

**Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης**

**Σχολή Τεχνολογικών Εφαρμογών  
Τμήμα Μηχανικών Πληροφορικής**

**Δημιουργία survey για τεχνολογίες case base  
reasoning στην εκπαίδευση**

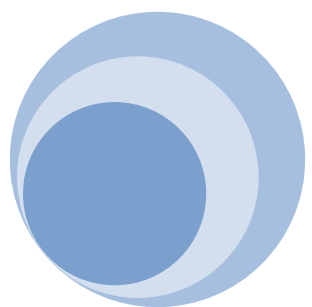
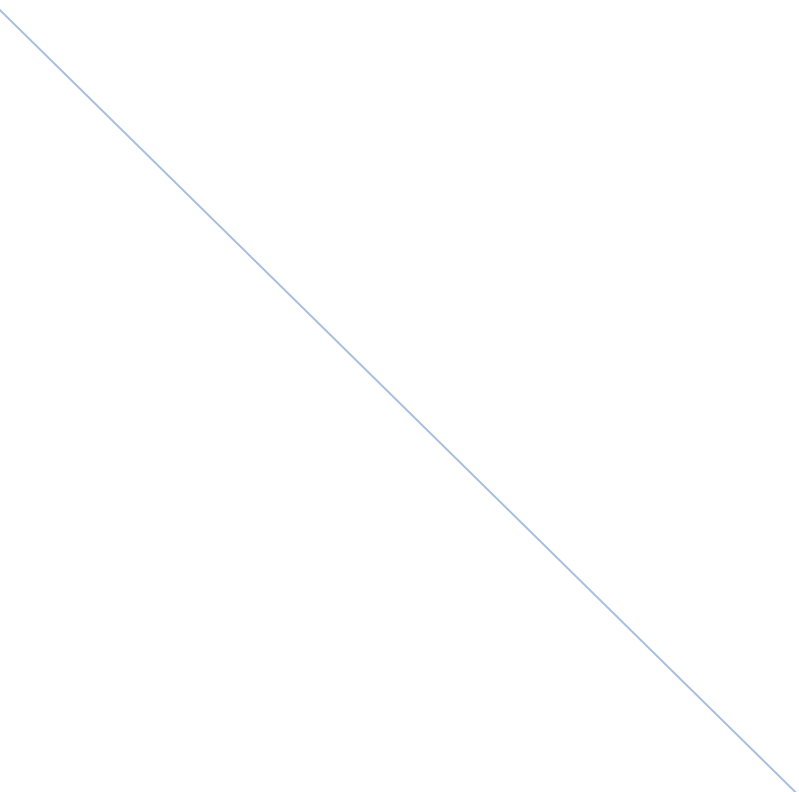
ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ



ΦΟΙΤΗΤΕΣ: ΣΚΟΡΙΔΑΚΗΣ ΑΝΑΣΤΑΣΙΟΣ ΑΜ 1248  
ΚΑΛΟΓΙΑΝΝΑΚΗΣ ΑΛΕΞΑΝΔΡΟΣ ΑΜ 1405

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Παπαδάκης Νικόλαος

**ΙΟΥΛΙΟΣ 2013**



# ΠΕΡΙΛΗΨΗ

Με τη παρούσα διπλωματική εργασία ο αναγνώστης έχει τη δυνατότητα να πληροφορηθεί για τη σημασία της δημιουργίας συστήματος για την ηλεκτρονική μάθηση (e-Learning) με χρήση της Ελληνικής γλώσσας ενός μαθήματος πληροφορικής. Σκοπός της εργασίας είναι να αναπτυχθεί ένα σύστημα που θα παρέχει την δυνατότητα διευκόλυνσης τόσο από τη σκοπιά του καθηγητή όσο και του μαθητή για ένα μάθημα που διδάσκετε στα ελληνικά. Σήμερα υπάρχουν πολλά εργαλεία στον τομέα της ηλεκτρονικής μάθησης για μαθήματα που διδάσκονται σε αγγλική γλώσσα. Με την εργασία αυτή θα αναπτυχθεί ένα σύστημα ως ενδιάμεσο των εργαλείων και στις πλατφόρμες ηλεκτρονικής μάθησης, ώστε να είναι εφικτή η χρήση των εργαλείων αυτών σε μαθήματα που διδάσκονται στην ελληνική γλώσσα. Τέλος θα πραγματοποιηθεί εγκατάσταση και έλεγχος του συστήματος σε υπάρχουσες πλατφόρμες ηλεκτρονικής μάθησης.

Ο όρος ηλεκτρονική μάθηση χαρακτηρίζει ένα διακριτό τύπο εκπαίδευσης, όπου στη διδασκαλία ως βασικό μέσο χρησιμοποιούνται ηλεκτρονικοί υπολογιστές, συνήθως δικτυωμένοι, με σκοπό την παραγωγή, προώθηση, παράδοση και διευκόλυνση εξατομικευμένης μάθησης οπουδήποτε και οποτεδήποτε, μέσα από ένα χρηστο-κεντρικό περιβάλλον. Το e-Learning καλύπτει πλήθος εφαρμογών και διαδικασιών, όπως είναι, ενδεικτικά, η μάθηση μέσω ηλεκτρονικών υπολογιστών και δικτύων υπολογιστών, οι εικονικές αίθουσες και η ψηφιακή συνεργατική μάθηση. Περιλαμβάνει τη διανομή του εκπαιδευτικού υλικού μέσω διαδικτύου, ενδοδικτύων (LAN/WAN), κασετών ήχου και βίντεο, δορυφορικής ή αλληλεπιδραστικής τηλεόρασης, cd-rom κ.ά. Πρόσφατα προτάθηκε να αντικατασταθεί από τον όρο Τεχνολογικά Ενισχυμένη Μάθηση.

Τα Συστήματα Διαχείρισης Μάθησης(ΣΔΜ) είναι λογισμικά που συνδυάζουν τη λειτουργικότητα των επικοινωνιών μέσω υπολογιστή, τις online μεθόδους παράδοσης διδακτικών υλικών και τα εργαλεία διαχείρισης της μαθησιακής διαδικασίας, παρέχοντας ένα ολοκληρωμένο διαδικτυακό περιβάλλον μάθησης. Ένα Σύστημα Διαχείρισης Μάθησης εγγράφει τους χρήστες, καταχωρεί τα μαθήματα σε καταλόγους, καταγράφει τα δεδομένα των σπουδαστών κ.ά. Ένα ΣΔΜ παρέχει σε ένα ίδρυμα τη δυνατότητα να αναπτύξει και να προσφέρει ηλεκτρονικά μαθησιακά υλικά στους εκπαιδευόμενους και εν συνεχεία να τους αξιολογήσει και να δημιουργήσει βάσεις δεδομένων, όπου θα καταγράφονται τα αποτελέσματα και η πρόοδός τους.

**Λέξεις Κλειδιά :** e-learning, Σύστημα Διαχείρισης Μάθησης, δίκτυο, μαθησιακό υλικό, stemming, Case Based Reasoning

## Ευχαριστίες

**Μ**ε το πέρας της διπλωματικής αυτής εργασίας , θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους εκείνους που μου προσέφεραν τη πολύτιμη βοήθεια τους στην εκπόνησή της. Αρχικά θα ήθελα να ευχαριστήσω τον κο. Παπαδάκη επιβλέποντα της διπλωματικής μου εργασίας για την εμπιστοσύνη που μου έδειξε με την ανάθεσή της, καθώς και την πολύτιμη βοήθεια, την καθοδήγηση και στήριξη που μου προσέφερε σε όλη τη διάρκεια της εκπόνησής της. Τέλος ένα μεγάλο ευχαριστώ στην οικογένειά μου που βρίσκεται στο πλευρό μου και με στηρίζει σε κάθε μου προσπάθεια όλα αυτά τα χρόνια, καθώς και τους εξαιρετικούς συναδέλφους και φίλους μου που στάθηκαν συνοδοιπόροι στο ταξίδι προς τη γνώση.



# Πίνακας περιεχομένων

## Περιεχόμενα

ΕΙΣΑΓΩΓΗ.....	10
ΚΕΦΑΛΑΙΟ 1.....	12
Βιβλιογραφική ανασκόπηση .....	12
1.1 Η εμφάνιση του E-learning .....	12
1.2 Απαραίτητες προϋποθέσεις για τη δημιουργία περιβάλλοντος E-learning .....	13
1.3 Μορφές της εξ αποστάσεως μάθησης. ....	14
1.3.1 Ορισμός ασύγχρονης τηλεκπαίδευσης.....	18
1.3.2 Σχεδιασμός ασύγχρονης πλατφόρμας e-learning .....	18
1.3.3 Ορισμός σύγχρονης τηλεκπαίδευσης.....	20
1.3.4 Δημιουργία εικονικής τάξης .....	20
1.4 Πλεονεκτήματα και μειονεκτήματα E-learning .....	21
1.4.1 Πλεονεκτήματα e-learning.....	21
1.4.2 Μειονεκτήματα e-learning .....	22
1.5. Που μπορεί να χρησιμοποιηθεί η τηλεκπαίδευση;.....	23
1.5.1 Εφαρμογές e-learning στην Τριτοβάθμια εκπαίδευση .....	23
1.5.2 Εφαρμογές E-learning στην πρωτοβάθμια και δευτεροβάθμια .....	24
1.5.3 Επιμόρφωση-Κατάρτιση και e-learning.....	25
Κεφάλαιο 2 .....	27
Η διαδικασία Αναγνώρισης Προτύπων (stemming) και οι υπάρχουσες τεχνικές .....	27
2.1 Εισαγωγή.....	27
2.2 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ.....	28
2.3 Dictionary-Based Τεχνική.....	28
2.4 Rule-Based Τεχνική .....	29
2.5 Light-Stemming Τεχνική.....	30

2.6 Corpus-Based Τεχνική .....	30
2.7 Stemming για μη αγγλικές γλώσσες .....	31
2.8 Ελληνικοί Stemmers.....	32
2.9 Suffix stripping αλγόριθμος .....	34
2.10 ΠΡΟΒΛΗΜΑ.....	36
2.11 ΑΠΟΤΕΛΕΣΜΑ .....	37
2.12 ΣΚΟΠΟΣ .....	37
2.13 ΜΕΘΟΔΟΣ .....	37
2.14 ΠΕΡΙΟΡΙΣΜΟΙ .....	38
Κεφάλαιο 3 .....	40
Η διαδικασία Αναγνώρισης Προτύπων στην Ελληνική Γλώσσα (Greek stemming).....	40
3.1. Η Ιστορική αναδρομή της Νέας Ελληνικής γλώσσας.....	40
3.2 Το ελληνικό αλφάβητο .....	41
3.3. ΟΡΙΣΜΟΙ .....	42
3.5. Stem σε Ελληνικές λέξεις .....	44
3.6. Παραδοχές για τον Ελληνικό Stemmer.....	46
3.7 Ο Stemmer για την Ελληνική γλώσσα .....	47
3.7.1. Ο Αλγόριθμος.....	47
Κεφάλαιο 4 .....	49
Η διαδικασία Αναγνώρισης Προτύπων στην Ελληνική Γλώσσα (Greek stemming).....	49
4.1. Εισαγωγή.....	49
4.2 Background Work and Relative Work.....	51
Jcolibri .....	51
4.3 Στατιστική (statistical) Μέθοδος.....	52
4.4 Η δική μας Επέκταση .....	52
4.5 Σύστημα .....	53
4.6 Αρχιτεκτονική.....	54

4.7 Μεθοδολογίες.....	55
4.7.1 Στάδιο προ-επεξεργασίας.....	55
4.7.2 Γλωσσική προ-επεξεργασία (Λεξιλογική Ανάλυση) .....	56
4.7.3 Αποκλεισμός Λέξεων( stopwords) .....	56
4.7.4 Στελέχωση Κειμένου (Stemming) .....	57
4.7.5 Ανάκτηση Πληροφορίας .....	58
4.7.6 Lucene .....	59
4.7.7 Υπολογισμός Σημαντικότητας Όρων(weighting) .....	59
4.7.8 Συναρτήσεις ομοιότητας .....	61
4.7.9 Μέθοδοι Διαχείρισης Της Γνώσης.....	64
4.7.10 Στατιστική Μέθοδος .....	64
4.7.11 Μέθοδος Ευρετηρίου .....	65
4.7.12 Μέθοδος Ευρετηρίου με τη χρήση Συνώνυμων Όρων.....	67
4.7.13 Μέθοδος Ευρετηρίου με τη χρήση Συνώνυμων Όρων θέτοντας βάρη. ....	68
4.8 Υλοποίηση.....	70
4.9 Efront .....	72
4.10 Επεξεργασία και Μετατροπή του αρχείου xml .....	74
4.11 Επεξεργασία Ερωτήματος.....	76
4.12 Μετρήσεις - Αξιολόγηση.....	81
4.13 Εφαρμογή στον κλάδο της Πληροφορικής.....	82
4.14 Αξιολόγηση Μεθοδολογιών στον κλάδο της Πληροφορικής.....	83
4.15 Εφαρμογή στον κλάδο της Ψυχολογίας .....	85
4.16 Αξιολόγηση Μεθοδολογιών στον τομέα της Ψυχολογίας .....	86
4.17 Αποτελέσματα Αξιολόγησης.....	87
5. Συμπεράσματα .....	94
Μελλοντική Δουλειά.....	<b>Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.</b>



5.1 Καταγραφή της επίδοσης των εκπαιδευόμενων .....	95
5.2 Προσαρμογή του συστήματος στην ελληνική γλώσσα.....	96
5.3 Προσαρμογή περιβάλλοντος Efront.....	96
5.4 Παροχή Εξατομικευμένης Πρόσβασης σε Υπηρεσίες εξ Αποστάσεως Εκπαίδευσης(e-Learning Personalization).....	97
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>	<b>98</b>

## ΕΙΣΑΓΩΓΗ

Στο πέρασμα των αιώνων συντελούνται στην Ιστορία τομές, οι οποίες, χωρίς να διακόπτουν την «εξέλιξη», επιφέρουν δραστικές αλλαγές στον τρόπο ζωής των ανθρώπων. Κάποιες φορές μάλιστα χαρακτηρίζονται, εξαιτίας της έντασης των συνεπειών τους, ως επαναστάσεις, όπως για παράδειγμα Γεωργική επανάσταση, Βιομηχανική, Υψηλής τεχνολογίας. Στις μέρες μας βιώνουμε την τρίτη γενιά της τεχνολογικής επανάστασης, στα πλαίσια της οποίας δικαίως μπορεί να υποστηριχθεί ότι «οι νέες τεχνολογίες –που έχουν κατακλύσει τη ζωή μας– επιδρούν στον τρόπο σκέψης και γενικότερα στις διαπροσωπικές σχέσεις καθώς και σε αυτές τις ίδιες τις προσδοκίες των ανθρώπων» (Παγγέ, 2004α).

Δημιουργείται έτσι μια κοινωνία όπου όλοι οφείλουν να είναι εξοικειωμένοι με τις ΤΠΕ, ενώ ο τεχνολογικά αναλφάβητος ρισκάρει να καταλήξει ο «κοινωνικά απόβλητος». Πόσοι όμως κατάφεραν μέχρι τώρα ή μπορούν ακόμα και σήμερα να ενταχθούν στην σύγχρονη αυτή κοινωνία; Οι άνθρωποι που στερούνται της δυνατότητας ή των ικανοτήτων πρόσβασης σε Η/Υ δεν είναι λίγοι. Βρισκόμαστε λοιπόν μπροστά στον άμεσο κίνδυνο επικράτησης του νέου φαινομένου του «ψηφιακού διΐσμου» (Αναστασιάδης, 2005), δηλαδή της ύπαρξης κοινωνιών δύο ταχυτήτων. Η ευνοητή λύση, για την αποφυγή ακραίων καταστάσεων, είναι ότι η μύηση στις νέες τεχνολογίες πρέπει να ξεκινήσει από τα πρώτα στάδια ανάπτυξης των παιδιών. Συνεπώς, θεωρείται δεδομένη η επιστράτευση της Εκπαίδευσης, η οποία «καλείται να ενσωματώσει δημιουργικά την πληροφορική κουλτούρα, να την εναρμονίσει με την ανθρωπιστική παιδεία και να την επαναδιαμορφώσει με τρόπο ώστε να ανταποκρίνεται στους στόχους που η ίδια η εκπαίδευση θέτει» (Βρύζας και Τσιτουρίδου 2005).

Το πρόβλημα που προκύπτει στην συγκεκριμένη περίπτωση είναι το πώς αλλά και ποιοι θα παρέχουν αυτή την Εκπαίδευση. Στο πρώτο σκέλος του ερωτήματος η πολιτεία απαντά με συνεχείς προσπάθειες, αφενός για εξοπλισμό των σχολείων με τα ανάλογα τεχνολογικά μέσα και αφετέρου για εκσυγχρονισμό του αναλυτικού προγράμματος, δεδομένου ότι μέχρι πρόσφατα περιοριζόταν στην χρήση του Η/Υ για την «Εκπαίδευση στην Πληροφορική», παραμελώντας την εισαγωγή της «Πληροφορικής στην Εκπαίδευση» (Καρτσιώτης & Καρατάσιος, 2008; Κυνηγός κ.α.,

2000). Η πρώτη δράση βέβαια είναι σαφώς απαραίτητη για την εξοικείωση των εκκολαπτόμενων πολιτών με την τεχνολογία και την εκμάθηση της χρήσης των υπολογιστικών εφαρμογών, που θα τους προετοιμάσουν για τις νέες και συνεχώς αυξανόμενες απαιτήσεις της ανταγωνιστικής αγοράς εργασίας, εξασφαλίζοντας παράλληλα την οικονομική και τεχνολογική ανάπτυξη της χώρας. Η δεύτερη δράση όμως, όπου οι ΝΤ ενσωματώνονται στην διδακτική διαδικασία, «σηματοδοτεί» κατά τους Ράπτη και Ράπτη (1999) «την απαρχή μιας περιόδου παιδαγωγικής αναγέννησης στη σχολική πραγματικότητα», προσφέροντας νέους τρόπους αξιοποίησης της τεχνολογίας και προσέγγισης της γνώσης και καλλιεργώντας πλήθος δεξιοτήτων των παιδιών (αντιληπτική ικανότητα, κριτική σκέψη, συνεργατικό πνεύμα κτλ.) (Τζιμογιάννης-Σιορέντα, 2007). Είναι όμως οι εκπαιδευτικοί στο σύνολο τους προετοιμασμένοι να αντιμετωπίσουν μια τέτοια πρόκληση; Η αρνητική απάντηση εδώ θεωρείται δεδομένη.

Στην παρούσα εισήγηση προτείνεται καταρχήν μια λύση για την επιμόρφωση των

εκπαιδευτικών με τον πιο σύγχρονο τρόπο εκπαίδευσης, την εξ αποστάσεως εκπαίδευση μέσω διαδικτύου, ήτοι το e-learning (τηλεκπαίδευση ή ηλεκτρονική μάθηση), της οποίας αναλύεται το «χρονικό εισβολής» στα εκπαιδευτικά δρώμενα.

Στη συνέχεια περιγράφεται και μια ανάλογη εφαρμογή, η κατασκευή και χρήση μιας πλατφόρμας ασύγχρονης τηλεκπαίδευσης στα πλαίσια μιας πρότασης επιμόρφωσης στη Μουσειακή Εκπαίδευση με το μοντέλο της υβριδικής μάθησης (blended learning).

## ΚΕΦΑΛΑΙΟ 1

# Βιβλιογραφική ανασκόπηση

### 1.1 Η εμφάνιση του E-learning

Σύμφωνα με τους Horton και Horton (2003) ως e-learning ορίζεται «κάθε χρήση των τεχνολογιών του Διαδικτύου και του Παγκόσμιου Ιστού για τη δημιουργία μαθησιακών εμπειριών που μπορεί να είναι προϊόν αυτορρυθμιζόμενης μάθησης με την συμβουλευτική καθοδήγηση από τον διδάσκοντα/επιμορφωτή». Η συγγραφέας Παντάνο- Ρόκου το θεωρεί ως την «χρήση των πολυμέσων και διαδικτύου για τη βελτίωση της ποιότητας εκπαίδευσης μέσω της διευκόλυνσης πρόσβασης σε πηγές πληροφοριών και υπηρεσίες αλλά και στην εξ αποστάσεως συνεργασία και ανταλλαγή απόψεων» (2005).

Το e-learning εμφανίζεται στις ΗΠΑ στα μέσα της δεκαετίας του '90 σε μορφή CBT (Computer Based Learning). Στην Ευρώπη γίνεται άμεσα αντιληπτή η αξία του, και το 2000 ψηφίζεται στη Λισσαβόνα το σχέδιο δράσης με τίτλο «e-learning: σχεδιάζοντας την εκπαίδευση του αύριο», ενώ στις 15 Μαΐου 2001 το e-learning αναγνωρίζεται από το Ευρωπαϊκό Κοινοβούλιο ως μια πρωτοβουλία που συμβάλει στην ενίσχυση της ιδέας ενός «ενιαίου ευρωπαϊκού χώρου εκπαίδευσης», που έρχεται να συμπληρώσει τον ευρωπαϊκό χώρο έρευνας και την ενιαία ευρωπαϊκή αγορά. Στη σύνοδο της Βαρκελώνης, ένα μόλις χρόνο μετά, αναγνωρίζεται η συμβολή του σε πολλούς τομείς, όπως στις προκλήσεις της κοινωνίας της γνώσης, στη βελτίωση ποιότητας εκπαίδευσης, στη διευκόλυνση πρόσβασης σε μαθησιακούς πόρους, στην αντιμετώπιση ειδικών αναγκών, στην ανάπτυξη αποτελεσματικότερης και αποδοτικότερης εκπαίδευσης και κατάρτισης στο χώρο εργασίας ([www.elearningeuropa.info](http://www.elearningeuropa.info)· Κεραμιδά & Ψιλέλης, 2005).

## 1.2 Απαραίτητες προϋποθέσεις για τη δημιουργία περιβάλλοντος E-learning

Η ανάπτυξη ενός περιβάλλοντος e-learning προϋποθέτει την ύπαρξη εκπαιδευτικού και διδακτικού σχεδιασμού, που θα ευνοούν την αποτελεσματικότερη αξιοποίηση των Νέων Τεχνολογιών, την οργάνωση των διδακτικών στρατηγικών και την βέλτιστη παρουσίαση του εκπαιδευτικού υλικού, επιτρέποντας την ανατροφοδότηση, την άμεση ανάδραση, την εξατομικευμένη μάθηση (Παντάνο-Ρόκου, 2005). Στην ουσία το ζητούμενο είναι η δημιουργία ενός περιβάλλοντος όπου θα προκαλείται η εσωτερική υποκίνηση του εκπαιδευόμενου, θα καλλιεργείται η συνεργατική μάθηση με τη δημιουργία στους συμμετέχοντες μιας «αίσθησης ακροατηρίου», θα εξασφαλίζεται η συνεχής υποστήριξη του εκπαιδευτή προς τον εκπαιδευόμενο, θα προσφέρεται η δυνατότητα αξιολόγησης (Μισετζής & Μαυροειδής, 2005). Με αυτό τον τρόπο, λοιπόν, οι μεθοδολογίες εξ αποστάσεως εκπαίδευσης σε συνδυασμό με περιβάλλοντα τηλεδιάσκεψης ή σύγχρονης συνεργατικής μάθησης μέσω διαδικτύου μπορούν να συνεισφέρουν στην ανύψωση του επιπέδου της εκπαίδευσης, να μειώσουν το κόστος της επανεκπαίδευσης, της κατάρτισης και της επιμόρφωσης (Παναγιωτακόπουλος, 2008).

Το προφίλ του «ιδανικού» ενήλικα εκπαιδευόμενου για τα δεδομένα του e-learning θα πρέπει να συνδυάζει διάφορα στοιχεία, όπως για παράδειγμα ισχυρή θέληση και ενθουσιασμό για το πρόγραμμα, οικογενειακή στήριξη, τεχνολογική επάρκεια, αλλά και δεξιότητες που αφορούν στην αυτορρυθμιζόμενη μάθηση και τη γρήγορη κατανόηση για την ταχύτερη εξέλιξη του προγράμματος (Παγγέ, 2004β).

Κεφαλαιώδους σημασίας όμως είναι και η ικανότητα και θέληση για συνεργασία, καθώς, όπως δείχνει και η εμπειρία από επιμορφώσεις ενηλίκων, απαραίτητη στα μαθήματα τηλεεκπαίδευσης στις εκπαιδευτικές επαγγελματικές κοινωνίες, είναι η δημιουργία «κοινότητας μάθησης», η οποία δε δημιουργείται ξαφνικά, αλλά ως προϊόν ενεργειών και προσπαθειών. Τοιουτοτρόπως, η εποικοδόμηση γνώσης συνιστά αποτέλεσμα διαδραστικού σχεδιασμού με τη συνεργασία όλων των συμμετεχόντων ανεξάρτητα από την απόσταση που τους χωρίζει (Lambropoulos, 2005· Mc Donald et al, 2007· Μπιγιλάκη, 2001· Χλαπάνης, 2005).

Ο ρόλος του εκπαιδευτή στο e-learning, από την άλλη, ποικίλει από ήπειρο σε ήπειρο (π.χ. στις ΗΠΑ είναι κόουτς, ενώ στην Ευρώπη και την Ασία έχει παιδαγωγικό ρόλο με έμφαση στη συμβουλευτική), αλλά και κατά περίπτωση τον συναντάμε ως κόουτς, διαχειριστή, διευκολυντή, σύμβουλο, αξιολογητή κτλ. ή και με συνδυασμό των ανωτέρω (Goga & Goga, 2007). Ο διαχειριστής της διαδικασίας τηλεκπαίδευσης (συνήθως ο εκπαιδευτής ή επιμορφωτής) αναλαμβάνει το διπλό ρόλο να διοικήσει την εικονική τάξη αλλά και να εκπαιδεύσει τους συμμετέχοντες μέσα από μια γκάμα εργαλείων, που τους δίνουν πολλαπλές δυνατότητες (π.χ. να φτιάχνει χώρους συζητήσεων, να αναρτά ανακοινώσεις, βαθμούς και εργασίες, να επικοινωνεί με εικόνα και ήχο σε πραγματικό χρόνο, να επιτρέπει στους συμμετέχοντες να παρακολουθήσουν μαθήματα με την εικόνα του και αυτή των διαφανειών του, να διαχειρίζεται το διδακτικό υλικό αλλά και να το συγγράφει, να δημιουργεί φύλλα αξιολόγησης, να στέλνει ηλεκτρονικά μηνύματα, κτλ.) (Παναγιωτακόπουλος, 2008·Σιδηρόπουλος κ.α., 2005). Στη συνέχεια εξετάζονται τα δύο πιο διαδεδομένα μοντέλα e-learning, η Ασύγχρονη και Σύγχρονη Τηλεκπαίδευση.

### **1.3 Μορφές της εξ αποστάσεως μάθησης.**

Η εξ αποστάσεως μάθηση έχει τρεις μορφές λειτουργίας, ανάλογα με τον τρόπο επικοινωνίας των συμμετεχόντων: η σύγχρονη, η ασύγχρονη και η μικτή (Horton, 2000:55 – 56, Βασιλάκης, 2006:12).

Στη σύγχρονη μορφή οι συμμετέχοντες έχουν τη δυνατότητα άμεσης επικοινωνίας σε «πραγματικό χρόνο» χρησιμοποιώντας τεχνολογίες τηλεδιάσκεψης, γι' αυτό και η διαδικασία μοιάζει αρκετά με τη συμβατική διδασκαλία. Αντίθετα, στην ασύγχρονη μορφή οι συμμετέχοντες δεν έχουν τη δυνατότητα άμεσης επικοινωνίας. Το εκπαιδευτικό υλικό είναι ψηφιακό και παρέχεται στον εκπαιδευόμενο είτε σταδιακά με την πορεία του μαθήματος, όπου ο εκπαιδευτής καθορίζει τη ροή της διδασκαλίας, είτε ολόκληρο στην έναρξη του μαθήματος, όπου ο εκπαιδευόμενος προγραμματίζει ατομικά το ρυθμό μάθησης, γι' αυτό και θεωρείται περισσότερο ευέλικτη. Αναφέρεται μάλιστα από τον Dr. Hester B. Jones ότι ο συνδυασμός σύγχρονης και ασύγχρονης

τηλεκπαίδευσης βελτιώνει την απόδοση των σπουδαστών κατά 8 – 9%. Ως εργαλεία ασύγχρονης και σύγχρονης επικοινωνίας μέσω διαδικτύου χρησιμοποιούνται το ηλεκτρονικό ταχυδρομείο (e – mail), οι λίστες συζητήσεων (mailing lists), οι ομάδες συζητήσεων (newsgroups), οι πίνακες ανακοινώσεων (BBS – Bulletin Board System), η συνομιλία με γραπτά μηνύματα σε πραγματικό χρόνο (chat), συχνά ενταγμένα σε συγκεκριμένο λογισμικό περιβάλλον. Θα πρέπει να σημειωθεί ότι παρόλο που ο Horton (2000:56) θεωρεί πως η σύγχρονη τηλεκπαίδευση ταιριάζει περισσότερο σε ακαδημαϊκούς χώρους, σχολεία κλπ, ενώ η ασύγχρονη είναι πιο κατάλληλη για τη βιομηχανία και τις επιχειρήσεις, μελέτες σε ελληνικά σχολεία της Δευτεροβάθμιας εκπαίδευσης (Manitsaris – Perdos – Pavlidis, 2006:217 – 218) έχουν δείξει ότι και η ασύγχρονη τηλεκπαίδευση είναι σημαντική για τη βελτίωση ή την υποβοήθηση της διδασκαλίας σε αυτά.

