Πολυτεχνείο Κρήτης

Σχεδιασμός και Υλοποίηση των Υποσυστημάτων Κατηγοριοποίησης & Αναζήτησης Εγγράφων και Διαχείρισης Διαγραμμάτων σε Εξατομικευμένα Συστήματα Νέων Κατ' Απαίτηση

Στέφανος Ι. Καρασαββίδης

Διπλωματική Εργασία που υποβλήθηκε
στα πλαίσια των απαιτήσεων
για την απόκτηση διπλώματος στο

Τμήμα Ηλεκτρονικών Μηχανικών
και Μηχανικών Υπολογιστών

Χανιά, Οκτώβριος 1998

Περίληψη

Η παρούσα εργασία περιγράφει το σχεδιασμό και την υλοποίηση των υποσυστημάτων κατηγοριοποίησης και αναζήτησης εγγράφων και διαχείρισης διαγραμμάτων χρηστών (User Profiles) του εξατομικευμένου συστήματος νέων κατ' απαίτηση «Hypermedia Custom News System». Σκοπός της ανάπτυξης των υποσυστημάτων ήταν η παροχή σύγχρονων υπηρεσιών νέων στους τελικούς χρήστες μέσω των τεχνολογιών εξατομικευμένων νέων και νέων κατ' απαίτηση, που απαιτούν τα σημερινά πληροφοριακά συστήματα στο διαδίκτυο. Η εργασία συμπληρώνει το περιβάλλον συγγραφής, διαχείρισης και παρουσίασης εγγράφων της εργασίας που παρουσιάστηκε στο [Πετρ98], για τη δημιουργία ενός ολοκληρωμένου συστήματος νέων.

Το Υποσύστημα Κατηγοριοποίησης Εγγράφων διαχειρίζεται τα λεξικά κατηγοριοποίησης των εγγράφων σύμφωνα με το μοντέλο κατηγοριοποίησης εγγράφων που παρουσιάστηκε αρχικά στην εργασία [Σκον98]. Το Υποσύστημα Αναζήτησης Εγγράφων προσφέρει ένα ολοκληρωμένο περιβάλλον αναζήτησης πληροφοριών που έχουν καταχωρηθεί στο σύστημα, επιτρέποντας τον υπολογισμό της σχετικότητας των αποτελεσμάτων σε σχέση με την υποβληθείσα ερώτηση. Το Υποσύστημα Διαχείρισης Διαγραμμάτων Χρηστών δίνει τη δυνατότητα παροχής εξατομικευμένων νέων στους τελικούς χρήστες.

Ιδιαίτερη έμφαση δόθηκε στην υποστήριξη αυθαίρετου αριθμού γλωσσών που αποτελεί ένα από τα σημαντικότερα χαρακτηριστικά του συστήματος. Συγκεκριμένα, παρέχεται η δυνατότητα αύξησης του αριθμού γλωσσών που υποστηρίζει το σύστημα χωρίς να χρειάζεται καμία αλλαγή σε κάποιο από τα υποσυστήματά του.

Στην εργασία αντιμετωπίστηκαν επίσης προβλήματα που έχουν σχέση με την πρόσβαση σε Βάση Δεδομένων ξεπερνώντας τους περιορισμούς που θέτει ένα δίκτυο υπολογιστών περιορισμένης πρόσβασης (Firewall), καθώς και προβλήματα αντιγραφής (replication) που δημιουργούνται στην προσπάθεια υποστήριξης πολλαπλών εξυπηρετητών.

Η υλοποίηση του συστήματος έγινε στη γλώσσα προγραμματισμού Java, που φιλοδοξεί να πραγματοποιήσει το όνειρο του «Write Once Run Everywhere» και να καλύψει όλα τα υπολογιστικά συστήματα.

Αφιέρωση

στον παππού μου που δεν πρόλαβε…

σ' αυτούς που νιώθουν
ότι οι αγωνίες
οι κόποι και οι θυσίες τους
βρίσκουν ανταπόκριση
με τις δικές μας επιτυχίες

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Σταύρο Χριστοδουλάκη για την επίβλεψή του στην ολοκλήρωση αυτής της διπλωματικής εργασίας και για τις εμπειρίες που μου πρόσφερε στα πλαίσια της εργασίας μου στο Ινστιτούτο Συστημάτων Πολλαπλών Μέσων (Multimedia Systems Institute of Crete – MUSIC) του Πολυτεχνείου Κρήτης.

Θα ήθελα να ευχαριστήσω προκαταβολικά τον καθηγητή κ. Βασίλειο Διγαλάκη και τον καθηγητή κ. Παναγιώτη Τριανταφύλλου για το χρόνο που διέθεσαν για την ανάγνωση της παρούσας διπλωματικής εργασίας και τις τυχόν παρατηρήσεις τους.

Επίσης ευχαριστώ τον Άλκη Σερβετά και τον Μανόλη Φραγκονικολάκη για την επίβλεψη και τη βοήθειά τους στο σχεδιασμό τμημάτων της διπλωματικής εργασίας.

Ιδιαίτερα θα ήθελα να ευχαριστήσω τον Θοδωρή Μαργαζά και τη Χρύσα Τσιναράκη για τον πολύτιμο χρόνο που αφιέρωσαν στην ανάγνωση της εργασίας και για τις χρήσιμες παρατηρήσεις τους, και ειδικά για τη συμπαράσταση και το ενδιαφέρον που έδειξαν. Επίσης ένα μεγάλο ευχαριστώ θα ήθελα να απευθύνω στο Μιχάλη Πετρόπουλο για την άριστη συνεργασία και συμβίωση στον ίδιο χώρο εργασίας.

Τέλος θα ήθελα να ευχαριστήσω τον για έξι χρόνια συγκάτοικο και φίλο Νίκο Χατζηχρυσάφη που μου συμπαραστάθηκε στις εύκολες και δύσκολες στιγμές των τελευταίων χρόνων.

Στέφανος Καρασαββίδης

Χανιά, Οκτώβριος 1998

Περιεχόμενα

1 Εισαγωγή

Τα τελευταία χρόνια η ανάπτυξη του Διαδικτύου ήταν, όπως αναμενόταν, ραγδαία. Στην ωρίμανση και την καθιέρωση του στην καθημερινή ζωή σημαντικό ρόλο έπαιξε η ανάγκη, ακόμα και του απλού πολίτη και όχι μόνο του εξειδικευμένου χρήστη, για πληρέστερη και αμεσότερη ενημέρωση. Ανάγκη την οποία καλούνται να καλύψουν τα συστήματα νέων.

Πρώτοι απ' όλους, στην τεχνολογία των συστημάτων νέων επένδυσαν οι παραδοσιακοί οργανισμοί που δραστηριοποιούνται στο χώρο της πληροφόρησης του κοινού, τα Μ.Μ.Ε. Εφημερίδες, περιοδικά, ραδιοφωνικοί και τηλεοπτικοί σταθμοί και γενικότερα όλες οι παραδοσιακές μορφές ενημέρωσης έκαναν την εμφάνιση τους στο Διαδίκτυο στα μέσα της δεκαετίας του 90, υπακούοντας στις εξελικτικές τάσεις της εποχής. Ακολούθησαν και άλλου τύπου εταιρείες ή οργανισμοί που απλά επιθυμούσαν να έχουν αμεσότερη επικοινωνία με τους πελάτες τους και το κοινό.

Τα πρώτα συστήματα νέων που έκαναν την εμφάνιση τους στο Διαδίκτυο χαρακτηρίζονταν από έλλειψη δυναμικά μεταβαλλόμενου περιεχομένου, εσωτερικής οργάνωσης, εύκολης αλληλεπίδρασης με τους αρθρογράφους, ενώ ο τελικός χρήστης είχε μόνο τη δυνατότητα απλής πλοήγησης μέσα σε ένα προκαθορισμένο περιβάλλον. Με λίγα λόγια, η διαδικασία συγγραφής και έκδοσης νέων στο Διαδίκτυο απείχε πολύ από τις συνηθισμένες μεθόδους παραγωγής νέων στα παραδοσιακά μέσα. Αυτές οι ανάγκες καλύπτονται με την ταυτόχρονη ανάπτυξη της τεχνολογίας σε διάφορους τομείς.

Το αυξημένο μέγεθος της πληροφορίας αναλαμβάνουν πλέον να οργανώσουν τα σύγχρονα συστήματα διαχείρισης βάσεων δεδομένων, που και αυτά πέρασαν με τη σειρά τους σε μια νέα φάση εξέλιξης, ακολουθώντας νέες κατευθύνσεις έρευνας και ανάπτυξης. Ενσωμάτωσαν μηχανισμούς επικοινωνίας με εξυπηρετητές του παγκόσμιου ιστού και φυλλομετρητές (browsers) για τη πλήρη εκμετάλλευση των δυνατοτήτων τους για την οργάνωση, αποθήκευση, ανάκτηση και έκδοση των παραγόμενων νέων, και πλέον αποτελούν τη βάση ανάπτυξης των συστημάτων πληροφόρησης στον παγκόσμιο ιστό.

Η γλώσσα προγραμματισμού Java, από την άλλη, παρέχοντας ανεξαρτησία πλατφόρμας εκτέλεσης, επέτρεψε την ανάπτυξη εφαρμογών ικανών να προσφέρουν αυξημένες δυνατότητες, τόσο στους αρθρογράφους όσο και στους τελικούς χρήστες.

Ξεπερνώντας έτσι το αρχικό στάδιο του πειραματισμού με το Διαδίκτυο, τα σημερινά συστήματα νέων προσπαθούν να προσαρμοστούν όσο το δυνατό καλύτερα στις ανάγκες των χρηστών, αλλά και των ανθρώπων που εργάζονται για την παραγωγή της πληροφορίας. Φθάνοντας έτσι στα σημερινά θέματα έρευνας και ανάπτυξης, αιχμή αποτελεί η τεχνολογία **εξατομίκευσης των νέων (custom news)**, καθώς και η τεχνολογία **νέων κατ' απαίτηση (news on demand)**. Η πρώτη έχει να κάνει με το είδος της πληροφορίας που θα παρουσιάζεται στον χρήστη, ενώ η δεύτερη έχει να κάνει με τον τρόπο που ο χρήστης θα βλέπει την πληροφορία αυτή.

Η τεχνολογία εξατομίκευσης των νέων επιτρέπει στον τελικό χρήστη να επιλέξει από τα υποσύνολα του μοντέλου κατηγοριοποίησης το περιεχόμενο της πληροφορίας που θα του παρέχεται, ανάλογα με τα ενδιαφέροντά του. Ο τελικός χρήστης μπορεί επίσης να καθορίσει την περίοδο ανανέωσης των νέων. Με τον τρόπο αυτό προσδιορίζει ακριβέστερα τις ειδήσεις που θέλει να του παρουσιάζονται από το σύστημα κάθε φορά που επιθυμεί να ενημερωθεί από αυτό.

Η τεχνολογία νέων κατ' απαίτηση επιτρέπει στον τελικό χρήστη να καθορίσει το χρόνο και τον τρόπο με τον οποίο θα παραλαμβάνει τα παραγόμενα νέα. Χρησιμοποιώντας κάποιο μηχανισμό ειδοποίησης έχει τη δυνατότητα να βρίσκεται χρονικά πλέον πολύ κοντά στην παραγωγή των νέων. Επίσης, μέσω ενός κατάλληλου εργαλείου μπορεί να οργανώνει την παρουσίαση των νέων με τρόπο που αυτός επιθυμεί.

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια της ανάπτυξης του συστήματος νέων **"Hypermedia Custom News System"**. Το σύστημα αυτό εσωκλείνει την τεχνολογία εξατομίκευσης των νέων και την τεχνολογία των νέων κατ' απαίτηση και έχει ως στόχο να ενσωματωθεί σε προγράμματα του Εργαστηρίου Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών Πολυμέσων (MU.S.I.C.) του τμήματος Η.Μ.Μ.Υ. του Πολυτεχνείου Κρήτης. Ειδικότερα, η εργασία επικεντρώνεται στην ανάπτυξη του **Υποσυστήματος Κατηγοριοποίησης Εγγράφων,** του **Υποσυστήματος Αναζήτησης Εγγράφων** καθώς και του

Υποσυστήματος Διαχείρισης Διαγραμμάτων Χρηστών (User Profiles).Η παρούσα εργασία συμπεριέλαβε ακόμη την επίλυση προβλημάτων που ανέκυψαν στην προσπάθεια πρόσβασης σε βάση δεδομένων που προστατεύεται από Firewall.

1.1 Αναγκαιότητα

Πολλά από τα προγράμματα που εκπονούνται στο MU.S.I.C. κινούνται στο χώρο των τουριστικών και πολιτιστικών συστημάτων πληροφοριών και έχουν ανάγκη από συστήματα νέων, τα οποία πρέπει να καλύπτουν όλες τις βασικές ανάγκες για παραγωγή και έκδοση νέων στο Διαδίκτυο και να υλοποιούν τεχνολογίες αιχμής επιθυμητές από τους τελικούς χρήστες. Επίσης, από τη φύση των προγραμμάτων επιβάλλεται η υποστήριξη αυθαίρετου αριθμού γλωσσών.

Η ανάγκη αυτή δεν μπορούσε να καλυφθεί με την αγορά ενός έτοιμου προϊόντος, γιατί οι δυνατότητες που προσφέρουν τα συστήματα αυτά είναι πολύ περιορισμένες. Συστήματα με αυξημένες δυνατότητες υπάρχουν στο Διαδίκτυο, αλλά έχουν αναπτυχθεί κυρίως για συγκεκριμένα ειδησεογραφικά πρακτορεία και σε συνεργασία με εταιρείες ανάπτυξης λογισμικού. Όλα τα έτοιμα συστήματα νέων που κυκλοφορούν είναι κλειστής αρχιτεκτονικής και δεν επιδέχονται αλλαγές από τρίτους.

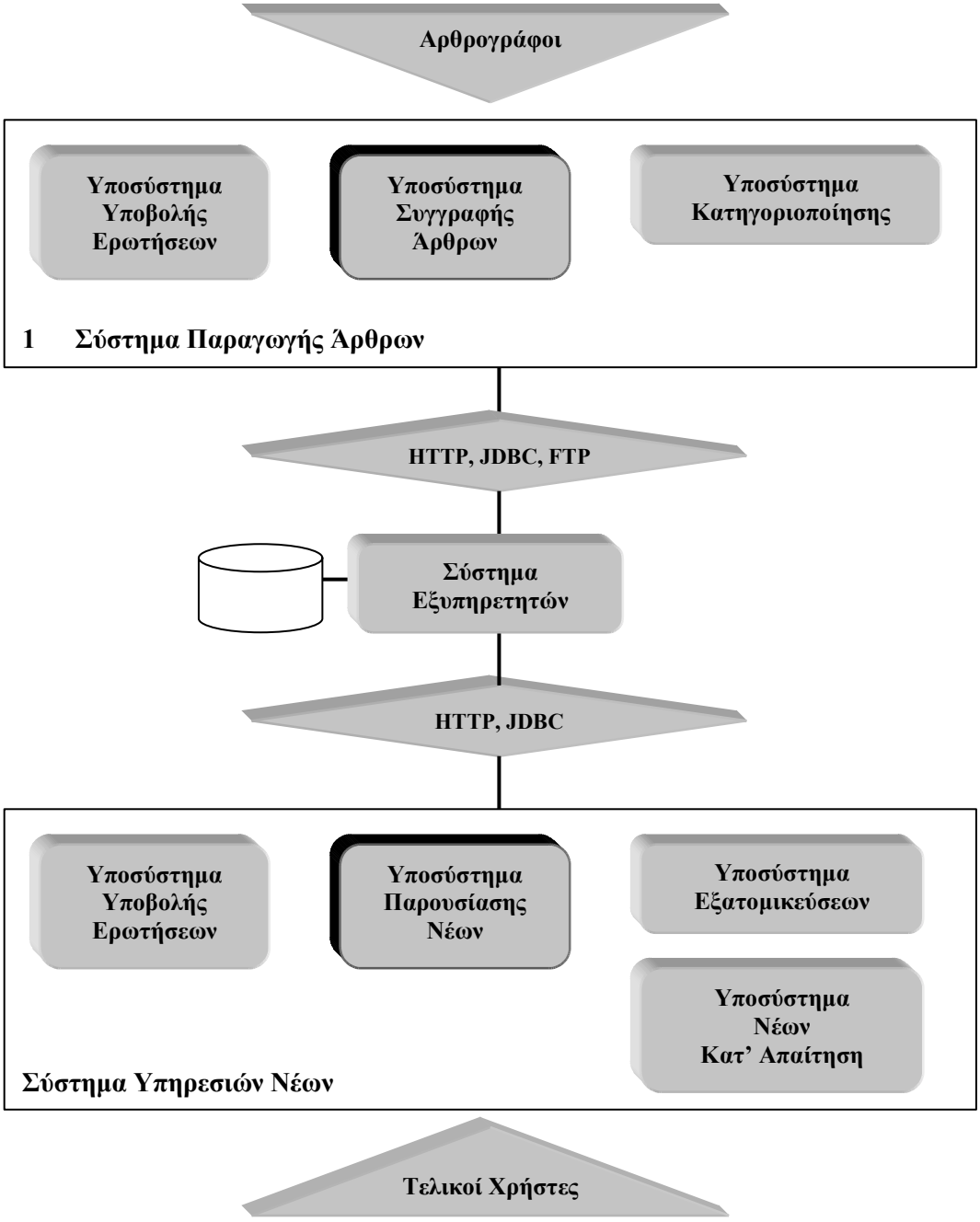
Το MU.S.I.C. έχει επίσης συμβάλλει στην ανάπτυξη ενός συστήματος εξατομικευμένων νέων κατ' απαίτηση, στα πλαίσια του ευρωπαϊκού ερευνητικού προγράμματος HyNoDe, αλλά χωρίς να έχει δικαίωμα χρήσης του όλου συστήματος. Οπότε η ανάπτυξη ενός συστήματος νέων από το ίδιο το MU.S.I.C. κρίθηκε επιβεβλημένη με ταυτόχρονη αναβάθμιση της τεχνολογίας υλοποίησης. Αυτή ήταν η αναγκαιότητα και ο σκοπός της δημιουργίας του συστήματος νέων **"Hypermedia Custom News System"**.

Οι παραπάνω διαπιστώσεις υπογραμμίζουν την αναγκαιότητα του συστήματος νέων **"Hypermedia Custom News System",** και παρουσιάζουν τη χρησιμότητα και συνεισφορά του. Ιδιαίτερα τον καιρό αυτό όπου η δραστηριότητα που παρατηρείται στον τομέα συστημάτων νέων, καθώς και ο μελλοντικά διαγραφόμενος περιορισμός των έντυπων μέσων ενημέρωσης, καθιστούν αναγκαία και πολύτιμη την απόκτηση της τεχνογνωσίας που απαιτεί ο σχεδιασμός και η υλοποίηση ενός τέτοιου συστήματος.

1.2 Το σύστημα “Hypermedia Custom News System”

Το Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών Πολυμέσων (MU.S.I.C.) συμμετέχει σε ερευνητικά προγράμματα της Ευρωπαϊκής Ένωσης με κυριότερα πεδία δράσης τον τουρισμό και τον πολιτισμό. Η ανάπτυξη των προγραμμάτων αυτών περιλαμβάνει και συστήματα νέων για την ενημέρωση των χρηστών τους.

Στόχος του συστήματος **“Hypermedia Custom News System”** είναι η δημιουργία ενός ολοκληρωμένου συστήματος νέων που θα ικανοποιεί τις κοινές απαιτήσεις των προγραμμάτων αυτών και χρησιμοποιεί τις τελευταίες καινοτομίες στον τομέα των συστημάτων νέων. Θα απευθύνεται στο σύγχρονο και απαιτητικό χρήστη του Διαδικτύου και στους ανθρώπους που εργάζονται για τη δημιουργία των νέων. Επίσης η υποστήριξη πολλαπλών γλωσσών είναι πολύ σημαντική, αφού το κοινό του συστήματος αναμένεται να προέρχεται από πολλές διαφορετικές χώρες.



Σχήμα 1-1 Γενική Αρχιτεκτονική του συστήματος εξατομικευμένων νέων κατ' απαίτηση "Hypermedia Custom News System"

Στο Σχήμα 1-1 απεικονίζεται γραφικά, σε υψηλό επίπεδο αφαίρεσης, η αρχιτεκτονική του συστήματος **"Hypermedia Custom News System"**. Χωρίζεται σε τρία κύρια τμήματα, τα οποία περιγράφονται περιληπτικά παρακάτω.

- Το **Σύστημα Παραγωγής Άρθρων** που είναι υπεύθυνο για την παραγωγή των νέων και περιλαμβάνει τρία υποσυστήματα. Μέσω του *Υποσυστήματος Συγγραφής Άρθρων* οι αρθρογράφοι έχουν τη δυνατότητα δημιουργίας, μορφοποίησης και διαχείρισης των άρθρων τους. Το *Υποσύστημα Κατηγοριοποίησης* τους παρέχει τη

δυνατότητα κατάταξης των άρθρων και αντικειμένων πολυμέσων (εικόνες, βίντεο κ.α.) στο μοντέλο κατηγοριοποίησης του συστήματος, ενώ το *Υποσύστημα Υποβολής Ερωτήσεων* τους βοηθά να εντοπίσουν υπάρχοντα άρθρα ή πολυμέσα που είναι αποθηκευμένα στο σύστημα για τον εμπλουτισμό του παραγόμενου άρθρου τους.

- Το **Σύστημα Εξυπηρετητών** αποτελείται από ένα σύνολο εξυπηρετητών αναγκαίων για τη σωστή και καλή λειτουργία του συστήματος. Συγκεκριμένα, περιλαμβάνει έναν εξυπηρετητή βάσης δεδομένων, του οποίου ο ρόλος είναι η οργανωμένη αποθήκευση, ταξινόμηση και ανάκτηση των δεδομένων, έναν εξυπηρετητή παγκόσμιου ιστού και έναν εξυπηρετητή μεταφοράς αρχείων. Με αυτούς αλληλεπιδρά το Σύστημα Παραγωγής Άρθρων για τη δημιουργία, αποθήκευση και ανάκτηση των άρθρων που δημιουργούνται από τους Αρθρογράφους, αλλά και το Σύστημα Υπηρεσιών Νέων για την ανάκτηση δεδομένων αναγκαίων για την ικανοποίηση των αιτήσεων των Τελικών Χρηστών. Τα δύο αυτά συστήματα επικοινωνούν με το Σύστημα Εξυπηρετητών μέσω των πρωτοκόλλων HTTP (Hypertext Transfer Protocol), JDBC (Java Database Connectivity) και FTP (File Transfer Protocol).

- Το **Σύστημα Υπηρεσιών Νέων** που είναι υπεύθυνο για την παροχή υπηρεσιών νέων στους τελικούς χρήστες και περιλαμβάνει και αυτό τρία υποσυστήματα. Μέσω του *Υποσυστήματος Παρουσίασης Νέων* οι Τελικοί Χρήστες έχουν τη δυνατότητα πλοήγησης στο μοντέλο κατηγοριοποίησης του συστήματος και επισκόπησης των άρθρων που περιέχονται σε αυτό. Το *Υποσύστημα Υποβολής Ερωτήσεων* τους επιτρέπει να αναζητήσουν άρθρα του συστήματος βάσει παραμέτρων που περιγράφουν τα ενδιαφέροντα τους. Το υποσύστημα αυτό είναι ουσιαστικά κοινό με αυτό του Συστήματος Παραγωγής Άρθρων, αλλά διαφέρει ο τόπος που εκτελεί τις διεργασίες του, δηλαδή στην πλευρά του χρήστη ή στην πλευρά του εξυπηρετητή. Το *Υποσύστημα Εξατομικεύσεων* είναι υπεύθυνο για τη διαχείριση των διαγραμμάτων που δημιουργούν οι Τελικοί Χρήστες και συνεργάζεται με το Υποσύστημα Παρουσίασης Νέων και το Υποσύστημα Υποβολής Ερωτήσεων για την εφαρμογή των προτιμήσεων που έχουν αποθηκευτεί στα διαγράμματα. Τέλος, το *Υποσύστημα Νέων Κατ' Απαίτηση* είναι υπεύθυνο για την ειδοποίηση του τελικού χρήστη, μέσω ειδικής εφαρμογής που

τρέχει στο σταθμό εργασίας του, για την έκδοση νέων άρθρων που ταιριάζουν με το διάγραμμά του.

1.3 Στόχοι της διπλωματικής εργασίας

Στόχος της παρούσας διπλωματικής εργασίας είναι η δημιουργία ενός ολοκληρωμένου περιβάλλοντος συγγραφής και διαχείρισης εγγράφων, καθώς επίσης και η υλοποίηση σύγχρονων υπηρεσιών νέων στους τελικούς χρήστες του συστήματος νέων **«Hypermedia Custom News System»** μέσω των τεχνολογιών εξατομικευμένων νέων και νέων κατ' απαίτηση.

Η εργασία επικεντρώνεται στα υποσυστήματα αναζήτησης και κατηγοριοποίησης εγγράφων και στο υποσύστημα διαχείρισης των διαγραμμάτων χρηστών, και συμπληρώνει την εργασία του [Πετρ98] για την ολοκλήρωση του συστήματος νέων. Επίσης επεκτείνεται στην επικοινωνία με βάση δεδομένων μέσα από σύστημα προστασίας δικτύου υπολογιστών (Firewall).

1.4 Δομή της διπλωματικής εργασίας

Στο κεφάλαιο 2 της διατριβής αναφέρονται αναλυτικά οι βασικές αρχές των σημερινών συστημάτων νέων και η κατηγοριοποίηση που γίνεται σύμφωνα με αυτές. Επίσης, γίνεται μια επισκόπηση των υπαρχόντων συστημάτων νέων στο Διαδίκτυο και των τεχνολογιών που χρησιμοποιούν, τόσο για το διεθνή όσο και για τον ελληνικό χώρο.

Στο κεφάλαιο 3 καταγράφεται η τεχνολογία που χρησιμοποιήθηκε για την υλοποίηση των τμημάτων του συστήματος, με κυριότερη έμφαση στη γλώσσα προγραμματισμού και την τεχνολογία Java.

Το κεφάλαιο 4 αποτελεί αναλυτική περιγραφή της αρχιτεκτονικής του συστήματος. Περιγράφει τους μηχανισμούς και τα τμήματα που περιλαμβάνει κάθε υποσύστημα, τη λειτουργικότητα τους και τον τρόπο αλληλεπίδρασής τους.

Στο κεφάλαιο 5 αναλύεται το μοντέλο της σχεσιακής βάσης δεδομένων που χρησιμοποιείται για την υποστήριξη του συστήματος και υλοποιείται από κατάλληλο Σύστημα Διαχείρισης Βάσεων Δεδομένων.

Στο κεφάλαιο 6 περιγράφεται το Υποσύστημα Κατηγοριοποίησης Εγγράφων που χρησιμοποιείται από τους αρθρογράφους για την κατάταξη των εγγράφων και από τους τελικούς χρήστες για την αναζήτηση επιθυμητών πληροφοριών.

Το κεφάλαιο 7 αναφέρεται στο Υποσύστημα Αναζήτησης Εγγράφων. Αναλύονται όλοι οι μηχανισμοί και λειτουργίες που αναπτύχθηκαν για την αναζήτηση των εγγράφων, και πως αυτοί χρησιμοποιούνται τόσο από τους τελικούς χρήστες, όσο και από τους αρθρογράφους. Παρουσιάζεται επίσης ο μηχανισμός για την εξαγωγή της σχετικότητας των εγγράφων σε σχέση με την υποβληθείσα ερώτηση.

Στο κεφάλαιο 8 παρουσιάζεται η διαχείριση των Διαγραμμάτων Χρηστών (User Profiles) και πως χρησιμοποιούνται στην αναζήτηση εγγράφων.

Στο κεφάλαιο 9 περιγράφεται ένα μοντέλο για την υποστήριξη πολυγλωσσικού περιεχομένου σε βάσεις δεδομένων και τα προβλήματα που συναντώνται στη χρησιμοποίηση κειμένων όταν απαιτείται η μεταφορά τους σε ακαθόριστο περιβάλλον παρουσίασης και μετατροπής.

Στο κεφάλαιο 10 περιγράφεται η πρόσβαση σε βάση δεδομένων διαπερνόντας το προστατευτικό σύστημα ενός εσωτερικού δικτύου υπολογιστών (Firewall) δια μέσω εξυπηρετητών παγκόσμιου ιστού.

Στο κεφάλαιο 11 θα αναλυθούν οι λόγοι που απαίτησαν την γενίκευση των μηχανισμών πρόσβασης σε βάση δεδομένων σε ένα νέο αντικείμενο χειρισμού συνδέσεων με βάση δεδομένων.

Το κεφάλαιο 122 αποτελεί σύντομη ανακεφαλαίωση της παρούσας διπλωματικής εργασίας και αναφέρεται στις μελλοντικές επεκτάσεις που μπορούν να ενσωματωθούν στο σύστημα.

2 Επισκόπηση των Υπαρχόντων Συστημάτων Νέων

Στο κεφάλαιο αυτό αναφέρονται συνοπτικά οι βασικές αρχές που διέπουν τα σημερινά συστήματα νέων στο Διαδίκτυο και οι βασικές κατηγορίες στις οποίες αυτά χωρίζονται. Στη συνέχεια περιγράφονται τα πιο δημοφιλή από τα υπάρχοντα συστήματα νέων καταγράφοντας τις δυνατότητες που παρέχουν στους χρήστες τους και σε ποια κατηγορία κατατάσσονται σύμφωνα με αυτές. Τα συστήματα αυτά προέρχονται τόσο από το διεθνή όσο και από τον ελληνικό χώρο.

2.1 Εισαγωγή στα Συστήματα Νέων

Τα συστήματα νέων στο Διαδίκτυο είναι μια νέα μορφή ενημέρωσης, που σκοπό έχει τη δυναμική, αλληλεπιδραστική και πιο ολοκληρωμένη ενημέρωση του κοινού από οποιαδήποτε άλλη μορφή πληροφόρησης (π.χ. τηλεόραση, ραδιόφωνο, εφημερίδες, περιοδικά). Η πρόσβαση στα συστήματα αυτά είναι δυνατή από οποιοδήποτε σημείο του κόσμου οποιαδήποτε στιγμή, διευρύνοντας έτσι το κοινό των παραδοσιακών μέσων και καθιστώντας δυνατή την ενημέρωση χρηστών απομακρυσμένων από την πηγή παραγωγής των ειδήσεων, χωρίς κανένα χρονικό περιορισμό. Επίσης η δυνατότητα οργάνωσης της πληροφορίας στα συστήματα αυτά επέτρεψε τη δημιουργία ειδησεογραφικών αρχείων στα οποία ο χρήστης μπορεί να ανατρέξει οποιαδήποτε στιγμή.

Ικανοποιώντας αρχικά τις παραπάνω αρχές, τα συστήματα νέων συνεχίζουν να ενσωματώνουν νέες τεχνολογίες και δυνατότητες, αφού με την πάροδο του χρόνου σημειώθηκε σημαντική πρόοδος και στον τομέα των δικτύων υπολογιστών και των τηλεπικοινωνιών. Μηχανισμοί αναζήτησης και κατάταξης των αποτελεσμάτων εφαρμόζονται στα ειδησεογραφικά αρχεία έτσι ώστε ο χρήστης να εντοπίζει την πληροφορία που τον ενδιαφέρει πολύ πιο γρήγορα. Τα άρθρα εμπλουτίζονται με δεδομένα πολυμέσων (π.χ. ήχος, video) και συσχετίζονται με άλλα άρθρα με σκοπό την πληρέστερη κάλυψη μιας είδησης. Ο χρήστης έχει τη δυνατότητα αλληλεπίδρασης με το σύστημα και καθορισμού των προσωπικών του προτιμήσεων (personalization), με σκοπό τη μείωση του χρόνου πλοήγησης και αναζήτησης πληροφορίας μέσα στο σύστημα. Δυνατή είναι επίσης και η κατ' απαίτηση (on

demand) ενημέρωση των χρηστών με τη χρήση μηχανισμών ειδοποίησης και αποστολής δεδομένων που έχουν αναπτυχθεί γι' αυτό το σκοπό.

Από τη στιγμή που οι τελικοί χρήστες αλληλεπιδρούν με το σύστημα, νέες δυνατότητες αποκτούν και οι εταιρείες που παρέχουν τα νέα. Στατιστικές μελέτες και δειγματοληψίες πάνω στις προτιμήσεις των χρηστών είναι εφικτές και μπορούν να αποδειχθούν ιδιαίτερα κερδοφόρες κυρίως για διαφημιστικούς σκοπούς. Επίσης δυνατή είναι και η εφαρμογή διαδικασιών χρέωσης των τελικών χρηστών για την πρόσβαση στο σύστημα.

Όπως βλέπουμε από τα παραπάνω, τα συστήματα νέων αυξάνουν συνεχώς τις παρεχόμενες υπηρεσίες τους. Και θα μπορούσαν να το κάνουν με γρηγορότερο ρυθμό αν δεν υπήρχε ο περιορισμός ταχύτητας μετάδοσης δεδομένων πάνω από το Διαδίκτυο. Ένας άλλος περιορισμός είναι ότι τα συστήματα νέων δεν αποτελούν ακόμα σίγουρη επένδυση, αφού η απαιτούμενη τεχνογνωσία για την ανάπτυξη τους στοιχίζει ακριβά, αλλά και το ευρύ κοινό δεν έχει εξοικειωθεί ακόμα με την ιδέα της ηλεκτρονικής ενημέρωσης. Τέλος, περιοριστική είναι και η έλλειψη προστασίας των δικαιωμάτων αναπαραγωγής των νέων και γενικότερα η ελλιπής ασφάλεια στο Διαδίκτυο.

2.2 Κατηγορίες Συστημάτων Νέων

Σύμφωνα με τα χαρακτηριστικά που αναφέρθηκαν παραπάνω τα σημερινά συστήματα νέων χωρίζονται κυρίως σε τρεις κατηγορίες, οι οποίες προκύπτουν από τις δυνατότητες που παρέχουν στους αρθρογράφους και στους τελικούς χρήστες. Πρόκειται για

- τα συμβατικά συστήματα νέων
- τα εξατομικευμένα συστήματα νέων, και
- τα εξατομικευμένα συστήματα νέων κατ' απαίτηση.

Η κάθε μία από αυτές περιγράφεται λεπτομερέστερα παρακάτω.

2.2.1 Συμβατικά Συστήματα Νέων

Τα συστήματα που κατατάσσονται στην κατηγορία των συμβατικών συστημάτων νέων, διαθέτουν ειδησεογραφικό αρχείο οργανωμένο στη δευτερεύουσα μνήμη και όχι σε βάση δεδομένων. Οι δυνατότητες των εργαλείων συγγραφής είναι αρκετά περιορισμένες και δεν υποστηρίζεται η ταυτόχρονη πρόσβαση από πολλούς αρθρογράφους. Το μοντέλο αναπαράστασης εγγράφου περιέχει μόνο τα κυριότερα χαρακτηριστικά ενός άρθρου, συνήθως σε μορφή απλού κειμένου. Η χωρική και χρονική κατηγοριοποίηση που ακολουθεί το σύστημα, υλοποιείται και αυτή στη δευτερεύουσα μνήμη σε μορφή φακέλων (directories). Η παρουσίαση των άρθρων και της δομής κατηγοριοποίησης τους πραγματοποιείται μέσω στατικών HTML σελίδων, γεγονός που καθιστά δυνατή την εφαρμογή ενός έτοιμου μηχανισμού αναζήτησης, ο οποίος όμως κοστίζει ακριβά και πιθανότατα να μην ικανοποιεί πλήρως τις απαιτήσεις των χρηστών του συστήματος. Ο τελικός χρήστης για να εντοπίσει το άρθρο που επιθυμεί πρέπει να διασχίσει τη δομή κατηγοριοποίησης του συστήματος και να το επιλέξει από μια λίστα υπαρχόντων.

Τα συστήματα αυτά αποτελούν συνήθως μεταφορά των νέων ενός άλλου μέσου ενημέρωσης (π.χ. εφημερίδας) σε ηλεκτρονική μορφή. Ο ρυθμός ανανέωσης των νέων είναι αργός, ενώ δεν παρέχονται εξατομικευμένες υπηρεσίες στους τελικούς χρήστες. Τέλος οι περιορισμένες δυνατότητες που παρέχουν τα συστήματα αυτά στον τελικό χρήστη καθιστούν την εμπορική τους εκμετάλλευση σχεδόν αδύνατη.

2.2.2 Εξατομικευμένα Συστήματα Νέων

Τα εξατομικευμένα συστήματα νέων μπορούν να χαρακτηριστούν ως σύγχρονα αφού υλοποιούν τις περισσότερες από τις τεχνολογίες αιχμής σήμερα. Όχι μόνο ολοκληρωμένα έγγραφα αλλά και αντικείμενα πολυμέσων οργανώνονται, αποθηκεύονται και ταξινομούνται σε κάποιον εξυπηρετητή βάσης δεδομένων. Τα εργαλεία συγγραφής παρέχουν τη δυνατότητα εισαγωγής και διαχείρισης πολυμέσων μέσα στα άρθρα, υποστηρίζοντας έτσι ένα πληρέστερο μοντέλο αναπαράστασης εγγράφου. Διαθέτουν ακόμα μηχανισμούς ταξινόμησης και αναζήτησης των νέων, παρέχοντας στους αρθρογράφους τη δυνατότητα κατηγοριοποίησης, ανεύρεσης και συσχέτισης των άρθρων μεταξύ τους. Επιπρόσθετα το μοντέλο κατηγοριοποίησης

άρθρων και πολυμέσων είναι μεταβλητό επιτρέποντας έτσι την επέκταση του εύρους της πληροφορίας που καλύπτει το σύστημα.

Την παρουσίαση του συστήματος αναλαμβάνουν να εκτελέσουν δυναμικά προς τον τελικό χρήστη κατάλληλοι μηχανισμοί παρουσίασης. Πέρα από τη δυνατότητα αναζήτησης και κατάταξης εγγράφων στον τελικό χρήστη παρέχεται ένας μηχανισμός αποθήκευσης των προσωπικών του προτιμήσεων (personalization), μειώνοντας έτσι κατά πολύ τον απαιτούμενο χρόνο πλοήγησης στο σύστημα για την ανεύρεση της επιθυμητής πληροφορίας. Ο τελικός χρήστης όταν εισέρχεται στο σύστημα βλέπει τα αποτελέσματα του φιλτραρίσματος που έγινε πάνω στα δεδομένα του συστήματος σύμφωνα με τις δικές του προτιμήσεις.

Τα παραπάνω συστήματα, παρέχουν τη δυνατότητα ταυτόχρονης πρόσβασης σε περισσότερους του ενός αρθρογράφους, ακόμα και περισσότερους του ενός ειδησεογραφικούς οργανισμούς, αυξάνοντας έτσι το ρυθμό ανανέωσης των ειδήσεων δραστικά σε σχέση με τα συμβατικά συστήματα νέων. Τέλος λόγω των αυξημένων δυνατοτήτων που παρέχουν στους τελικούς χρήστες υπάρχουν περιθώρια και για εμπορική εκμετάλλευση.

2.2.3 Εξατομικευμένα Συστήματα Νέων Κατ' απαίτηση

Τα εξατομικευμένα συστήματα νέων μπορεί να έχουν υψηλό ρυθμό ανανέωσης των νέων, αλλά δεν παραδίδουν τη νέα πληροφορία στους τελικούς χρήστες μόλις αυτή είναι διαθέσιμη. Αυτή την ανάγκη έρχεται να καλύψει η τεχνολογία νέων κατ' απαίτηση (news on demand), που αποτελεί σήμερα την αιχμή της τεχνολογίας των συστημάτων νέων. Ο τελικός χρήστης, αφού ορίσει τις προσωπικές του προτιμήσεις, μπορεί να απαιτήσει από το σύστημα να λαμβάνει ή να ειδοποιείται για τα εισερχόμενα νέα που καλύπτουν τα ενδιαφέροντα του μέσω κάποιου κατάλληλου μηχανισμού. Με αυτόν τον τρόπο ο τελικός χρήστης βρίσκεται πολύ κοντά, χρονικά, στην παραγωγή των νέων. Επιπλέον, στην περίπτωση που ο ρυθμός ανανέωσης του συστήματος είναι αρκετά γρήγορος, μπορούμε να πούμε ότι παρέχεται ενημέρωση σχεδόν σε πραγματικό χρόνο.

Οι μηχανισμοί ειδοποίησης που παρέχονται στους τελικούς χρήστες είναι κυρίως δύο ειδών. Είτε μέσω ηλεκτρονικού ταχυδρομείου (e-mail), όπου το σύστημα στέλνει στο

χρήστη ολόκληρο το άρθρο ή μια περίληψη μαζί με τη διεύθυνση όπου μπορεί να βρει το πλήρες άρθρο, είτε με τη χρήση ειδικών εφαρμογών ειδοποίησης που ο τελικός χρήστης πρέπει να εγκαταστήσει στο μηχάνημά του. Ο τελευταίος αυτός μηχανισμός παρέχει αρκετή ευελιξία στο χρήστη, είναι πλήρως παραμετρικός και ιδιαίτερα απλός και βολικός στη χρήση.

Τέτοιου είδους συστήματα προορίζονται κυρίως για εμπορική εκμετάλλευση διότι έχουν τη δυνατότητα παροχής εξειδικευμένων υπηρεσιών ακόμα και για κρίσιμες εφαρμογές (π.χ. χρηματιστήριο, καιρός κλπ). Τα συστήματα αυτά έχουν ήδη αρχίσει να κάνουν την εμφάνιση τους στο Διαδίκτυο και αναμένεται να φανούν αναγκαία από όλο και περισσότερους χρήστες τα επόμενα χρόνια.

