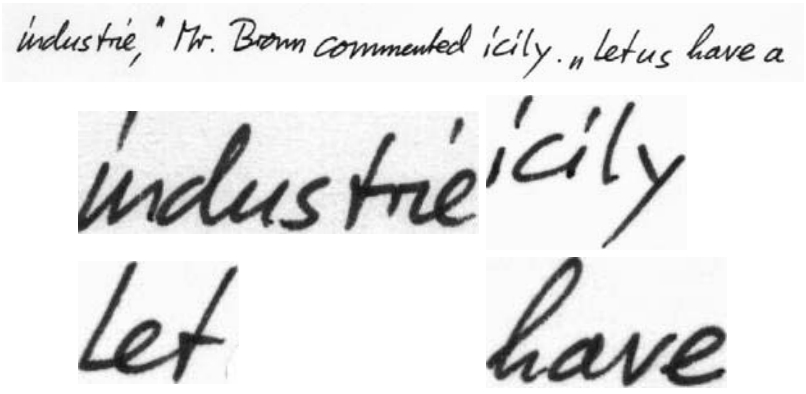
Οι κυριότερες βάσεις που υπάρχουν είναι οι εξής:

IAM

Κατάλληλη για να εκπαιδευθούν και να εξεταστούν τα συστήματα που είναι σε θέση να αναγνωρίσουν το αβίαστο χειρόγραφο αγγλικό κείμενο, δημιουργήθηκε η βάση δεδομένων IAM και δημοσιεύθηκε αρχικά στο *ICDAR το 1999*. Χρησιμοποιώντας αυτήν την βάση δεδομένων αναπτύχθηκε ένα σύστημα αναγνώρισης, βασισμένο στο HMM, για τις χειρόγραφες προτάσεις και δημοσιεύθηκε στο *ICPR το 2000*.

Η βάση δεδομένων περιέχει τις μορφές αβίαστου χειρόγραφου κειμένου, το οποίο ανιχνεύθηκε σε ανάλυση 300dpi και αποθηκεύτηκε σε μορφή εικόνων TIFF με 256 γκρίζα επίπεδα. Η παρακάτω εικόνα περιέχει τα δείγματα μιας πλήρους μορφής, μιας γραμμής του κειμένου και μερικών αποσπασματικών λέξεων.





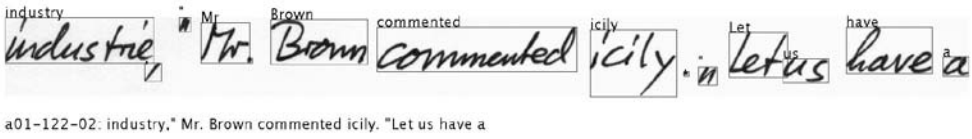
Από το σύστημα είναι διαθέσιμο τόσο το αρχείο εικόνας (τύπου TIFF), όσο και το αντίστοιχο αρχείου ετικετών (μορφής XML), που περιέχει το κείμενο του χειρόγραφου τμήματος και πληροφορίες για αυτό σε ASCII μορφή. Όλα τα κείμενα που παρέχονται στη βάση δεδομένων IAM χτίζονται χρησιμοποιώντας τις προτάσεις που παρέχονται από τη συλλογή κειμένων Lancaster-Oslo/Bergen η οποία περιέχει 500 τυπωμένα κείμενα με περίπου 2000 λέξεις το καθένα.

Χρησιμοποιώντας κατά προσέγγιση εκτιμήσεις το περιεχόμενο της βάσης δεδομένων μπορεί να συνοψιστεί ως εξής:

- 500 συγγραφείς, με δείγματα της γραφής τους
- 1500 σελίδες ανιχνευμένου κειμένου
- 10.000 απομονωμένες και ονομασμένες γραμμές κειμένων
- 100.000 απομονωμένες και ονομασμένες λέξεις

Πρέπει να σημειωθεί ότι είναι διαθέσιμες 27.000 απομονωμένες λέξεις αυτήν την περίοδο. Αυτές οι λέξεις έχουν εξαχθεί από 400 σελίδες σαρωμένου κειμένου χρησιμοποιώντας μία αυτόματη διαδικασία κατάτμησης και ελέγχθηκαν με το χέρι.

Με τη χρήση ενός τέτοιου αρχείου XML και του αντίστοιχου αρχείου εικόνας των γραμμών κειμένου και των απομονωμένων λέξεων, μπορούν να εξαχθούν οι απαραίτητες πληροφορίες που θα χρησιμοποιηθούν για να δημιουργήσουν τις εικόνες και να ελέγξουν την αυτόματη κατάτμηση όπως παρουσιάζεται παρακάτω:



a01-122-02: industry," Mr. Brown commented icily. "Let us have a

Εκτός από τις πληροφορίες κατάτμησης γραμμών και λέξεων, το αρχείο XML διαθέτει επίσης ποικίλες άλλες παραμέτρους. Για παράδειγμα η παρακάτω εικόνα παρουσιάζει τα παραπάνω στοιχεία συνδεδεμένα χρησιμοποιώντας οριοθετημένα πλαίσια, την κλίση της γραφής, καθώς και τις σαρωμένες γραμμές αναφοράς.

Unipen

Η έναρξη της κοινοπραξίας της UNIPEN άρχισε στο 11^{ο} διεθνές συνέδριο αναγνωρίσεων προτύπων (International Conference on Pattern Recognition) της IAPR (International Association for Pattern Recognition) και IEEE, το Σεπτέμβριο του 1992 όταν η πρώτη ανέλαβε να διερευνήσει τη δυνατότητα να δημιουργηθούν μεγάλες online βάσεις δεδομένων για έρευνα και την ανάπτυξη αναγνώρισης χειρόγραφων χαρακτήρων.

Το Μάρτιο του 1993, ένας πυρήνας των εμπειρογνωμόνων στη Online αναγνώριση χειρόγραφων χαρακτήρων, τόσο από μεγάλες επιχειρήσεις του χώρου, όσο και από πανεπιστήμια, έθεσε τα θεμέλια της UNIPEN. Σε συνεργασία με το Linguistic Data

Consortium και τη National Institute of Standards and Technologies (NIST) θα δημιουργούσαν ένα πρότυπο δεδομένων και συγκριτικής μέτρησης επιδόσεων (benchmark).

Η προσπάθεια αυτή ολοκληρώθηκε το Μάιο του 1993 και ελέγχθηκε ενδελεχώς τους επόμενους μήνες. Τα πρότυπα αυτά ελέγχθηκαν και χρησιμοποιήθηκαν από την AT&T και την NICI για τη συλλογή δεδομένων και τον έλεγχο συγκριτικής επίδοσης (benchmark) ήδη υπάρχοντος λογισμικού αναγνώρισης χειρόγραφων χαρακτήρων. Παράλληλα δημιουργήθηκε και λογισμικό από τη UNIPEN (UNIPEN software Tools), για την ανάλυση και επεξεργασία των δεδομένων. Η Unipen ανακοίνωσε την ίδρυσή της τον Μάρτιο του 1994, ενώ το 1999 μετονομάστηκε σε International Unipen Foundation (iUF).

Η Unipen εκτός από το πρότυπο δεδομένων που προσφέρει, και χρησιμοποιείται σαν standard από αρκετές βάσεις σήμερα, προσφέρει και δυνατότητες συγκριτικής μέτρησης επιδόσεων (benchmarking) για online βάσεις και εφαρμογές αναγνώρισης χειρόγραφων χαρακτήρων.

Η βάση δεδομένων της διαθέτει πάνω από 5 εκατομμύρια χαρακτήρες, πάνω από 2200 συγγραφείς, τα οποία έχουν δωριστεί στην κοινοπραξία από πάνω από 40 ιδρύματα και οργανισμούς. Η βάση της Unipen στην τελευταία της έκδοση χωρίζεται σε 10 κατηγορίες, όπως παρουσιάζονται στην παρακάτω εικόνα:

cat	nsegm	nfiles	
1a	15953	634	isolated digits
1b	28069	1423	isolated upper case
1c	61351	2145	isolated lower case
1d	17286	1222	isolated symbols (punctuations etc.)
2	122628	2735	isolated characters, mixed case
3	67352	1949	isolated characters in the context of words or texts
4	0	0	isolated printed words, not mixed with digits and symbols
5	0	0	isolated printed words, full character set
6	75529	3298	isolated cursive or mixed-style words (without digits and symbols)
7	85213	3393	isolated words, any style, full character set
8	14544	4563	text: (minimally two words of) free text, full character set

Οι κατηγορίες της Unipen Βάσης Δεδομένων.

Φόρμες εισαγωγής δεδομένων

Ιδιαίτερη αναφορά χρήζει και το γεγονός ότι το πρότυπο της Unipen έχει σχεδιαστεί για λήψη δεδομένων με ηλεκτρονική μορφή, δηλαδή με τη χρήση μιας ψηφιακής ταμπλέτας που καταγράφει την τροχιά και τις pen-up και pen-down συντεταγμένες της πένας.

Δυστυχώς εξαιτίας αυτού, αλλά και του γεγονότος ότι η Unipen έλαβε τα δεδομένα της από άλλους οργανισμούς, είναι αδύνατο να παρουσιάσουμε κάποια γενική φόρμα εισαγωγής δεδομένων.

CEDAR

Το CEDAR είναι ένα ερευνητικό κέντρο του πανεπιστημίου του Buffalo και του State University of New York. Το ερευνητικό κέντρο ξεκίνησε με την έρευνα σε αναγνώριση προτύπων το 1978 με την οικονομική υποστήριξη δημόσιων οργανισμών και κυρίως του US Postal Service. Το CEDAR ασχολείται με την ανάπτυξη λογισμικού σε διαφορετικούς ψηφιακούς τύπους εγγράφων, αλλά ειδικεύεται στην αναγνώριση γραφικού χαρακτήρα. Ιδιαίτερο ενδιαφέρον παρουσιάζει η βάση δεδομένων του CEDAR.

Η σε-απευθείας (online) σύνδεση χειρόγραφη βάση δεδομένων κειμένων CEDAR είναι μια βάση δεδομένων που αποτελείται από γραμμές κειμένου, χειρόγραφες σε μια ταμπλέτα γραψίματος από περίπου 200 συγγραφείς, και αποθηκευμένες σε ψηφιακή μορφή. Ο συνολικός αριθμός στοιχείων που περιλαμβάνονται στη βάση δεδομένων είναι 105.573. Η βάση δεδομένων περιέχει δείγματα από ελεύθερο και μεμονωμένο κείμενο, καθώς επίσης και δείγματα τα όποια είναι ένα μίγμα ελεύθερου και μεμονωμένου κειμένου. Επειδή περιέχει ολόκληρες γραμμές κειμένου, αντί για δείγματα μεμονωμένων λέξεων, η βάση δεδομένων μπορεί να χρησιμεύσει για να μελετηθεί ο χωρισμός λέξης και η αναγνώριση λέξεων από τα συμφραζόμενα, εκτός από την γενική αυτόματη αναγνώριση λέξεων. Επιπλέον, έχει γίνει συλλογή δειγμάτων ονόματος και διευθύνσεων χρησιμοποιώντας μια ενσωματωμένη συσκευή γραψίματος/επίδειξης (το Wacom PL100V). Οι διευθύνσεις περιλαμβάνουν αλφαριθμητικά (ονόματα οδών, πόλεις) και αριθμητικά (Ταχυδρομικούς Κώδικες) στοιχεία. Τα στοιχεία έχουν συλλεχθεί σε δύο φάσεις και χωρίζονται σε έξι υποκατηγορίες. Ο πίνακας 1 συνοψίζει το περιεχόμενο των στοιχείων σε έξι υποκατηγορίες και δύο συσκευές.

Device	Opaque Tablet			Integrated LCD Writing/Display Device			Total
Directory	ced0	ced1	ced2	ced3	ced4	ced5	
# of writers	22	73	125	132	95	95	NA
# of sentences	347	1,439	4,327	NA	NA	NA	6,113
# of words	3,348	14,013	41,252	5,808*	27,012	14,140	105,573
# of characters	13,734	57,835	169,524	48,576	206,650	57,453	553,772
# of truthed char.	13,734	57,835	0	48,576	0	0	120,145

Χαρακτηριστικά στοιχείων της Online Βάσης Δεδομένων CEDAR

Δύο συσκευές χρησιμοποιήθηκαν για τη συλλογή των στοιχείων. Η πρώτη είναι μια ψηφιακή ταμπλέτα, διασυνδεδεμένη με ένα σταθμό εργασίας SUN. Με βάση την προσεκτική εκτίμηση διάφορων κριτηρίων, θεωρήθηκε ότι μια ψηφιακή ταμπλέτα από κοινού με μια πέννα μελανιού θα παρήγαγε δείγματα γραφής περισσότερο «φυσικά». Η δεύτερη συσκευή είναι μια ενσωματωμένη συσκευή γραψίματος/επίδειξης, παρόμοια με εκείνους που χρησιμοποιούνται στους προσωπικούς ψηφιακούς βοηθούς (PDA).

Αυτή η βάση δεδομένων προσαρμόζεται στα πρότυπα και τις οδηγίες που περιγράφονται στο "First UNIPEN Benchmark of On-line Handwriting Recognizers", που παρουσιάσθηκε τον Ιούνιο του 2004, και το οποίο αναμένετε να χρησιμοποιηθεί εκτενώς από τους ερευνητές που αναπτύσσουν online συστήματα αναγνώρισης χαρακτήρων. Το online στοιχείο αποθηκεύεται σε "UNIPEN 1,0 format" και περιλαμβάνει τα αρχεία στοιχείων, τα αρχεία τεκμηρίωσης και τα λεξικά (lexicons). Όλες οι υποχρεωτικές πληροφορίες (όπως διευκρινίζεται στο πρότυπο UNIPEN) σχετικά με τη συσκευή και το συγγραφέα έχουν παρασχεθεί. Το τελευταίο καλύφθηκε με την συμπλήρωση ερωτηματολογίων από τους συγγραφείς. Η ακεραιότητα της μορφοποίησης στοιχείων έχει επικυρωθεί με τη χρησιμοποίηση των ελεγκτών σύνταξης που παρέχονται από την κοινοπραξία UNIPEN. Ημιαυτόματες και γραφικές διεπαφές χρησιμοποιήθηκαν για τον έλεγχο της ακεραιότητας των δεδομένων, και επιπλέον ελέγχθηκαν και από ανθρώπους.

Οι συγγραφείς για το πείραμα έχουν επιλεχτεί και από το SUNY/Buffalo και το γενικό πληθυσμό. Έγιναν ανακοινώσεις και στη συνέχεια οι εθελοντές κλήθηκαν να δώσουν μια συνοπτική εξέταση ικανότητας (για να δουν αν πληρούν τα ελάχιστα επίπεδα γραψίματος, ορθογραφίας, και ακουστικής δεξιότητας) για να καθορίσουν εάν ήταν κατάλληλοι για το πείραμα.

Φόρμες εισαγωγής Δεδομένων:
Ιδιαίτερο ενδιαφέρον παρουσιάζει ο τρόπος με τον οποίο έγινε η εισαγωγή δεδομένων στη βάση δεδομένων. Δεν έγινε μέσω ανίχνευσης χαρακτήρων από την επιφάνεια χαρτιού (scanning), αλλά χρησιμοποιώντας ψηφιακή ταμπλέτα. Ιδιαίτερο ενδιαφέρον χρήζουν και τα δείγματα που επιλέχθηκαν.

Οι λέξεις είναι κοινές αγγλικές λέξεις μαζί με τα κατάλληλα ουσιαστικά. Οι πιο συχνά χρησιμοποιημένες λέξεις (top 100) της αγγλικής γλώσσας αντιπροσωπεύονται επαρκώς. Αν και καταβλήθηκαν προσπάθειες να συμπεριληφθούν όσο το δυνατόν περισσότεροι χαρακτήρων που εμφανίζονται με χαμηλή συχνότητα, υπάρχει ακόμα μια δυσανάλογη αναλογία μεταξύ των χαρακτήρων που χρησιμοποιούνται πολύ συχνά και αυτών που χρησιμοποιούνται ελάχιστα.

Οι ακόλουθοι ASCII χαρακτήρες αντιπροσωπεύονται στα κείμενα: Κεφαλαία και μικρά γράμματα της αλφαβήτου, τα ψηφία από το 0 μέχρι και το 9, και σημεία στίξης συμπεριλαμβανομένων και των: ``. ,: $%"()" - ".

NIST

Τα αρχικά NIST National Institute of Standards and Technology των ΗΠΑ και αποτελεί ένα οργανισμό που ασχολείται με τη δημιουργία προτύπων σε όλους τους τομείς της τεχνολογίας. Ιδιαίτερο ενδιαφέρον παρουσιάζει η βάση NIST Special Database 19, η οποία ασχολείται με την αναγνώριση χειρόγραφων χαρακτήρων.

Η ειδική βάση δεδομένων 19 της NIST περιέχει το σύνολο των δεδομένων της NIST σε ότι αφορά την αναγνώριση χειρόγραφων χαρακτήρων και κειμένου. Εμπεριέχει ένα δείγμα 3.600 συγγραφέων, με 810.000 εικόνες χαρακτήρων απομονωμένες από την αρχική τους μορφή, με αληθή ταυτοποίηση για την κατηγοριοποίησή τους και λογισμικό για τη διαχείριση εικόνων. Η έκδοση αυτή καταργεί τις ειδικές βάσεις δεδομένων 1, 3 και 7 που ασχολούνταν με το ίδιο θέμα.

Η ειδική βάση δεδομένων 19 περιέχει, στα τμήματα hsf_0 έως hsf_3, τα δοκιμαστικά δεδομένα για το First Census Optical Character Recognition Conference, που προωθήθηκε από την NIST και το Bureau of the Census. Τα δεδομένα αυτά είχαν αρχικά δοθεί στο κοινό ως η ειδική βάση δεδομένων 3. Το τμήμα hsf_4 περιέχει τα δοκιμαστικά δεδομένα δοκιμής επίδοσης (benchmark test data) που χρησιμοποιήθηκαν στο συνέδριο και είχαν δοθεί στο κοινό ως η ειδική βάση δεδομένων 7.

Τα αποτελέσματα της διάσκεψης δημοσιεύτηκαν στην NIST Internal Report (IR) 4912. Συνοπτικά η ειδική βάση δεδομένων 19 περιέχει:

- Το τελικό και συνολικό δείγμα χειρόγραφων χαρακτήρων της NIST.
- Πλήρεις σελίδες HSF προτύπων από 3.600 συγγραφείς
- Ξεχωριστά πεδία με κεφαλαίους, πεζούς χαρακτήρες, ψηφία, και ελεύθερο κείμενο.
- Πάνω από 800.000 εικόνες με ελεγμένες από το χέρι ταξινομήσεις.
- Δοκιμαστικά και δεδομένα εκπαίδευσης OCR συστημάτων από το First Census OCR Systems Conference.
- Δυαδικές εικόνες scanned στα 11,8 dots per mm (300 dpi)
- Ενημερωμένο CCITT IV πηγαίο κώδικα συμπίεσης.
- Προγράμματα διαχείρισης βάσης δεδομένων.
- Οι εικόνες τις Ειδικής Βάσης Δεδομένων 19 αποτελούν ένα υπερσύνολο των εικόνων από τις δύο προηγούμενες εκδόσεις: Τις ειδικές βάσεις δεδομένων 1, 3 και 7 που διακόπηκαν.

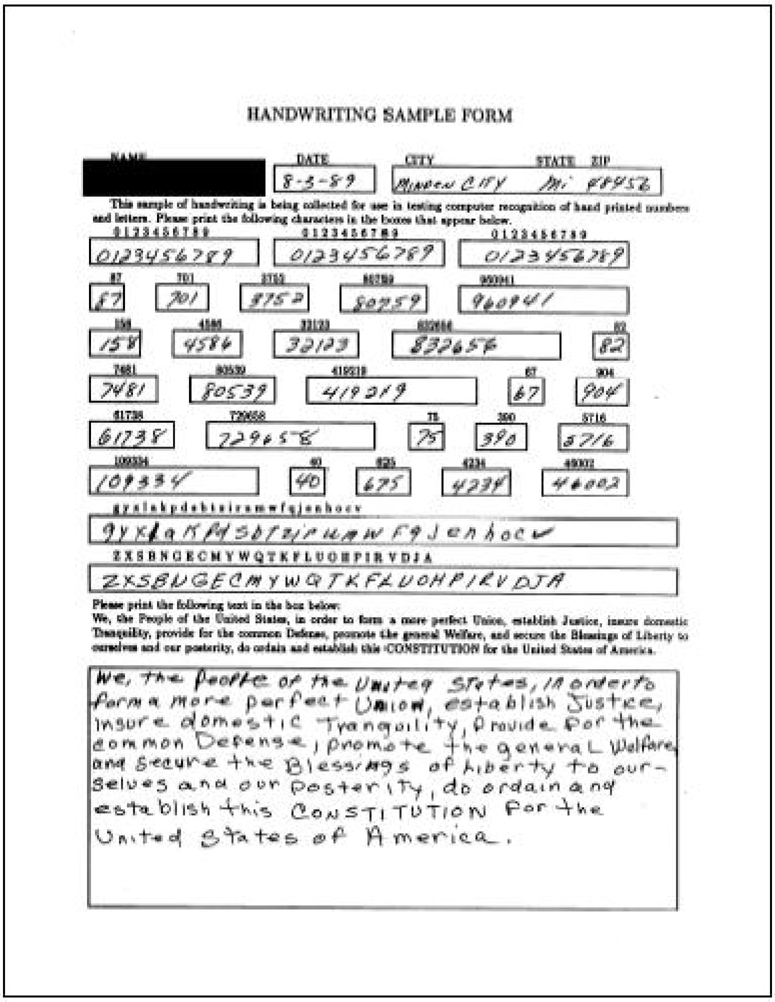
Οι βάση δεδομένων αυτή της NIST έχει ως στόχο την επεξεργασία εγγράφων με χειρόγραφους χαρακτήρες, αλλά και την εκμάθηση OCR εφαρμογών.

Φόρμες Εισαγωγής Δεδομένων:

Η ειδική βάση δεδομένων της NIST έχει μια φόρμα εισαγωγής δεδομένων που συμπληρώνετε από το χρήστη χρησιμοποιώντας πένα. Ο χρήστης καλείται να εισάγει στοιχεία σε πλαίσια. Πρέπει να συμπληρώσει ημερομηνία, πόλη, ταχυδρομικό κώδικα και το όνομά του σε πρώτη φάση. Έπειτα ακολουθεί η εισαγωγή κυρίως αριθμητικών

ψηφίων, τόσο σε ομάδες αριθμών, όσο και μεμονωμένα. Η εισαγωγή των μεμονωμένων ψηφίων από το 1 έως 9 γίνετε 3 φορές για να διασφαλιστεί καλύτερο δείγμα. Έπειτα ακολουθούν ομάδες αριθμών ώστε να υπάρχει δείγμα από ελεύθερο αριθμητικό χαρακτήρα του συγγραφέα.

Μετά τα αριθμητικά ψηφία ο χρήστης εισάγει 2 φορές μεμονωμένους χαρακτήρες και έπειτα το ελεύθερο κείμενο. Πρέπει να τονιστεί ότι η NIST προσφέρει κατάτμηση μόνο χαρακτήρες και όχι σε επίπεδο λέξεων.



HANDWRITING SAMPLE FORM

[redacted] DATE CITY STATE ZIP

This sample of handwriting is being collected for use in testing computer recognition of hand printed numbers and letters. Please print the following characters in the boxes that appear below.

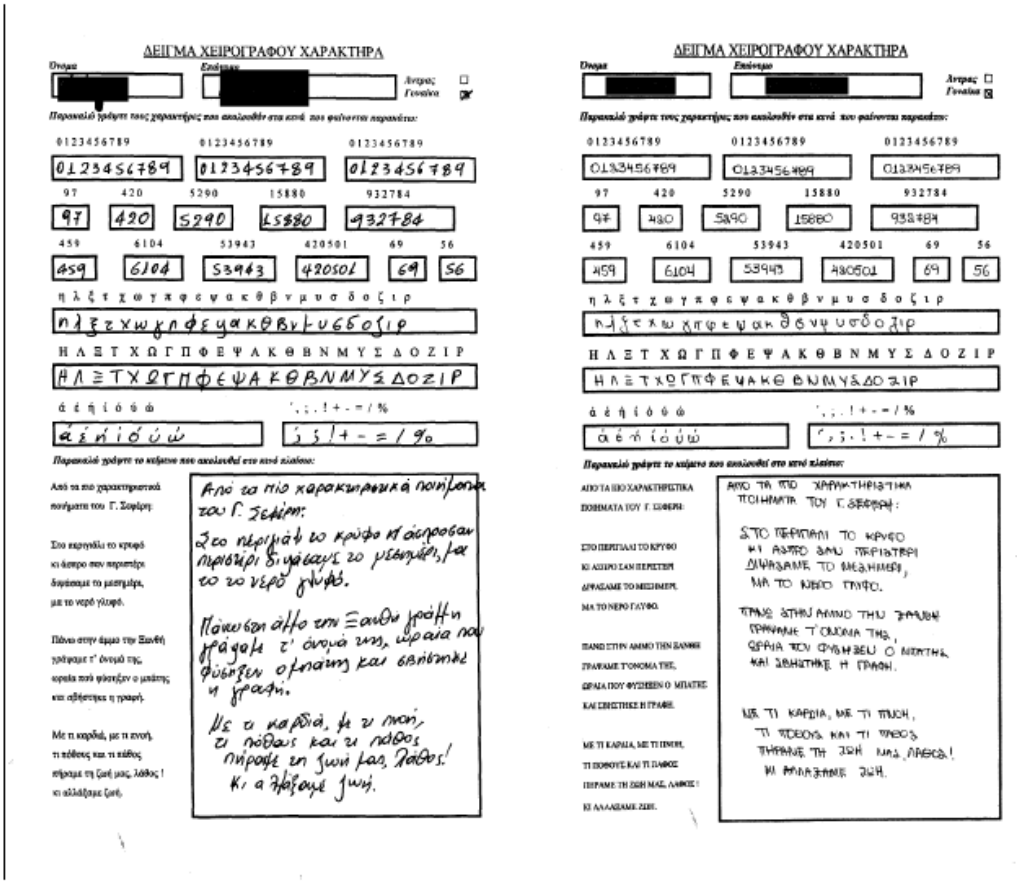
Please print the following text in the box below:
We, the People of the United States, in order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common Defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our posterity, do ordain and establish this CONSTITUTION for the United States of America.