Ως βάση ασύγχρονης μορφής εξ αποστάσεως μάθηση θα μπορούσε να θεωρηθεί ακόμη και μία απλή ιστοσελίδα, στην οποία ο εκπαιδευτής ανεβάζει το υλικό του μαθήματος και στη συνέχεια οι εκπαιδευόμενοι παραδίδουν τις εργασίες τους μέσω ηλεκτρονικού ταχυδρομείου. Ωστόσο σήμερα κάτι τέτοιο θα ήταν πολύ λίγο, καθώς με την εξέλιξη της τεχνολογίας, την αποκτηθείσα εμπειρία και τους ολοένα πιο απαιτητικούς χρήστες, μια ικανοποιητική πλατφόρμα για ασύγχρονη τηλεκπαίδευση πρέπει και μπορεί να υποστηρίξει, μέσω μιας ενιαίας και φιλικής προς τον χρήστη διεπιφάνειας, τις εξής βασικές λειτουργίες (Avgeriou – Papasalouros – Retalis – Skordalakis, 2003:3 και Αυγερίου – Παπασαλούρος – Ρετάλης – Ψαρομηλίγκος, 2005:132):

- Κατανομή της πληροφορίας (Information distribution)
- Διαχείριση του μαθησιακού υλικού (Management of learning material)
- Παροχή πολλαπλών δυνατοτήτων επικοινωνίας (Multiple communication facilities) και
- Διαχείριση τάξης (Class management)

Ειδικότερα θα πρέπει:

- Να υποστηρίξει κάποιου είδους πιστοποίηση των χρηστών αλλά και χωρισμό τους σε ομάδες, έτσι ώστε η ίδια πλατφόρμα να μπορεί να χρησιμοποιηθεί για περισσότερα από ένα μαθήματα.

- Να έχει φιλικό περιβάλλον για τους χρήστες (εκπαιδευόμενους και εκπαιδευτές).
- Να υποστηρίζει προσαρμογή του περιβάλλοντος ανάλογα με τις προτιμήσεις του κάθε χρήστη. Επίσης να κρατάει πληροφορίες για τον χρήστη (δημιουργία profiles) και να τον «βοηθάει» κατά την πλοήγηση.
- Το περιβάλλον να ανοίγει με απλό φυλλομετρητή (web browser), ώστε να είναι προσβάσιμο από παντού με οποιοδήποτε λειτουργικό σύστημα, και να μην είναι απαραίτητη η εγκατάσταση άλλου λογισμικού από τους χρήστες. Να παρέχει εύκολο τρόπο για την τοποθέτηση αφενός του υλικού του μαθήματος από τον εκπαιδευτή και αφετέρου των εργασιών των εκπαιδευομένων.
- Να υποστηρίζει την παρουσίαση αρχείων βίντεο, ήχου, εικόνων κλπ.
- Να δίνει τη δυνατότητα στους χρήστες τοπικής αποθήκευσης του υλικού του μαθήματος για επεξεργασία εκτός δικτύου.
- Το υλικό του μαθήματος να προσφέρεται και σε εύκολα εκτυπώσιμη μορφή, για τους χρήστες εκείνους που προτιμούν έντυπο υλικό.
- Να υποστηρίζει τη δημιουργία βημάτων συζήτησης (discussion forums) και δωματίων συζητήσεων (chat rooms), για την ασύγχρονη επικοινωνία εκπαιδευομένων και εκπαιδευτή, αλλά και για συζήτηση σε πραγματικό χρόνο.
- Να υλοποιεί ηλεκτρονικό ταχυδρομείο (e – mail), για την καλύτερη επικοινωνία των χρηστών, και να έχει ημερολόγιο με τις προθεσμίες και άλλα σημαντικά γεγονότα.
- Να υποστηρίζει την εύκολη δημιουργία διαγωνισμάτων (on – line tests) και να παρακολουθεί την πρόοδο των εκπαιδευομένων.

Σήμερα η εξ αποστάσεως μάθηση στην Ελλάδα χρησιμοποιεί κατά κύριο λόγο μια μικτή μορφή. Υπάρχουν κατάλληλα διαμορφωμένοι χώροι όπου γίνονται ηλεκτρονικά μαθήματα και μοιάζουν με τις πραγματικές τάξεις ενώ υπάρχει και ηλεκτρονικό υλικό το οποίο μπορεί να χρησιμοποιήσει κάποιος συμπληρωματικά στον παραδοσιακό τρόπο εκπαίδευσης. Η μικτή αυτή μορφή επιτυγχάνεται με τη χρήση συγκεκριμένων εφαρμογών. Οι εφαρμογές αυτές τυπικά ονομάζονται «Συστήματα Διαχείρισης Μάθησης» (ΣΔΜ – Learning Management Systems LMS). Βέβαια, οι ίδιες εφαρμογές μπορεί να έχουν διαφορετική ονομασία ανάλογα με τον τομέα στον οποίο στοχεύουν οι κατασκευαστές τους, όπως για παράδειγμα: «Συστήματα Διαχείρισης



Περιεχομένου» (Content Management Systems – CMS), «Συστήματα Διαχείρισης Μαθησιακού Περιεχομένου» (Learning Content Management Systems – LCMS), «Εικονικά Περιβάλλοντα Μάθησης» (Virtual Learning Environments), «Πλατφόρμες Ασύγχρονης Τηλεκπαίδευσης (τηλεμάθησης)», «Συστήματα Διαχείρισης Μαθημάτων» (Course Management Systems), «Περιβάλλοντα Ελεγχόμενης Μάθησης» (Managed Learning Environments – MLE), «Συστήματα Υποστήριξης της Μάθησης» (Learning Support Systems – LSS), «Διαδικτυακά Περιβάλλοντα Μάθησης» (On – line Learning Environments ή Networked Learning Environments), «Συνεργατικό Λογισμικό Μάθησης» (Collaborative Learning Software), «Εικονικές Τάξεις» (Virtual Classrooms), «Ολοκληρωμένα Μαθησιακά Περιβάλλοντα», «Ολοκληρωμένα Συστήματα Μάθησης» (Integrated Learning Systems), «Λογισμικό Διαδικτυακής Παράδοσης Μαθημάτων» (On – line Course Delivery Software), «Ολοκληρωμένα Συστήματα Παράδοσης Μαθημάτων» (Integrated Course Delivery Systems), «Πύλες Μάθησης» (Learning Portals), «Λύσεις Ηλεκτρονικής Μάθησης» (e – Learning Solutions), «Συστήματα Διαχείρισης Επιμόρφωσης» (Training Management System), «Συστήματα Διοίκησης Επιμόρφωσης» (Training Administration System), κ.λπ.

Γενικά, μπορούμε να πούμε ότι ένα Σύστημα Διαχείρισης Μάθησης είναι βασικά ένα πακέτο λογισμικού που οργανώνει, διαμοιράζει και καταγράφει τις εκπαιδευτικές δραστηριότητες μέσα από μια κεντρική διεπιφάνεια, σ' ένα τοπικό ή ευρύτερο δίκτυο (Lioudakis – Kalogiannakis – Psarros – Vassilakis, 2005:793 – 794). Οι βασικές λειτουργίες του περιλαμβάνουν:

- Ελεγχόμενη και ασφαλή πρόσβαση στα διδακτικά υλικά.
- Παρουσίαση της διδακτέας ύλης σε αυτοτελείς δομικές μονάδες (modules).
- Καταγραφή των στοιχείων και της προόδου του μαθητή.
- Υποστήριξη της on – line μάθησης και της πρόσβασης σε άλλες πηγές.
- Σύστημα αξιολόγησης του μαθητή.
- Επικοινωνία του μαθητή με τον εκπαιδευτή ή τους συμμαθητές του.
- Σχηματισμό ομάδων εκπαιδευομένων.
- Διασύνδεση με εξωτερικά συστήματα.

Στα θετικά τους χαρακτηριστικά περιλαμβάνονται οι δυνατότητες παρουσίασης του διδακτικού υλικού και διαχείρισης της μάθησης, οι δυνατότητες κινητοποίησης του

μαθητή, επικοινωνίας και συνεργασίας αλλά και αυτοαξιολόγησης, ενώ αρνητικά αξιολογείται το περιορισμένο εύρος δημιουργικότητας και η περιορισμένη διαδραστικότητα, καθώς και η έμφαση που συχνά δίνεται στον έλεγχο και την αξιολόγηση, περισσότερο ίσως από την ίδια τη μάθηση. Άλλες αδυναμίες τους οφείλονται στις ακατάλληλες μεθοδολογικές προσεγγίσεις, στον πλημμυρή σχεδιασμό των εφαρμογών ή και στον υψηλό βαθμό τεχνικής ικανότητας που απαιτείται από τους χρήστες. (Παναγιωτίδης, 2005:235).

### **1.3.1 Ορισμός ασύγχρονης τηλεεκπαίδευσης**

Η ασύγχρονη τηλεεκπαίδευση παρέχει ένα ευέλικτο μοντέλο εκπαίδευσης, καθώς δεν απαιτεί την παρουσία εκπαιδευτή και εκπαιδευόμενων σε ίδιο χώρο ή στον ίδιο χρόνο μπροστά από τον υπολογιστή, ενώ μπορεί να συντελείται με τη μορφή αυτοδιδασκαλίας, ημιαυτόνομης ή συνεργαζόμενης εκπαίδευσης (Μπαλαούρας). Η συγκεκριμένη μέθοδος, ανεξαρτήτως μορφής, προαπαιτεί την ύπαρξη ενός λογισμικού, το οποίο καλείται σύστημα διαχείρισης μαθησιακού υλικού (LMS=Learning Management System) ή, πιο απλά, πλατφόρμα ασύγχρονης τηλεεκπαίδευσης.

### **1.3.2 Σχεδιασμός ασύγχρονης πλατφόρμας e-learning**

Οι στοιχειώδεις απαιτήσεις τις οποίες οφείλει να ικανοποιεί μια τέτοια πλατφόρμα συνίστανται στο να υποστηρίζει την χρήση forum (για αμφίδρομη ασύγχρονη επικοινωνία πολλών μελών), chat rooms ( για συζητήσεις σε «πραγματικό χρόνο»), email (για προσωπική επικοινωνία), να παρέχει δυνατότητα εύκολης ανάρτησης εκπαιδευτικού υλικού από το διδάσκοντα και εργασιών από τους διδασκόμενους αλλά και αποθήκευσης του υλικού. Για τα καλύτερα παιδαγωγικά αποτελέσματα πρέπει, κατά το σχεδιασμό της, να ληφθούν υπόψη πολλές παράμετροι (Anouris, 2000· Belikov, 2007· Clark & Mayer, 2003· Polymenakou et all, 2005), εκ των οποίων οι σημαντικότερες είναι:

1. εμφάνιση του portal (π.χ. στυλ, ποιότητα και ποσότητα γραφικών, οπτική δομή κτλ.),

2. μέσα περιήγησης, που να διευκολύνουν την διαδικασία (π.χ. σχεδιάγραμμα προγράμματος, σχεδιάγραμμα και εικονική περιήγηση πλατφόρμας, δυνατότητα εύκολης μετάβασης σε «προηγούμενο», «επόμενο» ή «έξοδο», παροχή επιπλέον διευκρινιστικών κινήσεων με pop-up κείμενα κτλ.),
3. τρόπος δόμησης των πληροφοριών, όπου α. να είναι εμφανείς οι συσχετισμοί των διαφόρων screens (π.χ. το κείμενο πρέπει να σχετίζεται άμεσα με τα γραφικά, η ερώτηση δόκιμο είναι να βρίσκεται μαζί με την απάντηση, οι υπερσύνδεσμοι να μην καλύπτουν τη βασική πληροφορία) β. να υπάρχει συνοχή στον τρόπο παρουσίασης (π.χ. στο σχεδιασμό των συνδέσμων, την εμφάνιση και το σχεδιασμό των screens, την χρήση κοινών όρων κτλ.) και περιεχομένου του υλικού, γ. να υποστηρίζεται η διάδραση (π.χ. χρήση συνδέσμων στο βασικό κείμενο που να οδηγούν σε επιπλέον πληροφορίες),
4. εμφανής παροχή βοήθειας με ασύγχρονο και σύγχρονο τρόπο
5. ποιότητα και ποσότητα των παρεχόμενων πληροφοριών/γνώσεων,
6. δυνατότητα επικοινωνίας

Η παιδαγωγική δυναμική της ασύγχρονης τηλεκπαίδευσης συνίσταται, κατά τον Salmon (2001), σε πέντε βασικά στοιχεία: α. εύκολη πρόσβαση και παροχή κινήτρων, β. κοινωνικοποίηση, γ. ανταλλαγή πληροφοριών, δ. οικοδόμηση της γνώσης, ε. δια βίου εξέλιξη. Μέχρι πρόσφατα βασικό μειονέκτημα των περισσότερων πλατφορμών τηλεκπαίδευσης υπήρξε η αδυναμία επικοινωνίας τους με άλλα συστήματα, αποτρέποντας για παράδειγμα την ανταλλαγή εκπαιδευτικού υλικού, τη συλλογή δεδομένων το προφίλ των μαθητών κτλ. Η απάντηση σε αυτό το ζητούμενο δίνεται από το πρότυπο SCORM (Sharable Content Object Reference Model). Αυτό συνιστά την πιο αξιόλογη προσπάθεια τυποποίησης ορισμένων διαδικασιών στην δημιουργία εκπαιδευτικών περιβαλλόντων διαδικτύου, όντας ένα μοντέλο ευέλικτο, επιδεχόμενο βελτιώσεις που στοχεύει στην παροχή ποικίλων και λειτουργικών δυνατοτήτων για ευκολότερη πρόσβαση, δυνατότητα συνεργασίας μεταξύ συστημάτων και επαναχρησιμοποίησης εκπαιδευτικού υλικού σε διαφορετικά περιβάλλοντα (Σιδηρόπουλος κ.α., 2005).

### 1.3.3 Ορισμός σύγχρονης τηλεκπαίδευσης

Στη σύγχρονη τηλεκπαίδευση είναι αναγκαία η ταυτόχρονη συμμετοχή διδάσκοντα και διδασκομένων, αφού η ανταλλαγή απόψεων αλλά και η παροχή εκπαίδευσης πραγματοποιείται σε «πραγματικό χρόνο». Η αμφίδρομη επικοινωνία ήχου ή/και εικόνας επικουρείται με αποστολή αρχείων, μηνυμάτων, χρήση ηλεκτρονικού μαυροπίνακα κτλ. (Μπαλαούρας). Στα μέσα της σύγχρονης τηλεκπαίδευσης εντάσσονται η τηλεδιάσκεψη (αμφίδρομη επικοινωνία με ήχο και δυνατότητα για ανταλλαγή γραπτών μηνυμάτων) αλλά και η εξελιγμένη της μορφή, η βιντεοδιάσκεψη (αμφίδρομη επικοινωνία ήχου και εικόνας και δυνατότητα για ανταλλαγή γραπτών μηνυμάτων) (Παναγιωτακόπουλος, 2008).

### 1.3.4 Δημιουργία εικονικής τάξης

Τα απολύτως απαραίτητα εργαλεία για τη δημιουργία μιας «εικονικής τάξης» είναι η ύπαρξη αλληλεπιδραστικής οπτικοακουστικής επικοινωνίας, ηλεκτρονικού μαυροπίνακα, δυνατότητας διαμοιρασμού εφαρμογών, ενώ παράλληλα μπορούν να χρησιμοποιηθούν προβολή βίντεο, πλοήγηση σε διαδικτυακούς τόπους, χρήση προγραμμάτων προσομοίωσης κτλ. Στα πλαίσια αυτών των «διασκέψεων», προκειμένου να επιτευχθεί το καλύτερο αποτέλεσμα, αφενός οφείλει να τηρείται μια συγκεκριμένη σειρά διαδικασιών από την πρώτη φορά (γνωριμία συντονιστή/διδάσκοντα-συμμετεχόντων, γνωριμία συμμετεχόντων μεταξύ τους, εξάσκηση στη χρήση των εργαλείων) και αφετέρου απαιτείται αρκετή προετοιμασία από τον συντονιστή-διδάσκοντα και συνεχείς προσπάθειες για προσέλκυση του ενδιαφέροντος και παρότρυνση για συμμετοχή (με υποβολή ερωτήσεων, έναρξη συζήτησης, διευκρινιστικές παρεμβάσεις, προβολή εποπτικού υλικού με εφέ κίνησης).

Εντούτοις, διακοπές στη σύνδεση, αναποτελεσματικές σκηνοθετικές επιλογές (όπως τα πολλά παράθυρα που δεν επιτρέπουν την καλή εικόνα του διδάσκοντα), προβλήματα στην παρουσίαση υλικού λειτουργούν φυσικά ανασχετικά μαζί με την διάσπαση της προσοχής λόγω της απουσίας άμεσης επαφής, την κόπωση από τη

συνεχή προσήλωση στην ψηφιακή οθόνη, την τυποποιημένη επανάληψη παραστάσεων κτλ. (Ματθαίου κ.α., 2001· Μπιγιλάκη, 2001).

Συνεπώς, μόνο αν εξασφαλισθεί η τεχνολογική υποδομή και το οργανωτικό πλαίσιο που να διασφαλίζουν την ποιότητα της τηλε/βιντεοδιάσκεψης, το σύγχρονο αυτό περιβάλλον τηλεεκπαίδευσης δύναται να αξιοποιηθεί παιδαγωγικά και να φιλοξενήσει δραστηριότητες που αναδεικνύουν τον ενεργητικό και κατασκευαστικό χαρακτήρα της γνώσης (Καλογιαννάκης κ.α., 2007).

## **1.4 Πλεονεκτήματα και μειονεκτήματα E-learning**

Όπως προκύπτει από έρευνες (Μπαλά, 2005· Ρώσσης & Τζέπογλου, 2001· Sofos, 2005· Χατζηστεφανίδου & Πολυζώης, 2001) που έχουν γίνει σε διάφορους χώρους η εφαρμογή περιβάλλοντος εξ αποστάσεως κατάρτισης μέσω διαδικτύου (e-learning) παρέχει πλεονεκτήματα και μειονεκτήματα, τα σημαντικότερα εκ των οποίων συνοψίζονται στη συνέχεια.

### **1.4.1 Πλεονεκτήματα e-learning**

Σύμφωνα λοιπόν με τις παραπάνω έρευνες τα πλεονεκτήματα που προκύπτουν από τη χρήση e-learning:

- κατάρριψη περιορισμών χώρου (αν και πρέπει ο εκπαιδευόμενος να βρίσκεται σε χώρο με Η/Υ), χρόνου και ρυθμού
- υποστήριξη αναγκών ΑΜΕΑ
- εφαρμογή μαθητοκεντρικής μεθόδου
- δυνατότητα προσαρμογής με βάση τις ανάγκες και το επίπεδο του εκπαιδευόμενου
- παρότρυνση για ενεργητική συμμετοχή
- καλλιέργεια συνεργατικής μάθησης
- πρόσβαση μέσω διαδικτύου σε πλούσιο πληροφοριακό υλικό πάντα και παντού
- δυνατότητα παρακολούθησης διαλέξεων ειδικών
- εξοικείωση με τη χρήση των πιο σύγχρονων εφαρμογών των ΝΤ

- δυνατότητα στον εκπαιδευτικό εμπλουτισμού του μαθησιακού υλικού και χρήσης
- εφαρμογών NT, που το κάνουν πιο ελκυστικό, αλλά και επαναχρησιμοποίησης του
- δυνατότητα δημιουργίας βάσεων δεδομένων εφόσον το υλικό είναι διαθέσιμο στο διαδίκτυο
- ευκολότερη παρακολούθηση της προόδου των εκπαιδευόμενων και σωστότερη αξιολόγηση τους
- δυνατότητα συνεχούς βελτίωσης του περιεχομένου και της αποτελεσματικότητας
- εξοικονόμηση πόρων και μείωση κόστους για όλους τους συμμετέχοντες στη διαδικασία (εκπαιδευόμενους, εκπαιδευτές, φορείς)

#### 1.4.2 Μειονεκτήματα e-learning

Στα μειονεκτήματα της συγκεκριμένης μεθόδου οι έρευνες εντόπισαν τα ακόλουθα:

- μείωση της προσωπικής επικοινωνίας και επαφής
- προβλήματα από την έλλειψη εξοικείωσης με τις NT
- αυξημένες υποχρεώσεις του εκπαιδευτή/επιμορφωτή
- σπατάλη χρόνου για τη δημιουργία εικονικού περιβάλλοντος
- δυσκολίες στη χρήση και κόστος εξοπλισμού, συντήρησής του και πιθανής τεχνικής υποστήριξης
- άμεση εξάρτηση από την ομαλή λειτουργία του υπολογιστικού περιβάλλοντος και των τηλεπικοινωνιακών τεχνολογιών
- προβλήματα κατοχύρωσης πνευματικών δικαιωμάτων του εκπαιδευτικού υλικού
- τυποποίηση γνώσης

Σε ιδανικές περιπτώσεις συνιστάται για την άρση κάποιων από τις ανωτέρω δυσχέρειες να υπάρχει μια υπηρεσία σε θέση «τηλεμέντορα», που θα λειτουργεί εξ αποστάσεως μέσω διαδικτύου, παρέχοντας υποστήριξη και καθοδήγηση σε θέματα διδακτικά, πρακτικά (π.χ. εξεύρεση πηγών για έρευνα), μεθοδολογικά (π.χ. τρόπος

σχεδιασμού έρευνας), ή και συναφή με περαιτέρω επαγγελματική δραστηριότητα (π.χ. εύρεση θέσεων εργασίας) (Σαβρανίδης κ.α, 2003· Φειδάς κ.α., 2005)

## **1.5. Που μπορεί να χρησιμοποιηθεί η τηλεκαίδευση;**

### **1.5.1 Εφαρμογές e-learning στην Τριτοβάθμια εκπαίδευση**

Το e-learning δύναται να χρησιμοποιηθεί καταρχήν για ακαδημαϊκούς σκοπούς από τα ιδρύματα τριτοβάθμιας εκπαίδευσης ως εναλλακτικός τρόπος διδασκαλίας/μάθησης, κάνοντας πλέον πραγματικότητα τις σπουδές από το σπίτι και την απόκτηση ηλεκτρονικού πτυχίου ακόμα και από διακεκριμένα πανεπιστήμια (π.χ. Χάρβαρντ, Γέιλ, Πρίνστον, Κορνέλ, Κολούμπια, Στάνφορντ κτλ.). Ως κίνηση ξεκινά από τις ΗΠΑ, όπου μάλιστα στα τέλη του προηγούμενου αιώνα ιδρύονται τα πρώτα εικονικά πανεπιστήμια, τα οποία λειτουργούν εξ ολοκλήρου on-line (Παναγιωτακόπουλος, 2008). Στις μέρες μας όμως δεν είναι λίγα και τα ελληνικά ΑΕΙ (π.χ. ΕΜΠ, ΕΚΠΑ, ΑΣΟΟΕ, Παν/μιο Πατρών, Ιωαννίνων, Κρήτης κτλ.) που προσφέρουν δυνατότητα για σπουδές (συνήθως μεταπτυχιακές) με τη συγκεκριμένη μέθοδο ή και με συνδυασμό με την παραδοσιακή (όπως για παράδειγμα η απόπειρα του ΑΠΘ για δημιουργία «Πανεπιστημίου υβριδικού τύπου») (Γιάνναρου & Λακασάς). Πρόσφατα μάλιστα, στην 60η σύνοδο των πρυτάνεων προτάθηκε η δημιουργία μιας «ηλεκτρονικής τάξης» προσβάσιμης μέσω διαδικτύου για κάθε διδασκόμενο μάθημα στα ελληνικά πανεπιστήμια με στόχο την ενίσχυση της παραδοσιακής «διδασκαλίας αμφιθεάτρου» και την πλήρη αξιοποίηση των δυνατοτήτων των ΝΤ στην εκπαιδευτική διαδικασία (ΤΟ ΒΗΜΑ, 22/3/09).

Από τη χρήση του e-learning σε τριτοβάθμια εκπαιδευτικά ιδρύματα απορρέουν πολλά πλεονεκτήματα, όπως το ότι:

- α. προσεγγίζεται ένα ευρύτερο κοινό φοιτητών,
- β. ανταποκρίνεται στις ανάγκες των φοιτητών που δεν μπορούν να παρακολουθήσουν στο χώρο του πανεπιστημίου,
- γ. γίνεται εφικτή η εμπλοκή εξωτερικών ομιλητών,

- δ. διασυνδέονται φοιτητές από διαφορετικό κοινωνικό, πολιτιστικό, οικονομικό και κοινωνικό περιβάλλον,
- ε. επιτυγχάνεται η επικοινωνιακή χρήση της δικτυακής υποδομής του ιδρύματος και η αναβάθμιση των σπουδών του, στ. δίνεται δυνατότητα αξιοποίησης πολυμορφικού εκπαιδευτικού υλικού κτλ. (Μπαλτά, 2005· Myroni,2007)

### 1.5.2 Εφαρμογές E-learning στην πρωτοβάθμια και δευτεροβάθμια

Η εφαρμογή αυτή όμως τα τελευταία χρόνια αρχίζει να θεωρείται κατάλληλη προς χρήση και στις άλλες βαθμίδες της εκπαίδευσης, όπου ούτως ή άλλως αυξάνονται πλέον συνεχώς οι εκπαιδευτικές δραστηριότητες και δράσεις που αξιοποιούν τις ΝΤ, ή το διαδίκτυο ειδικότερα, για το διανθισμό της παραδοσιακής διδασκαλίας στη μαθητική κοινότητα στο σύνολο της, αλλά ιδιαίτερα σε μαθητές από περιφερειακές γεωγραφικές περιοχές (Αποστολάκη & Βλάχος, 1999). Είναι γεγονός ότι έχει αποδειχθεί πως, εφόσον κατά το σχεδιασμό των νέων αυτών περιβαλλόντων μάθησης ακολουθηθούν οι αρχές «ευχρηστίας» και «ευμάθειας» (Ζηβελίδης, 2003· Καρούλης, 2002), ήτοι προσεχθεί η μαθησιακή υποστήριξη (πόση εξήγηση μαθησιακού υλικού χρειάζεται ο μαθητής) και δοθεί βάρος στα προβλήματα κατανόησης (ποια μέρη του υλικού τον δυσκολεύουν να καταλάβει) τότε δύνανται να προσελκύουν το ενδιαφέρον, ευνοούν τη συμμετοχή των μαθητών και την καλλιέργεια συνεργατικού πνεύματος, οδηγούν στη βελτίωση της απόδοσης κτλ. Έχουν ήδη σχεδιασθεί ανάλογες πλατφόρμες για να επικουρήσουν τη διαδικασία (όπως π.χ. η e-Cosmos, Δαμιανάκης & Μάλαμος, 2002). Παράλληλα πληθαίνουν οι εφαρμογές ακόμα και σε μαθητές του Δημοτικού (Καρασαββίδης, 2003· Τριανταφυλλίδης & Κοφτερός, 2008) ενώ γίνονται εφικτές και σύνθετες διδακτικές παρεμβάσεις (όπως για παράδειγμα η εφαρμογή «Οδυσσέας», η επιστράτευση τηλεδιάσκεψης σε σχολεία Ελλάδας και Ομογένειας ή η δημιουργία μαθησιακού περιβάλλοντος που συνδυάζει τηλεδιδασκαλία και τηλεσυνεργασία σε σχολεία της Κύπρου, Anastasiades, 2003· Αναστασιάδης, 2007· Χαμπιαούρης κ.α., 2009).



Οι μαθητές δείχνουν ενδιαφέρον αλλά δεν έχουν πάντα την απαραίτητη αυτοπειθαρχία, δυσκολεύονται να απαντήσουν σε ερωτήσεις με e-mail κτλ. (Πέρδος κ.α., 2004). Βέβαια από την άλλη εκμεταλλεύονται το γεγονός ότι η ώρα δεν έχει τον περιορισμό των 45', μπορούν να σκεφτούν πριν απαντήσουν, να γνωρίσουν τις απαιτήσεις των άλλων και να τις δουν κριτικά, να κατανοήσουν την ύλη μέσα από τις απορίες των άλλων αλλά και να μελετήσουν κατ' ιδίαν αναπτύσσοντας πρωτοβουλία (Μάτος & Πούλιος, 2003· Πανούσης, 2007· Φεσάκης & Μαυρουδή, 2007).