2.3 Υπάρχοντα Συστήματα Νέων

Ο σχεδιασμός και η υλοποίηση του συστήματος "Hypermedia Custom News System" βασίστηκε αρχικά στην εμπειρία συναδέλφων, οι οποίοι είχαν συμμετάσχει σε ευρωπαϊκά ερευνητικά προγράμματα σχετικά με το αντικείμενο, και ειδικότερα στο πρόγραμμα **HyNoDe** (Hypermedia News On Demand). Το πρόγραμμα προέβλεπε την κατασκευή ενός συστήματος εξατομικευμένων νέων κατ' απαίτηση, στα πλαίσια του οποίου το MU.S.I.C. είχε αναλάβει το σχεδιασμό και την ανάπτυξη του εξυπηρετητή πολυμέσων και την υλοποίηση των μηχανισμών επικοινωνίας με αυτόν. Τα παραδοτέα του HyNoDe ήταν ιδιαίτερα πολύτιμο υλικό για τον αρχικό σχεδιασμό και τη μοντελοποίηση του συστήματος, από τη στιγμή που περιέγραφαν τις προδιαγραφές όλων των τμημάτων του προγράμματος. Αλλά και στη συνέχεια, κατά την πιλοτική φάση λειτουργίας του προγράμματος, πολλές ιδέες πάρθηκαν για τον τρόπο παρουσίασης του συστήματος στον τελικό χρήστη.

Δύο άλλα μεγάλα διεθνή και πρωτοποριακά συστήματα νέων που μελετήθηκαν κατά τη διάρκεια ανάπτυξης της παρούσας διπλωματικής εργασίας ήταν το "CNN Interactive" και το "MSNBC". Και τα δύο αναπτύχθηκαν στις Ηνωμένες Πολιτείες και καλύπτουν ειδησεογραφικά όλο τον κόσμο.

Σχήμα 2-1 Σελίδα παρουσίασης εξατομικευμένων νέων του συστήματος "CNN Custom News"

Το **"CNN Interactive"** αποτελεί τη δικτυακή μορφή ενημέρωσης του γνωστού αμερικανικού τηλεοπτικού καναλιού. Πέρα από το κυρίως σύστημα περιέχει και τέσσερα υποσυστήματα, τα τρία από τα οποία αφορούν πολιτικές (CNN/TIME Allpolitics), οικονομικές (CNN finance) και αθλητικές (CNN Sports Illustrated) ειδήσεις, ενώ το τέταρτο έχει να κάνει με τη διαχείριση των προσωπικών προτιμήσεων των τελικών χρηστών και ονομάζεται **"CNN Custom News"**. Αναπτύχθηκε σε συνεργασία με την Oracle και υλοποιεί την τεχνολογία εξατομικευμένων νέων. Το σύστημα υποστηρίζει και την τεχνολογία νέων κατ' απαίτηση υλοποιώντας όμως μόνο το μηχανισμό παράδοσης νέων μέσω ηλεκτρονικού ταχυδρομείου.

Το **"MSNBC"** προέκυψε από τη συνεργασία της Microsoft και του αμερικανικού δικτύου NBC και είναι και αυτό εξατομικευμένο σύστημα νέων κατ' απαίτηση. Πέρα από το μηχανισμό παράδοσης νέων μέσω ηλεκτρονικού ταχυδρομείου, το σύστημα παρέχει και ειδικό μηχανισμό ειδοποίησης του τελικού χρήστη, ο οποίος εγκαθίσταται στο σταθμό εργασίας του και τον ενημερώνει ανά τακτά χρονικά διαστήματα.

Σχήμα 2-2 Ο μηχανισμός ειδοποίησης τελικού χρήστη του συστήματος "MSNBC"

Στον ελληνικό χώρο το μόνο σύστημα που ξεφεύγει από τα συμβατικά είναι αυτό της Ναυτεμπορικής, το οποίο υποστηρίζει εξατομικευμένα νέα και περιλαμβάνει μηχανισμό παραλαβής νέων μέσω ηλεκτρονικού ταχυδρομείου. Επίσης το σύστημα αθλητικής ενημέρωσης "Sportline" διαθέτει πολύ καλή παρουσίαση νέων, ενσωματώνοντας πολυμέσα μέσα στο κείμενο των άρθρων. Από τα συμβατικά συστήματα νέων ξεχωρίζουν το "Μακεδονικό Πρακτορείο Ειδήσεων" και το "Flash 9.61 Interactive". Και τα δύο παρέχουν πλήρη ενημέρωση και καλαίσθητη παρουσίαση.

Από τα συστήματα που αναφέρθηκαν παραπάνω, κανένα δεν έχει τη δυνατότητα να υποστηρίξει αυθαίρετο αριθμό γλωσσών. Τα διεθνή συστήματα χρησιμοποιούν μόνο

αγγλικά, ενώ τα ελληνικά έχουν συνήθως δύο διαφορετικές εκδόσεις σε αγγλικά και ελληνικά.

2.4 Ανακεφαλαίωση

Στο κεφάλαιο αυτό έγινε μια προσπάθεια επισκόπησης των συστημάτων νέων που υπάρχουν στο Διαδίκτυο. Όπως φάνηκε τα συστήματα νέων στο Διαδίκτυο, με την πάροδο του χρόνου, συγκεντρώνουν όλο και περισσότερο το ενδιαφέρον, τόσο των τελικών χρηστών όσο και των ειδησεογραφικών οργανισμών. Μεγάλοι και παραδοσιακοί ειδησεογραφικοί οργανισμοί έχουν επενδύσει στην πληροφόρηση των πελατών τους από το Διαδίκτυο και διαφαίνεται ότι οι προοπτικές εξέλιξής τους είναι πολύ καλές, καθώς υπάρχει αρκετό επιστημονικό και αγοραστικό ενδιαφέρον προς αυτή την κατεύθυνση.

3 Πλατφόρμα Υλοποίησης

Στο κεφάλαιο αυτό περιγράφεται η πλατφόρμα υλοποίησης του συστήματος «Hypermedia Custom News System».

Το *«Hypermedia Custom News System»* έχει υλοποιηθεί χρησιμοποιώντας μια σειρά από διαφορετικές τεχνολογίες και στηρίζεται σε ένα σύνολο εξυπηρετητών, με κομβικό σημείο τον εξυπηρετητή Βάσεων Δεδομένων.

Κύριο μέλημα κατά διάρκεια του σχεδιασμού του συστήματος, ήταν το σύνολο των τεχνολογιών που επρόκειτο να χρησιμοποιηθούν, να βασίζεται σε διεθνή και ανοιχτά πρότυπα, με στόχο αφενός την όσο γίνεται μεγαλύτερη απεξάρτηση από κατασκευαστές και παροχείς λογισμικού και αφετέρου την ελαχιστοποίηση, από πλευράς τελικών χρηστών, της ανάγκης εγκατάστασης ειδικού λογισμικού προκειμένου να χρησιμοποιήσουν το σύστημα.

Στην πλευρά του τελικού χρήστη, οι παραπάνω στόχοι οδήγησαν στους γνωστούς φυλλομετρητές ως βάση πάνω στην οποία προσφέρονται οι υπηρεσίες προς τους χρήστες. Επίσης, ως κεντρική γλώσσα προγραμματισμού των εφαρμογών, τόσο στην πλευρά του εξυπηρετητή όσο και στην πλευρά του τελικού χρήστη, επιλέχθηκε η Java, ενώ για την επικοινωνία με τα διάφορα υποσυστήματα επιλέχθηκαν τα γνωστά πρωτόκολλα HTTP και FTP.

Στα επόμενα υποκεφάλαια ακολουθεί μία σύντομη περιγραφή των προϊόντων και τεχνολογιών που χρησιμοποιήθηκαν στην ανάπτυξη του *«Hypermedia Custom News System»*.

3.1 Microsoft SQL Server και IBM DB2

Ο Εξυπηρετητής Βάσης Δεδομένων είναι ίσως το σημαντικότερο τμήμα του συστήματος καθώς είναι υπεύθυνο για την διασφάλιση των παρεχόμενων υπηρεσιών και για τη συνέπεια του συστήματος.

Ο Microsoft SQL Server στην έκδοση 6.5 που χρησιμοποιήθηκε για το «Hypermedia Custom News System» και ο IBM DB2 στην έκδοση 2 που χρησιμοποιήθηκε για τις

πρώτες εκδόσεις του συστήματος νέων, είναι ολοκληρωμένα συστήματα διαχείρισης σχεσιακών βάσεων δεδομένων, τα κύρια χαρακτηριστικά του οποίου είναι η επεξεργασία (processing) και βελτιστοποίηση (optimization) των ερωτήσεων ανάκτησης, οι συνδιαλλαγές (transactions), ο ταυτοχρονισμός (concurrency) και η ανάνηψη (recovery).

Στα υποσυστήματα που υλοποιήθηκαν στη διπλωματική εργασία, πέραν των τυποποιημένων λειτουργιών για πρόσβαση σε ένα ΣΣΔΒΔ, δεν χρησιμοποιήθηκαν κάποια ιδιαίτερα χαρακτηριστικά των συγκεκριμένων προϊόντων της Microsoft και της IBM, με αποτέλεσμα το όλο σύστημα να μπορεί να μεταφερθεί εύκολα σε οποιοδήποτε άλλο ΣΣΔΒΔ.

3.2 Φυλλομετρητές, HTTP και FTP πρωτόκολλα και εξυπηρετητές

Το Διαδίκτυο έχει στιγματιστεί από την επέλαση των φυλλομετρητών και την χρησιμοποίηση αυτών ως βάση για κάθε είδους υπηρεσία που προσφέρεται δια μέσω του Διαδικτύου.

Η επιλογή των φυλλομετρητών ως περιβάλλοντος για τη χρησιμοποίηση του *«Hypermedia Custom News System»,* αποτέλεσε φυσική εξέλιξη κατά τη διάρκεια του σχεδιασμού του συστήματος.

Η ευελιξία που προσφέρει η HTML στην παρουσίαση πληροφορίας, σε συνδυασμό με το καλαίσθητο αποτέλεσμα που μπορεί να επιτευχθεί χρησιμοποιώντας αυτή τη γλώσσα για την παρουσίαση των εγγράφων, κατέστησαν το δίδυμο φυλλομετρητή-HTML ιδανικό συνδυασμό όσον αφορά τη διάθεση των εγγράφων στους τελικούς χρήστες.

Η δυνατότητα που προσφέρουν οι σημερινής τεχνολογίας φυλλομετρητές για την ενσωμάτωση, σε κοινές σελίδες HTML, και εφαρμογών οι οποίες είναι ικανές να αντεπεξέλθουν και σε πιο απαιτητικές εργασίες, οδήγησαν στην επιλογή των φυλλομετρητών ως περιβάλλον εργασίας ακόμα και για τις διαδικασίες της συγγραφής των άρθρων, αλλά και για την αναζήτηση εγγράφων από τους τελικούς χρήστες.

Τα πρωτόκολλα επικοινωνίας HTTP (Hypertext Transfer Protocol) και FTP (File Transfer Protocol) που είναι υπεύθυνα για την μεταφορά HTML σελίδων και αρχείων αντίστοιχα, και υποστηρίζονται από όλους τους σημερινούς φυλλομετρητές, χρησιμοποιήθηκαν και για τις ανάγκες της συγγραφής και παρουσίασης των εγγράφων. Για την υποστήριξη αυτών των πρωτοκόλλων απαιτείται η εγκατάσταση αντίστοιχου εξυπηρετητή, η οποία όμως δεν παρουσιάζει κάποιο ιδιαίτερο πρόβλημα, επειδή είναι ευρέως διαδεδομένοι και δεν περιορίζονται από συγκεκριμένα λειτουργικά συστήματα.

Τέλος, η ευρεία διάδοση των φυλλομετρητών στους χρήστες υπολογιστών, εγγυάται τη δυνατότητα πρόσβασης στο σύστημα από όλους τους εν δυνάμει χρήστες του Διαδικτύου.

3.3 Java

Η γλώσσα προγραμματισμού και το περιβάλλον Java θα μπορούσαν να χαρακτηριστούν με τις ακόλουθες λέξεις: απλή, αντικειμενοστραφές, ενήμερη για δίκτυα (network-savvy), διερμηνόμενη (interpreted), στιβαρή (robust), ασφαλή, ουδέτερη αρχιτεκτονικής, μεταφερόμενη (portable), υψηλής ταχύτητας, πολυνηματική (multithreaded), δυναμική. Για κάθε έναν από τους προηγούμενος χαρακτηρισμούς, θα δοθεί μία μικρή εξήγηση στις παρακάτω παραγράφους.

- **απλή,** διότι παρόλο που μοιάζει πολύ με τη C++ έχουν απαλειφθεί πολλές ιδιομορφίες και χαρακτηριστικά της C++, που συνήθως δε χρησιμοποιούνται αλλά προκαλούν σύγχυση στον προγραμματιστή. Για παράδειγμα δεν υπάρχει υπερφόρτωση τελεστών (operator overloading) και πολλαπλή κληρονομικότητα (multiple inheritance). Επιπλέον έχει προστεθεί η αυτόματη συλλογή αχρείαστης μνήμης (automatic garbage collection), απαλείφοντας έτσι ένα από τα σημαντικότερα προβλήματα που συναντάει ο προγραμματιστής της C ή της C++.

- **αντικειμενοστραφές,** καθώς χρησιμοποιεί ουσιαστικά το ίδιο μοντέλο που χρησιμοποιεί η C++ για τον αντικειμενοστραφή χαρακτηρισμό της.

- **ενήμερη για δίκτυα,** καθώς έχει σχεδιαστεί με δεδομένο τη χρησιμοποίηση των εφαρμογών σε δικτυακά περιβάλλοντα. Η γλώσσα έχει ενσωματωμένες πολλές

λειτουργίες και ρουτίνες για επικοινωνία σε δίκτυα υπολογιστών, καθιστώντας έτσι την πρόσβαση σε υπηρεσίες του δικτύου το ίδιο εύκολη με την πρόσβαση σε υπηρεσίες του τοπικού υπολογιστή.

- **διερμηνόμενη,** διότι μεταφράζεται σε γλώσσα μηχανής τη στιγμή της εκτέλεσης

- **στιβαρή,** διότι προνοεί για τον έλεγχο λαθών από τα πρώτα στάδια της ανάπτυξης μέχρι και το στάδιο της εκτέλεσης της εφαρμογής. Δεν παρέχει αριθμητικές πράξεις σε δείκτες μνήμης και παρέχει απόλυτη ασφάλεια όσον αφορά την πρόσβαση σε μνήμη εκτός του ιδίου προγράμματος. Επίσης η γλώσσα θέτει αυστηρά όρια στην αλλαγή τύπου (casting).

- **ασφαλής,** καθώς περιορίζει τις εφαρμογές που τρέχουν σε περιβάλλοντα δικτύων ως προς την πρόσβαση σε δεδομένα του υπολογιστή χωρίς την άδεια του χρήστη.

- **ουδέτερη αρχιτεκτονικής και μεταφερόμενη,** διότι δεν εξαρτάται από ιδιαιτερότητες των CPU ή των λειτουργικών συστημάτων καθώς παρέχει έναν πλήρως ορισμένο τρόπο μεταφοράς και εκτέλεσης των εφαρμογών.

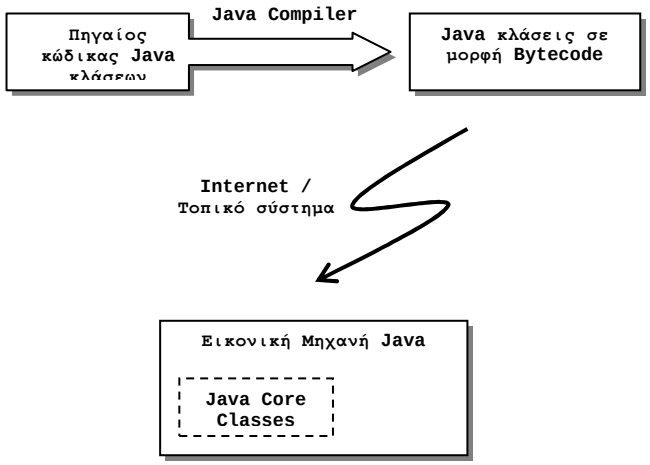
- **υψηλής ταχύτητας,** σε σχέση με άλλες ερμηνευόμενες γλώσσες, επειδή η μεταφερόμενη δομή του προγράμματος (bytecodes) είναι ήδη πολύ κοντά σε γλώσσα μηχανής. Σε συνδυασμό με Μεταγλωτιστές της Στιγμής (Just In Time Compilers, ο κώδικας Byte Code μεταγλωτίζεται στον πελάτη αντί να διερμηνεύεται), πλησιάζει την ταχύτητα συμβατικών προγραμμάτων C++.

- **πολυνηματική,** καθώς παρέχει τη δυνατότητα για εκτέλεση πολλών εργασιών «ταυτόχρονα» στην ίδια εφαρμογή εξασφαλίζοντας παράλληλα την αποφυγή ταυτόχρονης πρόσβασης σε πόρους του συστήματος με αναπτυγμένους τρόπους συγχρονισμού

- **δυναμική,** διότι φορτώνει και ενώνει τις χρησιμοποιούμενες κλάσεις την ώρα που θα χρειαστεί η πρόσβαση σε αυτές. Επίσης παρέχει τρόπους για έλεγχο του τύπου των αντικειμένων την ώρα της εκτέλεσης (run time type checking).

Σε αντίθεση με παραδοσιακές γλώσσες προγραμματισμού, το αποτέλεσμα της μεταγλώττισης ενός προγράμματος Java, δεν είναι κώδικας μηχανής αλλά ένα

ενδιάμεσο στάδιο που ονομάζεται Bytecode. Την ερμηνεία των Bytecodes και την εκτέλεση του κώδικα την αναλαμβάνει μία «Εικονική Μηχανή Java» (Java Virtual Machine, JVM). Εικονικές Μηχανές Java υπάρχουν διαθέσιμες από πολλούς κατασκευαστές, ενώ είναι ενσωματωμένες και στους γνωστούς φυλλομετρητές, δίνοντας σε αυτούς τη δυνατότητα να εκτελέσουν προγράμματα Java που μεταφέρονται μέσα από το Διαδίκτυο.



Σχήμα 3-1 Τα βήματα που ακολουθούνται για την εκτέλεση ενός Java προγράμματος. Διακρίνεται και η «Εικονική Μηχανή Java» που υλοποιεί τις βασικές κλάσεις της Java και μεταφράζει και εκτελεί τα λαμβανόμενα Bytecodes και τα εκτελεί.

Παρόλο που υπήρχαν από την αρχή της εμφάνισης της Java μεγάλες προσδοκίες από την αγορά για την εκπλήρωση της υπόσχεσης για ανεξαρτησία από την αρχιτεκτονική των υπολογιστών και από τα λειτουργικά συστήματα, η Java, για μεγάλο χρονικό διάστημα, δεν μπόρεσε να αντεπεξέλθει σε αυτές τις προσδοκίες, εμφανίζοντας ασυμβατότητες σε διαφορετικά υπολογιστικά συστήματα, ακόμα και μεταξύ διαφορετικών υλοποιήσεων στον ίδιο υπολογιστή. Σήμερα όμως η γλώσσα έχει φτάσει σε μεγάλο σημείο ωριμότητας και παρέχει πλέον ικανές συνθήκες για να μπορεί να χρησιμοποιηθεί για την ανάπτυξη σοβαρών εφαρμογών. Επίσης η Java αποκτά όλο και μεγαλύτερη σημασία όσον αφορά τη χρησιμοποίησή της και στην πλευρά του εξυπηρετητή.

3.3.1 Java Applets

Τα Applets είναι προγράμματα που φορτώνονται δυναμικά από έναν HTTP εξυπηρετητή και εκτελούνται στο περιβάλλον ενός φυλλομετρητή. Σε αντίθεση με απλές εφαρμογές Java, τα Applets έχουν περιορισμένη πρόσβαση στο σύστημα στο

οποίο εκτελούνται, εξασφαλίζοντας έτσι την ασφάλεια του τελικού χρήστη από ανεπιθύμητες ενέργειες, όπως την υποκλοπή προσωπικών πληροφοριών που βρίσκονται στον υπολογιστή του, ή την κακοήθη πρόκληση βλαβών (ιοί υπολογιστών). Με την έγκριση του χρήστη, μπορεί ακόμα και σε ένα Applet να δοθεί πρόσβαση σε επιπλέον τμήματα του υπολογιστή.

Τα προγράμματα σε μορφή Applets πέραν του περιορισμού στην πρόσβαση σε όλα τα τμήματα του υπολογιστή που εκτελούνται, έχουν πλήρες δυνατότητες, όπως κάθε άλλο πρόγραμμα σε άλλη γλώσσα προγραμματισμού, ακόμα και για την εκτέλεση πολύπλοκων υπολογισμών στην πλευρά του πελάτη.

Τα Applets ενσωματώνονται σε απλές HTML σελίδες και φορτώνονται δυναμικά μαζί με τα κείμενα και τις εικόνες της σελίδας. Αυτή η δυναμική επιτρέπει αλλαγή του κώδικα του Applet χωρίς να το καταλαβαίνει ο χρήστης. Με αυτόν τον τρόπο μπορούν να γίνουν διορθώσεις και επεκτάσεις στα προγράμματα με διαφανή (transparent) για τον τελικό χρήστη τρόπο.

3.3.2 Java Foundation Classes (JFC)

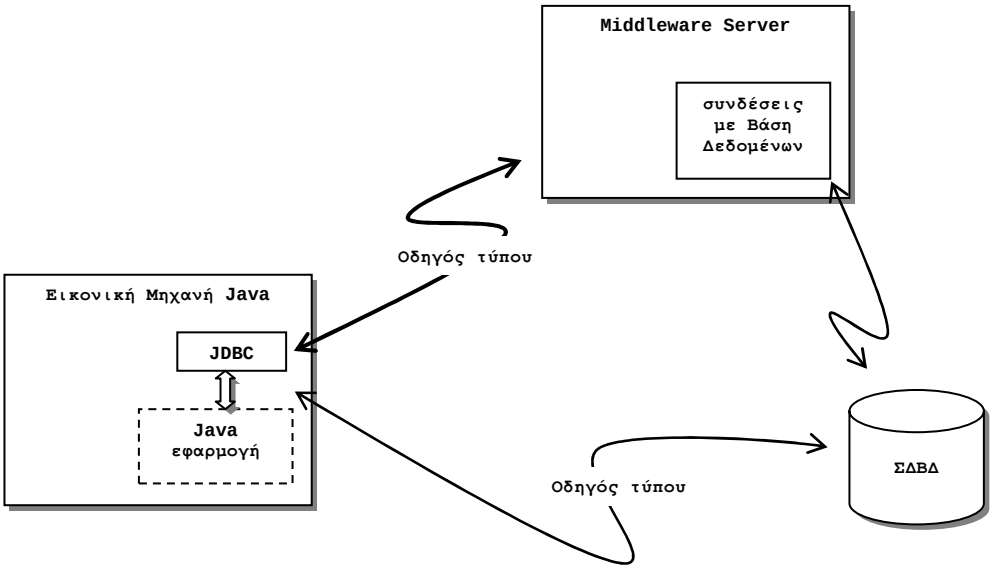
Οι Java Foundation Classes (JFC) αποτελούν μία συλλογή κλάσεων που αναπτύχθηκαν από κοινού από τη Netscape και τη Sun. Μερικά από τα χαρακτηριστικά της συλλογής αυτής είναι οι λειτουργίες για 2-D γραφικά, οι λειτουργίες για τη χρήση προγραμμάτων από άτομα με ειδικές ανάγκες (accessibility API), ο προσαρμοζόμενος τρόπος εμφάνισης και συμπεριφοράς (pluggable look & feel), η λειτουργικότητα «σύρε και άφησε» (drag & drop) και τα συστατικά Swing (Swing components).

Τα συστατικά Swing, σε συνδυασμό με τον προσαρμοζόμενο τρόπο εμφάνισης, αποτελούν την κύρια καινοτομία της JFC. Πρόκειται ουσιαστικά για μία εκτενή συλλογή από έτοιμα συστατικά για την επικοινωνία ανθρώπου- υπολογιστή (πίνακες, δέντρα, λίστες) που διευκολύνουν στη δημιουργία αποδοτικών και ισχυρών μηχανισμών για την αλληλεπίδραση με το χρήστη (User Interfaces), πράγμα που ως τώρα δεν ήταν δυνατόν να επιτευχθεί με τις βασικές λειτουργίες της Java. Λόγω της τεχνολογίας που χρησιμοποιούν, εξασφαλίζεται ίδια εμφάνιση των προγραμμάτων, ανεξαρτήτως της πλατφόρμας στην οποία εκτελούνται.

Για να χρησιμοποιηθεί η JFC, αυτή τη στιγμή πρέπει είτε να εγκατασταθούν οι κλάσεις στον υπολογιστή που πρόκειται να τρέξουν τα προγράμματα, είτε να μεταφερθούν μαζί με τις υπόλοιπες κλάσεις των προγραμμάτων. Επειδή όμως οι JFC κλάσεις έχουν ενσωματωθεί στην επόμενη έκδοση της Java ως κυρίως κλάσεις (Core Java Classes), αναμένεται η JFC να είναι συμπαγές συστατικό των «Εικονικών Μηχανών Java». Η επόμενη έκδοση της Java αναμένεται να εκδοθεί στα τέλη του του '98.

3.3.3 Java Database Connectivity (JDBC)

Το JDBC API ορίζει Java κλάσεις για την αναπαράσταση συνδέσεων σε Σχεσιακές Βάσεις Δεδομένων, δηλώσεων SQL, αποτελεσμάτων ερωτήσεων κλπ. Μέσω του JDBC ο προγραμματιστής μπορεί να εκτελεί ερωτήσεις SQL και να επεξεργάζεται τα αποτελέσματα αυτών. Το JDBC API είναι μέρος των κυρίων κλάσεων της Java (Java Core Classes).



Σχήμα 3-2 Ο ρόλος του JDBC στη σύνδεση με Βάσεις Δεδομένων. Φαίνεται επίσης η διαφορά ανάμεσα στους οδηγούς τύπου 3 και 4.

Υπάρχουν τέσσερις τύποι οδηγών που υλοποιούν το JDBC API:

1. Γέφυρα ανάμεσα JDBC και ODBC δίνοντας έτσι πρόσβαση σε σχεδόν όλες τις βάσεις δεδομένων δια μέσω του γνωστού και διαδεδομένου πρωτοκόλλου ODBC.

2. Οδηγοί που σε κάποιο τμήμα δεν είναι ανεξάρτητοι της αρχιτεκτονικής

3. Οδηγοί γραμμένοι σε Java που προσφέρονται από κατασκευαστές ενδιάμεσων εξυπηρετητών (middleware servers), οι οποίοι αναλαμβάνουν τη σύνδεση με πολλές Βάσεις Δεδομένων και δρουν ως ενδιάμεσοι σταθμοί ανάμεσα στους πελάτες και τους εξυπηρετητές για τη διοχέτευση αιτήσεων.

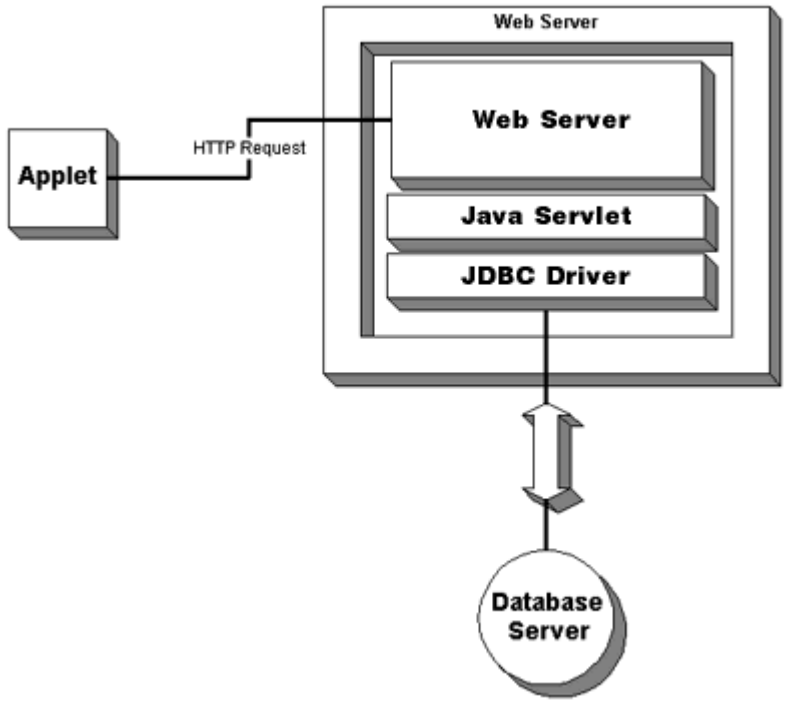
4. Οδηγοί γραμμένοι σε Java που υλοποιούν το πρωτόκολλο επικοινωνίας του ΣΔΒΔ και συνδέονται απευθείας σε αυτό.

Από τους παραπάνω τύπους οδηγών μόνο οι δύο τελευταίοι είναι τελείως ανεξάρτητοι από την πλατφόρμα στην οποία εκτελούνται οι εφαρμογές Java και είναι έτσι ικανές να υποστηρίζουν εφαρμογές που επικοινωνούν δια μέσω του Διαδικτύου από ακαθόριστου τύπου μηχανήματα.

Οι οδηγοί τύπου 3 έχουν ως κύριο πλεονέκτημα την εξοικονόμηση αδειών για σύνδεση στα ΣΔΒΔ, επειδή μπορούν με μία σύνδεση να εξυπηρετήσουν «ταυτόχρονα» πολλές αιτήσεις. Στα πλαίσια της διπλωματικής εργασίας χρησιμοποιήθηκαν οδηγοί τύπου 3 (Symantec, IDS Software) και τύπου 4 (Connect Software, IBM).

3.3.4 Java Servlets

Τα servlet είναι προγράμματα γραμμένα σε Java τα οποία εκτελούνται στην πλευρά του HTTP εξυπηρετητή. Η λειτουργικότητά τους είναι παρόμοια με αυτή των γνωστών CGI προγραμμάτων, δηλαδή προγραμμάτων που εκτελούνται από τον HTTP εξυπηρετητή κατόπιν κάποιου αιτήματος από κάποιο φυλλομετρητή, και το αποτέλεσμά τους επιστρέφεται στο φυλλομετρητή.



Σχήμα 3-3 Η αλληλεπίδραση των servlet με τις τη Βάση Δεδομένων και τους φυλλομετρητές.

Τα πλεονεκτήματα των servlet έναντι CGI προγραμμάτων, είναι κυρίως ότι όλες οι αιτήσεις σε ένα servlet εξυπηρετούνται από την ίδια διεργασία, δημιουργώντας κάθε φορά και ένα νέο thread. Δηλαδή δεν απαιτείται η δημιουργία μιας νέας διεργασίας με κάθε αίτηση στο servlet οπότε εξαλείφεται ένα μεγάλο πρόβλημα απόδοσης που υπάρχει στα παραδοσιακά CGI. Επιπλέον αυτού, στα servlet υπάρχει η δυνατότητα ύπαρξης κοινών μεταβλητών από περισσότερες αιτήσεις στο ίδιο servlet, το οποίο εξυπηρετεί κυρίως εφαρμογές που απαιτούν πρόσβαση σε Βάση Δεδομένων, οπότε μπορούν να χρησιμοποιούν τη σύνδεση στη Βάση Δεδομένων χωρίς να απαιτείται επανασύνδεση με τον εξυπηρετητή Βάσεων Δεδομένων κάθε φορά που εκτελείται το servlet. Η χρησιμοποίηση των servlet αντί κάποιου παραδοσιακού τρόπου εκτέλεσης προγραμμάτων στην πλευρά του εξυπηρετητή (CGI's) εποφελήθηκε, πέραν από τα πλεονεκτήματα που προσφέρει η Java, και από το γεγονός ότι χρησιμοποιήθηκε κοινός κώδικας με τις εφαρμογές που αναπτύχθηκαν στην πλευρά του πελάτη.

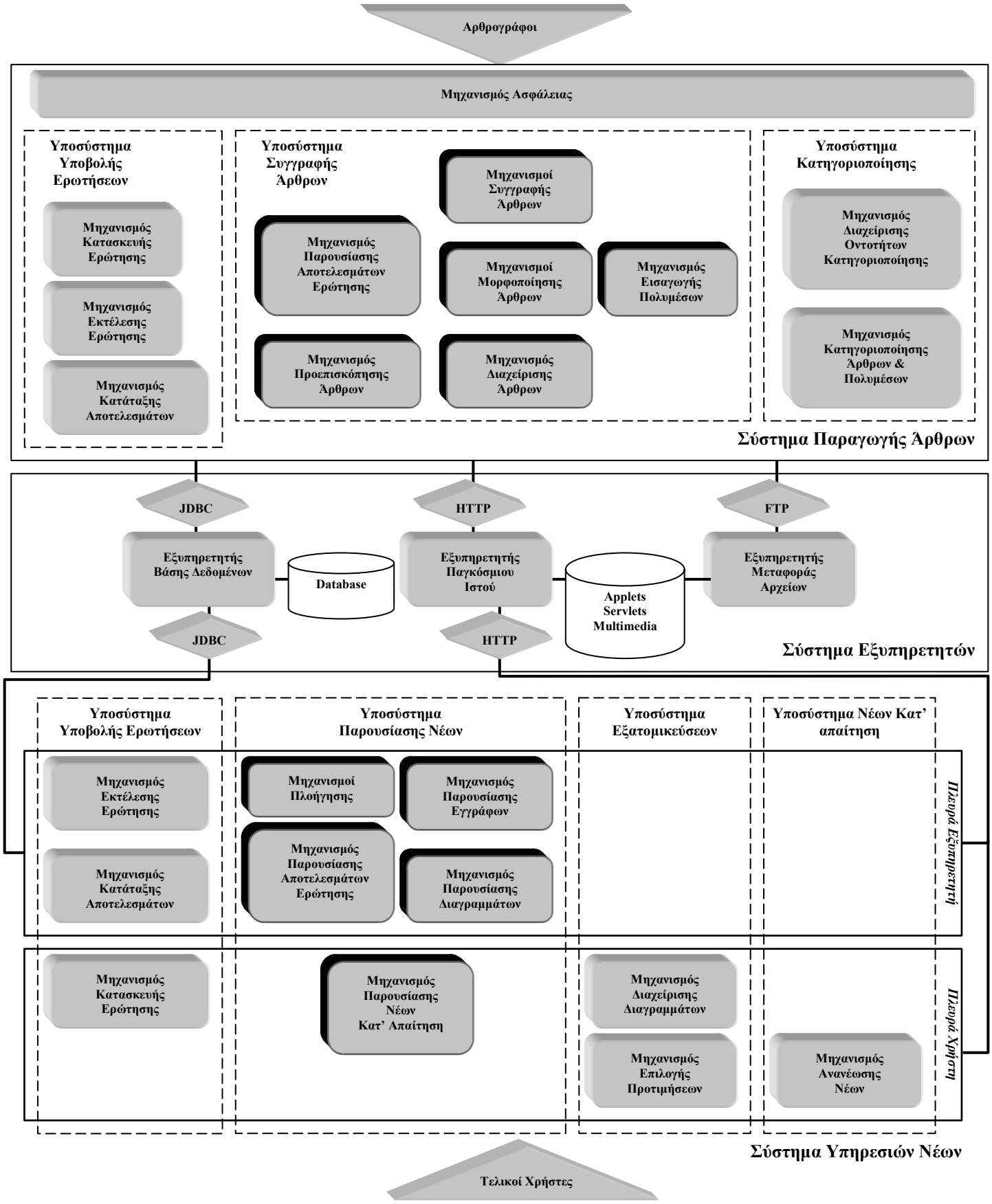
4 Η Αρχιτεκτονική του Συστήματος

Στο κεφάλαιο αυτό περιγράφεται η πλήρης αρχιτεκτονική του συστήματος "Hypermedia Custom News System" επεκτείνοντας την παρουσίαση που έγινε στο Σχήμα 1-1 του πρώτου κεφαλαίου.

Κύριο μέλημα της διαδικασίας σχεδιασμού η αρχιτεκτονική να χαρακτηρίζεται ως γενική, ώστε να υπάρχει η δυνατότητα εύκολης προσαρμογής της σε συστήματα με διαφορετικού είδους περιεχόμενο, αλλά και ανοικτή, ώστε να μπορεί να ενσωματώσει εξειδικευμένες επεκτάσεις, επιθυμητές ανά περίπτωση. Εξίσου σημαντικό χαρακτηριστικό της αρχιτεκτονικής είναι και το γεγονός ότι το σύστημα έχει τη δυνατότητα υποστήριξης αυθαίρετου αριθμού γλωσσών, κάτι που μέχρι σήμερα τουλάχιστον δεν έχει παρατηρηθεί σε άλλο σύστημα νέων.

Στο Σχήμα 3-1 φαίνεται ολοκληρωμένα η αρχιτεκτονική του συστήματος. Παραθέτονται γραφικά όλα τα υποσυστήματα και τμήματα του συστήματος που υλοποιήθηκαν, τόσο από την παρούσα διπλωματική εργασία όσο και από άλλες (βλ. [Πετρ98] και [Σκον98]). Επίσης διακρίνονται οι χρήστες του συστήματος και το τρόπος με τον οποίο αυτοί αλληλεπιδρούν με το σύστημα, καθώς επίσης και τα πρωτόκολλα επικοινωνίας που χρησιμοποιούνται για την επικοινωνία μεταξύ των κυριοτέρων τμημάτων του συστήματος.

Στα υποκεφάλαια που ακολουθούν περιγράφονται αναλυτικά όλα τα υποσυστήματα, και τα τμήματα αυτών, που διακρίνονται στο Σχήμα 3-1, καθώς επίσης και ο τρόπος που αλληλεπιδρούν μεταξύ τους, δίνοντας μεγαλύτερη έμφαση σε αυτά που υλοποιήθηκαν στα πλαίσια της παρούσας εργασίας.



Σχήμα 4-1 Αρχιτεκτονική του συστήματος νέων «Hypermedia Custom News System»

4.1 Σύστημα Παραγωγής Άρθρων

Το σύστημα παραγωγής άρθρων είναι υπεύθυνο για την παραγωγή και διαχείριση των άρθρων και αποτελείται από τρία κύρια υποσυστήματα και από ένα μηχανισμό ασφαλείας. Η εκτέλεση της εφαρμογής γίνεται εξ' ολοκλήρου στον σταθμό εργασίας του αρθρογράφου στην πλευρά του πελάτη, ο οποίος έχει τη δυνατότητα πρόσβασης στο σύστημα μέσω του Διαδικτύου. Στη συνέχεια περιγράφονται οι ανάγκες που καλύπτει κάθε υποσύστημα και μια σύντομη αναφορά στη λειτουργικότητα των τμημάτων τους.

4.1.1 Μηχανισμός Ασφαλείας

Ο Μηχανισμός Ασφαλείας επιτρέπει στον αρθρογράφο να κάνει χρήση του συστήματος αφού πρώτα επαληθεύσει την ταυτότητα του. Κάθε αρθρογράφος εφοδιάζεται με έναν κωδικό και ένα αναγνωριστικό (login, password) τα οποία του ζητούνται στην αρχή κάθε συνόδου του με το σύστημα. Ο μηχανισμός αυτός επικοινωνεί με όλα τα υποσυστήματα του συστήματος παραγωγής άρθρων για την κοινοποίηση της ταυτότητας του συνδεδεμένου χρήστη, έτσι ώστε να μπορεί το σύστημα να εγγυηθεί την ακεραιότητα της συγγραφικής εργασίας των αρθρογράφων. Η ταυτότητα του αρθρογράφου χρησιμοποιείται και στα υποσυστήματα κατηγοριοποίησης των εγγράφων για τον καθορισμό των δικαιωμάτων του χρήστη όσον αφορά τη μετατροπή των λεξικών κατηγοριοποίησης.

4.1.2 Υποσύστημα Συγγραφής Άρθρων

Το Υποσύστημα Συγγραφής Άρθρων αποτελείται από έξι κύριους μηχανισμούς, όπου ο καθένας καλύπτει συγκεκριμένες ανάγκες στη δημιουργία ενός άρθρου. Αρχικά ο *Μηχανισμός Συγγραφής Άρθρων* επιτρέπει στον αρθρογράφο να εισάγει τα βασικά χαρακτηριστικά ενός άρθρου (τίτλος, υπότιτλος, κυρίως κείμενο) και να τονίσει τα σημεία του κειμένου που επιθυμεί κάνοντας χρήση βασικών μεθόδων επισήμανσης (έντονα, πλάγια, υπογραμμισμένα γράμματα). Ο *Μηχανισμός Μορφοποίησης Άρθρων* παρέχει τη δυνατότητα ενσωμάτωσης πολυμέσων στο κείμενο του άρθρου, εισαγωγής συνδέσμων (hyperlinks) και συσχετισμού του με άλλα άρθρα (άρθρα που έχουν αποθηκευτεί το σύστημα). Τα πολυμέσα και τα άρθρα αυτά ο αρθρογράφος έχει τη δυνατότητα να τα εντοπίσει μέσω του Υποσυστήματος Υποβολής Ερωτήσεων και να τα επιλέξει μέσω του *Μηχανισμού Παρουσίασης Αποτελεσμάτων Ερώτησης*. Αν ο

αρθρογράφος επιθυμεί να εισάγει ένα νέο αντικείμενο πολυμέσου στο σύστημα και να το ενσωματώσει στο άρθρο που δημιουργεί, μπορεί να το πράξει κάνοντας χρήση του *Μηχανισμού Εισαγωγής Πολυμέσων*.