Η φόρμα εισαγωγής στοιχείων της NIST

GRUHD

Είναι η μοναδική έως τώρα ελληνική βάση δεδομένων η οποία περιέχει ελληνικούς χαρακτήρες. Η βάση περιέχει 1,760 φόρμες συμπληρωμένες από 1,000 συγγραφείς περιέχοντας 667,583 σύμβολα και 102,692 λέξεις στο σύνολο. Η βάση συνοδεύεται από λογισμικό το οποίο δίνει τη δυνατότητα στους χρήστες να εξάγουν τα στοιχεία που θέλουν ανάλογα με τα κριτήρια που θέτουν στη βάση. Αυτό υποβοηθάτε από τα στατιστικά στοιχεία των συγγραφέων που έχουν καταγραφεί.

Η φόρμα που χρησιμοποιήθηκε είναι παρόμοια με της NIST. Ο χρήστης κι εδώ καλείται να εισάγει στοιχεία σε πλαίσια. Αρχικά εισάγει στοιχεία όπως όνομα, επώνυμο, φύλο. Στη συμπληρώνει σε πεδία σύμβολα όπως αριθμούς και γράμματα. Τέλος υπάρχει ένα πεδίο για την συμπλήρωση ελεύθερου κειμένου.



Η φόρμα εισαγωγής στοιχείων της GRUHD

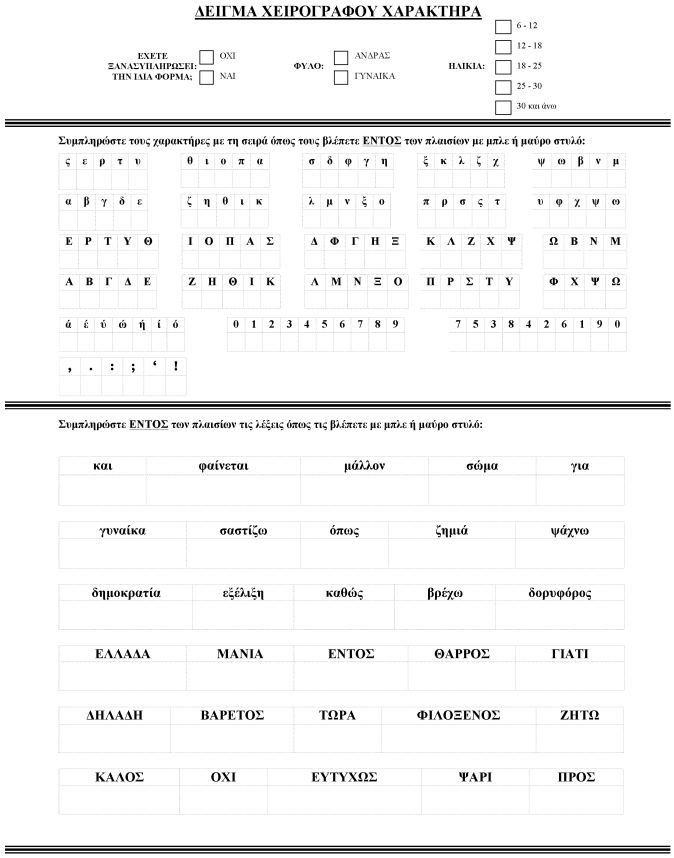
Υπάρχει ακόμα μια πληθώρα ΒΔ χειρόγραφων χαρακτήρων για διάφορες γλώσσες πέρα των λατινικών. Ιδιαίτερα για Kanji χαρακτήρες όπως η JAIST IIPL, η οποία ειδικεύεται σε ιαπωνικούς και κινέζικους χαρακτήρες ενώ έχει και προεκτάσεις για χειρόγραφους χαρακτήρες ατόμων με οπτικές δυσκολίες. Δεν γίνεται εκτενής αναφορά των βάσεων αυτών χωρίς αυτό να σημαίνει ότι δεν καλύπτουν ανάγκες όπως αυτές των λατινικών.

1.2 Φόρμα δείγματος χειρόγραφου χαρακτήρα της GCDB

1.2.1 Εισαγωγή:

Για την εισαγωγή δεδομένων στη βάση είναι απαραίτητο να σχεδιάσουμε μια φόρμα την οποία θα συμπληρώνουν οι συγγραφείς και από την οποία θα λαμβάνουμε το δείγμα του γραφικού τους χαρακτήρα. Οι χαρακτήρες αυτοί, μέσω μια ηλεκτρονικής συσκευής σάρωσης (scanner), θα αποθηκεύονται ηλεκτρονικά στη βάση δεδομένων και από εκεί θα είναι δυνατή η επεξεργασία αλλά και αξιοποίηση των δεδομένων αυτών μέσω μίας εφαρμογής client.

Ο σχεδιασμός της φόρμας εισαγωγής αποτελεί ένα καίριο σημείο της εργασίας αυτής, αφού θα καθορίσει ουσιαστικά τι είδους δεδομένα θα έχουμε στη βάση μας. Για το σχεδιασμό της φόρμας μας εμπνευστήκαμε από τις φόρμες εισαγωγής δεδομένων της NIST, ενώ χρησιμοποιήσαμε και αρκετές ιδέες από το είδος των δεδομένων που λάμβανε η CEDAR. Η παρούσα έκδοση της φόρμας δείγματος χειρόγραφου χαρακτήρα είναι η δεύτερη που σχεδιάζουμε. Το πρώτο μέρος περιέχει πληροφορίες, τις οποίες καλείται να συμπληρώσει ο συγγραφέας, που αφορούν στατιστικά στοιχεία. Το δεύτερο μέρος περιέχει τα πλαίσια όπου οι συγγραφείς καλούνται να συμπληρώσουν χαρακτήρες (είτε γραμμάτων είτε αριθμών) και το τρίτο είναι το πλαίσιο που πρέπει να συμπληρωθούν λέξεις.



ΔΕΙΓΜΑ ΧΕΙΡΟΓΡΑΦΟΥ ΧΑΡΑΚΤΗΡΑ

ΕΧΕΤΕ ΞΑΝΑΣΥΠΑΡΞΕΙ ΤΗΝ ΙΔΙΑ ΦΟΡΜΑ: ☐ ΟΧΙ ☐ ΝΑΙ

ΦΥΛΟ: ☐ ΑΝΔΡΑΣ ☐ ΓΥΝΑΙΚΑ

ΗΛΙΚΙΑ: ☐ 6 - 12 ☐ 12 - 18 ☐ 18 - 25 ☐ 25 - 30 ☐ 30 και άνω

Συμπληρώστε τους χαρακτήρες με τη σειρά όπως τους βλέπετε ΕΝΤΟΣ των πλαισίων με μπλε ή μαύρο στυλό:

ς ε ρ τ υ θ ι ο π α σ δ φ γ η ξ κ λ ζ χ ψ ω β ν μ

α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ ς τ υ φ χ ψ ω

Ε Ρ Τ Υ Θ Ι Ο Π Α Σ Δ Φ Γ Η Ξ Κ Λ Ζ Χ Ψ Ω Β Ν Μ

Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω

ά έ ό ώ ή ί ό 0 1 2 3 4 5 6 7 8 9 7 5 3 8 4 2 6 1 9 0

, . : ; ‘ !

Συμπληρώστε ΕΝΤΟΣ των πλαισίων τις λέξεις όπως τις βλέπετε με μπλε ή μαύρο στυλό:

και	φαίνεται	μάλλον	σώμα	για

γυναίκα	σαστίζω	όπως	ζημιά	ψάχνω

δημοκρατία	εξέλιξη	καθώς	βρέχω	δορυφόρος

ΕΛΛΑΔΑ	ΜΑΝΙΑ	ΕΝΤΟΣ	ΘΑΡΡΟΣ	ΓΙΑΤΙ

ΔΗΛΑΔΗ	ΒΑΡΕΤΟΣ	ΤΩΡΑ	ΦΙΛΟΞΕΝΟΣ	ΖΗΤΩ

ΚΑΛΟΣ	ΟΧΙ	ΕΥΤΥΧΩΣ	ΨΑΡΙ	ΠΡΟΣ

Η φόρμα της GCDB

Τα τρία μέρη της φόρμας διαχωρίζονται από δύο έντονες γραμμές οι οποίες διασχίζουν σε μήκος τη φόρμα. Οι γραμμές πέρα από τον εμφανή λόγο του διαχωρισμού των μερών της φόρμας εξυπηρετούν και στην εξαγωγή και ψηφιοποίηση των δεδομένων στη φόρμα. Όπως προαναφέραμε αφού η φόρμα συμπληρωθεί από τους συγγραφείς, οι πληροφορίες και τα σύμβολα που εμπεριέχονται πάνω σε αυτή θα ψηφιοποιηθούν με τη βοήθεια ηλεκτρονικής συσκευής σάρωσης (scanner). Σε αυτό το σημείο έγκεινται και μια πληθώρα προβλημάτων. Τα προβλήματα αυτά μπορούν να χωριστούν σε 2 μεγάλες κατηγορίες:

1. Ψηφιοποίηση των δεδομένων χωρίς απώλειες και θόρυβο.
2. Αντιστοίχηση των συμβόλου που σαρώνεται με το αντίστοιχο χαρακτήρα.

Οι γραμμές αυτές επιτελούν στην επίλυση του δεύτερου προβλήματος. Με τη χρήση των γραμμών αυτών έχουμε 2 σταθερά σημεία αναφοράς στη φόρμα μας, βάση των οποίων μπορούμε να προσδιορίσουμε τη σχετική θέση των συμβόλων του συγγραφέα. Αυτό σημαίνει ότι σε περίπτωση κλίσης ή κύρτωσης κατά την εκτύπωση ή την φωτοαντιγραφισή της αρχικής φόρμας το πρόγραμμα θα εξάγει της συντεταγμένες που αναμένει τα σύμβολα του συγγραφέα βάση των γραμμών, αποφεύγοντας έτσι να σαρωθεί μια πιθανώς κενή ή εσφαλμένη περιοχή, λόγω κακής εκτύπωσης ή φωτοαντιγράφισης.

Η λύση στο πρώτο πρόβλημα επιτυγχάνεται με τη χρήση των πλαισίων στα οποία πρέπει να συμπληρώσουν οι συγγραφείς τα γράμματα και λέξεις που τους ζητούνται. Επειδή οι συγγραφείς πρέπει να γράφουν μέσα στα πλαίσια και χωρίς να τα τέμνουν έχουμε μια ακριβή εικόνα του που θα βρίσκονται τα δεδομένα που θέλουμε να ψηφιοποιήσουμε. Επίσης το χρώμα των πλαισίων θα τυπωθεί σε ανοικτό γκρίζο, ώστε όταν γίνεται η σάρωση να μην λαμβάνεται υπόψη το πλαίσιο.

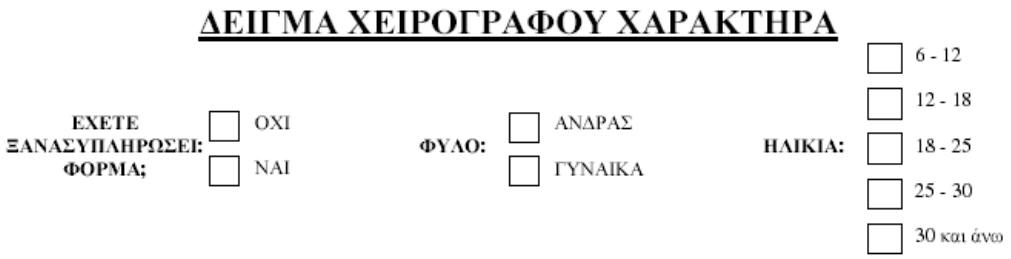
1.2.2 Πρώτο Τμήμα

Το πρώτο μέρος της φόρμας μας ζητά από το συγγραφέα να συμπληρώσει διάφορα προσωπικά στοιχεία, τα οποία επιθυμούμε να αποθηκεύσουμε στη βάση για στατιστικούς καθαρά λόγους. Τα στοιχεία αυτά αποτελούν κυρίως την ηλικία και το φύλο του συγγραφέα, καθώς αυτά είναι μια άμεση συνάρτηση του τρόπου γραφής του ανθρώπου. Αν και στις μέρες μας οι διαφορές στον τρόπο γραφής ανάμεσα στα δύο φύλα τείνουν να είναι πολύ μικρότερες από ότι στο παρελθόν, εξακολουθούμε να ζητάμε αυτήν την πληροφορία καθώς θα επιτρέπει καλύτερη επεξεργασία και αναζήτηση στα δεδομένα μας. Από την άλλη πλευρά όμως η ηλικία παίζει σημαντικό ρόλο στο τρόπο γραφής, καθώς άτομα σε μικρότερες ηλικίες δεν έχουν καταλήξει ακόμα στο ολοκληρωμένο και σταθερό γραφικό χαρακτήρα που έχουν οι ενήλικες. Συνεπώς η πληροφορία της ηλικίας κρίνεται απαραίτητη για τη βάση μας. Για την ηλικία δημιουργήσαμε πλαίσια με σύνολα ηλικιών. Έτσι έχουμε ένα πλαίσιο για 6 – 12 ετών, ένα για 12 – 18, ένα για 18 – 25, ένα για 25 – 30 και ένα για 30 και άνω. Ο λόγος που επιλέξαμε αυτά τα ηλικιακά πλαίσια είναι καθαρά για στατιστικούς λόγους.

Ένα ακόμα στοιχείο το οποίο επιθυμούσαμε να περιέχεται είναι το όνομα του συγγραφέα για να μπορέσουμε να επισυνάψουμε περισσότερα του ενός δείγματα στο ίδιο συγγραφέα σε περίπτωση που έχει συμπληρώσει ξανά τη φόρμα εισαγωγής. Στο συγκεκριμένο κομμάτι όμως αντιμετωπίσαμε μια πληθώρα προβλημάτων, τόσο λόγο του θεσμού προστασίας προσωπικών δεδομένων, όσο και από την αντίσταση των συγγραφέων. Θα μπορούσαμε να λάβουμε άδεια από την Αρχή Προστασίας

Προσωπικών Δεδομένων, την υπεύθυνη αρχή στην Ελλάδα, αλλά τελικά κρίναμε ότι η εισαγωγή του πεδίου του ονόματος θα προξενούσε περισσότερα προβλήματα παρά οφέλη. Σε αυτό συνέβαλε και η αρνητικότητα αρκετών συγγραφέων να συμπληρώσουν το όνομα τους σε ένα έγγραφο το οποίο πιθανώς να καταστήσει δυνατή την αναγνώριση του προσώπου τους από το γραφικό τους χαρακτήρα.

Για να επιλύσουμε το πρόβλημα που δημιουργήθηκε με την έλλειψη του ονόματος, δηλαδή κάποιος συγγραφέας να έχει συμπληρώσει και δεύτερη φόρμα, αρχικά αποφασίσαμε να περιλάβουμε και ένα αύξοντα αριθμό σε κάθε φόρμα και να ζητάμε από το συγγραφέα αν συμπληρώνει και δεύτερη φόρμα να εισάγει τον αύξοντα αριθμό της φόρμας που έχει ήδη συμπληρώσει. Η ιδέα απορρίφθηκε καθώς κρίθηκε ότι οι πλειονότητα, αν όχι όλοι οι συγγραφείς, δεν θα θυμούνται τον αριθμό αυτό, ενώ οποιοσδήποτε άλλος τρόπος να συνδέσουμε το συγγραφέα με τον αύξων αριθμό της φόρμας θα έπιπτε στα προαναφερθέντα προβλήματα. Κρίνοντας ότι ο αναμενόμενος αριθμός των συγγραφέων που θα συμπληρώσουν και δεύτερη φόρμα θα είναι σχετικά μικρός, αποφασίσαμε στη δεύτερη έκδοση να διατηρήσουμε μόνο το πλαίσιο όπου ρωτάμε το συγγραφέα αν έχει συμπληρώσει ξανά την ίδια φόρμα.

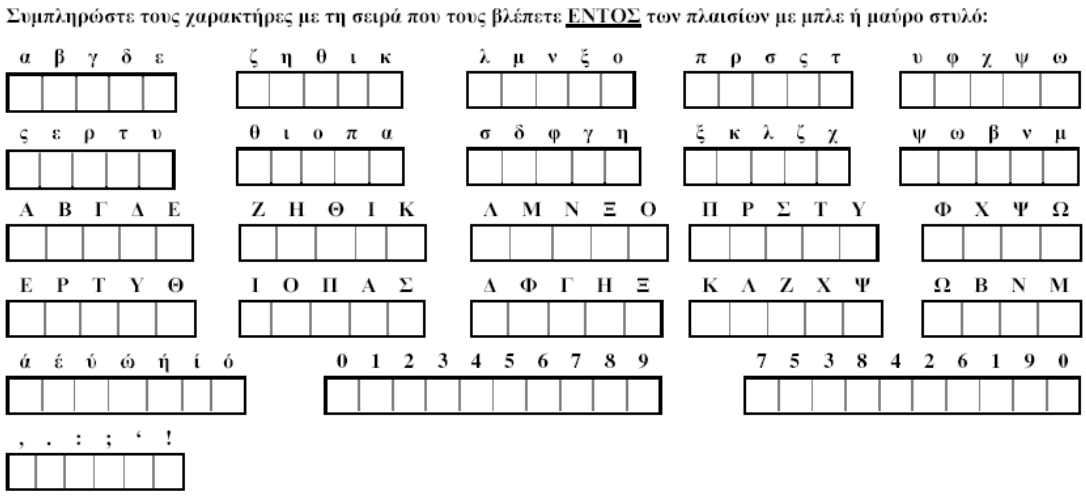


Το 1^{ο} μέρος της φόρμας.

Το πρώτο μέρος της φόρμας έχει ένα μικρό πλαίσιο, το όποιο πρέπει να συμπληρώσει είτε με "ν" είτε μαυρίζοντας την περιοχή του πλαισίου και το οποίο είναι το ερώτημα αν έχει συμπληρώσει ή όχι ίδια φόρμα στο παρελθόν. Με παρόμοια πλαίσια απαντάει και στις ερωτήσεις σχετικά με την ηλικία και το φύλο του ο συγγραφέας. Στην ηλικία υπάρχουν πολλαπλά πλαίσια με τα σύνολα ηλικιών που αναφέρονται παραπάνω. Σε όλα τα υπόλοιπα πλαίσια της φόρμας οι συγγραφείς πρέπει να γράψουν με έγχρωμη πένα μαύρου ή κυανού χρώματος.

1.2.3 Δεύτερο Τμήμα

Ακολουθεί το δεύτερο τμήμα της φόρμας, στο οποίο ο συγγραφέας καλείται να συμπληρώσει κενά πλαίσια με γράμματα και αριθμούς. Από αυτά η βάση μας θα αντλήσει ένα ιδιαίτερα σημαντικό τμήμα από τα δεδομένα της, τα δείγματα γραφικού χαρακτήρα σε επίπεδο συμβόλων. Σε οποιαδήποτε OCR (Optical Character Recognition – Αναγνώριση χειρόγραφων χαρακτήρων) εφαρμογή ή βάση εκμάθησης OCR εφαρμογών η αναγνώριση σε επίπεδο συμβόλου αποτελεί το θεμέλιο λίθο της εφαρμογής ώστε να υπάρξει και η δυνατότητα αναγνώρισης λέξεων. Πρέπει να τονίσουμε ότι στο τμήμα αυτό δώσαμε ιδιαίτερο βάρος στα γράμματα (κεφαλαία και πεζά) από ότι στους αριθμούς ή στα σημεία στίξης. Χαρακτηριστικό αυτού είναι και ο αριθμός των πλαισίων που προορίζονται για τα γράμματα σε σχέση με τον αριθμό των πλαισίων που προορίζονται για τους αριθμούς και τα σημεία στίξης, γεγονός που έγκειται στο εγγενή στόχο της βάσης μας.



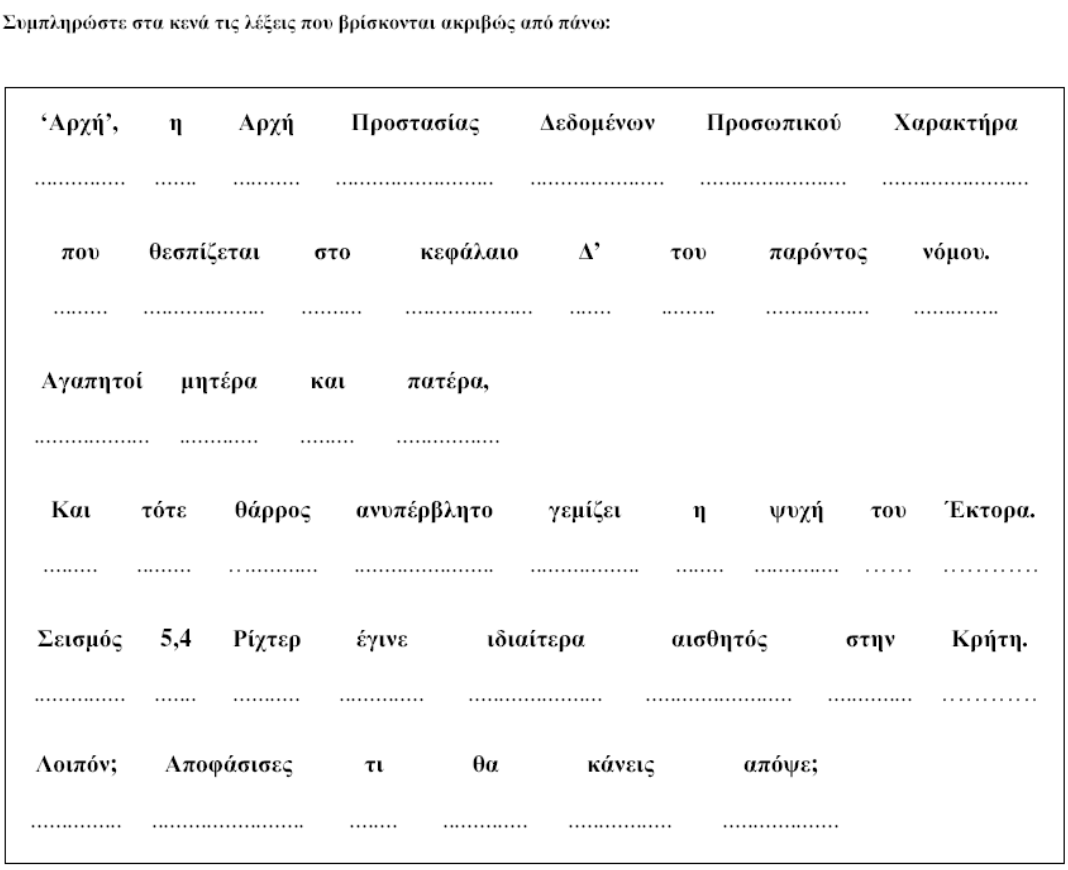
Το 2^{ο} μέρος της φόρμας.