Για να μπορέσουν, όμως, οι εκπαιδευτικοί να ανταποκριθούν στις απαιτήσεις αυτές χρειάζονται επιμόρφωση-κατάρτιση σε τεχνικό και παιδαγωγικό επίπεδο, την οποία αρκετές φορές απαιτούν οι ίδιοι (Δαπόντες & Τζιμόπουλος, 2001· Ιωάννου, 2004· Χλαπάνης & Δημητρακοπούλου, 2001).

### 1.5.3 Επιμόρφωση-Κατάρτιση και e-learning

Τοιουτοτρόπως καταλήγουμε στον επόμενο χώρο όπου επεκτείνεται η χρήση του elearning, ήτοι τον επαγγελματικό/επιχειρησιακό χώρο για την κατάρτιση των εργαζομένων. Στα σύγχρονα εργασιακά δεδομένα, ο κάθε εργαζόμενος οφείλει να επιδιώκει την δια βίου επιμόρφωση/κατάρτισή του, είτε στο χώρο της εργασίας του είτε σε άλλους φορείς, πάνω στο αντικείμενο του αλλά και σε γενικότερα (π.χ.εξοικείωση με NT), στοχεύοντας στην βελτίωση των γνώσεων του, την κοινωνική και οικονομική του αναβάθμιση, την αποφυγή του αποκλεισμού από την παραγωγική διαδικασία (Παγγέ, 2002· Παπαδάκης & Φραγκούλης, 2005). Από την πλευρά της πολιτείας ή της εργοδοσίας πρέπει να γίνεται απόπειρα εφαρμογής στο σύνολο της ιεραρχίας λαμβάνοντας υπόψη τις εξελίξεις στο χώρο της Διαχείρισης Γνώσης αλλά και το τεχνολογικό επίπεδο (Paradimitriou, 2007).

Παράγοντες που επηρεάζουν την αποτελεσματικότητα μιας επιμόρφωσης σχετίζονται με τον ατομικό προφίλ του κάθε εκπαιδευόμενου (ενδιαφέρον για το θέμα, κίνητρα,προδιάθεση, στάσεις) αλλά και με τις διδακτικές παραμέτρους (μέθοδος διδασκαλίας, στυλ συμπεριφοράς επιμορφωτή, είδος και ποικιλία χρησιμοποιούμενων διδακτικών τεχνικών και εκπαιδευτικών μεθόδων) (Ματσαγγούρας, 1998). Ιδιαίτερα

στον τομέα της επιμόρφωσης εκπαιδευτικών έχουν γίνει διάφορες απόπειρες χρήσης αρχικά της εξ αποστάσεως εκπαίδευσης (Αβραάμ & Μαυροειδής, 2001) και στη συνέχεια του elearning τόσο στην Ελλάδα (π.χ. επιμόρφωση στη μέθοδο project μέσω τηλεδιάσκεψης με τη χρήση του Πανελληνίου Σχολικού Δικτύου, Παπαδημητρίου, 2007) όσο και στο εξωτερικό (με την ύπαρξη και παραδειγμάτων διεθνών συνεργασιών π.χ. Αγγλίας-Φιλανδίας, Esko, 2005) με ιδιαίτερα θετικά αποτελέσματα (Deprez et al, 2005· Deprez, 2005· Paola et al, 2007· Roman & Arnost, 2005). Διαπιστώθηκε ότι για να είναι αποδοτική η επιμόρφωση με οποιαδήποτε διδακτική μέθοδο χρησιμοποιεί, θα πρέπει να συγκεντρώνει τουλάχιστον τα εξής χαρακτηριστικά (Κακαβάκης, 2005· Μαυροειδής & Φλωρίδης, 2005):

α. να αφορά

διαπιστωμένες ανάγκες,

β. να έχει μικρή διάρκεια και δυνατότητα επανάληψης,

γ. να είναι άμεσα σχετισμένη με την εκπαιδευτική έρευνα και τη σχολική πράξη,

δ. να προσφέρει ποικιλία επιμορφωτικών δραστηριοτήτων,

ε. να απευθύνεται σε μεγάλο αριθμό εκπαιδευτικών, στ. να εφαρμόζει σύγχρονες μεθόδους εκπαίδευσης ενηλίκων,

ζ. να εκμεταλλεύεται τις δυνατότητες ΝΤ, η. να είναι προσανατολισμένη στο μέλλον,

θ. να επιδέχεται συνεχή αξιολόγηση και βελτίωση.

Συνεπώς η πολιτεία αλλά και ιδιωτικοί φορείς βρίσκονται μπροστά σε μια πρόκληση: να σχεδιάσουν επιμορφωτικά προγράμματα εκμεταλλευόμενοι στο έπακρο τις δυνατότητες του e-learning προκειμένου να προετοιμάσουν επαρκώς τους εκπαιδευτικούς, αλλά και τους άλλους εργαζομένους, για τις απαιτήσεις του μέλλοντος.

## Κεφάλαιο 2

# Η διαδικασία Αναγνώρισης Προτύπων (stemming) και οι υπάρχουσες τεχνικές

### 2.1 Εισαγωγή

Στις μέρες μας, υπάρχουν πολλά εργαλεία που μας βοηθούν για την ανάκτηση πληροφοριών. Υπάρχουν ορισμένες ενδιαφέρουσες κατηγορίες των διαθέσιμων λογισμικών ανάκτησης πληροφοριών. Μπορούμε να βρούμε μια ποικιλία μηχανών αναζήτησης με εξελιγμένες παραμέτρους αναζήτησης, εξειδικευμένες μηχανές αναζήτησης για ανάκτηση εγγράφων μέσα από μια συλλογή εγγράφων, εξόρυξη δεδομένων και εργαλεία συσταδοποίησης καθώς και εργαλεία ταξινόμησης. Κατά την ανάπτυξή τους παρατηρούμε μια συνεχή εξειδίκευση στα χαρακτηριστικά της αναζήτησης. Οι μηχανές αυτές εξελίσσονται ολοένα και περισσότερο προσπαθώντας να καλύψουν τις ανάγκες του χρήστη για πρόσβαση σε συγκεκριμένες πληροφορίες.

Στη προσπάθεια να γίνουν πιο αποτελεσματικές οι μηχανές αναζήτησης ως προς την ανάκτηση πληροφορίας χρησιμοποιήθηκε η τεχνική Αναγνώρισης Προτύπου Λέξης (word stemming). Ο Lovins (1968) ορίζει έναν αλγόριθμο αναγνώρισης προτύπου ως «μια διαδικασία με στόχο τη μείωση όλων των λέξεων με την ίδια ρίζα σε μια κοινή μορφή, συνήθως αναλύοντας ετυμολογικά κάθε λέξη». Ο κύριος στόχος της διαδικασίας αυτής είναι να απομακρύνει όλα τα πιθανά παράγωγα και να μειώσει τη λέξη στη ρίζα του ( Dawson, 1974). Χρησιμοποιώντας την προαναφερθείσα διαδικασία πολλές σύγχρονες μηχανές αναζήτησης συνδέουν τις σύνθετες λέξεις με τη ρίζα της λέξης προκειμένου να διευρύνει την αναζήτηση υπό την έννοια ότι μπορεί να εξασφαλίσει το μέγιστο αριθμό των σχετικών ταιριασμάτων που περιλαμβάνονται στα αποτελέσματα της αναζήτησης. Η διαδικασία Αναγνώρισης Προτύπων (stemming) βρίσκει εφαρμογή στα εργαλεία μετάφρασης, στη περίληψη εγγράφων (Orasan, Pekar & Hasler 2004, Dalianis 2000) και στη ταξινόμηση κειμένου (Gaustad & Bouma 2002).

## 2.2 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Έχουν αναπτυχθεί πολλές θεωρίες και πειράματα για να αξιολογήσουν την αποτελεσματικότητα και τη σταθερότητα της διαδικασίας stemming στην ανάκτηση πληροφορίας. Ο Lennon (1981) εκπόνησε μια μελέτη αξιολόγησης για τις τεχνικές stemming και πώς αυτές επηρεάζουν την ακρίβεια της αναζήτησης, τονίζοντας ότι η διαδικασία σαν κίνητρο για πολλούς ερευνητές να ασχοληθούν με τη βελτίωση της διαδικασίας stemming.

Υπάρχουν πολλές τεχνικές οι οποίες αναπτύχθηκαν με το πέρασμα του χρόνου και χρησιμοποιούνται για τη διαδικασία Αναγνώρισης προτύπου λέξης (word stemming). Από τη πρώτη βασική προσέγγιση Dictionary-Based μέχρι και τη τελευταία Corpus-Based τεχνική, οι ερευνητές χρησιμοποιούν εναλλακτικούς κανόνες και σχηματισμούς για κάθε γλώσσα έτσι ώστε να αναπτύξουν έναν αξιόπιστο stemmer με πολλή υψηλή ακρίβεια.

## 2.3 Dictionary-Based Τεχνική

«Ιστορικά, οι περισσότεροι stemmers χαρακτηρίζονται είτε ως dictionary-based είτε ως αλγοριθμικοί» (Porter 2001). Οι dictionary-based stemmers ταιριάζουν κάθε λέξη με μια λέξη από ένα ορθό ψηφιακό λεξικό και αντιστοιχούν κάθε λέξη με τη ρίζα του (Carlberger et al. 2001). Στα πειράματα του Krovetz (1995) αυτή η απευθείας μέθοδος φάνηκε αποτελεσματική αλλά ανεπαρκής ως προς τις ξενόφερτες λέξεις οι οποίες και έχουν πιο περίπλοκο σχηματισμό. Αυτός ήταν και ο κύριος λόγος που τον οδήγησε να αξιολογήσει τους αλγοριθμικούς stemmers και να καταλήξει ότι «παρόλο τα λάθη που φαίνεται να κάνει, παράγει καλά πρακτικά αποτελέσματα». Επιπλέον, «οι dictionary-based stemmers απαιτούν συνεχή ανανέωση των λεξικών κάτι που αποτελεί πρόβλημα. «Ένα λεξικό το οποίο δημιουργήθηκε προκειμένου να υποστηρίξει έναν αλγόριθμο stemming προφανώς θα απαιτεί τεράστια ενημέρωση σε μερικά χρόνια. Επίσης ένα λεξικό το οποίο ήδη χρησιμοποιείται μάλλον θα είναι ήδη απαρχαιωμένο».

## 2.4 Rule-Based Τεχνική

Μία από τις πιο διαδεδομένες τεχνικής αναγνώρισης προτύπου με κύριο αντιπρόσωπό της τον αλγόριθμο ο οποίος παρουσιάστηκε από τον Porter (1980). Ο αλγόριθμος αφαιρεί επαναληπτικά τις καταλήξεις από μια λέξη μέχρι να εντοπίσει τη ρίζα της λέξης , χρησιμοποιώντας συγκεκριμένους κανόνες για την Αγγλική γλώσσα. Ακόμα και αν ο αλγόριθμος έχει κάποιους περιορισμούς , παρουσιάζει υψηλή ακρίβεια και υψηλό ποσοστό σχετικών αναφορών στην αναζήτησή μας. Ο stemmer του Lovin (1968) ακολουθεί την ίδια τεχνική αλλά δεν εφαρμόζει επαναληπτικά τους κανόνες και είναι αρκετά πιο συντηρητικός από τον αλγόριθμο του Porter. Στο ίδιο μονοπάτι βάδισαν και οι Paice , Husk παρουσιάζοντας έναν ακόμα Αγγλικό stemmer με διαφορετικούς κανόνες..Για τη Σκανδιναβική γλώσσα παρουσιάστηκαν rule-based stemmers το 2001 (Dalianis & Jongejan 2006). Τέλος, εφαρμόζοντας την ίδια φιλοσοφία υπάρχουν rule-based stemmers για τις παρακάτω γλώσσες :

- Αγγλικά
- Γαλλικά
- Ισπανικά
- Πορτογαλικά
- Ιταλικά
- Γερμανικά
- Ολλανδικά
- Σουηδικά
- Νορβηγικά
- Δανέζικα
- Ρώσικα
- Φινλανδικά

Οι παραπάνω stemmers και οι αλγόριθμοί τους μπορούν να βρεθούν στον ιστόχωρο “Tartarus” (<http://snowball.tartarus.org>) και ακολουθούν την SNOBOL (StriNg Oriented symBOLic Language) , μια γλώσσα που σχεδιάστηκε για τη δημιουργία stemming αλγόριθμους και χρησιμοποιεί strings ως σύμβολα . Οι stemmers οι οποίοι σχεδιάστηκαν με SNOBOL ονομάζονται «Snowball».

## 2.5 Light-Stemming Τεχνική

Σήμερα, υπάρχουν πολλοί rule-based αλγόριθμοι, δηλαδή αλγόριθμοι που η υλοποίησή τους βασίζεται σε κανόνες, και stemmers για διάφορες γλώσσες. Πολλές φορές χρησιμοποιείται διαφορετικός αλγόριθμος για κάθε γλώσσα προκειμένου να έχουμε μεγαλύτερη ακρίβεια στα αποτελέσματα. Για το λόγο αυτό έχουμε τους light-stemmers οι οποίοι αναφέρονται στη διαδικασία κατά την οποία αφαιρείται ένα μικρό σύνολο προθεμάτων και/ή καταλήξεων χωρίς να χρειαστεί να ασχοληθεί με κλίσεις ή να αναγνωρίσει πρότυπα και να βρεί τις ρίζες (Sughaiyer & Kharashi 2004).

## 2.6 Corpus-Based Τεχνική

Σύμφωνα με τον Porter (2001) οι αλγοριθμικοί και οι dictionary-based stemmers δεν είναι σαφή. Ένας αλγοριθμικός stemmer χρησιμοποιεί λίστες λέξεων έτσι ώστε να αφαιρεί τις καταλήξεις ή να τις εξαιρεί. Όσο πιο εξελιγμένος είναι ένας αλγόριθμος τόσο μεγαλύτερες είναι και οι λίστες. Αντιθέτως, ένας dictionary-based stemmer πρέπει αρχικά να αφαιρέσει μερικές βασικές καταλήξεις πριν ξεκινήσει τη διαδικασία αναζήτησης στο λεξικό. Προσπαθώντας να βελτιώσουμε την αποτελεσματικότητα αυτών των stemmers οδηγηθήκαμε στη rule-based τεχνική.

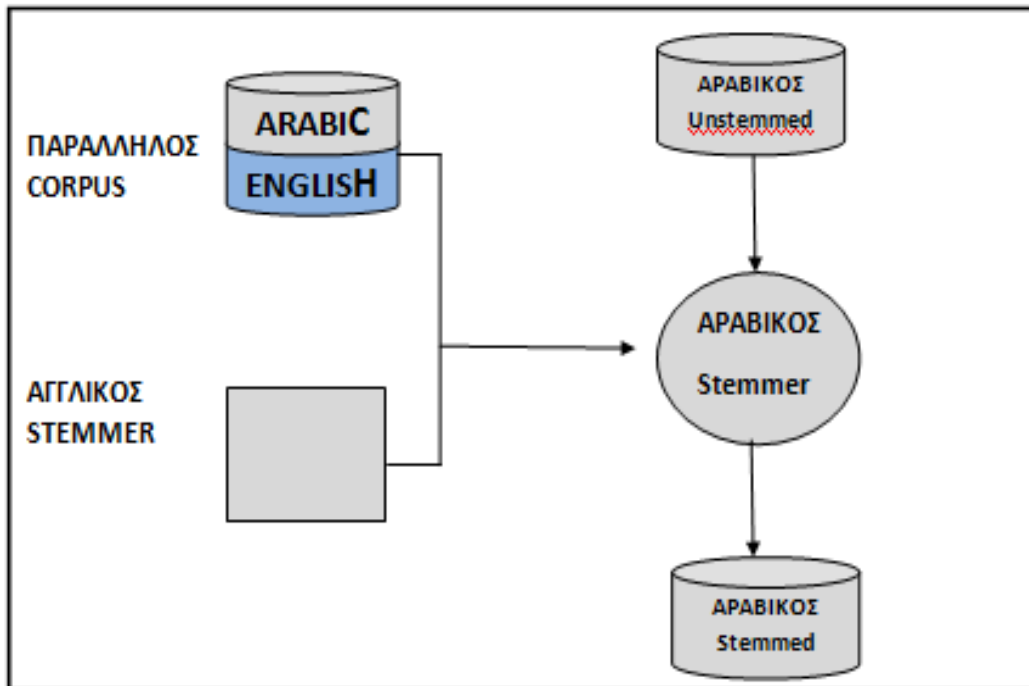
Νωρίτερα, η υβριδική αυτή άποψη εφαρμόστηκε σε πολλούς corpus-stemming αλγόριθμους με κύριους αντιπροσώπους τους Xu και Croft (1998). Η υπόθεση του έργου αυτού είναι ότι οι μορφές των λέξεων που θα πρέπει να εξομοιώνονται θα συν-εμφανίζονται στα έγγραφα του συνόλου. Ξεκινά με ένα σύνολο ακατέργαστων προκαταρκτικών κατηγοριών θεμάτων που έχουν δημιουργηθεί από έναν άλλον stemmer, ίσως τον Porter ή κάποιον άλλον που θα εξομοιώνει όλες τις λέξεις που αρχίζουν με τα ίδια τρία γράμματα. Στη συνέχεια, χρησιμοποιεί την ανάλυση συνύπαρξης της λέξης στην αρχική κλάση για να βρεί όλες εκείνες τις λέξεις που δε φαίνεται να ανήκουν μαζί. Οι corpus-based stemmers αποδείχτηκαν να παρέχουν μέτρια βελτίωση σε σχέση με τους ήδη υπάρχοντες rule-based stemmers. Όπως αναφέρουν στην έρευνά τους «η βασική ιδέα πίσω από αυτό το έργο είναι ότι μπορούμε να

χρησιμοποιήσουμε την ανάλυση συνύπαρξης των παραλλαγών μιας λέξης μέσα σε ένα συγκεκριμένο σύνολο προκειμένου να είναι σε θέση να γνωρίζει ποιες παραλλαγές ανήκουν στην ίδια ομάδα και ποιές όχι, ενώ ο stemmer του Porter δημιουργεί κλάσεις ανάλογα τη ρίζα της λέξης».

## 2.7 Stemming για μη αγγλικές γλώσσες

Για τις γλώσσες πλην της αγγλικής, υπάρχουν stemmers με διαφορετικούς κανόνες εφαρμογής για κάθε γλώσσα. Ένας πολύ γνωστός stemmer με υψηλή ακρίβεια και υψηλό ποσοστό συναφών αποτελεσμάτων (recall) είναι αυτός του Slovene (Porovic & Wilett 1992) ενώ ο Savoy (1993) παρουσίασε ένα διαφορετικό stemmer για τη γαλλική γλώσσα. Για τις σκανδιναβικές γλώσσες υπάρχει μια σύγκριση μεταξύ των CST (Center for Language Technology) και Euroling stemmers, τονίζοντας την εξέλιξη των αλγόριθμων stemming και των απαιτήσεων που προκύπτουν στο πεδίο της Ανάκτησης Πληροφοριών (Dalianis & Kongejan 1006)

Το να σχεδιάζεις έναν stemmer βασισμένο σε κανόνες για μια καινούρια, αυθαίρετη γλώσσα είναι μια διαδικασία πολύ χρονοβόρα και επιπλέον απαιτεί ειδικούς με γλωσσολογικές γνώσεις για τη συγκεκριμένη γλώσσα (Rogati et al. 2003). Για τη δημιουργία ενός καινούριου stemmer για την αραβική γλώσσα, χρησιμοποιείται μια τεχνική ενός παράλληλου συνόλου στο οποίο σύνολο εφαρμόζονται οι ήδη γνωστοί αλγόριθμοι για την αγγλική γλώσσα. Το παράλληλο σύνολο αποτελείται από μια συλλογή από ζευγάρια προτάσεων με την ίδια σημασία αλλά σε διαφορετικές γλώσσες (Rogati 2003). Συνήθως μεταφράζουμε ολόκληρα έγγραφα και τα ζευγάρια των προτάσεων ακολούθως αντικαθίστανται αυτόματα με τις έννοιές τους. Ένα μικρό παράλληλο σύνολο είναι διαθέσιμο όταν οι μεταφραστές δεν είναι διαθέσιμοι κάτι το οποίο είναι προτιμότερο για να κατασκευάσουμε έναν stemmer χρησιμοποιώντας ένα τέτοιο σύνολο. Στο Σχήμα 1 δίνεται μια γενική περιγραφή της παράλληλης corpus μεθόδου, η οποία χρησιμοποιείται από έναν Αραβικό stemmer.



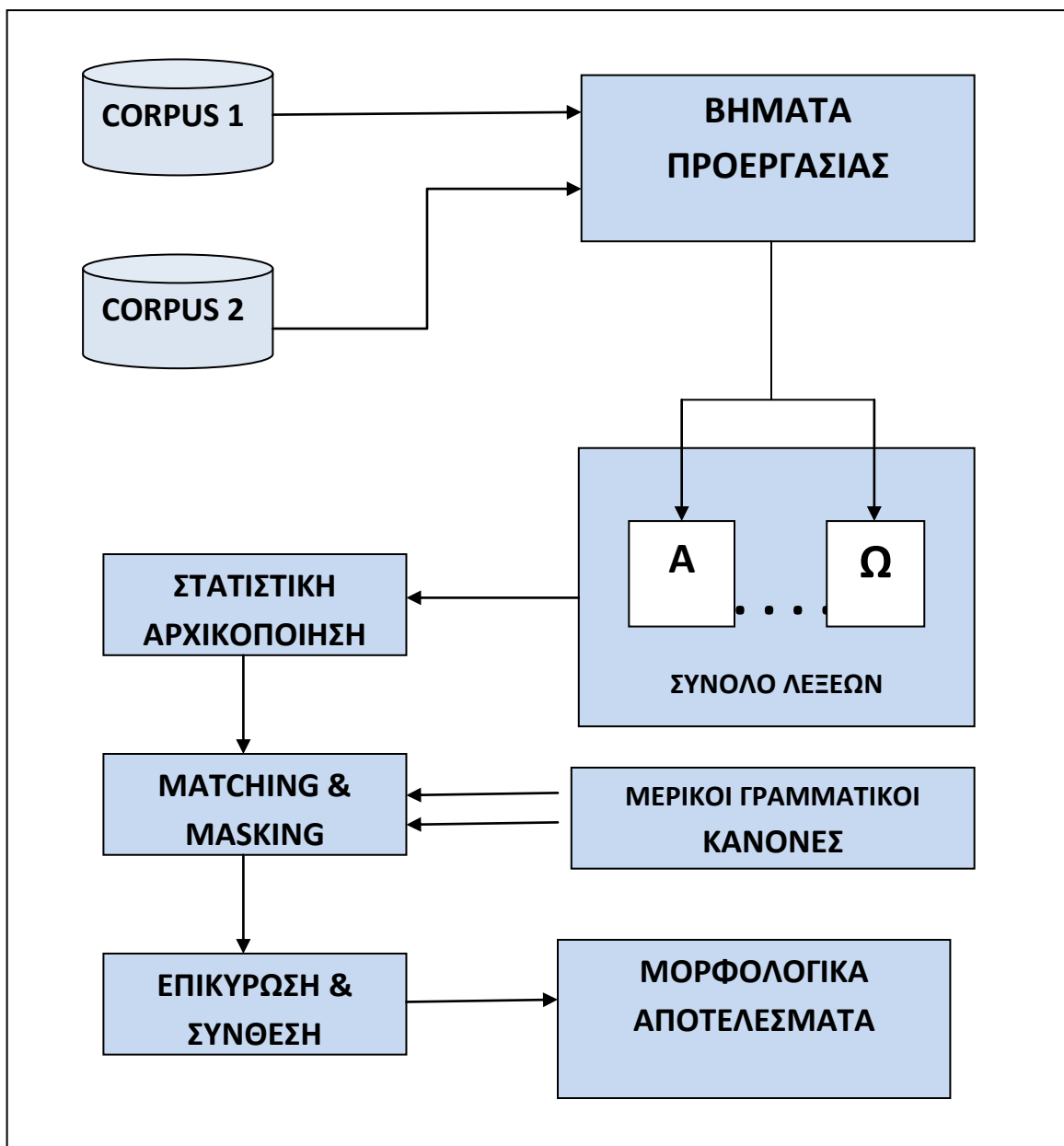
**ΣΧΗΜΑ 1.**

Ένας παράλληλος corpus stemmer είναι ανεξάρτητος από τη γλώσσα και έχει χρησιμοποιηθεί επιτυχώς και από άλλους ερευνητές (Yarowsky 2000, Diab & Resnik 2002)

## 2.8 Ελληνικοί Stemmers

Το 2001, οι Καραγιάννης και Ταμπουρατζής παρουσίασαν για το Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ) της Ελλάδας, ένα σύστημα το οποίο αυτόματα κατηγοριοποιεί μορφολογικά τις ελληνικές λέξεις που προέρχονται από ένα θέμα. Ο σκοπός αυτού του Αυτοποιημένου Μορφολογικού Επεξεργαστή, του οποίου η δομή παρουσιάζεται στο σχήμα 2, είναι να εφαρμόσει τμηματοποίηση στο σύνολο ως προς το θέμα και τις καταλήξεις με αυτοματοποιημένο τρόπο. Ο αλγόριθμος χρησιμοποιεί επαναληπτική matching-and-masking προσέγγιση βασισμένη σε κανόνες, και βασίζεται στο ταίριασμα των κομματιών διαφορετικών προτύπων. Εδώ, η διαδικασία βασίζεται σε ένα αρχικό σύνολο θεμάτων και καταλήξεων. Υπάρχει επίσης μια αξίωση σύμφωνα με την οποία κάθε λέξη αποτελείται από ένα θέμα και μια κατάληξη εξαιρώντας τις σύνθετες λέξεις.





**ΣΧΗΜΑ 2.**

Ακόμα και αν το σύστημα δεν είναι αμιγώς βασισμένο σε κανόνες, είναι επιτυχές ως προς τη διαδικασία stemming των ελληνικών λέξεων με ακρίβεια σχεδόν 95%. Μπορεί να διαχωρίσει την κατάληξη από το θέμα για μια δοσμένη λέξη και η συμπεριφορά του εξαρτάται από το κατά πόσο το λεξιλογικό θέμα και η κατάληξη είναι περίπλοκες. Ακόμα και έτσι όμως, είναι πολύ λίγοι οι γραμματικοί κανόνες οι οποίοι ακολουθούν την matching-and-maskng

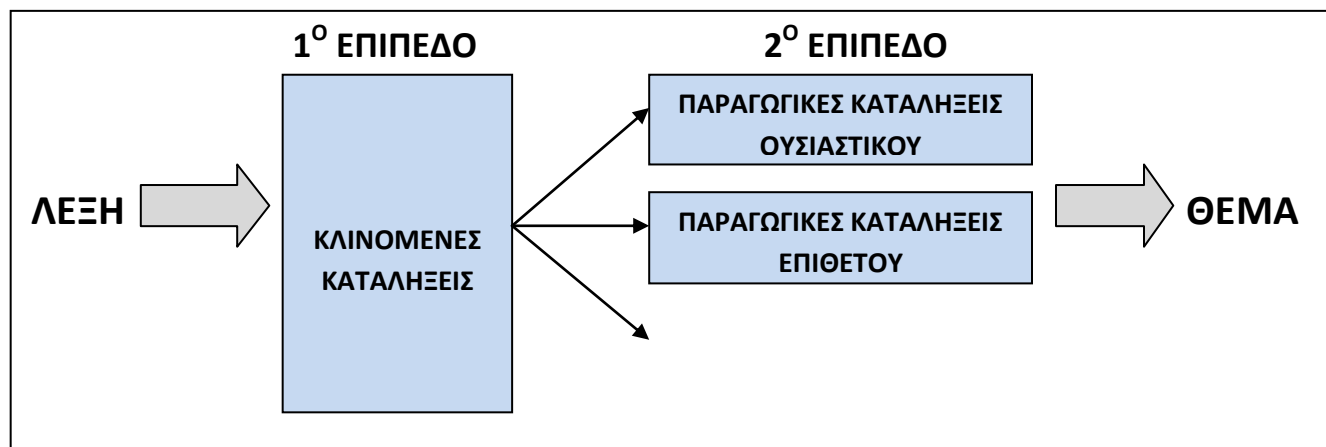
διαδικασία για να θεωρήσουμε ότι αναπτύξαμε έναν πλήρη stemming αλγόριθμο για την Ελληνική γλώσσα.