Ο *Μηχανισμός Προεπισκόπησης (Preview) Άρθρων και Πολυμέσων* παρέχει τη δυνατότητα προβολής των πολυμέσων στον αρθρογράφο και τη δυνατότητα προεπισκόπησης του ολόκληρου του άρθρου που δημιουργείται πριν αυτό εκδοθεί και αποθηκευτεί στη βάση δεδομένων. Τέλος ο *Μηχανισμός Διαχείρισης Άρθρων* είναι υπεύθυνος για τη μεταφορά και αποθήκευση ενός νέου άρθρου από το σύστημα παραγωγής άρθρων στη βάση δεδομένων, όπως επίσης και για την ανάκτηση ενός υπάρχοντος άρθρου αν ο αρθρογράφος επιθυμεί να το τροποποιήσει [Πετρ98].

4.1.3 Υποσύστημα Κατηγοριοποίησης

Το Υποσύστημα Κατηγοριοποίησης είναι ενσωματωμένο στο Υποσύστημα Συγγραφής Άρθρων και είναι υπεύθυνο για τη διαχείριση του μοντέλου κατηγοριοποίησης που ακολουθεί το σύστημα καθώς και για την κατάταξη σύμφωνα με αυτό των άρθρων και των αντικειμένων πολυμέσων που εισάγονται στο σύστημα. Τα επίπεδα κατηγοριοποίησης του συστήματος είναι δύο, κατηγορίες και υποκατηγορίες, ενώ διατηρείται και μια ιεραρχία λέξεων κλειδιών αυθαίρετου αριθμού επιπέδων, που χρησιμοποιούνται για τον χαρακτηρισμό άρθρων και αντικειμένων πολυμέσων.

Ο *Μηχανισμός Διαχείρισης Οντοτήτων Κατηγοριοποίησης* επιτρέπει στον αρθρογράφο την εισαγωγή, μετατροπή ή διαγραφή κατηγοριών, υποκατηγοριών ή λέξεων κλειδιών, έτσι ώστε να μπορεί αυτός να κατατάσσει τα άρθρα και τα πολυμέσα που εισάγει με μεγαλύτερη ακρίβεια και σαφήνεια. Ο *Μηχανισμός Κατηγοριοποίησης Άρθρων και Πολυμέσων* παρέχει τη δυνατότητα αντιστοίχησης άρθρων και πολυμέσων με λέξεις κλειδιά και κατάταξης αυτών σε κατηγορίες και υποκατηγορίες. Η τελευταία διαδικασία είναι ιδιαίτερα σημαντική, γιατί αν δεν εκτελεστεί με προσοχή από τον αρθρογράφο υπάρχει η πιθανότητα αδυναμίας εύρεσης του εισαχθέντος άρθρου ή μέσου στο σύστημα με τη χρήση του Υποσυστήματος Υποβολής Ερωτήσεων.

4.1.4 Υποσύστημα Υποβολής Ερωτήσεων

Το Υποσύστημα Υποβολής Ερωτήσεων παρέχει τη δυνατότητα αναζήτησης άρθρων και πολυμέσων που υπάρχουν στο σύστημα. Μέσω του *Μηχανισμού Κατασκευής Ερώτησης* ο αρθρογράφος επιλέγει τις λέξεις κλειδιά και τις κατηγορίες και υποκατηγορίες όπου πιστεύει ότι κατατάσσονται τα άρθρα ή πολυμέσα που επιθυμεί να εντοπίσει. Ο *Μηχανισμός Εκτέλεσης Ερώτησης* εξάγει τα αποτελέσματα της λογικής ερώτησης που απευθύνεται στη βάση δεδομένων, και ο *Μηχανισμός Κατάταξης Αποτελεσμάτων* κατατάσσει τα αποτελέσματα σύμφωνα με τη σχετικότητά τους ως προς την υποβληθείσα ερώτηση. Τέλος αυτά παραδίδονται στο Μηχανισμό Παρουσίασης Αποτελεσμάτων Ερώτησης του Υποσυστήματος Συγγραφής Άρθρων και παρουσιάζονται στον αρθρογράφο. Το υποσύστημα υποβολής ερωτήσεων διατίθεται ως αυτόνομη εφαρμογή ώστε να μπορούν οι τελικοί χρήστες να αναζητούν έγγραφα από το σύστημα, αλλά είναι ενσωματωμένο και στο Υποσύστημα Συγγραφής Άρθρων ώστε οι αρθρογράφοι να μπορούν να εντοπίσουν έγγραφα ώστε να τα συσχετίσουν με το έγγραφο που επιθυμούν να συντάξουν.

4.2 Σύστημα Εξυπηρετητών

Το σύστημα εξυπηρετητών αποτελείται από τρεις εξυπηρετητές, ο κάθε ένας από τους οποίους μπορεί να βρίσκεται σε οποιοδήποτε σημείο του Διαδικτύου, και οι οποίοι εξυπηρετούν αιτήσεις που δέχονται τόσο από το Σύστημα Παραγωγής Άρθρων όσο και από το Σύστημα Υπηρεσιών Νέων από τους τελικούς χρήστες. Ο ρόλος του κάθε ενός από αυτούς περιγράφεται αναλυτικά παρακάτω.

4.2.1 Εξυπηρετητής Βάσης Δεδομένων

Ο Εξυπηρετητής Βάσης Δεδομένων είναι ίσως το σημαντικότερο τμήμα του συστήματος καθώς είναι υπεύθυνο για τη διασφάλιση των παρεχόμενων υπηρεσιών και για τη συνέπεια του συστήματος. Σε αυτόν υλοποιείται το διάγραμμα οντοτήτων – σχέσεων που σχεδιάστηκε για τις ανάγκες του συστήματος. Εκεί αποθηκεύονται σε δομημένη μορφή τα δεδομένα του συστήματος και από αυτόν εξυπηρετούνται οι αιτήσεις αποθήκευσης και ανάκτησης δεδομένων που δέχεται από όλα τα υποσυστήματα. Στην ουσία είναι ένα ολοκληρωμένο σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων, τα κύρια χαρακτηριστικά του οποίου είναι η επεξεργασία

(processing) και βελτιστοποίηση (optimization) των ερωτήσεων ανάκτησης, οι συνδιαλλαγές (transactions), ο ταυτοχρονισμός (concurrency) και η ανάνηψη (recovery). Τα υποσυστήματα επικοινωνούν με τον Εξυπηρετητή Βάσης Δεδομένων μέσω του πρωτοκόλλου JDBC (Java DataBase Connectivity), της τεχνολογίας δηλαδή σύνδεσης των Java Εφαρμογών με βάσεις δεδομένων.

4.2.2 Εξυπηρετητής Παγκόσμιου Ιστού

Ο Εξυπηρετητής Παγκόσμιου Ιστού είναι υπεύθυνος για την παροχή των υπηρεσιών νέων στους τελικούς χρήστες, ενεργοποιώντας τους κατάλληλους μηχανισμούς παρουσίασης μέσω των servlets (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης) που εκτελούνται στην πλευρά του εξυπηρετητή. Επίσης παρέχει τις απαραίτητες Java κλάσεις στους χρήστες του συστήματος, έτσι ώστε αυτοί να έχουν πρόσβαση στα υποσυστήματα, που τρέχουν στο σταθμό εργασίας τους, μέσω των Java Applets. Οι αρθρογράφοι και οι τελικοί χρήστες επικοινωνούν με τον Εξυπηρετητή Παγκόσμιου Ιστού κάνοντας χρήση του πρωτοκόλλου μεταφοράς υπερκειμένου HTTP (HyperText Transfer Protocol) που χρησιμοποιούν οι φυλλομετρητές παγκόσμιου ιστού (Web Browsers).

4.2.3 Εξυπηρετητής Μεταφοράς Αρχείων

Ο Εξυπηρετητής Μεταφοράς Αρχείων είναι υπεύθυνος για την υποστήριξη του Μηχανισμού Εισαγωγής Αντικειμένων Πολυμέσων, μέσω του οποίου οι αρθρογράφοι εισάγουν νέα αντικείμενα πολυμέσων στο σύστημα. Συγκεκριμένα, κατά την ολοκλήρωση της διαδικασίας εισαγωγής ενός νέου αντικειμένου στο σύστημα αποστέλλεται αίτηση μεταφοράς του περιεχομένου του στον Εξυπηρετητή Μεταφοράς Αρχείων και τότε αρχίζει η μεταφορά του σε αυτόν. Μόλις αυτή ολοκληρωθεί, το μέσο έχει αποθηκευτεί στη δευτερεύουσα μνήμη και η διαδικασία εισαγωγής του θεωρείται επιτυχημένη. Ο Μηχανισμός Εισαγωγής Πολυμέσων επικοινωνεί με τον Εξυπηρετητή Μεταφοράς Αρχείων με το πρωτόκολλο μεταφοράς αρχείων FTP (File Transfer Protocol).

4.3 Σύστημα Υπηρεσιών Νέων

Το σύστημα αυτό είναι υπεύθυνο για την παροχή των υπηρεσιών του συστήματος στους τελικούς χρήστες και αποτελείται από τέσσερα κυρίως υποσυστήματα. Κάποια τμήματα των υποσυστημάτων αυτών εκτελούνται στο σταθμό εργασίας του τελικού χρήστη, ο οποίος έχει τη δυνατότητα πρόσβασης στο σύστημα μέσω του Διαδικτύου, ενώ άλλα εκτελούνται στην πλευρά του συστήματος εξυπηρετητών. Στη συνέχεια περιγράφονται οι ανάγκες που καλύπτει κάθε υποσύστημα και μια σύντομη αναφορά στη λειτουργικότητα των τμημάτων τους.

4.3.1 Υποσύστημα Παρουσίασης Νέων

Το Υποσύστημα Παρουσίασης Νέων ([Πετρ98]) αναλαμβάνει την παρουσίαση του περιεχομένου του συστήματος στον τελικό χρήστη με τον καλύτερο δυνατό τρόπο, έτσι ώστε η αλληλεπίδραση αυτών των δύο να είναι όσο πιο φιλική και εύκολη γίνεται. Συγκεκριμένα καθορίζει το περιεχόμενο και κατασκευάζει τις HTML σελίδες που παρουσιάζονται στον τελικό χρήστη ανάλογα με τις αιτήσεις που αυτός υποβάλλει μέσω του εξυπηρετητή παγκόσμιου ιστού. Αποτελείται από πέντε κύριους μηχανισμούς, όπου ο καθένας καλύπτει συγκεκριμένες ανάγκες παρουσίασης.

Ο *Μηχανισμός Πλοήγησης* επιτρέπει στον τελικό χρήστη την επισκόπηση των κατηγοριών και υποκατηγοριών του συστήματος, καθώς και μια σύντομη περιγραφή των νεότερων άρθρων που αυτές περιέχουν. Ο *Μηχανισμός Παρουσίασης Αποτελεσμάτων Ερώτησης* αναλαμβάνει την παρουσίαση σύντομων περιγραφών των άρθρων που προκύπτουν από τις ερωτήσεις που ο τελικός χρήστης υποβάλλει στο σύστημα μέσω του Υποσυστήματος Υποβολής Ερωτήσεων.

Ο *Μηχανισμός Παρουσίασης Διαγραμμάτων* εξασφαλίζει ότι η ενημέρωση των τελικών χρηστών από το σύστημα φιλτράρεται σύμφωνα με τις προσωπικές τους προτιμήσεις, τις οποίες ο τελικός χρήστης έχει δηλώσει και αποθηκεύσει μέσω του Υποσυστήματος Εξατομικεύσεων. Ο *Μηχανισμός Παρουσίασης Νέων Κατ΄ Απαίτηση* μοιάζει αρκετά με τον Μηχανισμό Παρουσίασης Διαγραμμάτων, μόνο που αυτός ενεργοποιείται στην πλευρά του τελικού χρήστη, και φροντίζει για την περιληπτική παρουσίαση σε αυτόν μόνο των νεώτερων άρθρων που τον ενδιαφέρουν. Τέλος, ο *Μηχανισμός Παρουσίασης Εγγράφων* πραγματοποιεί την ολοκληρωμένη παρουσίαση

των χαρακτηριστικών ενός άρθρου. Οι παραπάνω μηχανισμοί ενεργοποιούνται στην πλευρά του συστήματος εξυπηρετητών, όπως φαίνεται και στο Σχήμα 3-1.

4.3.2 Υποσύστημα Υποβολής Ερωτήσεων

Το Υποσύστημα Υποβολής Ερωτήσεων παρέχει τη δυνατότητα αναζήτησης άρθρων του συστήματος από τους τελικούς χρήστες και είναι παρόμοιο με αυτό που παρουσιάστηκε στο σύστημα παραγωγής άρθρων. Μόνο που εδώ όλοι οι μηχανισμοί του δεν εκτελούνται στην πλευρά του τελικού χρήστη, όπως φαίνεται και στο Σχήμα 3-1. Συγκεκριμένα ο *Μηχανισμός Κατασκευής Ερώτησης* είναι ο μόνος που εκτελείται στην πλευρά του χρήστη και του επιτρέπει να επιλέξει τις προτιμήσεις του και τα ενδιαφέροντά του, ακολουθώντας το μοντέλο κατηγοριοποίησης των εγγράφων του συστήματος.

Ο *Μηχανισμός Εκτέλεσης Ερώτησης* και ο *Μηχανισμός Κατάταξης Αποτελεσμάτων* ενεργοποιούνται στην πλευρά του συστήματος εξυπηρετητών και εξάγουν καταταγμένα τα άρθρα που προέκυψαν από την ερώτηση. Το γεγονός αυτό βελτιώνει κατά πολύ την απόδοση του συστήματος, αφού η επεξεργασία και η εκτέλεση της ερώτησης γίνεται τοπικά. Στη συνέχεια τα εξαγόμενα αποτελέσματα παραδίδονται στο Μηχανισμό Παρουσίασης Αποτελεσμάτων Ερώτησης, που βρίσκεται επίσης στην πλευρά του συστήματος εξυπηρετητών, και ο οποίος παράγει τις HTML σελίδες που παρουσιάζονται στον τελικό χρήστη.

4.3.3 Υποσύστημα Εξατομικεύσεων

Το Υποσύστημα Εξατομικεύσεων προσδίδει στο σύστημα το χαρακτηρισμό του εξατομικευμένου, χρησιμοποιώντας την αντίστοιχη τεχνολογία. Αποτελείται από δύο κύριους μηχανισμούς, όπου ο καθένας καλύπτει συγκεκριμένες ανάγκες εξατομίκευσης των νέων. Ο *Μηχανισμός Διαχείρισης Διαγραμμάτων Χρηστών* (User Profiles) επιτρέπει στον τελικό χρήστη τη δημιουργία ενός νέου ή τη μετατροπή του υπάρχοντος διαγράμματος του, ενώ μέσω του *Μηχανισμού Επιλογής Προτιμήσεων* ο χρήστης επιλέγει τις κατηγορίες και τις υποκατηγορίες, καθώς και τις λέξεις κλειδιά, που πιστεύει ότι περιέχουν και χαρακτηρίζουν νέα τα οποία καλύπτουν τα προσωπικά του ενδιαφέροντα. Και οι δύο παραπάνω μηχανισμοί ενεργοποιούνται στην πλευρά του τελικού χρήστη, όπως φαίνεται και στο Σχήμα 3-1.

4.3.4 Υποσύστημα Νέων Κατ' Απαίτηση

Το Υποσύστημα Νέων Κατ' Απαίτηση προσδίδει στο σύστημα το χαρακτηρισμό του κατ' απαίτηση, και αποτελείται από ένα μηχανισμό. Ο *Μηχανισμός Ανανέωσης Νέων* ενεργοποιείται στην πλευρά του τελικού χρήστη και περιοδικά συνδέεται με το σύστημα για την αναζήτηση νεοεκδοθέντων άρθρων που ταιριάζουν με το διάγραμμά του. Σε περίπτωση ύπαρξης τέτοιων άρθρων ειδοποιεί ανάλογα τον τελικό χρήστη.

4.4 Επιπλέον Μηχανισμοί Συστημάτων Νέων

Στην παρούσα εργασία αναπτύχθηκε και ένας μηχανισμός που αφορά μία γενικότερη αντιμετώπιση του προβλήματος που προκύπτει στην περίπτωση που είτε οι τελικοί χρήστες είτε ο εξυπηρετητής της βάσης δεδομένων προστατεύεται από ένα Firewall (διάταξη που απαγορεύει την απ' ευθείας σύνδεση μεταξύ υπολογιστών που βρίσκονται σε διαφορετική πλευρά του Firewall). Αναπτύχθηκαν μηχανισμοί για την πρόσβαση στη βάση δεδομένων δια μέσω του HTTP πρωτοκόλλου που υλοποιείται από τους εξυπηρετητές παγκόσμιου ιστού και τους φυλλομετρητές (browsers).

4.5 Ανακεφαλαίωση

Στο κεφάλαιο αυτό περιγράφηκε πλήρως η αρχιτεκτονική του συστήματος «Hypermedia Custom News System». Όπως η αρχιτεκτονική του συστήματος καλύπτει όλες τις βασικές ανάγκες ενός συστήματος εξατομικευμένων νέων κατ' απαίτηση και παρέχει όλη τη λειτουργικότητα που ο τελικός χρήστης μπορεί να ζητήσει από ένα τέτοιο σύστημα. Επίσης, φαίνονται και τα τμήματα που υλοποιεί η παρούσα διατριβή και η σημαντικότητα τους μέσα στο σύστημα. Συνοψίζοντας, τα κυριότερα πλεονεκτήματα της αρχιτεκτονικής του συστήματος είναι τα παρακάτω:

- Υποστηρίζει αυθαίρετο αριθμό γλωσσών.

- Είναι γενική ώστε να προσαρμόζεται εύκολα σε συστήματα με διαφορετικού είδους περιεχόμενο, επειδή ο σχεδιασμός, η αναζήτηση, η κατηγοριοποίηση και η παρουσίαση δεν εξαρτώνται από το περιεχόμενο των εγγράφων που διαχειρίζεται.

- Είναι ανοικτή ώστε να ενσωματώνει εύκολα υπηρεσίες και λειτουργικότητα απαραίτητες σε εξειδικευμένα συστήματα.

- Παρέχει ένα ολοκληρωμένο σύστημα παραγωγής νέων στους αρθρογράφους, με πολλές δυνατότητες μορφοποίησης και διαχείρισης των άρθρων τους.

- Υλοποιεί ένα εξελιγμένο σύστημα κατηγοριοποίησης και αναζήτησης άρθρων και πολυμέσων.

- Διαθέτει κατάλληλους μηχανισμούς παρουσίασης για την πληρέστερη και ευκολότερη ενημέρωση των τελικών χρηστών από το σύστημα.

- Υποστηρίζει τις τεχνολογίες εξατομίκευσης των νέων και των νέων κατ' απαίτηση με τη δημιουργία και παρουσίαση διαγραμμάτων από και προς τους τελικούς χρήστες αντίστοιχα.

Σύμφωνα με τα παραπάνω χαρακτηριστικά και την παρουσίαση κάποιων υπαρχόντων συστημάτων νέων του κεφαλαίου 2 «Επισκόπηση Υπαρχόντων Συστημάτων», μπορεί να γίνει η σύγκριση του πίνακα 4-1 για τα χαρακτηριστικά του συστήματος που υλοποιήθηκε, του συστήματος νέων του CNN και του συστήματος νέων της καθημερινής.

	Σύστημα που αναπτύχθηκε	CNN interactive	Ναυτεμπορική
πολυγλωσσικό	_	a	a
υποστηριζόμενα έγγραφα	_	`	`
ευκολία προσαρμογής για άλλου είδους έγγραφα	_	a	a
ανοικτό	_	-	-
ολοκληρωμένο σύστημα παραγωγής στο Διαδίκτυο	_	a	a
μοντέλο κατηγοριοποίησης	_	a	a
παρουσίαση αποτελεσμάτων και εγγράφων	_	_	a
εξατομικευμένο	`	_	`
κατ᾽ απαίτηση	_	a	`

_ Πολύ καλή υποστήριξη

` Μέτρια υποστήριξη

a Ελάχιστη υποστήριξη

5 Η Βάση Δεδομένων

Στο κεφάλαιο αυτό παρουσιάζεται το εννοιολογικό και φυσικό σχήμα (conceptual και physical) της βάσης δεδομένων του συστήματος "Hypermedia Custom News System". Μέσω της ανάλυσης απαιτήσεων καθορίζονται οι οντότητες του συστήματος και οι σχέσεις μεταξύ τους, έτσι όπως αυτές προκύπτουν από την αρχιτεκτονική που παρουσιάστηκε στο προηγούμενο κεφάλαιο, και με σκοπό να υποστηρίζουν όλα τα υποσυστήματα και τμήματα του συστήματος. Από αυτήν προκύπτει το διάγραμμα οντοτήτων – σχέσεων, από το οποίο καθορίζεται με τη σειρά του το σχεσιακό μοντέλο της βάσης δεδομένων του συστήματος που υλοποιείται στον εξυπηρετητή βάσεων δεδομένων.

5.1 Ανάλυση Απαιτήσεων

Στο υποκεφάλαιο αυτό παρουσιάζονται οι οντότητες που προέκυψαν από την αρχιτεκτονική του συστήματος. Οι οντότητες αυτές είναι οι εξής:

- **Η οντότητα γλωσσών (languages entity)**, που περιέχει την περιγραφή και την κωδικοποίηση των γλωσσών που υποστηρίζει το σύστημα και είναι το πρώτο και σημαντικότερο βήμα για την υποστήριξη αυθαίρετου αριθμού γλωσσών.

- **Η οντότητα αρθρογράφων (authors entity)**, που κρατά πληροφορίες για τους αρθρογράφους που αναγνωρίζει το σύστημα και αποτελεί τη βάση ανάπτυξης του μηχανισμού ασφαλείας (βλέπε [Πετρ98]).

- **Η οντότητα κατηγοριών (categories entity)** , που αποθηκεύει το λεξικό των κατηγοριών και υποκατηγοριών σύμφωνα με το μοντέλο κατηγοριοποίησης που ακολουθεί το σύστημα. Η οντότητα αυτή είναι το πρώτο από τα δύο βασικά χαρακτηριστικά κατηγοριοποίησης των εγγράφων.

- **Η οντότητα λέξεων - κλειδιών (keywords entity**, που**)** είναι η δεύτερη οντότητα υποστήριξης του μοντέλου κατηγοριοποίησης και εσωκλείει την ιεραρχία των λέξεων κλειδιών που διαθέτει το σύστημα για το χαρακτηρισμό των άρθρων και των αντικειμένων πολυμέσων.

- **Η οντότητα στατιστικών (statistics entity), που** κρατά στατιστικά στοιχεία για τον αριθμό εγγράφων ανά κατηγορία, υποκατηγορία και λέξη κλειδί και έχει να κάνει με το μοντέλο υπολογισμού σχετικότητας που υλοποιεί το σύστημα (βλ. και [Σκον98]). Για τις κατηγορίες και τις λέξεις κλειδιά που υπάρχουν στο σύστημα η οντότητα ενημερώνεται μέσω δυο σχέσεων με τις αντίστοιχες οντότητες.

- **Η οντότητα πολυμέσων (media entity)**, που μπορεί να περιγράψει τρία είδη αντικειμένων πολυμέσων, εικόνες, ήχους και video. Περιέχει κάθε χαρακτηριστικό αναγκαίο και χρήσιμο για την αποθήκευση και παρουσίαση των αντικειμένων αυτών, είτε αυτό είναι ενσωματωμένο σε κάποιο άρθρο είτε προβάλλονται ανεξάρτητα. Συνδέεται με τις οντότητες κατηγοριών και λέξεων κλειδιών έτσι ώστε κάθε αντικείμενο να κατηγοριοποιείται μέσα στο σύστημα. Αυτός ο συσχετισμός είναι αναγκαίος για την αναζήτηση και τον εντοπισμό των αντικειμένων πολυμέσων από το υποσύστημα υποβολής ερωτήσεων (βλέπε [Πετρ98]).

- **Η οντότητα εγγράφων (documents entity), που** περιέχει όλα τα αναγκαία χαρακτηριστικά για τη διαχείριση και την παρουσίαση ενός εγγράφου. Συνδέεται με την οντότητα γλωσσών για τον καθορισμό της γλώσσας στην οποία είναι γραμμένο κάθε έγγραφο και με την οντότητα αρθρογράφων για τον καθορισμό του συγγραφέα του. Και αυτή η οντότητα συνδέεται με τις οντότητες κατηγοριών και λέξεων κλειδιών ώστε κάθε έγγραφο να κατηγοριοποιείται μέσα στο σύστημα. Το υποσύστημα υποβολής ερωτήσεων είναι άλλωστε υπεύθυνο για την αναζήτηση και αντικειμένων πολυμέσων και εγγράφων. Μέσω δύο σχέσεων με τον εαυτό της η οντότητα μπορεί να περιγράψει τα άρθρα που συσχετίζονται με κάθε έγγραφο καθώς και τις μεταφράσεις του κάθε άρθρου. Τέλος υπάρχει σύνδεση της οντότητας αυτής με την οντότητα αντικειμένων πολυμέσων για τον καθορισμό των πολυμέσων που περιέχει κάθε έγγραφο και για τον τρόπο που παρουσιάζονται μέσα σε αυτό. Τόσο το υποσύστημα συγγραφής άρθρων όσο και το υποσύστημα παρουσίασης νέων στηρίζονται κατά πρώτο λόγο σε αυτή την οντότητα, και κατά δεύτερο στην οντότητα πολυμέσων (βλέπε [Πετρ98]).

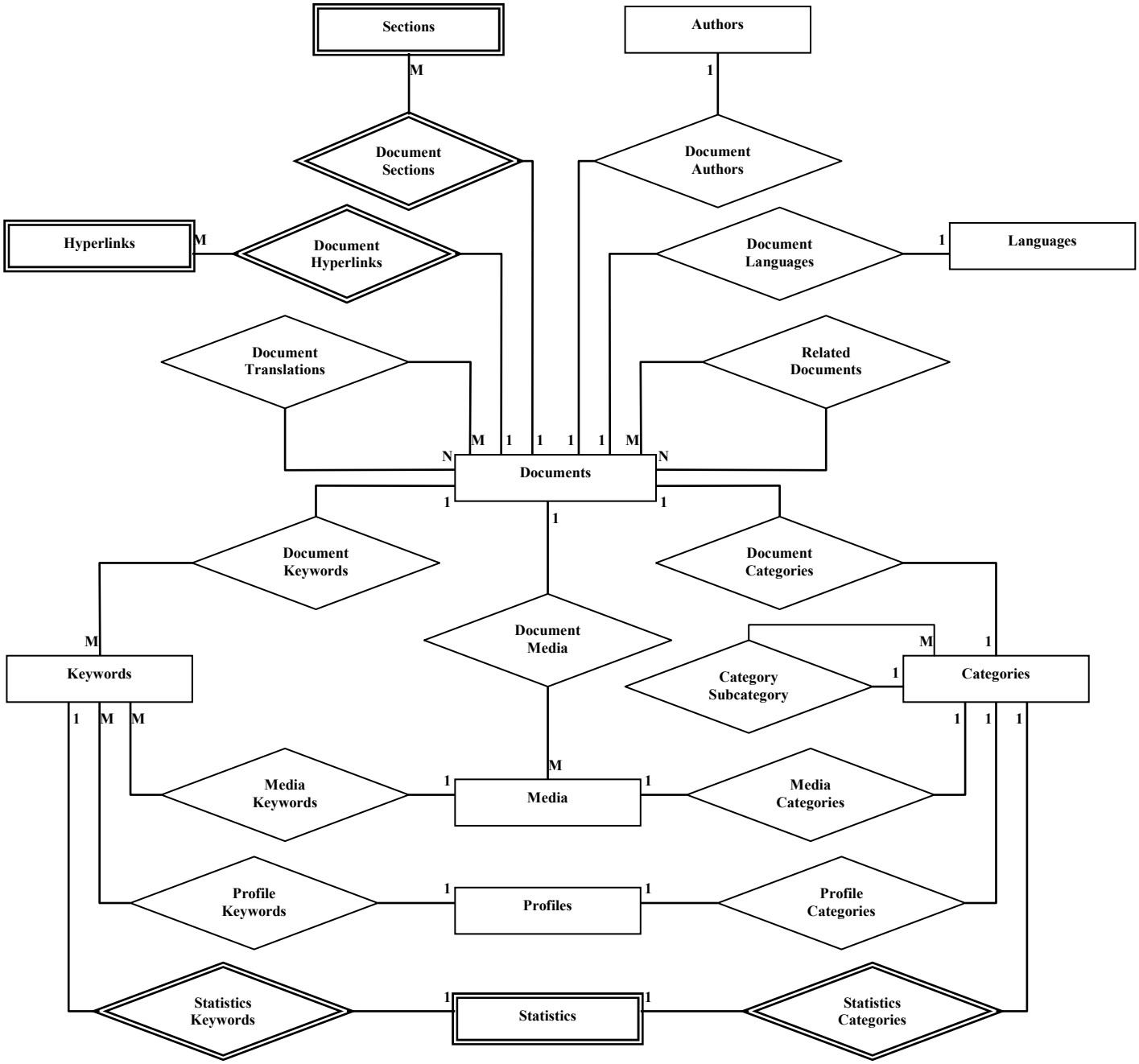
- **Η οντότητα ενοτήτων (sections entity)**, που περιέχει τις ενότητες στις οποίες μπορεί να χωρίζεται ένα έγγραφο και συνδέεται με την οντότητα εγγράφων για το σκοπό αυτό (βλέπε [Πετρ98]).

- **Η οντότητα διασυνδέσεων (hyperlinks entity)**, που περιέχει τις διασυνδέσεις που μπορεί να περιέχονται μέσα στο κείμενο ενός εγγράφου, και παραπέμπουν τον αναγνώστη σε κάποια διεύθυνση στο Διαδίκτυο ή σε κάποιο αντικείμενο πολυμέσων. Συνδέεται και αυτή με την οντότητα εγγράφων (βλέπε [Πετρ98]).

- **Η οντότητα διαγραμμάτων (profiles entity)**, που αποτελεί τα προσωπικά διαγράμματα των τελικών χρηστών. Αυτά αποτελούνται από τις προσωπικές πληροφορίες του χρήστη και από την περιγραφή των ενδιαφερόντων του. Τις προτιμήσεις του ο χρήστης τις εκφράζει επιλέγοντας κατηγορίες και λέξεις κλειδιά, οπότε η οντότητα αυτή πρέπει να συνδέεται με τις οντότητες κατηγοριών και λέξεων κλειδιών. Τα υποσυστήματα εξατομικεύσεων και νέων κατ' απαίτηση στηρίζονται αποκλειστικά στην οντότητα διαγραμμάτων.

5.2 Διάγραμμα Οντοτήτων – Σχέσεων

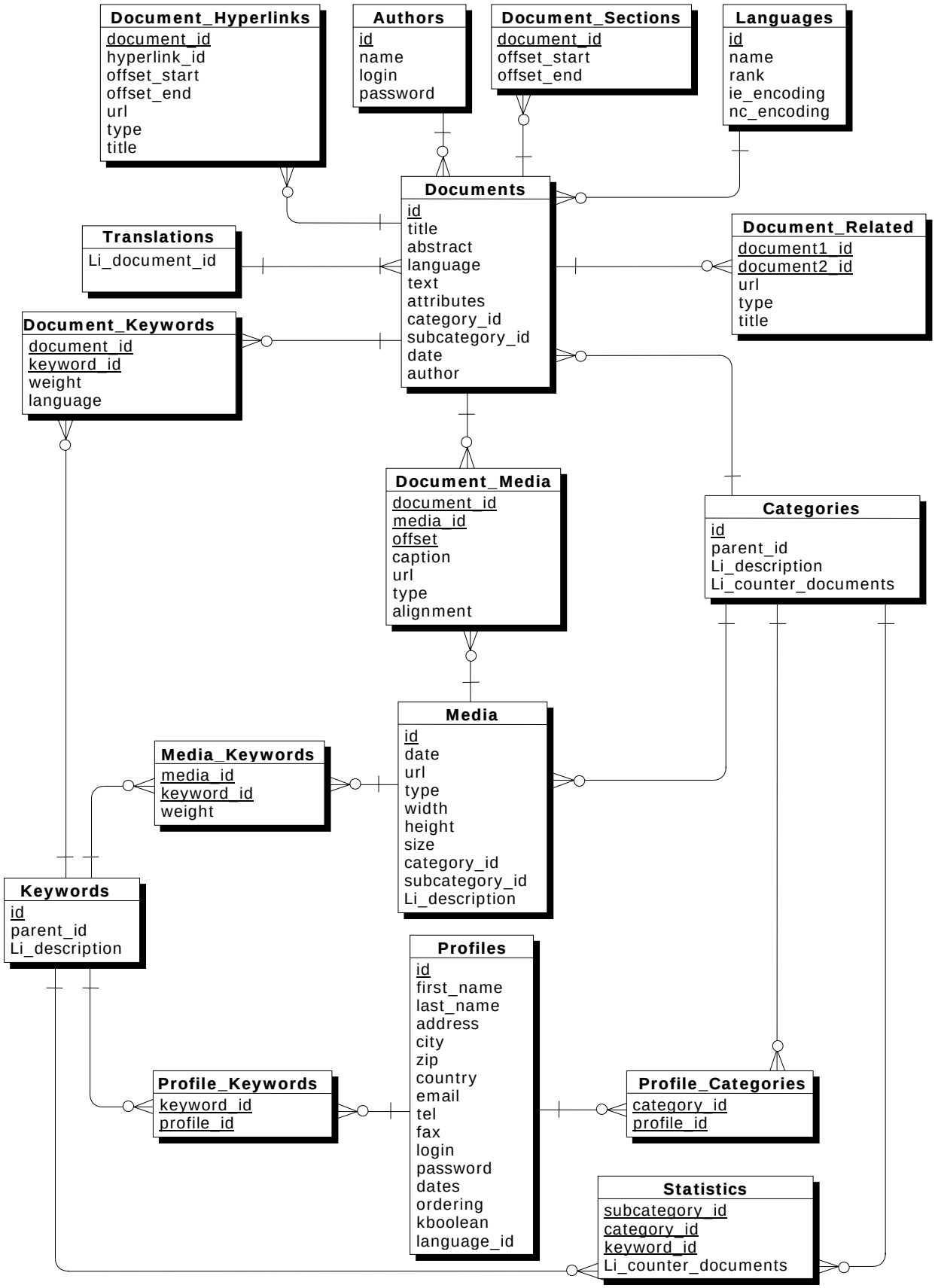
Το διάγραμμα οντοτήτων – σχέσεων που προκύπτει από την παραπάνω ανάλυση απαιτήσεων φαίνεται στο Σχήμα 4-1. Στο σχήμα αυτό φαίνονται όλες οι οντότητες που περιγράφηκαν παραπάνω, και οι οποίες υποστηρίζουν όλα τα υποσυστήματα και τους μηχανισμούς αυτών όπως απεικονίζονται στην αρχιτεκτονική του συστήματος. Επίσης φαίνονται οι σχέσεις που συνδέουν τις οντότητες μεταξύ τους και ο τύπος της κάθε σχέσης.



Σχήμα 5-1 Διάγραμμα οντοτήτων – σχέσεων του συστήματος νέων «Hypermedia Custom News System»

5.3 Το Σχεσιακό Μοντέλο της Βάσης Δεδομένων

Από το διάγραμμα οντοτήτων – σχέσεων, που περιγράφηκε στο προηγούμενο υποκεφάλαιο, εξάγεται το σχεσιακό μοντέλο της βάσης δεδομένων του συστήματος, το οποίο φαίνεται στο Σχήμα 4-2. Διακρίνονται οι πίνακες που προέκυψαν από τις οντότητες και από τις σχέσεις αυτών, μαζί με όλα τα πεδία τους.



Σχήμα 5-2 Σχεσιακό μοντέλο της βάσης δεδομένων του συστήματος νέων «Hypermedia Custom News System»

5.4 Ανακεφαλαίωση

Στο κεφάλαιο αυτό τέθηκαν οι απαιτήσεις για το σύστημα νέων και σχεδιάστηκαν οι το διάγραμμα οντοτήτων - σχέσεων καθώς και το σχεσιακό μοντέλο της Βάσης Δεδομένων που προκύπτει από το διάγραμμα αυτό.

6 Υποσύστημα Κατηγοριοποίησης Εγγράφων

Η κατηγοριοποίηση των εγγράφων σύμφωνα με κάποιο συγκεκριμένο μοντέλο προσδιορισμού του περιεχομένου αυτών, αποτελεί τη βάση για την επιτυχή αναζήτηση της πληροφορίας που ενδιαφέρει τους τελικούς χρήστες. Ειδικά σε συστήματα νέων που μπορούν να διαχειρίζονται πληροφορίες από μία πληθώρα θεμάτων, οι μηχανισμοί που υποστηρίζονται από το σύστημα νέων για την κατηγοριοποίηση της πληροφορίας κρίνουν την ποιότητα όλου του συστήματος, επειδή σε αυτούς στηρίζεται ολόκληρος ο μηχανισμός αναζήτησης πληροφορίας. Ένα ελλιπές μοντέλο κατηγοριοποίησης έχει ως συνέπεια το σύστημα νέων να είναι άχρηστο στους χρήστες του, επειδή οι τελευταίοι δε θα έχουν τα απαραίτητα συστατικά για την αναζήτηση πληροφορίας. Από την άλλη μεριά, ένα πλήρες μοντέλο κατηγοριοποίησης παρέχει στους τελικούς χρήστες τις προϋποθέσεις για της επιτυχή αναζήτηση της πληροφορίας που επιθυμούν.

Σε αυτό το κεφάλαιο θα περιγραφεί το μοντέλο κατηγοριοποίησης των εγγράφων που επιλέχτηκε για το «Hypermedia Custom News System», και θα αναλυθούν τα εργαλεία και οι μηχανισμοί που αναπτύχθηκαν για την υποστήριξή του.

6.1 Ανάλυση Απαιτήσεων

Οι απαιτήσεις που υπάρχουν, όσον αφορά το μοντέλο κατηγοριοποίησης, εστιάζονται κυρίως στην όσο το δυνατόν καλύτερη υποστήριξη του υποσυστήματος αναζήτησης. Το μοντέλο κατηγοριοποίησης πρέπει

- να παρέχει τη δυνατότητα σαφής περιγραφής του περιεχομένου κάποιου εγγράφου.

- να παρέχει στον τελικό χρήστη η απαιτούμενη ευελιξία για την κατασκευή των ερωτήσεων αναζήτησης πληροφορίας.

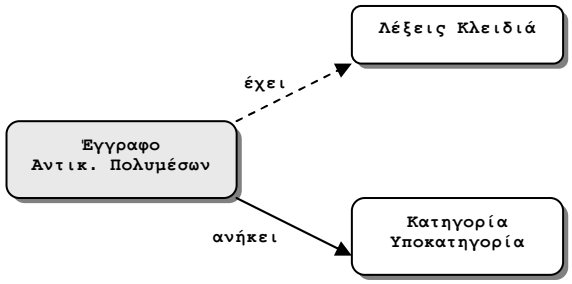
- να υποστηρίζει τη σύγκριση των αποτελεσμάτων αναζήτησης ως προς τη σχετικότητά τους σε σχέση με την υποβληθείσα ερώτηση.

6.2 Μοντέλο Κατηγοριοποίησης Εγγράφων

Το μοντέλο για την κατηγοριοποίηση των εγγράφων βασίζεται στο μοντέλο που χρησιμοποιήθηκε στο σύστημα HyNoDe ([Σκον98]). Ορίζονται δύο έννοιες για την κατηγοριοποίηση των εγγράφων: η έννοια της *λέξης κλειδί (Keyword)* και η έννοια της *κατηγορίας (Category)*:

- **Λέξη Κλειδί**: ένα έγγραφο χαρακτηρίζεται από λέξεις κλειδιά, οι οποίες προσδιορίζουν το περιεχόμενο του εγγράφου. Κάθε έγγραφο μπορεί να χαρακτηρίζεται από απεριόριστο αριθμό λέξεων κλειδιών έτσι ώστε να μπορούν να προσδιοριστούν με ακρίβεια όλες οι πτυχές του εγγράφου.