Η βάση αυτή σχεδιάστηκε με σκοπό να καλύψει ένα κενό που παρατηρείται στον τομέα των OCR βάσεων εκμάθησης και εφαρμογών. Ενώ οι βάσεις και εφαρμογές για αριθμούς, λατινικούς και Kanji χαρακτήρες είναι ιδιαίτερα ανεπτυγμένη, δεν ισχύει κάτι αντίστοιχο για τους ελληνικούς χαρακτήρες. Η ανάγκη για μια τέτοια βάση κρίνεται ιδιαίτερα σημαντική εξαιτίας της ευρείας χρήσης των ελληνικών χαρακτήρων σε μαθηματικά και τεχνικά κείμενα. Έτσι και στη φόρμα μας δίνεται ιδιαίτερο βάρος στους χαρακτήρες αυτούς. Ιδιαίτερη αναφορά χρίει και το ότι η εισαγωγή των χαρακτήρων γίνεται δύο φόρες. Υπάρχουν 2 μεγάλες ομάδες για τους χαρακτήρες (τόσο πεζούς όσο και κεφαλαίους). Οι ομάδες αυτές χωρίζονται σε μικρότερες για τη διευκόλυνση του συγγραφέα και η κάθε ομάδα περιέχει πλαίσια των 5 γραμμάτων. Στη πρώτη μεγάλη ομάδα ο συγγραφέας εισάγει τα γράμματα με τη σειρά που συναντούνται στο ελληνικό αλφάβητο (α,β,γ,δ,ε κτλ), ενώ στη δεύτερη μεγάλη ομάδα τα γράμματα είναι ανακατεμένα. Ο λόγος για τον οποίο ζητείται από το συγγραφέα να συμπληρώσει δύο φόρες τα ίδια σύμβολα είναι το διαφορετικό είδος γραφής το οποίο συναντάται σε κάθε περίπτωση.

Παρατηρείται σε διάφορες φόρμες άλλων ΒΔ όπως της NIST ότι διαφέρει ο τρόπος με τον οποίο ο άνθρωπος γράφει ένα κείμενο, όταν αυτό το κείμενο είναι γνωστό και τετριμμένο σε αυτόν και όταν το κείμενο είναι άγνωστο. Όταν αντιμετωπίζουν ένα γνωστό και τετριμμένο κείμενο οι συγγραφείς (και η αλφάβητος αποτελεί το πιο χαρακτηριστικό παράδειγμα) η γραφή είναι πιο χαλαρή και ελεύθερη. Αντίθετα όταν έχουν άγνωστο κείμενο (αναδιαταγμένοι με τυχαία σειρά χαρακτήρες) τείνουν να δίνουν προσοχή στο κείμενο τους και ο τρόπος γραφής τους να είναι πιο αυστηρός και προσεγμένος. Επειδή επιζητούμε το καλύτερο δυνατό δείγμα στη φόρμα μας υπάρχουν και οι δύο κατηγορίες.

1.2.4 Τρίτο Τμήμα

Το τρίτο και τελευταίο τμήμα στοχεύει στην αποθήκευση λέξεων στη βάση μας. Σκοπός του τμήματος αυτού είναι η αποθήκευση ολόκληρων λέξεων στη βάση, η οποίες μπορούν να χρησιμοποιηθούν για την εκπαίδευση OCR εφαρμογών. Είναι γεγονός ότι ο τρόπος γραφής διαφέρει σημαντικά από ένα μεμονωμένο γράμμα, σε ένα γράμμα μέσα σε μία λέξη. Η ρευστότητα κατά τη συγγραφή μίας λέξης είναι πιθανό να προκαλέσει προβλήματα αναγνώρισης για κάποιο χαρακτήρα. Γι αυτό δείγματα λέξεων και χαρακτήρων μέσα σε λέξεις είναι απαραίτητα και ιδιαίτερα βοηθητικά.


Συμπληρώστε στα κενά τις λέξεις που βρίσκονται ακριβώς από πάνω:

'Αρχή',	η	Αρχή	Προστασίας	Δεδομένων	Προσωπικού	Χαρακτήρα
.....	.....	.....	.....	.....	.....	.....

που	θεσπίζεται	στο	κεφάλαιο	Δ'	του	παρόντος	νόμου.
.....	.....	.....	.....	.....	.....	.....	.....

Αγαπητοί	μητέρα	και	πατέρα,
.....	.....	.....	.....

Και	τότε	θάρρος	ανυπέρβλητο	γεμίζει	η	ψυχή	του	Έκτορα.
.....	.....	.....	.....	.....	.....	.....	.....	.....

Σεισμός	5,4	Ρίχτερ	έγινε	ιδιαίτερα	αισθητός	στην	Κρήτη.
.....	.....	.....	.....	.....	.....	.....	.....

Λοιπόν;	Αποφάσισες	τι	θα	κάνεις	απόψε;
.....	.....	.....	.....	.....	.....

Το 3^{ο} μέρος της φόρμας - αρχικά.

Αρχικά σκεφτήκαμε να χρησιμοποιήσουμε μικρές προτάσεις και διακεκομμένες γραμμές ανά λέξη για να συμπληρώσει ο συγγραφέας, όπως παρουσιάζεται στη παραπάνω εικόνα. Οι γραμμές αυτές λειτουργούσαν σαν οδηγός για το συγγραφέα για να τον υποβοηθήσουν στην ευθυγράμμιση της λέξεις και να μας δώσουν τη δυνατότητα να υπολογίσουμε περίπου το σημείο όπου θα σαρώσουμε (scan) τη λέξη. Είχαμε επιλέξει ολόκληρες προτάσεις για να συμπληρώσει ο συγγραφέας, σε αντίθεση με τη τελική φόρμα που έχουμε μόνο λέξεις. Η επιλογή των προτάσεων έγινε για να έχουμε πιο φυσικό κείμενο και δείγμα γραφής.

Οι προτάσεις που είχαμε επιλέξει ήταν 5 συνολικά. Πιο συγκεκριμένα επιλέξαμε μια πρόταση από τις εξής κατηγορίες:

1. Νομικό Έγγραφο.
2. Ανεπίσημη επιστολή ηλεκτρονικού ταχυδρομείου.

3. Λογοτεχνικό κείμενο.
4. Δελτίο τύπου εφημερίδας.
5. Καθημερινή συζήτηση.

Συμπληρώστε ΕΝΤΟΣ των πλαισίων τις λέξεις που βρίσκονται ακριβώς από πάνω με μπλε ή μάυρο στυλό:

και	να	αλλά	ως	για

αυτό	από	όμως	μέσα	προς

ναι	όχι	καθώς	βρέχω	δηλαδή

Ελλάδα	λοιπόν	εντός	ζημιά	ξέρω

σαστίζω	τώρα	σύμφωνα	φαίνεται	ψάχνω

συνεπώς	εξέλιξη	γυναίκα	δημοκρατία	σώμα

Το 3^{ο} μέρος της φόρμας - τελικά.

Στη τελική φόρμα ο στόχος μας σε ότι αφορά τις λέξεις μετατοπίστηκε από τη φυσικότητα του κειμένου στην πρακτικότητα και τη ευκολία διεξαγωγής αλλαγών στις λέξεις. Χαρακτηριστικά της δεύτερης έκδοσης είναι πλέον δεν υπάρχουν γραμμές οδηγοί, αλλά πλαίσια παρόμοια με αυτά του δεύτερου τμήματος. Επιπλέον δεν υπάρχουν αυτοτελείς προτάσεις, αλλά ένα σύνολο διαφορετικών λέξεων. Ο λόγος για την αλλαγή αυτή είναι ότι επιζητούσαμε να έχουμε λέξεις με συγκεκριμένο αριθμό γραμμάτων σε συγκεκριμένα σημεία. Πχ θέλαμε η πρώτη λέξη να έχει 2 γράμματα, ή η δεύτερη λέξη 3. Στόχος μας με την αλλαγή αυτή είναι η γρήγορη και εύκολη αντικατάσταση των λέξεών μας για την πιθανή μελλοντική αναναίωση της ΒΔ, χωρίς να χρειάζονται πολλές αλλαγές στη φόρμα ή το πρόγραμμα εισαγωγής και ψηφιοποίησης των δεδομένων.

Οι λέξεις που έχουμε επιλέξει είναι τέτοιες ώστε να βρίσκονται μέσα στις πιο συχνά χρησιμοποιούμενες λέξεις τις ελληνικής γλώσσας, χωρίς αυτό να σημαίνει ότι δεν υπάρχουν και λέξεις που συναντώνται με μικρή συχνότητα στη ελληνική γλώσσα. Επίσης περιέχονται όλα τα γράμματα του ελληνικού αλφάβητου στις λέξεις αυτές, ώστε να έχουμε έγκυρο και ενημερωμένο δείγμα.

Παρατηρούμε ότι σε σχέση με τη φόρμα της GRUHD υπάρχουν σημαντικές βελτιώσεις:

- Εξασφάλιση του απορρήτου προσωπικών δεδομένων, καθώς δεν αναφέρονται πουθενά προσωπικά στοιχεία όπως το ονοματεπώνυμο συγγραφέα.

- Χρήση ξεχωριστών πλαισίων για κάθε γράμμα/λέξη που αποσκοπεί στη δημιουργία μιας αυτόματης διαδικασίας καταχώρησης τους στη ΒΔ (περισσότερες λεπτομέρειες στο κεφάλαιο 4).
- Το αλφάβητο με σειρά και ανακατεμένο καθώς και λέξεις που περιέχουν όλα τα γράμματα αυτού εξασφαλίζοντας έτσι ένα αξιόπιστο δείγμα γραφής.

2. ΟΡΓΑΝΩΣΗ ΚΑΙ ΣΧΕΔΙΑΣΜΟΣ ΤΗΣ GCDB (Greek Characters DataBase)

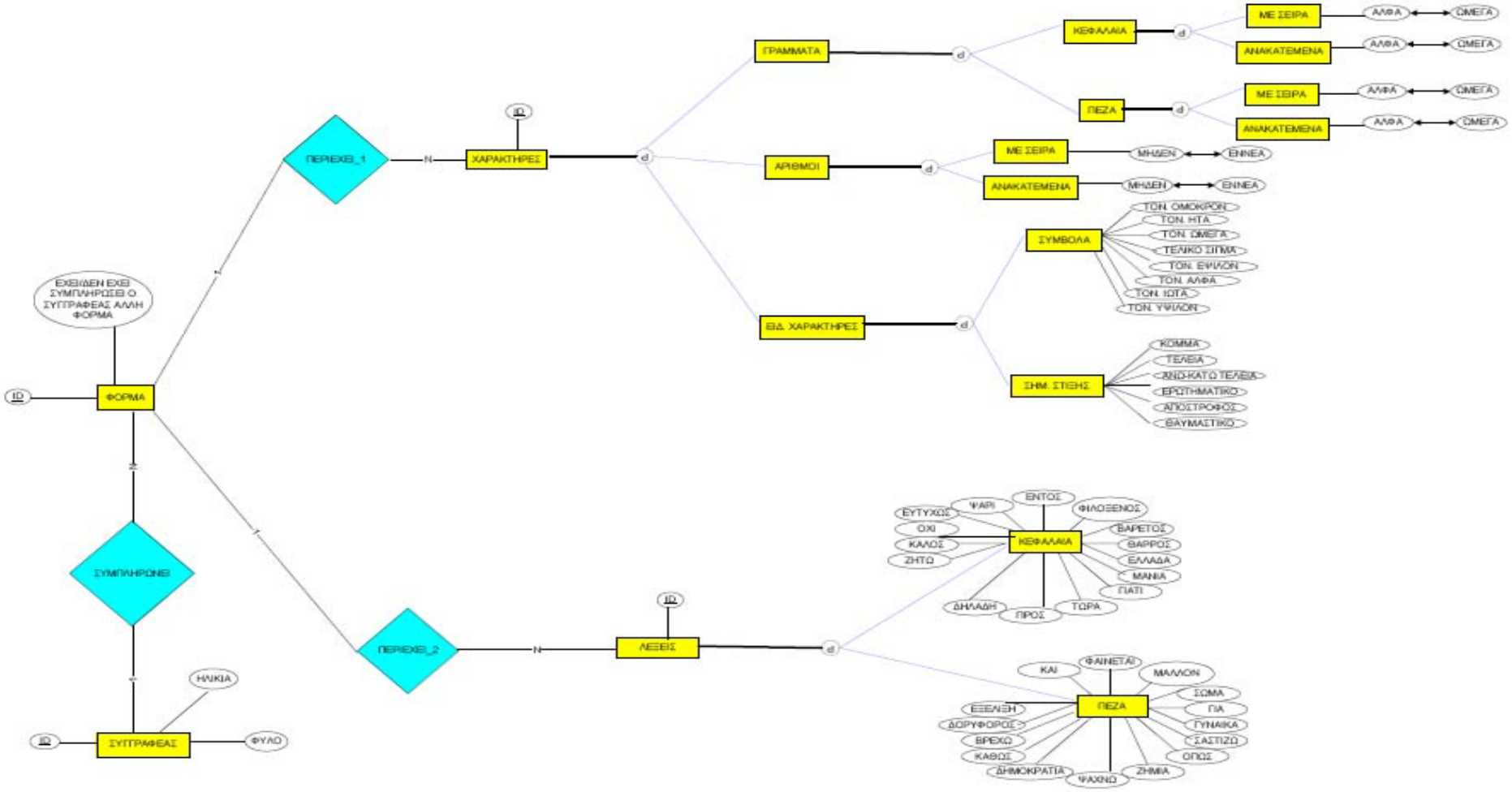
Η οργάνωση και ο σχεδιασμός της βάσης στηρίζονται στη φόρμα συλλογής χαρακτήρων και στα δεδομένα που μπορούμε να καταχωρήσουμε από αυτή.

Η φόρμα χωρίζεται σε τρεις περιοχές:

1. Η $1^{η}$ περιοχή αφορά στοιχεία του συγγραφέα όπως αν έχει συμπληρώσει ή όχι παρόμοια φόρμα, το φύλο και την ηλικία.
2. Η $2^{η}$ περιοχή περιέχει τα δεδομένα που μας ενδιαφέρουν. Δηλαδή τους όλους τους χαρακτήρες του αλφάβητου, τους αριθμούς, τους τονισμένους χαρακτήρες και τα σημεία στίξης.
3. Τέλος η $3^{η}$ περιοχή περιέχει λέξεις σε κεφαλαία και πεζά. Η επιλογή αυτών των λέξεων έγινε με κριτήριο την συχνότητα εμφάνισης τους στην ελληνική γλώσσα και την κάλυψη όλων των χαρακτήρων αυτής. Σκοπός μας είναι αργότερα από αυτές τις λέξεις να πάρουμε ένα ακόμα πιο αξιόπιστο δείγμα χειρόγραφων χαρακτήρων.

Οι οντότητες που χρειαζόμαστε για την οργάνωση αυτών των δεδομένων είναι ΣΥΓΓΡΑΦΕΑΣ, ΦΟΡΜΑ, ΧΑΡΑΚΤΗΡΕΣ και ΛΕΞΕΙΣ. Μελετάμε ξεχωριστά τις οντότητες ΣΥΓΓΡΑΦΕΑΣ και ΦΟΡΜΑ, παρόλο που από κάθε φόρμα παίρνουμε στοιχεία ενός συγγραφέα, γιατί μπορούμε να καταχωρίσουμε παραπάνω από μία φόρμες σε κάθε συγγραφέα. Όσο για την οντότητα ΛΕΞΕΙΣ θεωρούμε ότι τα δεδομένα της θα είναι ολόκληρες εικόνες λέξεων που χρειάζονται διαφορετική επεξεργασία από αυτές της οντότητας ΧΑΡΑΚΤΗΡΕΣ. Οι οντότητες και οι μεταξύ τους σχέσεις παρουσιάζονται στο παρακάτω διάγραμμα οντοτήτων-συσχετίσεων.

2.1 ΔΙΑΓΡΑΜΜΑ ΟΝΤΟΤΗΤΩΝ - ΣΥΣΧΕΤΙΣΕΩΝ



Οι οντότητες είναι οι εξής:

- ΣΥΓΓΡΑΦΕΑΣ περιέχει τα πεδία:
    - ID κλειδί της οντότητας που χαρακτηρίζει τη μοναδικότητα κάθε εγγραφής
    - ΗΛΙΚΙΑ είναι η πληροφορία για την ηλικία του συγγραφέα.
    - ΦΥΛΟ είναι η πληροφορία για το φύλο του συγγραφέα.
- ΦΟΡΜΑ περιέχει τα πεδία:
    - ID κλειδί της οντότητας που χαρακτηρίζει τη μοναδικότητα κάθε εγγραφής
    - ΕΧΕΙ/ΔΕΝ ΕΧΕΙ ΣΥΜΠΛΗΡΩΣΕΙ Ο ΣΥΓΓΡΑΦΕΑΣ ΑΛΛΗ ΦΟΡΜΑ
- ΧΑΡΑΚΤΗΡΕΣ περιέχει τα πεδία:
    - ID κλειδί της οντότητας που χαρακτηρίζει τη μοναδικότητα κάθε εγγραφής
    - Εξειδικεύεται σε:
        - ΓΡΑΜΜΑΤΑ (Είναι το δείγμα χαρακτήρων του ελληνικού αλφάβητου):
            - Εξειδικεύεται σε:
                - ΚΕΦΑΛΑΙΑ (Είναι το δείγμα κεφαλαίων χαρακτήρων του ελληνικού αλφάβητου):
                    - Εξειδικεύεται σε:
                        - ΜΕ ΣΕΙΡΑ (Είναι το δείγμα χαρακτήρων που συμπληρώθηκε με αλφαβητική σειρά) περιέχει τα πεδία:
                            - ΑΛΦΑ εικόνα του χειρόγραφου χαρακτήρα άλφα.
                            - ΒΗΤΑ εικόνα του χειρόγραφου χαρακτήρα βήτα.
                            - ΓΑΜΜΑ εικόνα του χειρόγραφου χαρακτήρα γάμμα.
                            - ΔΕΛΤΑ εικόνα του χειρόγραφου χαρακτήρα δέλτα.
                            - ΕΨΙΛΟΝ εικόνα του χειρόγραφου χαρακτήρα έψιλον.
                            - ΖΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ζήτα.
                            - ΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ήτα.
                            - ΘΗΤΑ εικόνα του χειρόγραφου χαρακτήρα θήτα.
                            - ΙΩΤΑ εικόνα του χειρόγραφου χαρακτήρα ιώτα.
                            - ΚΑΠΑ εικόνα του χειρόγραφου χαρακτήρα κάπα.
                            - ΛΑΜΔΑ εικόνα του χειρόγραφου χαρακτήρα λάμδα.

- ΜΙ εικόνα του χειρόγραφου χαρακτήρα μι.
- ΝΙ εικόνα του χειρόγραφου χαρακτήρα νι.
- ΞΙ εικόνα του χειρόγραφου χαρακτήρα ξι.
- ΟΜΙΚΡΟΝ εικόνα του χειρόγραφου χαρακτήρα όμικρον.
- ΠΙ εικόνα του χειρόγραφου χαρακτήρα πι.
- ΡΟ εικόνα του χειρόγραφου χαρακτήρα ρο.
- ΣΙΓΜΑ εικόνα του χειρόγραφου χαρακτήρα σίγμα.
- ΤΑΦ εικόνα του χειρόγραφου χαρακτήρα ταφ.
- ΥΨΙΛΟΝ εικόνα του χειρόγραφο χαρακτήρα ύψιλον.
- ΦΙ εικόνα του χειρόγραφου χαρακτήρα φι.
- ΧΙ εικόνα του χειρόγραφου χαρακτήρα χι.
- ΨΙ εικόνα του χειρόγραφου χαρακτήρα ψι.
- ΩΜΕΓΑ εικόνα του χειρόγραφου χαρακτήρα ωμέγα.

- ΑΝΑΚΑΤΕΜΕΝΑ (Είναι το δείγμα χαρακτήρων που συμπληρώθηκε με ανακατεμένη σειρά) περιέχει τα πεδία:
  - ΑΛΦΑ εικόνα του χειρόγραφου χαρακτήρα άλφα.
  - ΒΗΤΑ εικόνα του χειρόγραφου χαρακτήρα βήτα.
  - ΓΑΜΜΑ εικόνα του χειρόγραφου χαρακτήρα γάμμα.
  - ΔΕΛΤΑ εικόνα του χειρόγραφου χαρακτήρα δέλτα.
  - ΕΨΙΛΟΝ εικόνα του χειρόγραφου χαρακτήρα έψιλον.
  - ΖΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ζήτα.
  - ΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ήτα.
  - ΘΗΤΑ εικόνα του χειρόγραφου χαρακτήρα θήτα.
  - ΙΩΤΑ εικόνα του χειρόγραφου χαρακτήρα ιώτα.
  - ΚΑΠΑ εικόνα του χειρόγραφου χαρακτήρα κάπα.
  - ΛΑΜΔΑ εικόνα του χειρόγραφου χαρακτήρα λάμδα.
  - ΜΙ εικόνα του χειρόγραφου χαρακτήρα μι.
  - ΝΙ εικόνα του χειρόγραφου χαρακτήρα νι.
  - ΞΙ εικόνα του χειρόγραφου χαρακτήρα ξι.
  - ΟΜΙΚΡΟΝ εικόνα του χειρόγραφου χαρακτήρα όμικρον.

        - ΠΙ εικόνα του χειρόγραφου χαρακτήρα πι.
        - ΡΟ εικόνα του χειρόγραφου χαρακτήρα ρο.
        - ΣΙΓΜΑ εικόνα του χειρόγραφου χαρακτήρα σίγμα.
        - ΤΑΦ εικόνα του χειρόγραφου χαρακτήρα ταφ.
        - ΥΨΙΛΟΝ εικόνα του χειρόγραφο χαρακτήρα ύψιλον.
        - ΦΙ εικόνα του χειρόγραφου χαρακτήρα φι.
        - ΧΙ εικόνα του χειρόγραφου χαρακτήρα χι.
        - ΨΙ εικόνα του χειρόγραφου χαρακτήρα ψι.
        - ΩΜΕΓΑ εικόνα του χειρόγραφου χαρακτήρα ωμέγα.
- ΠΕΖΑ (Είναι το δείγμα πεζών χαρακτήρων του ελληνικού αλφάβητου):
    - Εξειδικεύεται σε:
        - ΜΕ ΣΕΙΡΑ (Είναι το δείγμα χαρακτήρων που συμπληρώθηκε με αλφαβητική σειρά) περιέχει τα πεδία:
            - ΑΛΦΑ εικόνα του χειρόγραφου χαρακτήρα άλφα.
            - ΒΗΤΑ εικόνα του χειρόγραφου χαρακτήρα βήτα.
            - ΓΑΜΜΑ εικόνα του χειρόγραφου χαρακτήρα γάμμα.
            - ΔΕΛΤΑ εικόνα του χειρόγραφου χαρακτήρα δέλτα.
            - ΕΨΙΛΟΝ εικόνα του χειρόγραφου χαρακτήρα έψιλον.
            - ΖΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ζήτα.
            - ΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ήτα.
            - ΘΗΤΑ εικόνα του χειρόγραφου χαρακτήρα θήτα.
            - ΙΩΤΑ εικόνα του χειρόγραφου χαρακτήρα ιώτα.
            - ΚΑΠΑ εικόνα του χειρόγραφου χαρακτήρα κάπα.
            - ΛΑΜΔΑ εικόνα του χειρόγραφου χαρακτήρα λάμδα.
            - ΜΙ εικόνα του χειρόγραφου χαρακτήρα μι.
            - ΝΙ εικόνα του χειρόγραφου χαρακτήρα νι.
            - ΞΙ εικόνα του χειρόγραφου χαρακτήρα ξι.
            - ΟΜΙΚΡΟΝ εικόνα του χειρόγραφου χαρακτήρα όμικρον.
            - ΠΙ εικόνα του χειρόγραφου χαρακτήρα πι.
            - ΡΟ εικόνα του χειρόγραφου χαρακτήρα ρο.

- ΣΙΓΜΑ εικόνα του χειρόγραφου χαρακτήρα σίγμα.
- ΤΑΦ εικόνα του χειρόγραφου χαρακτήρα ταφ.
- ΥΨΙΛΟΝ εικόνα του χειρόγραφο χαρακτήρα ύψιλον.
- ΦΙ εικόνα του χειρόγραφου χαρακτήρα φι.
- ΧΙ εικόνα του χειρόγραφου χαρακτήρα χι.
- ΨΙ εικόνα του χειρόγραφου χαρακτήρα ψι.
- ΩΜΕΓΑ εικόνα του χειρόγραφου χαρακτήρα ωμέγα.

- ΑΝΑΚΑΤΕΜΕΝΑ (Είναι το δείγμα χαρακτήρων που συμπληρώθηκε με ανακατεμένη σειρά) περιέχει τα πεδία:
  - ΑΛΦΑ εικόνα του χειρόγραφου χαρακτήρα άλφα.
  - ΒΗΤΑ εικόνα του χειρόγραφου χαρακτήρα βήτα.
  - ΓΑΜΜΑ εικόνα του χειρόγραφου χαρακτήρα γάμμα.
  - ΔΕΛΤΑ εικόνα του χειρόγραφου χαρακτήρα δέλτα.
  - ΕΨΙΛΟΝ εικόνα του χειρόγραφου χαρακτήρα έψιλον.
  - ΖΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ζήτα.
  - ΗΤΑ εικόνα του χειρόγραφου χαρακτήρα ήτα.
  - ΘΗΤΑ εικόνα του χειρόγραφου χαρακτήρα θήτα.
  - ΙΩΤΑ εικόνα του χειρόγραφου χαρακτήρα ιώτα.
  - ΚΑΠΑ εικόνα του χειρόγραφου χαρακτήρα κάπα.
  - ΛΑΜΔΑ εικόνα του χειρόγραφου χαρακτήρα λάμδα.
  - ΜΙ εικόνα του χειρόγραφου χαρακτήρα μι.
  - ΝΙ εικόνα του χειρόγραφου χαρακτήρα νι.
  - ΞΙ εικόνα του χειρόγραφου χαρακτήρα ξι.
  - ΟΜΙΚΡΟΝ εικόνα του χειρόγραφου χαρακτήρα όμικρον.
  - ΠΙ εικόνα του χειρόγραφου χαρακτήρα πι.
  - ΡΟ εικόνα του χειρόγραφου χαρακτήρα ρο.
  - ΣΙΓΜΑ εικόνα του χειρόγραφου χαρακτήρα σίγμα.
  - ΤΑΦ εικόνα του χειρόγραφου χαρακτήρα ταφ.
  - ΥΨΙΛΟΝ εικόνα του χειρόγραφο χαρακτήρα ύψιλον.
  - ΦΙ εικόνα του χειρόγραφου χαρακτήρα φι.