Ο Αυτοποιημένος Μορφολογικός Επεξεργαστής δεν σταματά στη διαδικασία matching-and-masking. Συνεχίζει στη σύνθεση των αποτελεσμάτων από τη διαδικασία stemming και μετά από 4 βήματα επιστρέφει έναν αριθμό πιθανών αποτελεσμάτων. Για να επιλέξει το αποτέλεσμα εκείνο που αντιπροσωπεύει ορθότερα το τμήμα, χρησιμοποιεί ένα κριτήριο κατάταξης το οποίο με τη σειρά του χρησιμοποιεί τον υπάρχων λεξικό ILSP προκειμένου να κάνει συγκρίσεις. Περισσότερες πληροφορίες για το σύστημα και την αξιολόγησή του παρουσιάζονται στη σχετική ερευνητική μελέτη (Ταμπουρατζής & Καραγιάννης, 2001).

## 2.9 Suffix stripping αλγόριθμος

Οι Καλαμπούκης & Νικολαΐδης παρουσίασαν για το Ερευνητικό Κέντρο του Οικονομικού Πανεπιστημίου Αθηνών μια ευθέως εργασία του Ελληνικού Stemmer.

Το 1995 οι προαναφερθέντες δημοσίευσαν τον πρώτο αλγόριθμο αφαίρεσης καταλήξεων για την Ελληνική γλώσσα. Ο αλγόριθμος αυτός σχεδιάστηκε για την ανάκτηση πληροφορίας από Ελληνικά κείμενα και ασχολείται με κλίσεις και παράγωγα της Ελληνικής γλώσσας. Χρησιμοποιούν μια λίστα καταλήξεων και έχουν εφαρμόσει έναν επαναληπτικό αλγόριθμο καταλήξεων με δύο επίπεδα. Το πρώτο επίπεδο αφορά την ανάλυση των κλίσεων ενώ το δεύτερο ασχολείται με την αφαίρεση των παραγωγικών καταλήξεων σύμφωνα με τις κατηγορίες της γραμματικής. Έχουν δημιουργήσει τρεις διαφορετικούς πίνακες καταλήξεων που αντιστοιχούν στις τρεις βασικές κατηγορίες : ουσιαστικό, επίθετο και ρήμα. Οι καταλήξεις ελέγχονται σύμφωνα με την κατηγορία και απομακρύνονται βάσει του πίνακα καταλήξεων σε δύο βήματα. Στο Σχήμα 3 δίνεται μια γενική επισκόπηση του συστήματος.



### ΣΧΗΜΑ 3. “TZK Αλγόριθμος”

Ο “TZK αλγόριθμος”, όπως αναφέρεται στη σχετική μελέτη, απομακρύνει 65 καταλήξεις και στα δύο επίπεδα και αυτό διότι όπως παραδέχτηκαν «συμπεριέλαβαν μόνο ένα μικρό σύνολο καταλήξεων διότι είχαν φτάσει σε ένα σημείο όπου όταν προσέθεταν περισσότερους κανόνες για να βελτιώσουν την απόδοσή του σε ένα σημείο, υποβαθμιζόταν η απόδοσή του σε κάποιο άλλο σημείο. Το γεγονός αυτό αποτελεί και το βασικό περιορισμό του αλγορίθμου εφόσον στην Ελληνική γλώσσα υπάρχουν τουλάχιστον 166 διαφορετικές κλινόμενες καταλήξεις (Τριανταφυλλίδης, 1941). Επιπροσθέτως, ο αλγόριθμος δουλεύει μόνο με κεφαλαία γράμματα για να ξεπεράσει το πρόβλημα του τονισμού το οποίο επηρεάζει την σημασία και την ορθογραφία της λέξης.

Ένα άλλο κρίσιμο σημείο του αλγορίθμου είναι το γεγονός ότι η απομάκρυνση της παραγωγικής κατάληξης βασίζεται στο τι μέρος του λόγου είναι η λέξη. Αφού όμως δεν υπάρχει κανένα εργαλείο μορφολογικής ανάλυσης, ο πίνακας καταλήξεων στο 1ο επίπεδο αφαίρεσης δεν είναι αρκετός να διαχωρίσει αν μια λέξη είναι ουσιαστικό, επίθετο ή ρήμα.

Παρ’ όλα τ’ αυτά, ο αλγόριθμος φαίνεται να είναι αποδεκτός για τα ελληνικά κείμενα. Σύμφωνα με τη πρώτη αξιολόγηση το 1995, ο αλγόριθμος εφαρμόστηκε σε δύο σύνολα εγγράφων, ένα το οποίο αφορούσε την ιατρική επιστήμη και ένα το οποίο αφορούσε την επιστήμη των υπολογιστών με 7959 διαφορετικές λέξεις συνολικά. Το ποσοστό του λάθους ανήλθε στο 10% με ικανοποιητική ακρίβεια και ποσοστό εύρεσης σχετικών εγγράφων.

Το 1999 οι ίδιοι ερευνητές ,Καλαμπουκής & Νικολαΐδης, αξιολόγησαν τους αλγόριθμους stemming με Νέα Ελληνικά χρησιμοποιώντας μια διαφορετική προσέγγιση. Χρησιμοποίησαν το σύστημα Αποθήκευσης και Ανάκτησης (SMART, STORAGE MANAGEMENT AND RETRIEVAL) το οποίο εφευρέθηκε στο Πανεπιστήμιο Cornell. Πρόσθεσαν μερικές καινούριες και τα τροποποιημένες διαδικασίες του συστήματος SMART για να μπορέσουν να διαχειριστούν τα Ελληνικά κείμενα και συμπεριέλαβαν μια “stopword” λίστα των πιο συχνά εμφανιζόμενων λέξεων και εξέτασαν τρεις αλγόριθμους συνολικά : τον “TZK

αλγόριθμο” (Καλαμπούκης & Νικολαΐδης, 1995) , τον “infl\_only αλγόριθμο” ο οποίος αφαιρεί μόνο 19 κλινόμενες καταλήξεις και μία καινούρια τροποποιημένη έκδοση του TZK αλγόριθμου. Η αξιολόγηση αυτών των αλγορίθμων έδειξε ότι «η διαδικασία stemming είναι μια διαδικασία συσταδοποίησης βασισμένη στο κορμό (corpus) και γι’ αυτό για να μην καταλήξουμε σε μια ανεπιθύμητη συγχώνευση θα πρέπει να ενσωματώσουμε corpus-based στατιστική για να παγιδεύσουμε το νόημα των όρων». Στην μελέτη τους που δημοσιεύτηκε το 1999 παρουσιάζεται η εκτεταμένη αξιολόγηση καθώς και πιο συγκεκριμένα στατιστικά τέστ.

## 2.10 ΠΡΟΒΛΗΜΑ

Κάθε φυσική γλώσσα έχει τα δικά της χαρακτηριστικά και στοιχεία. Προφανώς λοιπόν είναι δύσκολο να εφαρμόσουμε το ίδιο πρότυπο stemming και τους ίδιους κανόνες stemming για όλες τις λέξεις και σε όλες τις γλώσσες, δημιουργώντας έναν γενικό αλγόριθμο βασισμένο σε κανόνες. Διαφορετικές καταλήξεις και προθέματα καθώς και εξαιρέσεις απαιτούν έναν ειδικό χειρισμό και έναν προσεκτικό σχηματισμό με συγκεκριμένες νόρμες εφαρμοσμένες σε μελετημένες γλώσσες.

Όπως αναφέραμε και παραπάνω, στα μέσα του '90 παρουσιάστηκαν μερικές μέθοδοι stemming για τα ελληνικά κείμενα. Οι μέθοδοι αυτοί αποτελούν μέρος μιας εκτεταμένης εργασίας πάνω σε μορφολογική ανάλυση και ανάκτηση γνώσης από μια ποικιλία κειμένων και δε μπορούν να θεωρηθούν ως stemmers βασισμένοι σε κανόνες, ακόμα και αν ο “TZK αλγόριθμος” είναι ένα σύνολο κανόνων.

Σύμφωνα με την έρευνα όσον αφορά το ελληνικό stemming . οι Καλαμπούκης & Νικολαΐδης το 1995 αλλά και ο Ταμπουρατζής το 2001 συμφώνησαν πως συγκεκριμένοι γραμματικοί κανόνες μπορούν να βελτιώσουν την αποδοτικότητα της ανάκτησης γνώσης από ελληνικά κείμενα. Η ανάπτυξη ενός ελληνικού stemmer με εκτεταμένους γραμματικούς κανόνες θα αποτελέσει λύση του υπάρχοντος προβλήματος του προηγούμενου περιορισμένου αλγόριθμου.

## 2.11 ΑΠΟΤΕΛΕΣΜΑ

Το αποτέλεσμα αυτής της διατριβής είναι ένας εκτεταμένος αλγόριθμος stemming για την Ελληνική γλώσσα. Όλοι αυτοί οι γραμματικοί κανόνες μπορούν να αφαιρέσουν αποτελεσματικά συγκεκριμένες καταλήξεις μιας δοσμένης λέξης. Ο αλγόριθμος είναι γραμμένος σε JavaScript ως μια εφαρμογή του διαδικτύου και λειτουργεί μέσω μιας απλής ιστοσελίδας.

## 2.12 ΣΚΟΠΟΣ

Ο λόγος για τον οποίο παρουσιάζουμε αυτόν τον αλγόριθμο είναι να καλύψουμε το κενό του υπάρχοντος αλγόριθμου και να κατασκευάσουμε έναν πιο αποδοτικό ελληνικό stemmer. Θα κατασκευάσουμε ένα σύστημα βασισμένοι στη προηγούμενη έρευνα, το οποίο θα απομακρύνει αποτελεσματικά τις καταλήξεις των ελληνικών λέξεων. Ένας ακριβής ελληνικός stemmer μπορεί να χρησιμοποιηθεί για πολλούς σκοπούς στους τομείς της Ανάκτησης Γνώσης και της Μορφολογικής Ανάλυσης.

Το σύστημα αυτό θα μας βοηθήσει να αποκτήσουμε περισσότερες και καλύτερες επιτυχίες κατά την αναζήτηση και ανάκτηση πληροφορίας. Τα εξελιγμένα στοιχεία της διαδικασίας stemming έχουν αναβαθμίσει τα χαρακτηριστικά της αναζήτησης για όλες εκείνες τις γλώσσες των οποίων οι stemmers έχουν εξελιχθεί. Επιπλέον, σύμφωνα με μελέτες και μετρήσεις των ελληνικών stemmers, ένας ελληνικός stemmer στο διαδίκτυο θα δίνει στους χρήστες πιο συγκεκριμένα αποτελέσματα αναζήτησης όπως ανέφερε και ο Καλαμπούκης στην ερευνά του (1995) για τους αλγόριθμους stemming για τα Νέα Ελληνικά.

## 2.13 ΜΕΘΟΔΟΣ

Προκειμένου να αναπτύξουμε τον ελληνικό stemmer ακολουθούμε τον αλγόριθμο Porter (Porter 1980) ως πιο αξιόπιστο. Φυσικά, ο αλγόριθμος αυτός κατασκευάστηκε για την Αγγλική γλώσσα και δεν μπορούμε να εφαρμόσουμε τους ίδιους κανόνες στον ελληνικό stemmer. Αλλά ο ελληνικός stemmer προσπαθεί να ακολουθήσει την απλότητα και την ευθύτητα του Porter.

Η έρευνα βασίζεται στη προηγούμενη εργασία για το ελληνικό stemming και την αποτελεσματικότητά του. Το εργαλείο ανάπτυξης είναι η γλώσσα Java, μία γλώσσα ανοιχτού κώδικα, ελεύθερα διαθέσιμη στο διαδίκτυο. Τέλος, για την αξιολόγηση του stemmer

χρησιμοποιήσαμε ένα ελληνικό λεξικό το οποίο μας διέθεσε ευγενικά το Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών «ΔΗΜΟΚΡΙΤΟΣ».

## 2.14 ΠΕΡΙΟΡΙΣΜΟΙ

Για να παράγουμε καλύτερους stemmers ίσως προσθέσουμε μερικούς περιορισμούς στον αλγόριθμο stemming. Κάποιες υποθέσεις είναι επίσης απαραίτητες για να προσδιορίσουμε την έρευνα. Διαφορετικά βρισκόμαστε αντιμέτωποι με μια χρονοβόρα διαδικασία προσπαθώντας να αντιμετωπίσουμε εκτεταμένους κανόνες όπως αυτοί της Ελληνικής γλώσσας.

Αρχικά, θα χρησιμοποιήσουμε μόνο κεφαλαία γράμματα στη προσπάθειά μας να επιλύσουμε το πρόβλημα του τονισμού (Καλαμπούκης & Νικολαΐδης 1995). Χρησιμοποιώντας μόνο κεφαλαία γράμματα μερικές λέξεις μπορούν να προφερθούν διαφορετικά με διαφορετικές σημασίες κάθε φορά. Υπάρχουν όμως, λίγες τέτοιες λέξεις που δεν επηρεάζουν σημαντικά την αποτελεσματικότητα του αλγόριθμου stemming.

Στην Ελληνική γλώσσα τα προθέματα μπορούν να αλλάξουν ριζικά την έννοια και τη σημασιολογία των λέξεων. Για το λόγο αυτό δεν έχουμε ασχοληθεί με την αφαίρεση των προθεμάτων. Εκτός των γενικών προθεμάτων, υπάρχουν περιπτώσεις στην Ελληνική γλώσσα όπου οι λέξεις είναι ομόηχες. Τα ρήματα τα οποία αρχίζουν με σύμφωνο, παίρνουν το γράμμα “ε” σαν πρόθεμα στους παρελθοντικούς χρόνους (Τριανταφυλλίδης 1941). Σε αυτούς τους χρόνους το θέμα αλλάζει, και για το λόγο αυτό υπάρχουν δύο θέματα για κάθε ρήμα.

ΕΝΕΣΤΩΤΑΣ/ΜΕΛΛΟΝΤΑΣ	ΑΟΡΙΣΤΟΣ	ΘΕΜΑ ΕΝΕΣΤΩΤΑ	ΘΕΜΑ ΑΟΡΙΣΤΟΥ
ΔΕΝΩ	ΕΔΕΣΑ	ΔΕΝ	ΔΕΣ
ΦΕΥΓΩ	ΕΦΥΓΑ	ΦΕΥΓ	ΕΦΥΓ
ΠΑΙΖΩ	ΕΠΑΙΞΑ	ΠΑΙΖ	ΕΠΑΙΞ

**ΠΙΝΑΚΑΣ 1. ΘΕΜΑ ΕΝΕΣΤΩΤΑ ΚΑΙ ΘΕΜΑ ΑΟΡΙΣΤΟΥ ΤΗΣ ΙΔΙΑΣ ΛΕΞΗΣ**

Ένας ακριβής ελληνικός stemmer θα πρέπει να αντιμετωπίζει και τις παραγωγικές και τις κλινόμενες καταλήξεις (Καλαμπούκης 1995). Αλλά η Ελληνική γλώσσα είναι πλούσια σε παράγωγα. Αυτό σημαίνει ότι υπάρχουν πολλές λέξεις που παράγονται από το ίδιο θέμα. Και αν λάβει κανείς υπόψη του ότι υπάρχουν δέκα παραγωγικές καταλήξεις και κάθε μια από αυτές έχει πενήντα κλινόμενες καταλήξεις, καταλήγουμε σε μια λίστα με 500 περίπου λέξεις να ανήκουν στην ίδια κατηγορία και να έχουν το ίδιο θέμα. Θέλοντας να εφαρμόσουμε τον ελληνικό stemmer σε μια μηχανή αναζήτησης, κάτι τέτοιο δε θα ήταν χρήσιμο διότι έχουμε πολύ γενικευμένους κανόνες. Γι' αυτό θα ασχοληθούμε μόνο με τις κλινόμενες καταλήξεις.

Στο τρίτο κεφάλαιο της παρούσας εργασίας παρουσιάζονται αναλυτικά οι περιορισμοί του αλγόριθμου του ελληνικού stemming.

## Κεφάλαιο 3

# Η διαδικασία Αναγνώρισης Προτύπων στην Ελληνική Γλώσσα (Greek stemming)

### 3.1. Η Ιστορική αναδρομή της Νέας Ελληνικής γλώσσας

Η γλώσσα που χρησιμοποιούμε οι Έλληνες σήμερα είναι διαφορετική από αυτή που χρησιμοποιούσαμε παλιότερα. Βασίζεται στην Αρχαία Ελληνική γλώσσα που χρησιμοποιούσαν τον 5ο αιώνα π.Χ. Την εποχή εκείνη η Αθήνα άκμαζε τόσο πολιτικά όσο και πνευματικά με αποτέλεσμα όλο και περισσότερες φυλές οι οποίες κατοικούσαν στην Ελλάδα να υιοθετήσουν τη γλώσσα αυτή. Μέρι την εποχή του Μεγάλου Αλεξάνδρου, οι κάτοικοι της Ελλάδας, της Περσίας, της Αιγύπτου και της Μέσης Ανατολής μιλούσαν την Αρχαία Ελληνική γλώσσα . Κατά τη διάρκεια της Ελληνιστικής αυτής περιόδου , παρουσιάστηκαν πολλές μίξεις με αποτέλεσμα να αλλοιωθεί η γλώσσα κυρίως η προφορική. Κατά τη διάρκεια της Βυζαντινής Αυτοκρατορίας, η Ελληνική γλώσσα αλλοιώθηκε για ακόμα μια φορά τόσο στη σύνταξη όσο και στη γραμματική στη προσπάθεια να απλοποιήσουν την γραπτή γλώσσα και να είναι ίδια με την προφορική.

Μετά την Τουρκοκρατία (1453) , η Ελληνική γλώσσα είναι κυρίως προφορική και παραμένει στη μορφή αυτή για σχεδόν 400 χρόνια. Οι άνθρωποι χρησιμοποιούν ένα είδος διαλέκτου τελείως διαφορετική από την κλασσική Ελληνική και εμφανώς επηρεασμένη από την οθωμανική διάλεκτο.

Μετά την απελευθέρωση (1821) το Ελληνικό έθνος χρειάζεται μια καινούρια επίσημη γλώσσα , πριν την επανάσταση της υπόλοιπης Ευρώπης. Στις αρχές του 19ου αιώνα , η Ελλάδα έχει χωριστεί σε δύο ομάδες. Αυτοί που επιθυμούν μια γλώσσα παρόμοια με την αρχαία Ελληνική και από την άλλη μεριά υπήρχε η κατηγορία των ανθρώπων που προτιμούσαν μια γλώσσα περισσότερο απλοποιημένη ,πιο κοντά στην προφορική που χρησιμοποιούνταν εκείνη τη περίοδο. Μετά από πολύ καιρό διαμαχιών καθιερώθηκε ως επίσημη γλώσσα η «Καθομιλουμένη» η οποία παρουσιάστηκε από τον Αδαμάντιο Κοραή. Η «Καθομιλουμένη» είναι ένας συνδυασμός της αρχαίας και της νέας ελληνικής γλώσσας. Ορίστηκε ως επίσημη γλώσσα της Ελλάδας μέχρι το 1976.



Για λόγους απλοποίησης , το ελληνικό κοινοβούλιο αποδέχτηκε τη νεοελληνική γλώσσα το 1976 , τη «Δημοτική» ,ως την επίσημη γλώσσα της Ελλάδας.

### 3.2 Το ελληνικό αλφάβητο

Το αλφάβητο της νέας ελληνικής γλώσσας είναι το ίδιο με αυτό της αρχαίας ελληνικής και βρίσκει τις ρίζες του στο Αιγυπτιακό και στο Φοινικικό αλφάβητο. Παρουσιάζεται στο σχήμα 4.

<i>alpha</i>	<i>beta</i>	<i>gamma</i>	<i>delta</i>	<i>epsilon</i>	<i>zeta</i>	<i>eta</i>	<i>theta</i>
α	β	γ	δ	ε	ζ	η	θ
A	B	Γ	Δ	E	Z	H	Θ
<i>iota</i>	<i>kappa</i>	<i>lambda</i>	<i>mu</i>	<i>nu</i>	<i>xi</i>	<i>omicron</i>	<i>pi</i>
ι	κ	λ	μ	ν	ξ	ο	π
I	K	Λ	M	N	Ξ	O	Π
<i>rho</i>	<i>sigma</i>	<i>tau</i>	<i>upsilon</i>	<i>phi</i>	<i>chi</i>	<i>psi</i>	<i>omega</i>
ρ	σ	τ	υ	φ	χ	ψ	ω
P	Σ	T	Υ	Φ	Χ	Ψ	Ω

ΣΧΗΜΑ 4 . ΤΟ ΕΛΛΗΝΙΚΟ ΑΛΦΑΒΗΤΟ

Το Ελληνικό αλφάβητο περιέχει 24 γράμματα , 7 φωνήεντα (Α, Ε, Η, Ι, Ο, Υ) και 17 σύμφωνα (Β, Γ, Δ, Ζ, Θ, Κ, Λ, Μ, Ν, Ξ, Π, Ρ, Σ, Τ, Φ, Χ, Ψ). Ο χαρακτήρας “ς” ονομάζεται «τελικό σίγμα» όταν είναι στο τέλος της λέξης και μόνο όταν η λέξη γράφεται με μικρά γράμματα.

Κάθε λέξη με περισσότερες από δύο συλλαβές παίρνει τόνο σε ένα φωνήεν και βάσει συγκεκριμένων κανόνων. Σε μερικές περιπτώσεις βάζουμε τόνο σε μονοσύλλαβες λέξεις και άλλοτε βάζουμε 2 τόνους σε μία λέξη, πάντα όταν η λέξη γράφεται με μικρά γράμματα.

### 3.3. ΟΡΙΣΜΟΙ

Πριν δώσουμε μια σύντομη εξήγηση των πιο σημαντικών κανόνων της ελληνικής γραμματικής, θα ήταν χρήσιμο να ορίσουμε μερικούς σημαντικούς όρους που συναντώνται συχνά σε αυτή την εργασία.

- **Συλλαβή** ονομάζεται το μέρος της λέξης που αποτελείται από τουλάχιστον ένα φωνήεν ή ένα φωνήεν που ακολουθείται από ένα ή περισσότερα σύμφωνα.
- **Θέμα** ονομάζεται το αρχικό μέρος της λέξης και παραμένει αναλλοίωτο
- **Παράγωγο** είναι μία λέξη η οποία προέρχεται από μία άλλη προσθέτοντας κάποιο πρόθεμα
- **Παραγωγική κατάληξη** είναι μια κατάληξη μιας παράγωγης λέξης
- **Ρίζα** ονομάζεται μία λέξη η οποία δε μπορεί να παραχθεί από καμία άλλη. Δημιουργείται εάν ένα επίθημα προστίθεται στη ρίζα ή στο αρχικό θέμα
- **Σύνθετη** χαρακτηρίζεται μία λέξη η οποία προέρχεται από δύο λέξεις προσθέτοντας τα θέματά τους.
- **Κατάληξη** είναι το τελευταίο μέρος της λέξης το οποίο μπορεί να αλλάζει.

Στον πίνακα 2 παρουσιάζονται μερικές λέξεις σε διαφορετικές μορφές. Η λέξη «ΧΑΡΑΚΤΗΡΑΣ» είναι η ρίζα. Αποτελείται από το θέμα «ΧΑΡΑΚΤΗΡ» και την κατάληξη «ΑΣ». Το ίδιο συμβαίνει και για τον πληθυντικό αριθμό αλλά με την κατάληξη «ΕΣ».

Για τη λέξη «ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ», έχει το ίδιο θέμα «ΧΑΡΑΚΤΗΡ» και τη κατάληξη «ΙΣΤΙΚΟ». Η λέξη αυτή είναι παράγωγο.

Συνεπώς, μία ρίζα μπορεί να έχει πολλές και διαφορετικές παραγωγικές καταλήξεις με τις οποίες αλλάζει η έννοια της λέξης.

ΛΕΞΗ	ΘΕΜΑ	ΠΑΡΑΓΩΓΙΚΗ ΚΑΤΑΛΗΞΗ	ΚΑΤΑΛΗΞΗ
ΧΑΡΑΚΤΗΡΑΣ	ΧΑΡΑΚΤΗΡ-	-	ΑΣ
ΧΑΡΑΚΤΗΡΕΣ	ΧΑΡΑΚΤΗΡ-	-	ΕΣ
ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ	ΧΑΡΑΚΤΗΡ-	ΙΣΤΙΚΟ	Ο
ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	ΧΑΡΑΚΤΗΡ-	ΙΣΤΙΚΑ	Α
ΧΑΡΑΚΤΗΡΙΖΩ	ΧΑΡΑΚΤΗΡ-	ΙΖΩ	Ω
ΧΑΡΑΚΤΗΡΙΖΕΙΣ	ΧΑΡΑΚΤΗΡ-	ΙΖΕΙΣ	ΕΙΣ
ΧΑΡΑΚΤΗΡΙΣΜΟΣ	ΧΑΡΑΚΤΗΡ-	ΙΣΜΟΣ	ΟΣ
ΧΑΡΑΚΤΗΡΙΣΜΟΙ	ΧΑΡΑΚΤΗΡ-	ΙΣΜΟΙ	ΟΙ

ΠΙΝΑ  
ΚΑΣ 2

·  
3.4.  
Γραμ  
ματι  
κή

Η

Ελληνική γλώσσα έχει δέκα διαφορετικούς τύπους λέξεων : ουσιαστικό, επίθετο, αντωνυμία, ρήμα, επίθετο, επίρρημα, μετοχή, πρόθεση, σύνδεσμο και επιφώνημα. Το άρθρο, το ουσιαστικό, το επίθετο, η αντωνυμία, το ρήμα και η μετοχή είναι κλητικές καταλήξεις και έχουν πολλούς τύπους. Το ουσιαστικό κλίνεται ανάλογα τον αριθμό (ενικός, πληθυντικός). Το επίθετο αλλάζει ανάλογα το γένος (αρσενικό, θηλυκό, ουδέτερο) και τον αριθμό όπως παραπάνω. Για κάθε γένος και αριθμό υπάρχουν τέσσερις διαφορετικοί τύποι για τα ουσιαστικά και για τα επίθετα (ονομαστική, γενική, αιτιατική, κλητική). Το ρήμα έχει δύο αριθμούς όπως τα παραπάνω αλλά και χρόνος και φωνές (ενεργητική, παθητική).

Λαμβάνοντας υπόψη τα ουσιαστικά, υπάρχουν 39 διαφορετικές καταλήξεις σε όλες τις μορφές τους. Προσθέτοντας και τις διαφορετικές περιπτώσεις των επιθέτων υπάρχουν ακόμα 17 διαφορετικές καταλήξεις. Προσμετρώντας και τις 110 πιθανές καταλήξεις των ρημάτων το συνολικό πλήθος των καταλήξεων είναι 166, όπως φαίνεται και στο σχήμα 5.