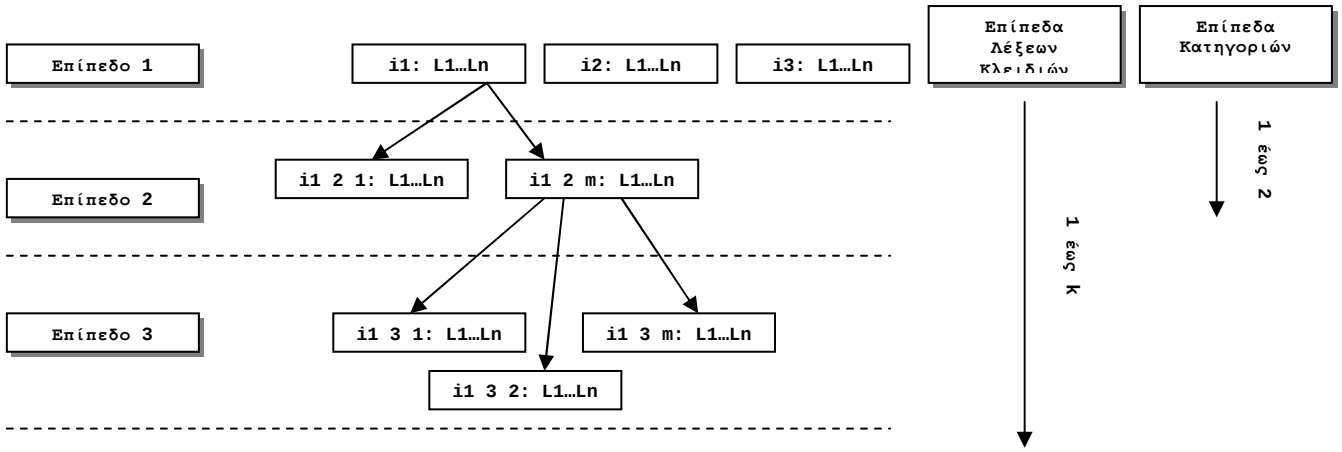
- **Κατηγορία**: κάθε έγγραφο μπορεί να ανήκει είτε σε μία γενική κατηγορία εγγράφων, είτε σε μία εξειδίκευση μιας γενικής κατηγορίας (υποκατηγορία).



Σχήμα 6-1 Τα έγγραφα χαρακτηρίζονται από απεριόριστο αριθμό Λέξεων Κλειδιών και ανίκουν ή σε μία κατηγορία, ή σε μία εξειδίκευση μιας κατηγορίας

Στο Σχήμα 6-1 φαίνεται η σχέση των εγγράφων και των αντικειμένων πολυμέσων με τις λέξεις κλειδιά και τις κατηγορίες και υποκατηγορίες.

Οι λέξεις κλειδιά και οι κατηγορίες είναι οργανωμένες σε μία ιεραρχική δομή, όπου κάθε κόμβος της ιεραρχίας μπορεί να έχει παιδιά, τα οποία αποτελούν εξειδίκευση του κόμβου. Οι λέξεις κλειδιά και οι κατηγορίες δημιουργούν με αυτόν τον τρόπο δύο δέντρα, όπως φαίνεται και στο παρακάτω σχήμα.

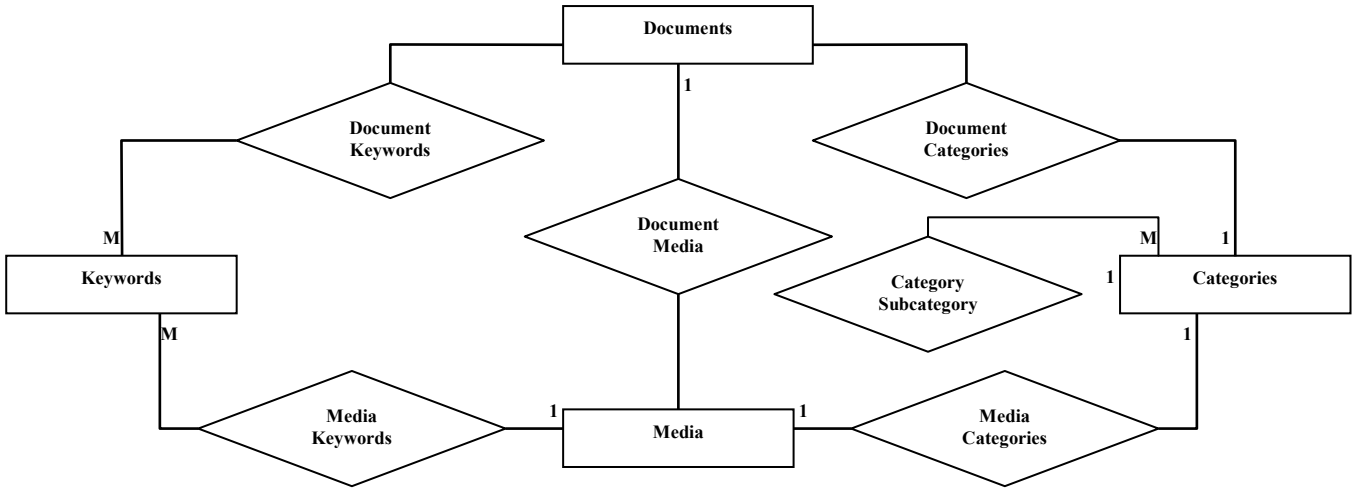


Σχήμα 6-2 Η ιεραρχική δομή των λέξεων κλειδιών και των κατηγοριών. Τα *L1* ως *Ln* είναι οι διαθέσιμες μεταφράσεις του κάθε κόμβου στις γλώσσες τους συστήματος. Το σύμβολο *i* συμβολίζει τους κόμβους του δέντρου. Για τον κόμβο *i1* φαίνονται και 2 επίπεδα εξειδικεύσεων. Επίσης διακρίνεται το επιτρεπτό βάθος των επιπέδων, όπου για τις κατηγορίες επιτρέπεται μόνο ένα επίπεδο εξειδικεύσεων, ενώ για τις λέξεις κλειδιά δεν υπάρχει περιορισμός.

Το σύνολο των λέξεων της ιεραρχίας λέξεων κλειδιών και το σύνολο των κατηγοριών της ιεραρχίας των κατηγοριών, συνθέτουν το λεξικό λέξεων κλειδιών και το λεξικό κατηγοριών του συστήματος, αντίστοιχα. Όπως φαίνεται και στο παραπάνω σχήμα, οι λέξεις κλειδιά μπορούν να έχουν απεριόριστο βάθος εξειδικεύσεων, ενώ οι κατηγορίες μόνο ένα επίπεδο εξειδικεύσεων. Όπως όλο το σύστημα, έτσι και οι κόμβοι των λεξικών υποστηρίζουν απεριόριστο αριθμό μεταφράσεων για κάθε λέξη και κατηγορία, όπως θα περιγραφεί στο κεφάλαιο για την Υποστήριξη Πολυγλωσσικού Περιεχομένου.

Τα λεξικά των λέξεων κλειδιών και των κατηγοριών είναι διαθέσιμα στον τελικό χρήστη για την κατασκευή ερωτήσεων αναζήτησης εγγράφων, όπως θα περιγραφεί στο κεφάλαιο για το Υποσύστημα Αναζήτησης Εγγράφων, καθώς και στον αρθρογράφο ώστε να μπορεί να χαρακτηρίσει τα έγγραφά του.

Το τμήμα του διαγράμματος Οντοτήτων-Σχέσεων που αναφέρεται στις κατηγορίες και στις λέξεις κλειδιά, και είχε παρουσιαστεί στο κεφάλαιο για τη Βάση Δεδομένων, επαναλαμβάνεται στο Σχήμα 6-3.



Σχήμα 6-3 Το διάγραμμα Οντοτήτων-Σχέσεων που αναφέρεται στις κατηγορίες και στις λέξεις κλειδιά, και στις σχέσεις τους με τα έγγραφα του συστήματος

Επειδή ένα έγγραφο μπορεί να χαρακτηριστεί από απεριόριστο αριθμό λέξεων κλειδιών, υπάρχει ο κίνδυνος, στην προσπάθεια να καλυφθούν όλες οι πτυχές του περιεχομένου του εγγράφου, να επιλεχτούν πολλές λέξεις κλειδιά, ακόμα και αν κάποιες από αυτές δεν έχουν μεγάλη σημασία ως προς το κύριο περιεχόμενο του εγγράφου. Η αναζήτηση κάποιου εγγράφου σε αυτήν την περίπτωση, θα είχε ως αποτέλεσμα την επιστροφή μεγάλου αριθμού εγγράφων χωρίς να υπάρχει τρόπος αναγνώρισης και διαχωρισμού ανάλογα με τη σημαντικότητα σε σχέση με τις επιλεγμένες λέξεις κλειδιά. Για να μπορεί να εξαχθεί κάποια μετρική σχετικότητας των αποτελεσμάτων μιας ερώτησης σε σχέση με την υποβληθείσα ερώτηση, αλλά και για να μπορεί ο αρθρογράφος να χαρακτηρίσει τα έγγραφα με μεγαλύτερη ακρίβεια, εισήχθηκε στο χαρακτηρισμό των εγγράφων με λέξεις κλειδιά και η έννοια της βαρύτητας μιας λέξης κλειδί. Αυτό σημαίνει ότι όταν επιλέγεται μία λέξη κλειδί για κάποιο έγγραφο, η επιλογή αυτή συνοδεύεται και από τη σημασία που έχει αυτή η λέξη για το συγκεκριμένο έγγραφο.

Έτσι η σχέση που συνδέει τα έγγραφα με τις λέξεις κλειδιά, έχει και ένα πεδίο για το βάρος της λέξης στο έγγραφο, όπως φαίνεται στον πίνακα *document_keywords* στο Σχήμα 5-2 που παρουσιάζει το σχεσιακό μοντέλο της Βάσης Δεδομένων. Στο ίδιο σχήμα φαίνεται ότι για κάθε μία από τις κατηγορίες (πίνακας *categories*) υπάρχει και ένα πεδίο που κρατάει τον αριθμό των εγγράφων που ανήκουν στη συγκεκριμένη κατηγορία, ενώ υπάρχει και η σχέση *statistics* στην οποία αποθηκεύονται στατιστικά στοιχεία για κάθε κατηγορία και για κάθε λέξη κλειδί. Τα παραπάνω δεδομένα χρησιμοποιούνται για τον υπολογισμό της σχετικότητας των εγγράφων του

αποτελέσματος μιας ερώτησης σε σχέση με την υποβληθείσα ερώτηση, όπως θα περιγραφεί στο επόμενο κεφάλαιο και αναφέρεται στο [Σκον98].

6.3 Εργαλεία Διαχείρισης και Κατηγοριοποίησης

Στο «Hypermedia Custom News System» απαιτεί σε διάφορα τμήματά του την πρόσβαση στις λέξεις κλειδιά και στις κατηγορίες. Πέρα από την ανάγκη ύπαρξης μεθόδων ανάκτησης των λεξικών από τη βάση δεδομένων, υπάρχει και ανάγκη αναπαράστασης των λεξικών σε δενδρική μορφή, τόσο εσωτερικά στις εφαρμογές, όσο και στο επίπεδο αλληλεπίδρασης με τους χρήστες (User Interfaces). Χρειάζεται ακόμα λειτουργικότητα για τη διαχείριση των λεξικών, συμπεριλαμβανομένων και λειτουργιών για εισαγωγή, διαγραφή και μετατροπή λέξεων κλειδιών και κατηγοριών.

Επειδή οι παραπάνω λειτουργίες χρησιμοποιούνται με όμοιο τρόπο σε πολλά σημεία του Συστήματος, δημιουργήθηκε η ανάγκη για γενίκευση όλων των λειτουργιών και απομόνωσή τους σε όσο το δυνατόν μεγαλύτερο βαθμό, έτσι ώστε να μπορούν να ενσωματωθούν χωρίς αλλαγές σε οποιοδήποτε σημείο είναι απαραίτητο.

6.3.1 Βασικές Λειτουργίες

Στις βασικές λειτουργίες που υλοποιήθηκαν στο σύστημα κατηγοριοποίησης των εγγράφων, περιλαμβάνονται μηχανισμοί για ανάκτηση των δεδομένων για τις λέξεις κλειδιά και τις κατηγορίες από τη βάση δεδομένων, την αναπαράσταση των κόμβων των λεξικών σε κλάσεις Java και την κατασκευή των ιεραρχικών δομών για το χειρισμό των δέντρων. Στη δομή του δέντρου, αναπτύχθηκαν λειτουργίες για εισαγωγή και διαγραφή παιδιών κάποιου κόμβου, για την αναζήτηση κόμβων με βάση το μοναδικό κωδικό τους και για την επεξεργασία των τιμών των χαρακτηριστικών των κόμβων για την περαιτέρω χρησιμοποίησή τους.

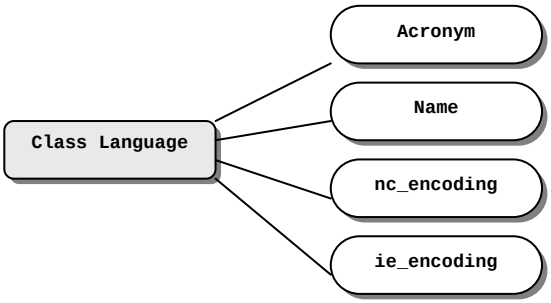
Υλοποιήθηκαν ακόμα λειτουργίες για την ανάκτηση των δεδομένων για τις υποστηριζόμενες γλώσσες του συστήματος και για την αναζήτηση πληροφορίας με βάση το ακρωνύμιο μιας γλώσσας.

Τέλος αναπτύχθηκε μια γενικευμένη προσέγγιση για το χειρισμό της σύνδεσης με τη βάση δεδομένων, η οποία θα περιγραφεί στο κεφάλαιο 11 Πρόσβαση σε Βάση Δεδομένων με JDBC.

6.3.1.1 Υποστήριξη Γλωσσών

Για την υποστήριξη πολλών γλωσσών (βλέπε και κεφάλαιο 9: Υποσύστημα Υποστήριξης Πολυγλωσσικού Περιεχομένου) χρειάζεται κατ' αρχήν η ανάκτηση της πληροφορίας από τη βάση δεδομένων. Για κάθε υποστηριζόμενη γλώσσα η πληροφορία που διατίθεται είναι το όνομα της γλώσσας, το ακρωνύμιο της γλώσσας και οι κωδικοί κωδικοποίησης χαρακτήρων που χρησιμοποιούνται για τη μετατροπή κειμένων από 8-bit σε 16-bit Unicode χαρακτήρες.

Μετά την ανάκτηση της πληροφορίας από τη βάση δεδομένων, κατασκευάζεται για κάθε γλώσσα ένα αντικείμενο όπως φαίνεται στο Σχήμα 6-4



Σχήμα 6-4 Η κλάση *Language* που κρατάει τις απαραίτητες πληροφορίες για κάθε υποστηριζόμενη γλώσσα του συστήματος

Για κάθε ένα από τα χαρακτηριστικά της κλάσης υπάρχουν οι αντίστοιχες *getXXX()* και *setXXX()* μέθοδοι για την αλλαγή και ανάκτηση των τιμών τους.

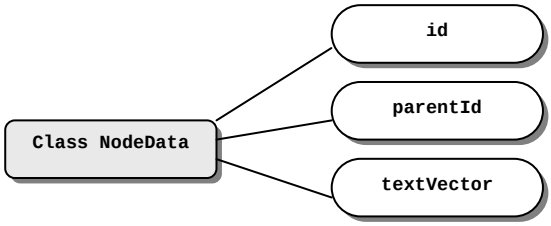
Το σύνολο των αντικειμένων για τις γλώσσες χρησιμοποιείται για την κατασκευή μιας λίστας, η οποία παρέχει μεθόδους για την ανάκτηση πληροφοριών για τη γλώσσα με βάση κάποιον δείκτη. Αυτή η λίστα χρησιμοποιείται αυτούσια για την αναπαράσταση των γλωσσών στο επίπεδο αλληλεπίδρασης (User Interface) σε όλες τις εφαρμογές, και ο δείκτης με βάση τον οποίον γίνεται η ανάκτηση πληροφορίας, είναι αυτός της επιλεγμένης γλώσσας.

6.3.1.2 Αναπαράσταση Κατηγοριών και Λέξεων Κλειδιών

Παρόμοια με τη διαδικασία που ακολουθείται για τις υποστηριζόμενες γλώσσες, για τις κατηγορίες και για τις λέξεις κλειδιά πρέπει πρώτα να ανακτηθεί η πληροφορία από τη βάση δεδομένων, να οργανωθεί σε κατάλληλες δομές για κάθε κόμβο και να

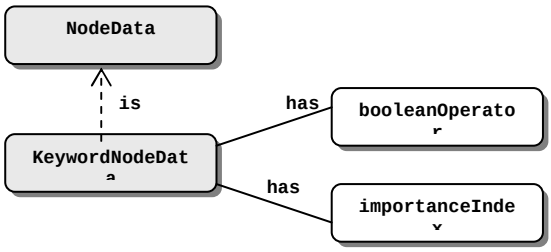
σχηματιστεί η ιεραρχική δομή των λεξικών. Η δυσκολία που συναντάται σε αυτή την περίπτωση, είναι η πιο πολύπλοκη δομή του ιεραρχικού μοντέλου των λεξικών.

Αρχικά γίνεται η ανάκτηση των κατηγοριών και των λέξεων κλειδιών από τη βάση δεδομένων, και δημιουργείται για κάθε μία από αυτές ένα αντικείμενο όπως αυτό που φαίνεται στο Σχήμα 6-5



Σχήμα 6-5 Αντικείμενα που χρησιμοποιούνται για την αναπαράσταση των κατηγοριών και των λέξεων κλειδιά

Ειδικά για τις λέξεις κλειδιά, δημιουργούνται αντικείμενα του τύπου όπως φαίνεται στο Σχήμα 6-6, και έχουν επιπλέον στοιχεία, απαραίτητα μόνο για τις λέξεις κλειδιά.



Σχήμα 6-6 Ειδικά για τις λέξεις κλειδιά, δημιουργούνται αντικείμενα τύπου *KeywordNodeData* επειδή για αυτά απαιτείται η φύλαξη επιπλέον πληροφορίας

Τα πεδία των δύο κλάσεων, που περιγράφηκαν πριν, είναι τα εξής:

id	Το μοναδικό αναγνωριστικό των λέξεων κλειδιών ή κατηγοριών
parentId	το μοναδικό αναγνωριστικό του πατέρα κάποιου κόμβου
textVector	διάνυσμα με τις μεταφράσεις του όρου (κατηγορίας ή λέξης κλειδί) για κάθε γλώσσα του συστήματος
booleanOperator	Οι υποστηριζόμενοι δυαδικοί τελεστές (AND και OR)
importanceIndex	η βαρύτητα της λέξης κλειδί

Πίνακας 6-1 Τα χαρακτηριστικά των κλάσεων NodeData και KeywordNodeData

Για τις παραπάνω δύο κλάσεις υπάρχουν οι κατάλληλες μέθοδοι για την αλλαγή των τιμών των χαρακτηριστικών τους, και κυρίως για την ανάκτηση της μετάφρασης ενός κόμβου σε μία συγκεκριμένη γλώσσα.

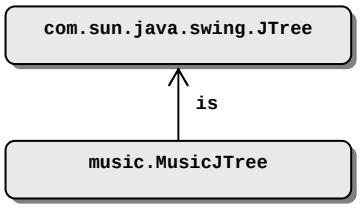
Το σύνολο των αντικειμένων των λέξεων κλειδιών και των κατηγοριών, χρησιμοποιείται για την κατασκευή των δενδρικών δομών που αναπαριστούν τις ιεραρχίες των κατηγοριών και των λέξεων κλειδιών.

6.3.2 Στοιχεία του Επιπέδου Αλληλεπίδρασης (User Interfaces)

Πέραν των βασικών λειτουργιών, όσον αφορά την ανάκτηση των κατηγοριών και των λέξεων κλειδιών, καθώς και την υποστήριξη πολλών γλωσσών, αναπτύχθηκαν στοιχεία επικοινωνίας ανθρώπων- υπολογιστή (User Interface Components) που χρησιμοποιούν τις παραπάνω λειτουργίες για την απεικόνιση (visualization) των δεδομένων σε εφαρμογές. Αυτό περιλαμβάνει την εμφάνιση της δενδρικής δομής των λεξικών των κατηγοριών και των λέξεων κλειδιών, καθώς και την εμφάνιση των υποστηριζόμενων γλωσσών του συστήματος. Πέραν της εμφάνισης των δεδομένων, υλοποιήθηκαν και λειτουργίες για την εύκολη διαχείριση και επιλογή των δεδομένων, καθώς και APIs (Application Programming Interfaces) για την αλληλεπίδραση των στοιχείων μεταξύ τους, έτσι ώστε να είναι δυνατή η επαναχρησιμοποίησή τους ως συστατικών σε άλλες εφαρμογές.

6.3.2.1 Εμφάνιση Ιεραρχίας Λεξικών και Διαχείριση Κόμβων

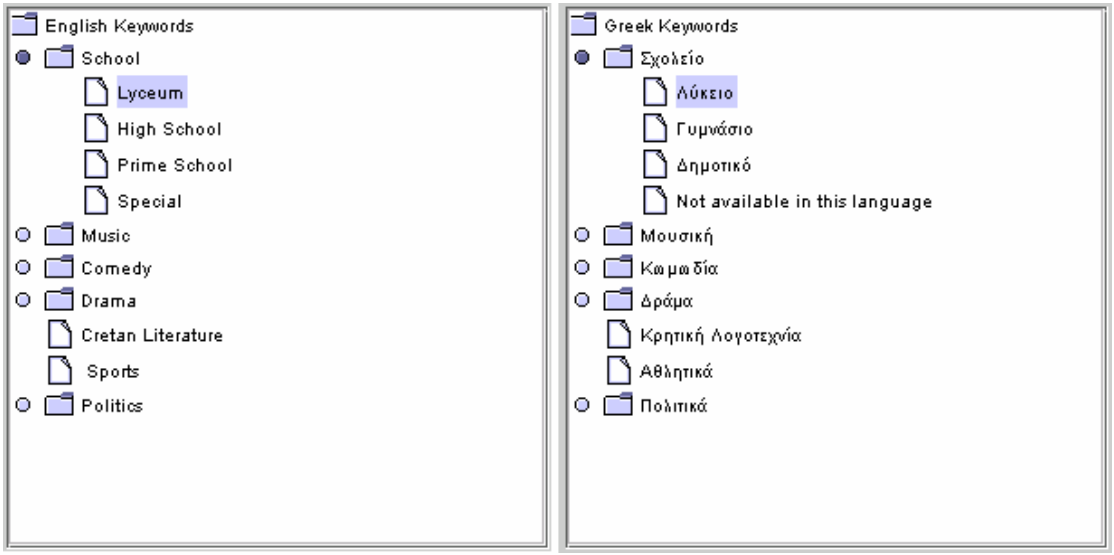
Για την εμφάνιση της δενδρικής δομής των λεξικών, επεκτάθηκε το συστατικό (component) που παρέχεται στο πακέτο Swing (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης) με τις επιπλέον λειτουργίες που απαιτούνται από το συγκεκριμένο σύστημα.



Σχήμα 6-7 Η κλάση *music.MusicJTree* που υλοποιήθηκε για την εμφάνιση πολυγλωσσικών κόμβων σε ιεραρχική δομή επεκτείνει (extends) την κλάση του Swing.

Οι κόμβοι του δένδρου μπορεί να είναι κλάσεις του τύπου *NodeData*, ενώ υπάρχουν οι κατάλληλες μέθοδοι για την αλλαγή γλώσσας εμφάνισης. Έτσι, όταν πρόκειται να εμφανιστεί το δένδρο κάποιου λεξικού, ως είσοδος λαμβάνεται η δενδρική δομή που έχει κατασκευαστεί από τα δεδομένα της βάσης δεδομένων, και εμφανίζεται κάθε κόμβος, με ετικέτα στη γλώσσα που είναι επιλεγμένη (βλέπε Εικόνα 6-1).

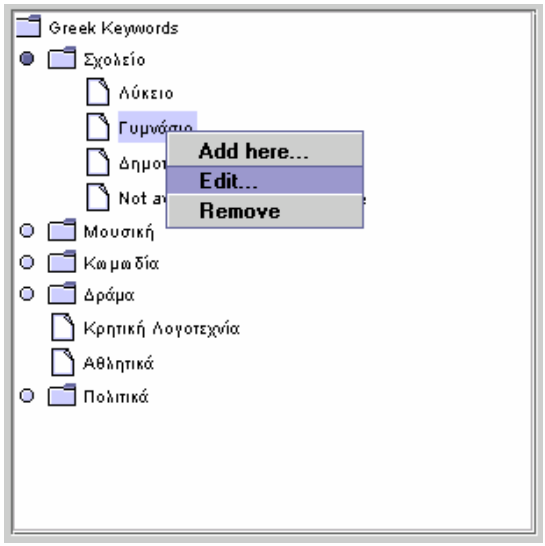
Το κείμενο που πρέπει να εμφανιστεί κάθε φορά, λαμβάνεται από την κλάση *NodeData* δίνοντας ως είσοδο το δείκτη της επιλεγμένης γλώσσας. Η αλλαγή της



επιλεγμένης γλώσσας εκφυλίζεται με αυτό τον τρόπο στην απλή αλλαγή ενός δείκτη.

Εικόνα 6-1 Στην εικόνα εμφανίζονται δύο στιγμιότυπα του δέντρου του ίδιου λεξικού, με επιλεγμένη στα αριστερά την αγγλική γλώσσα και στα δεξιά την ελληνική γλώσσα. Όπως φαίνεται στο αριστερό δέντρο, υπάρχει στην 6η γραμμή ένας κόμβος για τον οποίο δεν υπάρχει ελληνική μετάφραση, και εμφανίζεται άντ' αυτού ένα επεξηγηματικό κείμενο

Πέραν των παραπάνω, το συστατικό (component) για την εμφάνιση των λεξικών, επεκτάθηκε ώστε να υποστηρίζει και τα γνωστά PopUp Menus (μενού που εμφανίζονται σε οποιοδήποτε σημείο της οθόνης, πατώντας το δεξί κουμπί του ποντικιού) για να διευκολυνθεί η διαχείριση των κόμβων και των λεξικών. Στην Εικόνα 6-2 εμφανίζεται ένα στιγμιότυπο με ανοιγμένο το μενού PopUp.



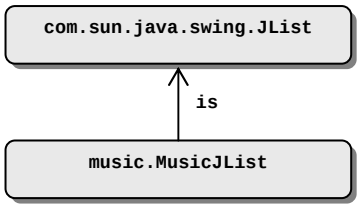
Εικόνα 6-2 Το δέντρο ενός λεξικού κατηγοριοποίησης με ανοιγμένο το PopUp Menu

6.3.2.2 Εμφάνιση Επιλεγμένων Κόμβων

Σε διάφορα σημεία του παρόντος συστήματος νέων, υπάρχει η ανάγκη εμφάνισης ενός υποσυνόλου των λεξικών των κατηγοριών ή των λέξεων κλειδιών σε μορφή λίστας. Αυτό συμβαίνει κατά την επιλογή των λέξεων κλειδιών που χαρακτηρίζουν ένα έγγραφο, αλλά και κατά τη διαδικασία επιλογής των λέξεων κλειδιών και των κατηγοριών για την κατασκευή ερωτήσεων αναζήτησης εγγράφων.

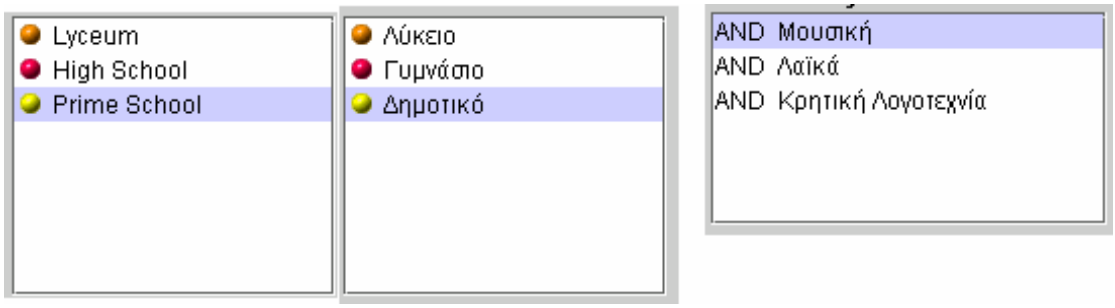
Όμοια με το συστατικό (component) για την εμφάνιση της δεντρικής μορφής των λεξικών, τα δεδομένα που πρέπει να εμφανιστούν φυλάσσονται σε αντικείμενα των κλάσεων *NodeData* και *KeywordNodeData* και παρέχονται μεταφράσεις σε πολλές γλώσσες. Η ιδιαιτερότητα που υπάρχει σε αυτήν την περίπτωση είναι ότι κατά την κατηγοριοποίηση οι λέξεις κλειδιά χαρακτηρίζονται και από τη βαρύτητα που έχουν για το συγκεκριμένο έγγραφο, ενώ κατά την κατασκευή ερώτησης αναζήτησης συνοδεύονται και από τους τελεστές AND ή OR.

Για την εμφάνιση των επιλεγμένων κόμβων, επεκτάθηκε το συστατικό (component) που παρέχεται στο πακέτο Swing (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης) με τις απαιτούμενες λειτουργίες για την εμφάνιση των στοιχείων για τη βαρύτητα και τους τελεστές.



Σχήμα 6-8 Η κλάση *music.MusicList* που υλοποιήθηκε για την εμφάνιση επιλεγμένων κόμβων των λεξικών βασίζεται στην κλάση του Swing.

Στην Εικόνα 6-3 εμφανίζονται στιγμιότυπα του συστατικού για την εμφάνιση επιλεγμένων κόμβων. Φαίνεται η αλλαγή γλώσσας εμφάνισης και τα εικονίδια διαφορετικών χρωμάτων που υποδηλώνουν τη βαρύτητα της λέξης κλειδί.. Επίσης φαίνονται στο τρίτο τμήμα οι τελεστές AND με τους οποίους θα συνδεθούν οι λέξεις κλειδιά για την κατασκευή μιας ερώτησης



Εικόνα 6-3 Στιγμιότυπα από το συστατικό για την εμφάνιση επιλεγμένων κόμβων των λεξικών.

6.3.3 Συστατικό Διαχείρισης Κατηγοριών κ' Λέξεων Κλειδιών και Κατηγοριοποίησης Εγγράφων

Οι βασικές λειτουργίες που περιγράφηκαν παραπάνω, μαζί με τα συστατικά (components) για την εμφάνιση της ιεραρχίας των λεξικών και των επιλεγμένων κόμβων, χρησιμοποιήθηκαν στην ανάπτυξη ενός νέου, συνδυασμένου συστατικού, το οποίο δίνει τη δυνατότητα επιλογής λέξεων κλειδιών με συγκεκριμένο βάρος καθώς και μιας κατηγορίας ή υποκατηγορίας. Το νέο συστατικό διαθέτει επίσης συγκεκριμένο και σαφώς ορισμένο API για την επικοινωνία με άλλες εφαρμογές ή συστατικά. Συγκεκριμένα παρέχει δυνατότητα για αρχικοποίηση των λεξικών από τη βάση δεδομένων, απεπιλογή οποιονδήποτε λέξεων κλειδιών ή κατηγοριών. Παρέχει ακόμα μεθόδους για την ανάκτηση των επιλεγμένων λέξεων κλειδιών και κατηγορίας σε μία συγκεκριμένη δομή, καθώς και μεθόδους για προεπιλογή συγκεκριμένων κόμβων. Τέλος δίνεται η δυνατότητα να καθοριστεί κατά πόσο επιτρέπονται πράξεις

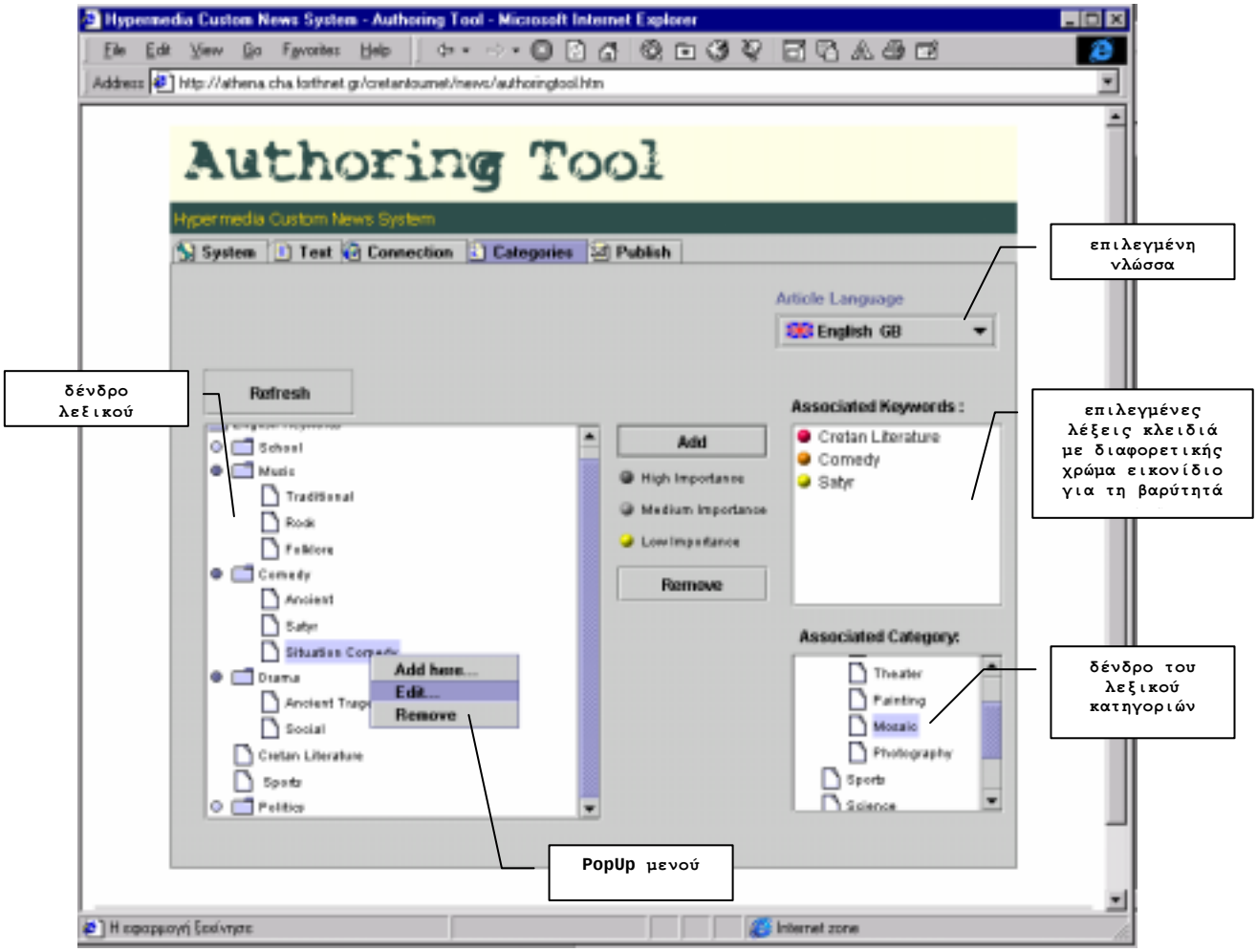
διαχείρισης των κόμβων των δέντρων. Αν οι πράξεις αυτές δεν επιτρέπονται, απενεργοποιούνται τα PopUp Menu των δέντρων.

Το παραπάνω συστατικό χρησιμοποιήθηκε αυτούσιο στο υποσύστημα συγγραφής άρθρων του «Hypermedia Custom News System» ([Πετρ98]) για την κατηγοριοποίηση των εγγράφων και των αντικειμένων πολυμέσων. Στην περίπτωση ενός νέου εγγράφου, ο χρήστης δια μέσω του συστατικού κατηγοριοποίησης επιλέγει κάποιες λέξεις κλειδιά, προσδίνοντας σε αυτές και κάποιο βάρος, και επιλέγει και μία κατηγορία ή υποκατηγορία. Χρησιμοποιώντας το API του συστατικού, το υποσύστημα συγγραφής άρθρων γνωρίζει τα χαρακτηριστικά κατηγοριοποίησης που επέλεξε ο αρθρογράφος, και ενημερώνει κατάλληλα τη βάση δεδομένων. Στην περίπτωση της μετατροπής ενός υπάρχοντος εγγράφου, χρησιμοποιώντας το API του συστατικού, το υποσύστημα συγγραφής άρθρων προεπιλέγει τους κόμβους με τους οποίους είχε χαρακτηριστεί το έγγραφο, ενώ στη συνέχεια ο αρθρογράφος έχει τη δυνατότητα να μετατρέψει τις τιμές των χαρακτηριστικών αυτών.

Ειδικά για τα αντικείμενα πολυμέσων που διαχειρίζεται το υποσύστημα συγγραφής άρθρων, χρησιμοποιείται το ίδιο συστατικό για το χαρακτηρισμό τους, με τη διαφορά ότι σε αυτή την περίπτωση δεν επιλέγονται βάρη στην επιλογή των λέξεων κλειδιών.

Τέλος, ανάλογα με τις δυνατότητες πρόσβασης του αρθρογράφου που χρησιμοποιεί το σύστημα, το υποσύστημα συγγραφής άρθρων, ενεργοποιεί ή απενεργοποιεί τη δυνατότητα διαχείρισης των κόμβων των λεξικών.

Στην Εικόνα 6-4 φαίνεται το συστατικό κατηγοριοποίησης εγγράφων κ' διαχείρισης λεξικών, όπως αυτό έχει ενσωματωθεί στο υποσύστημα συγγραφής άρθρων.



Εικόνα 6-4 Το συστατικό κατηγοριοποίησης εγγράφων και διαχείρισης λεξικών, όπως έχει ενσωματωθεί στο υποσύστημα συγγραφής εγγράφων

6.4 Ανακεφαλαίωση

Σε αυτό το κεφάλαιο παρουσιάστηκε το Υποσύστημα Κατηγοριοποίησης που είναι υπεύθυνο για τη διαχείριση του μοντέλου κατηγοριοποίησης που ακολουθεί το σύστημα καθώς και για την κατάταξη σύμφωνα με αυτό των άρθρων και των αντικειμένων πολυμέσων που εισάγονται στο σύστημα. Το Υποσύστημα Κατηγοριοποίησης είναι ενσωματωμένο στο Υποσύστημα Συγγραφής Άρθρων.

Τα επίπεδα κατηγοριοποίησης του συστήματος είναι δύο, κατηγορίες και υποκατηγορίες, ενώ διατηρείται και μια ιεραρχία λέξεων κλειδιών αυθαίρετου αριθμού επιπέδων, που χρησιμοποιούνται για το χαρακτηρισμό άρθρων και αντικειμένων πολυμέσων.

Περιγράφηκαν τα συστατικά αλληλεπίδρασης που αναπτύχθηκαν για την υποστήριξη των λεξικών σε επίπεδο εφαρμογής και οι βασικές λειτουργίες και οι κλάσεις για την ανάκτηση των δεδομένων και τη διαχείρισή τους από τις εφαρμογές.

Όλες οι λειτουργίες αναπτύχθηκαν με γνώμονα την ανεξαρτησία τους από τη συγκεκριμένη εφαρμογή και τη δυνατότητα επαναχρησιμοποίησής τους από άλλες εφαρμογές.

7 Υποσύστημα Αναζήτησης Εγγράφων

Στον τελικό χρήστη του εξατομικευμένου συστήματος νέων κατ' απαίτηση «Hypermedia Custom News System» που αναπτύχθηκε, το σύνολο του συστήματος κρίνεται από το περιεχόμενο των εγγράφων που του παρουσιάζονται έπειτα από την εκτέλεση μιας ερώτησης με τις παραμέτρους που αυτός έχει επιλέξει. Το αποτέλεσμα των ερωτήσεων εξαρτάται βεβαίως από τις επιλογές που κάνει ο τελικός χρήστης για την αναζήτηση, αλλά τα θεμέλια των επιλογών στηρίζονται στο μοντέλο κατηγοριοποίησης με το οποίο κατατάσσονται τα έγγραφα και το οποίο περιγράφηκε στο προηγούμενο κεφάλαιο, αλλά και στα εργαλεία που του προσφέρονται για να εκφράσει τις επιθυμίες του. Συμπερασματικά, η αποδοτικότητα των εργαλείων αναζήτησης που προσφέρονται στον τελικό χρήστη, κρίνει σε μεγάλο βαθμό τη χρησιμότητα του συστήματος νέων.

Οι χρήστες του υποσυστήματος αναζήτησης εγγράφων, μπορεί να είναι είτε οι αρθρογράφοι που επιθυμούν να αναζητήσουν έγγραφα που έχουν ήδη καταχωρηθεί ώστε να τα συνδέσουν με το έγγραφο που συνθέτουν, είτε οι τελικοί χρήστες που επιθυμούν να πληροφορηθούν για θέματα που τους ενδιαφέρουν.

Σε αυτό το κεφάλαιο θα περιγραφεί το υποσύστημα αναζήτησης εγγράφων του εξατομικευμένου συστήματος νέων κατ' απαίτηση. Θα αναλυθούν οι μηχανισμοί που υλοποιήθηκαν για την αναζήτηση εγγράφων και πως αυτοί χρησιμοποιούνται στα διάφορα υποσυστήματα, και θα περιγραφεί το Επίπεδο Αλληλεπίδρασης (User Interface) που χρησιμοποιείται από τους τελικούς χρήστες αλλά και από τους αρθρογράφους.

7.1 Ανάλυση απαιτήσεων

Όπως αναφέρθηκε, υπάρχουν δύο είδη χρηστών του υποσυστήματος αναζήτησης εγγράφων:

1. οι **αρθρογράφοι** που χρησιμοποιούν το υποσύστημα συγγραφής άρθρων ([Πετρ98]) και επιθυμούν μέσα από το υποσύστημα συγγραφής άρθρων να αναζητήσουν έγγραφα του συστήματος με σκοπό να τα χρησιμοποιήσουν ως συσχετιζόμενα άρθρα, ή ακόμα και να τα αλλάξουν

2. οι **τελικοί χρήστες** που επιθυμούν ένα εύχρηστο εργαλείο για την επιλογή των προτιμήσεών τους ή την αναζήτηση πολύ συγκεκριμένης πληροφορίας

Οι δύο αυτές ομάδες δεν έχουν ακριβώς τις ίδιες ανάγκες για την αναζήτηση εγγράφων. Και τα δύο είδη χρηστών απαιτούν ένα εργαλείο από το οποίο μπορούν να επιλέξουν τις λέξεις κλειδιά και τις κατηγορίες από το μοντέλο κατηγοριοποίησης, για τις οποίες ενδιαφέρονται. Επίσης χρειάζεται να μπορούν να δουν τα δεδομένα σε όλες τις υποστηριζόμενες γλώσσες του συστήματος, χωρίς να υπάρχει κάποιος ιδιαίτερος περιορισμός.