        - ΧΙ εικόνα του χειρόγραφου χαρακτήρα χι.
        - ΨΙ εικόνα του χειρόγραφου χαρακτήρα ψι.
        - ΩΜΕΓΑ εικόνα του χειρόγραφου χαρακτήρα ωμέγα.
- ΑΡΙΘΜΟΙ (Είναι το δείγμα αριθμητικών χαρακτήρων):
    - Εξειδικεύεται σε:
        - ΜΕ ΣΕΙΡΑ (Είναι το δείγμα χαρακτήρων που συμπληρώθηκε με τη σειρά 0-9) περιέχει τα πεδία:
            - ΜΗΔΕΝ εικόνα του χειρόγραφου χαρακτήρα μηδέν.
            - ΕΝΑ εικόνα του χειρόγραφου χαρακτήρα ένα.
            - ΔΥΟ εικόνα του χειρόγραφου χαρακτήρα δύο.
            - ΤΡΙΑ εικόνα του χειρόγραφου χαρακτήρα τρία.
            - ΤΕΣΣΕΡΑ εικόνα του χειρόγραφου χαρακτήρα τέσσερα.
            - ΠΕΝΤΕ εικόνα του χειρόγραφου χαρακτήρα πέντε.
            - ΕΞΙ εικόνα του χειρόγραφου χαρακτήρα έξι.
            - ΕΠΤΑ εικόνα του χειρόγραφου χαρακτήρα επτά.
            - ΟΚΤΩ εικόνα του χειρόγραφου χαρακτήρα οχτώ.
            - ΕΝΝΕΑ εικόνα του χειρόγραφου χαρακτήρα εννέα.
        - ΑΝΑΚΑΤΕΜΕΝΑ (Είναι το δείγμα χαρακτήρων που συμπληρώθηκε με ανακατεμένη σειρά) περιέχει τα πεδία:
            - ΜΗΔΕΝ εικόνα του χειρόγραφου χαρακτήρα μηδέν.
            - ΕΝΑ εικόνα του χειρόγραφου χαρακτήρα ένα.
            - ΔΥΟ εικόνα του χειρόγραφου χαρακτήρα δύο.
            - ΤΡΙΑ εικόνα του χειρόγραφου χαρακτήρα τρία.
            - ΤΕΣΣΕΡΑ εικόνα του χειρόγραφου χαρακτήρα τέσσερα.
            - ΠΕΝΤΕ εικόνα του χειρόγραφου χαρακτήρα πέντε.
            - ΕΞΙ εικόνα του χειρόγραφου χαρακτήρα έξι.
            - ΕΠΤΑ εικόνα του χειρόγραφου χαρακτήρα επτά.
            - ΟΚΤΩ εικόνα του χειρόγραφου χαρακτήρα οχτώ.
            - ΕΝΝΕΑ εικόνα του χειρόγραφου χαρακτήρα εννέα.
- ΕΙΔ. ΧΑΡΑΚΤΗΡΕΣ (Είναι το δείγμα ιδιαίτερων χαρακτήρων που περιέχονται μόνο στην ελληνική γλώσσα):

- Εξειδικεύεται σε:
    - ΣΥΜΒΟΛΑ (Είναι το δείγμα τονισμένων και ιδιαίτερων χαρακτήρων) περιέχει τα πεδία:
        - ΤΟΝ. ΟΜΙΚΡΟΝ εικόνα του χειρόγραφου χαρακτήρα του τονισμένου όμικρον.
        - ΤΟΝ. ΗΤΑ εικόνα του χειρόγραφου χαρακτήρα του τονισμένου ήτα.
        - ΤΟΝ. ΩΜΕΓΑ εικόνα του χειρόγραφου χαρακτήρα του τονισμένου ωμέγα.
        - ΤΟΝ. ΕΨΙΛΟΝ εικόνα του χειρόγραφου χαρακτήρα του τονισμένου έψιλον.
        - ΤΟΝ. ΑΛΦΑ εικόνα του χειρόγραφου χαρακτήρα του τονισμένου άλφα.
        - ΤΟΝ. ΙΩΤΑ εικόνα του χειρόγραφου χαρακτήρα του τονισμένου ιώτα.
        - ΤΕΛΙΚΟ ΣΙΓΜΑ εικόνα του χειρόγραφου χαρακτήρα του τελικού σίγμα.
    - ΣΗΜ. ΣΤΙΞΗΣ (Είναι το δείγμα χαρακτήρων των σημείων στίξης της ελληνικής γλώσσας) περιέχει τα πεδία:
        - ΚΟΜΜΑ εικόνα του χειρόγραφου χαρακτήρα κόμμα(,).
        - ΤΕΛΕΙΑ εικόνα του χειρόγραφου χαρακτήρα τελεία(.).
        - ΑΝΩ-ΚΑΤΩ ΤΕΛΕΙΑ εικόνα του χειρόγραφου χαρακτήρα άνω-κάτω τελεία(:).
        - ΕΡΩΤΗΜΑΤΙΚΟ εικόνα του χειρόγραφου χαρακτήρα ερωτηματικό(;).
        - ΑΠΟΣΤΡΟΦΟΣ εικόνα του χειρόγραφου χαρακτήρα απόστροφος(΄).
        - ΘΑΥΜΑΣΤΙΚΟ εικόνα του χειρόγραφου χαρακτήρα θαυμαστικό(!).
- ΛΕΞΕΙΣ περιέχει τα πεδία:
    - ID κλειδί της οντότητας που χαρακτηρίζει τη μοναδικότητα κάθε εγγραφής
    - Εξειδικεύεται σε:
        - ΚΕΦΑΛΑΙΑ (Είναι το δείγμα λέξεων με κεφαλαίους χαρακτήρες) περιέχει τα πεδία:
            - ΔΗΛΑΔΗ εικόνα της συγκεκριμένης λέξης.
            - ΠΡΟΣ εικόνα της συγκεκριμένης λέξης.
            - ΤΩΡΑ εικόνα της συγκεκριμένης λέξης.
            - ΓΙΑΤΙ εικόνα της συγκεκριμένης λέξης.
            - ΜΑΝΙΑ εικόνα της συγκεκριμένης λέξης.
            - ΕΛΛΑΔΑ εικόνα της συγκεκριμένης λέξης.
            - ΘΑΡΡΟΣ εικόνα της συγκεκριμένης λέξης.
            - ΒΑΡΕΤΟΣ εικόνα της συγκεκριμένης λέξης.

- ΦΙΛΟΞΕΝΟΣ εικόνα της συγκεκριμένης λέξης.
- ΕΝΤΟΣ εικόνα της συγκεκριμένης λέξης.
- ΨΑΡΙ εικόνα της συγκεκριμένης λέξης.
- ΕΥΤΥΧΩΣ εικόνα της συγκεκριμένης λέξης.
- ΟΧΙ εικόνα της συγκεκριμένης λέξης.
- ΚΑΛΟΣ εικόνα της συγκεκριμένης λέξης.
- ΖΗΤΩ εικόνα της συγκεκριμένης λέξης.

▪ ΠΕΖΑ (Είναι το δείγμα λέξεων με πεζούς χαρακτήρες) περιέχει τα πεδία:

- ΚΑΙ εικόνα της συγκεκριμένης λέξης.
- ΦΑΙΝΕΤΑΙ εικόνα της συγκεκριμένης λέξης.
- ΜΑΛΛΟΝ εικόνα της συγκεκριμένης λέξης.
- ΣΩΜΑ εικόνα της συγκεκριμένης λέξης.
- ΓΙΑ εικόνα της συγκεκριμένης λέξης.
- ΓΥΝΑΙΚΑ εικόνα της συγκεκριμένης λέξης.
- ΣΑΣΤΙΖΩ εικόνα της συγκεκριμένης λέξης.
- ΟΠΩΣ εικόνα της συγκεκριμένης λέξης.
- ΖΗΜΙΑ εικόνα της συγκεκριμένης λέξης.
- ΨΑΧΝΩ εικόνα της συγκεκριμένης λέξης.
- ΔΗΜΟΚΡΑΤΙΑ εικόνα της συγκεκριμένης λέξης.
- ΚΑΘΩΣ εικόνα της συγκεκριμένης λέξης.
- ΒΡΕΧΩ εικόνα της συγκεκριμένης λέξης.
- ΔΟΡΥΦΟΡΟΣ εικόνα της συγκεκριμένης λέξης.
- ΕΞΕΛΙΞΗ εικόνα της συγκεκριμένης λέξης.

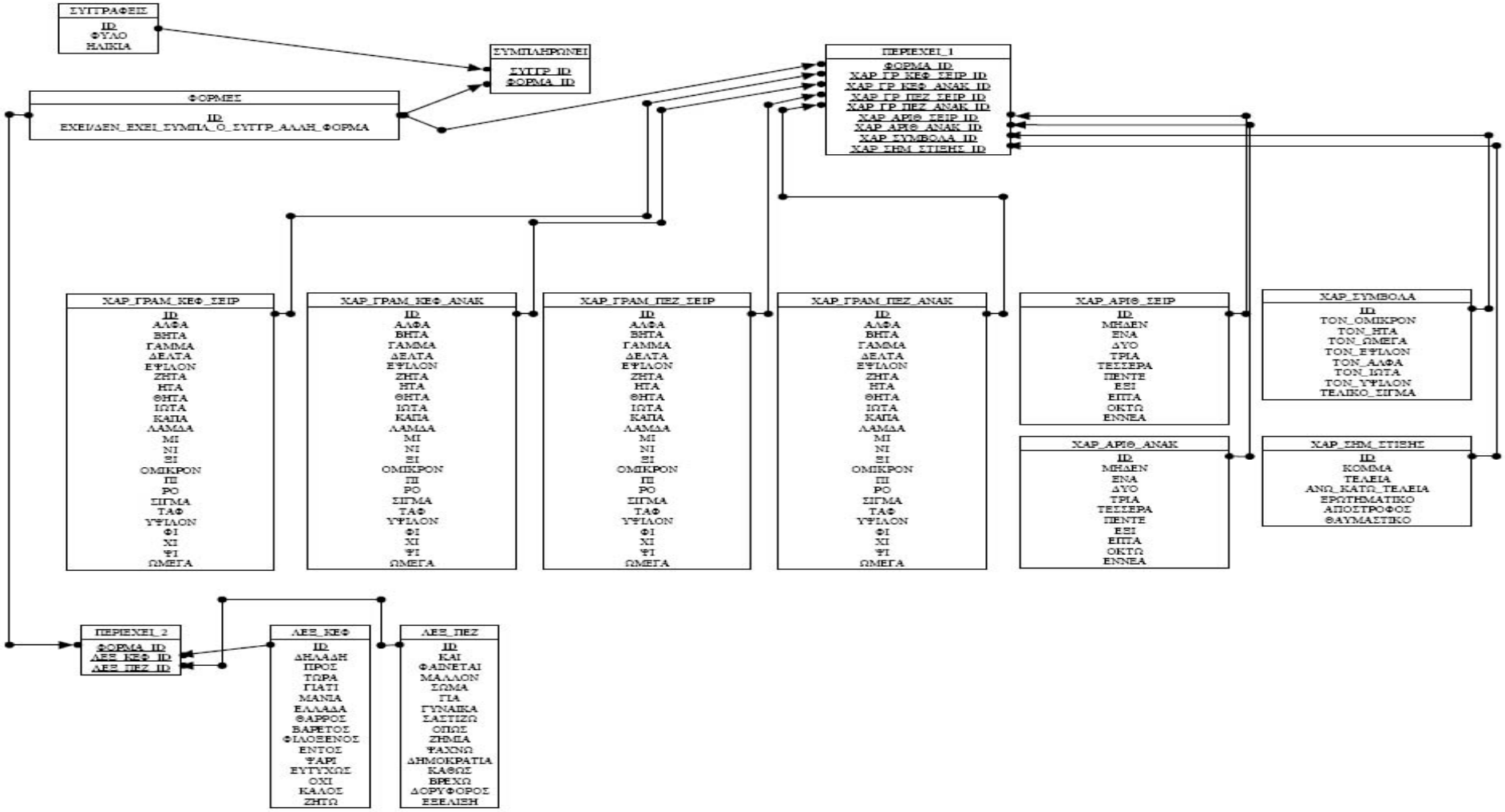
Οι σχέσεις μεταξύ των οντοτήτων είναι:

▪ ΣΥΜΠΛΗΡΩΝΕΙ: Είναι 1-Ν σχέση μεταξύ των οντοτήτων ΣΥΓΓΡΑΦΕΑΣ και ΦΟΡΜΑ αντίστοιχα. Δηλαδή κάθε συγγραφέας μπορεί να συμπληρώσει πολλές φόρμες αλλά κάθε φόρμα έχει συμπληρωθεί σίγουρα από ένα και μόνο συγγραφέα.

- ΠΕΡΙΕΧΕΙ_1: Είναι 1-Ν σχέση μεταξύ των οντοτήτων ΦΟΡΜΑ και ΧΑΡΑΚΤΗΡΕΣ αντίστοιχα. Σε κάθε φόρμα αντιστοιχούν πολλές εγγραφές της οντότητας ΧΑΡΑΚΤΗΡΕΣ (μια από κάθε εξειδίκευση).
- ΠΕΡΙΕΧΕΙ_2: Είναι 1-Ν σχέση μεταξύ των οντοτήτων ΦΟΡΜΑ και ΛΕΞΕΙΣ αντίστοιχα. Σε κάθε φόρμα αντιστοιχούν το πολύ δύο (Nmax=2) εγγραφές της οντότητας ΛΕΞΕΙΣ (μια από κάθε εξειδίκευση).

2.2 ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ

Η μετατροπή του διαγράμματος ΟΣ σε σχεσιακό έγινε με τη λογική ότι για κάθε φόρμα θα υπήρχε τουλάχιστον μια εγγραφή από κάθε εξειδίκευση. Επομένως θεωρήσαμε κάθε εξειδίκευση ως ξεχωριστή σχέση και οι συσχετίσεις με τη σειρά τους θα είναι ενδιάμεσες σχέσεις που θα περιέχουν ως ξένα κλειδιά τα πρωτεύοντα κλειδιά των σχέσεων που τους αντιστοιχούν. Στηριζόμαστε στο ότι μπορούν να υλοποιηθούν «γρήγοροι» αλγόριθμοι αναζήτησης πάνω στο συγκεκριμένο μοντέλο και η βάση μπορεί να επεκταθεί εύκολα προσθέτοντας μόνο τις καινούργιες σχέσεις που πιθανόν να χρειαστούμε. Στην επόμενη σελίδα απεικονίζεται το μοντέλο που προέκυψε.



ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ της GCDB

Οι σχέσεις που έχουν προκύψει είναι οι εξής:

- ΣΥΓΓΡΑΦΕΙΣ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΣΥΓΓΡΑΦΕΑΣ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΦΥΛΟ: BOOLEAN-Type οι τιμές που μπορεί να πάρει είναι
            1. Άνδρας
            2. Γυναίκα
        - ΗΛΙΚΙΑ: Οι τιμές που μπορεί να πάρει είναι
            1. 6-12
            2. 12-18
            3. 18-25
            4. 25-30
            5. 30 και άνω
    - Πρωτεύων κλειδί: ID
- ΦΟΡΜΕΣ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΦΟΡΜΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΕΧΕΙ/ΔΕΝ_ΕΧΕΙ_ΣΥΜΠΛ_Ο_ΣΥΓΓΡ_ΑΛΛΗ_ΦΟΡΜΑ: BOOLEAN-Type οι τιμές που μπορεί να πάρει είναι
            1. ΝΑΙ
            2. ΟΧΙ
    - Πρωτεύων κλειδί: ID
- ΣΥΜΠΛΗΡΩΝΕΙ: Είναι η σχέση που αντιστοιχεί στην συσχέτιση ΣΥΜΠΛΗΡΩΝΕΙ.
    - Πεδία:
        - ΣΥΓΓΡ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΣΥΓΓΡΑΦΕΙΣ.
        - ΦΟΡΜΑ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΦΟΡΜΕΣ.
    - Πρωτεύων κλειδί: ΣΥΓΓΡ_ID, ΦΟΡΜΑ_ID
- ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΣΕΙΡ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΓΡΑΜΜΑΤΑ -> ΚΕΦΑΛΑΙΑ -> ΜΕ ΣΕΙΡΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΑΛΦΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΒΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΑΜΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΕΛΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΖΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΘΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΙΩΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΚΑΠΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΛΑΜΔΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΜΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΝΙ: Οι τιμές που παίρνει είναι τύπου binary-image.

        - ΞΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΜΙΚΡΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΡΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΣΙΓΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΑΦ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΥΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΦΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΧΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΨΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΩΜΕΓΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΑΝΑΚ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΓΡΑΜΜΑΤΑ -> ΚΕΦΑΛΑΙΑ -> ΑΝΑΚΑΤΕΜΕΝΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΑΛΦΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΒΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΑΜΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΕΛΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΖΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΘΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΙΩΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΚΑΠΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΛΑΜΔΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΜΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΝΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΞΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΜΙΚΡΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΡΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΣΙΓΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΑΦ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΥΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΦΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΧΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΨΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΩΜΕΓΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΣΕΙΡ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΓΡΑΜΜΑΤΑ -> ΠΕΖΑ -> ΜΕ ΣΕΙΡΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΑΛΦΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΒΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΑΜΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.

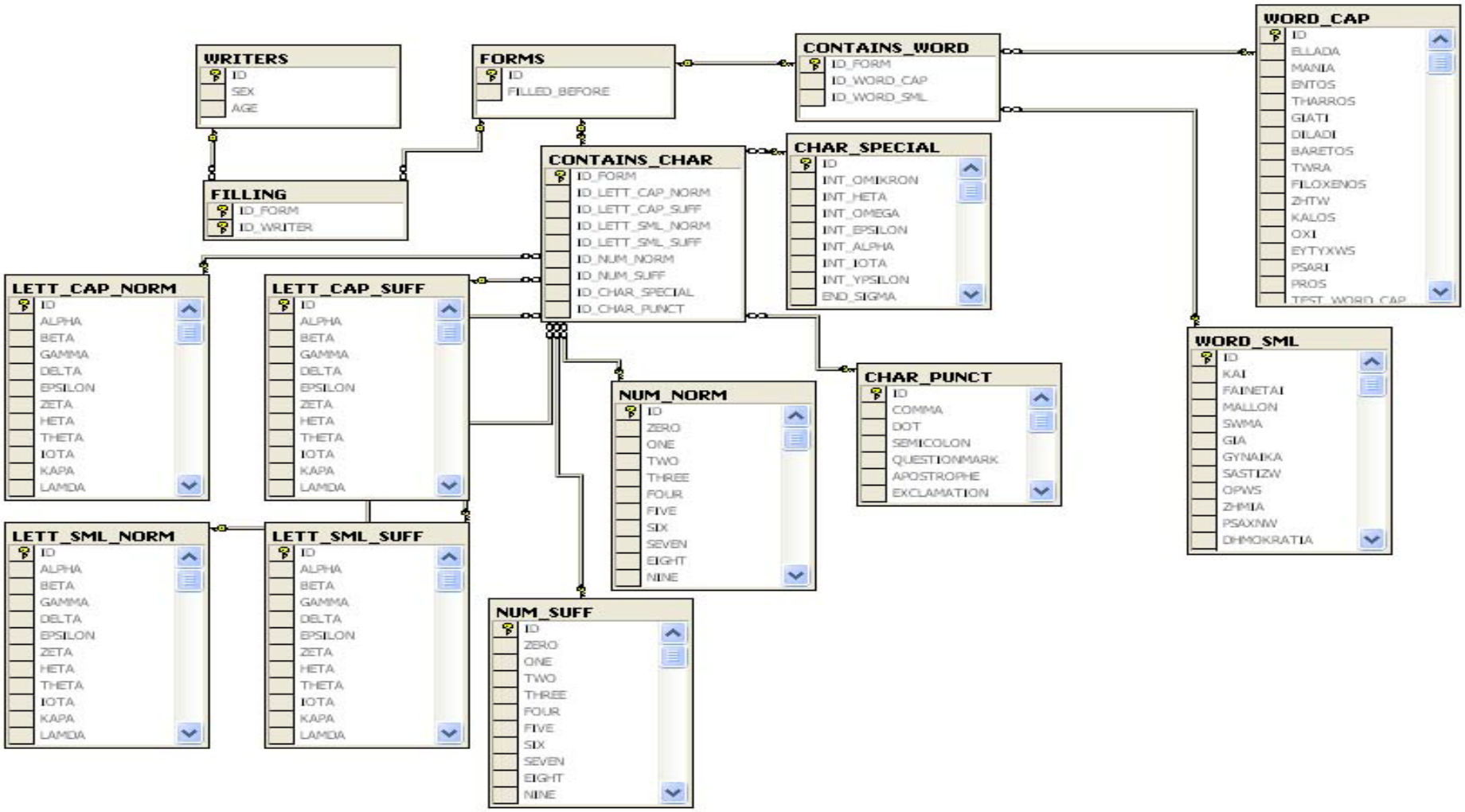
        - ΔΕΛΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΖΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΘΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΙΩΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΚΑΠΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΛΑΜΔΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΜΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΝΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΞΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΜΙΚΡΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΡΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΣΙΓΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΑΦ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΥΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΦΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΧΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΨΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΩΜΕΓΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΑΝΑΚ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΓΡΑΜΜΑΤΑ -> ΠΕΖΑ -> ΑΝΑΚΑΤΕΜΕΝΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΑΛΦΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΒΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΑΜΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΕΛΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΖΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΘΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΙΩΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΚΑΠΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΛΑΜΔΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΜΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΝΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΞΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΜΙΚΡΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΡΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΣΙΓΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΑΦ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΥΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΦΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΧΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΨΙ: Οι τιμές που παίρνει είναι τύπου binary-image.

        - ΩΜΕΓΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΑΡΙΘ_ΣΕΙΡ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΑΡΙΘΜΟΙ -> ΜΕ ΣΕΙΡΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΜΗΔΕΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΝΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΥΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΡΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΕΣΣΕΡΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΕΝΤΕ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΞΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΠΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΚΤΩ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΝΝΕΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΑΡΙΘ_ΑΝΑΚ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΑΡΙΘΜΟΙ -> ΑΝΑΚΑΤΕΜΕΝΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΜΗΔΕΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΝΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΥΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΡΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΕΣΣΕΡΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΕΝΤΕ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΞΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΠΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΚΤΩ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΝΝΕΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΣΥΜΒΟΛΑ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΕΙΔ. ΧΑΡΑΚΤΗΡΕΣ -> ΣΥΜΒΟΛΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΤΟΝ_ΟΜΙΚΡΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΟΝ_ΗΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΟΝ_ΩΜΕΓΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΟΝ_ΕΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΟΝ_ΑΛΦΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΟΝ_ΙΩΤΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΟΝ_ΥΨΙΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.