ΑΔΕΣ	ΗΣΑΜΕ	ΙΕΣΤΕ	ΟΥΜΕ	ΕΣΑΙ	ΙΟΥΜΑΣΤΕ	ΗΘΟΥΝ	ΟΥΝ
ΑΔΩΝ	ΑΝΕ	ΗΚΑ	ΗΘΟΥΜΕ	ΕΣ	ΙΟΥΝΤΑΙ	ΗΘΩ	ΟΥΝΤΑΙ
ΕΔΕΣ	ΗΚΑΝΕ	ΗΚΕΣ	ΗΘΟΥΜΕ	ΕΤΑΙ	ΙΟΥΝΤΑΝ	Ο	ΟΥΝΤΑΝ
ΕΔΩΝ	ΗΘΗΚΑΝΕ	ΗΚΕ	ΑΤΑ	Ι	Η	ΟΙ	ΟΥΣ
ΟΥΔΕΣ	ΑΤΑΝΕ	ΗΘΗΚΑ	ΑΤΟΣ	ΙΕΜΑΙ	ΗΔΕΣ	ΟΜΑΙ	ΟΥΣΑΝ
ΟΥΔΩΝ	ΟΥΣΑΝΕ	ΗΘΗΚΕΣ	ΑΤΩΝ	ΙΕΜΑΣΤΕ	ΗΔΩΝ	ΟΜΑΣΤΑΝ	ΟΥΣΑΤΕ
ΕΩΣ	ΗΣΑΝΕ	ΗΘΗΚΕ	Α	ΙΕΤΑΙ	ΗΘΕΙ	ΟΜΟΥΝ	Υ
ΕΩΝ	ΙΟΥΝΤΑΝΕ	ΟΥΣΑ	ΑΓΑΤΕ	ΙΕΣΑΙ	ΗΘΕΙΣ	ΟΜΟΥΝΑ	ΥΣ
ΙΑ	ΙΟΥΝΤΑΝΕ	ΟΥΣΕΣ	ΑΓΑΝ	ΙΕΣΑΣΤΕ	ΗΘΕΙΤΕ	ΟΝΤΑΙ	Ω
ΙΟΥ	ΙΟΥΝΤΑΝΕ	ΟΥΣΕ	ΑΕΙ	ΙΟΜΑΣΤΑΝ	ΗΘΗΚΑΤΕ	ΟΝΤΑΝ	ΩΝ
ΙΩΝ	ΟΝΤΑΝΕ	ΑΓΑ	ΑΜΑΙ	ΙΟΜΟΥΝ	ΗΘΗΚΑΝ	ΟΝΤΟΥΣΑΝ	ΩΤΕΡ
ΙΚΟ	ΟΤΑΝΕ	ΑΓΕΣ	ΑΝ	ΙΟΜΟΥΝΑ	ΗΘΟΥΝ	ΟΣ	ΩΤΕΡ
ΙΚΟΥ	ΟΥΝΤΑΝΕ	ΑΓΕ	ΑΣ	ΙΟΝΤΑΝ	ΗΘΩ	ΟΣΑΣΤΑΝ	ΩΤΑΤ
ΙΚΑ	ΕΤΕ	ΗΕΑ	ΑΣΑΙ	ΙΟΝΤΟΥΣΑΝ	ΗΚΑΤΕ	ΟΣΑΣΤΕ	ΩΤΑΤ
ΙΚΩΝ	ΗΣΕΤΕ	ΗΕΕ	ΑΤΑΙ	ΙΟΣΑΣΤΑΝ	ΗΚΑΝ	ΟΣΟΥΝ	ΥΤΕΡ
ΑΜΕ	ΟΝΤΑΣ	ΗΘΟΥ	ΑΩ	ΙΟΣΑΣΤΕ	ΗΣ	ΟΣΟΥΝΑ	ΥΤΑΤ
ΗΚΑΜΕ	ΩΝΤΑΣ	ΗΕΤΕ	Ε	ΙΟΣΟΥΝ	ΗΣΑΝ	ΟΤΑΝ	ΕΣΤΕΡ
ΗΘΗΚΑΜΕ	ΟΜΑΣΤΕ	ΟΥΝΕ	ΕΙ	ΙΟΣΟΥΝΑ	ΗΣΑΤΕ	ΟΥ	ΕΣΤΑΤ
ΑΓΑΜΕ	ΙΟΜΑΣΤΕ	ΗΘΟΥΝΕ	ΕΙΣ	ΙΟΥΝΤΑΝ	ΗΣΕΙ	ΟΥΜΑΙ	ΙΑΣ
ΟΥΣΑΜΕ	ΕΣΤΕ	ΗΘΟΥΝΕ	ΕΙΤΕ	ΙΟΥΜΑΙ	ΗΣΕΣ	ΟΥΜΑΣΤΕ	ΙΕΣ
ΙΟΙ	ΙΟΥΣ	ΑΤΕ	ΤΕ	ΗΣΕΙΣ	ΑΣΤΕ		

### ΣΧΗΜΑ 5. ΟΙ 166 ΚΑΤΑΛΗΞΕΙΣ

#### 3.5. Stem σε Ελληνικές λέξεις

Όπως προαναφέραμε για το θέμα (stem) είναι το στατικό μέρος μίας λέξης. Μερικές λέξεις στην Ελληνική γλώσσα έχουν δύο διαφορετικά θέματα. Για τα ουσιαστικά και τα επίθετα, το θέμα αυξάνεται κατά μία συλλαβή. Και αυτό δεν αποτελεί εξαίρεση για την Ελληνική γλώσσα. Στον παρακάτω πίνακα παραθέτουμε μερικά παραδείγματα.

ΕΝΙΚΟΣ ΑΡΙΘΜΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ ΑΡΙΘΜΟΣ	ΘΕΜΑΤΑ	ΚΑΤΑΛΗΞΕΙΣ
ΚΥΜΑ	ΚΥΜΑΤΑ	ΚΥΜ, ΚΥΜΑΤ	Α
ΓΙΑΓΙΑ	ΓΙΑΓΙΑΔΕΣ	ΓΙΑΓΙ, ΓΙΑΓΙΑΔ	Α, ΕΣ
ΑΛΕΠΟΥ	ΑΛΕΠΟΥΔΕΣ	ΑΛΕΠ, ΑΛΕΠΟΥΔ	ΟΥ, ΕΣ

### ΠΙΝΑΚΑΣ 3. ΔΙΑΦΟΡΕΤΙΚΑ ΘΕΜΑΤΑ ΟΥΣΙΑΣΤΙΚΩΝ

Σύμφωνα με την Ελληνική γραμματική, αυτά τα δύο θέματα συμφωνούν εάν η παραπάνω συλλαβή θεωρείται ως μέρος της κατάληξης. Στη περίπτωση αυτή έχουμε το ίδιο θέμα αλλά περισσότερες συλλαβές, όπως φαίνεται στον πίνακα 4.

ΕΝΙΚΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ	ΘΕΜΑΤΑ	ΚΑΤΑΛΗΞΕΙΣ
--------	-------------	--------	------------

ΑΡΙΘΜΟΣ	ΑΡΙΘΜΟΣ		
ΚΥΜΑ	ΚΥΜΑΤΑ	ΚΥΜ	Α, ΑΤΑ
ΓΙΑΓΙΑ	ΓΙΑΓΙΑΔΕΣ	ΓΙΑΓΙ	Α, ΑΔΕΣ
ΑΛΕΠΟΥ	ΑΛΕΠΟΥΔΕΣ	ΑΛΕΠ	ΟΥ, ΟΥΔΕΣ

#### ΠΙΝΑΚΑΣ 4. ΙΔΙΑ ΘΕΜΑΤΑ

Κάτι τέτοιο όμως δεν ισχύει και για τα ρήματα. Κάθε ρήμα έχει δύο θέματα για κάθε φωνή. Το «παρόν θέμα» και το «παρελθοντικό θέμα». Οι χρόνοι σχηματίζονται σύμφωνα με τα παραπάνω δύο θέματα. Επιπλέον, τα ρήματα τα οποία ξεκινούν με σύμφωνο παίρνουν το γράμμα “ε” σε κάποιους παρελθοντικούς χρόνους. Στον πίνακα 5 παρουσιάζονται οι σχηματισμοί του ρήματος «ΔΕΝΩ».

ΧΡΟΝΟΙ	ΡΗΜΑ	ΘΕΜΑΤΑ	ΚΑΤΑΛΗΞΕΙΣ
ΠΑΡΟΝΤΙΚΟΙ ΧΡΟΝΟΙ	ΔΕΝΩ	ΔΕΝ	Ω
	ΘΑ ΔΕΝΩ	ΔΕΝ	Ω
	ΔΕΝΕ	ΔΕΝ	Ε
	ΔΕΝΟΝΤΑΣ	ΔΕΝ	ΟΝΤΑΣ
ΠΑΡΕΛΘΟΝΤΙΚΟΙ ΧΡΟΝΟΙ	ΕΔΕΣΑ	ΔΕΣ	Α
	ΝΑ ΔΕΣΩ	ΔΕΣ	Ω
	ΔΕΣΕ	ΔΕΣ	Ε
	ΔΕΣΕΙ	ΔΕΣ	ΕΙ

#### ΠΙΝΑΚΑΣ 5. ΔΙΑΦΟΡΕΤΙΚΑ ΘΕΜΑΤΑ ΡΗΜΑΤΩΝ

### 3.6. Παραδοχές για τον Ελληνικό Stemmer

Πριν συνεχίσουμε με την παρουσίαση του αλγόριθμου stemming θα ήταν σημαντικό να σημειώσουμε τις παραδοχές για την ανάπτυξη του συγκεκριμένου stemmer. Επειδή η Ελληνική γλώσσα περιέχει πολλά επιθέματα, η διαδικασία της αφαίρεσης της κατάληξης μπορεί να αποβεί χρονοβόρα και ο αλγόριθμος μπορεί να είναι αρκετά περίπλοκος αφού χρησιμοποιεί πολλές εξαιρέσεις και σώματα (corpus). Από την άλλη μεριά, εάν χρησιμοποιήσουμε γενικούς κανόνες ίσως μπορέσουμε να μειώσουμε δύο διαφορετικά σημασιολογικά λέξεις στην ίδια ρίζα (overstemming).

Ο αλγόριθμος δουλεύει για τις ελληνικές λέξεις όταν είναι γραμμένες με κεφαλαία γράμματα όπως τονίσαμε και στην παράγραφο 1.6...Χρησιμοποιώντας κεφαλαία γράμματα αντιμετωπίζουμε το πρόβλημα του τονισμού το οποίο εμφανίζεται όταν χρησιμοποιούμε μικρά γράμματα αφού ο τόνος αλλάζει θέση ανάλογα το σχηματισμό της λέξης.

Ο αλγόριθμος αφαιρεί μόνο τις κλινόμενες καταλήξεις μιας λέξης. Σύμφωνα με τον αυστηρό ορισμό της λέξης “stemming”, πρέπει να αφαιρεθούν και οι παραγωγικές αλλά και οι κλινόμενες καταλήξεις. Ο “TZK αλγόριθμος” έχει μια ορθολογική προσέγγιση στο σημείο αυτό και προσπαθεί να αντιμετωπίσει και τις δύο μορφές καταλήξεων, παραγωγικές και κλινόμενες. Αλλά όπως αναφέραμε και παραπάνω, στους περιορισμούς του συστήματος, υπάρχουν πάρα πολλές λέξεις οι οποίες προέρχονται από το ίδιο θέμα. Για το λόγο αυτό είναι περισσότερο λογικό να θεωρήσουμε ως θέμα μία λέξη χωρίς να απομακρύνουμε την παραγωγική του κατάληξη. Οι κλινόμενες καταλήξεις επηρεάζουν πολύ περισσότερο την ανάκτηση πληροφορίας στα ελληνικά κείμενα σε σχέση με τα αγγλικά, αφού η Ελληνική γλώσσα είναι περισσότερο κλινόμενη.

Η διαδικασία stemming εφαρμόζεται για τις λέξεις εκείνες στις οποίες αλλάζουν οι καταλήξεις τους. Για αυτό το λόγο προσπαθήσαμε να αναπτύξουμε έναν αλγόριθμο stemming μόνο για τους κλινόμενους τύπους λέξεων. Και αφού τα κυρίως κλινόμενα μέρη του λόγου της Ελληνικής γλώσσας είναι το ουσιαστικό, το επίθετο και το ρήμα, θα ασχοληθούμε μόνο με αυτά.

Ο αλγόριθμος ξεχωρίζει το «παροντικό» και το «παρελθοντικό» θέμα των ρημάτων. Όμως δεν μπορούμε να μειώσουμε ένα ρήμα που βρίσκεται στον ενεστώτα με το ίδιο ρήμα όταν βρίσκεται

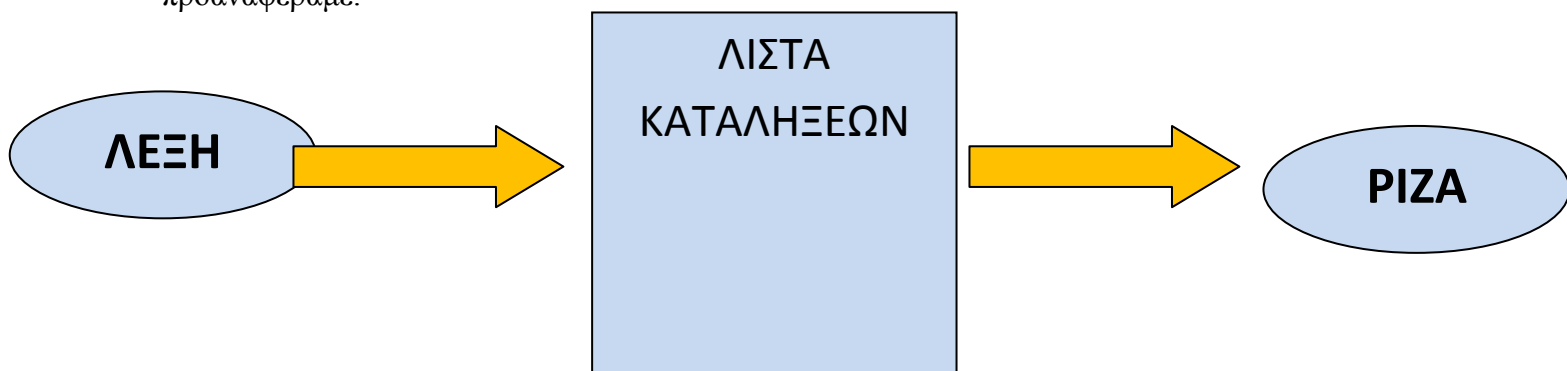
στον αόριστο , στο ίδιο θέμα. Και αυτό διότι τα θέματα είναι διαφορετικά για κάθε χρόνο του ρήματος. Η παραδοχή αυτή όμως δεν ακολουθεί πιστά τον αυστηρό ορισμό του “stemming”. Αλλά εδώ έχουμε να αντιμετωπίσουμε ένα συγκεκριμένο γραμματικό φαινόμενο της Ελληνικής γλώσσας. Αφαιρώντας μόνο τις κλινόμενες καταλήξεις, μειώνουμε ένα ρήμα στο θέμα του, είτε το ρήμα βρίσκεται στον ενεστώτα είτε στον αόριστο.

### 3.7 Ο Stemmer για την Ελληνική γλώσσα

#### 3.7.1. Ο Αλγόριθμος

Ο αλγόριθμος stemming επικεντρώνεται κυρίως στην αφαίρεση των καταλήξεων για οποιαδήποτε σύνθετη Ελληνική λέξη. Λαμβάνοντας υπόψη την Ελληνική γραμματική , καταλήξαμε σε μια λίστα με 166 διαφορετικές καταλήξεις για τις 3 βασικά μέρη του λόγου : ουσιαστικό, επίθετο και ρήμα. Επειδή ο Ελληνικός stemmer δε μπορεί να διαχωρίσει τους τύπους των λέξεων, η προσέγγισή μας είναι η απευθείας αφαίρεση των καταλήξεων.

Η βασική ιδέα είναι να φιλτράρουμε τη λέξη χρησιμοποιώντας μια λίστα καταλήξεων η οποία περιέχει όλες τις πιθανές καταλήξεις. Η λίστα αυτή μπορεί να δημιουργηθεί αφού έχουμε μελετήσει προσεκτικά την Ελληνική γραμματική και περιλαμβάνει τις 166 καταλήξεις που προαναφέραμε.



**ΣΧΗΜΑ 6 : ΓΕΝΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΤΟΥ ΕΛΛΗΝΙΚΟΥ Stemmer**

Βέβαια, με το σύστημα αυτό αντιμετωπίζουμε πρόβλημα αν κάποιες από τις καταλήξεις της λίστας επηρεάζει τη λέξη με λάθος τρόπο, απομακρύνοντας το λάθος μέρος της λέξης αφού το θεωρεί σαν κατάληξη. Για παράδειγμα θέλουμε να απομακρύνουμε τη κατάληξη “Α” και “ΑΔΕΣ” για τη λέξη “ΜΑΜ-Α” και “ΜΑΜ-ΑΔΕΣ”. Έτσι θα έχουμε την ίδια ρίζα “ΜΑΜ” και

για τις δύο περιπτώσεις ενικού και πληθυντικού αριθμού. Εφαρμόζοντας το γενικό κανόνα για ένα διαφορετικό σύνολο λέξεων όπως “ΟΜΑΔ-Α” και “ΟΜΑΔ-ΕΣ” θα μειώσουμε τις δύο λέξεις σε διαφορετικές ρίζες αφού ο αλγόριθμος θα βρεί σαν ρίζα του πληθυντικού αριθμού την “ΟΜ” και όχι “ΟΜΑΔ”.

Προκειμένου να αντιμετωπίσουμε το πρόβλημα αυτό το οποίο απαντάται συχνά στην Ελληνική γλώσσα, μπορούμε να δημιουργήσουμε μια διαφορετική λίστα για την κατάληξη “ΑΔΕΣ” και να «παγιδεύσουμε» όλες τις λέξεις ή τουλάχιστον τις περισσότερες που επηρεάζονται λανθασμένα από αυτόν τον κανόνα. Στην ίδια λίστα μπορούμε να συμπεριλάβουμε και την κατάληξη “ΑΔΩΝ” όπου είναι η κατάληξη των ίδιων λέξεων αλλά στη γενική πτώση. Ένας αλγόριθμος ο οποίος αντιμετωπίζει τέτοιες εξαιρέσεις είναι περισσότερο ακριβής και εφαρμόσιμος από έναν αλγόριθμο με γενικευμένους κανόνες.

Φυσικά, υπάρχουν καταλήξεις που δε δημιουργούν «σύγχυση» και μπορούν εύκολα να αφαιρεθούν. Για παράδειγμα, εάν φτιάξουμε έναν κανόνα ο οποίος αφαιρεί την κατάληξη “ΑΣ” για οποιαδήποτε λέξη χωρίς να συγχέεται με άλλες λέξεις. Αυτός ο γενικός κανόνας καλύπτει ένα μεγάλο πλήθος των Ελληνικών λέξεων.

Τέλος,, οι κανόνες είτε γενικοί είτε εξειδικευμένοι , εφαρμόζονται κάθε φορά για τη μεγαλύτερη ως προς το πλήθος των χαρακτήρων κατάληξη της λίστας. Οπότε, όταν βρίσκουμε την κατάληξη “Α” και “ΑΤΑ” στη λίστα των καταλήξεων , η λέξη “ΚΥΜΑΤΑ” θα μειωθεί στην κατάληξη “ΚΥΜ” και όχι “ΚΥΜΑΤ”.



## Κεφάλαιο 4

# Η διαδικασία Αναγνώρισης Προτύπων στην Ελληνική Γλώσσα (Greek stemming)

### 4.1. Εισαγωγή

Προτείνουμε ένα σύστημα που χρησιμοποιείται ως ενδιάμεσο της επικοινωνίας του εκπαιδευτή και εκπαιδευόμενου που εφαρμόζει τεχνικές του Textual Cased Based Reasoning(TCBR) μια υποκατηγορία των Case Based Reasoning(CBR). Το σύστημα αυτό χρησιμοποιείται ως ενδιάμεσο της επικοινωνίας μεταξύ εκπαιδευτή και εκπαιδευόμενου που εφαρμόζει τεχνικές συλλογιστικής βασισμένες σε περιπτώσεις κειμένου, ώστε να αυτοματοποιηθεί η διαδικασία απάντησης των αποριών του εκπαιδευόμενου προς όφελος τόσο του εκπαιδευτή όσο και του εκπαιδευόμενου. Για τον εκπαιδευτή η αυτοματοποίηση της διαδικασίας αυτής μειώνει σημαντικά το φόρτο εργασίας και ενασχόλησής του, και για τους εκπαιδευόμενους ελαχιστοποιείται η αναμονή της απάντησής του. Το σύστημα βασίζεται στη χρήση της προϋπάρχουσας γνώσης για την επίλυση των αποριών από τους εκπαιδευόμενους. Έτσι σε μια καινούργια απορία ελέγχεται αν υπάρχει γνώση για την επίλυσή της από προηγούμενες περιπτώσεις που είχαν επιλυθεί στο παρελθόν από τον εκπαιδευτή. Αν το σύστημα εντοπίσει προϋπάρχουσα γνώση για την επίλυση της απορίας την επιστρέφει στον εκπαιδευόμενο χωρίς την μεσολάβηση του εκπαιδευτή. Στην αντίθετη περίπτωση που δεν εντοπιστεί πληροφορία ώστε να επιλύει την απορία, την επιλύει ο εκπαιδευτής, και αυτή η γνώση αποθηκεύεται στο σύστημα για την επίλυση παρόμοιων αποριών στο μέλλον. Η εφαρμογή του συστήματος μπορεί να πραγματοποιηθεί για τη διδασκαλία μαθημάτων όλων των θεματικών ενοτήτων.

Σκοπός είναι η δημιουργία ενός συστήματος που θα αυτοματοποιήσει τη διαδικασία επίλυσης αποριών των εκπαιδευόμενων με τη γνώση που προϋπάρχει. Πηγή γνώσης του συστήματος αποτελούν όλες οι λυμένες απορίες των εκπαιδευόμενων από τον εκπαιδευτή. Όλη αυτή η γνώση επεξεργάζεται και αποθηκεύεται σε αρχεία κειμένου που αποτελούν τη βάση του συστήματος. Όταν τίθεται μια καινούργια απορία από κάποιον εκπαιδευόμενο τότε το σύστημα

ελέγχει αν υπάρχει γνώση για την επίλυση της από τη βάση του συστήματος. Όταν υπάρχει γνώση που μπορεί να επιλύσει όλο ή ένα μέρος της απορίας τότε επιστρέφεται στον εκπαιδευόμενο χωρίς την μεσολάβηση του εκπαιδευτή, στην αντίθετη περίπτωση καλείτε ο εκπαιδευτής να την επιλύσει και η γνώση αυτή αφού επεξεργαστεί αποθηκεύεται στη βάση ώστε να χρησιμοποιηθεί στην αντιμετώπιση παρόμοιων περιπτώσεων.

Στόχος του συστήματος είναι να μπορεί να εφαρμοστεί για τη διδασκαλία όλων των μαθημάτων ανεξάρτητα τη θεματική ενότητα που ανήκουν, δηλαδή να είναι θεματικά ανεξάρτητο. Για την επίτευξη του στόχου σχεδιάστηκαν τέσσερις μέθοδοι διαχείρισης γνώσης για τις οποίες πραγματοποιήθηκε αξιολόγηση από δυο διαφορετικές θεματικές ενότητες. Για κάθε θεματική ενότητα δημιουργήθηκε ένα πλήθος περιπτώσεων και ερωτήσεων όπου εξετάστηκαν οι τέσσερις αυτές μέθοδοι διαχείρισης. Από την αξιολόγηση επιλέχτηκε η μέθοδος της οποίας τα αποτελέσματα ήταν τα πιο ικανοποιητικά για την εφαρμογή της στο σύστημα που σχεδιάστηκε.

Το σύστημα που αναπτύχθηκε προτείνει ένα σύστημα αυτόματων απαντήσεων εφόσον υπάρχει γνώση για αυτές. Για την επίτευξη αυτού του σκοπού ενσωματώθηκαν στο σύστημα που προτείνουμε τεχνικές TCBR. Για την υλοποίηση του συστήματος επιλέχτηκε το πλαίσιο Jcolibri που περιγράφηκε παραπάνω, και πάνω σε αυτό ενσωματώθηκαν όλες οι λειτουργίες που υλοποιήθηκαν για την ολοκλήρωση της συγκεκριμένης εργασίας. Στο προηγούμενο κεφάλαιο περιγράφηκαν οι βασικές μέθοδοι ανάκτησης που παρέχει το Jcolibri. Στο συγκεκριμένο σύστημα χρησιμοποιήθηκε η στατιστική μέθοδος ανάκτησης και πάνω σε αυτή υλοποιήθηκε το σύστημα που προτείνουμε. Ο κύριος λόγος που επιλέξαμε τη στατιστική μέθοδο ανάκτησης είναι ότι η πηγή γνώσης αναφέρεται σε κείμενα τα οποία είναι σε αδόμητη μορφή όπως και οι αναρτήσεις που δημοσιεύονται στο φόρουμ. Έχοντας λοιπόν το βασικό πλαίσιο, υλοποιήσαμε ένα ολοκληρωμένο σύστημα που παίρνει τη γνώση που υπάρχει από το φόρουμ, και την αποθηκεύει σε αρχείο κειμένου που θεωρείται η βάση του συστήματος. Στη συνέχεια όταν δημοσιεύεται μια καινούργια ερώτηση ελέγχεται η βάση για παρόμοια περίπτωση και επιστρέφεται στον εκπαιδευόμενο εφόσον υπάρχει γνώση για την απορία του. Στην αντίθετη περίπτωση ο εκπαιδευτής θα πρέπει να δώσει τη λύση του συγκεκριμένου προβλήματος, και το πρόβλημα μαζί με τη λύση να αποθηκευτεί στη βάση για να χρησιμοποιηθεί στο μέλλον σε ανάλογη περίπτωση.

## 4.2 Background Work and Relative Work

### Jcolibri

Το JColibri είναι ένα αντικειμενοστραφές πλαίσιο ανοικτού λογισμικού για την δημιουργία συστημάτων CBR που περιλαμβάνει τις περισσότερες εργασίες που περιλαμβάνονται στον κύκλο ζωής των CBR συστημάτων και έχει αναπτυχθεί στη γλώσσα προγραμματισμού Java. Το Jcolibri είναι το μόνο σύστημα ανοιχτού κώδικα που προσφέρει βασικές εργασίες για την υποκατηγορία των TCBR. Πιο συγκεκριμένα, έχει χωρίσει τα συστήματα TCBR σε δύο κατηγορίες. Ο κύριος διαχωρισμός που γίνεται στις δύο κατηγορίες αυτές αφορά τη δομή του κειμένου, αν είναι δηλαδή δομημένο ή αδόμητο.

Όταν η πηγή γνώσης είναι με τη μορφή κειμένου, είναι δύσκολο να οριστούν γενικές λειτουργίες για τον κύκλο ζωής CBR που να είναι κοινές σε όλες τις διαδικασίες. Έτσι δεν έχουν αναπτυχθεί πολλά συστήματα που να περιλαμβάνουν βασικές λειτουργίες για ένα σύστημα TCBR. Το Jcolibri είναι το μοναδικό που ανέπτυξε ένα πλαίσιο για σύστημα TCBR, που περιλαμβάνει όμως μόνο μεθόδους για το βήμα της ανάκτησης.

Υπάρχουν δύο μεγάλες κατηγορίες αλγορίθμων ανάκτησης (Retrieval) που μπορούν να χρησιμοποιηθούν σε συστήματα TCBR[**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**],[**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**].

1. Η πρώτη κατηγορία βασίζεται σε τεχνικές ΙΕ που αναγνωρίζουν τα χαρακτηριστικά του κειμένου, και εφαρμόζουν τον αλγόριθμο ομοιότητας πλησίον-γειτόνων (Nearest Neighbor similarity). Με τη συγκεκριμένη τεχνική επιλέγεται η περίπτωση που τα χαρακτηριστικά γνωρίσματα ταιριάζουν στο μεγαλύτερο βαθμό με την παρούσα περίπτωση. Στην συγκεκριμένη μέθοδο χρησιμοποιείται γνώση σε μορφή κειμένου που έχει μια σταθερή δομή.
2. Η δεύτερη κατηγορία βασίζεται σε αλγόριθμους που χρησιμοποιούνται από της μηχανές αναζήτησης και που βασίζονται στο Μοντέλο Διανυσματικού Χώρου (Vector Space Model). Αυτό το μοντέλο θεωρεί τόσο τις αποθηκευμένες περιπτώσεις όσο και την τρέχουσα ως διανύσματα ενός πολυδιάστατου χώρου, οι διαστάσεις των οποίων είναι οι λέξεις που χρησιμοποιούνται για την αναπαράσταση του κειμένου. Αυτή η μέθοδος συνηθίζεται να χρησιμοποιείται όταν οι πηγές γνώσεων είναι σε μορφή κειμένου χωρίς κάποια σταθερή δομή.

Η πρώτη κατηγορία συνηθίζεται να αποκαλείται σημασιολογική (semantic) ανάκτηση, επειδή προσπαθεί να αναλύσει την σημασιολογία του κειμένου, και η δεύτερη αποκαλείται στατιστική (statistical) ανάκτηση επειδή λαμβάνει υπόψη κυρίως τη συχνότητα των όρων.

### 4.3 Στατιστική (statistical) Μέθοδος

Η Στατιστική μέθοδος βασίζεται σε διαδικασίες που χρησιμοποιούνται στην μηχανή αναζήτησης Apache Lucene[**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**]. Στο Apache Lucene εφαρμόζεται ένας συνδυασμός του Μοντέλου Διανυσματικού Χώρου(Vector Space Model) με το μοντέλο λογικής (Boolean) για να οριστεί πόσο σχετικό είναι ένα καινούργιο ερώτημα με κάποια περίπτωση του κειμένου.

Το μοντέλο Διανυσματικού χώρου είναι ένας τρόπος οργάνωσης αρχείων. Σύμφωνα με αυτό το μοντέλο κάθε κείμενο αναπαρίσταται σαν ένα διάνυσμα στο οποίο η κάθε «διάσταση» δείχνει τη συμμετοχή κάθε όρου που υπάρχει στο κείμενο, και συγκρίνεται σε σχέση με το πλήθος των εμφανίσεών του σε όλα τα έγγραφα. Το Boolean μοντέλο βασίζεται στη θεωρία συνόλων και αποτελεί ένα από τα πρώτα μοντέλα που χρησιμοποιήθηκαν στα συστήματα ανάκτησης πληροφορίας. Τα ερωτήματα αναπαρίστανται με λογικές εκφράσεις της άλγεβρας Boole, χρησιμοποιώντας τους λογικούς τελεστές. Το βασικό χαρακτηριστικό του Λογικού μοντέλου είναι ότι υποστηρίζει την επακριβή ταύτιση(exact match). Αυτό σημαίνει ότι τα έγγραφα του αποτελέσματος θα πρέπει να ικανοποιούν πλήρως τη λογική έκφραση του ερωτήματος, δηλαδή η λογική έκφραση θα πρέπει να είναι αληθής για κάθε έγγραφο.