Άλλοι παράμετροι που απαιτούνται από κοινού από τις δύο ομάδες, είναι χρονικοί περιορισμοί ως προς την ημερομηνία συγγραφής των άρθρων καθώς και παράμετροι για τη σειρά εμφάνισης των αποτελεσμάτων. Ένα πολύ σημαντικό σημείο είναι η δυνατότητα υπολογισμού κάποιας μετρικής για τη σχετικότητα των εγγράφων που επιστρέφονται από την εκτέλεση μίας ερώτησης, σε σχέση με την εκτελεσθείσα ερώτηση, δηλαδή το βαθμό εκπλήρωσης των παραμέτρων της ερώτησης από ένα έγγραφο.

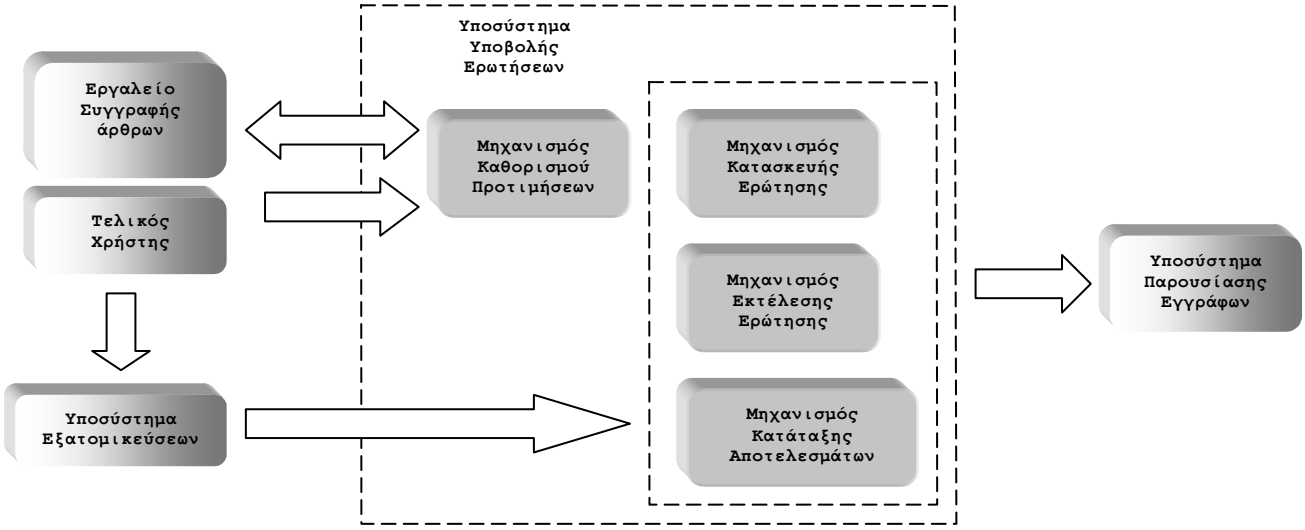
Πέραν των παραπάνω απαιτήσεων, ειδικά οι αρθρογράφοι χρειάζονται και τη δυνατότητα αναζήτησης, πέραν των κλασικών εγγράφων, και αντικειμένων πολυμέσων που διαχειρίζεται το σύστημα νέων. Το τελευταίο είναι ιδιαίτερα χρήσιμο όταν οι αρθρογράφοι επιθυμούν να εμπλουτίσουν το έγγραφο με διάφορα αντικείμενα πολυμέσων (εικόνες, video κλπ.) αλλά δεν έχουν οι ίδιοι διαθέσιμα κατάλληλα αρχεία πολυμέσων. Σε αυτή την περίπτωση μπορούν να αναζητήσουν πολυμέσα που είναι ήδη καταχωρημένα στο σύστημα και να τα χρησιμοποιήσουν. Επιπρόσθετα οι αρθρογράφοι χρειάζονται κάποια λειτουργικότητα για να περιορίσουν τα αποτελέσματα μιας ερώτησης σε σχέση με τον αρθρογράφο, δηλαδή να απαιτήσουν να εξαχθούν μόνο τα έγγραφα που έχουν συνθέσει οι ίδιοι.

Τέλος, επειδή το υποσύστημα αναζήτησης εγγράφων στην περίπτωση των αρθρογράφων είναι ενσωματωμένο στο υποσύστημα συγγραφής άρθρων, η έξοδος κάποιας ερώτησης πρέπει να έχει διαφορετική μορφή από την έξοδο στην περίπτωση του απλού χρήστη, επειδή τα δεδομένα του αποτελέσματος επεξεργάζονται και εμφανίζονται με διαφορετικό τρόπο στο υποσύστημα συγγραφής άρθρων.

Όμοια με το υποσύστημα κατηγοριοποίησης εγγράφων, και εδώ υπάρχει η ανάγκη ανεξαρτησίας του εργαλείου επιλογής προτιμήσεων ώστε να μπορεί να ενσωματωθεί και σε άλλες εφαρμογές. Σε επόμενο κεφάλαιο θα περιγραφεί η χρησιμοποίηση του εργαλείου στην εφαρμογή για τα διαγράμματα των χρηστών (User Profiles).

7.2 Μηχανισμοί για την Αναζήτηση Εγγράφων

Στο Σχήμα 7-1 διακρίνονται τα διάφορα τμήματα του υποσυστήματος αναζήτησης εγγράφων, και ο τρόπος με τον οποίο αυτά επικοινωνούν μεταξύ τους.



Σχήμα 7-1 Τα τμήματα του υποσυστήματος αναζήτησης εγγράφων και η αλληλεπίδραση μεταξύ τους και με τους χρήστες

Το υποσύστημα αποτελείται από το Μηχανισμό Καθορισμού Προτιμήσεων, το Μηχανισμό Κατασκευής Ερωτήσεων, το Μηχανισμό Εκτέλεσης Ερώτησης και το Μηχανισμό Κατάταξης Αποτελεσμάτων. Οι χρήστες του υποσυστήματος, που μπορεί να είναι είτε οι αρθρογράφοι είτε οι τελικοί χρήστες, αλληλεπιδρούν με το υποσύστημα δια μέσω του Μηχανισμού Καθορισμού Προτιμήσεων όπου επιλέγουν τις λέξεις κλειδιά και τις κατηγορίες για την αναζήτηση των εγγράφων. Στην περίπτωση που η αναζήτηση γίνει με βάση κάποιο διάγραμμα χρήστη (User Profile) δεν υπάρχει αυτό το στάδιο, επειδή οι χρήστες έχουν καθορίσει τις προτιμήσεις τους κατά τη διάρκεια δημιουργίας του διαγράμματος. Το αποτέλεσμα της αναζήτησης, που είναι ένα σύνολο εγγράφων, δίνεται στο Υποσύστημα Παρουσίασης Εγγράφων προς παρουσίαση όταν πρόκειται για τους τελικούς χρήστες, ή επιστρέφεται στο υποσύστημα συγγραφής άρθρων για περαιτέρω επεξεργασία.

7.2.1 Μηχανισμός Καθορισμού Προτιμήσεων

Ο μηχανισμός καθορισμού προτιμήσεων περιλαμβάνει όλες τις απαραίτητες μεθόδους για τον καθορισμό των κριτηρίων αναζήτησης εγγράφων. Στο μηχανισμό αυτό εμπεριέχονται και οι λειτουργίες για την ανάκτηση των ιεραρχικών δομών των λεξικών από τη βάση δεδομένων, όπως περιγράφηκε εκτενώς στο κεφάλαιο για την κατηγοριοποίηση εγγράφων. Αξίζει να σημειωθεί ότι χρησιμοποιούνται ακριβώς οι ίδιες δομές για την αναπαράσταση των κόμβων των λεξικών που χρησιμοποιούνται και για την κατηγοριοποίηση εγγράφων.

7.2.2 Μηχανισμός Κατασκευής Ερωτήσεων

Ο μηχανισμός κατασκευής ερωτήσεων συνθέτει την SQL ερώτηση με βάση τις επιλογές που έχουν γίνει από το μηχανισμό καθορισμού προτιμήσεων. Ο τύπος της ερώτησης διαφέρει ανάλογα με το αν πρόκειται για ερώτηση που γίνεται από τον τελικό χρήστη, ή αν πρόκειται για ερώτηση από αρθρογράφο. Στη δεύτερη περίπτωση η αρχική ερώτηση τροποποιείται για να συμπεριλάβει και τα πολυμέσα του συστήματος νέων, όπως θα περιγραφεί παρακάτω.

7.2.3 Μηχανισμός Εκτέλεσης Ερωτήσεων

Ο μηχανισμός εκτέλεσης ερωτήσεων αναλαμβάνει την επικοινωνία με τον εξυπηρετητή βάσης δεδομένων για την εκτέλεση της SQL ερώτησης και την ανάκτηση των αποτελεσμάτων. Μετά την ανάκτηση, τα αποτελέσματα οργανώνονται σε κατάλληλες δομές για την περαιτέρω επεξεργασία τους.

7.2.4 Μηχανισμός Κατάταξης Αποτελεσμάτων

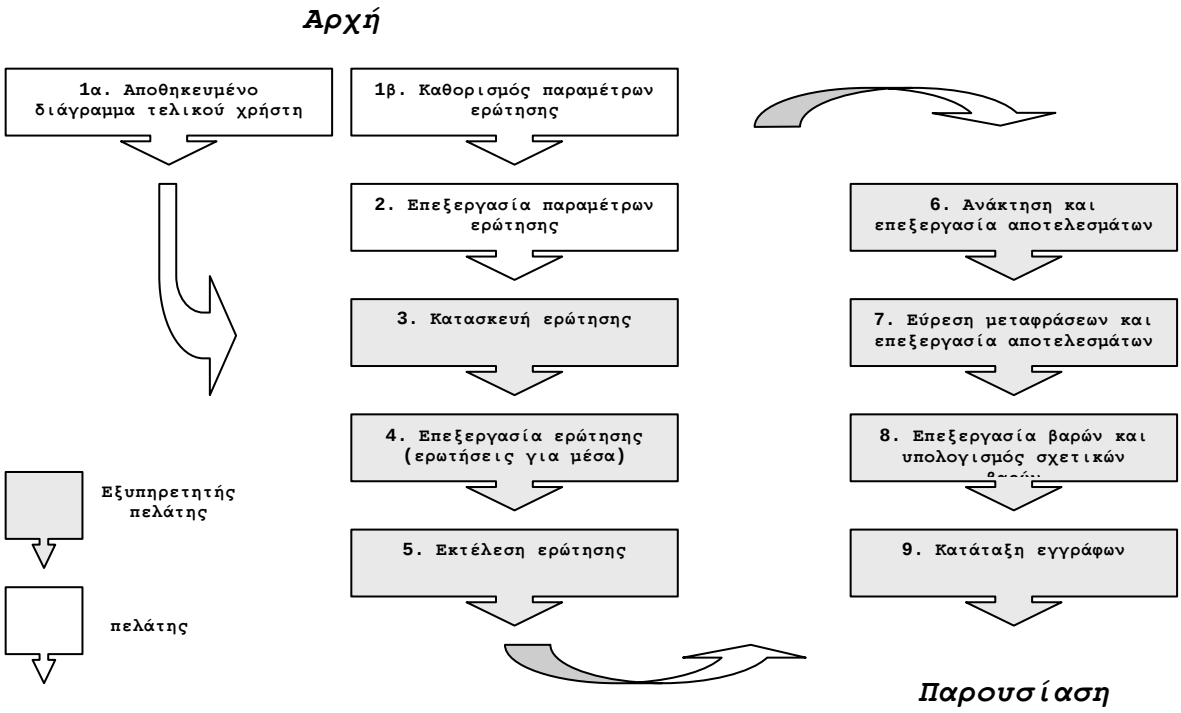
Έχοντας το σύνολο των αντικειμένων που εκπληρώνουν τα κριτήρια αναζήτησης, τα αποτελέσματα ομαδοποιούνται και κατατάσσονται σύμφωνα με τις παραμέτρους της ερώτησης. Αυτό περιλαμβάνει τον υπολογισμό της σχετικότητας κάθε εγγράφου, αλλά και την ομαδοποίηση ανά γλώσσα και μετάφραση.

7.2.5 Βήματα για την Αναζήτηση Εγγράφων

Η διαδικασία που ακολουθείται για την αναζήτηση εγγράφων περιλαμβάνει έναν αριθμό βημάτων, που αναλόγως με την περίπτωση εκτελούνται είτε στη μεριά του

πελάτη είτε στη μεριά του εξυπηρετητή. Η περίπτωση που διαφοροποιεί τον τόπο εκτέλεσης είναι αν πρόκειται για τον αρθρογράφο ή για τον τελικό χρήστη, που όπως αναφέρθηκε έχουν διαφορετικές ανάγκες ως προς τη χρησιμοποίηση των αποτελεσμάτων.

Στο παρακάτω σχήμα φαίνονται τα βήματα που ακολουθούνται για την υποβολή μιας ερώτησης αναζήτησης εγγράφων προς το σύστημα. Τα κουτιά με άσπρο φόντο αναφέρονται σε διαδικασίες που εκτελούνται αποκλειστικά στην πλευρά του πελάτη, ενώ τα κουτιά με γκρίζο φόντο, εκτελούνται ανάλογα με την περίπτωση είτε στον πελάτη είτε στον εξυπηρετητή.



Σχήμα 7-2 Διαδικασία εργασιών για την υποβολή ερώτησης. Τα σκιαγραμμένα τμήματα εκτελούνται είτε στην πλευρά του εξυπηρετητή είτε του πελάτη, ενώ τα υπόλοιπα τμήματα πάντα στην πλευρά του πελάτη.

Στη συνέχει θα περιγραφούν αναλυτικά όλα τα βήματα για την αναζήτηση εγγράφων.

1. Στην αρχή χρειάζεται να καθοριστούν οι παράμετροι για την υπό εκτέλεση ερώτηση. Αυτές οι παράμετροί μπορεί να είναι ήδη αποθηκευμένοι σε κάποιο διάγραμμα χρήστη (User Profile), οπότε με βάση τον κωδικό και το αναγνωριστικό (login και password) του χρήστη ανακτώνται από τη βάση

δεδομένων. Σε διαφορετική περίπτωση πρέπει να ακολουθηθεί η διαδικασία καθορισμού των παραμέτρων από το χρήστη, με τη βοήθεια της κατάλληλης εφαρμογής, που θα περιγραφεί παρακάτω. Οι παράμετροι της ερώτησης περιλαμβάνουν λέξεις κλειδιά, κατηγορίες, λογική σύνδεση AND ή OR για τις λέξεις κλειδιά, χρονικούς περιορισμούς συγγραφής των εγγράφων, επιθυμία κατάταξης, και ειδικά για την περίπτωση των αρθρογράφων περιλαμβάνονται και επιλογή για απλά έγγραφα ή αντικείμενα πολυμέσων και περιορισμοί για το συγγραφέα των εγγράφων.

2. Στην περίπτωση που η επιλογή των παραμέτρων της ερώτησης γίνει από την αρχή από τους χρήστες, απαιτείται μία επεξεργασία των παραμέτρων για τα περαιτέρω βήματα. Στην περίπτωση του αρθρογράφου, αυτή η επεξεργασία συνίσταται στην απλή μετατροπή τους σε μορφή που χρησιμοποιείται από τις μεθόδους αναζήτησης. Στην περίπτωση του τελικού χρήστη, οι παράμετροι πρέπει να προετοιμαστούν για να μεταφερθούν με κατάλληλο τρόπο στον εξυπηρετητή, ο οποίος θα αναλάβει την περαιτέρω εκτέλεση της ερώτησης. Η προετοιμασία περιλαμβάνει την κωδικοποίηση των παραμέτρων σε μορφή που να μπορούν να μεταφερθούν μέσω της HTTP GET μεθόδου, και τη μετατροπή των κωδικών των παραμέτρων σε αλφαριθμητικούς χαρακτήρες. Για παράδειγμα, χρειάζεται να μετατραπούν οι κωδικοί των επιλεγμένων λέξεων κλειδιών σε αλφαριθμητικούς χαρακτήρες, ενώ τα στοιχεία για τους χρονικούς περιορισμούς που περιλαμβάνουν και ειδικούς χαρακτήρες, πρέπει να κωδικοποιηθούν ώστε να μπορούν να αποτελούν χαρακτήρες για κατασκευή URL. Το τελευταίο είναι αναγκαίο επειδή υπάρχουν περιορισμοί στους χαρακτήρες που μπορούν να χρησιμοποιηθούν για την κατασκευής ενός URL.

3. Από τις παραμέτρους της ερώτησης κατασκευάζεται η SQL ερώτηση που πρέπει να εκτελεστεί για την ανάκτηση των εγγράφων. Η ερώτηση λαμβάνει υπ' όψιν όλες τις παραμέτρους που αναφέρθηκαν στο βήμα ένα. Αυτό που ενδιαφέρει σε αυτό το στάδιο είναι να ανακτηθούν όλα τα έγγραφα που ικανοποιούν τις παραπάνω συνθήκες, ανεξάρτητα από τη γλώσσα των εγγράφων.

4. Η ερώτηση που κατασκευάστηκε στο βήμα τρία, λαμβάνει υπ' όψιν τις σχέσεις της βάσης δεδομένων που αναφέρονται στα πλήρη έγγραφα. Όπως αναφέρθηκε και στο κεφάλαιο για τη βάση δεδομένων, η πληροφορία για τα αντικείμενα

πολυμέσων φυλάσσεται σε διαφορετικές σχέσεις, που διαφέρουν από τις σχέσεις για τα έγγραφα, όσον αφορά τις ερωτήσεις αναζήτησης, κυρίως στην ονομασία. Στην περίπτωση λοιπόν που η αναζήτηση γίνεται από τον αρθρογράφο, και αυτός επέλεξε να συμπεριλάβει στα αποτελέσματα και τα αντικείμενα πολυμέσων, η αρχική ερώτηση που κατασκευάστηκε πρέπει να τροποποιηθεί ελαφρώς ώστε ως αποτέλεσμα να δίνει τα αντικείμενα πολυμέσων.

5. Το βήμα της εκτέλεσης της ερώτησης περιλαμβάνει τη σύνδεση με τη βάση δεδομένων, τη δημιουργία των αντικειμένων εκτέλεσης της ερώτησης και μεταφοράς των δεδομένων. Η SQL έκφραση παραδίδεται στον εξυπηρετητή, ο οποίος αναλαμβάνει την εκτέλεσή της.

6. Μετά την εκτέλεση της ερώτησης από τη βάση δεδομένων, στην εφαρμογή επιστρέφεται ένα ειδικό αντικείμενο με τα αποτελέσματα της ερώτησης. Τα δεδομένων σε αυτό το βήμα ανακτώνται μέσω αυτού του αντικειμένου και επεξεργάζονται και οργανώνονται σε διανύσματα για την παροχή πρόσβασης σε αυτά.

7. Τα αποτελέσματα της ερώτησης, όπως αναφέρθηκε, δεν εξαρτώνται από τη γλώσσα στην οποία είναι γραμμένα τα έγγραφα. Αυτό σημαίνει κατ' αρχήν ότι υπάρχουν έγγραφα σε γλώσσες που δε ζήτησε ο χρήστης και ότι υπάρχουν περισσότερα του ενός έγγραφα που μπορεί να είναι το ένα μετάφραση του άλλου. Σε αυτό το βήμα, με δεδομένο το σύνολο των εγγράφων, εκτελείται μία ερώτηση στη βάση δεδομένων με την οποία επιστρέφονται τα έγγραφα ομαδοποιημένα αρχικά ανά ομάδα μεταφράσεων εγγράφων, και στη συνέχεια επιστρέφονται και οι μεταφράσεις των εγγράφων που δεν ήταν στο αποτέλεσμα της αρχικής ερώτησης. Η ερώτηση που εκτελείται είναι της μορφής

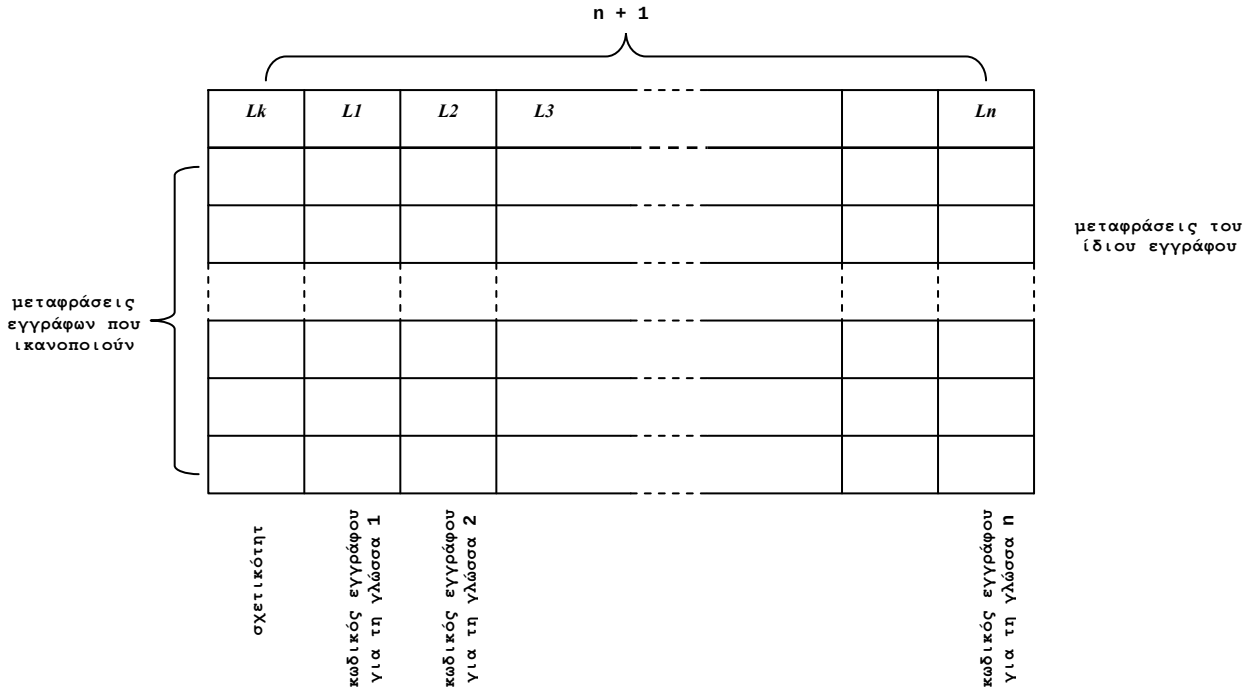
   select distinct ***L1***, ***L2***, …, ***Ln***
   from translations
   where (***L1*** in ***S***) OR (***L2*** in ***S***) OR ..... OR (***Ln*** in ***S***)

   όπου ***L1*** ως ***Ln*** είναι τα ακρωνύμια των υποστηριζόμενων γλωσσών και ***S*** το σύνολο των κωδικών των εγγράφων από την πρώτη ερώτηση. Κάθε εγγραφή που επιστρέφεται από την παραπάνω ερώτηση αντιστοιχεί σε ομάδα εγγράφων που το

ένα αποτελεί μετάφραση του άλλου, ενώ εξασφαλίζεται ότι τουλάχιστον ένα από τα έγγραφα κάθε εγγραφής, εκπληρώνει τα κριτήρια αναζήτησης.

8. Σε αυτό το βήμα υπολογίζονται από τα βάρη των λέξεων κλειδιών για κάθε έγγραφο και από τα συνολικά στατιστικά στοιχεία του συστήματος η σχετικότητα κάθε εγγράφου και αποθηκεύεται σε κατάλληλη δομή για την περαιτέρω επεξεργασία. Η διαδικασία για τον υπολογισμό της σχετικότητας θα περιγραφεί στο υποκεφάλαιο 7.3.

9. Στο τελευταίο βήμα, τα έγγραφα κατατάσσονται σύμφωνα με τη σχετικότητα που έχουν συναρτήσει των παραμέτρων της ερώτησης, και αποθηκεύονται σε μεταβλητές οι οποίες είναι προσβάσιμες από άλλα τμήματα του συστήματος. Η μορφή της δομής στην οποία αποθηκεύονται τα αποτελέσματα φαίνεται στο Σχήμα 7-3.



Σχήμα 7-3 Η δομή του τελικού αποτελέσματος μιας αναζήτησης. Ουσιαστικά πρόκειται για έναν πίνακα με στοιχεία κωδικούς εγγράφων. Ο πίνακας έχει *n*+1 στήλες, όπου *n* είναι ο αριθμός των υποστηριζόμενων γλωσσών. *Lk* είναι το ακρωνύμιο της γλώσσας ως προς την οποία έγινε η αναζήτηση, ενώ *L1* ως *Ln* τα ακρωνύμια των υποστηριζόμενων γλωσσών. Η πρώτη στήλη περιέχει τη βαρύτητα του εγγράφου, εφόσον υπάρχει στη γλώσσα αναζήτησης, ενώ οι υπόλοιπες τον κωδικό κάθε εγγράφου του ίδιου συνόλου μεταφράσεων.

7.3 Σχετικότητα Εγγράφων

Η έννοια «σχετικότητα εγγράφων» αναφέρεται σε δύο είδη πληροφορίας για τα αποτελέσματα μιας ερώτησης

1. το βαθμό εκπλήρωσης των κριτηρίων αναζήτησης βάση του περιεχομένου κάθε εγγράφου

2. τη σχέση κάθε εγγράφου με τα υπόλοιπα έγγραφα του αποτελέσματος αναζήτησης ως προς την εκπλήρωση των κριτηρίων αναζήτησης

Το πρώτο σημείο αναφέρεται ουσιαστικά κατά πόσο ένα έγγραφο που επιστράφηκε από το σύστημα νέων, όντως ενδιαφέρει το χρήστη, σύμφωνα με τις προτιμήσεις που δήλωσε στην ερώτηση. Το δεύτερο σημείο συγκρίνει τα έγγραφα του αποτελέσματος μεταξύ τους, δηλώνει δηλαδή πιο έγγραφο του αποτελέσματος αναζήτησης ενδιαφέρει περισσότερο το χρήστη.

Όπως αναφέρθηκε στο κεφάλαιο για την κατηγοριοποίηση εγγράφων, κάθε έγγραφο χαρακτηρίζεται από λέξεις κλειδιά και ανήκει σε μία κατηγορία ή υποκατηγορία. Η συσχέτιση με μία λέξη κλειδί συνοδεύεται με την επιλογή κάποιου βάρους της λέξης για το συγκεκριμένο έγγραφο. Παράλληλα με το βάρος κάθε λέξης κλειδί, αποθηκεύονται στη βάση δεδομένων και άλλα στοιχεία, που είναι ο αριθμός των εγγράφων που έχουν ενταχθεί σε κάθε κατηγορία/ υποκατηγορία και ο αριθμός των εγγράφων που έχουν ενταχθεί σε κάθε κατηγορία/ υποκατηγορία και συγκεκριμένη γλώσσα και περιέχουν τη συγκεκριμένη λέξη κλειδί. Τα παραπάνω στοιχεία αποθηκεύονται στη στατιστική βάση δεδομένων του συστήματος, όπως φαίνεται στο Σχήμα 5-2 για το σχεσιακό μοντέλο της βάσης δεδομένων. Η στατιστική βάση δεδομένων ενημερώνεται με κάθε αλλαγή στα έγγραφα του συστήματος και χρησιμοποιείται για τον υπολογισμό της σχετικότητας των εγγράφων, όπως θα περιγραφεί παρακάτω.

7.3.1 Υπολογισμός Σχετικότητας Εγγράφου

Προκειμένου να υπολογιστεί η σχετικότητα κάποιου εγγράφου, υπολογίζεται το σχετικό βάρος των λέξεων κλειδιών που χαρακτηρίζουν το έγγραφο περιέχονται στην

ερώτηση. Η σχέση που χρησιμοποιείται για τον υπολογισμό του σχετικού βάρους είναι η εξής:

$\mathbf{w_{ij} = tf_{ij} * \{ 1 + log(N / df_i) \}}$ όπου

$\mathbf{w_{ij}}$ σχετικό βάρος της λέξης ***i*** στο έγγραφο ***j***

$\mathbf{tf_{ij}}$ το βάρος της λέξης ***i*** στο έγγραφο ***j***

N ο συνολικός αριθμός εγγράφων που ανήκουν στην ίδια κατηγορία/ υποκατηγορία

$\mathbf{df_i}$ ο συνολικός αριθμός εγγράφων στην κατηγορία/ υποκατηγορία που περιέχουν τη λέξη ***i***

με ***i*** από 1 ως ***k***

με ***j*** από 1 ως ***l***

και ***l*** ο αριθμός των εγγράφων που προέκυψαν από το δίτιμο (boolean) τμήμα της ερώτησης

Για κάθε έγγραφο προστίθενται τα σχετικά βάρη των λέξεων της ερώτησης, οπότε έχουμε για κάθε έγγραφο ένα μέτρο σχετικότητας με την υποβληθείσα ερώτηση. Αυτά τα μέτρα κανονικοποιούνται με τη ρίζα του αθροίσματος των τετραγώνων των σχετικών βαρών επί του αριθμού των λέξεων κλειδιών της ερώτησης, δηλαδή με $\sqrt{\sum W_{ij}^2 \cdot k}$ και το αποτέλεσμα είναι ο επί τοις εκατό βαθμός σχετικότητας του κάθε εγγράφου σε σχέση με την υποβληθείσα ερώτηση [Salton89].

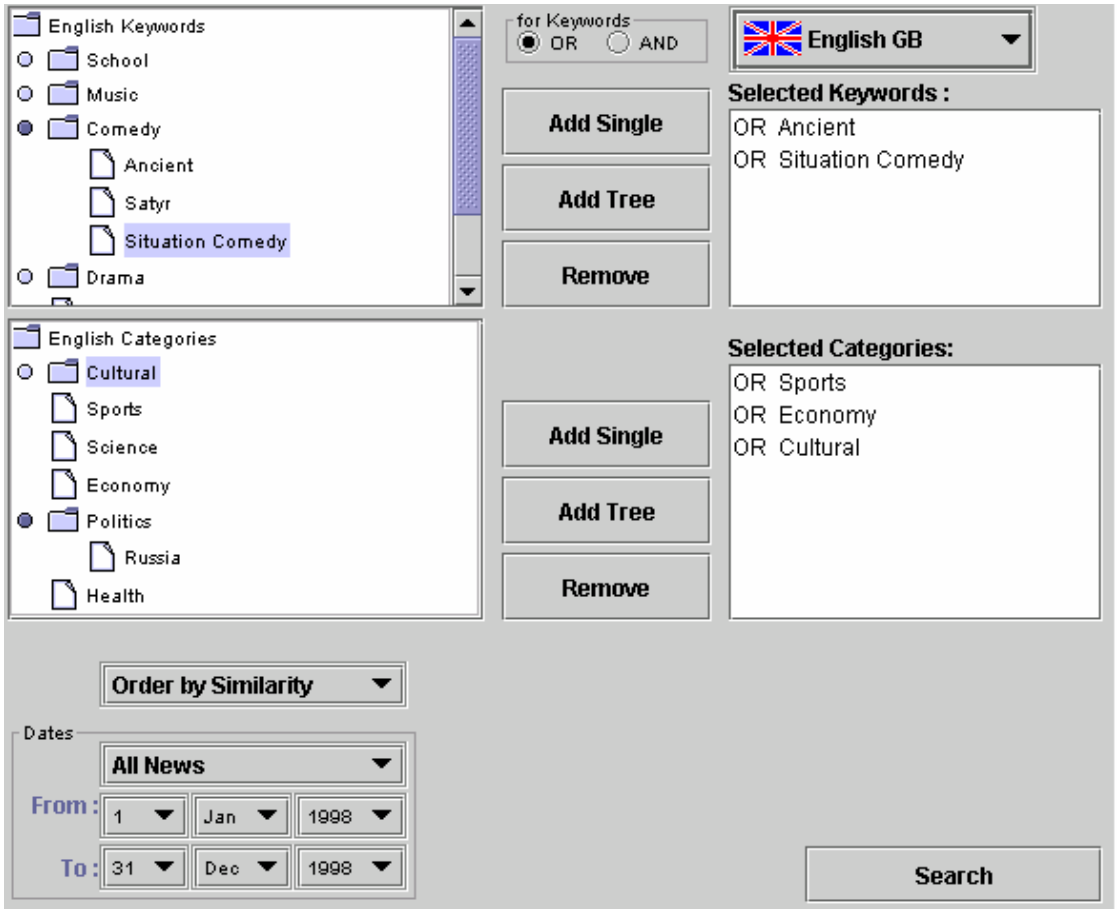
7.4 Συστατικό Επιλογής Προτιμήσεων και Αναζήτησης Εγγράφων

Στο κεφάλαιο για την κατηγοριοποίηση εγγράφων αναλύθηκαν οι βασικές λειτουργίες που απαιτούνται για τη διαχείριση των λεξικών των λέξεων κλειδιών και κατηγοριών και την επιλογή των χαρακτηριστικών των εγγράφων. Επίσης αναφέρθηκε η ενσωμάτωση του συστατικού (component) που αναπτύχθηκε στο εργαλείο συγγραφής εγγράφων.

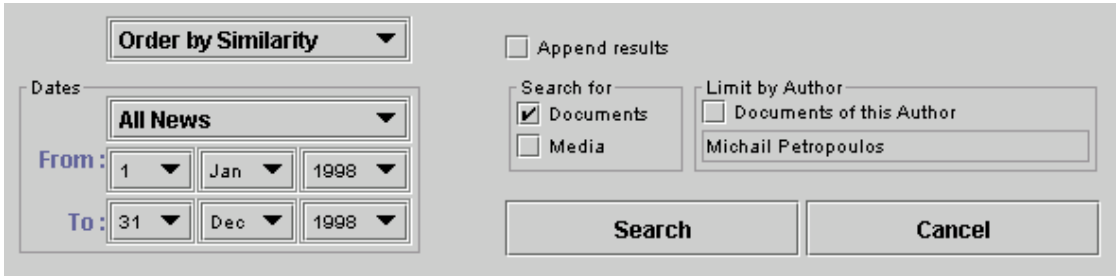
Σε αυτή την ενότητα θα περιγραφεί το συστατικό επιλογής προτιμήσεων και αναζήτησης εγγράφων που αναπτύχθηκε για την εκτέλεση ερωτήσεων αναζήτησης εγγράφων προς το «Hypermedia Custom News System».

Το συστατικό αυτό εμπεριέχει όλες τις λειτουργίες και συλλογές κλάσεων που υλοποιήθηκαν για την κατηγοριοποίηση εγγράφων, και αφορούν την ανάκτηση των λεξικών από τη βάση δεδομένων, την οργάνωσή τους σε κατάλληλες δομές και την εμφάνιση των ιεραρχικών δομών σε κατάλληλη μορφή στην εφαρμογή. Αυτές περιλαμβάνουν τις κλάσεις για την αποθήκευση των δεδομένων κάθε κόμβου των λεξικών των λέξεων κλειδιών και των κατηγοριών, τις μεθόδους για τη σύνδεση με τη βάση δεδομένων και την ανάκτηση της πληροφορίας για τα λεξικά και τα συστατικά για την εμφάνιση των ιεραρχικών δομών και των επιλεγμένων κόμβων των λεξικών. Για την υλοποίηση της εφαρμογής επιλογής προτιμήσεων και αναζήτησης εγγράφων, υλοποιήθηκαν κυρίως οι λειτουργίες και οι μηχανισμοί που είναι απαραίτητοι για την εκτέλεση των ερωτήσεων και την ανάκτηση των αποτελεσμάτων, καθώς και για τον υπολογισμό της σχετικότητας των αποτελεσμάτων. Όλες οι λειτουργίες ενσωματώθηκαν σε ένα συνδυαστικό συστατικό (compound component) που μπορεί να χρησιμοποιηθεί είτε αυτόνομα, είτε να ενσωματωθεί σε άλλη εφαρμογή. Χαρακτηριστικά αναφέρεται ότι το συστατικό που υλοποιήθηκε χρησιμοποιείται χωρίς αλλαγές στην εφαρμογή συγγραφής εγγράφων, στην εφαρμογή διαχείρισης διαγραμμάτων χρηστών και αυτόνομα ως εφαρμογή αναζήτησης εγγράφων.

Στην Εικόνα 7-1 και Εικόνα 7-2 φαίνεται το συστατικό επιλογής προτιμήσεων και αναζήτησης εγγράφων, όπως αυτό χρησιμοποιείται σαν αυτόνομη εφαρμογή από τους τελικούς χρήστες και από τους αρθρογράφους.



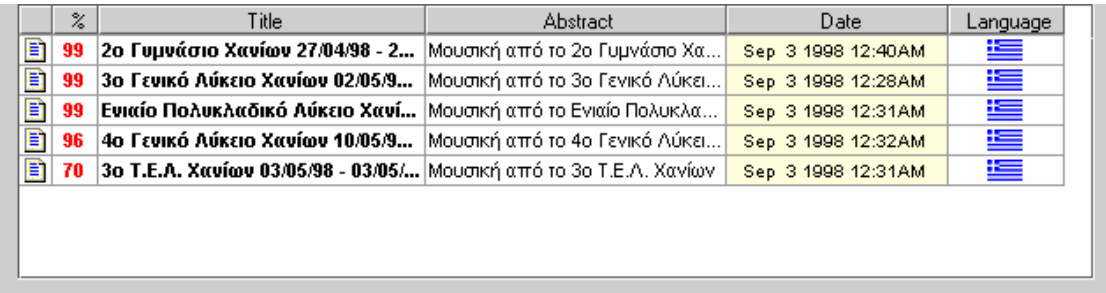
Εικόνα 7-1 Το συστατικό επιλογής προτιμήσεων και αναζήτησης εγγράφων ως αυτόνομη εφαρμογή



Εικόνα 7-2 Το κάτω τμήμα του συστατικού επιλογής προτιμήσεων και αναζήτησης εγγράφων, όπως χρησιμοποιείται από τους αρθρογράφους

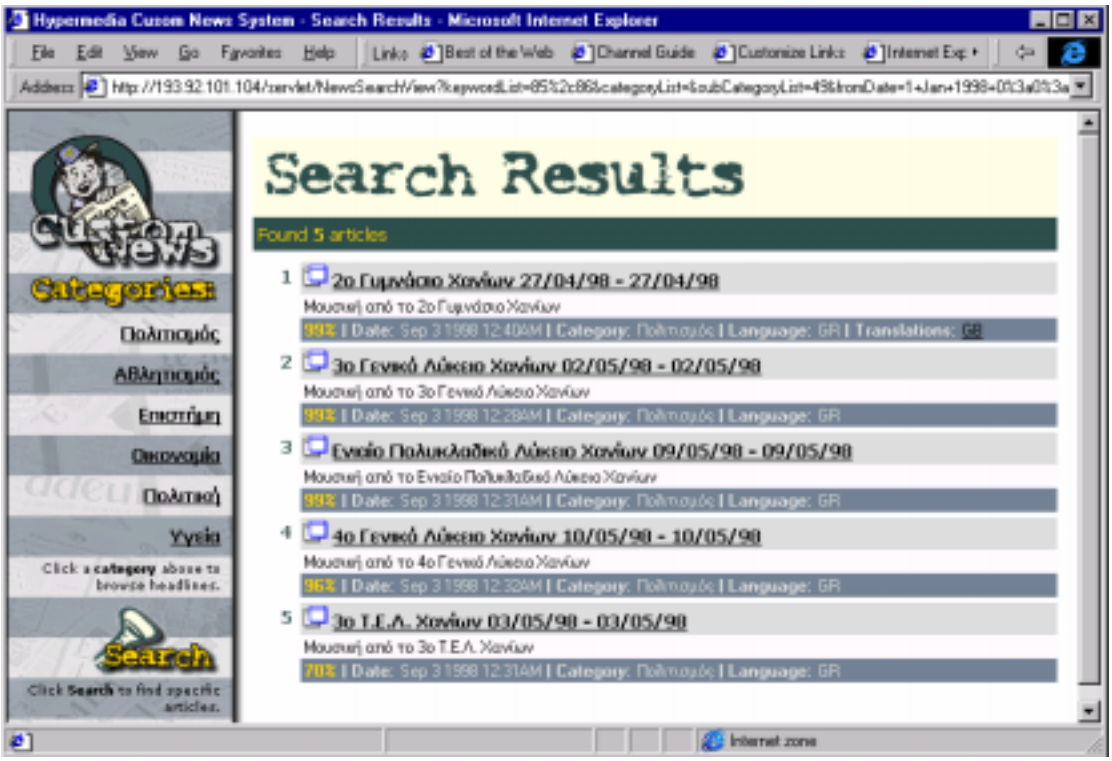
Στην Εικόνα 7-1 διακρίνονται τα δύο δέντρα των λεξικών των λέξεων κλειδιών και των κατηγοριών, και οι δύο λίστες για τις επιλογές που έκανε ο χρήστης. Ειδικά στην περίπτωση του αρθρογράφου (Εικόνα 7-2) φαίνονται οι επιπλέον επιλογές που έχει στη διάθεσή του για την αναζήτηση εγγράφων. Ο αρθρογράφος μπορεί να επιλέξει εάν θέλει απλά έγγραφα ή αν θέλει πολυμέσα, μπορεί να περιορίσει τα αποτελέσματα στα έγγραφα που έχει συνθέσει ο ίδιος και τέλος μπορεί να επιλέξει να προστεθούν τα

αποτελέσματα στα αποτελέσματα που ήδη υπάρχουν από προηγούμενη αναζήτηση. Το κουμπί *«Cancel»* είναι απαραίτητο επειδή ο αρθρογράφος χρησιμοποιεί το συστατικό από ξεχωριστό παράθυρο της εφαρμογής συγγραφής άρθρων.