        - ΤΕΛΙΚΟ_ΣΙΓΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΧΑΡ_ΣΗΜ_ΣΤΙΞΗΣ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΧΑΡΑΚΤΗΡΕΣ -> ΕΙΔ.ΧΑΡΑΚΤΗΡΕΣ -> ΣΗΜ. ΣΤΙΞΗΣ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΚΟΜΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΕΛΕΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΑΝΩ_ΚΑΤΩ_ΤΕΛΕΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΡΩΤΗΜΑΤΙΚΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΑΠΟΣΤΡΟΦΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΘΑΥΜΑΣΤΙΚΟ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΠΕΡΙΕΧΕΙ_1: Είναι η σχέση που αντιστοιχεί στη συσχέτιση ΠΕΡΙΕΧΕΙ_1.
    - Πεδία:
        - ΦΟΡΜΑ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΦΟΡΜΕΣ.
        - ΧΑΡ_ΓΡ_ΚΕΦ_ΣΕΙΡ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΣΕΙΡ.
        - ΧΑΡ_ΓΡ_ΚΕΦ_ΑΝΑΚ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΑΝΑΚ.
        - ΧΑΡ_ΓΡ_ΠΕΖ_ΣΕΙΡ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΣΕΙΡ.
        - ΧΑΡ_ΓΡ_ΠΕΖ_ΑΝΑΚ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΑΝΑΚ.
        - ΧΑΡ_ΑΡΙΘ_ΣΕΙΡ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΑΡΙΘ_ΣΕΙΡ.
        - ΧΑΡ_ΑΡΙΘ_ΑΝΑΚ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΑΡΙΘ_ΑΝΑΚ.
        - ΧΑΡ_ΣΥΜΒΟΛΑ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΣΥΜΒΟΛΑ.
        - ΧΑΡ_ΣΗΜ_ΣΤΙΞΗΣ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΧΑΡ_ΣΗΜ_ΣΤΙΞΗΣ.
    - Πρωτεύων κλειδί: ΦΟΡΜΑ_ID
- ΛΕΞ_ΚΕΦ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΛΕΞΕΙΣ -> ΚΕΦΑΛΑΙΑ
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΔΗΛΑΔΗ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΠΡΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΤΩΡΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΙΑΤΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΜΑΝΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΛΛΑΔΑ: Οι τιμές που παίρνει είναι τύπου binary-image.

        - ΘΑΡΡΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΒΑΡΕΤΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΦΙΛΟΞΕΝΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΝΤΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΨΑΡΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΥΤΥΧΩΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΧΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΚΑΛΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΖΗΤΩ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΛΕΞ_ΠΕΖ: Είναι η σχέση που αντιστοιχεί στην οντότητα ΛΕΞΕΙΣ -> ΠΕΖΑ.
    - Πεδία:
        - ID: Μοναδικός ακέραιος αριθμός που χαρακτηρίζει κάθε εγγραφή της σχέσης.
        - ΚΑΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΦΑΙΝΕΤΑΙ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΜΑΛΛΟΝ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΣΩΜΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΓΥΝΑΙΚΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΣΑΣΤΙΖΩ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΟΠΩΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΖΗΜΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΨΑΧΝΩ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΗΜΟΚΡΑΤΙΑ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΚΑΘΩΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΒΡΕΧΩ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΔΟΡΥΦΟΡΟΣ: Οι τιμές που παίρνει είναι τύπου binary-image.
        - ΕΞΕΛΙΞΗ: Οι τιμές που παίρνει είναι τύπου binary-image.
    - Πρωτεύων κλειδί: ID
- ΠΕΡΙΕΧΕΙ_2: Είναι η σχέση που αντιστοιχεί στη συσχέτιση ΠΕΡΙΕΧΕΙ_2.
    - Πεδία:
        - ΦΟΡΜΑ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΦΟΡΜΕΣ.
        - ΛΕΞ_ΚΕΦ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΛΕΞ_ΚΕΦ.
        - ΛΕΞ_ΠΕΖ_ID: Ξένο κλειδί του πεδίου ID της σχέσης ΛΕΞ_ΠΕΖ.
    - Πρωτεύων κλειδί: ΦΟΡΜΑ_ID

Στο σχεσιακό μοντέλο παρουσιάζεται η δομή της βάσης και η οργάνωση των δεδομένων. Για την υλοποίηση χρησιμοποίησαμε το εργαλείο Access της Microsoft. Στις παρακάτω εικόνα βλέπουμε την GCDB υλοποιημένη σε Access.



Το σχεσιακό μοντέλο υλοποιημένο.

ΜΕΤΑΤΡΟΠΕΣ ΟΡΩΝ:

- WRITERS = ΣΥΓΓΡΑΦΕΙΣ
  - ID = ID
  - SEX = ΦΥΛΟ
  - AGE = ΗΛΙΚΙΑ
- FORMS = ΦΟΡΜΕΣ
  - ID = ID
  - FILLED_BEFORE = ΕΧΕΙ/ΔΕΝ_ΕΧΕΙ_ΣΥΜΠΛ_Ο_ΣΥΓΓΡ_ΑΛΛΗ_ΦΟΡΜΑ
- FILLING = ΣΥΜΠΛΗΡΩΝΕΙ
  - ID_FORM = ΦΟΡΜΑ_ID
  - ID_WRITER = ΣΥΓΓΡ_ID
- LETT_CAP_NORM = ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΣΕΙΡ
  - ID = ID
  - ALPHA = ΑΛΦΑ
  - BETA = ΒΗΤΑ
  - GAMMA = ΓΑΜΜΑ
  - DELTA = ΔΕΛΤΑ
  - EPSILON = ΕΨΙΛΟΝ
  - ZETA = ΖΗΤΑ
  - HETA = ΗΤΑ
  - THETA = ΘΗΤΑ
  - IOTA = ΙΩΤΑ
  - KAPA = ΚΑΠΑ
  - LAMDA = ΛΑΜΔΑ
  - MI = MI
  - NI = NI
  - KS =ΞΙ
  - OMIKRON = ΟΜΙΚΡΟΝ
  - PII = ΠΙ
  - RO = ΡΟ
  - SIGMA = ΣΙΓΜΑ
  - TAU = ΤΑΦ
  - YPSILON = ΥΨΙΛΟΝ
  - FI = ΦΙ
  - XI = ΧΙ
  - PSI = ΨΙ
  - OMEGA = ΩΜΕΓΑ
- LETT_CAP_SUFF = ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΠΕΖ
  - ID = ID
  - ALPHA = ΑΛΦΑ
  - BETA = ΒΗΤΑ
  - GAMMA = ΓΑΜΜΑ
  - DELTA = ΔΕΛΤΑ
  - EPSILON = ΕΨΙΛΟΝ
  - ZETA = ΖΗΤΑ
  - HETA = ΗΤΑ
  - THETA = ΘΗΤΑ
  - IOTA = ΙΩΤΑ

    - KAPA = ΚΑΠΑ
    - LAMDA = ΛΑΜΔΑ
    - MI = ΜΙ
    - NI = ΝΙ
    - KS =ΞΙ
    - OMIKRON = ΟΜΙΚΡΟΝ
    - PII = ΠΙ
    - RO = ΡΟ
    - SIGMA = ΣΙΓΜΑ
    - TAU = ΤΑΦ
    - YPSILON = ΥΨΙΛΟΝ
    - FI = ΦΙ
    - XI = ΧΙ
    - PSI = ΨΙ
    - OMEGA = ΩΜΕΓΑ
- LETT_SML_NORM = ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΣΕΙΡ
    - ID = ID
    - ALPHA = ΑΛΦΑ
    - BETA = ΒΗΤΑ
    - GAMMA = ΓΑΜΜΑ
    - DELTA = ΔΕΛΤΑ
    - EPSILON = ΕΨΙΛΟΝ
    - ZETA = ΖΗΤΑ
    - HETA = ΗΤΑ
    - THETA = ΘΗΤΑ
    - IOTA = ΙΩΤΑ
    - KAPA = ΚΑΠΑ
    - LAMDA = ΛΑΜΔΑ
    - MI = ΜΙ
    - NI = ΝΙ
    - KS =ΞΙ
    - OMIKRON = ΟΜΙΚΡΟΝ
    - PII = ΠΙ
    - RO = ΡΟ
    - SIGMA = ΣΙΓΜΑ
    - TAU = ΤΑΦ
    - YPSILON = ΥΨΙΛΟΝ
    - FI = ΦΙ
    - XI = ΧΙ
    - PSI = ΨΙ
    - OMEGA = ΩΜΕΓΑ
- LETT_SML_SUFF = ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΑΝΑΚ
    - ID = ID
    - ALPHA = ΑΛΦΑ
    - BETA = ΒΗΤΑ
    - GAMMA = ΓΑΜΜΑ
    - DELTA = ΔΕΛΤΑ
    - EPSILON = ΕΨΙΛΟΝ
    - ZETA = ΖΗΤΑ
    - HETA = ΗΤΑ

    - THETA = ΘΗΤΑ
    - IOTA = ΙΩΤΑ
    - KAPA = ΚΑΠΑ
    - LAMDA = ΛΑΜΔΑ
    - MI = ΜΙ
    - NI = ΝΙ
    - KS =ΞΙ
    - OMIKRON = ΟΜΙΚΡΟΝ
    - PII = ΠΙ
    - RO = ΡΟ
    - SIGMA = ΣΙΓΜΑ
    - TAU = ΤΑΦ
    - YPSILON = ΥΨΙΛΟΝ
    - FI = ΦΙ
    - XI = ΧΙ
    - PSI = ΨΙ
    - OMEGA = ΩΜΕΓΑ
- NUM_NORM = ΧΑΡ_ΑΡΙΘ_ΣΕΙΡ
    - ID = ID
    - ZERO = ΜΗΔΕΝ
    - ONE = ΕΝΑ
    - TWO = ΔΥΟ
    - THREE = ΤΡΙΑ
    - FOUR = ΤΕΣΣΕΡΑ
    - FIVE = ΠΕΝΤΕ
    - SIX = ΕΞΙ
    - SEVEN = ΕΠΤΑ
    - EIGHT = ΟΚΤΩ
    - NINE = ΕΝΝΕΑ
- NUM_SUFF = ΧΑΡ_ΑΡΙΘ_ΑΝΑΚ
    - ID = ID
    - ZERO = ΜΗΔΕΝ
    - ONE = ΕΝΑ
    - TWO = ΔΥΟ
    - THREE = ΤΡΙΑ
    - FOUR = ΤΕΣΣΕΡΑ
    - FIVE = ΠΕΝΤΕ
    - SIX = ΕΞΙ
    - SEVEN = ΕΠΤΑ
    - EIGHT = ΟΚΤΩ
    - NINE = ΕΝΝΕΑ
- CHAR_SPECIAL = ΧΑΡ_ΣΥΜΒΟΛΑ
    - ID = ID
    - INT_OMIKRON = ΤΟΝ_ΟΜΙΚΡΟΝ
    - INT_HETA = ΤΟΝ_ΗΤΑ
    - INT_EPSILON = ΤΟΝ_ΕΨΙΛΟΝ
    - INT_ALPHA = ΤΟΝ_ΑΛΦΑ
    - INT_IOTA = ΤΟΝ_ΙΩΤΑ
    - INT_YPSILON = ΤΟΝ ΥΨΙΛΟΝ
    - END_SIGMA = ΤΕΛΙΚΟ_ΣΙΓΜΑ

- CHAR_PUNCT = ΧΑΡ_ΣΗΜ_ΣΤΙΞΗΣ
  - ID = ID
  - COMMA = ΚΟΜΜΑ
  - DOT = ΤΕΛΕΙΑ
  - SEMICOLON = ΑΝΩ_ΚΑΤΩ_ΤΕΛΕΙΑ
  - QUESTIONMARK = ΕΡΩΤΗΜΑΤΙΚΟ
  - APOSTROPHE = ΑΠΟΣΤΡΟΦΟΣ
  - EXCLAMATION = ΘΑΥΜΑΣΤΙΚΟ
- CONTAINS_CHAR = ΠΕΡΙΕΧΕΙ_1
  - ID_FORM = ΦΟΡΜΑ_ID
  - ID_LETT_CAP_NORM = ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΣΕΙΡ_ID
  - ID_LETT_CAP_SUFF = ΧΑΡ_ΓΡΑΜ_ΚΕΦ_ΑΝΑΚ_ID
  - ID_LETT_SML_NORM = ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΣΕΙΡ_ID
  - ID_LETT_SML_ SUFF = ΧΑΡ_ΓΡΑΜ_ΠΕΖ_ΑΝΑΚ_ID
  - ID_NUM_NORM = ΧΑΡ_ΑΡΙΘ_ΣΕΙΡ_ID
  - ID_NUM_SUFF = ΧΑΡ_ΑΡΙΘ_ΑΝΑΚ_ID
  - ID_CHAR_SPECIAL = ΧΑΡ_ΣΥΜΒΟΛΑ_ID
  - ID_CHAR_PUNCT = ΧΑΡ_ΣΗΜ_ΣΤΙΞΗΣ_ID
- WORD_CAP = ΛΕΞ_ΚΕΦ
  - ID = ID
  - ELLADA =ΕΛΛΑΔΑ
  - MANIA = ΜΑΝΙΑ
  - ENTOS = ΕΝΤΟΣ
  - THARROS = ΘΑΡΡΟΣ
  - GIATI = ΓΙΑΤΙ
  - DILADI = ΔΗΛΑΔΗ
  - BARETOS = ΒΑΡΕΤΟΣ
  - TWRA = ΤΩΡΑ
  - FILOXENOS = ΦΙΛΟΞΕΝΟΣ
  - ZHTW = ΖΗΤΩ
  - KALOS = ΚΑΛΟΣ
  - OXI = ΟΧΙ
  - EYTYXWS = ΕΥΤΥΧΩΣ
  - PSARI = ΨΑΡΙ
  - PROS = ΠΡΟΣ
- WORD_SML = ΛΕΞ_ΠΕΖ
  - ID = ID
  - KAI = ΚΑΙ
  - FAINETAI = ΦΑΙΝΕΤΑΙ
  - MALLON = ΜΑΛΛΟΝ
  - SWMA = ΣΩΜΑ
  - GIA = ΓΙΑ
  - GYNAIKA = ΓΥΝΑΙΚΑ
  - SASTIZW = ΣΑΣΤΙΖΩ
  - OPWS = ΟΠΩΣ
  - ZHMIA = ΖΗΜΙΑ
  - PSAXNW = ΨΑΧΝΩ
  - DHMOKRATIA = ΔΗΜΟΚΡΑΤΙΑ
  - EKSELIKSI = ΕΞΕΛΙΞΗ
  - KATHWS = ΚΑΘΩΣ

    - BREXW = ΒΡΕΧΩ
    - DORYFOROS = ΔΟΡΥΦΟΡΟΣ
- CONTAINS_WORD = ΠΕΡΙΕΧΕΙ_2
    - ID_WORD_CAP = ΛΕΞ_ΚΕΦ_ID
    - ID_WORD_SML = ΛΕΞ_ΠΕΖ_ID

3. Λογισμικό της GCDB

Για την έρρυθμη χρήση της βάσης δεδομένων καθίσταται αναγκαία και η δημιουργία του κατάλληλου λογισμικού, το οποίο θα αναλαμβάνει την εισαγωγή και την εξαγωγή των δεδομένων στη βάση. Το λογισμικό το οποίο δημιουργήθηκε αποτελείται από δύο επιμέρους τμήματα: το λογισμικό εισαγωγής δεδομένων στη βάση και το λογισμικό εξαγωγής δεδομένων.

Το λογισμικό εισαγωγής στη βάση δεδομένων αναλαμβάνει να πάρει μία εικόνα της σάρωσης της φόρμας εισαγωγής δεδομένων, να εντοπίσει τα επιμέρους στοιχεία της (γράμματα, αριθμούς, λέξεις και στοιχεία συγγραφέα) και να τα εισάγει στη βάση δεδομένων στα κατάλληλα πεδία. Η εισαγωγή γίνεται αυτόματα και το μόνο που χρειάζεται από το χρήστη είναι να προσδιορίσει τα αρχεία σάρωσης των φορμών εισαγωγής.

Το λογισμικό εξαγωγής δεδομένων από τη βάση είναι μια απλή εφαρμογή που στέλνει ένα ερώτημα (query) στη βάση δεδομένων και αποθηκεύει τα αποτελέσματα σε ένα φάκελο που προσδιορίζει ο χρήστης.

Τόσο η αποθήκευση όσο και η εξαγωγή των δεδομένων γίνεται σε εικόνες της μορφής bitmap(.bmp).

Η ανάπτυξη του λογισμικού έγινε με Object Pascal στην πλατφόρμα Delphi 6.02 της Borland, με τη χρήση των παρακάτω υποσυστήματα/δομοστοιχεία (components):

- Βιβλιοθήκη psvDialogs από Serhiy Perevoznyk, η οποία εμπεριέχει components βασισμένα σε μη καταγεγραμμένες κλήσεις συστήματος και διαδικασίες των windows.
- ImageEn 2.2.3 από HiComponents, που περιέχει components και διαδικασίες επεξεργασίας εικόνων.

Επιπλέον έγιναν αλλαγές στο αρχείο της Borland: HelpIntfs.pas το οποίο χρησιμοποιείται για την διασύνδεση εφαρμογών με αρχεία βοήθειας (.hlp και .chm μορφής).

Η ανάπτυξη του αρχείου εγκατάστασης έγινε με τη χρήση του Inno Setup 5 (έκδοση 5.1.12) και του βοηθητικού προγράμματος ISTools (έκδοση 5.1.8). Επιπλέον στο ISTools χρησιμοποιήθηκαν τα παρακάτω αρθρώματα:

- isxdl.dll, προσθήκη δυνατοτήτων διασύνδεσης με το διαδίκτυο για τη λήψη αρχείων.
- ISCrypt.dll, προσθήκη δυνατοτήτων κρυπτογράφησης.

3.1 Περιγραφή Εφαρμογής Εισαγωγής Δεδομένων

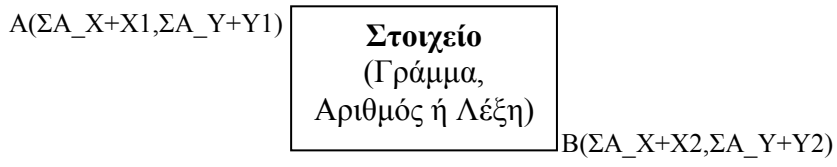
Η εφαρμογή εισαγωγής δεδομένων στη βάση με την ονομασία GCDBInput είναι το τμήμα λογισμικού που αναλαμβάνει να εντοπίσει και αποθηκεύσει τα τμήματα που περιέχουν δείγματα γραφικού χαρακτήρα από μία φόρμα εισαγωγής και να τα εισάγει στα κατάλληλα πεδία στη βάση δεδομένων.

Η εφαρμογή αρχικά λαμβάνει ως είσοδο μία εικόνα της σάρωσης της φόρμας εισαγωγής δεδομένων που έχει συμπληρωθεί από έναν συγγραφέα. Έπειτα αναλαμβάνει τη μετατροπή της εικόνας σε ασπρόμαυρη. Η μετατροπή αυτή γίνεται με βάση ένα κατώφλι (threshold), το οποίο ορίζεται παραμετρικά από την εφαρμογή. Όλα τα εικονοστοιχεία (pixels) που βρίσκονται άνω του ορίου θα μετατραπούν σε λευκά, και όλα κάτω του ορίου θα μετατραπούν σε μαύρα.

Εφόσον έχει ολοκληρωθεί η μετατροπή σε ασπρόμαυρη εικόνα το πρόγραμμα επιχειρεί να ευθυγραμμίσει την εικόνα. Για να το πετύχει αυτό η εφαρμογή χρησιμοποιεί έναν αλγόριθμο ο οποίος υπολογίζει την κλίση σε μοίρες και έπειτα περιστρέφει την εικόνα. Για τον υπολογισμό της ροπής σε μοίρες χρησιμοποιεί προοδευτικές περιστροφές της εικόνας κατά 0,1 μοίρες μέχρι να εντοπίσει την καλύτερη κλίση.

Μετά την ευθυγράμμιση της εικόνας η εφαρμογή προχωρά στην εύρεση του σημείου αρχής των συντεταγμένων κοπής, ώστε να μπορέσει να εντοπίσει και να αποθηκεύσει τα δείγματα γραφικού χαρακτήρα που εμπεριέχονται στη φόρμα που έχει σαρωθεί. Ο εντοπισμός του σημείου αρχής των συντεταγμένων γίνεται βάση της μεγάλης μαύρης γραμμής που διατρέχει οριζόντιως κάθε φόρμα την οποία βρίσκει ψάχνοντας το ποσοστό των μαύρων εικονοστοιχείων σε κάθε γράμμη της φόρμας μέχρι να βρει τις πρώτες 4 γραμμές όπου το ποσοστό αυτών είναι τουλάχιστον 90%. Το αρχικό σημείο βρίσκεται στην αρχή της μαύρης αυτής γραμμής. Όταν θα εντοπιστεί το πρόγραμμα θα μπορεί να εντοπίσει και να κόψει τα επιμέρους στοιχεία της φόρμας βάση συντεταγμένων που βρίσκονται σε ξεχωριστό αρχείο τύπου ini και να τα αποθηκεύσει στα αντίστοιχα πεδία στη βάση δεδομένων μας.

Η κοπή γίνεται βάση συντεταγμένων που βρίσκονται αποθηκευμένες σε βοηθητικό αρχείο. Για κάθε στοιχείο (γράμμα, αριθμό ή λέξη) υπάρχουν 2 ζευγάρια συντεταγμένων, τα X1,Y1 και X2,Y2, τα οποία ορίζουν την απόσταση του άνω αριστερού και κάτω δεξιού τμήματος του ορθογωνίου που περικλείει κάθε στοιχείο από το σημείο αρχής (ΣΑ).



Παρόμοια λογική ακολουθεί το πρόγραμμα και για τη λήψη πληροφοριών για τα δεδομένα του ίδιου συγγραφέα, που από εδώ και στο εξής θα αναφέρονται ως κριτήρια. Τα δεδομένα αυτά, τα οποία έχουν και προαναφερθεί, είναι η ηλικία, το φύλο και το γεγονός αν έχει συμπληρώσει ή όχι ξανά τη φόρμα μας. Σε αντίθεση με τα δεδομένα γραφικού χαρακτήρα που λαμβάνονται απλά ως εικόνες και ζητούνται

ως εικόνες, τα δεδομένα αυτά θα πρέπει να μετατραπούν από εικόνα σε αριθμητικά και Αλήθειας-Ψεύδους (Boolean).

Η μορφή με την οποία συμπληρώνονται από το συγγραφέα είναι σε κουτιά όπου ο συγγραφέας καλείται να σημειώσει στο κατάλληλο κουτί. Έτσι υπάρχουν 2 κουτιά για το αν έχει συμπληρώσει ή όχι ξανά, 2 κουτιά για το φύλλο και 5 κουτιά για την ηλικία, ανάλογα με το ηλικιακό κατηγορία στην οποία υπάγεται ο συγγραφέας. Οι ηλικιακές κατηγορίες είναι:

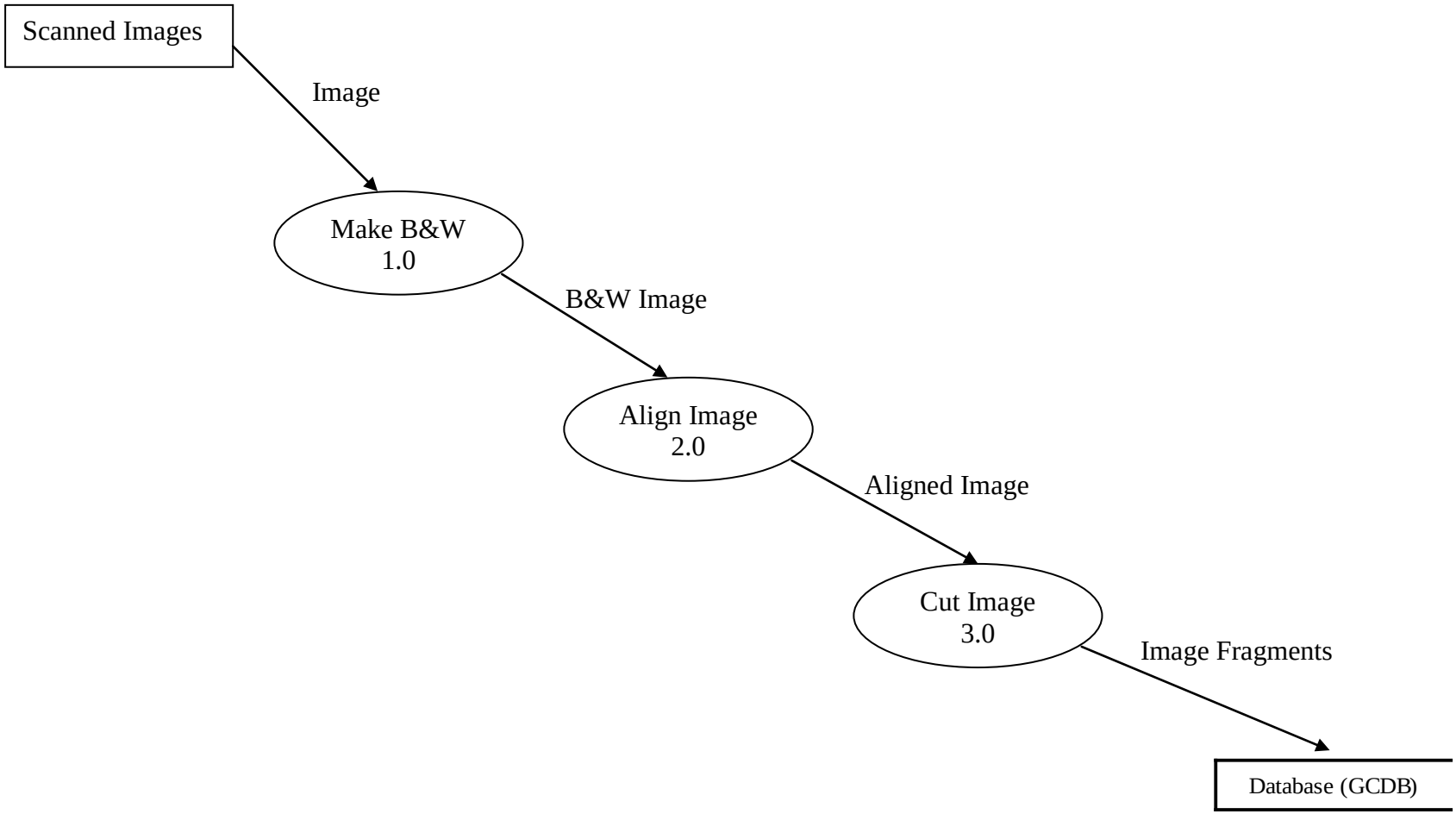
- 6 – 12
- 12 – 18
- 18 – 25
- 25 – 30
- 30+

Στη βάση δεδομένων δεν είναι επιθυμητό να αποθηκεύονται εικόνες για τα δεδομένα αυτά καθώς θα ήταν αδύνατη η εύρεση φορμών με βάση τα παραπάνω στοιχεία συγγραφέα. Έτσι μετά την κοπή και πριν την αποθήκευση η εφαρμογή θα πρέπει να τα μετατρέψει σε αριθμητικά και Αλήθειας-Ψεύδους (Boolean) δεδομένα.