Το Apache Lucene προσθέτει ορισμένες δυνατότητες και βελτιώσεις πάνω σε αυτά τα μοντέλα. Τα κύρια πλεονεκτήματα της μεθόδου είναι τα σωστά αποτελέσματα και η δυνατότητα να εφαρμοστεί σε μη δομημένα κείμενα. Το μεγάλο μειονέκτημα είναι η έλλειψη σημασιολογικής γνώσης των κειμένων.

### 4.4 Η δική μας Επέκταση

Στο δικό μας σύστημα επεκτείνουμε την αρχική στατιστική μέθοδο που προσφέρει το Jcolibri για αδόμητο κείμενο με κάποιες σημασιολογικές μεθόδους. Η στατιστική μέθοδος

ουσιαστικά μας παρέχει μια μέθοδο ανάκτησης που δε χρησιμοποιεί κανένα στάδιο προ-επεξεργασίας του κειμένου. Παρέχει όλη τη βιβλιοθήκη για τη χρήση συναρτήσεων ομοιότητας. Τις συναρτήσεις ομοιότητας τις εφαρμόζει κάνοντας έλεγχο όλου του κειμένου όλων των περιπτώσεων και επιστρέφεται αυτή με την μεγαλύτερη βαθμολογία ομοιότητας.

Για την υλοποίηση του συστήματος μια διαδικασία ελέγχου όλου του κειμένου και η χρήση των συναρτήσεων ομοιότητας σε αυτό, για τον εντοπισμό της ιδανικής περίπτωσης, δεν ήταν πλήρως ικανοποιητική. Ο κύριος λόγος ήταν ότι όλο το κείμενο δεν έχει σημαντική πληροφορία αλλά ένα μέρος αυτού. Φυσικά σε αδόμητα κείμενα δεν είναι εύκολο και δεν υπάρχουν πλήρως αυτοματοποιημένες διαδικασίες για τον εντοπισμό της σημαντικής πληροφορίας καθώς η μία περίπτωση από την άλλη, είναι τελείως διαφορετική. Για την μεγαλύτερη αξιοπιστία του συστήματος σχεδιάστηκαν και υλοποιήθηκαν κάποιες μεθοδολογίες που υπάγονται στη διαχείριση γνώσης για το βήμα της ανάκτησης. Η κάθε μια σχεδιάστηκε για να καλύπτει ένα φάσμα απαιτήσεων. Προκειμένου να εντοπιστεί η μεθοδολογία που εντοπίζει τα πιο αξιόπιστα αποτελέσματα πραγματοποιήθηκαν μετρήσεις έτσι ώστε να καταλήξουμε στη χρήση της πιο κατάλληλης.

Πριν παρουσιαστούν όμως οι μέθοδοι που χρησιμοποιήσαμε, στο κεφάλαιο αυτό θα περιγράψουμε το στάδιο προ-επεξεργασίας τόσο του κειμένου (οι περιπτώσεις που έχουν αποθηκευτεί) όσο και του ερωτήματος που υλοποιήθηκε. Το στάδιο προ-επεξεργασίας περιλαμβάνει κάποια βήματα που εφαρμόζονται σε όλες τις μεθοδολογίες που αναπτύχθηκαν. Το στάδιο αυτό είναι πάρα πολύ χρήσιμο, καθώς μπορεί να εξασφαλίσει μεγαλύτερη εγκυρότητα.

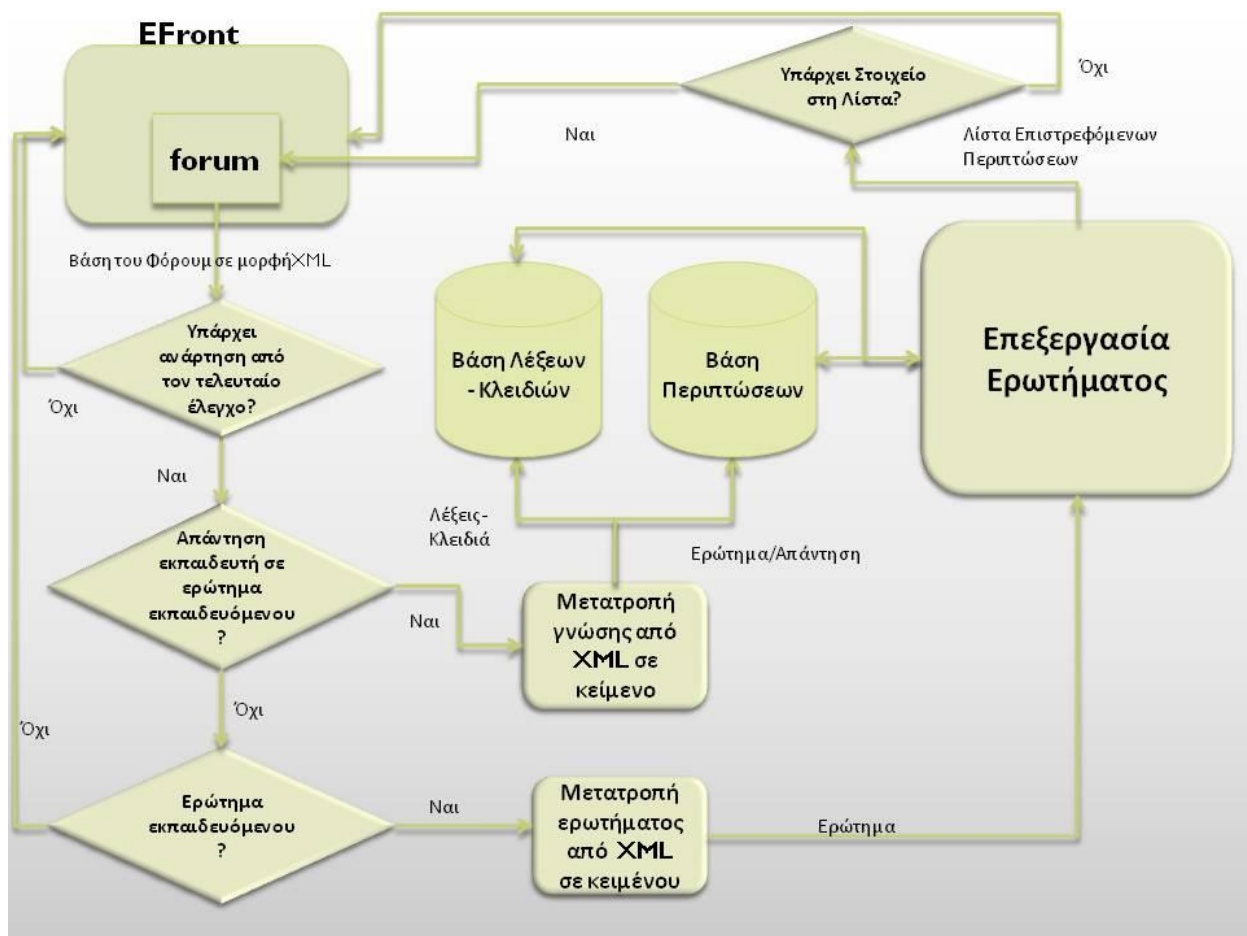
## 4.5 Σύστημα

Για την υλοποίηση του συστήματος μια διαδικασία ελέγχου όλου του κειμένου και η χρήση των συναρτήσεων ομοιότητας σε αυτό, για τον εντοπισμό της ιδανικής περίπτωσης, δεν ήταν πλήρως ικανοποιητική. Ο κύριος λόγος ήταν ότι όλο το κείμενο δεν έχει σημαντική πληροφορία αλλά ένα μέρος αυτού. Φυσικά σε αδόμητα κείμενα δεν είναι εύκολο και δεν υπάρχουν πλήρως αυτοματοποιημένες διαδικασίες για τον εντοπισμό της σημαντικής πληροφορίας καθώς η μία περίπτωση από την άλλη, είναι τελείως διαφορετική. Για την μεγαλύτερη αξιοπιστία του συστήματος σχεδιάστηκαν και υλοποιήθηκαν κάποιες μεθοδολογίες

που υπάγονται στη διαχείριση γνώσης για το βήμα της ανάκτησης. Η κάθε μια σχεδιάστηκε για να καλύπτει ένα φάσμα απαιτήσεων. Προκειμένου να εντοπιστεί η μεθοδολογία που εντοπίζει τα πιο αξιόπιστα αποτελέσματα πραγματοποιήθηκαν μετρήσεις έτσι ώστε να καταλήξουμε στη χρήση της πιο κατάλληλης.

Πριν παρουσιαστούν όμως οι μέθοδοι που χρησιμοποιήσαμε, στο κεφάλαιο αυτό θα περιγράψουμε το στάδιο προ-επεξεργασίας τόσο του κειμένου (οι περιπτώσεις που έχουν αποθηκευτεί) όσο και του ερωτήματος που υλοποιήθηκε. Το στάδιο προ-επεξεργασίας περιλαμβάνει κάποια βήματα που εφαρμόζονται σε όλες τις μεθοδολογίες που αναπτύχθηκαν. Το στάδιο αυτό είναι πάρα πολύ χρήσιμο, καθώς μπορεί να εξασφαλίσει μεγαλύτερη εγκυρότητα.

#### 4.6 Αρχιτεκτονική



## 4.7 Μεθοδολογίες

### 4.7.1 Στάδιο προ-επεξεργασίας

Το στάδιο προ-επεξεργασίας περιλαμβάνει κάποια συγκεκριμένα βήματα που πρέπει να εκτελεστούν πριν γίνει ο έλεγχος ομοιότητας τόσο για τα στοιχεία της βάσης όσο και για το ερώτημα. Η προ-επεξεργασία κειμένου βασίζεται στην ιδέα ότι όλες οι λέξεις του κειμένου δεν είναι κατάλληλες για την παράσταση του περιεχομένου. Βασικός στόχος της είναι να βελτιώσει την αποτελεσματικότητα και αποδοτικότητα της ανάκτησης, καθώς και να παρέχει ένα έλεγχο στο λεξιλόγιο κυρίως σε ότι αφορά τη μείωσή του, με αποτέλεσμα και τη μείωση του μεγέθους του κειμένου. Τα βασικά βήματα από το στάδιο της προ-επεξεργασίας που ενσωματώθηκαν στο σύστημά μας παρουσιάζονται στην Εικόνα 1 και περιγράφονται αναλυτικά στη συνέχεια.



Εικόνα 1 - Στάδια Προ-επεξεργασίας

#### 4.7.2 Γλωσσική προ-επεξεργασία (Λεξιλογική Ανάλυση)

Σε αυτή τη διαδικασία (η οποία ονομάζεται και Tokenization) εντοπίζουμε όλες τις λέξεις του κειμένου και απομακρύνουμε μη λεκτικά σύμβολα (όπως τα σημεία στίξης) ή άλλα (όπως αριθμούς) τα οποία ενδεχομένως να θεωρούμε ότι δεν έχουν ιδιαίτερη σημασία στα έγγραφα. Ακόμη, σε αυτή τη φάση πραγματοποιείται και case folding, δηλαδή μετατρέπονται όλοι οι χαρακτήρες ενός εγγράφου στο ίδιο σχήμα, έτσι ώστε η αναπαράσταση όλων των χαρακτήρων να γίνεται σε μια τυποποιημένη μορφή, με μικρά ή κεφαλαία γράμματα.



Εικόνα 2 - Παράδειγμα Λεξιλογικής Ανάλυσης

#### 4.7.3 Αποκλεισμός Λέξεων( stopwords)

Οι λέξεις αποκλεισμού είναι λέξεις που έχουν πολύ μικρή διακριτική ικανότητα όπως άρθρα, αντωνυμίες, κτητικές αντωνυμίες, κλπ. Οι λέξεις αυτές εμφανίζονται πολύ συχνά στο περιεχόμενο ενός εγγράφου και περιέχουν πολύ μικρή πληροφορία για το περιεχόμενο του εγγράφου στο οποίο εμφανίζονται. Έτσι, η αφαίρεση των stop-words αποτελεί την αφαίρεση μη περιγραφικών λέξεων από τα αρχεία κείμενου διατηρώντας το νόημα των προτάσεων. Οι κοινές λέξεις αποκλεισμού υπάρχει περίπτωση να υπάρχουν πολύ λίγες φορές στο κείμενο, όμως και σε αυτή την περίπτωση θα πρέπει να απομακρυνθούν, καθώς δεν αποτελούν χρήσιμη πληροφορία για το κείμενο. Αυτό έχει σαν αποτέλεσμα να διατηρήσει τις ουσιώδεις λέξεις του κειμένου, και να προσφέρει τη μείωση της προσπάθειας εύρεσης ομοιότητας κάνοντας τη μετέπειτα εργασία περισσότερο χρήσιμη και αποτελεσματική.

Οι stop-words εξαρτώνται από τη φυσική γλώσσα κι έτσι είναι διαφορετικές για κάθε γλώσσα στην οποία είναι γραμμένο ένα κείμενο. Οι stop-words λέξεις από ένα κείμενο θα



μπορούσαμε να πούμε ότι είναι μια τεχνική αφαίρεσης δεδομένων η οποία απαλλάσσει το κείμενο από ένα σημαντικό μέγεθος μη-χρήσιμης πληροφορίας.

#### 4.7.4 Στελέχωση Κειμένου (Stemming)

Το stemming είναι η διαδικασία της μετατροπής μια λέξης στη ρίζα της. Στις περισσότερες γλώσσες οι λέξεις μπορεί να εμφανίζονται με πολλούς τρόπους: για παράδειγμα τα ουσιαστικά έχουν ενικό και πληθυντικό αριθμό ή και πτώσεις(π.χ. αυτοκίνητα, αυτοκίνητο, αυτοκινήτων) ενώ τα ρήματα κλίνονται σε διαφορετικά πρόσωπα και έχουν διάφορους χρόνους και φωνές (ενεργητική, παθητική, μετοχές) πχ κάνω, κάνουν, έκαναν. Έτσι όταν ίδιες λέξεις με το ίδιο νόημα εμφανίζονται με διαφορετική ορθογραφία γίνεται υποβίβαση στη ρίζα τους, και οι λέξεις αυτές θεωρούνται όμοιες. Είναι εύκολα λοιπόν αντιληπτό ότι το stemming είναι απαραίτητο να διεξαχθεί σε όλα τα αρχεία κειμένου πριν το στάδιο της εφαρμογής των συναρτήσεων ομοιότητας. Με την εύρεση κοινής ρίζας μεταξύ των λέξεων, οι αντίστοιχοι όροι συγχωνεύονται.

Έχουν αναπτυχθεί πολλοί και διάφοροι αλγόριθμοι στελέχωσης κειμένου, ο πρώτος δημοσιευμένος αλγόριθμος που αναπτύχθηκε ήταν αυτός του Lovins το 1968 [**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**]. Αποτέλεσε πηγή έμπνευσης για πολλά μετέπειτα συστήματα αυτόματης μορφολογικής ανάλυσης παρόλο που θεωρείται σχετικά απλός. Στη συνέχεια υλοποιήθηκαν και πολλοί άλλοι όπως ο αλγόριθμος Nice, ο οποίος αναπτύχθηκε για το σύστημα IRIS του πανεπιστημίου North Carolina. Ωστόσο ένας από τους πιο διαδεδομένους αλγορίθμους είναι ο αλγόριθμος του Porter και αναπτύχθηκε για την αγγλική γλώσσα, αν και σήμερα έχει εφαρμοστεί και σε άλλες γλώσσες. Ο αλγόριθμος του Porter χρησιμοποιεί μια σειρά κανόνων, που προσθέτουν ή αφαιρούν τμήματα από μια λέξη με συγκεκριμένη σειρά ανάλογα με το συνδυασμό των γραμμάτων που αναγνωρίζουν. Ο συγκεκριμένος αλγόριθμος έχει υψηλά ποσοστά ακρίβειας και ανάκλησης. Ο Martin Porter εξέδωσε μια εφαρμογή ελεύθερου λογισμικού το 2000 ενώ αργότερα επεκτάθηκε με τη δημιουργία ενός πακέτου Showball. Στο σύστημά μας χρησιμοποιήθηκε το συγκεκριμένο πακέτο για τη διαδικασία stemming(<http://snowball.tartarus.org>).

#### 4.7.5 Ανάκτηση Πληροφορίας

Η ανάκτηση πληροφορίας αφορά μια επιστήμη που μελετά προβλήματα που έχουν σχέση με την αναπαράσταση, αποθήκευση, προσπέλαση και οργάνωση πληροφοριακών αντικειμένων. Η συγκεκριμένη μέθοδος αρχικά εφαρμόστηκε σε μικρές συλλογές εγγράφων και στη συνέχεια εξελίχτηκε έτσι ώστε να χρησιμοποιείται στο διαδίκτυο και σε άλλες σχεσιακές βάσεις δεδομένων. Ένας μηχανισμός ανάκτησης πληροφορίας έχει δύο στόχους:

- Αποτελεσματικότητα
- Απόδοση

Τρία είναι τα κλασσικά μοντέλα ανάκτησης πληροφορίας το Λογικό Μοντέλο(Boolean), το Διανυσματικό Μοντέλο και το Πιθανοτικό Μοντέλο.

Πριν όμως περιγραφούν τα μοντέλα που χρησιμοποιήθηκαν θα δώσουμε έναν τυπικό και ακριβή ορισμό για το τι είναι ένα μοντέλο Ανάκτησης Πληροφορίας. Ένα μοντέλο ανάκτησης πληροφορίας είναι η τετράδα  $[D, Q, F, R(q_i, d_j)]$  όπου:

- $D$  είναι ένα σύνολο από λογικές αναπαραστάσεις για τα κείμενα της συλλογής
- $Q$  είναι ένα σύνολο από λογικές αναπαραστάσεις για τις πληροφοριακές ανάγκες του χρήστη. Αυτές οι αναπαραστάσεις καλούνται ερωτήματα
- $F$  είναι ένα υπόβαθρο για την μοντελοποίηση της αναπαράστασης των κειμένων, των ερωτημάτων και των σχέσεων μεταξύ τους
- $R(q_i, d_j)$  είναι μια συνάρτηση κατάταξης, η οποία συνδέει έναν πραγματικό αριθμό με ένα ερώτημα  $q_i \in Q$  και μια αναπαράσταση κειμένου  $d_j \in D$ . Μια τέτοια κατάταξη ορίζει μια διάταξη πάνω στα κείμενα πάντα με βάση το ερώτημα  $q_i$ .

Ο ορισμός αυτός δίνει την περιγραφή καθορισμού ενός μοντέλου Ανάκτησης Πληροφορίας. Στην αρχή ορίζεται ένας τρόπος αναπαράστασης για τα κείμενα και την πληροφοριακή ανάγκη του χρήστη. Έπειτα καθορίζεται ένα υπόβαθρο στο οποίο θα μπορούν αυτές οι αναπαραστάσεις να μοντελοποιηθούν. Το υπόβαθρο αυτό, θα πρέπει να μπορεί να παρέχει και τον μηχανισμό κατάταξης. Για παράδειγμα στο Λογικό μοντέλο, το υπόβαθρο αυτό αποτελείται από τις αναπαραστάσεις των κειμένων και των ερωτήσεων ως σύνολα, και τις κλασσικές πράξεις πάνω στα σύνολα. Αντίστοιχα στο Διανυσματικό Μοντέλο, το υπόβαθρο

αποτελείται από τις διανυσματικές αναπαραστάσεις κειμένων στον t-διάστατο διανυσματικό χώρο και τις επιτρεπτές αλγεβρικές πράξεις πάνω σε διανύσματα.

#### 4.7.6 Lucene

Το Lucene ουσιαστικά είναι μια βιβλιοθήκη Ανάκτησης Πληροφοριών. Αυτό μας δίνει τη δυνατότητα ενός τρόπου ευρετηρίασης και αναζήτησης στο σύστημα. Αφορά ένα λογισμικό ανοιχτού κώδικα υλοποιημένο σε java. Με τη χρήση της βιβλιοθήκης αυτής μπορεί να κατασκευαστεί εύκολα ευρετήριο και να πραγματοποιείται αναζήτηση σε οτιδήποτε δεδομένα είναι σε μορφή κειμένου.

#### 4.7.7 Υπολογισμός Σημαντικότητας Όρων(weighting)

Με τη μέθοδο του ανεστραμμένου καταλόγου, μπορούμε να αναζητήσουμε εύκολα και γρήγορα ένα σύνολο όρων. Ένας όρος που εμφανίζεται σε δύο κείμενα δεν έχει την ίδια σπουδαιότητα και για τα δύο. Έτσι θα πρέπει να εκχωρηθεί ένα βάρος σε κάθε όρο του καταλόγου, αναλόγως τη σημασία του για την αναγνώριση του περιεχομένου.

Για να λειτουργεί αποτελεσματικά το σύστημα ο καθορισμός του βάρους ενός όρου στο έγγραφο που εμφανίζεται ορίζεται ως σημαντικός όταν σχετίζεται με το περιεχόμενο του εγγράφου ώστε να ανακτηθεί όταν είναι χρήσιμο, και να ξεχωρίζει το έγγραφο από τα υπόλοιπα έτσι ώστε να εξασφαλίζεται η ανάκτηση εγγράφων που είναι κατάλληλα. Ο λόγος του αριθμού των σχετικών εγγράφων που ανακτήθηκαν προς το συνολικό αριθμό σχετικών εγγράφων ορίζεται ως ανάκληση, και το ποσοστό των εγγράφων που είναι σχετικά μεταξύ αυτών που έχουν ανακτηθεί ορίζεται ως ακρίβεια.

Για να μπορέσει να υπολογιστεί το βάρος του κάθε όρου στα έγγραφα, έτσι ώστε να ικανοποιούνται οι προϋποθέσεις για την αποτελεσματικότητα του συστήματος, θα χρησιμοποιηθεί το σχήμα TF-IDF(term frequency – inverse document frequency). Με το συγκεκριμένο σχήμα ένας όρος θεωρείται σημαντικός εφόσον η συχνότητα εμφάνισής του σε αυτό είναι μεγάλη, δηλαδή όταν εμφανίζεται πολλές φορές στο έγγραφο, τότε αυτό ο όρος σχετίζεται με το περιεχόμενου του εγγράφου, και έτσι ικανοποιούμε την πρώτη απαίτηση της

αποτελεσματικότητας. Αυτή η πληροφορία αναπαριστάται με τη χρήση ενός παράγοντα συχνότητας όρου (term frequency) που συμβολίζεται  $tf$ .

Ο αριθμός των εμφανίσεων κάποιου όρου σε μία περίπτωση ουσιαστικά παριστάνει τη σπουδαιότητα του όρου αυτού στη συγκεκριμένη περίπτωση. Αν όμως ο όρος αυτός εμφανίζεται σε πολλές περιπτώσεις τότε μειώνεται η σημαντικότητά του. Άρα το βάρος ενός τέτοιου όρου θα πρέπει να είναι μικρό, αφού δεν αποτελεί αντιπροσωπευτικό όρο για καμία περίπτωση όπου εμφανίζεται. Για να υπολογιστεί και αυτή η παράμετρος χρησιμοποιείται ο παράγοντας υπολογισμού βαρών που ονομάζεται αντίστροφη συχνότητα εγγράφων (inverse document frequency) και συμβολίζεται  $idf$ . Με τον παράγοντα αυτό ικανοποιούμε και την δεύτερη απαίτηση της αποτελεσματικότητας του συστήματος.

Όπως αναφέρθηκε το βάρος κάποιου όρου εξαρτάται από δύο παράγοντες, την συχνότητα του όρου και την αντίστροφη συχνότητα εγγράφου. Έτσι το βάρος κάποιου όρου  $t$  που περιλαμβάνεται στο έγγραφο  $d$  συμβολίζεται  $w_{t,d}$  και ισούται:

$$w_{t,d} = tf_{t,d} * idf_t$$

Η συχνότητα του όρου  $tf_{t,d}$  ορίζεται ως ο αριθμός των φορών που εμφανίζεται ο όρος  $t$  στο έγγραφο  $d$ . Η αντίστροφη συχνότητα ορίζεται ως:

$$idf_t = 1 + \log\left(\frac{numDocs}{docFreq + 1}\right)$$

Όπου  $N$  το πλήθος των εγγράφων, και  $ft$  ο συνολικός αριθμός εμφανίσεων του όρου  $t$  στη συλλογή εγγράφων.

Αν χρησιμοποιήσουμε τον παραπάνω τύπο για τον υπολογισμό του βάρους, τότε όροι που εμφανίζονται σε έγγραφα μεγάλου μεγέθους, ενδεχομένως να έχουν και μεγαλύτερο βάρος γιατί αυξάνεται η πιθανότητα ύπαρξής του σε αυτό. Έτσι για να μη γίνεται διάκριση μεταξύ μικρών και μεγάλων εγγράφων ορίζουμε το κανονικοποιημένο βάρος που μπορεί να οριστεί ως εξής:

$$w_{t,d} = \frac{tf_{t,d} * idf_d}{\sqrt{\sum_d (w_{t,d}^2)}}$$

#### 4.7.8 Συναρτήσεις ομοιότητας

Όπως έχει είδη αναφερθεί, στη στατιστική μέθοδο χρησιμοποιείται η βιβλιοθήκη Apache Lucene που εφαρμόζει ένα συνδυασμό του Μοντέλου Διανυσματικού Χώρου (Vector Space Model) με το Μοντέλο Λογικής (Boolean) για να οριστεί πόσο σχετικό είναι ένα καινούργιο ερώτημα με κάποια περίπτωση του κειμένου. Πρώτα χρησιμοποιεί το μοντέλο Λογικής για να περιορίσει τα έγγραφα που πρέπει να βαθμολογηθούν. Το διανυσματικό μοντέλο στηρίζεται στην ιδέα ότι όσο πιο πολλές φορές εμφανίζεται ένας όρος ερωτήματος σε ένα έγγραφο σε σχέση με τον αριθμό που εμφανίζεται συνολικά σε όλα τα έγγραφα, τόσο πιο σχετικό θεωρείται με το ερώτημα.

Το μοντέλο λογικής ( Boolean) στηρίζεται στη θεωρία συνόλων, τα ερωτήματα αναπαρίστανται με λογικές εκφράσεις της Άλγεβρας Boole χρησιμοποιώντας τους λογικούς τελεστές AND (σύζευξη), OR (διάζευξη) και NOT (άρνηση). Το βασικό χαρακτηριστικό αυτού του μοντέλου είναι ότι υποστηρίζει επακριβή ταύτιση, δηλαδή το αποτέλεσμα θα πρέπει να ικανοποιεί τη λογική έκφραση του ερωτήματος. Ένα ερώτημα απαρτίζεται από έναν ή περισσότερους όρους, που προκύπτουν μετά από την προ-επεξεργασία του ερωτήματος. Ο συνδυασμός των όρων του ερωτήματος με τους λογικούς τελεστές σχηματίζουν μια λογική πρόταση του ερωτήματος, που σύμφωνα με το μοντέλο λογικής, ένας όρος είτε υπάρχει σε μία περίπτωση είτε όχι. Άρα για κάθε όρο του ερωτήματος τι μπορούμε να δώσουμε μία τιμή σε κάθε περίπτωση  $d_j$ , 0 ή 1. Η τιμή 0 σημαίνει ότι ο όρος  $d_j$  δεν εμφανίζεται στο κείμενο της περίπτωσης  $d_j$  και η τιμή 1 σημαίνει ότι ο όρος υπάρχει τουλάχιστον μία φορά στο κείμενο της περίπτωσης.

Το διανυσματικό μοντέλο χώρου (vector space model –VSM) στηρίζεται στη διανυσματική αναπαράσταση των εγγράφων και ερωτημάτων. Κάθε άξονας στο χώρο αντιστοιχεί σε ένα χαρακτηριστικό (attribute), δηλαδή σε έναν όρο/λέξη, με αποτέλεσμα η συντεταγμένη κάθε διανύσματος ως προς έναν άξονα να χαρακτηρίζει την εμφάνιση του όρου (στον οποίο αντιστοιχεί ο άξονας) στο συγκεκριμένο διάνυσμα-αρχείο κειμένου, και μάλιστα να αποτελεί ένα «βάρος» του όρου ως προς το συγκεκριμένο κείμενο. Τα βάρη που χρησιμοποιούνται για κάθε attribute είναι πραγματικές τιμές και μπορεί να είναι απλά η συχνότητα εμφάνισης της λέξης.