	%	Title	Abstract	Date	Language
	99	**2ο Γυμνάσιο Χανίων 27/04/98 - 2...**	Μουσική από το 2ο Γυμνάσιο Χα...	Sep 3 1998 12:40AM	
	99	**3ο Γενικό Λύκειο Χανίων 02/05/9...**	Μουσική από το 3ο Γενικό Λύκει...	Sep 3 1998 12:28AM	
	99	**Ενιαίο Πολυκλαδικό Λύκειο Χανί...**	Μουσική από το Ενιαίο Πολυκλα...	Sep 3 1998 12:31AM	
	96	**4ο Γενικό Λύκειο Χανίων 10/05/9...**	Μουσική από το 4ο Γενικό Λύκει...	Sep 3 1998 12:32AM	
	70	**3ο Τ.Ε.Λ. Χανίων 03/05/98 - 03/05/...**	Μουσική από το 3ο Τ.Ε.Λ. Χανίων	Sep 3 1998 12:31AM	

Εικόνα 7-3 Αποτελέσματα αναζήτησης όπως παρουσιάζονται στον αρθρογράφο



Εικόνα 7-4 Αποτελέσματα αναζήτησης όπως παρουσιάζονται στον τελικό χρήστη

Στην Εικόνα 7-3 και Εικόνα 7-4 φαίνονται τα αποτελέσματα αναζήτησης, όπως αυτά παρουσιάζονται στον αρθρογράφο και στον τελικό χρήστη ([Πετρ98]). Όπως αναφέρθηκε, η αναζήτηση από την εφαρμογή συγγραφής των εγγράφων έχει διαφορετικές ανάγκες όσον αφορά τα αποτελέσματα, επειδή απαιτείται η χρησιμοποίηση και επεξεργασία τους μέσα στην εφαρμογή. Όπως φαίνεται στην εικόνα 6-3, δημιουργείται από τα αποτελέσματα ένας πίνακας μέσα στην εφαρμογή, ενώ τα αποτελέσματα στον τελικό χρήστη παρουσιάζονται σε μορφή HTML.

7.5 Ανακεφαλαίωση

Σε αυτό το κεφάλαιο αναλύθηκαν οι απαιτήσεις που υπάρχουν για την αναζήτηση εγγράφων στο «Hypermedia Custom News System» και περιγράφηκαν οι μηχανισμοί και οι λειτουργίες που υλοποιήθηκαν για την υποστήριξη των απαιτήσεων αυτών. Επίσης δόθηκε μία περιγραφή των βημάτων που απαιτούνται για την ολοκλήρωση μιας ερώτησης προς το σύστημα και αναλύθηκαν και η έννοια της σχετικότητας των αποτελεσμάτων προς την υποβληθείσα ερώτηση και ο τρόπος χρήσης του μοντέλου κατηγοριοποίησης των εγγράφων για τον υπολογισμό της σχετικότητας. Τέλος περιγράφηκε το εργαλείο που αναπτύχθηκε για να υποστηρίξει τον καθορισμό των προτιμήσεων του τελικού χρήστη, αλλά και των αρθρογράφων, και τονίστηκε η αναγκαιότητα που υπάρχει για ένα γενικό εργαλείο το οποίο θα μπορεί να χρησιμοποιηθεί σε πολλές εφαρμογές χωρίς τροποποιήσεις.

8 Υποσύστημα Διαγραμμάτων Χρηστών (User Profiles)

Στα προηγούμενα κεφάλαια παρουσιάστηκαν το μοντέλο κατηγοριοποίησης και οι εφαρμογές που χρησιμοποιούνται για την αναζήτηση εγγράφων. Με αυτά τα δεδομένα, προκειμένου οι τελικοί χρήστες να δουν την πληροφορία που τους ενδιαφέρει, πρέπει κάθε φορά να εκκινήσουν την εφαρμογή αναζήτησης εγγράφων, να επιλέξουν τις προτιμήσεις τους και να ενεργοποιήσουν τους μηχανισμούς αναζήτησης ώστε τελικά να παρουσιαστούν σε αυτούς τα έγγραφα που επιθυμούν.

Η παραπάνω διαδικασία είναι αναπόφευκτη στην περίπτωση που οι χρήστες αναζητούν κάθε φορά πολύ συγκεκριμένα έγγραφα, ή γενικώς διαφοροποιούν συχνά τις προτιμήσεις τους. Στη γενική περίπτωση όμως που αναφερόμαστε σε χρήστες που έχουν ένα συγκεκριμένο και σχετικά αμετάβλητο πεδίο ενδιαφέροντος, οπότε θα έπρεπε κάθε φορά να κατασκευάσουν το ίδιο σύνολο προτιμήσεων στην εφαρμογή αναζήτησης, αυτή η διαδικασία αποδεικνύεται ιδιαίτερα χρονοβόρα και μη φιλική προς το χρήστη. Επίσης αντιμετωπίζεται το πρόβλημα ότι μη γνωρίζοντας το σύστημα ποια έγγραφα έχει δει ο συγκεκριμένος χρήστης που κάνει την ερώτηση, θα πρέπει να του παρουσιάσει ως αποτέλεσμα όλα τα έγγραφα, ακόμα και αν τα περισσότερα τα έχει ήδη δει ο χρήστης. Με αυτό τον τρόπο χάνεται η έννοια του «Συστήματος Νέων», δηλαδή το να παρουσιάζονται στο χρήστη μόνο τα έγγραφα που ναι μεν εμπίπτουν στο είδος των εγγράφων που επιθυμεί να βλέπει, και που δεν τα έχει δει ακόμα, δηλαδή αποτελούν γι' αυτόν νέα.

Τα διαγράμματα χρηστών (User Profiles) καλούνται να αντιμετωπίσουν το παραπάνω πρόβλημα, αποθηκεύοντας τις προτιμήσεις που έχει ο κάθε χρήστης ώστε να μη χρειάζεται κάθε φορά να επιλεχθούν από την αρχή. Αυτό σημαίνει ότι ο κάθε χρήστης που επιθυμεί να έχει ένα αποθηκευμένο σύνολο προτιμήσεων στο σύστημα, περνάει από ένα στάδιο στο οποίο επιλέγει τις προτιμήσεις για τα έγγραφα που θέλει να του παρουσιάζονται. Αυτές αποθηκεύονται στο σύστημα και ανακτώνται κάθε φορά που εισέρχεται ο συγκεκριμένος χρήστης στο σύστημα, οπότε είναι γνωστές οι προτιμήσεις του χρήστη, και μπορούν να του παρουσιαστούν απ' ευθείας τα έγγραφα που τον ενδιαφέρουν.

Πέραν αυτού όμως, τα διαγράμματα χρηστών μπορούν να χρησιμοποιηθούν και για άλλες λειτουργίες. Κατ' αρχήν, γνωρίζοντας ποιος χρήστης έχει εισέλθει στο σύστημα, μπορεί να αποθηκεύεται στη βάση δεδομένων πληροφορία για το ποια άρθρα ο χρήστης έχει ήδη δει, οπότε δε θα είχε νόημα να του εμφανιστούν εκ' νέου. Επίσης μπορεί κάθε χρήστης, πέραν των προτιμήσεων για τα έγγραφα που τον ενδιαφέρουν, να επιλέγει και τις προτιμήσεις του όσον αφορά τον τρόπο με τον οποίο θέλει να του παρουσιάζονται τα έγγραφα, και γενικώς να εξατομικεύσει ολόκληρο το περιβάλλον του συστήματος που του παρουσιάζεται κάθε φορά που εισέρχεται σε αυτό.

Γνωρίζοντας τους χρήστες που εισέρχονται στο σύστημα και τις προτιμήσεις τους, μπορεί κανείς παρακολουθώντας τη συμπεριφορά των χρηστών μέσα στο σύστημα να βγάλει συμπεράσματα για την αποτελεσματικότητα των επιλογών του χρήστη, και να προτείνει στο χρήστη αλλαγή του διαγράμματός του ώστε αυτό να καλύπτει περισσότερο τα ενδιαφέροντά του. Αυτό επιτυγχάνεται αναλύοντας τον τρόπο με τον οποίο έχουν κατηγοριοποιηθεί τα έγγραφα που τελικά βλέπει ο χρήστης σε σχέση με τα έγγραφα που επιστρέφονται στο χρήστη από το διάγραμμα που είχε κατασκευάσει. Για παράδειγμα, βλέποντας το σύστημα ότι ο χρήστης παρόλο που έχει επιλέξει στο διάγραμμά του μία κατηγορία εγγράφων, τα οποία αποδεικνύεται ότι ποτέ δεν τα βλέπει, θα μπορούσε να προτείνει το σύστημα στον χρήστη να βγάλει από το διάγραμμά του τη συγκεκριμένη κατηγορία, επειδή ούτως ή άλλως δεν ενδιαφέρεται γ' αυτά τα έγγραφα. Ακόμα και στην περίπτωση που αυτή η ανάλυση των επιλογών δε γίνεται αυτόματα, μπορεί με κατάλληλες επιλογές ο χρήστης ο ίδιος να έχει τη δυνατότητα να κρίνει τα επιστρεφόμενα από το σύστημα άρθρα ως προς τη σχετικότητα της πληροφορίας των με τις προτιμήσεις του, οπότε και πάλι υπάρχει μία ένδειξη για την αποτελεσματικότητα του συγκεκριμένου διαγράμματος να επιστρέψει τα έγγραφα που πράγματι επιθυμεί ο χρήστης.

Στα πλαίσια της παρούσας διπλωματικής εργασίας υλοποιήθηκε ένας μηχανισμός διαγραμμάτων που αποθηκεύει τις προτιμήσεις των χρηστών, επιτρέποντας την αναζήτηση εγγράφων χωρίς την αναγκαιότητα του εκ' νέου καθορισμού των προτιμήσεων, κάθε φορά που οι χρήστες εισέρχονται στο σύστημα. Ως στόχος για την υλοποίηση των μηχανισμών για τη διαχείριση και χρήση των διαγραμμάτων είχε τεθεί να σχεδιαστούν με τέτοιο τρόπο ώστε να είναι ανοιχτά σε μελλοντικές επεκτάσεις,

στις οποίες θα μπορούν να υλοποιηθούν περισσότερες υπηρεσίες, όπως αυτές που αναφέρθηκαν παραπάνω.

8.1 Ανάλυση Απαιτήσεων

Οι απαιτήσεις που τέθηκαν στα πλαίσια της διπλωματικής εργασίας όσον αφορά τα διαγράμματα χρηστών, εστιάζονται κυρίως στον τρόπο καθορισμού των προτιμήσεων και την φιλικότητα του συστήματος προς τους χρήστες που έχουν ένα διάγραμμα και θέλουν να εισέλθουν στο σύστημα.

Για κάθε διάγραμμα απαιτείται η ύπαρξη κάποιου κωδικού χρήστη (login- User Id) που να συνοδεύεται από ένα αναγνωριστικό (password), ώστε να είναι σαφές στο σύστημα ποιος εισέρχεται σε αυτό, και να μπορεί να διασφαλιστεί η προστασία των χρηστών από ανεπιθύμητες ενέργειες τρίτων που θα ήθελαν να γνωρίζουν τα ενδιαφέροντα των χρηστών για ιδίο όφελος. Η αποθήκευση των προτιμήσεων των χρηστών, πρέπει να συνοδεύεται και από κάποια προσωπικά στοιχεία, ώστε να είναι δυνατή η επικοινωνία με αυτούς σε περίπτωση που παρουσιαστεί κάποιο πρόβλημα.

Όσον αφορά την είσοδο στο σύστημα, αυτή απαιτείται να γίνεται με την πληκτρολόγηση του κωδικού και του αναγνωριστικού του χρήστη χωρίς περαιτέρω διαδικασία. Η ανανέωση δε της σελίδας των νέων που παρουσιάζονται στο χρήστη, δεν πρέπει να απαιτεί την εκ’ νέου πληκτρολόγηση των παραπάνω στοιχείων.

8.2 Δημιουργία και Διαχείριση Διαγραμμάτων

Όλες οι πληροφορίες για τα διαγράμματα των χρηστών αποθηκεύονται σε ειδικές σχέσεις της Βάσης Δεδομένων του συστήματος νέων. Για τη δημιουργία ή τη μετατροπή διαγραμμάτων, αναπτύχθηκαν δύο διαφορετικές εφαρμογές. Η πρώτη ακολουθεί το σχεδιασμό όλων των εφαρμογών που παρουσιάστηκαν μέχρι τώρα και είναι γραμμένη σε Java. Η δεύτερη είναι μία έκδοση όπου ο χρήστης χρησιμοποιεί HTML σελίδες για τη δημιουργία και μετατροπή των διαγραμμάτων (HTML Front End), ενώ στην πλευρά του εξυπηρετητή εκτελούνται Java μέθοδοι με τη χρησιμοποίηση servlets (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης). Η ανάπτυξη της δεύτερης εφαρμογής κρίθηκε αναγκαία για την υποστήριξη διαγραμμάτων

χρηστών σε φυλλομετρητές που δεν υποστηρίζουν το πλήρες σύνολο της Java Τεχνολογίας.

8.2.1 Βάση Δεδομένων για τα Διαγράμματα Χρηστών

Όλα τα στοιχεία που είναι απαραίτητα για τα διαγράμματα χρηστών φυλάσσονται σε σχέσεις της βάσης δεδομένων. Τα διαγράμματα περιλαμβάνουν κατ' αρχήν την πληροφορία για τις προτιμήσεις των χρηστών, αλλά και προσωπικές πληροφορίες. Όπως φαίνεται και στο διάγραμμα του σχήματος 4.2, οι σχέσεις όπου αποθηκεύονται αυτές οι πληροφορίες είναι αυτές που φαίνονται στον πίνακα 8-1.

Profiles	
id	primary key
first_name	*όνομα*
last_name	*επώνυμο*
address	διεύθυνση
city	πόλη
zip	ταχ. κώδικας
country	χώρα
email	ηλεκτρονική διεύθυνση
tel	τηλέφωνο
fax	FAX
login	*κωδικός*
password	*αναγνωριστικό*
dates	*χρονικοί περιορισμοί*
ordering	*κατάταξη κατά ημερομηνία ή σχετικότητα*
kboolean	*σύνδεση λέξεων κλειδιών με λογικό OR ή λογικό AND*
language_id	*γλώσσα προτίμησης*

Profile_Categories	
category_id	primary key κατηγοριών
profile_id	primary key διαγράμματος

Profile_Keywords	
keyword_id	primary key λέξεων κλειδιών
profile_id	primary key διαγράμματος

profile_document_viewed	
profile_id	primary key διαγράμματος
document_id	primary key εγγράφου
date	ημερομηνία, ώρα

Πίνακας 8-1 Οι σχέσεις που χρησιμοποιούνται για την αποθήκευση των πληροφοριών για τα διαγράμματα χρηστών

Οι σχέσεις *Profiles*, *Profiles_Categories* και *Profiles_keywords* χρησιμοποιούνται για τα διαγράμματα χρηστών καθ' αυτά, ενώ η σχέση *profile_document_viewed* χρησιμοποιείται για τη φύλαξη των εγγράφων που έχουν ήδη δει οι χρήστες, και τα οποία δεν έχει νόημα να παρουσιαστούν στο συγκεκριμένο χρήστη την επόμενη φορά που θα επισκεφθεί τη σελίδα των αποτελεσμάτων.

Στη σχέση *Profiles* τα πεδία με πλάγια γράμματα είναι απαραίτητα σε κάθε διάγραμμα (όνομα χρήστη, κωδικός και αναγνωριστικό όσον αφορά τα προσωπικά στοιχεία) ενώ τα υπόλοιπα είναι προαιρετικά. Στη σχέση *profile_document_viewed* υπάρχει και ένα πεδίο *date* όπου αποθηκεύεται η ακριβής ημερομηνία και ώρα που ο χρήστης είχε πρόσβαση σε κάποιο έγγραφο.

8.2.2 Java Έκδοση της Εφαρμογής Δημιουργίας και Μετατροπής Διαγραμμάτων

Για τη δημιουργία νέων διαγραμμάτων και τη μετατροπή υπαρχόντων από τους τελικούς χρήστες, έχει αναπτυχθεί μία Java εφαρμογή στα πρότυπα των εφαρμογών για την κατηγοριοποίηση και την αναζήτηση εγγράφων. Η εφαρμογή χρησιμοποιεί για την επιλογή των προτιμήσεων του χρήστη, το συστατικό (component) που περιγράφηκε στο κεφάλαιο για την αναζήτηση εγγράφων, και το οποίο χρησιμοποιείται αυτούσιο (με παραλλαγές στο user interface) από τους αρθρογράφους και από τους τελικούς χρήστες κατά την αναζήτηση εγγράφων. Στην προκειμένη περίπτωση, η λειτουργία του συστατικού περιορίζεται στον απλό καθορισμό των προτιμήσεων, οι οποίες με τη σειρά τους αποθηκεύονται στις σχέσεις που περιγράφηκαν παραπάνω.

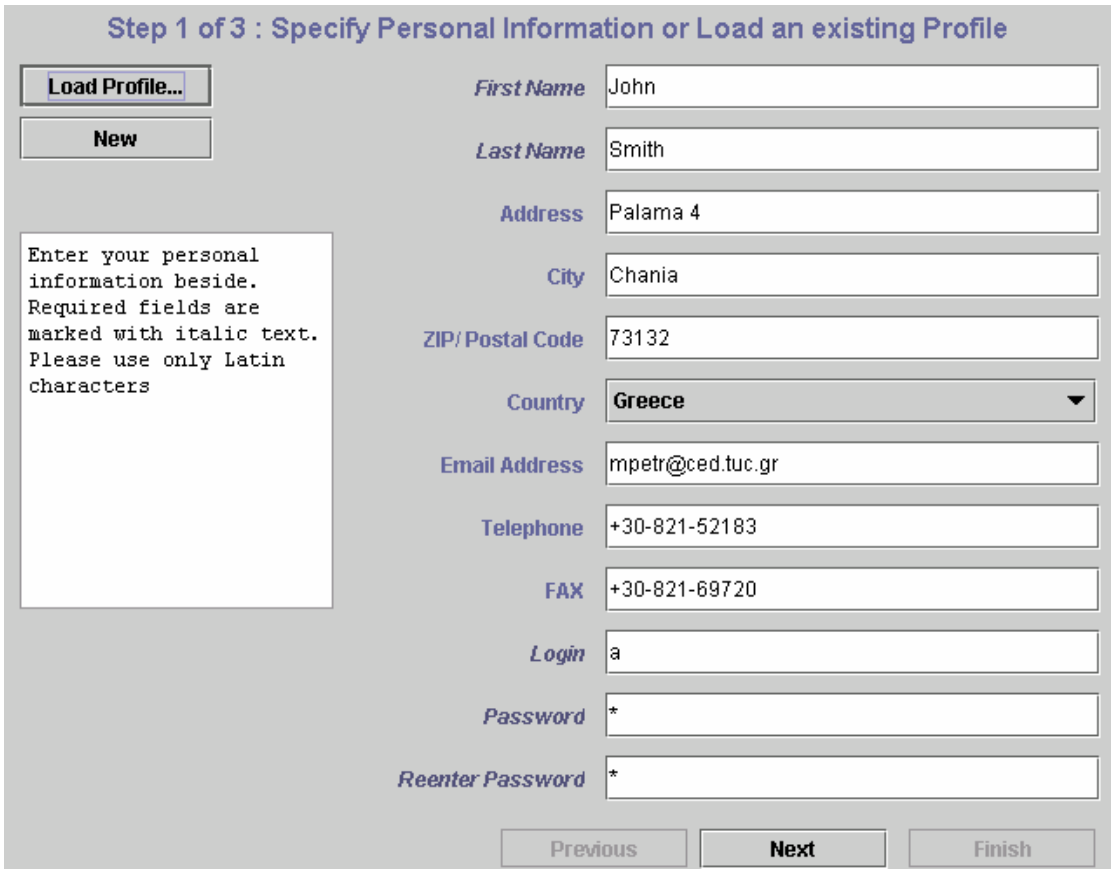
Η εφαρμογή έχει τρία διαφορετικά τμήματα.:

1. Το πρώτο αναφέρεται στα προσωπικά δεδομένα του χρήστη και δίνει τη δυνατότητα για ανάκτηση ήδη υπάρχοντος διαγράμματος,

2. το δεύτερο αναφέρεται στον καθορισμό των προτιμήσεων του χρήστη, και

3. το τρίτο ανακεφαλαιώνει τις επιλογές που έχουν γίνει.

Στο πρώτο στάδιο (βλέπε Εικόνα 8-1) υπάρχουν επιλογές για δημιουργία νέου ή μετατροπή ενός υπάρχοντος διαγράμματος. Στην περίπτωση που επιλεχτεί από το χρήστη η μετατροπή ενός διαγράμματος, εμφανίζεται ένας διάλογος όπου απαιτείται να δοθεί ο κωδικός και το αναγνωριστικό του χρήστη (Login και Password). Απαραίτητα στοιχεία για να προχωρήσει στο επόμενο στάδιο, είναι να υπάρχουν εγγραφές για το όνομα, τον κωδικό και το αναγνωριστικό του χρήστη. Σε αντίθετη

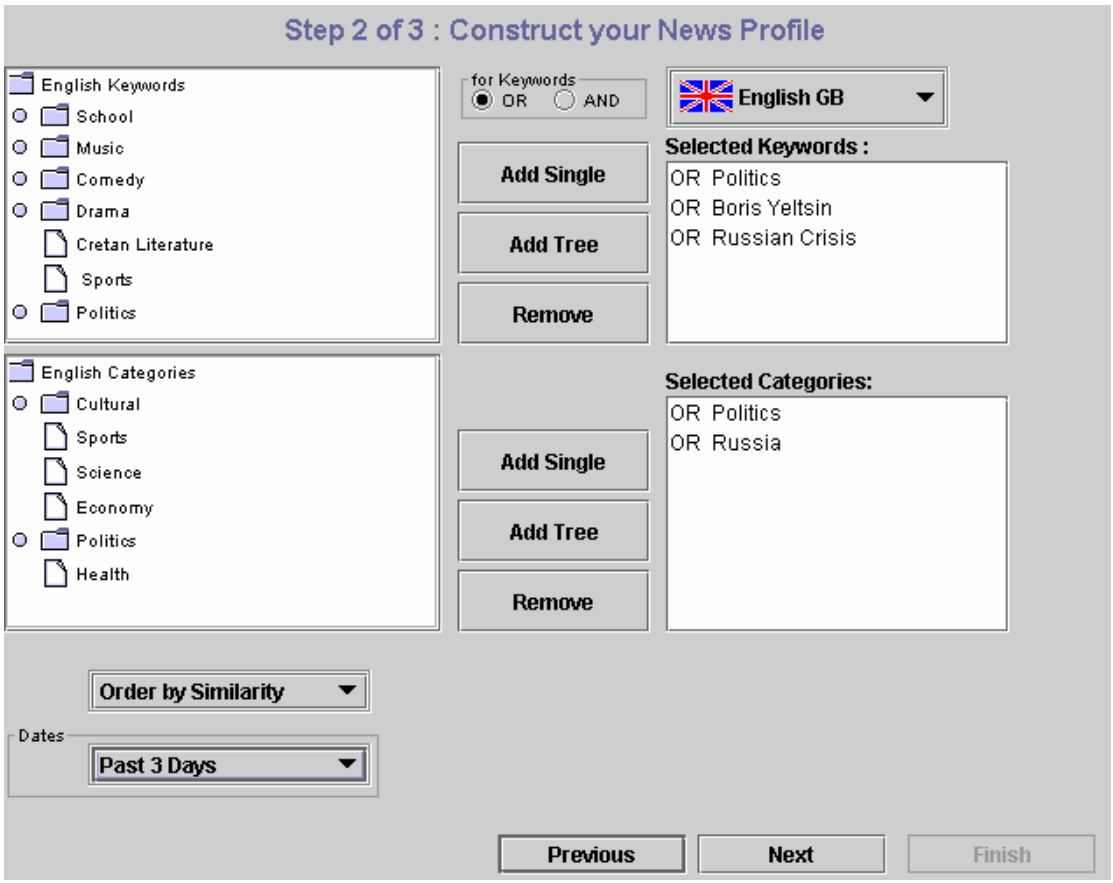


περίπτωση εμφανίζεται το αντίστοιχο μήνυμα λάθους.

Εικόνα 8-1 Το πρώτο στάδιο της δημιουργίας ή μετατροπής διαγραμμάτων χρηστών αναφέρεται στα προσωπικά δεδομένα του χρήστη

Όπως αναφέρθηκε, στο δεύτερο στάδιο (βλέπε Εικόνα 8-2) ο χρήστης επιλέγει τις προτιμήσεις για έγγραφα που τον ενδιαφέρουν, σύμφωνα με το μοντέλο κατηγοριοποίησης που ακολουθεί το σύστημα. Στην περίπτωση που γίνεται μετατροπή υπάρχοντος διαγράμματος, στα πεδία θα έχουν επιλεχτεί οι παλιές προτιμήσεις του χρήστη, όπως αυτές είχαν αποθηκευτεί στη βάση δεδομένων. Όπως

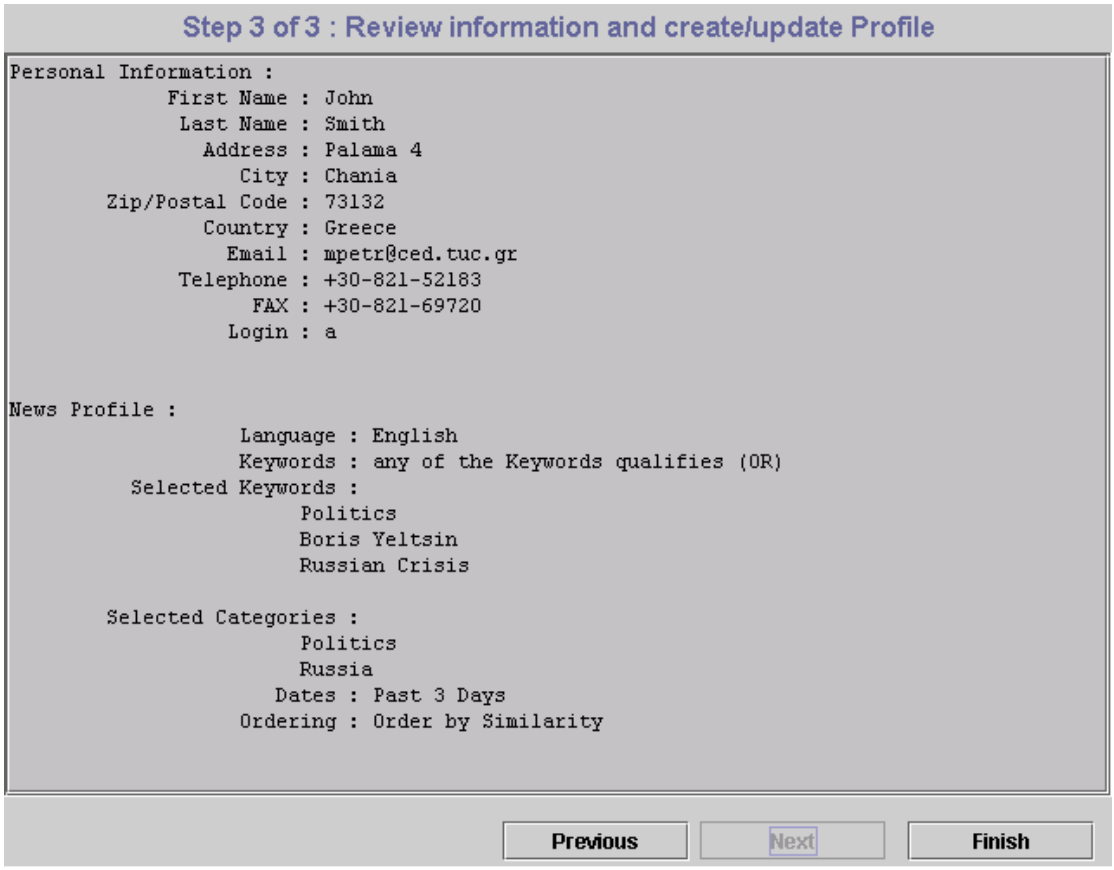
φαίνεται και στην εικόνα, το δεύτερο στάδιο είναι ουσιαστικά το ίδιο με το συστατικό



για την αναζήτηση εγγράφων από τους τελικούς χρήστες και τους αρθρογράφους.

Εικόνα 8-2 Στο δεύτερο στάδιο γίνεται η επιλογή των προτιμήσεων των χρηστών για τις πληροφορίες και τα έγγραφα που τους ενδιαφέρουν

Τέλος, στο τρίτο στάδιο (βλέπε Εικόνα 8-3) της εφαρμογής για τα διαγράμματα χρηστών, παρουσιάζονται στον χρήστη συνοπτικά οι επιλογές που έχει κάνει. Σε περίπτωση που συμφωνεί με τις επιλογές, μπορεί να τις αποθηκεύσει στη βάση δεδομένων, οπότε δημιουργείται είτε ένα νέο διάγραμμα, είτε ενημερώνονται οι εγγραφές του παλιού διαγράμματος.



Εικόνα 8-3 Στο τρίτο στάδιο της εφαρμογής, ο χρήστης έχει τη δυνατότητα να ελέγξει τα στοιχεία που έδωσε.

8.2.3 HTML Έκδοση της Εφαρμογής Δημιουργίας κ' Μετατροπής Διαγραμμάτων

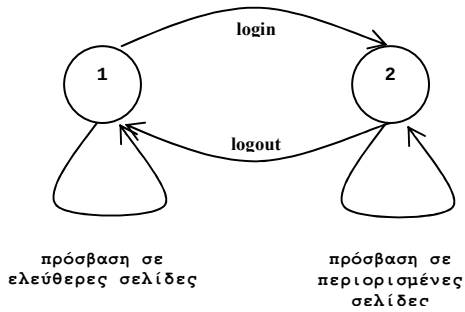
Για τους λιγότερο απαιτητικούς χρήστες, αναπτύχθηκε μία δεύτερη έκδοση της εφαρμογής για τα διαγράμματα χρηστών, η οποία χρησιμοποιεί απλή HTML. Όσον αφορά τα προσωπικά στοιχεία των χρηστών έχει τις ίδιες δυνατότητες με την Java έκδοση, αλλά όσον αφορά την επιλογή των προτιμήσεων, υποστηρίζει μόνο την επιλογή του πρώτου επιπέδου των κατηγοριών.

Όπως αναφέρθηκε, ο χρήστης επικοινωνεί με το σύστημα δια μέσω HTML σελίδων. Στην πλευρά του εξυπηρετητή η επεξεργασία των αιτήσεων γίνονται με Java servlets (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης). Επειδή η εφαρμογή επεκτείνεται σε περισσότερες από μία σελίδες και το HTTP πρωτόκολλό δεν είναι ενήμερο καταστάσεων (stateless) απαιτήθηκε η χρησιμοποίηση των HTTP-Συνόδων (Sessions) που υποστηρίζονται από τους εξυπηρετητές παγκόσμιου ιστού (WWW Servers) και τους φυλλομετρητές. Με αυτόν το μηχανισμό ο εξυπηρετητής είναι σε θέση να

προσδιορίσει εάν μία αίτηση που έρχεται από έναν φυλλομετρητή ανήκει στην ίδια σύνοδο. Αυτό διευκολύνει την επεξεργασία των αιτήσεων από τη μεριά του εξυπηρετητή, επειδή έτσι δε χρειάζεται με κάθε αίτηση να στέλνονται και τα στοιχεία αναγνώρισης του χρήστη, παρά μόνο μια φορά στην αρχή.

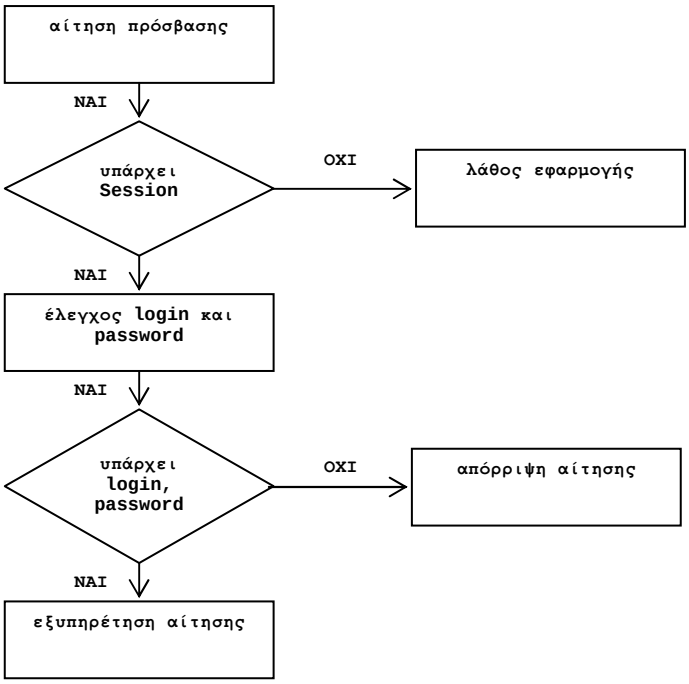
8.2.3.1 Σύνοδοι (Sessions) στην HTML έκδοση

Οι εξυπηρετητές παγκόσμιου ιστού διαχειρίζονται τις προσβάσεις στα αντικείμενά τους (HTML σελίδες, εικόνες, αρχεία κ.α.) σε συνόδους. Αυτό σημαίνει ότι μπορεί να αναγνωριστεί αν μία πρόσβαση έχει προηγηθεί από άλλες προσβάσεις με τον ίδιο φυλλομετρητή από το ίδιο μηχάνημα. Συνήθως οι προσβάσεις στα αντικείμενα των εξυπηρετητών δεν υφίστανται περιορισμούς και δεν απαιτείται κάποιο στάδιο αναγνώρισης του πελάτη. Στην περίπτωση που όμως σε κάποια αντικείμενα υπάρχει περιορισμός πρόσβασης, απαιτείται αρχικά η επιτυχής αναγνώριση του χρήστη από τον εξυπηρετητή, οπότε οι ακολουθούμενες προσβάσεις, είναι ελεύθερες για το περιορισμένο τμήμα.



Σχήμα 8-1 Οι δύο καταστάσεις των χρηστών του συστήματος νέων. Η κατάσταση 2 αναφέρεται ειδικά για την HTML έκδοση, όπου η πρόσβαση σε ορισμένες σελίδες προϋποθέτει την επιτυχή είσοδο (login) στο σύστημα.

Στη δική μας περίπτωση, η αναγνώριση του χρήστη επιτυγχάνεται όταν αυτός δώσει ένα σωστό συνδυασμό από κωδικό χρήστη και αναγνωριστικό. Αυτό που επιθυμούμε είναι ο χρήστης που έχει αναγνωριστεί, να έχει ελεύθερη πρόσβαση στα τμήματα της HTML έκδοσης της εφαρμογής διαγραμμάτων, που αφορούν το διάγραμμα του χρήστη. Έχοντας από τον εξυπηρετητή παγκόσμιου ιστού την πληροφορία για τις συνόδους των φυλλομετρητών, αρκεί να αποθηκεύονται οι επιτυχείς είσοδοι και να συνδυάζονται με τις συνόδους, για να γνωρίζουμε εάν πρέπει να εξυπηρετηθεί μία συγκεκριμένη αίτηση.



Σχήμα 8-2 Η διαδικασία που ακολουθείται στην περίπτωση αίτησης από πελάτες σε περιοχές περιορισμένης πρόσβασης

Στη πράξη το servlet που είναι υπεύθυνο για τις αιτήσεις από την HTML έκδοση της εφαρμογής διαγραμμάτων, έχει μεταβλητές όπου φυλάει τις πληροφορίες για τις επιτυχείς εισόδους στο σύστημα μαζί με τη σύνοδο στην οποία ανήκει η κάθε είσοδος. Στην περίπτωση που έρχεται μία αίτηση, το servlet παίρνει από τον εξυπηρετητή παγκόσμιου ιστού τη σύνοδο στην οποία ανήκει η αίτηση. Μόνο στην περίπτωση που έχει γίνει επιτυχή είσοδος στο σύστημα από αυτή τη σύνοδο, και η αίτηση αφορά το χρήστη της εισόδου, η αίτηση εξυπηρετείται, ενώ σε αντίθεση περίπτωση απορρίπτεται.

8.3 Χρησιμοποίηση Διαγραμμάτων

Προκειμένου να χρησιμοποιηθεί ένα διάγραμμα, ο χρήστης δίνει στο σύστημα νέων τον κωδικό και το αναγνωριστικό του, ενώ το σύστημα αναλαμβάνει την ανάκτηση των προτιμήσεων του χρήστη από τη βάση δεδομένων, και την κατάλληλη κατασκευή της ερώτησης αναζήτησης εγγράφων. Το αποτέλεσμα αυτής της διαδικασίας, δε διαφέρει από τον παραδοσιακό τρόπο αναζήτησης, με τη διαφορά ότι τα αποτελέσματα περιορίζονται ως προς τα άρθρα που δεν έχει ανακτήσει ακόμα ο συγκεκριμένος χρήστης με το διάγραμμά του.

8.3.1 Αναζήτηση Εγγράφων μέσω Διαγραμμάτων Χρηστών

Η εξυπηρέτηση των αιτήσεων μέσω διαγραμμάτων χρηστών γίνεται στην πλευρά του εξυπηρετητή από ένα κατάλληλα διαμορφωμένο Java servlet (βλέπε κεφάλαιο 3 Πλατφόρμα Υλοποίησης). Η παράδοση των κωδικού χρήστη και του αναγνωριστικού (login, password) που είναι απαραίτητα στο servlet, γίνεται μέσω των HTTP GET ή PUT μεθόδων. Οι HTTP GET και PUT μέθοδοι είναι τυποποιημένοι τρόποι για παράδοση παραμέτρων σε προγράμματα που εκτελούνται μέσα από τον εξυπηρετητή παγκόσμιου ιστού στην πλευρά του εξυπηρετητή. Επειδή τα servlet ακολουθούν παρόμοια φιλοσοφία, χρησιμοποιούνται οι διαδεδομένες HTML φόρμες για να πληκτρολογηθούν τα απαιτούμενα στοιχεία και να παραδοθούν στο servlet.

Το servlet ελέγχει εάν ο κωδικός χρήστη υπάρχει στη σχέση των διαγραμμάτων, και συγκρίνει το αναγνωριστικό που δόθηκε με αυτό που είναι αποθηκευμένο στη βάση δεδομένων. Στην περίπτωση που η σύγκριση είναι επιτυχής, ανακτώνται από τη βάση δεδομένων οι προτιμήσεις του χρήστη και παραδίνονται στο υποσύστημα υποβολής ερωτήσεων. Η διαδικασία που ακολουθείται από αυτό το σημείο και πέρα δε διαφέρει από τη διαδικασία αναζήτησης εγγράφων, εκτός του ότι τα αποτελέσματα περιορίζονται στα έγγραφα που δεν έχει ανακτήσει ο χρήστης δια μέσω του διαγράμματός του, σύμφωνα με τη σχέση *profile_document_viewed* της βάσης δεδομένων, που φαίνεται στο σχήμα 7-1.

Τέλος, η σχέση *profile_document_viewed* ενημερώνεται όταν από τα αποτελέσματα ερώτησης διαγράμματος χρηστών, φτάνει μία αίτηση στο σύστημα για την εμφάνιση κάποιου εγγράφου, οπότε και προστίθεται το αυτό στο διάγραμμα του χρήστη.

8.4 Ανακεφαλαίωση

Σε αυτό το κεφάλαιο αναπτύχθηκε η έννοια του διαγράμματος σε συστήματα νέων και οι υπηρεσίες που μπορούν να συνδεθούν με αυτό. Στη συνέχεια παρουσιάστηκαν οι απαιτήσεις που τέθηκαν για τα διαγράμματα χρηστών στην παρούσα διπλωματική εργασία και περιγράφηκαν οι εφαρμογές που αναπτύχθηκαν για τη δημιουργία και μετατροπή διαγραμμάτων. Τέλος, δόθηκε μία σύντομη περιγραφή της διαδικασίας που ακολουθείται για τη χρησιμοποίηση ενός διαγράμματος.

9 Υποσύστημα Υποστήριξης Πολυγλωσσικού Περιεχομένου

Με την έννοια *πολυγλωσσικό περιεχόμενο* εννοείται ότι η πληροφορία των εγγράφων μπορεί να διατεθεί σε περισσότερες από μία γλώσσα. Επίσης η αναζήτηση για κάποιο έγγραφο μπορεί να πραγματοποιηθεί χρησιμοποιώντας όρους που είναι διαθέσιμοι σε περισσότερες από μία γλώσσες.