Εφόσον η συμπλήρωση γίνεται σημειώνοντας σε ένα κουτί, η εφαρμογή μετράει για κάθε ένα από τα παραπάνω στοιχεία τον αριθμό των μαύρων εικονοστοιχείων (pixel) που βρίσκονται σε κάθε ένα από τα κουτιά τους. Για παράδειγμα, για το φύλλο μετράει τα μαύρα εικονοστοιχεία στα κουτιά 6-12, 12-18, 18-25, 25-30 και 30+, ενώ για το φύλλο στα κουτιά Άνδρας και Γυναίκα. Το κουτί που θα έχει επιλεχθεί από το συγγραφέα θα έχει τα περισσότερα μαύρα εικονοστοιχεία, αφού ο συγγραφέας σημείωσε με μελάνι το κουτί που τον αντιπροσωπεύει. Όταν η εφαρμογή εντοπίσει ποιο κουτί έχει τα περισσότερα μαύρα εικονοστοιχεία τότε αποθηκεύει στη βάση δεδομένων την κατάλληλη τιμή.

Η εφαρμογή έχει σχεδιαστεί ώστε να μπορεί να συνδέεται με διάφορες βάσεις δεδομένων με μικρές αλλαγές στις παραμέτρους. Έχει δοκιμαστεί η λειτουργία της με MS Access και MS SQL Server. Η σύνδεση με Access γίνεται με βάση τους Microsoft Jet 4.0 OLE DB Drivers και με τον SQL Server με τους Microsoft OLE DB Provider for SQL Server Drivers.

Ακολουθεί το διάγραμμα ροής δεδομένων της εφαρμογής εισαγωγής δεδομένων καθώς και η δομημένη περιγραφή και λεξικό δεδομένων:


Scanned Images
Image
Make B&W
1.0
B&W Image
Align Image
2.0
Aligned Image
Cut Image
3.0
Image Fragments
Database (GCDB)

Δομημένη Περιγραφή

Όνομα Επεξεργασίας: Make B&W
Κωδικός: 1.0
Input: Image
Output: B&W Image
Trigger: Είσοδος εικόνας.

Περιγραφή
Όταν εισέλθει μία εικόνα στο σύστημα τότε:
Προσπάθησε:
Begin
  Μετέτρεψε εικόνα σε BMP
  Μετέτρεψε την εικόνα σε ασπρόμαυρη.
End
Αλλιώς (* Αν αποτύχεις*)
Begin
  Εμφάνισε μήνυμα λάθους.
End

<u>Παρατηρήσεις</u>

1. **Τρόποι Επεξεργασίας:** Θα πρέπει το σύστημα να είναι ανεκτικό και αν μπορεί αν χειριστεί τόσο έγχρωμες όσο και εικόνες στην κλίμακα του γκρι.
2. **Απαιτήσεις Ασφάλειας:** Δεν έχει εφαρμογή.
3. **Ποσοτικά Στοιχεία:** Δεν έχει εφαρμογή.
4. **Προτεραιότητες:** Δεν έχει εφαρμογή.

Όνομα Επεξεργασίας: Align Image
Κωδικός: 2.0
Input: B&W Image
Output: Aligned Image
Trigger: Είσοδος εικόνας

Περιγραφή
Όταν εισέλθει μια εικόνα στο σύστημα τότε:
Προσπάθησε:
Begin
  Υπολόγισε την κλίση της.
  Κάνε περιστροφή της εικόνας κατά μοίρες ίσες με την κλίση της εικόνας.
End
Αλλιώς (* Αν αποτύχεις *)
Begin
  Εμφάνισε μήνυμα λάθους.
End

Παρατηρήσεις

1. **Τρόποι Επεξεργασίας:** Αν βρεθεί ότι μια εικόνα δεν έχει κλίση (δηλαδή έχει κλίση 0 μοίρες), τότε πάλι γίνεται περιστροφή κατά 0 μοίρες (Ουσιαστικά δεν περιστρέφεται)
2. **Απαιτήσεις Ασφάλειας:** Δεν έχει εφαρμογή.
3. **Ποσοτικά Στοιχεία:** Δεν έχει εφαρμογή.
4. **Προτεραιότητες:** Δεν έχει εφαρμογή.

Όνομα Επεξεργασίας: Cut Image
Κωδικός: 3.0
Input: Aligned Image
Output: Image Fragments
Trigger: Είσοδος εικόνας

Περιγραφή
Όταν εισέλθει μία εικόνα στο σύστημα βρες τα επιμέρους τμήματα που πρέπει να κοπούν και απέκοψε τα τμήματα αυτά από την εικόνα.

Παρατηρήσεις:

1. **Τρόποι Επεξεργασίας:** Δεν γίνεται έλεγχος για το αν τα στοιχεία που κόβει ή όχι είναι τα ζητούμενα. Τα στοιχεία κόβονται βάση συντεταγμένων.
2. **Απαιτήσεις Ασφάλειας:** Δεν έχει εφαρμογή.
3. **Ποσοτικά Στοιχεία:** Αποκόπτονται 161 μικρότερες εικόνες από κάθε εικόνα (24*4 Γράμματα, 9 Ειδικοί χαρακτήρες, 6 Στοιχεία στίξης, 10*2 Αριθμοί και 15*2 Λέξεις).
4. **Προτεραιότητες:** Δεν έχει εφαρμογή.

Λεξικό Δεδομένων

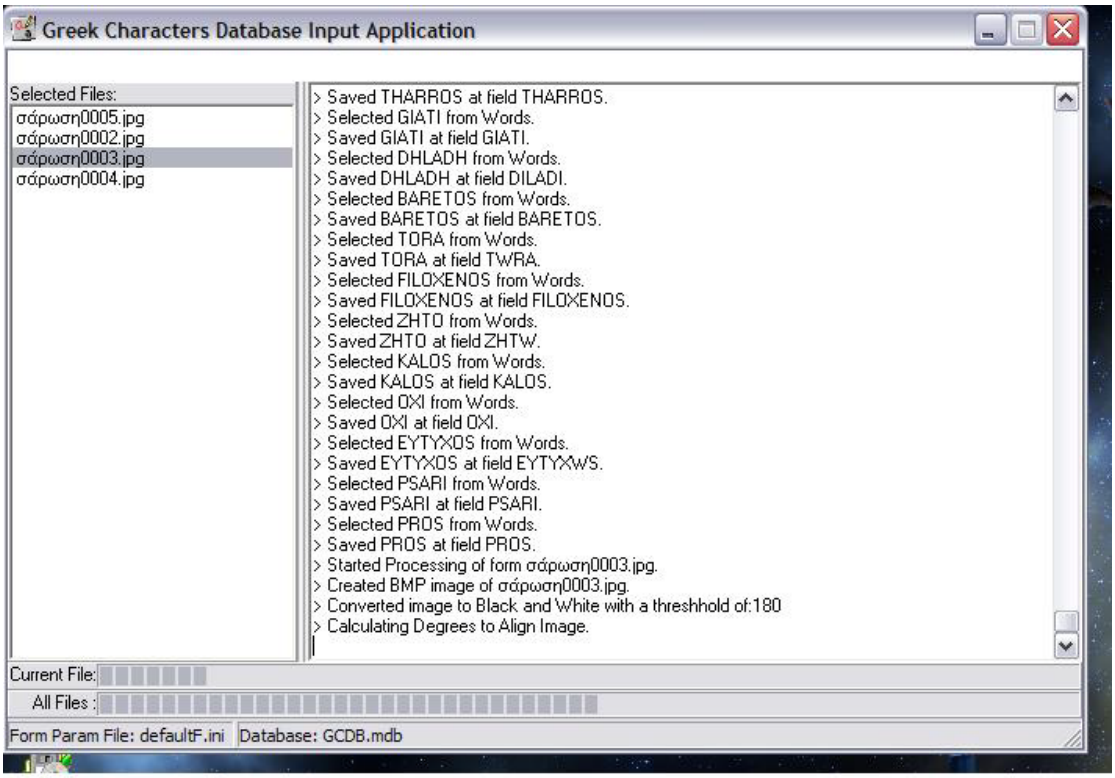
Scanned Images: Το σύνολο των εικόνων που έχουν σαρωθεί από τις φόρμες που έχουν συμπληρώσει οι συγγραφείς.
Image: Μία από τις αρχικές, πρωτότυπες εικόνες που έχουν σαρωθεί από τις φόρμες.
B&W Image: Η ασπρόμαυρη εικόνα της Image που προέκυψε μετά από τη διεργασία Make B&W.
Aligned Image: Η ευθυγραμμισμένη ασπρόμαυρη εικόνα της αρχικής φόρμας που προέκυψε από τη διεργασία Align Image.
Image Fragments: Το σύνολο των μικρότερων εικόνων που έχουν αποκοπεί από την Aligned Image. Αποτελούν τα επιμέρους πεδία που έχουν συμπληρώσει οι συγγραφείς στην αρχική φόρμα.
Database (GCDB): Η βάση δεδομένων στην οποία αποθηκεύονται οι εικόνες που αποκόπηκαν από τη διεργασία Cut Image.

Η εφαρμογή εισαγωγής αποτελείται από τα παρακάτω αρχεία:

- GCDBInput.exe Κυρίως εκτελέσιμο αρχείο
- GCDBInput.ini Αρχείο που περιέχει τις απαραίτητες παραμέτρους της εφαρμογής.
- defaultF.ini ή άλλο όνομα. Πρόκειται για το αρχείο που περιγράφει μία φόρμα και εμπεριέχει πληροφορίες όπως το κατώφλι μετατροπής σε ασπρόμαυρη εικόνα, συντεταγμένες για κάθε στοιχείο, καθώς και μια πληθώρα άλλων πληροφοριών.
- GCDBInput.chm Το αρχείο βοήθειας της εφαρμογής.
- GCDBInput.log Ένα αρχείο κειμένου που περιέχει το αρχείο καταγραφής της εφαρμογής (Application Log File), σε περίπτωση που αυτή έχει ενεργοποιηθεί από τις επιλογές.

Περιγραφή γραφικού περιβάλλοντος εφαρμογής.

Το γραφικό περιβάλλον της εφαρμογής εισαγωγής αποτελείται από 2 τμήματα και μία μπάρα κάτω από τα τμήματα αυτά. Το ένα τμήμα αποτελεί μία λίστα με τα αρχεία τα οποία έχουν φορτωθεί στην εφαρμογή και είναι υπό επεξεργασία ή αναμένουν επεξεργασία. Στο δεύτερο τμήμα βρίσκεται το αρχείο καταγραφής όπου, εφόσον έχει ενεργοποιηθεί η αντίστοιχη επιλογή, εμφανίζονται πληροφορίες για την τρέχουσα ενέργεια της εφαρμογής.



Κάτω από τα δύο τμήματα βρίσκεται η διπλή μπάρα προόδου. Η άνωθεν μπάρα δείχνει την πρόοδο του τρέχοντος αρχείου και η κάτωθεν της συνολικής εργασίας.

Η εφαρμογή περιέχει 3 μενού επιλογών με υποεπιλογές στο καθέ ένα από αυτά. Αυτά με τη σειρά που εμφανίζονται είναι:
Μενού File με τις επιλογές:

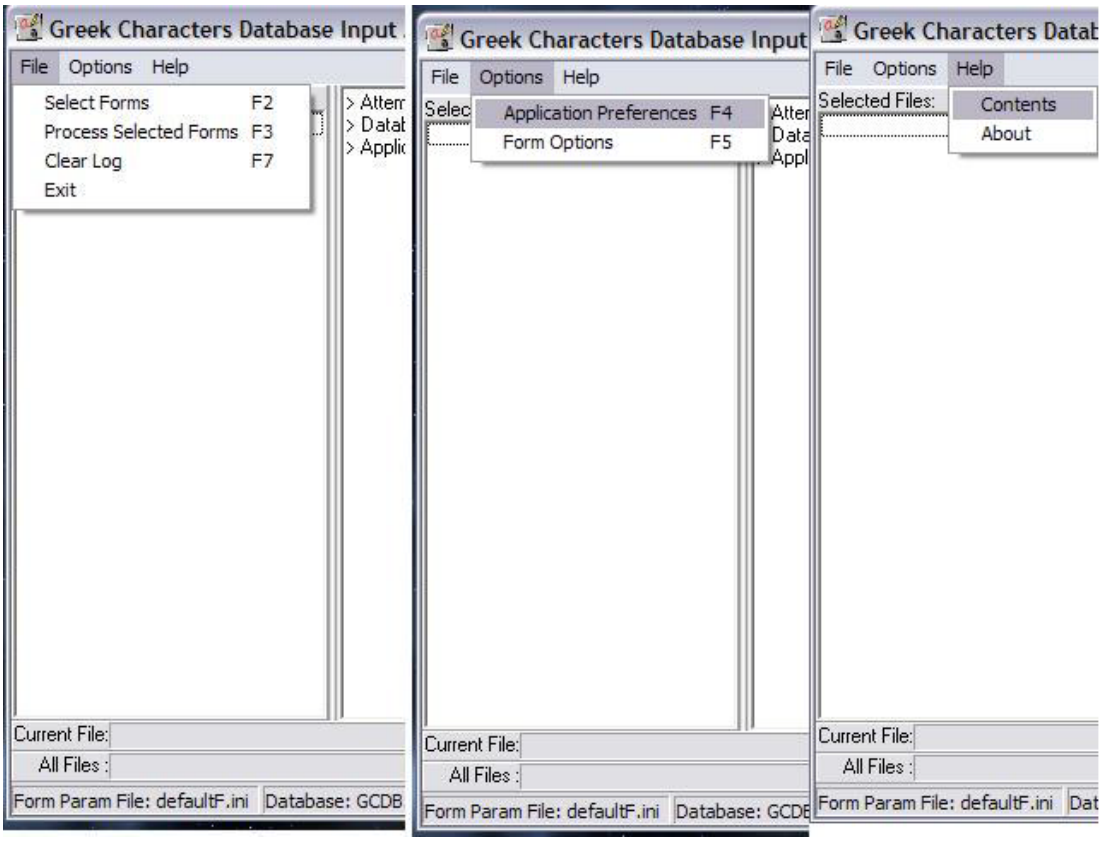
- Select Forms
- Process Selected Forms
- Clear Log
- Exit

Μενού Options με τις επιλογές:

- Application Preferences
- Form Options

Μενού Help με τις επιλογές:

- Contents
- About



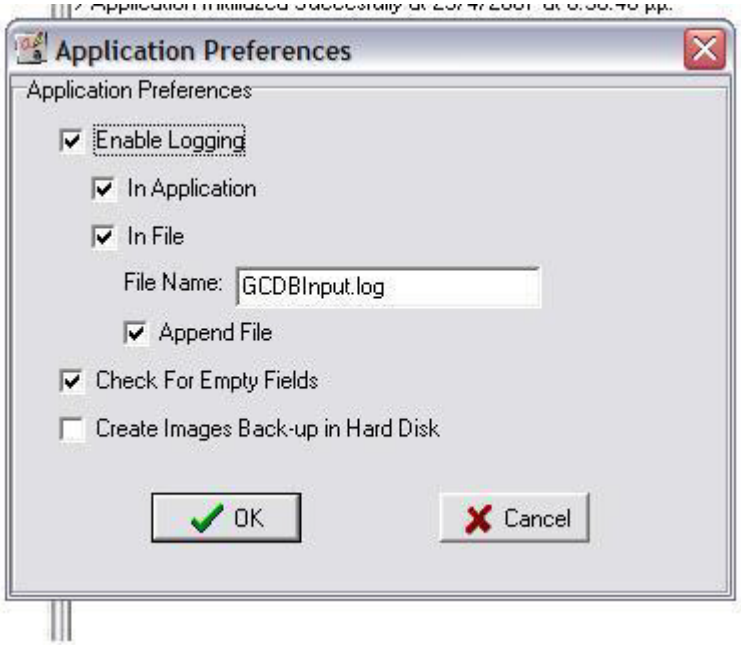
Η επιλογή Select Forms ανοίγει ένα παράθυρο διαλόγου όπου ο χρήστης καλείται να επιλέξει το αρχείο/α εικόνας/ων που θέλει να επεξεργαστεί με την εφαρμογή. Μετά την επιλογή τα ονόματα των αρχείων θα εμφανιστούν στο αριστερό τμήμα.

Η επιλογή Process Selected Forms αρχίζει την επεξεργασία που περιγράφικε παραπάνω για όσες εικόνες έχουν επιλεγεί μέσω της Select Form. Με την έναρξη της επεξεργασίας πληροφορίες για την τρέχουσα ενέργεια θα αρχίσουν αν εμφανίζονται στο δεξί τμήμα, ενώ παράλληλα θα αρχίσουν αν γεμίζουν και οι μπάρες προόδου.

Η επιλογή Clear Log καθαρίζει όλες τις πληροφορίες από το πλαίσιο καταγραφής. Οι πληροφορίες αυτές καθαρίζονται μόνο από το τμήμα της εφαρμογής και όχι από το αρχείο καταγραφής (αν φυσικά είναι ενεργοποιημένη η καταγραφή σε αρχείο).

Η επιλογή Application Preferences εμφανίζει ένα νέο παράθυρο όπου υπάρχουν ρυθμίσεις και επιλογές σχετικές με την εφαρμογή. Οι επιλογές που υπάρχουν στο νέο

παράθυρο αφορούν την καταγραφή (Logging) , αν η εφαρμογή θα κάνει έλεγχο κενών πεδίων και αν θα δημιουργεί αντίγραφα των εικόνων που αποθηκεύει στη βάση. Ο έλεγχος κενών πεδίων έγκειται στο να ελέγχει η εφαρμογή αν κάποιο από τα πεδία στην εικόνα της φόρμας είναι πιθανώς κενό. Αν, παραδείγματος χάρη, ένας συγγραφέας δεν συμπλήρωσε μία από τις λέξεις και την άφησε κενή.

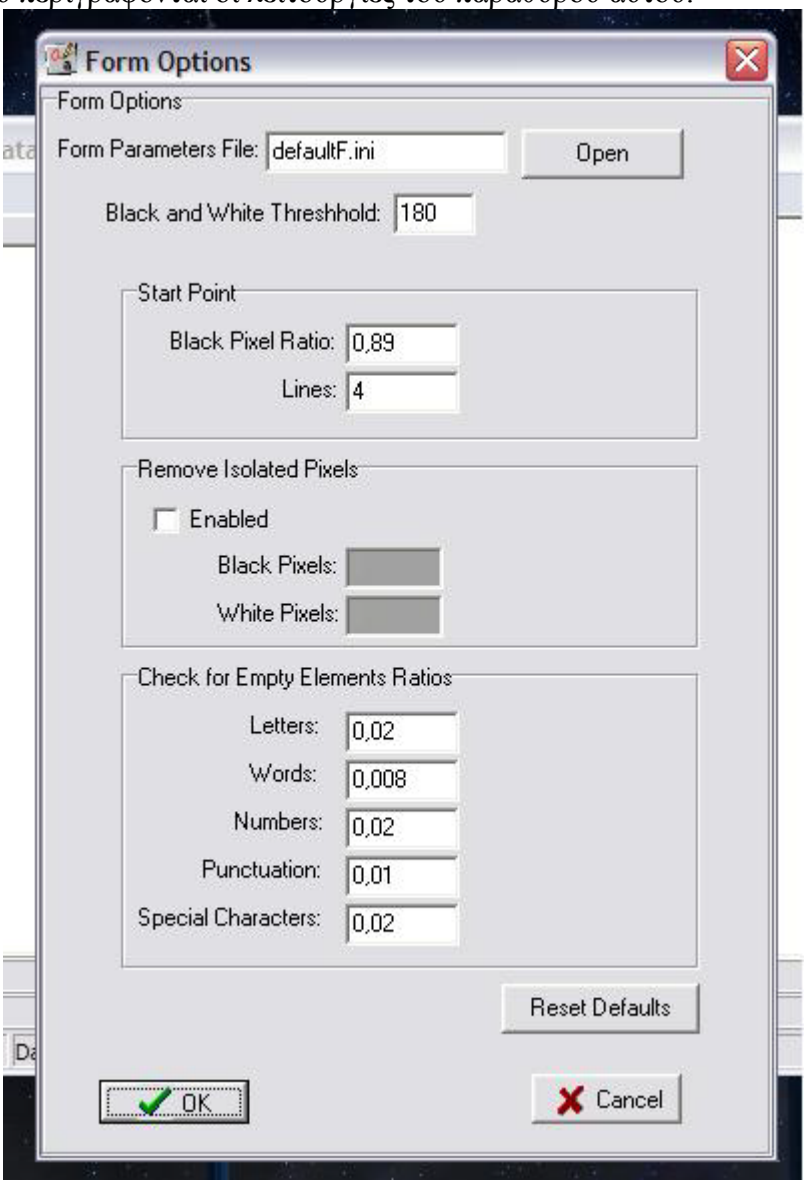


Ο έλεγχος αυτός επιτυγχάνεται με τη μέτρηση των μαύρων εικονοστοιχείων (pixel) της κάθε πλαισίου όπου αναμένει το πρόγραμμα ότι υπάρχουν δεδομένων (γραφή συγγραφέα), την εύρεση της αναλογίας των μαύρων προς των συνολικών εικονοστοιχείων και τη σύγκριση του αριθμού αυτό με ένα απόλυτο ελάχιστο που βρίσκεται αποθηκευμένο στο αρχείο περιγραφής της φόρμας. Για την εύρεση των ελάχιστων αυτών αναλογιών έγιναν μετρήσεις σε περίπου 100 διαφορετικές εικόνες φορμών εισαγωγής. Τα δεδομένα αυτά παρατίθενται στο τεχνικό παράρτημα.

Η επιλογή Form Options περιέχει ρυθμίσεις σχετικά με τις εικόνες των φορμών εισαγωγής. Οι αλλαγές που γίνονται σε αυτό το παράθυρο είναι για προχωρημένους χρήστες και οποιαδήποτε αλλαγή είναι πιθανόν να δημιουργήσει πρόβλημα στην επεξεργασία των εικόνων. Οι επιλογές που προσφέρονται είναι η αλλαγή του αρχείου περιγραφής της φόρμας (Form Parameters File – Αρχείο παραμέτρων φορμών), αλλαγή του κατωφλιού για τη μετατροπή της εικόνας ασπρόμαυρη, αλλαγές στις παραμέτρους εύρεσης του αρχικού σημείου συντεταγμένων, επιλογές καθαρισμού μεμονωμένων μαύρων και λευκών εικονοστοιχείων, καθώς και τις ελάχιστες αναλογίες που προαναφέρθηκαν για τον έλεγχο κενών πεδίων. Περισσότερες πληροφορίες και οδηγίες για το αρχείο περιγραφής της φόρμας βρίσκονται στο αντίστοιχο τμήμα του τεχνικού παραρτήματος.

Η επιλογή Contents ανοίγει το ευρετήριο του αρχείου βοήθειας και τέλος η επιλογή About εμφανίζει ένα παράθυρο με τα πνευματικά δικαιώματα της εφαρμογής. Αξίζει να σημειωθεί ότι ο χρήστης μπορεί να έχει πρόσβαση στο αρχείο βοήθειας οποιαδήποτε στιγμή πατώντας το πλήκτρο F1. Μάλιστα αναλόγως με το που βρίσκεται όταν πατήσει το πλήκτρο η βοήθεια θα ανοίξει στο αντίστοιχο σημείο. Πχ,

αν πατήσει F1 ενώ βρίσκεται στο παράθυρο Form Options θα ανοίξει η βοήθεια στο σημείο όπου περιγράφονται οι λειτουργίες του παραθύρου αυτού.