Η συνάρτηση ομοιότητας του Lucene βασίζεται στο σχήμα TF-IDF (term frequency – inverse document frequency) που περιγράψαμε στον τρόπο υπολογισμού σημαντικότητας των όρων. Το Lucene μοντέλο εφαρμόζει έναν συνδυασμό των δύο μοντέλων που περιγράφηκαν, του Λογικού και του Διανυσματικού Χώρου. Στην αρχή χρησιμοποιείται το Λογικό μοντέλο για να μπορέσει να περιορίσει τα έγγραφα που πρέπει να βαθμολογηθούν. Πιο συγκεκριμένα εντοπίζει όλα τα έγγραφα που περιέχουν έστω και ένα όρο που περιλαμβάνεται στο ερώτημα μετά την προ-επεξεργασία. Τα έγγραφα αυτά είναι που θα πρέπει να βαθμολογηθούν με την συνάρτηση ομοιότητας για να εντοπιστεί το πιο κατάλληλο. Τα έγγραφα που δεν περιέχουν κανέναν όρο και δεν επιλέγονται από το Λογικό μοντέλο βαθμολογούνται απευθείας με την τιμή μηδέν. Τα έγγραφα που επιλέχθηκαν από το Λογικό Μοντέλο θα βαθμολογηθούν εφαρμόζοντας το Μοντέλο Διανυσματικού Χώρου. Ο υπολογισμός της βαθμολογίας που αφορά την ομοιότητα ενός εγγράφου  $d$  σε σχέση με ένα ερώτημα  $q$  υπολογίζεται με το εσωτερικό γινόμενο μεταξύ των διανυσμάτων του εγγράφου και του ερωτήματος. Η βαθμολογία ομοιότητας του Lucene υπολογίζεται ως:

$$score(q, d) = coord(q, d) \times queryNorm(q) \times \sum_{t \in q} [tf(t \text{ in } d) \times idf(t)^2 \times t.getBoost() \times lengthNorm(d)]$$

Όπου

- $tf(t \text{ in } d)$  συσχετίζεται με τη συχνότητα του όρου  $t$  μέσα στο κείμενο  $d$ . Τα έγγραφα που κατέχουν μεγαλύτερο πλήθος εμφανίσεων ενός όρου λαμβάνουν υψηλότερη βαθμολογία. Ο προκαθορισμένος υπολογισμός είναι:

$$tf(t \text{ in } d) = \sqrt{frequency}$$

- $idf(t)$  είναι η ανάστροφη συχνότητα εγγράφου. Συσχετίζεται η τιμή αυτή με το αντίστροφο του  $docFreq$ , δηλαδή του πλήθους εμφανίσεων στα οποία εμφανίζεται ο όρος. Έτσι οι πιο σπάνιοι όροι στη συλλογή εγγράφων δίνουν μια υψηλότερη συνεισφορά στη συνολική βαθμολογία. Ο προκαθορισμένος υπολογισμός είναι:

$$idf(t) = 1 + \log\left(\frac{numDocs}{docFreq + 1}\right)$$

- $coord(q, d)$  είναι ένας παράγοντας που εξαρτάται από το πόσους όρους του ερωτήματος  $q$  περιλαμβάνονται στο έγγραφο  $d$ . Το έγγραφο που περιέχει τους περισσότερους όρους του ερωτήματος λαμβάνει μεγαλύτερη βαθμολογία από αυτά που περιλαμβάνουν λιγότερους.

- $queryNorm(q)$  είναι ένας παράγοντας κανονικοποίησης που χρησιμοποιείται για να κάνει τις βαθμολογίες μεταξύ διαφορετικών ερωτημάτων συγκρίσιμες. Ο παράγοντας αυτός δεν επηρεάζει την κατάταξη των εγγράφων (δεδομένου ότι όλα τα έγγραφα που κατατάσσονται πολλαπλασιάζονται με τον ίδιο συντελεστή). Ο προκαθορισμένος υπολογισμός είναι:

$$queryNorm(q) = queryNorm(sumOfSquaredWeights) = \frac{1}{sumOfSquaredWeights^{1/2}}$$

όπου  $sumOfSquaredWeights$  το άθροισμα των τετραγώνων του βάρους που ισούται:

$$sumOfSquaredWeights = \sum_{t \in q} (idf(t) \times t.getBoost())^2$$

- $t.getBoost()$  είναι ένας παράγοντας ώθησης κάποιου όρου του ερωτήματος που θα επιλέξει ο χρήστης. Στην ουσία δίνεται η δυνατότητα να προσδιορίσει ο χρήστης τη σημαντικότητα κάποιου όρου του ερωτήματος καθώς και το βαθμό της σημαντικότητας του. Ο τρόπος με τον οποίο εφαρμόζεται είναι με τη χρήση του συμβόλου  $\wedge$  ακολουθούμενο από έναν αριθμό που προσδιορίζει τη σπουδαιότητα του όρου που αναζητά. Η προεπιλεγμένη τιμή για τους όρους του ερωτήματος είναι 1.
- $LengthNorm(d)$  είναι ένας παράγοντας κανονικοποίησης που εξαρτάται από το αντίστροφο του συνολικού πλήθους των όρων του εγγράφου  $d$ . Ουσιαστικά συσχετίζεται με το μέγεθος του κειμένου  $d$ .
- Εφαρμόζοντας τη συγκεκριμένη συνάρτηση ομοιότητας εξάγονται τα εξής γενικευμένα συμπεράσματα:
- Έγγραφα τα οποία περιέχουν στο κείμενό τους όλους τους όρους του ερωτήματος θεωρούνται πολύ καλά.
- Λέξεις που ο αριθμός εμφάνισής του σε όλη τη συλλογή εγγράφων είναι μικρός, έχουν μεγαλύτερη αξία από αυτές που εμφανίζονται συχνά.
- Τα έγγραφα που η έκτασή τους είναι μικρή είναι καλύτερα από τα μεγαλύτερα αν περιέχουν ακριβώς τις ίδιες λέξεις του ερωτήματος και στην ίδια συχνότητα.

#### 4.7.9 Μέθοδοι Διαχείρισης Της Γνώσης

Για την ανάγκη επίλυσης του προβλήματος θα δοκιμάσουμε τέσσερις διαδικασίες(αλγόριθμους) που έχουν δημιουργηθεί στα πλαίσια της συγκεκριμένης εργασίας. Στη συνέχεια θα αξιολογήσουμε τις διαδικασίες αυτές με δύο βάσεις που έχουν δημιουργηθεί στα πλαίσια δύο μαθημάτων διαφορετικών ειδικοτήτων. Οι βάσεις αυτές επιλέξαμε να είναι από διαφορετικές επιστήμες για να είναι πιο ακριβή τα αποτελέσματα. Οι δύο τομείς που επιλέξαμε είναι του τομέα επιστήμης υπολογιστών και ψυχολογίας. Η αξιολόγηση θα καθορίσει την επιλογή της πιο αξιόπιστης και έγκυρης διαδικασίας που θα εφαρμοστεί. Οι μεθοδολογίες που έχουν δημιουργηθεί για τις ανάγκες της εργασίας παρουσιάζονται στη συνέχεια.

#### 4.7.10 Στατιστική Μέθοδος

Η πρώτη μεθοδολογία που αναπτύχθηκε είναι αυτή της στατιστικής ανάκτησης με τη γενικότερη ιδέα που περιγράφεται στα συστήματα Textual Case Based Reasoning. Στην συγκεκριμένη διαδικασία έχουμε μία βάση από όλες τις περιπτώσεις. Στη βάση αυτή έχει αποθηκευτεί όλη η γνώση, δηλαδή το πρόβλημα μαζί με την λύση του. Πρόβλημα στην προκειμένη κατάσταση εννοούμε την απορία κάποιου εκπαιδευόμενου, και λύση την απάντηση έτσι όπως έχει οριστεί από τον εκπαιδευτή. Οπότε στη βάση υπάρχουν αποθηκευμένες όλες οι ερωτήσεις που έχουν αναρτηθεί από εκπαιδευόμενους, μαζί με την απάντηση που έχει δοθεί από τον εκπαιδευτή και μόνο, εφόσον υπάρχει.

Όταν δημοσιεύεται ένα καινούργιο ερώτημα, αρχικά δημιουργούνται τα Lucene Document για κάθε μία περίπτωση που υπάρχει στη βάση, στη συνέχεια πραγματοποιείται όλη η διαδικασία προ-επεξεργασίας του ερωτήματος και των περιεχομένων των περιπτώσεων που έχουν αποθηκευτεί στα Lucene Document. Έπειτα δημιουργείται ο ανεστραμμένος κατάλογος που αναπαριστά τους όρους που περιέχονται στις περιπτώσεις. Έχοντας ολοκληρωθεί ο ανεστραμμένος κατάλογος μπορούν να εφαρμοστούν οι συναρτήσεις ομοιότητας για να δοθεί η βαθμολογία ομοιότητας για κάθε περίπτωση σε σχέση με το ερώτημα που τέθηκε. Η βαθμολογία αυτή αφορά την ομοιότητα που έχει κάθε μια περίπτωση με το ερώτημα. Στη συνέχεια ταξινομούνται όλες οι περιπτώσεις σε σχέση με τη βαθμολογία ομοιότητας που έχει λάβει το καθένα. Οι περιπτώσεις που επιλέγονται είναι αυτές που η βαθμολογία τους είναι μεγαλύτερη ή ίση από την τιμή 0,6.

Επιστρέφεται στο χρήστη η περίπτωση με τη μεγαλύτερη βαθμολογία και από δίπλα εμφανίζονται στη σειρά όλες οι περιπτώσεις με βαθμολογία μεγαλύτερη ή ίση του 0,6 με την ένδειξη «Μήπως εννοείτε: ». Έτσι το σύστημα επιστρέφει στον εκπαιδευόμενο την περίπτωση που θεωρεί πιο πιθανή σχετικά με το ερώτημα που έθεσε, χωρίς όμως να παραλείψει και τις



άλλες περιπτώσεις που έχουν εξίσου μεγάλη βαθμολογία. Ο λόγος που επιλέχτηκε ο αριθμός 0,6 ως ένας φραγμός στη βαθμολογία, σχετίζεται από τα αποτελέσματα των μετρήσεων, το οποίο θα περιγραφεί σε επόμενα κεφάλαια.

Το αρνητικό της συγκεκριμένης μεθόδου είναι ότι η αναζήτηση της πιο κατάλληλης περίπτωσης γίνεται στη βάση με όλες τις περιπτώσεις, δηλαδή σε όλο το κείμενο του ερωτήματος και της απάντησης. Το κείμενο της περίπτωσης μπορεί να περιέχει και όρους που δεν προσδιορίζουν κάποια σημαντική πληροφορία για τη συγκεκριμένη περίπτωση, και να δώσουν μη ακριβείς βαθμολογίες. Έτσι αποφασίστηκε να αναπτυχθούν και άλλες μέθοδοι για τη διαχείριση της γνώσης που παρουσιάζονται στη συνέχεια.

#### 4.7.11 Μέθοδος Ευρετηρίου

Προκειμένου να καλύψουμε τα αρνητικά της προηγούμενης μεθόδου στη συγκεκριμένη μεθοδολογία σκοπεύουμε να καλύψουμε δύο βασικές διαδικασίες με στόχο την αποδοτικότερη ανάκληση περιπτώσεων.

Αρχικά σκοπεύουμε να εντοπίσουμε τα σημαντικά χαρακτηριστικά της κάθε περίπτωσης. Σε κάθε πρόβλημα οι περιπτώσεις αποτελούνται από έναν αριθμό χαρακτηριστικών. Από αυτά τα χαρακτηριστικά, κάποια είναι πολύ σημαντικά, κάποια άλλα λιγότερο, κάποια μπορεί να είναι αδιάφορα, ενώ υπάρχει και η πιθανότητα, να υπάρχουν χαρακτηριστικά τα οποία αντί να διευκολύνουν τη διάκριση των περιπτώσεων να την κάνουν δυσκολότερη. Η εύρεση των σημαντικών χαρακτηριστικών μπορεί να γίνεται από κάποιον ειδικό, είτε να γίνεται και αυτόματα.

Στη συνέχεια καλούμαστε να εντοπίσουμε και να επανακτήσουμε παρόμοιες περιπτώσεις από την βάση περιπτώσεων. Εδώ, παρουσιάζεται πρόβλημα όταν το μέγεθος της βάσης περιπτώσεων γίνει αρκετά μεγάλο. Σε μια τεράστια βάση το κόστος ανάκτησης της πιο όμοιας περίπτωσης γίνεται πολύ υψηλό με αποτέλεσμα η χρήση του συστήματος να είναι ασύμφορη. Ο ανεστραμμένος κατάλογος που θα δημιουργηθεί για τους όρους των περιπτώσεων θα είναι πάρα πολύ μεγάλος, και πέρα από τη μνήμη που θα δεσμεύει, η αναζήτηση κάποιου όρου σε αυτό θα είναι μια χρονοβόρα διαδικασία. Έτσι η προτεινόμενη λύση είναι η κατάλληλη δεικτοδότηση των περιπτώσεων από τα χαρακτηριστικά αυτών. Αυτό θα έχει σαν αποτέλεσμα η εύρεση και κατ' επέκταση η δημιουργία του ανεστραμμένου καταλόγου να πραγματοποιείται από μια βάση πολύ μικρότερου μεγέθους μόνο με τα χαρακτηριστικά των περιπτώσεων.

Έτσι δημιουργήθηκε η μέθοδος του ευρετηρίου (indexing). Για κάθε μία από τις περιπτώσεις (ερωτήσεις – απαντήσεις) θα δηλώνονται όλες οι λέξεις-κλειδιά που τη χαρακτηρίζουν. Οι λέξεις-κλειδιά για κάθε μία περίπτωση αναπαριστά τα χαρακτηριστικά του κειμένου. Τις λέξεις-κλειδιά αρχικά θα τις προσδιορίζει ο καθηγητής αφού απαντήσει στην

απορία που τέθηκε. Με την μέθοδο του ευρετηρίου, η εύρεση της κατάλληλης περίπτωσης σε μια καινούργια απορία που τίθεται θα γίνεται με τον έλεγχο των λέξεων-κλειδιών που έχουν οριστεί για κάθε περίπτωση, και όχι σε ολόκληρο το κείμενο. Έτσι επιτυγχάνονται δύο βασικά πλεονεκτήματα:

- ▶ Το πρώτο είναι ότι η εύρεση πλέον δεν γίνεται σε ολόκληρο το κείμενο, αλλά σε ένα πολύ μικρότερο μέγεθος. Έτσι μειώνεται αισθητά ο χρόνος εκτέλεσης του αλγορίθμου κυρίως όταν με την πάροδο του χρόνου η γνώση του συστήματος μεγαλώνει.
- ▶ Είναι πιο αξιόπιστο, γιατί οι λέξεις-κλειδιά προσδιορίζουν την ουσία του κειμένου και αυτό που περιγράφει. Έτσι εξασφαλίζουμε ότι όταν στο μέλλον τεθεί μια καινούργια απορία που έχει απαντηθεί στο παρελθόν θα χρησιμοποιηθούν κάποιοι από τους όρους αυτούς ή συνώνυμοι. Αυτός ήταν και ο λόγος που επιλέχθηκε τις λέξεις κλειδιά να τις προσδιορίζει ο εκπαιδευτής για κάθε μία περίπτωση, καθώς είναι ο πιο κατάλληλος να προσδιορίσει τις λέξεις-κλειδιά για κάθε μία περίπτωση καλύτερα από οποιονδήποτε.

Για την υλοποίηση της συγκεκριμένης διαδικασίας χρησιμοποιούνται δύο βάσεις δεδομένων, στην πρώτη υπάρχουν όλες οι περιπτώσεις(ερωτήσεις – απαντήσεις) που έχουν δοθεί στο φόρουμ, και στη δεύτερη βάση όλες οι λέξεις-κλειδιά (χαρακτηριστικά) που έχουν δοθεί ομαδοποιημένες σε σχέση με την περίπτωση που απευθύνονται. Οι περιπτώσεις που αποθηκεύονται είναι αυτές στις οποίες ο εκπαιδευτής έχει δώσει απάντηση σε κάποια ερώτηση εκπαιδευόμενου. Στην πρώτη βάση αποθηκεύεται η ερώτηση μαζί με την απάντηση, και στη δεύτερη βάση αποθηκεύονται σε αντίστοιχες θέσεις οι λέξεις-κλειδιά που έχει ορίσει ο εκπαιδευτής για τη συγκεκριμένη περίπτωση. Η πρώτη βάση ονομάζεται "Βάση Περιπτώσεων", και η δεύτερη "Βάση Λέξεων-Κλειδιών". Στην ορολογία των συστημάτων TCBR ως βάση θεωρείται ένα αρχείο κειμένου, που αποθηκεύεται όλη η γνώση. Η αντιστοιχία των δύο βάσεων ορίζεται ως η θέση αποθήκευσής τους στις δύο βάσεις. Το διαχωριστικό αναγνώρισης μεταξύ των περιπτώσεων είναι τρεις κενές γραμμές στο αρχείο κειμένου που είναι αποθηκευμένες.

Έτσι η διαδικασία εύρεσης της πιο ιδανικής περίπτωσης στη συγκεκριμένη μεθοδολογία εκτελείτε ως εξής:

- Δημιουργούνται τα Lucene Document από τη βάση με τις λέξεις-κλειδιά.
- Πραγματοποιείται η προ-επεξεργασία των Lucene Document και του ερωτήματος.
- Στη συνέχεια δημιουργείται ο ανεστραμμένος κατάλογος για τους όρους των Lucene Document όπως αυτοί προκύπτουν από την προ-επεξεργασία.

- Εφαρμόζονται οι συναρτήσεις ομοιότητας για την απόδοση της βαθμολογίας σε σχέση με την ομοιότητα των λέξεων-κλειδιών της κάθε περίπτωσης με το τρέχον ερώτημα και αποδίδεται σε κάθε μία ομάδα λέξεων-κλειδιών που αντιστοιχούν σε μία περίπτωση μία βαθμολογία
- Ταξινομείται η λίστα με της ομάδες των περιπτώσεων σε σχέση με την βαθμολογία που κατέχουν.
- Γίνεται αντιστοίχιση των ομάδων λέξεων-κλειδιών με τις περιπτώσεις και επιστρέφονται αυτές οι περιπτώσεις που κατέχουν την μεγαλύτερη βαθμολογία.

Η βαθμολογία αυτή αντιστοιχεί στην ομοιότητα που έχει κάθε ένα σύνολο λέξεων-κλειδιών που εκφράζουν κάποια περίπτωση με το ερώτημα που τέθηκε. Επιλέγουμε όλες τις περιπτώσεις που έχουν βαθμολογία μεγαλύτερη ή ίση από 0,6 σε μία λίστα για τους λόγους που προαναφέρθηκαν. Επομένως επιστρέφεται στο χρήστη η περίπτωση με τη μεγαλύτερη βαθμολογία και από δίπλα εμφανίζονται στη σειρά όλες οι περιπτώσεις με βαθμολογία μεγαλύτερη ή ίση του 0,6 με την ένδειξη «Μήπως εννοείτε: ». Με τον τρόπο αυτό εξασφαλίζουμε ότι αν υπάρχει γνώση για ένα ερώτημα, αν δεν είναι αυτή που επιστρέφεται θα είναι μέσα στην λίστα των περιπτώσεων με την ένδειξη «Μήπως εννοείτε: ». Έτσι θα διασφαλίζουμε ότι εφόσον υπάρχει γνώση, αυτή θα εντοπιστεί και θα επιστραφεί. Φυσικά αυτό είναι το ζητούμενο για κάθε μία μεθοδολογία.

#### 4.7.12 Μέθοδος Ευρετηρίου με τη χρήση Συνώνυμων Όρων

Η συγκεκριμένη μεθοδολογία αναπτύχθηκε για την προσθήκη σημασιολογίας στο σύστημά μας εξασφαλίζοντας τα πλεονεκτήματα της προηγούμενης μεθόδου. Έτσι προτείνεται μία μεθοδολογία που εφαρμόζει τη μέθοδο του ευρετηρίου χρησιμοποιώντας κάποιο σημασιολογικό λεξικό για την εύρεση των συνώνυμων όρων. Το σημασιολογικό λεξικό που επιλέχθηκε είναι το WordNet και αφορά την αγγλική γλώσσα. Το WordNet είναι μία εφαρμογή ανοιχτού κώδικα που αναπτύσσεται από το Πανεπιστήμιο του Princeton. Η επιλογή της χρήσης του συγκεκριμένου λεξικού προέκυψε λόγω της ευρείας χρήσης του από πολλά συστήματα ως γνωσιακή βάση. Με την χρήση του ηλεκτρονικού λεξικού WordNet σκοπός είναι να βρεθούν όλες οι λέξεις που έχουν την ίδια σημασιολογία με τους όρους του ερωτήματος.

Πιο αναλυτικά όταν τίθεται ένα ερώτημα αρχικά δημιουργούνται τα Lucene Document από τη βάση με τις λέξεις-κλειδιά. Στη συνέχεια εφαρμόζεται η προ-επεξεργασία των Lucene Document και δημιουργείται ο ανεστραμμένος κατάλογος από τους όρους των λέξεων κλειδιών όπως προκύπτουν από την προ-επεξεργασία. Η προ-επεξεργασία του ερωτήματος εφαρμόζεται

με τη διαφορά ότι δεν περιλαμβάνει το βήμα της στελέχωσης κειμένου, γιατί έτσι δεν θα μπορεί να πραγματοποιηθεί η εύρεση των συνώνυμών τους. Εφόσον το ερώτημα πλέον περιλαμβάνει μόνο τους χρήσιμους όρους(από αυτούς που ήταν εφικτό να εντοπίσουμε), εφαρμόζουμε το λεξικό WordNet έτσι ώστε να ανακτήσουμε όλους τους συνώνυμους όρους τους. Έπειτα εκτελείται και το τελευταίο βήμα της προ-επεξεργασίας με τη διαδικασία του stemming τόσο στους όρους του ερωτήματος όσο και στους συνώνυμους όρους τους που εντοπίστηκαν από το λεξικό.

Το λεξικό WordNet κάποιες φορές επιστρέφει έναν όρο παραπάνω από μια φορές, και επιπλέον επιστρέφει πολύ συχνά όρους που έχουν κοινή ρίζα, έτσι έχει κατασκευαστεί μια συνάρτηση η οποία απαλείφει όμοιους όρους που εμφανίζονται στη πρόταση. Τη συνάρτηση αυτή την εφαρμόζουμε στην πρόταση που περιλαμβάνει το αρχικό ερώτημα μετά από όλη την προ-επεξεργασία συμπεριλαμβανομένου και του stemming, μαζί με τους συνώνυμους όρους τους.

Στο τελευταίο βήμα εφαρμόζεται η συνάρτηση ομοιότητας για να αποδοθούν οι βαθμοί ομοιότητας σε κάθε περίπτωση σε σχέση με το ερώτημα μαζί με τους συνώνυμους όρους τους και ταξινομούνται βάσει αυτής της βαθμολογίας. Επιλέγουμε όλες τις περιπτώσεις που έχουν βαθμολογία μεγαλύτερη ή ίση από 0,6 όπως και στις προηγούμενες μεθόδους. Έτσι επιστρέφεται στο χρήστη η περίπτωση που αντιστοιχεί στη μεγαλύτερη βαθμολογία και από δίπλα εμφανίζονται στη σειρά όλες οι περιπτώσεις με βαθμολογία μεγαλύτερη ή ίση του 0,6 με την ένδειξη «Μήπως εννοείτε: ».

Appendix

#### **4.7.13 Μέθοδος Ευρετηρίου με τη χρήση Συνώνυμων Όρων θέτοντας βάρη.**

Στη μέθοδο ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη στόχος είναι να συνδυαστούν οι μέθοδος του ευρετηρίου και η μέθοδος του ευρετηρίου με τη χρήση συνώνυμων όρων. Στην τελευταία μέθοδο χρησιμοποιείται η ιδέα του ευρετηρίου εντοπίζοντας και τους συνώνυμους όρους του ερωτήματος. Το τελικό ερώτημα προκύπτει από τους αρχικούς όρους και τους συνώνυμους τους. Αυτό σημαίνει ότι οι συνώνυμοι όροι επηρεάζουν τη βαθμολογία των περιπτώσεων σχετικά με την ομοιότητά τους σε ένα μεγαλύτερο ποσοστό από τους όρους που αρχικά είχε θέσει ο χρήστης. Ο λόγος είναι ότι το πλήθος των συνώνυμων για έναν όρο του ερωτήματος είναι περισσότεροι του ενός, και σε κάποιες περιπτώσεις μπορεί να φτάνει και το δεκαπλάσιο του πλήθους των αρχικών όρων. Για να περιοριστεί το ποσοστό της βαθμολογίας που αφορά τους συνώνυμους όρους του ερωτήματος, λαμβάνοντας τους όμως υπόψη αναπτύχθηκε η μέθοδος του ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη.

Στη συγκεκριμένη μέθοδο εκτελούνται δύο βασικές διεργασίες για να μπορέσει να βρεθεί το ιδανικό αποτέλεσμα για το ερώτημα. Η πρώτη διαδικασία είναι αυτή του ευρετηρίου που περιγράφηκε παραπάνω. Υπολογίζεται και επιστρέφεται η λίστα με τα πιο όμοια αποτελέσματα από τη βάση με τις λέξεις-κλειδιά. Η κάθε μία περίπτωση από την λίστα έχει και ένα αριθμό ομοιότητας που κυμαίνεται μεταξύ του 0-1, με 1 τη μέγιστη ομοιότητα και 0 ότι δεν έχει καμία απολύτως ομοιότητα. Έτσι με την μέθοδο του ευρετηρίου βρίσκουμε τις πιο όμοιες περιπτώσεις με τον βαθμό ομοιότητας που αντιστοιχεί για την κάθε μια από αυτές σε σχέση πάντα με το ερώτημα που τέθηκε.

Η δεύτερη διαδικασία αφορά την μέθοδο ευρετηρίου με τη χρήση συνώνυμων όρων. Επιλέξαμε και στην συγκεκριμένη μεθοδολογία να εντοπίσουμε τους συνώνυμους όρους του ερωτήματος για να δώσουμε βαρύτητα στη σημασιολογική ερμηνεία των όρων. Τα βήματα εκτέλεσης της συγκεκριμένης διαδικασίας είναι παρόμοια με αυτά της μεθόδου του ευρετηρίου με τη χρήση συνώνυμων όρων, με τη διαφορά ότι στο τελικό ερώτημα που θα ελεγχτεί με όλες τις περιπτώσεις δεν περιλαμβάνονται οι όροι του αρχικού ερωτήματος αλλά μόνο η συνώνυμοί τους.

Αρχικά δημιουργούνται τα Lucene Document από τη βάση των λέξεων-κλειδιών, και πραγματοποιείται η διαδικασία της προ-επεξεργασίας σε αυτά. Έπειτα κατασκευάζεται ο ανεστραμμένος κατάλογος από τους όρους που προκύπτουν. Σε ότι αφορά το ερώτημα γίνεται η διαδικασία προ-επεξεργασίας του ερωτήματος χωρίς να συμπεριληφθεί αυτή της στελέχωσης κειμένου (Stemming). Στη συνέχεια εντοπίζονται οι συνώνυμοι όροι του ερωτήματος εφαρμόζοντας το λεξικό WordNet. Στους συνώνυμους αυτούς όρους εφαρμόζεται τέλος και το τελευταίο κομμάτι για την ολοκλήρωση της προ-επεξεργασίας που περιλαμβάνει τη διαδικασία της στελέχωσης. Επιπλέον εφαρμόζεται και η συνάρτηση απάλειψης όμοιων όρων που εμφανίζονται στη πρόταση του ερωτήματος. Το τελικό ερώτημα στη μέθοδο αυτή προκύπτει μόνο από τους επεξεργασμένους συνώνυμους όρους, χωρίς να συμπεριλαμβάνονται και οι αρχικοί όροι του ερωτήματος. Ο αλγόριθμος ομοιότητας εφαρμόζεται χρησιμοποιώντας τον ανεστραμμένο κατάλογο με τις πληροφορίες της βάσης και το τελικό ερώτημα. Σε κάθε ομάδα λέξεων-κλειδιών αποδίδεται μια βαθμολογία που περιγράφει την ομοιότητα του με το τελικό ερώτημα. Η βαθμολογημένη λίστα ταξινομείται με βάση τη βαθμολογία ομοιότητας που τους αποδόθηκε από τη συνάρτηση ομοιότητας. Τέλος πραγματοποιείται η αντιστοίχιση των λέξεων-κλειδιών με τις περιπτώσεις

Εφόσον έχει υπολογιστεί η ομοιότητα του αρχικού ερωτήματος και των συνωνύμων τους με όλες τις ομάδες λέξεων κλειδιών, μπορούμε να υπολογίσουμε την τελική βαθμολογία ομοιότητας για την κάθε μία περίπτωση. Σκοπός της συγκεκριμένης μεθοδολογίας είναι να συνδυάσει το βαθμό ομοιότητας, όταν το ερώτημα αποτελείται από τους επεξεργασμένους αρχικούς όρους του ερωτήματος, και το βαθμό ομοιότητας όταν το ερώτημα αποτελείται μόνο

από τους συνώνυμους όρους του αρχικού ερωτήματος. Η διαδικασία με την οποία θα πραγματοποιηθεί αυτός ο συνδυασμός θα είναι με τη χρήση βαρών

Στην τελική απόφαση για την/τις περιπτώσεις που έχουν τη μεγαλύτερη ομοιότητα με το ερώτημα θα ορίσουμε κάποιο βάρος στα αποτελέσματα της πρώτης διαδικασίας και κάποιο βάρος στα αποτελέσματα της δεύτερης. Προκειμένου τα τελικά αποτελέσματα της βαθμολογίας να είναι κανονικοποιημένα στο διάστημα  $[0,1]$  και τα βάρη που θα πρέπει να ορίσουμε θα είναι στο διάστημα αυτό. Χρειάζεται η βαθμολογία των τελικών αποτελεσμάτων να ανήκει σε αυτό το διάστημα για να υπάρχει μια αντιστοιχία με τις προηγούμενες μεθοδολογίες έτσι ώστε να ολοκληρωθεί με αξιοπιστία η σύγκριση όλων των μεθοδολογιών.