Ως στόχος για το παρόν σύστημα είχε τεθεί η υποστήριξη πολυγλωσσικού περιεχομένου με τέτοιο τρόπο ώστε οι εφαρμογές που θα χρησιμοποιούν το σύστημα να είναι ικανές να χρησιμοποιήσουν τη Βάση Δεδομένων, χωρίς να γνωρίζουν τον ακριβή αριθμό των υποστηριζόμενων γλωσσών, ούτε να διαθέτουν λεπτομέρειες για αυτές.

Αυτό σημαίνει την πλήρη ανεξαρτησία των εφαρμογών που προσπελαύνουν τη Βάση Δεδομένων από τις υποστηριζόμενες γλώσσες, αυξάνοντας έτσι στο μέγιστο την ευελιξία του συστήματος ως προς την προσαρμογή της Βάσης Δεδομένων στη διαχείριση και προσθαφαίρεση γλωσσών.

Τα προβλήματα που συναντώνται στην προσπάθεια υποστήριξης πολυγλωσσικού περιεχομένου εντοπίζονται σε τέσσερα διαφορετικά επίπεδα

1. στο σχεδιασμό της Βάσης Δεδομένων
2. στον τρόπο με τον οποίο η Βάση Δεδομένων χειρίζεται δεδομένα κειμένου σε συνάρτηση με το χρησιμοποιούμενο λειτουργικό σύστημα
3. στις εφαρμογές που χειρίζονται τα κείμενα της Βάσης Δεδομένων
4. στην παρουσίαση των εγγράφων σε συνάρτηση με το χρησιμοποιούμενο λειτουργικό σύστημα και το φυλλομετρητή

Κάθε ένα από τα παραπάνω επίπεδα απαιτεί ξεχωριστή αντιμετώπιση ώστε να ελαχιστοποιηθούν οι εξαρτήσεις από τις ιδιαιτερότητες των εξυπηρετητών, και των πελατών και του χρησιμοποιούμενου κάθε φορά λειτουργικού συστήματος.

9.1 Σχεδιασμός Βάσης Δεδομένων

Όπως αναφέρθηκε στην Εισαγωγή, ένας από τους κύριους στόχους που τέθηκαν ήταν η ευελιξία του Συστήματος όσον αφορά την προσθαφαίρεση των υποστηριζόμενων γλωσσών και η ανεξαρτησία των εφαρμογών που χειρίζονται τα δεδομένα της Βάσης Δεδομένων από τον αριθμό των υποστηριζόμενων γλωσσών.

9.1.1 Γενικό μοντέλο για την υποστήριξη πολλών γλωσσών

Για να ανεξαρτητοποιηθεί η ονοματολογία των πεδίων των σχέσεων της Βάσης Δεδομένων, προστέθηκε μία σχέση η οποία περιγράφει τις υποστηριζόμενες γλώσσες. Όπως φαίνεται στον πίνακα 9-1, που περιγράφει τη σχέση *Languages*, υπάρχει για κάθε γλώσσα εκτός από το πλήρες όνομά της, και ένα ακρωνύμιο το οποίο χρησιμοποιείται στην ονομασία των πεδίων των υπολοίπων σχέσεων, όταν απαιτείται η ύπαρξη πεδίων για κάθε υποστηριζόμενη γλώσσα.

Έτσι οι εφαρμογές που προσπελαύνουν τη Βάση Δεδομένων, διαβάζοντας τη σχέση *Languages* γνωρίζουν τον αριθμό των υποστηριζόμενων γλωσσών καθώς και την ακριβή ονομασία των πολυγλωσσικών πεδίων.

Languages	**περιγραφή**
id	ακρωνύμιο γλώσσας
name	πλήρες όνομα
rank	σειρά προτεραιότητας εμφάνισης
ie_encoding	κωδικός κωδικοποίησης για IE
nc_encoding	κωδικός κωδικοποίησης

Πίνακας 9-1 Η σχέση που περιγράφει τις υποστηριζόμενες γλώσσες

Στο παρών σύστημα υποστηρίζονται έξι γλώσσες, και τα περιεχόμενα της σχέσης *Languages* φαίνονται στον πίνακα 9-2

id	name	rank	ie_encoding	nc_encoding
GR	Greek	3	cp1253	iso-8859-7
N	Norwegian	6	cp1252	iso-8859-1
F	French	4	cp1252	iso-8859-1
D	German	2	cp1252	iso-8859-1

GB	English	1	cp1252	iso-8859-1
IT	Italian	5	cp1252	iso-8859-1

Πίνακας 9-2 Τα περιεχόμενα της σχέσης Languages για τις έξι γλώσσες που υποστηρίζονται αυτή τη στιγμή από το υπάρχων σύστημα

Σύμφωνα με τα παραπάνω, η σχέση που αποθηκεύει τα δεδομένα για τις λέξεις κλειδιά, η οποία έχει για κάθε υποστηριζόμενη γλώσσα και ένα πεδίο, φαίνεται στον πίνακα 9-3

Keywords	**περιγραφή**
id	κωδικός λέξης κλειδί
parent_id	κωδικός πατέρα
GR_desription	ελληνική μετάφραση
F_desription	γαλλική μετάφραση
D_desription	γερμανική μετάφραση
IT_desription	ιταλική μετάφραση
GB_desription	αγγλική μετάφραση
N_desription	νορβηγική μετάφραση

Πίνακας 9-3 Η σχέση της Βάσης Δεδομένων που περιγράφει τις λέξεις κλειδιά. Φαίνεται ο τρόπος της ονοματολογίας των πεδίων για την υποστήριξη πολλών γλωσσών

Με αυτόν τον τρόπο, οι εφαρμογές γνωρίζοντας τον αριθμό των γλωσσών από τη σχέση *Languages* και τα ακρωνύμια για κάθε γλώσσα, προσθέτοντας στην αρχή του βασικού ονόματος του πεδίου το ακρωνύμιο της συγκεκριμένης γλώσσας έχουν τα πλήρη ονόματα των πολυγλωσσικών πεδίων.

Το πεδίο *rank* στη σχέση *Languages* χρησιμοποιείται για τη διατεταγμένη παρουσίαση των υποστηριζόμενων γλωσσών στις εφαρμογές, σύμφωνα με τις εκτιμήσεις που έγιναν όσον αφορά τους τελικούς χρήστες του συστήματος. Τα πεδία *ie_encoding* και *nc_encoding* θα περιγραφούν στο υποκεφάλαιο 9.2.

9.1.2 Πρόσθεση επιπλέον γλωσσών

Η πρόσθεση μίας επιπλέον γλώσσας απαιτεί την πρόσθεση της αντίστοιχης εγγραφής στη σχέση *Languages* με την οποία θα περιγράφεται η γλώσσα ως προς το όνομά της, το ακρωνύμιό της, τη σειρά εμφάνισης και τα υπόλοιπα στοιχεία. Παράλληλα

χρειάζεται να προστεθεί σε κάθε σχέση που έχει πεδία για κάθε υποστηριζόμενη γλώσσα, και ένα πεδίο για τη νέα γλώσσα. Το όνομα αυτού του πεδίου θα είναι το ακρωνύμιο της γλώσσας ακολουθούμενο από το βασικό όνομα του πεδίου, όπως περιγράφηκε στο προηγούμενο υποκεφάλαιο.

Η παραπάνω διαδικασία μπορεί να αυτοματοποιηθεί με τη χρησιμοποίηση κατάλληλου μηχανισμού πυροδότησης (trigger), έτσι ώστε με την πρόσθεση μίας εγγραφής στη σχέση *Languages* αυτόματα το trigger να προσθέτει τα πεδία στις διάφορες σχέσεις.

Με αυτόν τον τρόπο αυτοματοποιείται πλήρως η διαχείριση των υποστηριζόμενων γλωσσών.

9.1.3 Υποστήριξη μεταφράσεων εγγράφων

Έχοντας μοντελοποιήσει την οργάνωση των δεδομένων, όσον αφορά τις υποστηριζόμενες γλώσσες, με τον τρόπο που περιγράφηκε στα προηγούμενα υποκεφάλαια, υπάρχει η δυνατότητα για τη συγγραφή εγγράφων σε περισσότερες από μία γλώσσες. Τα έγγραφα αυτά δεν σχετίζονται μεταξύ τους, ακόμα και αν αποτελούν μετάφραση του ίδιου κειμένου. Οι τελικοί χρήστες, αναζητώντας στο Σύστημα Νέων έγγραφα σε περισσότερες από μία γλώσσες, είναι δυνατόν να πάρουν ως απάντηση περισσότερα έγγραφα τα οποία όμως αναφέρονται στο ίδιο νέο, με μόνη διαφορά τη χρησιμοποίηση διαφορετικής γλώσσας.

Από τα παραπάνω προκύπτει η ανάγκη ύπαρξης κάποιας συσχέτισης μεταξύ των εγγράφων που αποτελούν το ένα μετάφραση του άλλου. Για την αντιμετώπιση αυτού του προβλήματος, εισήχθηκε στη Βάση Δεδομένων μία επιπλέον σχέση η οποία συσχετίζει αυτά τα έγγραφα. Η δομή της σχέσης για τις έξι γλώσσες που υποστηρίζονται αυτή τη στιγμή, φαίνεται στον παρακάτω πίνακα.

Translations	**περιγραφή**
GR	κωδικός εγγράφου της ελληνικής μετάφρασης
N	κωδικός εγγράφου της νορβηγικής μετάφρασης
F	κωδικός εγγράφου της γαλλικής μετάφρασης
D	κωδικός εγγράφου της γερμανικής μετάφρασης

GB	κωδικός εγγράφου της αγγλικής μετάφρασης
IT	κωδικός εγγράφου της ιταλικής μετάφρασης

Πίνακας 9-4 Η σχέση Translations που συσχετίζει μεταφράσεις εγγράφων

Κάθε εγγραφή στη σχέση *Translations* αντιπροσωπεύει ένα έγγραφο του οποίου το περιεχόμενο είναι διαθέσιμο σε κάποιο αριθμό γλωσσών. Όπως φαίνεται από τον πίνακα, ακολουθήθηκε και εδώ η ονοματολογία για τα πεδία της σχέσης που περιγράφηκε και στα προηγούμενα υποκεφάλαια, ώστε να μην απαιτείται από τις εφαρμογές η ακριβή γνώση των ονομάτων των πεδίων. Έτσι κάθε πεδίο έχει ως όνομα το ακρωνύμιο της γλώσσας που αντιπροσωπεύει.

Έχοντας τώρα ως αποτέλεσμα κάποιας αναζήτησης ένα σύνολο κωδικών εγγράφων ***S,*** τα οποία αντιστοιχούν σε πολλές γλώσσες, οι κωδικοί ομαδοποιούνται ανά μεταφρασμένα έγγραφα με την ακόλουθη SQL ερώτηση

select distinct ***L1***, ***L2***, …, ***Ln***
from translations
where (***L1*** in ***S***) OR (***L2*** in ***S***) OR ..... OR (***Ln*** in ***S***)

όπου ***L1*** ως ***Ln*** είναι τα ακρωνύμια των υποστηριζόμενων γλωσσών.(βλέπε κεφάλαιο 7: Αναζήτηση Εγγράφων).

Από τις εγγραφές που επιστρέφονται, στον τελικό χρήστη παρουσιάζεται μόνο το έγγραφο στην γλώσσα που αυτός επιθυμεί, ενώ μπορεί να δοθεί η δυνατότητα να παρουσιαστούν στο χρήστη και οι μεταφράσεις του εγγράφου.

9.2 Αναπαράσταση και χειρισμός κειμένου

Η αποθήκευση, η αναπαράσταση και ο χειρισμός δεδομένων κειμένου σε ηλεκτρονικούς υπολογιστές, βασίζονται σε μία πληθώρα κανονισμών από οργανισμούς τυποποίησης. Μερικά από τα θέματα που καλύπτονται στους κανονισμούς αυτούς είναι

- ο ορισμός ενός μικρότερου δυνατού αριθμού στοιχείων κειμένου προς κωδικοποίηση

- η ανάθεση ενός μοναδικού αριθμητικού κωδικού σε κάθε στοιχείο

- η παροχή βασικών κανόνων για την κωδικοποίηση και ερμηνεία κειμένων έτσι ώστε να είναι δυνατή η ανάγνωση και η επεξεργασία τους από προγράμματα ηλεκτρονικών υπολογιστών

Οι κανονισμοί αυτοί συνήθως δεν ορίζουν τη μορφή που θα έχουν τα στοιχεία του κειμένου (glyphs) όταν αυτά προβάλλονται στην οθόνη ή εκτυπώνονται σε χαρτί. Οι κανονισμοί ορίζουν με ποιον τρόπο πρέπει να ερμηνεύονται οι χαρακτήρες και όχι πως αυτοί πρέπει να ζωγραφιστούν στην οθόνη. Αυτή την εργασία την αναλαμβάνουν τα υποσυστήματα του κάθε λειτουργικού συστήματος.

Στη συνέχεια θα γίνει μία σύντομη παρουσίαση των κανονισμών ASCII, ISO-8859-x και UNICODE.

9.2.1 Κανονισμός ASCII

Η τυποποίηση κατά ASCII (American Standard Code for Information Interchange) αναπτύχθηκε από το American National Standards Institute (ANSI) για να ορίσει τον τρόπο με τον οποίο οι υπολογιστές διαβάζουν και αποθηκεύουν χαρακτήρες κειμένου. Ο ASCII είναι ένας κώδικας που χρησιμοποιεί 7-bit για την κωδικοποίηση των χαρακτήρων, προσφέροντας έτσι ένα σύνολο από 128 χαρακτήρες. Σε αυτούς συμπεριλαμβάνονται γράμματα, αριθμοί, σημεία στίξης και κωδικοί ελέγχου (για παράδειγμα ο χαρακτήρας που συμβολίζει το τέλος μιας γραμμής).

Κάθε χαρακτήρας αναπαρίσταται από έναν αριθμό. Για παράδειγμα το κεφαλαίο *W* αντιστοιχεί στον αριθμό 87 ενώ το μικρό *w* στον αριθμό 119.

Οι 128 δυνατοί χαρακτήρες καλύπτονται από το λατινικό αλφάβητο, τους αριθμούς, τα βασικά σημεία στίξης και τους χαρακτήρες ελέγχου. Οι χαρακτήρες αυτοί μπορεί για τις αγγλόφωνες χώρες να μην αποτελούν περιορισμό, αλλά για ένα διεθνές περιβάλλον όπως είναι το Διαδίκτυο αποτελούν τροχοπέδη για εφαρμογές που απευθύνονται σε ένα ευρύτερο κοινό από πολλές χώρες.

9.2.2 Κανονισμός ISO-8859-x

Η απαίτηση για υποστήριξη χαρακτήρων εκτός των αυστηρά λατινικών χαρακτήρων, οδήγησε στην επέκταση του 7-bit κώδικα που χρησιμοποιεί η τυποποίηση κατά ASCII σε 8-bit. Αυτό διπλασιάζει τους διαθέσιμους χαρακτήρες σε 256 δίνοντας έτσι χώρο για την κάλυψη ενός ευρύτερου φάσματος χωρών.

Στις επιπλέον 128 θέσεις μπορούν να καλυφθούν όλοι οι ειδικοί χαρακτήρες με τα σημεία στίξης που έχουν χώρες όπως η Γαλλία, η Γερμανία και οι Σκανδιναβικές χώρες. Δεν υπάρχει όμως αρκετός χώρος για αλφάβητα που μοιάζουν ελάχιστα με το λατινικό αλφάβητο, όπως είναι το ελληνικό ή το ρωσικό αλφάβητο.

Για να καλυφθεί το παραπάνω κενό επινοήθηκε η έννοια της κωδικοσελίδας (code page). Όλες οι κωδικοσελίδες έχουν κοινούς χαρακτήρες από τους κωδικούς 0 έως 127 ενώ διαφέρουν στους υπόλοιπους 128 χαρακτήρες. Αυτό επιτρέπει με μία κωδικοσελίδα να αναπαρασταθεί το πλήρες[1] λατινικό αλφάβητο μαζί με κάποιους άλλους χαρακτήρες που βρίσκονται στις θέσεις 128 έως 255.

Τέτοια σύνολα κωδικοσελίδων έχουν αναπτύξει πολλοί οργανισμοί. Γνωστά σύνολα είναι αυτά της IBM (π.χ. Codepage 437 για ελληνικά), της Microsoft (Windows-cp1253 για ελληνικά) και ο Διεθνή Οργανισμός Τυποποίησης (iso-8859-7 για ελληνικά).

Ο τελευταίος έχει καθιερωθεί τα τελευταία χρόνια, έχοντας υποστήριξη από πολλούς κατασκευαστές υπολογιστικών συστημάτων και εφαρμογών. Μερικές από τις πιο γνωστές κωδικοσελίδες είναι η iso-8859-1 ή Latin-1 που περιλαμβάνει τους περισσότερους χαρακτήρες που βασίζονται στο λατινικό αλφάβητο, συμπεριλαμβανομένων χαρακτήρων του γαλλικού και του γερμανικού αλφαβήτου, και των αλφαβήτων σκανδιναβικών χώρων, η κωδικοσελίδα iso-8859-7 που περιλαμβάνει το ελληνικό αλφάβητο και η κωδικοσελίδα iso-8859-5 για το κυριλλικό αλφάβητο.

[1] Υπάρχουν και κωδικοσελίδες που δεν έχουν κανέναν κοινό χαρακτήρα με το λατινικό αλφάβητο

9.2.3 Κανονισμός UNICODE

Όπως φάνηκε στην προηγούμενη ενότητα, οι κωδικοσελίδες αντιμετώπισαν μεν το πρόβλημα της εσωτερικής αναπαράστασης χαρακτήρων πέραν των λατινικών, αλλά παρόλα αυτά δεν έλυσαν το πρόβλημα της αποθήκευσης κειμένου. Για παράδειγμα, εάν κάποια εφαρμογή διαβάσει από κάποιο αρχείο έναν χαρακτήρα (8-bit) που αντιστοιχεί στον κωδικό 0xD1, δεν γνωρίζει σε ποια κωδικοσελίδα αντιστοιχεί αυτός ο χαρακτήρας. Στην περίπτωση της κωδικοσελίδας iso-8859-1 (Latin-1) πρόκειται για το χαρακτήρα *Ñ* ενώ στην περίπτωση της κωδικοσελίδας iso-8859-7 (Greek) για το χαρακτήρα *Ρ*.

Επιπλέον αυτού, η εξάπλωση των ηλεκτρονικών υπολογιστών και σε ευρύτερα στρώματα του πληθυσμού και κυρίως η εξάπλωση του Διαδικτύου (Internet) δημιούργησαν την ανάγκη για μεγαλύτερη απλοποίηση στην αναπαράσταση και διαχείριση κειμένου. Έτσι προωθήθηκαν κανονισμοί όπως το UNICODE που θα περιγραφεί στη συνέχεια.

Το UNICODE αποτελεί έναν ολοκληρωμένο κανονισμό για την κωδικοποίηση χαρακτήρων για την αναπαράσταση κειμένου και την επεξεργασία του από ηλεκτρονικούς υπολογιστές [Unicode]. Προσφέρει ένα συνεπή (consistent) τρόπο για την κωδικοποίηση πολυγλωσσικού, απλού κειμένου και βάζει έτσι τάξη στο χάος που υπάρχει μέχρι σήμερα και δυσκολεύει την ανταλλαγή αρχείων κειμένου ανά τον κόσμο.

Ο σχεδιασμός του UNICODE βασίστηκε στην απλότητα και τη συνέπεια του ASCII κανονισμού, αλλά υπερβαίνει τον περιορισμό της κωδικοποίησης λατινικών μόνο χαρακτήρων. Χρησιμοποιεί αντί των 7-bit 16-bit για την κωδικοποίηση και παρέχει έτσι τη δυνατότητα για αναπαράσταση περισσότερων των 65000 χαρακτήρες.

Οι χαρακτήρες που κωδικοποιούνται στους αριθμούς 0x0000 έως 0x00FF συμπίπτουν με την κωδικοποίηση iso-8859-1, δηλαδή το πλήρες δυτικό σύνολο χαρακτήρων.

9.2.4 Αποθήκευση και ανάκτηση πολυγλωσσικού κειμένου σε Βάση Δεδομένων με JDBC

Τα περισσότερα λειτουργικά συστήματα χρησιμοποιούν κωδικοποίηση των χαρακτήρων κατά ASCII ενώ για την αναπαράσταση μη λατινικών χαρακτήρων, χρησιμοποιούν είτε κάποιο διεθνές πρότυπο όπως το iso-8859-x είτε κάποιο πρότυπο που έχει αναπτυχθεί ειδικά για μια συγκεκριμένη πλατφόρμα.

Έτσι για παράδειγμα η Microsoft στο λειτουργικό της σύστημα Windows95 χρησιμοποιεί το δικό της πρότυπο κωδικοσελίδων με ονομασία *windows cp125x*, ενώ τα Windows NT χρησιμοποιούν Unicode κωδικοποίηση για την αναπαράσταση κειμένων. Σε κάθε περίπτωση, πέραν του βασικού τρόπου που χρησιμοποιεί κάθε λειτουργικό σύστημα, είναι προετοιμασμένο και για άλλες κωδικοποιήσεις.

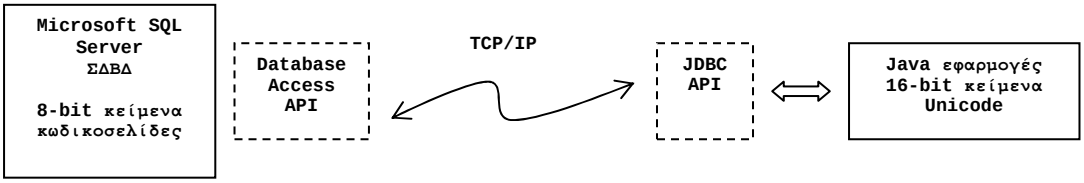
Πέραν της κωδικοποίησης χαρακτήρων που χρησιμοποιούν τα λειτουργικά συστήματα εσωτερικά για την αναπαράσταση κειμένων, προκύπτουν άλλα δύο ζητήματα που επηρεάζουν το σωστό χειρισμό κειμένου

1. **αποθήκευση κειμένου σε αρχεία:** αναφέρεται στον τρόπο με τον οποίο τα κείμενα αποθηκεύονται στη δευτερεύουσα μνήμη

2. **υποστήριξη κειμένου από τις εφαρμογές:** αναφέρεται στις διάφορες εφαρμογές που καλούνται να χειριστούν τα κείμενα

Παρόλο που το λειτουργικό σύστημα του Εξυπηρετητή Βάσεων Δεδομένων είναι τα Windows NT, τα οποία είναι ικανά να χειριστούν Unicode χαρακτήρες, ο Microsoft SQL Server, δηλαδή το ΣΔΒΔ που επιλέχτηκε, αποθηκεύει δεδομένα κειμένου χρησιμοποιώντας 8-bit ανά χαρακτήρα. Επίσης, οι πελάτες της Βάσης Δεδομένων, που είναι κυρίως Java εφαρμογές, χρησιμοποιούν εσωτερικά Unicode χαρακτήρες για την αναπαράσταση κειμένου, και αυτό σε ένα περιβάλλον λειτουργικού συστήματος το οποίο δεν είναι προκαθορισμένο. Τέλος, το JDBC (Java Database Connectivity) που χρησιμοποιείται για την πρόσβαση στη βάση δεδομένων, δεν παρέχει λειτουργικότητα επιρροής στην πρόσβαση σε κείμενα και την μετατροπή τους σε 16-bit Unicode Java χαρακτήρες.

Δεδομένου του παραπάνου σκηνικού, η αξίωση για την υποστήριξη πολλών γλωσσών ταυτόχρονα στην ίδια εφαρμογή, συναντάει σχεδόν ανυπέρβλητα προβλήματα, όταν απαιτούνται, εκτός της εισαγωγής του κειμένου στην εφαρμογή, η αποθήκευση και η επανανάκτησή του.



Σχήμα 9-1 Το σχήμα δείχνει τις δύο καταστάσεις στις οποίες βρίσκεται το κείμενο. Είναι είτε σε μορφή 8-bit στην πλευρά της Βάσης Δεδομένων, είτε σε μορφή 16-bit στην πλευρά των Java εφαρμογών

Το πρόβλημα έγκειται στο ότι:

1. οι εφαρμογές Java δεν γνωρίζουν σε ποια γλώσσα είναι γραμμένα τα κείμενα τα οποία ανακτώνται ώστε να χρησιμοποιήσουν τον κατάλληλο πίνακα αντιστοίχησης για τη μετατροπή των χαρακτήρων, και

2. γνωρίζοντας τη γλώσσα του υπό ανάκτηση κειμένου, οι άμεσοι μέθοδοι που προσφέρει το JDBC για την ανάκτηση κειμένου δε είναι παραμετροποιημένοι ως προς τις μεθόδους αντιστοίχησης χαρακτήρων.

Όπως αναφέρθηκε στο υποκεφάλαιο 9-2, οι εφαρμογές γνωρίζουν τον αριθμό των υποστηριζόμενων γλωσσών, και επίσης γνωρίζουν σε ποια γλώσσα αντιστοιχεί κάποιο πολυγλωσσικό πεδίο μιας σχέσης.

Η παραδοχή που έγινε για την αντιμετώπιση του πρώτου σκέλους του προβλήματος είναι ότι έχοντας ένα πεδίο της βάσης δεδομένων το οποίο αντιστοιχεί σε κάποια συγκεκριμένη γλώσσα, τα κείμενα που αποθηκεύονται σε αυτό το πεδίο, είναι όλα και αποκλειστικά γραμμένα σε αυτήν τη γλώσσα. Αυτή η παραδοχή δεν επιτρέπει για παράδειγμα την εισαγωγή μιας γαλλικής λέξης (μία λέξη που να χρησιμοποιεί ειδικούς γαλλικούς χαρακτήρες) σε ένα πεδίο της βάσης δεδομένων που αντιστοιχεί στην ελληνική γλώσσα.

Με την παραπάνω παραδοχή, και έχοντας υπ' όψιν τη σχέση *Languages* της Βάσης Δεδομένων από τα προηγούμενα υποκεφάλαια, η οποία περιγράφει τις

υποστηριζόμενες γλώσσες του συστήματος, οι εφαρμογές είναι σε θέση να γνωρίζουν τη γλώσσα του υπό ανάκτηση κειμένου.

Η κωδικοσελίδα που πρέπει να χρησιμοποιηθεί από τις εφαρμογές για τη μετατροπή των 8-bit χαρακτήρων σε 16-bit χαρακτήρες, αποθηκεύεται στη σχέση *Languages* στα πεδία *nc_encoding* και *ie_encoding* όπως φαίνεται στους πίνακες 9-3 και 9-4. Το πεδίο *nc_encoding* αποθηκεύει την κωδικοσελίδα στην οποία αντιστοιχεί η συγκεκριμένη γλώσσα, ενώ το πεδίο *ie_encoding* χρησιμοποιείται για τον ίδιο σκοπό ειδικά όμως όταν το Java Applet τρέχει στο περιβάλλον του φυλλομετρητή της Microsoft (Internet Explorer) επειδή η υλοποίηση της Java της Microsoft δεν δέχεται τους iso-8859-x κανονισμούς όταν πρόκειται να μετατρέψει 8-bit σε 16-bit χαρακτήρες.

Για να αντιμετωπιστεί το δεύτερο σκέλος του προβλήματος, δηλαδή τη μη υποστήριξη επιλογής κωδικοποίησης κατά την ανάκτηση κειμένων από το JDBC, χρειάστηκε να παρακαμφθούν οι μέθοδοι που προσφέρει το JDBC για την ανάκτηση κειμένου, και να αντικατασταθούν από μεθόδους στις οποίες μπορεί να γίνει μετατροπή του κειμένου από κωδικοποίηση σε κωδικοποίηση έτσι ώστε να συνυπολογίζεται στη μετατροπή και η γλώσσα στην οποία βρίσκεται το κείμενο.

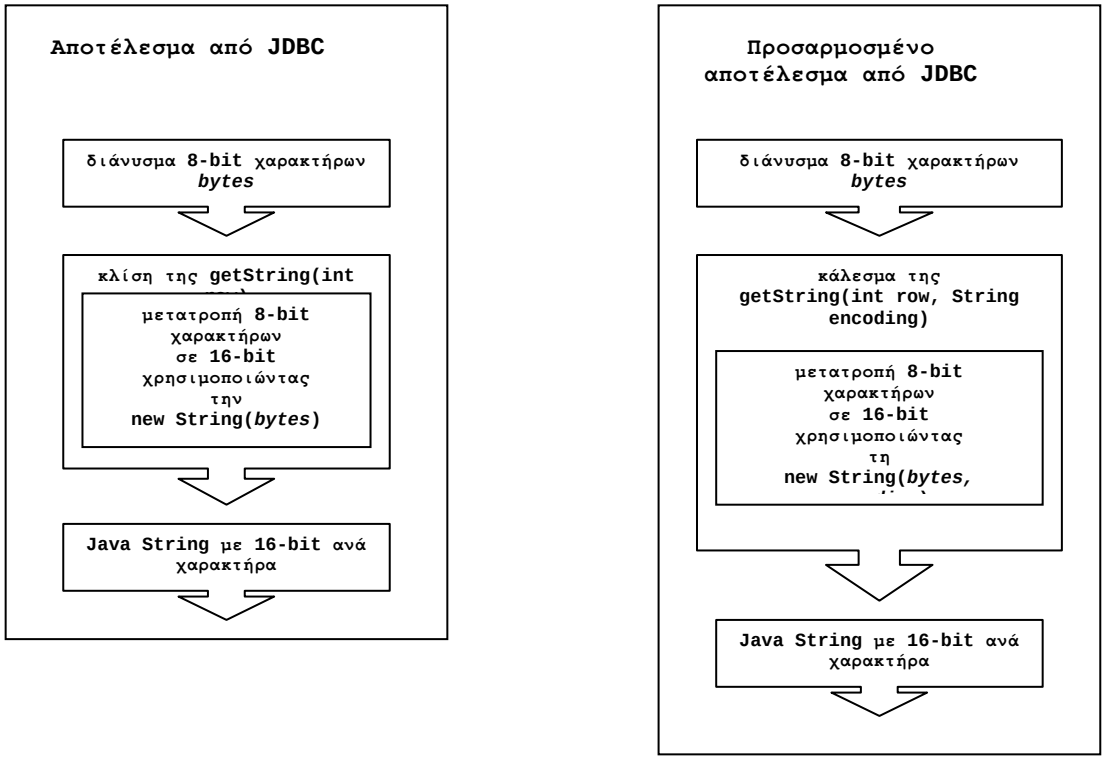
9.2.4.1 Ανάκτηση δεδομένων κειμένου με JDBC

Έχοντας το αποτέλεσμα μιας ερώτησης, το JDBC API προσφέρει τη μέθοδο *getString(int row)* για να επιστρέψει το κείμενο του πεδίου *row*. Για να υποστηριχθεί ο μηχανισμός για την μετατροπή των χαρακτήρων, δημιουργήθηκε μία ξεχωριστή Java κλάση, η οποία χρησιμοποιεί το αποτέλεσμα μιας ερώτησης αλλά επιτρέπει τη χρησιμοποίηση μεθόδων όπως

```
getString(int row, String encoding)
```

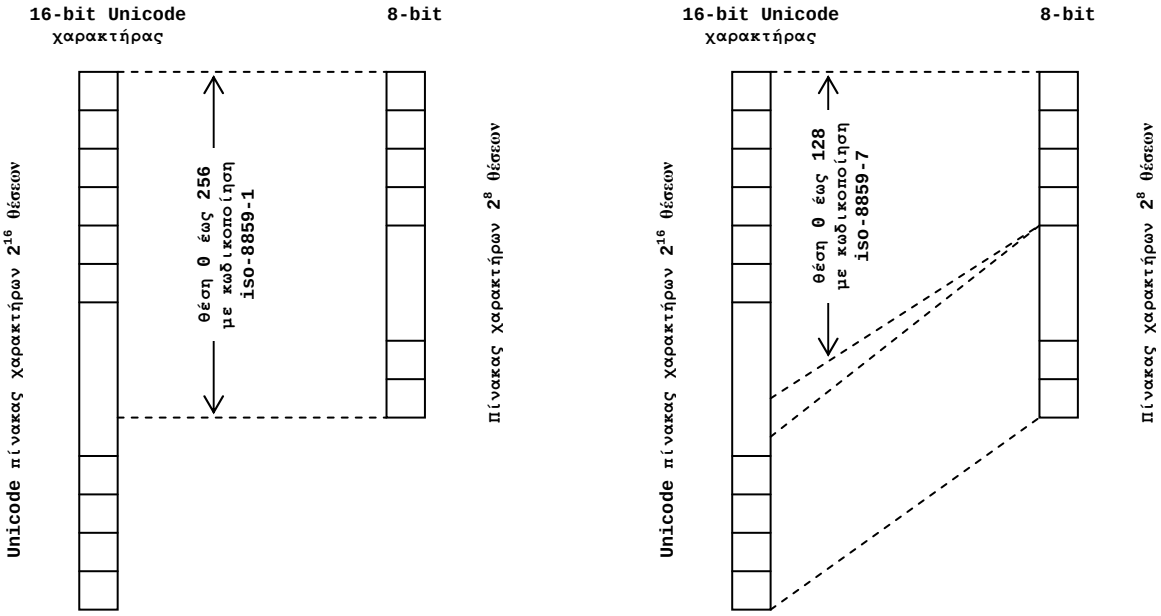
όπου *encoding* το όνομα της κωδικοσελίδας του κειμένου (για παράδειγμα iso-8859-7 για ελληνικά).

Στο παρακάτω διάγραμμα φαίνονται τα βήματα που γίνονται για τη σωστή μετατροπή των κειμένων.



Σχήμα 9-2 Η διαφορετική προσέγγιση που γίνεται για τη μετατροπή 8-bit χαρακτήρων σε 16-bit χαρακτήρες από το κανονικό JDBC API και από την προσαρμογή του ώστε να επιτρέπει την ύπαρξη παραμέτρου για την κωδικοποίηση που θα χρησιμοποιήσει

Η αντιστοίχηση που γίνεται καλώντας τη μέθοδο *new String(bytes, encoding)* φαίνεται σχηματικά στο επόμενο διάγραμμα.



Σχήμα 9-3 Στο σχήμα παρουσιάζεται η διαφορά αντιστοίχησης 8-bit χαρακτήρων σε 16-bit χαρακτήρες στην περίπτωση που πρόκειται για iso-8859-1 (Latin-1) χαρακτήρες, οπότε έχουμε ένα προς ένα αντιστοίχηση, και στην περίπτωση κάποιας άλλης κωδικοσελίδας, οπότε οι πρώτες 128 θέσεις αντιστοιχούνται στις πρώτες 128 θέσεις του Unicode χαρακτήρες, ενώ οι υπόλοιπες θέσεις αντιστοιχούνται σε άλλα σημεία

9.2.4.2 Ενημέρωση δεδομένων κειμένου με JDBC

Όσον αφορά την αντίστροφη διαδικασία, δηλαδή την ενημέρωση υπαρχόντων κειμένων ή την εισαγωγή νέων, χρησιμοποιούνται οι μηχανισμοί που προσφέρει το JDBC για την εκτέλεση SQL ερωτήσεων.

Για την εισαγωγή μιας νέας εγγραφής ή την ενημέρωση κάποιας άλλης, χρησιμοποιούνται οι γνωστές εντολές της SQL INSERT και UPDATE. Σε αυτές τις εκφράσεις SQL, υπάρχουν και οι τιμές των πεδίων που πρέπει να ενημερωθούν. Οι SQL εκφράσεις είναι σε μορφή Java 16-bit Strings, οπότε προκύπτει πάλι το ερώτημα της μετατροπής των τιμών των πεδίων (16-bit Strings) σε χαρακτήρες που θα αποθηκευτούν στη Βάση Δεδομένων (8-bit χαρακτήρες).

Παρόλο που η αντιστοίχηση χαρακτήρων Unicode σε iso-8859-x είναι μονοσήμαντη, και δε θα έπρεπε να απαιτείται ιδιαίτερη μέριμνα, στην πράξη η εκτέλεση μιας τέτοιας SQL έκφρασης (που περιέχει χαρακτήρες πέραν των λατινικών) δεν οδηγεί σε σωστά αποτελέσματα.

Το πρόβλημα έγκειται ανάμεσα στον οδηγό πρόσβασης που χρησιμοποιείται από το JDBC για την πρόσβαση στο Microsoft SQL Server, και στον ίδιο τον SQL Server. Αντί να γίνει απλή αντιστοίχηση των Unicode χαρακτήρων σε 8-bit, κόβονται απλώς τα περισσότερο σημαντικά 8-bit από τους χαρακτήρες και χρησιμοποιούνται οι υπόλοιποι ως έχουν.

Επειδή δεν υπάρχει πρόσβαση στον οδηγό πρόσβασης του JDBC, οι SQL εκφράσεις υπόκεινται σε ειδική μεταχείριση πριν την εκτέλεσή τους. Έτσι, έχοντας το String της SQL έκφρασης, πρώτα μετατρέπεται σε διάνυσμα από bytes, χρησιμοποιώντας την κωδικοσελίδα της γλώσσας που μας ενδιαφέρει, και επανασχηματίζεται από τα bytes το SQL String, χρησιμοποιώντας κωδικοσελίδα Latin-1 (δηλαδή καμία αλλαγή).

Μετά την παραπάνω διαδικασία, τα τμήματα της SQL έκφρασης έχουν μείνει αναλλοίωτα, επειδή περιέχουν μόνο λατινικούς χαρακτήρες, ενώ τα τμήματα που έχουν τις τιμές των πεδίων είναι μετασχηματισμένα με σωστό τρόπο ώστε όταν κοπούν τα πάνω 8-bit των χαρακτήρων για να εκτελεστεί η έκφραση, οι τιμές να βρίσκονται στη σωστή τους μορφή.

9.3 Παρουσίαση πολυγλωσσικού κειμένου

Έχοντας αποθηκεύσει τα κείμενα των εγγράφων στη Βάση Δεδομένων με τη σωστή μορφή, η παρουσίασή τους δε συναντάει κάποια ιδιαίτερη δυσκολία.

Την παρουσίαση των εγγράφων την αναλαμβάνει το υποσύστημα παρουσίασης εγγράφων, το οποίο διαβάζει από τη Βάση Δεδομένων όλες τις σχετικές με ένα έγγραφο πληροφορίες και κατασκευάζει τις κατάλληλες HTML δομές.

Κάθε φυλλομετρητής που θα κληθεί να παρουσιάσει την HTML σελίδα, έχει τους δικούς του μηχανισμούς για την επιλογή της κωδικοσελίδας στην οποία θα εμφανίσει τα κείμενα. Συνήθως υπάρχει μία επιλογή στο φυλομετρητή που καθορίζει σε ποια κωδικοσελίδα ο χρήστης επιθυμεί να δει τα κείμενα.

Πέραν αυτού όμως, μπορεί να καθοριστεί από το ίδιο το HTML έγγραφο η προς χρησιμοποίηση κωδικοσελίδα, εισάγοντας στην αρχή του εγγράφου τις κατάλληλες HTML δομές. Έτσι εισάγοντας τη δομή

```
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=ISO-8859-7">
```

στην αρχή του HTML εγγράφου, ο φυλλομετρητής γνωρίζει ότι πρόκειται να δείξει ένα HTML έγγραφο, χρησιμοποιώντας την κωδικοσελίδα iso-8859-7.

Με αυτόν τον τρόπο απαλλάσσεται ο χρήστης από την επιλογή της κωδικοσελίδας από τα μενού του φυλλομετρητή..

9.4 Ανακεφαλαίωση

Στο κεφάλαιο αυτό παρουσιάστηκε η έννοια του πολυγλωσσικού περιεχομένου που υποστηρίζει το «Hypermedia Custom News System». Έγινε μία σύντομη αναφορά στους τρόπους κωδικοποίησης δεδομένων κειμένου και στη δυσκολία που υπάρχει στην ταυτόχρονη χρησιμοποίηση και χειρισμό κειμένων πολλών γλωσσών.

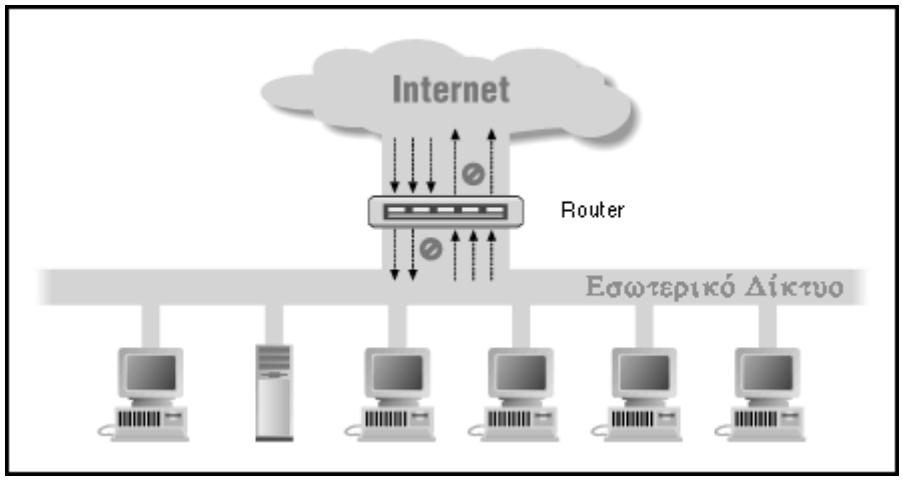
Προτάθηκε μία λύση που μοντελοποιεί τις υποστηριζόμενες γλώσσες στο σύστημα και περιγράφηκαν οι μέθοδοι που χρησιμοποιήθηκαν για τη συνεπή ανάκτηση και αποθήκευση πολυγλωσσικών κειμένων σε Βάση Δεδομένων με τη χρησιμοποίηση της Java και του JDBC.