Πρέπει να σημειωθεί τα περισσότερα δεδομένα που χρειάζεται η εφαρμογή τα λαμβάνει παραμετρικά από τα 2 .ini αρχεία. Τα δεδομένα αυτά περιλαμβάνουν:

- Όλες τις ρυθμίσεις τις εφαρμογής που προσφέρονται και από το παράθυρο των επιλογών.
- Το όνομα του αρχείου περιγραφής της φόρμας.
- Το όνομα του αρχείου βοήθειας.
- Τη σύνδεση με τη βάση δεδομένων.
- Όλες τις πληροφορίες για τη επεξεργασία των εικόνων, καθώς και τις συντεταγμένες κοπής κάθε στοιχείου.

3.2 Περιγραφή Εφαρμογής Εξαγωγής Δεδομένων

Η εφαρμογή εξαγωγής δεδομένων είναι ένα ιδιαίτερα απλό κομμάτι λογισμικού που κάνει ένα επερώτημα (query) στη βάση δεδομένων και λαμβάνει δεδομένα. Η χρήση του λογισμικού αυτού δεν είναι αναγκαία, σε αντίθεση μην εφαρμογή εισαγωγής δεδομένων, και οποιοσδήποτε μπορεί να κατασκευάσει μία εφαρμογή για να λαμβάνει δεδομένα από τη βάση.

Ο σκοπός της συγκεκριμένης εφαρμογής είναι περισσότερο για να παρουσιάσουμε την ορθή λειτουργία της βάσης, καθώς προσφέρει μόνο τη δυνατότητα λήψης των δεδομένων με τη μορφή εικόνων σε φάκελους στο σκληρό δίσκο, περιορίζοντας έτσι τις δυνατότητες για λήψη δεδομένων για συστηματική εκμάθηση άλλων εφαρμογών.

Η λειτουργία του προγράμματος είναι σχετικά απλή. Ο χρήστης επιλέγει το φάκελο που θέλει να αποθηκευτούν τα αποτελέσματα του επερωτήματός του και έπειτα κάνοντας απλές επιλογές κριτηρίων η εφαρμογή χτίζει το επερώτημα και λαμβάνει τα δεδομένα.

Ο χρήστης έχει επίσης την επιλογή να λαμβάνει τα δεδομένα ταξινομημένα με δύο τρόπους:

- Με βάση το όνομα της φόρμας. Στην περίπτωση αυτή δημιουργείται ένας φάκελος για κάθε φόρμα που πληροί τα κριτήρια που δόθηκαν από τον χρήστη. Μέσα σε κάθε φάκελο υπάρχει μία εικόνα, μορφής Bitmap, για κάθε στοιχείο της φόρμας με όνομα του κάθε στοιχείου.
- Με βάση το όνομα του στοιχείου. Στην περίπτωση αυτή δημιουργείται ένας φάκελος για κάθε στοιχείο, με τα αντίστοιχα ονόματα. Μέσα στο φάκελο υπάρχει μία εικόνα του αντίστοιχου κριτηρίου για κάθε φόρμα που πληροί τα κριτήρια, μορφής Bitmap, με το όνομα της φόρμας.

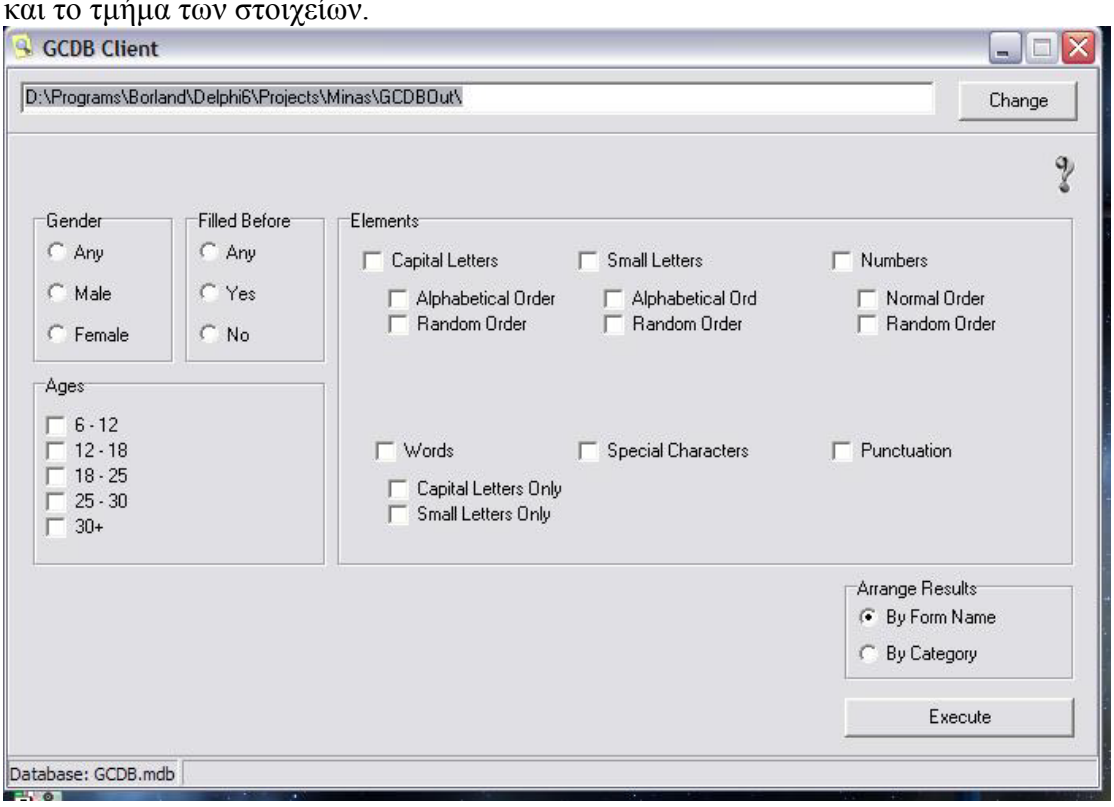
Τέλος θα πρέπει να σημειωθεί ότι για κάθε επερώτημα δημιουργείται ένας ξεχωριστός, μοναδικός φάκελος, για να αποθηκευτεί. Για να εξασφαλίσουμε τη μοναδικότητα του ονόματος του φακέλου χρησιμοποιούμε την τρέχουσα ώρα και ημερομηνία στο όνομα του.

Πρέπει επίσης να σημειωθεί ότι η εφαρμογή λειτουργεί διαφορετικά σε ότι αφορά τα επερωτήματα των λέξεων και των γραμμάτων/αριθμών. Επειδή συχνά οι χρήστες κάνουν επερωτήματα σε μόνο μία από τις 2 κατηγορίες (και όχι και στις 2 ταυτόχρονα) η εφαρμογή εκτελεί 2 διαφορετικά επερωτήματα κάθε φορά. Ένα για τις λέξεις και ένα για τα γράμματα/αριθμούς. Στην περίπτωση βέβαια που ο χρήστης επιλέξει μόνο τη μία από τις 2 κατηγορίες στοιχείων, λόγου χάρη μόνο λέξεις, τότε γίνεται ένα επερώτημα. Δύο επερωτήματα θα γίνουν στην περίπτωση που επιλέξει λέξεις και γράμματα/αριθμούς. Αυτό γίνεται για να αυξηθεί η ταχύτητα και η απόδοση της εφαρμογής.

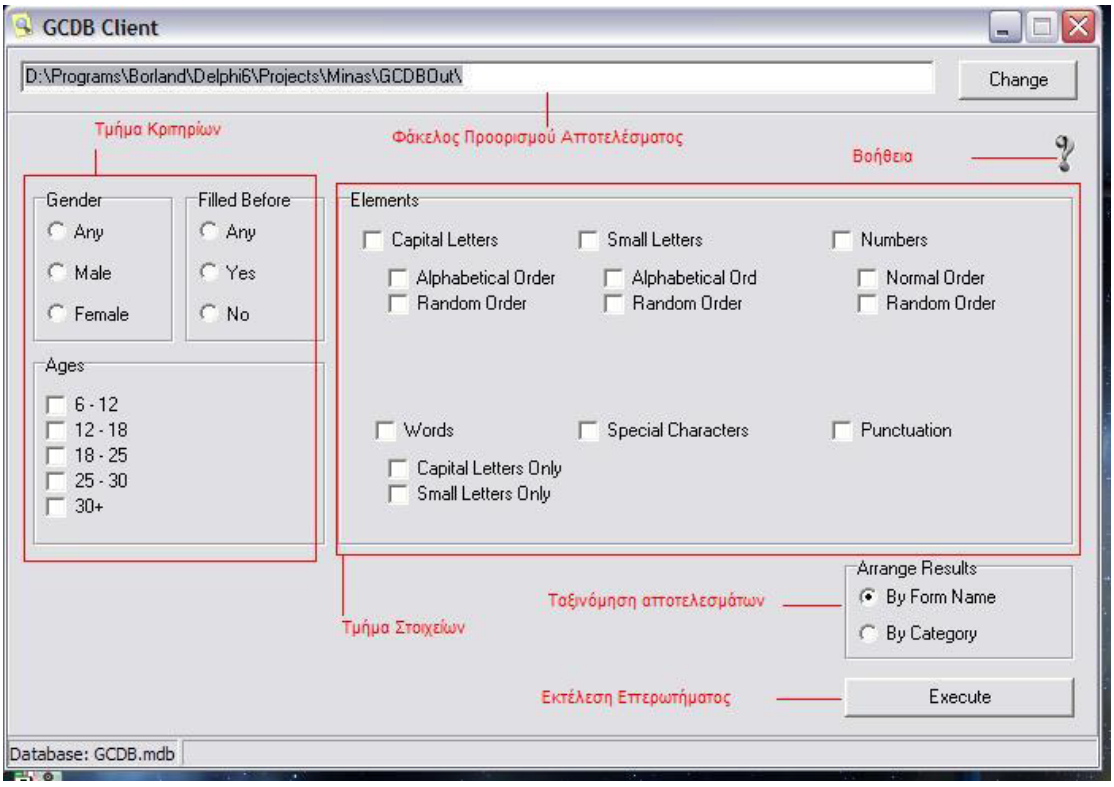
<u>Περιγραφή Γραφικού Περιβάλλοντος Εφαρμογής</u>

Το γραφικό περιβάλλον της εφαρμογής είναι ιδιαίτερα απλό. Στο πάνω μέρος βρίσκεται ο τρέχων επιλεγμένος κατάλογος και ένα κουμπί που δίνει τη δυνατότητα στο χρήστη να αλλάξει κατάλογο. Ακολουθεί το κυρίως μέρος της εφαρμογής, το

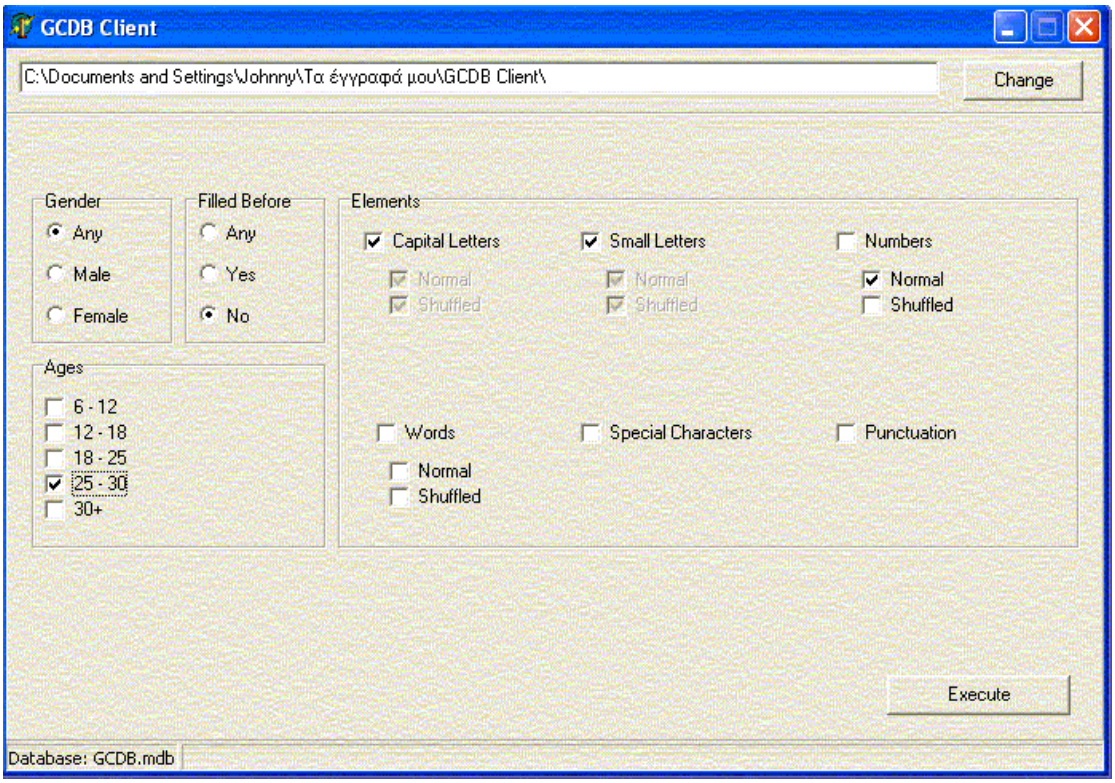
οποίο περιέχει τις επιλογές που πρέπει να κάνει ο χρήστης για να δημιουργηθεί το επερώτημα και το οποίο χωρίζεται σε δύο κυρίως τμήματα. Το τμήμα των κριτηρίων και το τμήμα των στοιχείων.



Το πρώτο τμήμα περιλαμβάνει τα κριτήρια με βάση των οποίων θα γίνει η επιλογή των φορμών. Τα κριτήρια αυτά σχετίζονται με τους συγγραφείς και αφορούν το φύλο, την ηλικία και αν έχει συμπληρώσει ξανά την ίδια φόρμα.



Το δεύτερο τμήμα αποτελεί τα τμήμα των στοιχείων και περιέχει όλα τα στοιχεία τα οποία καλείται να συμπληρώσει ένας συγγραφέας σε μία φόρμα. Ο χρήστης καλείται να διαλέξει για ποια από αυτά τα στοιχεία η εφαρμογή θα επιστρέψει δεδομένα (εικόνες). Λόγου χάρη μπορεί να επιλέξει μόνο κεφαλαία γράμματα σε τυχαία σειρά. Κάθε επιλογή του χρήστη προστίθεται στο επερώτημα με το λογικό AND. Επίσης αν επιλέξει μια κατηγορία ο χρήστης θα συμπεριληφθούν και όλες οι υποκατηγορίες.



Παράδειγμα επιλογής στοιχείων

Το επερώτημα που δημιουργείται από τις επιλογές του παραπάνω παραδείγματος είναι ως εξής:

```
SELECT   dbo.LETT_CAP_NORM.* AND dbo.LETT_CAP_SUFF.* AND
         dbo.LETT_SML_NORM.* AND dbo.LETT_SML_SUFF.* AND
         dbo.NUM_NORM.*
FROM     dbo.CONTAINS_CHAR INNER JOIN
         dbo.FORMS ON dbo.CONTAINS_CHAR.ID_FORM = dbo.FORMS.ID
         INNER JOIN
         dbo.FILLING ON dbo.FORMS.ID = dbo.FILLING.ID_FORM INNER
         JOIN
         dbo.WRITERS ON dbo.FILLING.ID_WRITER = dbo.WRITERS.ID
WHERE    dbo.FORMS.FILLED_BEFORE=No AND dbo.WRITERS.AGE=4
```

Στο τέλος υπάρχει η τελική επιλογή για την ταξινόμηση των αποτελεσμάτων και το κουμπί που θα δώσει εντολή στην εφαρμογή να εκτελέσει τα επερωτήματα και να επιστρέψει τα αποτελέσματα στο φάκελο που έχει ορίσει ο χρήστης.

Όπως και με την εφαρμογή εισαγωγής υπάρχει και εδώ αρχείο βοήθειας, στο οποίο μπορεί να έχει πρόσβαση ο χρήστης από το κουμπί του αγγλικού ερωτηματικού το οποίο βρίσκεται κάτω από το κουμπί αλλαγής φακέλου προορισμού.

ΤΕΧΝΙΚΟ ΠΑΡΑΡΤΗΜΑ:

ΟΔΗΓΙΕΣ ΕΠΕΚΤΑΣΗΣ ΕΦΑΡΜΟΓΗΣ

GCDB Έκδοση 1.0

Επέκταση ήδη υπάρχοντος πίνακα στη βάση:

Βάση Δεδομένων:

1. Προσθήκη πεδίων στον επιλεγμένο πίνακα.

Αρχείο INI:

1. Εντοπίστε την κατηγορία στην οποία υπάγεται ο πίνακας που έχετε αλλάξει. Οι κατηγορίες είναι οι ακόλουθες: [CRITERIA], [Letters], [Special Chars], [Numbers], [Punct], [Words]
2. Στο τέλος τυχών εγγραφών κάτω από την κατηγορία αυτή προσθέστε εγγραφές για πεδία που έχετε εισάγει στη βάση δεδομένων.

Θα χρειαστείτε 4 εγγραφές για κάθε πεδίο. Ακολουθήστε τον κανόνα: ONOMA_X1, ONOMA_Y1, ONOMA_X2, ONOMA_Y2. Είναι καλό όπου ONOMA να επιλέξετε ένα αλφαριθμητικό που να έχει σχετίζεται με το όνομα του πεδίου στη βάση.
Εισάγετε τις εγγραφές και την τιμή που αντιπροσωπεύει τις συντεταγμένες X1,Y2 και X2,Y2 στη φόρμα ακολουθώντας τον κανόνα: ONOMA_X1=<Τιμή X1>, ONOMA_Y1=<Τιμή Y1>, ONOMA_X2=<Τιμή X2>, ONOMA_Y2=<Τιμή Y2>, όπου <Τιμή ZZ> ένας ΑΚΕΡΑΙΟΣ αριθμός, που αντιπροσωπεύει την αντίστοιχη συντεταγμένη. (Ο αριθμός είναι σε pixels).

Πηγαίος Κώδικας:

1. Αλλαγή στην **procedure** CutLetters:

   a. Εύρεση του τμήματος που αντιστοιχεί στον πίνακα της βάσης δεδομένων.
   Υπάρχουν διάφοροι τρόποι να το εντοπίσει κάποιος. Ένας είναι τα σχόλια. Ένας δεύτερος είναι στα σύνολα των procedures CutElement να κοιτάξει το τελευταίο όρισμα. Τα TAdoDataSet components (που ουσιαστικά κάνουν το query με τη βάση δεδομένων) ακολουθούν τον κανόνα ονομασίας ado_<ONOMA_ΠΙΝΑΚΑ> (π.χ. Το ado_LETT_CAP_NORM αντιστοιχεί στον πίνακα LETT_CAP_NORM).

   b. Προσθήκη CutElement για το νέο στοιχείο του πίνακα.
   Η **procedure** CutElement Ορίζεται ως εξής: CutElement(Category_in_INI_File, Element_Name_in_INI_File, Field_Name_in_DB, TAdoDataSet_Name).
   Εισάγεται μία επιπλέον CutElement για κάθε καινούριο πεδίο του πίνακα, όπου:

   Category_in_INI_File : είναι η κατηγορία κάτω από τις οποίες βρίσκονται οι εγγραφές για τα πεδία που έχετε εισάγει στο Form Ini File.
   Element_Name_in_INI_File : είναι το όνομα της εγγραφής που έχετε εισάγει στο INI File. Σε καμία περίπτωση μην προσθέσετε τα _ZZ προσθέματα που υπάρχουν μετά το όνομα.

Field_Name_in_DB : Το όνομα του πεδίου στη βάση δεδομένων.
TAdoDataSet_Name : Το αντίστοιχο DataSet για τον πίνακα στον οποίο βρίσκονται τα νέα πεδία. Απλά αντιγράψτε το όνομα από τις προηγούμενες CutElement που ανήκουν στο ίδιο σύνολο.
ΠΡΟΣΟΧΗ: Οι CutElement που θα εισάγετε θα πρέπει να βρίσκονται **πριν** το **try** που βρίσκεται στον κώδικα κάτω από τις CutElement

Παράδειγμα:
Σαν παράδειγμα θα κάνουμε την προσθήκη 2 λέξεων με κεφαλαία γράμματα.
Βάση Δεδομένων:

- Προσθέτουμε στον πίνακα WORD_CAP δύο νέα πεδία, τα LEKSI1 και LEKSI2 τύπου OLE Object

Αρχείο INI:

- Βρίσκουμε την κατηγορία [WORDS].
- Στο τέλος προσθέτουμε τα εξής:
  LEKSI1_X1=1048
  LEKSI1_Y1=769
  LEKSI1_X2=1070
  LEKSI1_Y2=775
  LEKSI_2_X1=1250
  LEKSI_2_Y1=769
  LEKSI_2_X2=1290
  LEKSI_2_Y2=775

Όπως παρατηρείτε τα ονόματα στο Form INI αρχείο μπορούν διαφέρουν από τα ονόματα στα πεδία της βάσης.
Πηγαίος Κώδικας:

- Το τμήμα που αντιστοιχεί στον πίνακα για τις λέξεις με κεφαλαία γράμματα είναι το σύνολο με τις CutElement που το τελευταίο όρισμά τους είναι το ado_WORD_CAP
- Στο τέλος των ήδη υπαρχόντων εγγραφών προσθέτουμε μια εγγραφή για κάθε πεδίο.
  Προσθέτουμε:
  CutElement('Words','LEKSI1','LEKSI1',ado_WORD_CAP);
  CutElement('Words','LEKSI_2','LEKSI2',ado_WORD_CAP);

Δημιουργία νέου πίνακα στη βάση:

Βάση Δεδομένων:

1. Δημιουργία του νέου πίνακα.
   a. Προσθήκη ενός πεδίου ID που θα είναι autonumber και θα αποτελεί πρωτεύων κλειδί και index του πίνακα. Χρησιμοποιήστε κάποιον άλλο πίνακα ως αναφορά.
   ΠΡΟΣΟΧΗ: Το πεδίο ID θα πρέπει να είναι πάντα το **πρώτο** πεδίο του πίνακα.
   b. Προσθήκη των πεδίων εικόνων που θα είναι τύπου OLE Object.
2. Ενημέρωση αντίστοιχου ενδιάμεσου πίνακα. Υπάρχουν 2 ενδιάμεσοι πίνακες:

CONTAINS_WORD: για λέξεις.
CONTAINS_CHAR: για χαρακτήρες (συμπεριλαμβανομένου και αριθμών ή στοιχείων στίξης).

   a. Διαλέξτε τον ενδιάμεσο πίνακα που αντιστοιχεί στον πίνακα που δημιουργήσατε και προσθέστε ένα πεδίο με το όνομα: ID_<Όνομα πίνακα> όπου <Όνομα πίνακα> το όνομα του πίνακα που έχετε δημιουργήσει. Χρησιμοποιήστε ένα ήδη υπάρχον πεδίο του πίνακα ως αναφορά.
   b. Δημιουργήστε σχέση 1 προς πολλά ανάμεσα στο ID_<Όνομα πίνακα> και στο πεδίο ID του πίνακά σας. Χρησιμοποιείστε μια ήδη υπάρχουσα σχέση της βάσης ως αναφορά.

   Παρατήρηση: Τα queries του output εκτελούνται με βάση τους 2 αυτούς ενδιάμεσους πίνακες.

Αρχείο INI:

1. Αν πιστεύεται ότι ο πίνακας υπάγεται σε κάποια από τις ήδη υπάρχουσες κατηγορίες, εντοπίστε την. Ειδάλλως δημιουργήστε μια καινούρια κατηγορία. Οι ήδη υπάρχουσες κατηγορίες είναι: [CRITERIA], [Letters], [Special Chars], [Numbers], [Punct], [Words]
   Αν χρειαστεί να δημιουργήσετε κατηγορία ακολουθήστε τον κανόνα: [ΟΝΟΜΑ_ΚΑΤΗΓΟΡΙΑΣ] όπου ΟΝΟΜΑ_ΚΑΤΗΓΟΡΙΑΣ το όνομα που επιθυμείτε.
2. Στο τέλος τυχών εγγραφών κάτω από την κατηγορία αυτή προσθέστε εγγραφές για πεδία που έχετε εισάγει στη βάση δεδομένων.
   Θα χρειαστείτε 4 εγγραφές για κάθε πεδίο. Ακολουθήστε τον κανόνα: ΟΝΟΜΑ_Χ1, ΟΝΟΜΑ_Υ1, ΟΝΟΜΑ_Χ2, ΟΝΟΜΑ_Υ2. Είναι καλό όπου ΟΝΟΜΑ να επιλέξετε ένα αλφαριθμητικό που να έχει σχετίζεται με το όνομα του πεδίου στη βάση.
   Εισάγετε τις εγγραφές και την τιμή που αντιπροσωπεύει τις συντεταγμένες Χ1,Υ2 και Χ2,Υ2 στη φόρμα ακολουθώντας τον κανόνα: ΟΝΟΜΑ_Χ1=<Τιμή Χ1>, ΟΝΟΜΑ_Υ1=<Τιμή Υ1>, ΟΝΟΜΑ_Χ2=<Τιμή Χ2>, ΟΝΟΜΑ_Υ2=<Τιμή Υ2>, όπου <Τιμή ΖΖ> ένας ΑΚΕΡΑΙΟΣ αριθμός, που αντιπροσωπεύει την αντίστοιχη συντεταγμένη. (Ο αριθμός είναι σε pixels).

Πηγαίος Κώδικας:

1. Προσθήκη TAdoDataSet Component.
   a. Από τη μπάρα των components του Delphi επιλέξτε το tab ADO
   b. Κάντε κλικ σε ένα AdoDataSet Component και έπειτα κάντε κλικ οπουδήποτε πάνω στη φόρμα της εφαρμογής.
   c. Επιλέξτε το AdoDataSet Component που δημιουργήσατε και πηγαίνετε στον Object Inspector (με το F11 ή με το ποντίκι).

d. Στα properties του Object Inspector βρείτε τα:
   i. CommandText : Πατήστε το κουμπί με τις ... τελείες που βρίσκεται δίπλα του. Στο παράθυρο που θα ανοίξει δημιουργήστε ένα SQL Query που να καλεί όλα τα στοιχεία του πίνακα που δημιουργήσατε. Μπορείτε να το κάνετε είτε με τα κουμπιά και το inferface του παραθύρου που άνοιξε, είτε πληκτρολογώντας απευθείας στο χώρο SQL. (Το query σας θα πρέπει να είναι της μορφής: Select * from <Table_Name> Όπου <Table_Name> το όνομα του πίνακά σας. Πατήστε OK μόλις τελειώσετε.
   ii. Connection: Από το drop Down list επιλέξτε το db.
   iii. Name : Αλλάξτε το όνομα (το default όνομα είναι AdoDataSet1) σε ένα δικό σας. Είναι επιθυμητό να ακολουθήσετε τον κανόνα: ado_<Όνομα Πίνακα>, όπου <Όνομα Πίνακα> το όνομα του πίνακα που αντιστοιχεί το component.

2. Αλλαγή στην procedure CutLetters

Θα πρέπει να δημιουργήσετε μια δική σας ομάδα CutElement κλήσεων. Είναι επιθυμητό να προσθέσετε ένα σχόλιο στη αρχή της ομάδας που να υποδεικνύει σε πιο πίνακα αντιστοιχεί. Η ομάδας σας θα πρέπει ΟΠΩΣΔΗΠΟΤΕ να βρίσκεται πάνω από την ενημέρωση του αντίστοιχου ενδιάμεσου πίνακα στην οποία ανήκει. Τα σημεία στα οποία γίνονται οι ενημερώσεις των ενδιάμεσων πινάκων μπορούν να εντοπιστούν να εντοπιστούν από τα σχόλια:

//Δημιουργία ενδιάμεσου πίνακα CONTAINS_CHAR και //Δημιουργία ενδιάμεσου πίνακα CONTAINS_CHAR – End
//Δημιουργία ενδιάμεσου πίνακα CONTAINS_WORD και //Δημιουργία ενδιάμεσου πίνακα CONTAINS_WORD – End

a. Εντοπίστε την αρχή της δημιουργίας του ενδιάμεσου πίνακα στον οποίο υπάγεται ο πίνακας που δημιουργήσατε. Από πάνω εισάγεται του εισάγετε τον εξής κώδικα:

Παρατήρηση: Θεωρούμε ως ado_X το όνομα του TAdoDataSet Component που δημιουργήσατε ανωτέρω και το <CUTELEMENT PROCEDURES> το σύνολο των κλήσεων των CutElement Procedures που περιγράφεται παρακάτω.

```
if not ado_X.Active then ado_X.Active:=True;
ado_X.Insert;
<CUTELEMENT PROCEDURES>
try
  ado_X.Post;
except
  AppendLog('Error: Unable to insert Values in Database in table X.');
end;
pbCurrent.StepIt;
```

b. Δημιουργήστε τις CutElement Procedures για τα πεδία του πίνακά σας.
   Η procedure CutElement Ορίζεται ως εξής: CutElement(Category_in_INI_File, Element_Name_in_INI_File, Field_Name_in_DB, TAdoDataSet_Name).
   Εισάγεται μία επιπλέον CutElement για κάθε καινούριο πεδίο του πίνακα, όπου:

Category_in_INI_File : είναι η κατηγορία κάτω από τις οποίες βρίσκονται οι εγγραφές για τα πεδία που έχετε εισάγει στο Form Ini File.
Element_Name_in_INI_File : είναι το όνομα της εγγραφής που έχετε εισάγει στο INI File. Σε καμία περίπτωση μην προσθέσετε τα _ZZ προσθέματα που υπάρχουν μετά το όνομα.
Field_Name_in_DB : Το όνομα του πεδίου στη βάση δεδομένων.
TAdoDataSet_Name : Το αντίστοιχο DataSet για τον πίνακα στον οποίο βρίσκονται τα νέα πεδία. Απλά αντιγράψτε το όνομα από τις προηγούμενες CutElement που ανήκουν στο ίδιο σύνολο.
ΠΡΟΣΟΧΗ: Οι CutElement που θα εισάγετε θα πρέπει να βρίσκονται **πριν** το **try** που βρίσκεται στον κώδικα κάτω από τις CutElement

c. Έπειτα θα πρέπει να κάνετε και κάποιες προσθήκες στον κώδικα ενημέρωσης του ενδιάμεσου πίνακα. Θεωρούμε το ado_EP το όνομα που αντιστοιχεί στο AdoDataSet του ενδιάμεσου πίνακα που αντιστοιχεί στον πίνακα σας.
Ο κώδικας που υπάρχει στο αρχείο είναι της μορφής:

```
if not ado_EP.Active then ado_EP.Active:=True;
ado_EP.Insert;
ado_EP.Fields[0].Value:=ado_FORMS.Fields[0].Value;
ado_EP.Fields[1].Value:= ...
.
.
.
try
  ado_EP.Post;
except
  AppendLog('Error: Unable to insert Values in Database in table EP.');
end;
```

Πριν από το try και μετά από τις ό λ ες άλλες κλήσεις θα πρέπει να προσθέσετε μία ακόμα γραμμή:

```
ado_EP.Fields[Z].Value:=ado_X.Fields[0].Value;
```

Όπου Z ο αριθμός της σειράς που βρίσκεται το πεδίο που εισάγατε στον ενδιάμεσο πίνακα, αριθμώντας τα πεδία του ενδιάμεσου πίνακα από το 0. (Θεωρητικά θα πρέπει να είναι όσο ο αριθμός της αμέσως προηγούμενης κλήσης +1).
Και όπου ado_X το όνομα του TAdoDataSet component που δημιουργήσατε παραπάνω.

3. Ενημέρωση Progress Bar [Current]:
   a. Εντοπίστε την Procedure Tfmain.OpenForms;
   b. Εντοπίστε τον κώδικα:

```
//Progress Bars initialization
pbCurrent.Max:=X*10;
```

Παρατήρηση: Αν δεν έχουν γίνει αλλαγές, από default θα είναι:
pbCurrent.Max:=20*10;

c. Αυξήστε τον αριθμό Χ αριθμό του pbCurrent.Max κατά ένα. (Για κάθε νέο πίνακα ο αριθμός αυξάνεται κατά 1. -Ουσιαστικά ο πρώτος αριθμός είναι ίσος με το σύνολο των pbCurrent.StepIt; κλήσεων σε όλο το πρόγραμμα -)

Παράδειγμα:
Σαν παράδειγμα θα κάνουμε την προσθήκη ενός νέου πίνακα με το όνομα ENG_LETT_CAP_NORM με δύο πεδία.
Βάση Δεδομένων:

- Δημιουργία πίνακα ENG_LETT_CAP_NORM
  Δημιουργία πεδίων στον πίνακα:
  - Όνομα Πεδίου: ID [Πρωτεύων κλειδί]
    Τύπος: AutoNumber
    Field Size: Long Integer
    Indexed: Yes (No Dublicates)
    New Values: Increment
    Required: Yes
  - Όνομα Πεδίου: A
    Τύπος: OLE Object
    Required: No
  - Όνομα Πεδίου: B
    Τύπος: Ole Object
    Required: No
- Προσθήκη πεδίου στον ενδιάμεσο πίνακα: CONTAINS_CHAR
  - Όνομα Πεδίου: ID_ENG_LETT_CAP_NORM
  - Τύπος: Number (Long Integer)
  - Required: No
- Δημιουργία σχέσης 1 προς πολλά ανάμεσα στο CONTAINS_CHAR και ENG_LETT_CAP_NORM

Αρχείο INI:

- Βρίσκουμε την κατηγορία [Letters].
- Στο τέλος προσθέτουμε τα εξής:
  ENG_CAP_A_X1=570
  ENG_CAP_A_Y1=335
  ENG_CAP_A_X2=575
  ENG_CAP_A_Y2=340
  ENG_CAP_B_X1=580
  ENG_CAP_B_Y1=335
  ENG_CAP_B_X2=585
  ENG_CAP_B_Y2=340

Πηγαίος Κώδικας:

- Δημιουργούμε το ένα νέο AdoDataSet με
  - CommandText : Select * from ENG_LETT_CAP_NORM
  - Connection : db
  - Name : ado_ENG_LETT_CAP_NORM
- Δημιουργία ομάδας CutElements:

```
//Αγγλικά Κεφαλαία σε σειρά γράμματα
if not ado_ENG_LETT_CAP_NORM.Active then
  ado_ENG_LETT_CAP_NORM.Active:=True;
```

```
    ado_ENG_LETT_CAP_NORM.Insert;
    CutElement('Letters','ENG_CAP_A','A',ado_ENG_LETT_CAP_NORM);
    CutElement('Letters','ENG_CAP_B','B',ado_ENG_LETT_CAP_NORM);
    try
      ado_ENG_LETT_CAP_NORM.Post;
    except
      AppendLog('Error: Unable to insert Values in Database in table
ENG_LETT_CAP_NORM.');
    end;
    pbCurrent.StepIt;
```

- Ενημέρωση ενδιάμεσου πίνακα:
  Προσθέτουμε τη γραμμή:

```
    ado_CONTAINS_CHAR.Fields[9].Value:=ado_ENG_LETT_CAP_NORM.Field
s[0].Value;
```

Το τελικό αποτέλεσμα πρέπει να είναι:

```
    //Δημιουργία ενδιάμεσου πίνακα CONTAINS_CHAR
    if not ado_CONTAINS_CHAR.Active then
ado_CONTAINS_CHAR.Active:=True;
    ado_CONTAINS_CHAR.Insert;
    ado_CONTAINS_CHAR.Fields[0].Value:=ado_FORMS.Fields[0].Value;
    ado_CONTAINS_CHAR.Fields[1].Value:=ado_LETT_CAP_NORM.Fields[0].V
alue;
    ado_CONTAINS_CHAR.Fields[2].Value:=ado_LETT_CAP_SUFF.Fields[0].Val
ue;
    ado_CONTAINS_CHAR.Fields[3].Value:=ado_LETT_SML_NORM.Fields[0].V
alue;
    ado_CONTAINS_CHAR.Fields[4].Value:=ado_LETT_SML_SUFF.Fields[0].Val
ue;
    ado_CONTAINS_CHAR.Fields[5].Value:=ado_NUM_NORM.Fields[0].Value;
    ado_CONTAINS_CHAR.Fields[6].Value:=ado_NUM_SUFF.Fields[0].Value;
    ado_CONTAINS_CHAR.Fields[7].Value:=ado_CHAR_PUNCT.Fields[0].Value;
    ado_CONTAINS_CHAR.Fields[8].Value:=ado_CHAR_SPECIAL.Fields[0].Valu
e;
    ado_CONTAINS_CHAR.Fields[9].Value:=ado_ENG_LETT_CAP_NORM.Field
s[0].Value;
    try
      ado_CONTAINS_CHAR.Post;
    except
      AppendLog('Error: Unable to insert Values in Database in table
CONTAINS_CHAR.');
    end;
    //Δημιουργία ενδιάμεσου πίνακα CONTAINS_CHAR – End
```

- Ενημέρωση Progress Bar [Current]:

```
    pbCurrent.Max:=21*10;
```

ΟΔΗΓΙΕΣ ΔΗΜΙΟΥΡΓΙΑΣ ΚΑΙ ΑΛΛΑΓΗΣ ΦΟΡΜΑΣ

GCDB Έκδοση 1.0

Υπάρχει η δυνατότητα αλλαγής ή και ακόμα δημιουργίας εκ νέου φόρμας εισαγωγής. Απαραίτητη προϋπόθεση για να επεξεργαστεί το πρόγραμμα μια φόρμα είναι η ύπαρξη μιας έντονης παχιάς οριζόντιας μαύρης γραμμής που να διαχωρίζει τη φόρμα στα δύο.
Το πρόγραμμα εντοπίζει την πρώτη μαύρη γραμμή τέτοιου είδους και επιλέγει το πρώτο μαύρο pixel της (της πρώτης γραμμής pixel αν έχει πολλές γραμμές pixel η γραμμή μας) ως το σημείο (0,0). Δηλαδή το σημείο από το οποίο υπολογίζονται όλες οι συντεταγμένες.

Βήμα 1:
Κατασκευάστε τη φόρμα σας και τοποθετήστε την μαύρη γραμμή.

Βήμα 2:
Δημιουργήστε ένα αρχείο <Όνομα>.ini όπου <Όνομα> μπορείτε να βάλετε ότι εσείς επιθυμείτε.

Βήμα 3:
Ανοίξτε το αρχείο και γράψτε:

[Form]
BWThreshhold=X
AlignType=Y
Selection=Z

Όπου:

X: Το όριο σε pixels με βάση το οποίο θα γίνεται η μετατροπή της εικόνας από έγχρωμη σε ασπρόμαυρη. Το X λαμβάνει τιμές από το 0 έως το 255, όπου 0 το μαύρο και 255 το λευκό.
Ο αλγόριθμος που χρησιμοποιείται από την εφαρμογή μετατρέπει την εικόνα σε αποχρώσεις του γκρι και έπειτα όλα τα επίπεδα που είναι μικρότερα από το ThreshHold θα γίνουν Μαύρο, ενώ όλα όσα είναι μεγαλύτερα θα μετατραπούν σε λευκά.
Αν τεθεί τιμή -1 τότε ο αλγόριθμος μετατροπής της εικόνας σε ασπρόμαυρη επιλέγει την μέση τιμή όλων των επιπέδων της εικόνας.
Θέτοντας τιμή -2 τότε ο αλγόριθμος επιλέγει τιμή ThreshHold βάση ενός αλγορίθμου μέγιστης Εντροπίας (Maximum Entropy Algorith).

Y: Ο τρόπος με τον οποίο γίνεται ο υπολογισμός κλίσης της εικόνας της φόρμας από την εφαρμογή (Skew Detection). Το Y λαμβάνει τιμές από το 1 έως το 3.
Ο αλγόριθμος που χρησιμοποιούμε χρησιμοποιεί διαδοχικές περιστροφές της εικόνας (με κλίση 0.1 μοίρα ανά περιστροφή) μέχρι να εντοπίσει τον καλύτερο προσανατολισμό του εγγράφου.
Δίνοντας τιμή 1 χρησιμοποιείται ο αλγόριθμος σε ολόκληρη την εικόνα για να υπολογιστεί η γωνία κλίσης. [ΑΡΓΗ ΜΕΘΟΔΟΣ]

Δίνοντας τιμή 2 χρησιμοποιείται ο αλγόριθμος σε μίας μικρότερης κλίμακας εικόνα (thumbnail) της αρχικής εικόνας [ΓΡΗΓΟΡΗ ΜΕΘΟΔΟΣ]
Δίνοντας τιμή 3 χρησιμοποιείται ο αλγόριθμος στα πρώτα ZxIMAGE_WIDTH pixel της εικόνας (για ολόκληρη την εικόνα), όπου Z η παράμετρος του Selection [ΓΡΗΓΟΡΗ ΜΕΘΟΔΟΣ]
Ενδείκνυται η χρήση της παραμέτρου 3.

Z: Μία τιμή σε pixels για το πόσα εικονοστοιχεία (pixels) θα επιλεχθούν για την εκτέλεση του αλγόριθμου υπολογισμού κλίσης, αν επιλεχθεί η μέθοδος 3. ΔΕΝ Συνιστάτε να εισάγετε πολύ μικρές ή πολύ μεγάλες τιμές. Για μία εικόνα μεγέθους Α4 συνίστανται τιμές από 300 έως 600.

Παρατήρηση: Αν παραληφθούν κάποιες από τις παραπάνω παραμέτρους η ακόμα και ολόκληρο το τμήμα Form τότε η εφαρμογή λαμβάνει τις παρακάτω (default) τιμές:
BWThreshhold=-1
AlignType=3
Selection=500

Βήμα 4:
Στο αρχείο γράψτε:

[StartPoint]
Ratio=X
Lines=Y

Οι τιμές αυτές χρησιμοποιούνται από το πρόγραμμα για να εντοπίσει το πρώτο σημείο της μαύρης γραμμής και να το χρησιμοποιήσει σαν αρχή των συντεταγμένων. Ο αλγόριθμος που έχει υλοποιηθεί υπολογίζει για κάθε γραμμή την αναλογία μαύρων εικονοστοιχείων ως προς τα συνολικά εικονοστοιχεία (black pixel ratio (Black Pixels / Pixel Sum)
Αν κάποια γραμμή έχει αναλογία μεγαλύτερη ή ίση της τιμής X τότε είναι υποψήφια για τη μαύρη γραμμή.
Μόλις ο αλγόριθμος εντοπίσει Y ΣΥΝΕΧΟΜΕΝΕΣ υποψήφιες γραμμές (δηλαδή γραμμές που καλύπτουν την παραπάνω συνθήκη) τότε το πρώτο μαύρο εικονοστοιχείο που ακολουθείται από τουλάχιστον 10 μαύρα εικονοστοιχεία της πρώτης από τις Y γραμμές είναι το εικονοστοιχείο που αποτελεί τις αρχές των συντεταγμένων.

X: λαμβάνει τιμές από 0 έως 1 με 2 δεκαδικά στοιχεία. Η εφαρμογή χρησιμοποιεί έχει κατασκευαστεί για τα Ελληνικά δεδομένα, συνεπώς η υποδιαστολή γράφεται με κόμμα και όχι με τελεία.

Y: λαμβάνει ακέραιες θετικές τιμές.

Παρατήρηση: Αν παραληφθούν κάποιες από τις παραπάνω παραμέτρους η ακόμα και ολόκληρο το τμήμα StartPoint τότε η εφαρμογή λαμβάνει τις παρακάτω (default) τιμές:

Ratio=0,89
Lines=4

[Clean Isolated Pixels]
Enabled=X
Black=Y
White=Z

Το πρόγραμμα προτού εισάγει τις εικόνες στη βάση δεδομένων έχει τη δυνατότητα να κάνει καθαρισμό απομονωμένων εικονοστοιχείων (pixels).

X: λαμβάνει τιμές 0 και 1. Για 0 δεν γίνεται καθαρισμός των εικονοστοιχείων, για 1 γίνεται.

Y: λαμβάνει ακέραιες θετικές τιμές.

[Check Empty Parameters]
Letters=X
Words=Y
Punct=Z
Numbers=W
Special Chars=V

Το πρόγραμμα εισαγωγής έχει την δυνατότητα να ελέγχει το ποσοστό μαύρων εικονοστοιχείων προς τα συνολικά εικονοστοιχεία σε ένα πεδίο εικόνας το οποίο πρόκειται να εισαχθεί στη βάση. Αν το ποσοστό αυτό είναι χαμηλότερο του ποσοστού που αναφέρεται ανωτέρω τότε το πρόγραμμα θεωρεί ότι η εικόνα είναι κενή και δεν την εισάγει στη βάση. Κάθε κατηγορία έχει το δικό της ποσοστό.

X: λαμβάνει τιμές ανάμεσα στο 0 και 1. Αν εισαχθεί τιμή 0 τότε εισάγονται όλες οι εικόνες, ενώ αν εισαχθεί τιμή 1 τότε δεν εισάγεται καμία εικόνα. Αναφέρεται στις εικόνες γραμμάτων.

Y: λαμβάνει τιμές ανάμεσα στο 0 και 1. Αν εισαχθεί τιμή 0 τότε εισάγονται όλες οι εικόνες, ενώ αν εισαχθεί τιμή 1 τότε δεν εισάγεται καμία εικόνα. Αναφέρεται στις εικόνες λέξεων.

Z: λαμβάνει τιμές ανάμεσα στο 0 και 1. Αν εισαχθεί τιμή 0 τότε εισάγονται όλες οι εικόνες, ενώ αν εισαχθεί τιμή 1 τότε δεν εισάγεται καμία εικόνα. Αναφέρεται στις εικόνες σημείων στίξης.

W: λαμβάνει τιμές ανάμεσα στο 0 και 1. Αν εισαχθεί τιμή 0 τότε εισάγονται όλες οι εικόνες, ενώ αν εισαχθεί τιμή 1 τότε δεν εισάγεται καμία εικόνα. Αναφέρεται στις εικόνες αριθμών.

Y: λαμβάνει τιμές ανάμεσα στο 0 και 1. Αν εισαχθεί τιμή 0 τότε εισάγονται όλες οι εικόνες, ενώ αν εισαχθεί τιμή 1 τότε δεν εισάγεται καμία εικόνα. Αναφέρεται στις εικόνες ειδικών χαρακτήρων.

<u>Βήμα 5:</u>
Αφού πλέον έχουμε τελειώσει με τις τεχνικές πληροφορίες της φόρμας θα πρέπει να καταγραφούν οι συντεταγμένες των στοιχείων της φόρμας (γραμμάτων, λέξεων, αριθμών, κτλ). Τα στοιχεία αυτά θα πρέπει να συμφωνούν με τα πεδία της βάσης. Για κάθε στοιχείο θα πρέπει να εισαχθούν 2 ζεύγη συντεταγμένων. Θεωρώντας το στοιχείο ως ένα ορθογώνιο με γωνίες Α, Β, Γ, Δ:

Οι συντεταγμένες που απαιτούνται είναι αυτές των σημείων Α και Δ, ορίζοντας τις συντεταγμένες του Α ως X1,Y1 και του Δ ως X2,Y2
Οι ήδη υπάρχουσες κατηγορίες είναι: [Letters], [Words], [Punct], [Special Chars] και [Numbers], αλλά μπορείτε να δημιουργήσετε και δικές σας αν κρίνεται ότι αυτές δεν επαρκούν.
ΠΡΟΣΟΧΗ: Αν δημιουργήσετε δική σας κατηγορία θα πρέπει να την προσθέσετε στο [Check Empty Parameters] μαζί με την αντίστοιχη τιμή της.
Κάτω από την αντίστοιχη κατηγορία θα πρέπει να προσθέσετε τα στοιχεία της φόρμας την εξής μορφή:
<Όνομα>_X1=X
<Όνομα>_Y1=Y
<Όνομα>_X2=Z
<Όνομα>_Y2=W

Όπου <Όνομα> το όνομα του επιθυμείται να δώσετε στο στοιχείο σας και X=X1 συντεταγμένη, Y=Y1 συντεταγμένη, Z=X2 συντεταγμένη και W=Y συντεταγμένη

ΠΡΟΣΟΧΗ: Κάτω από την ίδια κατηγορία δεν μπορείτε να έχετε το ίδιο όνομα για 2 στοιχεία.

Βήμα 6:
Ανοίξτε την εφαρμογή GCDBInput. Πηγαίνετε στο μενού Options και επιλέξτε την επιλογή Form Options (ή πατήστε το πλήκτρο F5).
Στο νέο παράθυρο κάντε κλικ στο κουμπί Open και επιλέξτε το αρχείο που δημιουργήσατε.

ΠΡΟΣΟΧΗ: Το καινούριο αρχείο θα πρέπει να βρίσκεται στο φάκελο του προγράμματος (το φάκελο που περιέχει το αρχείο GCDBInput.exe). Αν το αρχείο είναι σε άλλο φάκελο τότε θα παρουσιαστεί πρόβλημα.

ΒΙΒΛΙΟΓΡΑΦΙΑ:

- Σύστημα αυτόματης επεξεργασίας εγγράφου και αναγνώρισης χειρόγραφων χαρακτήρων συνεχόμενης γραφής (Διδακτορική Διατριβή Εργίνας Καβαλλιεράτου)

Σχετικοί σύνδεσμοι:

http://www.iam.unibe.ch/~zimmerma/iamdb/iamdb.html
http://www.cedar.buffalo.edu/Databases/
http://unipen.nici.kun.nl/
http://www.nist.gov/srd/PDFfiles/nistsd19.pdf
ftp://sequoyah.ncsl.nist.gov/pub/databases/catalog.txt