Παρόλο που η λύση που προτάθηκε δεν παρουσιάζει προβλήματα, ο χειρισμός πολυγλωσσικού κειμένου παραμένει ακόμα ένα ανοιχτό θέμα. Η υποστήριξη του κανονισμού Unicode από Βάσεις Δεδομένων, είναι ικανή να λύσει τα περισσότερα προβλήματα που συναντώνται σήμερα, έχοντας όμως και την κατάλληλη υποστήριξη από μεριάς εφαρμογών και λειτουργικών συστημάτων.

10 Υποσύστημα Ανάκτησης Πληροφοριών σε Δίκτυα Περιορισμένης Πρόσβασης

Η ευρεία εξάπλωση του Διαδικτύου έφερε στο προσκήνιο σημαντικά προβλήματα ασφαλείας των υπολογιστικών συστημάτων που είναι συνδεδεμένα απευθείας στο δίκτυο. Πέραν των αδυναμιών των λειτουργικών συστημάτων που προέρχονται από την καθεαυτή σχεδίασή τους, καθημερινά εμφανίζονται μηχανισμοί για την εισβολή σε κάποιο υπολογιστικό σύστημα που γίνεται δυνατή από ελλείψεις στην υλοποίηση των υπηρεσιών που προσφέρονται δια μέσου δικτύων επικοινωνίας. Κακόβουλοι προγραμματιστές, εκμεταλλευόμενοι αδυναμίες στην υλοποίηση κάποιας υπηρεσίας, καταφέρνουν και εισβάλλουν σε υπολογιστικά συστήματα, αποκτώντας πρόσβαση στα δεδομένα του συστήματος, ενώ δεν είναι λίγες οι περιπτώσεις όπου οι εισβολείς εσκεμμένα προκαλούν και βλάβες.

Ένας δημοφιλής τρόπος προστασίας από τέτοιου είδους κινδύνους είναι να αποκοπούν οι υπολογιστές που δεν προσφέρουν υπηρεσίες στο Διαδίκτυο από τον κύριο κορμό του εξωτερικού δικτύου. Στο παρακάτω σχήμα φαίνεται μία τέτοια προσέγγιση.



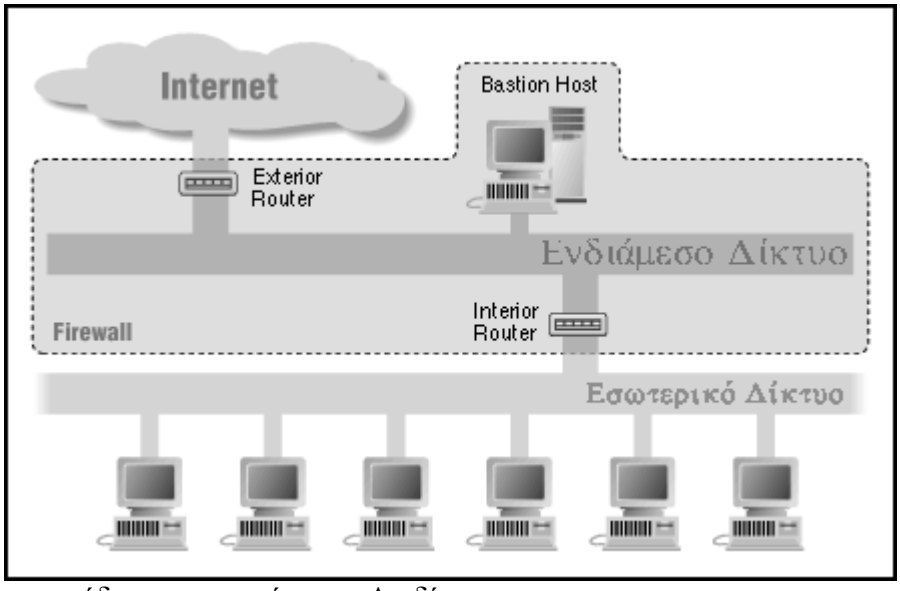
Σχήμα 10-1 Στο σχήμα φαίνεται μία απλή οργάνωση για την προστασία ενός εσωτερικού δικτύου. Το Δίκτυο απομονώνεται από το Διαδίκτυο από τον Router ο οποίος δεν επιτρέπει να μπουν προς τα μέσα όλα τα είδη πακέτων.

Συστήματα που υλοποιούν μία πολιτική περιορισμένης πρόσβασης σε ένα δίκτυο, γενικώς αποκαλούνται Firewall. Όπως φαίνεται στο παραπάνω σχήμα, ο

δρομολογητής (Router) που συνδέει το εσωτερικό δίκτυο με το Διαδίκτυο, δεν επιτρέπει όλα τα πακέτα να εισέλθουν στο εσωτερικό δίκτυο.

Αυτό που συνήθως γίνεται, είναι να αποκλείονται όλα τα πακέτα προς τους εσωτερικούς υπολογιστές, εκτός από πακέτα που απευθύνονται σε συγκεκριμένες υπηρεσίες συγκεκριμένων υπολογιστών. Η διάκριση των πακέτων γίνεται συνήθως από την IP διεύθυνση του μηχανήματος στο οποίο απευθύνονται, αλλά και από το Port του μηχανήματος δια μέσω του οποίου θέλουν να επικοινωνήσουν. :Αφήνοντας ελεύθερη πρόσβαση στο Port από το οποίο εξυπηρετεί ένα μηχάνημα αιτήσεις για μία συγκεκριμένη υπηρεσία, δε δημιουργείται πρόβλημα για τη συγκεκριμένα υπηρεσία.

Στο παρακάτω σχήμα φαίνεται μία άλλη οργάνωση ενός Firewall, όπου όλοι οι υπολογιστές του εσωτερικού δικτύου είναι προστατευμένοι δια μέσω του εσωτερικού Router, ενώ υπάρχει και ένας υπολογιστής στο ενδιάμεσο δίκτυο, ο οποίος μπορεί να



προσφέρει ανεμπόδιστα υπηρεσίες στο Διαδίκτυο.

Σχήμα 10-2 Μία δεύτερη αρχιτεκτονική για ένα Firewall, όπου υπάρχει ένας υπολογιστής εκτός του προστατευόμενου δικτύου για να μπορεί να προσφέρει υπηρεσίες ανεμπόδιστα προς τα έξω.

Αναφερόμαστε στην περίπτωση που η μόνη υπηρεσία που επιτρέπεται να περάσει μέσα από το Firewall είναι για τον εξυπηρετητή παγκόσμιου ιστού, από τον οποίο οι τελικοί αλλά και οι κανονικοί χρήστες έχουν πρόσβαση στις σελίδες του συστήματος Νέων. Το πρόβλημα που δημιουργείται με αυτή την αρχιτεκτονική είναι ότι η

πρόσβαση στον εξυπηρετητή Βάσης Δεδομένων που χρησιμοποιείται, απαιτεί την ελεύθερη πρόσβαση από ένα επιπλέον, ακαθόριστο Port, πράγμα που δεν επιτρέπεται από την πολιτική ασφαλείας των οργανισμών όπου είναι εγκατεστημένοι.

Το παραπάνω γεγονός αποκλείει την επικοινωνία με τον εξυπηρετητή Βάσης Δεδομένων απ' ευθείας από τους πελάτες μέσω JDBC. Ο μόνος τρόπος πρόσβασης προς το εσωτερικό του Firewall είναι μέσω του εξυπηρετητή παγκόσμιου ιστού, οπότε δημιουργήθηκε η ανάγκη για την υλοποίηση ενός μηχανισμού πρόσβασης στη βάση δεδομένων, έχοντας ως μεσολαβητή αυτό τον εξυπηρετητή.

10.1 Ανάλυση Απαιτήσεων

Οι ανάγκες των εφαρμογών για τις οποίες απαιτήθηκε η ανάπτυξη του μηχανισμού πρόσβασης στη Βάση Δεδομένων μέσα από Firewall, είναι σαφώς καθορισμένες. Ουσιαστικά απαιτείται η δυνατότητα εκτέλεσης SQL εκφράσεων και η ανάκτηση των αποτελεσμάτων που προκύπτουν από την εκτέλεσή τους.

Το σύνολο των SQL εκφράσεων που μπορεί να εκτελούνται μπορεί να ομαδοποιηθεί σε συγκεκριμένο αριθμό τύπων εκφράσεων. Διακρίνονται οι SQL εκφράσεις που εκτελούνται χωρίς να επιστρέφουν ως αποτέλεσμα εγγραφές της Βάσης Δεδομένων, αλλά μόνο έναν κωδικό για επιτυχή ή αποτυχημένη εκτέλεση, και σε εκφράσεις που επιστρέφουν εγγραφές της Βάσης Δεδομένων. Ο δεύτερος τύπος εκφράσεων διαφέρει ως προς τον αριθμό των στηλών που επιστρέφονται. Τέλος υπάρχουν και ειδικοί τύποι προσβάσεων στη βάση δεδομένων, όπου απαιτείται η εκτέλεση περισσότερων εκφράσεων ως μέρος μιας συναλλαγής (transaction).

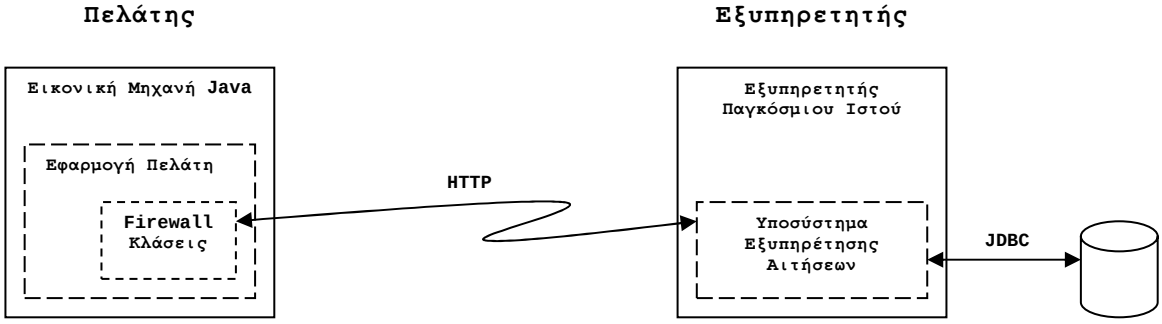
Σε κάθε περίπτωση, καλύπτεται μόνο μεταφορά εγγραφών σε μορφή κειμένου. Αυτό σημαίνει ότι για αριθμητικές τιμές, πριν τη μεταφορά μετατρέπεται η τιμή σε αλφαριθμητικούς χαρακτήρες.

Η εκτέλεση των SQL εκφράσεων, καθώς και η ανάκτηση των αποτελεσμάτων δεν πρέπει να διαφέρει σημαντικά από το συνηθισμένο τρόπο εκτέλεσης όταν χρησιμοποιείται απ' ευθείας JDBC σύνδεση. Αυτό σημαίνει ότι το API των κλάσεων δεν πρέπει να διαφέρει σημαντικά από το API του JDBC.

Τέλος, πρέπει η πρόσβαση στη Βάση Δεδομένων μέσω του εξυπηρετητή παγκόσμιου ιστού, να είναι ελεγχόμενη, δηλαδή πρέπει να υπάρχει ένα στάδιο αναγνώρισης του αιτούντος, πριν εξυπηρετηθεί καθεαυτή η αίτηση.

10.2 Αρχιτεκτονική Υποσυστήματος

Στο επόμενο σχήμα απεικονίζεται η αρχιτεκτονική του υποσυστήματος που είναι υπεύθυνο για την εκτέλεση των εκφράσεων SQL , χωρίς να περιορίζεται από τους μηχανισμούς ασφαλείας ενός Firewall. Στην πλευρά του πελάτη ο προγραμματιστής έχει στη διάθεσή του μία συλλογή κλάσεων υψηλού επιπέδου που δίνουν την αίσθηση μίας απλής σύνδεσης με τη βάση δεδομένων.



Σχήμα 10-3 Το σχήμα δείχνει την αρχιτεκτονική του Υποσυστήματος που είναι υπεύθυνο για την εκτέλεση ερωτήσεων σε Βάση Δεδομένων ξεπερνώντας τους περιορισμούς ενός Firewall. Φαίνεται ότι η εφαρμογή του πελάτη επικοινωνεί με το υποσύστημα εξυπηρέτησης αιτήσεων με το τυποποιημένο πρωτόκολλο των WWW εξυπηρετητών, το HTTP.

Στην πλευρά του εξυπηρετητή, η πρόσβαση στη βάση δεδομένων και την ανάκτηση πληροφοριών εκτελείται από το Υποσύστημα Εξυπηρέτησης Αιτήσεων, το οποίο είναι μεν αυτόνομη εφαρμογή, αλλά εκτελείται στα πλαίσια του εξυπηρετητή παγκόσμιου ιστού. Η καθεαυτή επικοινωνία με τη βάση δεδομένων γίνεται μέσω JDBC.

Ο εξυπηρετητής παγκόσμιου ιστού παίζει το ρόλο του ενδιάμεσου φορέα για την επικοινωνία των πελατών με το Υποσύστημα Εξυπηρέτησης Αιτήσεων. Όπως φαίνεται και στο σχήμα, αυτή η επικοινωνία γίνεται με το πρωτόκολλο HTTP που χρησιμοποιείται από τους εξυπηρετητές παγκόσμιου ιστού, έτσι ώστε να μην υπάρχει πρόβλημα από το Firewall, επειδή όπως αναφέρθηκε το Firewall επιτρέπει τη διέλευση πακέτων από και προς τον εξυπηρετητή παγκόσμιου ιστού.

10.3 Κλάσεις Πρόσβασης από Firewall

Σε επίπεδο προγραμματιστή, έχει σχεδιαστεί και αναπτυχθεί μία κλάση, η οποία αναλαμβάνει την μεταφορά των SQL εκφράσεων στον εξυπηρετητή παγκόσμιου ιστού, και την επεξεργασία των αποτελεσμάτων που προέρχονται από το Υποσύστημα Εξυπηρέτησης Αιτήσεων.

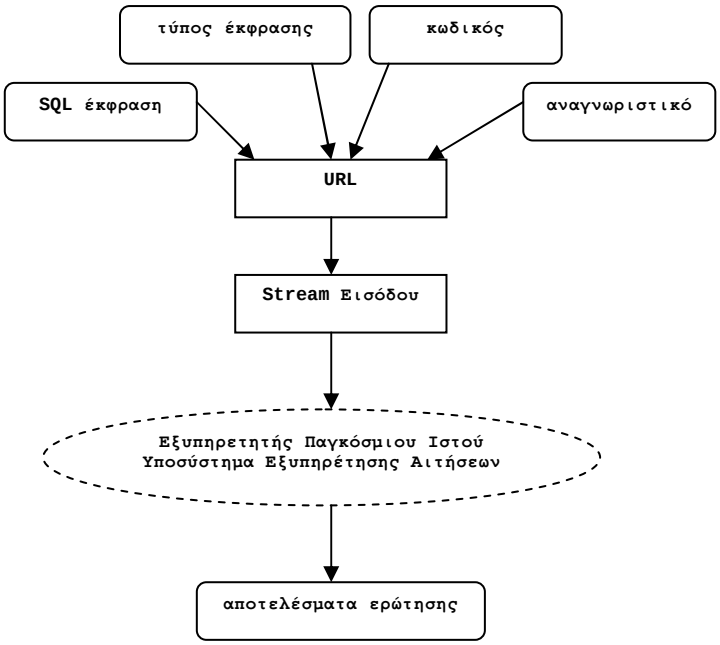
Ως παραμέτρους, η κλάση παίρνει την προς εκτέλεση SQL έκφραση, καθώς και τον τύπο της SQL έκφρασης, όπως αναφέρθηκε παραπάνω. Αυτός ο τύπος παραδίνεται στο Υποσύστημα Εξυπηρέτησης Αιτήσεων, το οποίο διακλαδίζεται στην αντίστοιχη με τον τύπο υπορουτίνα.

Όπως αναφέρθηκε, χρησιμοποιείται το HTTP πρωτόκολλο για την επικοινωνία με το Υποσύστημα Εξυπηρέτησης Αιτήσεων, και συγκεκριμένα χρησιμοποιούνται οι μηχανισμοί που προσφέρουν οι εξυπηρετητές παγκόσμιου ιστού για την εκτέλεση προγραμμάτων (CGI, servlets κλπ.). Η παράδοση παραμέτρων σε τέτοιου είδους προγράμματα, γίνεται δια μέσω των γνωστών HTTP GET και PUT μεθόδων.

Στην πλευρά του πελάτη, η κλάση που είναι υπεύθυνη για την επικοινωνία με το Υποσύστημα Εξυπηρέτησης Αιτήσεων, κατασκευάζει ένα URL που δείχνει στην εφαρμογή του υποσυστήματος, και έχει ως GET παραμέτρους την SQL έκφραση, τον τύπο της έκφρασης καθώς και τον κωδικό και το αναγνωριστικό πρόσβαση στη Βάση Δεδομένων (login, password).

Από το URL η κλάση αναλαμβάνει να κατασκευάσει μία Java Ροή Εισόδου (Java Input Stream) και να διαβάσει τα αποτελέσματα της ερώτησης, τα οποία προέρχονται από το Υποσύστημα Εξυπηρέτησης Αιτήσεων.

Στο Σχήμα 10-4 φαίνονται τα δεδομένα εισόδου της κλάσης, και πως αυτά καταλήγουν στο αποτέλεσμα από το Υποσύστημα Εξυπηρέτησης Αιτήσεων.



Σχήμα 10-4 Στο σχήμα φαίνεται η λειτουργία της κλάσης που είναι υπεύθυνη από την μεριά του πελάτη για την εκτέλεση εκφράσεων SQL μέσα από Firewall

Τα αποτελέσματα της ερώτησης είναι μία δομημένη ακολουθία χαρακτήρων. Η ακολουθία αποτελείται από τα αποτελέσματα της ερώτησης, και έναν αριθμό που δηλώνει την επιτυχή εκτέλεση ή το είδος του λάθους που συναντήθηκε. Όλα τα στοιχεία διαχωρίζονται από μία συγκεκριμένη ακολουθία χαρακτήρων που χρησιμοποιείται για το διαχωρισμό των στοιχείων.

Η ίδια η κλάση παρέχει ένα API για την ανάκτηση των δεδομένων στην πλευρά του πελάτη. Οι μέθοδοι που αποτελούν αυτό το API φαίνονται στον πίνακα 11-1 που ακολουθεί:

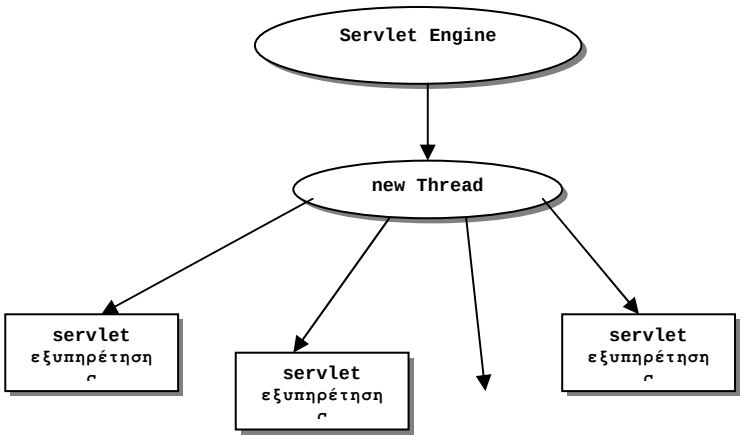
getError()	επιστρέφει τον κωδικό λάθους
wasSuccessfull()	επιστρέφει εάν έγινε επιτυχή εκτέλεση ή υπήρχε κάποιο λάθος
getNext()	επιστρέφει το επόμενο στοιχείο των δεδομένων του αποτελέσματος

Πίνακας 10-1 Το API για την ανάκτηση των δεδομένων από το αποτέλεσμα της ερώτησης δια μέσω Firewall

10.4 Υποσύστημα Εξυπηρέτησης Αιτήσεων

Το Υποσύστημα Εξυπηρέτησης Αιτήσεων για τις Ερωτήσεις μέσα από Firewall, έχει υλοποιηθεί σαν Java servlet (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης).

Διαχειρίζεται τη σύνδεση με τη Βάση Δεδομένων του συστήματος νέων και δέχεται κλήσεις με HTTP PUT μεθόδους από τα Applets διαχείρισης των λεξικών του συστήματος. Για τη διασφάλιση της πρόσβασης από εξουσιοδοτημένους χρήστες μόνο, πρέπει με κάθε αίτηση να δίνεται ένας κωδικός χρήστη και το αναγνωριστικό του κωδικού.



Σχήμα 10-5 Η εξυπηρέτηση των αιτήσεων από τα servlet γίνεται από ξεχωριστό Thread.

Για κάθε αίτηση που φτάνει στη μηχανή εκτέλεσης των servlet, δημιουργείται ένα ξεχωριστό Thread.

Αρχικά γίνεται από το Thread έλεγχος στον κωδικό χρήστη και το αναγνωριστικό εάν επιτρέπεται η πρόσβαση σε αυτό. Στη θετική περίπτωση, ελέγχονται οι παράμετροι που δόθηκαν και διακλαδίζεται ανάλογα με τον τύπο της SQL έκφρασης στην αντίστοιχη ρουτίνα εξυπηρέτησης.

Τα αποτελέσματα της εκτέλεσης της έκφρασης γράφονται στην έξοδο του servlet, σύμφωνα με τη δομημένη μορφή που περιγράφηκε παραπάνω.

10.5 Ανακεφαλαίωση

Σε αυτό το κεφάλαιο περιγράφηκαν τα προβλήματα που προκύπτουν στην περίπτωση που οι χρήστες του συστήματος βρίσκονται προστατευμένοι πίσω από ένα Firewall, όσον αφορά την πρόσβαση στη βάση δεδομένων.

Για την αντιμετώπιση του προβλήματος μεταφέρθηκε η καθ' εαυτή πρόσβαση στη βάση δεδομένων στον εξυπηρετητή, και δημιουργήθηκαν οι κατάλληλοι μηχανισμοί

για επικοινωνία με αυτόν δια μέσω του HTTP πρωτοκόλλου που επιτρέπεται από το Firewall να περάσει.

11 Πρόσβαση σε Βάση Δεδομένων με JDBC

Όλες οι εφαρμογές που παρουσιάστηκαν στα προηγούμενα κεφάλαια έχουν ως κοινό χαρακτηριστικό την εκτενή χρήση των υπηρεσιών των βάσεων δεδομένων, όπου αποθηκεύονται όλες οι πληροφορίες του συστήματος νέων. Η χρήση της Java ως γλώσσας προγραμματισμού επιβάλει τη χρήση του JDBC πρωτοκόλλου (βλέπε κεφάλαιο 3: Πλατφόρμα Υλοποίησης) για τη σύνδεση, την ανάκτηση και μετατροπή δεδομένων από τη βάση δεδομένων.

Παρόλο που οι λειτουργίες που προσφέρονται από τις κλάσεις που υλοποιούν το πρωτόκολλο είναι αρκετές για την αποδοτική χρησιμοποίησή του, η άμεση χρησιμοποίησή τους δυσχεραίνεται από μία σειρά παραγόντων που συναντήθηκαν κατά τη διάρκεια του σχεδιασμού του συστήματος νέων. Τέτοιοι παράγοντες είναι για παράδειγμα το πρόβλημα που υπάρχει όσον αφορά την ανάκτηση και μετατροπή πολυγλωσσικού κειμένου (βλέπε κεφάλαιο 9: Υποσύστημα Υποστήριξης Πολυγλωσσικού Περιεχομένου), η ανάγκη χρήσης περισσότερων της μίας σύνδεσης με τη βάση δεδομένων για αποδοτικότερη εξυπηρέτηση των πελατών, καθώς και άλλα προβλήματα.

Για την αντιμετώπιση των παραπάνω ζητημάτων σχεδιάστηκε και υλοποιήθηκε μία νέα κλάση που καλείται να αντικαταστήσει την κυρίως κλάση σύνδεσης με τη βάση δεδομένων. Παράλληλα με αυτήν σχεδιάστηκε και υλοποιήθηκε και μία νέα κλάση δια μέσω της οποίας γίνεται ο χειρισμός των αποτελεσμάτων της πρόσβασης, που αντικατέστησε την αντίστοιχη κλάση του JDBC.

Οι δύο παραπάνω κλάσεις χρησιμοποιούν εσωτερικά το JDBC πρωτόκολλο καθώς και τα με αυτό συσχετισμένα αντικείμενα, αλλά παρέχουν μία σειρά από επιπλέον δυνατότητες που θα περιγραφούν στα επόμενα υποκεφάλαια.

11.1 Μέθοδοι για κωδικοποιημένη ανάκτηση κειμένου

Όπως περιγράφηκε στο κεφάλαιο 9 « Υποσύστημα Υποστήριξης Πολυγλωσσικού Περιεχομένου», έχοντας το αποτέλεσμα μιας ερώτησης SQL, η ανάκτηση μιας στήλης σε μορφή κειμένου δημιουργεί προβλήματα κωδικοποίησης των χαρακτήρων. Συγκεκριμένα, χρησιμοποιείται ανεξάρτητα από την κωδικοποίηση στην οποία

βρίσκεται το κείμενο, πάντα η ίδια κωδικοσελίδα για την μετατροπή των 8-bit χαρακτήρων του αποτελέσματος σε 16-bit χαρακτήρες που χρησιμοποιεί η Java. Αυτό έχει σαν αποτέλεσμα την αλλοίωση των χαρακτήρων, στην περίπτωση που αυτές οι δύο κωδικοσελίδες δε συμπίπτουν. Στο σύστημα που αναπτύχθηκε και το οποίο φιλοδοξεί να υποστηρίζει πολλαπλές γλώσσες ανεξάρτητα με την κωδικοποίησή τους, η παραπάνω συμπεριφορά είναι απαράδεκτη.

Για την αντιμετώπιση του παραπάνω προβλήματος υλοποιήθηκε μία μέθοδο η οποία μετατρέπει τους χαρακτήρες ανάλογα με έναν κωδικό κωδικοποίησης που δέχεται ως παράμετρο. Η μέθοδος έχει τη μορφή

```
getString(resultrow, code-page)
```

και επιστρέφει το περιεχόμενο της στήλης *resultrow* αφού πρώτα έχει κατασκευάσει το αποτέλεσμα χρησιμοποιώντας κωδικοποίηση κατά *code-page*.

11.2 Μηχανισμοί προστασίας από ταυτόχρονη πρόσβαση

Κατά τη διάρκεια ανάπτυξης του συστήματος, χρησιμοποιήθηκαν διάφοροι οδηγοί, που προσφέρονται από διάφορες εταιρίες (Connect Software, Symantec, IDS Software, IBM), για την πρόσβαση σε Βάση Δεδομένων MS SQL. Ανάλογα με τον οδηγό αλλά και με την έκδοση του οδηγού, υπήρχε διαφορετική συμπεριφορά όσον αφορά τη χρησιμοποίηση μιας σύνδεσης για την εκτέλεση πολλών SQL εκφράσεων ταυτόχρονα.

Συγκεκριμένα, υπάρχουν μερικοί οδηγοί οι οποίοι δεν επιτρέπουν την ταυτόχρονη εκτέλεση περισσότερων SQL εκφράσεων την ίδια στιγμή, επιστρέφοντας λάθος. Το λάθος που επιστρέφεται δεν οφείλεται στο ΣΔΒΔ, το οποίο έχει την ικανότητα να επεξεργαστεί πολλές αιτήσεις ταυτόχρονα, αλλά στην υλοποίηση του πρωτοκόλλου επικοινωνίας με το ΣΔΒΔ από τους κατασκευαστές των οδηγών. Για την αντιμετώπιση του παραπάνω ζητήματος απαιτήθηκε κάποιος μηχανισμός ελέγχου ταυτοχρονισμού στην πρόσβαση στη σύνδεση με τη βάση δεδομένων, ώστε να αποκλείονται οι ταυτόχρονες προσβάσεις, όταν αυτές δεν επιτρέπονται από τον οδηγό.

Αυτός ο έλεγχος υλοποιήθηκε χρησιμοποιώντας στις κλάσεις πρόσβασης στη Βάση Δεδομένων τους μηχανισμούς για συγχρονισμένη πρόσβαση που προσφέρει η Java. Όλες οι προσβάσεις στα αντικείμενα της κλάσης πρόσβασης συγχρονίζονται έτσι ώστε κάθε στιγμή να υπάρχει μόνο μία πρόσβαση στη σύνδεση με τη βάση δεδομένων.

Αυτή η προστασία είναι ιδιαίτερα σημαντική στους μηχανισμούς αναζήτησης εγγράφων που εκτελούνται στην πλευρά του εξυπηρετητή, επειδή εκεί υπάρχει η δυνατότητα να δέχονται πολλές αιτήσεις εξυπηρέτησης, από όλους τους χρήστες που επιθυμούν να αναζητήσουν έγγραφα.

11.3 Διαχείριση Πολλαπλών Ταυτόχρονων Συνδέσεων

Ειδικά στην περίπτωση των μηχανισμών που εκτελούνται στην πλευρά του εξυπηρετητή (αναζήτηση εγγράφων, παρουσίαση εγγράφων κλπ.) όταν έχουμε συνωστισμό αιτήσεων από τους τελικούς χρήστες, η χρησιμοποίηση μίας και μοναδικής σύνδεσης δημιουργεί προβλήματα απόδοσης στο σύστημα.

Για το λόγο αυτό υλοποιήθηκε στην κλάση πρόσβασης στη βάση δεδομένων, η δυνατότητα να διαχειρίζεται περισσότερες της μίας σύνδεσης στη βάση δεδομένων. Ο αριθμός των συνδέσεων εξαρτάται από μία παράμετρο που διαβάζεται κατά την εκκίνηση από συγκεκριμένο αρχείο, δίνοντας έτσι τη δυνατότητα να προσαρμοστεί ανάλογα με τις απαιτήσεις του συστήματος και τον αριθμό αδειών που έχουν προμηθευτεί από τον κατασκευαστή της βάσης δεδομένων.

Με κάθε αίτηση που γίνεται σε έναν από τους μηχανισμούς που εκτελούνται στην πλευρά του εξυπηρετητή, ελέγχεται ποια από τις συνδέσεις με τη βάση δεδομένων εξυπηρετεί τις λιγότερες αιτήσεις, και ανατίθεται σε αυτή η εξυπηρέτηση της συγκεκριμένης αίτησης.

11.4 Διαχείριση λαθών

Ένα από τα σημαντικότερα θέματα σε εφαρμογές που χρησιμοποιούνται από πολλούς, ανεκπαίδευτους στην εφαρμογή, χρήστες, είναι ένας ενιαίος τρόπος αντιμετώπισης προβλημάτων που ανακύπτουν.

Το JDBC, όταν προκύπτει κάποιο λάθος κατά τη διάρκεια εκτέλεσης μιας SQL έκφρασης ή της ανάκτησης δεδομένων, «πετάει» μία Exception, που δίνει διάφορες πληροφορίες για το είδος τους λάθους. Ανάλογα με το λάθος που έχει γίνει, πρέπει να δοθεί στον χρήστη κάποιο μήνυμα, ή να αντιμετωπισθεί εσωτερικά στην εφαρμογή. Για την αντιμετώπιση τέτοιων λαθών, αναπτύχθηκε και υλοποιήθηκε μία ειδική κλάση λάθους που χρησιμοποιείται από τις κλάσεις σύνδεσης με τη βάση δεδομένων και την κλάση για ανάκτηση των αποτελεσμάτων, η οποία αναλόγως με την περίπτωση επιστρέφει κάποιο επεξηγηματικό μήνυμα ή προσπαθεί να αντιμετωπίσει το πρόβλημα.

Μία περίπτωση αντιμετώπισης προβλήματος, είναι η περίπτωση να χαθεί για κάποιο λόγο, όπως η παρατεταμένη αποκοπή από το Διαδύκτιο, η σύνδεση με την βάση δεδομένων. Σε αυτή την περίπτωση είναι επιθυμητή η επανασύνδεση με τη βάση δεδομένων, χωρίς την επανεκκίνηση της εφαρμογής και το ενδεχόμενο χάσιμο της εργασίας που έχει γίνει μέχρι εκείνη τη στιγμή από το χρήστη. Ακριβώς αυτή η πολιτική υλοποιήθηκε στο αντικείμενο πρόσβασης στη βάση δεδομένων. Γίνεται αυτόματη επανασύνδεση με τη βάση δεδομένων, χωρίς να επηρεαστεί η λειτουργία της εφαρμογής που χρησιμοποιεί τη σύνδεση.

11.5 Ανακεφαλαίωση

Στο κεφάλαιο περιγράφηκαν τα προβλήματα που αντιμετωπίστηκαν κατά τη διάρκεια σχεδιασμού των εφαρμογών για το σύστημα «Hypermedia Custom News System». Η ανάγκη για ελαχιστοποίηση του κώδικα όσον αφορά τις επαναλαμβανόμενες λειτουργίες, όπως είναι η αντιγραφή κάθε εκτελεσθείσας έκφρασης, καθώς και η ανάγκη για μία γενικευμένη αντιμετώπιση ζητημάτων όπως διαχείριση λαθών, οδήγησαν στο σχεδιασμό και την ανάπτυξη μίας γενικής χρήσεως κλάσης για την πρόσβαση σε ΣΔΒΔ.

12 Συνεισφορά της Διπλωματικής Εργασίας και Μελλοντικές Επεκτάσεις

Στην παρούσα διπλωματική εργασία υλοποιήθηκαν οι αναγκαίοι μηχανισμοί για την υποστήριξη κατηγοριοποίησης των εγγράφων, ερωτήσεων αναζήτησης εγγράφων και διαχείριση διαγραμμάτων χρηστών στο εξατομικευμένο σύστημα νέων κατ' απαίτηση «Hypermedia Custom News System». Οι παραπάνω μηχανισμοί συμπληρώνουν την εργασία [Πετρ98] και δημιουργούν με αυτόν τον τρόπο ένα ολοκληρωμένο περιβάλλον συγγραφής και διαχείρισης εγγράφων με δυνατότητες κατηγοριοποίησης και αναζήτησης εγγράφων καθώς και δημιουργίας διαγραμμάτων χρηστών.

Ιδιαίτερη έμφαση δόθηκε στην υποστήριξη πολλαπλών γλωσσών, με μοντελοποίηση σε επίπεδο βάσης δεδομένων της έννοιας του πολυγλωσσικού περιεχομένου και την ανεξαρτησία των υποσυστημάτων από τον αριθμό και το είδος των υποστηριζόμενων γλωσσών. Σε ένα διεθνές περιβάλλον, όπως είναι το Διαδίκτυο, αυτό το χαρακτηριστικό κρίνεται ιδιαίτερα σημαντικό και είναι η πρώτη φορά που υποστηρίζονται με έναν γενικευμένο τρόπο αυθαίρετος αριθμός γλωσσών.

Στη εργασία αντιμετωπίστηκε επίσης το πρόβλημα στην πρόσβαση σε βάση δεδομένων όταν αυτή βρίσκεται προστατευμένη πίσω από ένα Firewall, με την μετάθεση των προσβάσεων στη βάση δεδομένων στον εξυπηρετητή και την επικοινωνία με αυτόν δια μέσω εξυπηρετητή παγκόσμιου ιστού.

Για την ανάπτυξη όλων των τμημάτων του συστήματος χρησιμοποιήθηκε εξ' ολοκλήρου η γλώσσα προγραμματισμού Java που φιλοδοξεί να υλοποιήσει την προσδοκία για ανεξαρτησία από περιβάλλοντα και πλατφόρμα εκτέλεσης των εφαρμογών.

12.1 Μελλοντικές Επεκτάσεις

Η γενική και ανοικτή αρχιτεκτονική του συστήματος νέων «Hypermedia Custom News System» επιτρέπει την εύκολη προσθήκη νέων λειτουργιών και υπηρεσιών μερικές εκ των οποίων θα μπορούσαν να είναι οι ακόλουθες:

- Επιστροφή Ανάδρασης άρθρων από τους τελικούς χρήστες (Relevance Feedback) με σκοπό την αυτόματη εξαγωγή και υποβολή προς το χρήστη προτάσεων για αλλαγή του διαγράμματός του

- Υπηρεσία βαθμολόγησης των άρθρων από τους τελικούς χρήστες με σκοπό τη βελτίωση της ποιότητας των παρεχόμενων πληροφοριών

- Ανάπτυξη HTML έκδοσης για το εργαλείο αναζήτησης ώστε και χρήστες με λιγότερο εξελιγμένους φυλλομετρητές, ή χρήστες με ιδιαίτερα χαμηλές ταχύτητες σύνδεσης να μπορούν να χρησιμοποιήσουν τις υπηρεσίες του συστήματος.

- Περαιτέρω επέκταση της HTML έκδοσης της εφαρμογής για τη διαχείριση των διαγραμμάτων χρηστών, ώστε να συμπεριλαμβάνονται οι πλήρες δυνατότητες που προσφέρει η Java έκδοση.

- Δυνατότητα ανάκτησης των προτιμήσεων αναζήτησης από το αποθηκευμένο διάγραμμα κάποιου χρήστη μέσα από το εργαλείο αναζήτησης, ώστε να είναι ευκολότερη η αναζήτηση με τη μετατροπή υπάρχοντος διαγράμματος.

- Δυνατότητα επιλογής από το χρήστη εάν θέλει να του επανεμφανιστεί ένα έγγραφο που το έχει ήδη επισκεφθεί, και αντίστοιχη δυνατότητα να επιλέξει τη μη επανεμφάνιση εγγράφου ακόμα και στην περίπτωση που δεν το έχει επισκεφθεί.

- Ανάπτυξη εφαρμογής «News Ticker» στην αρχική σελίδα του συστήματος, το οποίο θα εμφανίζει τις κυριότερες ειδήσεις.

- Επέκταση του μηχανισμού αναγνώρισης χρήστη ώστε να είναι δυνατή η λεπτομερέστερη παροχή ελευθεριών ή περιορισμό των χρηστών.

- Αλλαγή ΣΔΒΔ ώστε να αποθηκεύονται τα κείμενα σε Unicode μορφή και να μη χρειάζεται η μετατροπή των κειμένων από 8-bit σε 16-bit και αντίστροφα.

- Καλύτερη χρησιμοποίηση των συνδέσεων με τη Βάση Δεδομένων ώστε να κλείνουν σε περίπτωση εκτεταμένης αχρηστίας

13 Ευρετήριο λέξεων

Διάγραμμα Profile

Διαδίκτυο Internet

Διασύνδεση Interface

Εξατομικευμένα Νέα Personalized News

Ιστοσελίδα Web Page

Λέξεις Κλειδιά Keywords

Μορφοποίηση Format

Νέα κατ' απαίτηση News On-demand

Παγκόσμιος Ιστός World Wide Web

Τελικός Χρήστης End User

Φυλλομετρητής Browser

Χαρακτηριστικά Attributes

Hypermedia News On Demand HyNoDe

HyperText Markup Language HTML

Universal Resource Locator URL

Αναγνώριση Χρηστών: User Authentication

Δικαιώματα: Privileges

14 Κατάλογος Σχημάτων

15 Κατάλογος Πινάκων

16 Κατάλογος Εικόνων

Βιβλιογραφία

[Πετρ98] Πετρόπουλος Μιχάλης, Συγγραφή και Παρουσίαση Εγγράφων σε Εξατομικευμένα Συστήματα Νέων Κατ' Απαίτηση, Διπλωματική εργασία που υποβλήθηκε στο Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης, 1998.

[Σκον98] Σκόνδρας Παναγιώτης, Σχεδιασμός και Υλοποίηση Μηχανισμών Προώθησης και Ανέλκυσης (Push & Pull) σε Σύστημα Νέων Κατ' Απαίτηση, Διπλωματική εργασία που υποβλήθηκε στο Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης, 1998.

[Salton89] Gerald Salton, Automatic Text Processing, Addison-Wesley 1989

[Hyn1] INTRACOM, HyNoDe, Specification of the HyNoDe Initial Prototype, 1997.

[Hyn2] INTRACOM, HyNoDe, Preliminary market survey for the NoD application, 1996.

[Hyn3] INTRACOM, HyNoDe, Definition of NoD Scenarios, 1996.

[Hyn4] INTRACOM, HyNoDe, Identification of HyNoDe Actors, 1996.

[Hyn5] IBM, HyNoDe, Specification of an HTML-based News Document structure, 1996.

[Hyn6] ETNOTEAM, HyNoDe, Browsing tool adapted to the document structure, 1997.

[Hyn7] COSI, HyNoDe, Specification of the News Authoring Tool (HTML based), 1997.

[O'Neil94] O'Neil Patrick, Database, Principles, Programming, Performance, Morgan Kaufmann Publishers, 1994.

[Unicode] The Unicode Consortium, The Unicode Standard, http://www.unicode.org/unicode/standard/standard.html

[Cass97] Cassady-Dorion Luke, Industrial Strength Java, New Riders Publishing, 1997.

[Java1] Java Technology - http://www.javasoft.com/

[Java2] Java Developer Connection - http://developer.javasoft.com/developer/

[Sol97] Solomon David S., Microsoft SQL Server 6.5 Unleashed, HW Sams, 1997.

[Souk] Soukup P., Inside Microsoft SQL Server 6.5, Microsoft Press, 1997.

[Swi1] The Swing Connection - http://java.sun.com/products/jfc/tsc/

[Firew1] Firewall Design, http://www.sunworld.com/swol-01-1996/swol-01-firewall.html