Αν θεωρήσουμε  $x$  το βάρος της πρώτης διαδικασίας και  $y$  το βάρος της δεύτερης τότε θα πρέπει να ισχύει :  $x + y = 1$ . Για να προσδιοριστούν οι τιμές του  $x$  και του  $y$ , θα πρέπει να γίνουν οι κατάλληλες μετρήσεις έτσι ώστε να εντοπίσουμε τις τιμές αυτές που θα ικανοποιούν (αν είναι εφικτό) όλες τις περιπτώσεις και αν όχι όλες όσο το δυνατόν περισσότερες. Οπότε οι τιμές  $x$  και  $y$  θα προσδιοριστούν στα τελικά συμπεράσματα των μετρήσεων.

Με τη μέθοδο αυτή η τελική βαθμολογία για κάθε περίπτωση πλέον θα υπολογίζεται ως εξής:

$$V = x * w_1 + y * w_2$$

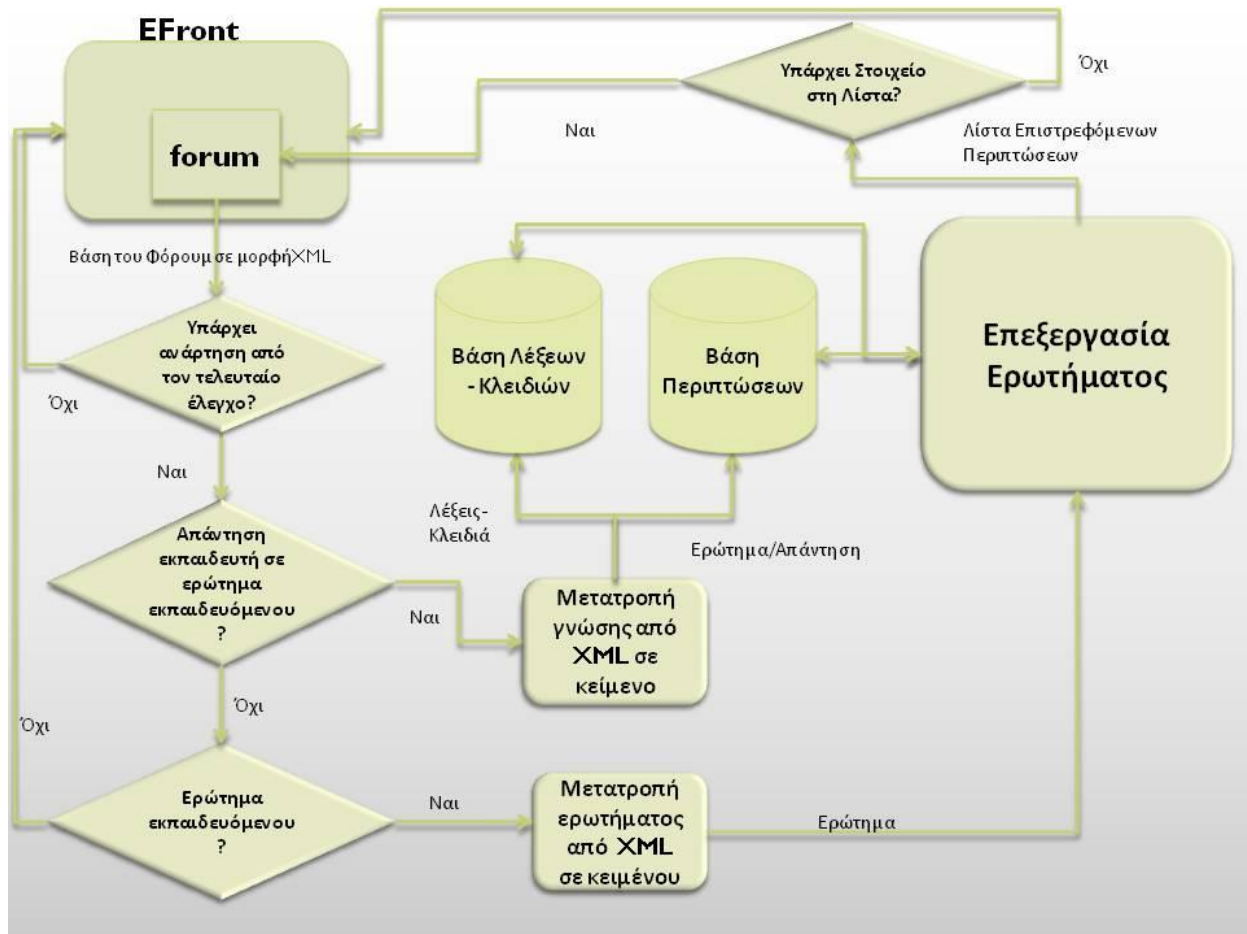
Όπου  $v$  η τελική βαθμολογία για κάθε μία πρόταση,  $x$  το βάρος για τη βαθμολογία της μεθόδου του ευρετηρίου,  $w_1$  η βαθμολογία ομοιότητας με τη μέθοδο του ευρετηρίου,  $y$  το βάρος για τη βαθμολογία της μεθόδου του ευρετηρίου με την χρήση μόνο συνώνυμων όρων,  $w_2$  η βαθμολογία ομοιότητας με τη μέθοδο του ευρετηρίου χρησιμοποιώντας μόνο συνώνυμους όρους.

## 4.8 Υλοποίηση

Στην Εικόνα 3 που ακολουθεί παρουσιάζεται η βασική σχεδίαση του συστήματος, που περιλαμβάνει όλα τα τμήματα που υλοποιήθηκαν προκειμένου να κατασκευαστεί ένα σύστημα ως ένα ενδιάμεσο εργαλείο επικοινωνίας μεταξύ εκπαιδευόμενου και εκπαιδευτή παρέχοντας υπηρεσίες εξατομικευμένης εξ αποστάσεως διδασκαλίας. Η αρχιτεκτονική που παρουσιάζεται περιγράφει την υλοποίηση της μεθοδολογίας του ευρετηρίου. Οι υπόλοιπες μεθοδολογίες έχουν

μικρές αλλαγές ως προς την υλοποίησή τους σε σχέση με την αρχιτεκτονική που παρουσιάζεται και αναφέρονται στο προηγούμενο κεφάλαιο.

Καθώς δεν υπάρχει κάποια θεματική δέσμευση με κάποιο εκπαιδευτικό αντικείμενο μπορεί να εφαρμοστεί για την ενίσχυση της διδασκαλίας σε μια πληθώρα θεματικών ενοτήτων.



Εικόνα 3 - Αρχιτεκτονική Συστήματος

Με τη συγκεκριμένη εφαρμογή είναι εφικτό να καταγράφονται παράλληλα οι απορίες των εκπαιδευόμενων έτσι ώστε να εξάγονται συμπεράσματα σχετικά με τα σημεία που παρουσιάστηκαν προβλήματα επίδοσης. Με την επέκταση αυτή θα μπορεί ο εκπαιδευτής να έχει μια πραγματική εικόνα σχετικά με τα προβλήματα κατανόησης κάποιων ενοτήτων από τους εκπαιδευόμενους, για να μπορέσει να επανεξετάσει τη μέθοδο διδασκαλίας των συγκεκριμένων ενοτήτων καθώς και το εκπαιδευτικό υλικό.

## 4.9 Efront

Το efront αποτελεί μία μοντέρνα πλατφόρμα ηλεκτρονικής μάθησης που χρησιμοποιείται για την αποδοτικότερη διεξαγωγή του μαθήματος HY-100 Εισαγωγή στην Επιστήμη Υπολογιστών. Το efront βασίζεται εξ ολοκλήρου σε τεχνολογίες Ιστού, συνεπώς για τη χρήση του δεν απαιτείται παρά μόνο ένας τυπικός web browser. Προβλέπει την ύπαρξη διακεκριμένων ρόλων χρηστών. Παρέχει εργαλεία συγγραφής περιεχομένου, καθώς και εξελιγμένες δυνατότητες δημιουργίας διαγωνισμάτων και αυτόματης διόρθωσής τους. Το υλικό μπορεί να οργανωθεί χρονικά, με βάση τον προγραμματισμό του καθηγητή, αλλά και να καθοριστεί η ροή του υλικού στους μαθητές, χρησιμοποιώντας ευέλικτους κανόνες πρόσβασης. Οι δυνατότητες επικοινωνίας της πλατφόρμας είναι εκτεταμένες, με υποστήριξη Φόρουμ, προσωπικών μηνυμάτων, Chat.

Το efront υποστηρίζει τρία είδη ρόλων: Διαχειριστής, Καθηγητής και Μαθητής. Κάθε ένας από αυτούς αντιμετωπίζεται με εντελώς διαφορετικό τρόπο και έχει στη διάθεσή του ένα ξεχωριστό περιβάλλον. Κάθε χρήστης εντάσσεται υποχρεωτικά σε ένα και μόνο ρόλο. Για την επικοινωνία εκπαιδευτή – εκπαιδευόμενου στο μάθημα της Εισαγωγής στην Επιστήμη των Υπολογιστών χρησιμοποιείται το φόρουμ. Έτσι όταν κάποιος εκπαιδευόμενος έχει κάποια απορία τη δημοσιεύει στο φόρουμ. Στο παραδοσιακό μοντέλο τις απορίες των εκπαιδευόμενων που είχαν δημοσιευτεί στο φόρουμ καλούνταν να τις λύσει ο εκπαιδευτής.

Το σύστημα που υλοποιήθηκε λαμβάνει όλες τις πληροφορίες που έχουν αναρτηθεί από τη βάση του φόρουμ από ένα αρχείο xml. Το αρχείο αυτό περιέχει όλες τις δημοσιεύσεις που έχουν αναρτηθεί στο φόρουμ αριθμώντας την κάθε μία ανάρτηση με ένα μοναδικό αύξοντα αριθμό(id) ανάλογα με την σειρά που αναρτήθηκε. Έτσι ανά πάσα χρονική στιγμή το σύστημα μπορεί να γνωρίζει μέχρι ποιον αριθμό ανάρτησης έχει ελέγξει για να μη βρεθεί σε μία ατέρμονη ακολουθία. Το αρχείο xml για κάθε μια ανάρτηση περιέχει και όλη την πληροφορία που το διέπει. Εκτός από το μοναδικό αύξοντα αριθμό, που περιέχεται στην ετικέτα(tag) “id”, περιέχει πληροφορία όπως το είδος του χρήστη που τη δημοσίευσε, τον τίτλο του θέματος κ.ά. Κάθε μία πληροφορία περιέχεται στην ανάλογη ετικέτα.

Για τη σωστή λειτουργία του συστήματος, θα πρέπει να ελέγχει όλες τις αναρτήσεις που δημοσιεύτηκαν ανά ένα χρονικό διάστημα. Το χρονικό διάστημα που ορίστηκε είναι 2 λεπτά. Η υλοποίηση της διαδικασίας αυτής έγινε με τη χρήση μίας συνάρτησης «*timer*», που ανά δύο λεπτά λαμβάνει από τη βάση όλη την πληροφορία σε ένα xml αρχείο. Κάθε φορά που λαμβάνει το καινούργιο αρχείο, ξεκινάει τη διαδικασία εκτέλεσης των βημάτων όπως παρουσιάζεται και στην Εικόνα 3.

Για να γίνει πιο σαφές στην Εικόνα 4 που ακολουθεί προβάλλεται ένα στιγμιότυπο του φόρουμ μια δεδομένη στιγμή, που περιλαμβάνει δύο ερωτήσεις που έχουν αναρτηθεί από τους εκπαιδευόμενους και στη μία από αυτές έχει δοθεί απάντηση από τον εκπαιδευτή. Οι απαντήσεις



μπορεί να δίνονται με δύο τρόπους είτε με απλή απάντηση είτε με απάντηση σε παράθεση όπως φαίνεται και στο παράδειγμα που ακολουθεί. Η διαφορά της μίας από την άλλη μέθοδο είναι ότι η μία περιλαμβάνει και το κείμενο της ερώτησης που δίνεται η απάντηση ενώ η άλλη όχι. Το σύστημα προβλέπει και τις δύο αυτές μεθόδους.



Εικόνα 4 - Στιγμιότυπο του Φόρουμ στο Efront

Στο φόρουμ του μαθήματος μπορεί να υπάρξει παραπάνω από ένα θέμα. Μια απορία κάποιος εκπαιδευόμενος τη θέτει σε κάποιο θέμα που υπάρχει ήδη ή δημιουργεί ένα καινούργιο που να σχετίζεται με την απορία του και τη δημοσιεύει εκεί. Κάθε θέμα του φόρουμ έχει έναν τίτλο που ορίζεται κατά τη διάρκεια δημιουργίας του και δηλώνει τη γενική περιγραφή του. Ανεξάρτητα με πόσα θέματα έχει το φόρουμ και πόσες αναρτήσεις έχουν πραγματοποιηθεί σε κάθε ένα από αυτά, το xml αρχείο που λαμβάνει το σύστημα από τη βάση του Efront είναι ένα, και η σειρά των αναρτήσεων που περιέχει το αρχείο αφορά αποκλειστικά τη χρονική σειρά με την οποία αναρτήθηκαν. Κάθε μία καταχώρηση στο xml αρχείο περιγράφει μία ανάρτηση που έχει πραγματοποιηθεί, συμπεριλαμβάνοντας το θέμα στο οποίο περιλαμβάνεται η ανάρτηση.

Έτσι για την πληροφορία που έχει καταγραφεί στο φόρουμ για το παραπάνω παράδειγμα, το xml αρχείο που θα ανακτηθεί από τη βάση έχει τη μορφή που εμφανίζεται στην Εικόνα 5. Κάθε μία ανάρτηση αποτελεί ένα ξεχωριστό element με τίτλο "f\_messages" για το αρχείο xml. Στο κύριο αυτό element συμπεριλαμβάνονται όλα τα sub-element που είναι:

- id - Η ετικέτα «id» όπως έχουμε ήδη αναφέρει αφορά έναν μοναδικό αύξοντα αριθμό για κάθε μία ανάρτηση.
- f\_topics\_ID – Είναι ο αριθμός που αντιστοιχεί στο θέμα που περιλαμβάνεται η συγκεκριμένη ανάρτηση.
- title – ο τίτλος της ανάρτησης.
- body – Το κείμενο της ανάρτησης.
- timestamp – Χρονική σήμανση.
- users\_LOGIN – το όνομα χρήστη του συγγραφέα της ανάκτησης

```

- <f_messages>
  <id>17</id>
  <f_topics_ID>7</f_topics_ID>
  <title>List</title>
  <body><p>What is a list ?</p></body>
  <timestamp>1286703217</timestamp>
  <users_LOGIN>filiop</users_LOGIN>
  <replyto>0</replyto>
  <rank>0</rank>
</f_messages>
- <f_messages>
  <id>18</id>
  <f_topics_ID>7</f_topics_ID>
  <title>Using Lists as Stacks</title>
  <body><p>Can i use lists as stacks?</p></body>
  <timestamp>1286703309</timestamp>
  <users_LOGIN>filiop</users_LOGIN>
  <replyto>0</replyto>
  <rank>0</rank>
</f_messages>
- <f_messages>
  <id>19</id>
  <f_topics_ID>7</f_topics_ID>
  <title>Re: List</title>
  <body><p>[quote]What is a list ?[/quote]</p> <p>Python knows a number of <em>compound</em> data types, used to group together other values. The most versatile is the <em>list</em>, which can be written as a list of comma-separated values (items) between square brackets. List items need not all have the same type.</p> <div class="highlight-python"> <div class="highlight"> <pre><span class="gp">&gt;&gt;&gt;</span><span class="s">'spam'</span><span class="p">,</span><span class="s">'eggs'</span><span class="p">,</span><span class="mi">1234</span><span class="p">]</span><span class="mi">100</span><span><span class="p">,</span><span class="p">,</span><span class="mi">1234</span><span class="p">]</span><span class="gp">&gt;&gt;&gt;</span><span class="n">a</span><span class="go">['spam', 'eggs', 100, 1234]</span></pre></div></div></body>
  <timestamp>1286703397</timestamp>
  <users_LOGIN>professor</users_LOGIN>
  <replyto>0</replyto>
  <rank>0</rank>
</f_messages>

```

Εικόνα 5 - Αρχείο xml που εξαγεται από τη βάση του φόρουμ

## 4.10 Επεξεργασία και Μετατροπή του αρχείου xml

Έχοντας το σύστημα ανακτήσει το αρχείο xml από τη βάση του φόρουμ, εντοπίζει όλες τις αναρτήσεις που πραγματοποιήθηκαν στο διάστημα από τον προηγούμενο έλεγχο. Η επιλογή των αναρτήσεων γίνεται βάσει του μοναδικού αύξοντα αριθμού που αποτελεί και την ταυτότητα για κάθε ανάκτηση. Κάθε φορά που ολοκληρώνεται η διαδικασία, θα αποθηκεύεται ο μεγαλύτερος αριθμός έτσι ώστε στη επόμενη ανάκτηση του αρχείου xml να ξεκινάει ο έλεγχος των αναρτήσεων από τον

ακριβώς επόμενο μεγαλύτερο. Ο αριθμός αυτός κάθε φορά θα προσδιορίζει την τελευταία ανάρτηση που ελέγχτηκε από το σύστημα.

Ποιο συγκεκριμένα από το αρχείο xml ελέγχουμε μόνο τις ανακτήσεις που η ταυτότητά τους είναι μεγαλύτερη από αυτή που είχαμε ελέγξει την προηγούμενη φορά. Την πρώτη φορά η μεταβλητή που κρατάει αυτή την πληροφορία είναι αρχικοποιημένη με την τιμή μηδέν. Στην περίπτωση που δεν έχει πραγματοποιηθεί κάποια ανάκτηση στο χρονικό αυτό διάστημα το σύστημα τερματίζει τη λειτουργία του.

Στην αντίθετη περίπτωση πραγματοποιούνται δύο βασικοί έλεγχοι:

- Αν υπάρχει ανάρτηση από κάποιον εκπαιδευόμενο χωρίς να έχει δοθεί απάντηση από τον εκπαιδευτή.
- Αν υπάρχει απάντηση(ανάρτηση) από τον καθηγητή σε κάποια απορία του εκπαιδευόμενου.

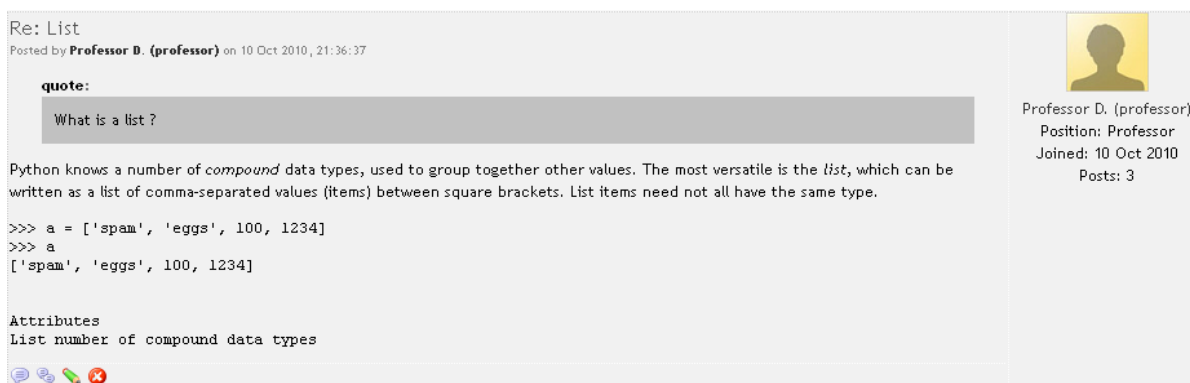
Για την εφαρμογή των ελέγχων θα πρέπει να εντοπιστεί ο ρόλος του χρήστη που είναι ιδιοκτήτης της ανάρτησης. Από το xml αρχείο γνωρίζουμε μόνο το "users\_LOGIN" δηλαδή το όνομα χρήστη για τους ιδιοκτήτες των αναρτήσεων. Στη βάση του Efront υπάρχει ο πίνακας «users» που περιέχει όλους τους χρήστες του συστήματος και όλες τις πληροφορίες για αυτούς, μαζί με τα users\_LOGIN και το ρόλο τους. Έτσι από τον πίνακα «users» βρίσκουμε το ρόλο όλων των χρηστών που αντιστοιχούν στο «users\_LOGIN» που είναι γνωστό από το αρχείο xml και μοναδικό για τον καθένα χρήστη.

Εφόσον γνωρίζουμε τους ρόλους των χρηστών που είναι ιδιοκτήτες των αναρτήσεων που ελέγχουμε, πραγματοποιούνται οι δύο έλεγχοι. Αν ο πρώτος έλεγχος είναι αληθής τότε η ανάρτηση/σεις αποτελεί το *ερώτημα* του συστήματος, και γίνεται η μετατροπή του από xml σε κείμενο. Από τη μετατροπή αυτή αποθηκεύεται όλη η χρήσιμη γνώση και η υπόλοιπη αγνοείται. Χρήσιμη γνώση για το ερώτημα που αποτελεί είσοδο του συστήματος αρχικά είναι το κείμενό της που περιέχεται στην ετικέτα «body», πέρα όμως από το κείμενο αποθηκεύονται σ' έναν πίνακα και ο αριθμός του θέματος καθώς και ο τίτλος δηλαδή η πληροφορία που περιέχεται στις ετικέτες f\_topics\_ID, title. Αυτή η πληροφορία θα χρησιμοποιηθεί εφόσον ολοκληρωθεί η εύρεση της πιο κατάλληλης περίπτωσης(αν υπάρχει) για να επιστραφεί ως απάντηση στο φόρουμ του efront.

Αν ο δεύτερος έλεγχος είναι αληθής τότε η πληροφορία αυτή αποτελεί γνώση του συστήματος και πρέπει να αποθηκευτεί στη βάση. Έτσι μετατρέπεται η πληροφορία αυτή σε κείμενο κρατώντας δύο στοιχεία. Τα στοιχεία αυτά είναι το κείμενο της ερώτησης του εκπαιδευόμενου, και το κείμενο της ανάρτησης με την απάντηση που έχει δοθεί από τον εκπαιδευτή δημιουργώντας έτσι μία περίπτωση(case). Στο κείμενο της απάντησης του καθηγητή θα περιλαμβάνονται και οι λέξεις κλειδιά της περίπτωσης, γιατί όπως αναφέρθηκε θα τις

προσδιορίζει ο εκπαιδευτής αφού ολοκληρώνει την απάντησή του. Η διαδικασία που ορίστηκε έτσι ώστε να γίνεται ο εντοπισμός τους από το σύστημα είναι μία σύμβαση με τον εκπαιδευτή. Μετά την ολοκλήρωση της απάντησής του ο καθηγητής θα πληκτρολογεί όλες τις λέξεις-κλειδιά της περίπτωσης αυτής κάτω από έναν τίτλο που ονομάζεται “Attributes” όπως παρουσιάζεται και στην Εικόνα 6 που ακολουθεί.

Με τον τρόπο αυτό το σύστημα θα διαχωρίζει την απάντηση του καθηγητή σε δύο τμήματα, το πρώτο θα αφορά την απάντησή του για την ερώτηση που αναφέρεται, και το δεύτερο τις λέξεις κλειδιά που χαρακτηρίζουν την περίπτωση. Η ερώτηση του εκπαιδευόμενου με το πρώτο τμήμα της απάντησης του εκπαιδευτή ορίζουν μια περίπτωση, που αποτελεί γνώση για το σύστημα και αποθηκεύεται στη βάση περιπτώσεων. Το δεύτερο τμήμα της απάντησης αποτελεί τα χαρακτηριστικά της περίπτωσης αυτής και αποθηκεύεται σε αντίστοιχη θέση στη βάση περιπτώσεων.



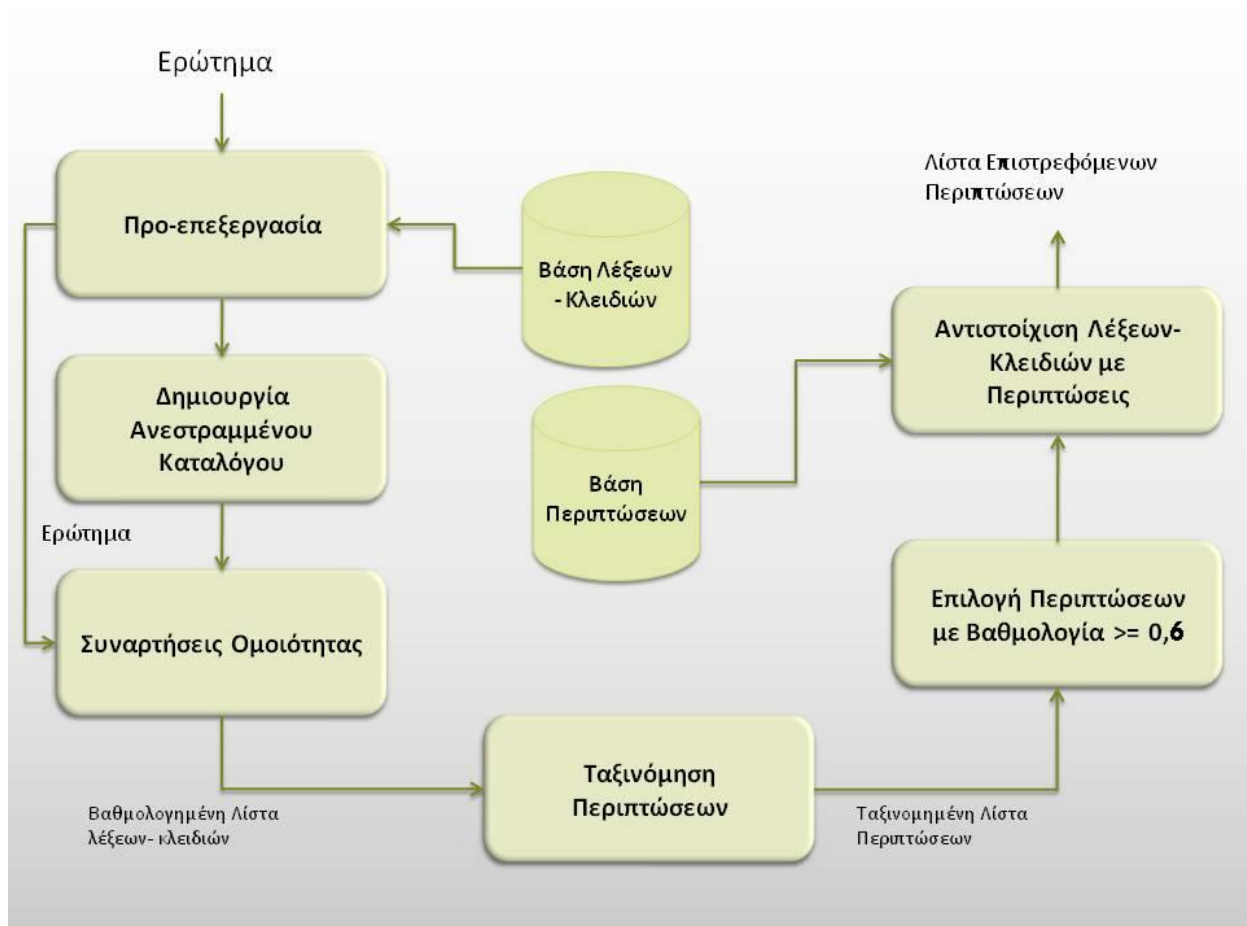
Εικόνα 6 - Στιγμιότυπο Φόρουμ προσθέτοντας τις Λέξεις-Κλειδιά

#### 4.11 Επεξεργασία Ερωτήματος

Το επόμενο βήμα για τον υπολογισμό της ιδανικής περίπτωσης αφού έχει εντοπιστεί το ερώτημα, είναι η συνολική επεξεργασία του ερωτήματος. Η αρχιτεκτονική της επεξεργασίας του ερωτήματος παρουσιάζεται στην Εικόνα 7 που ακολουθεί. Η βάση ουσιαστικά αποτελείται από ένα αρχείο κειμένου, που περιέχει όλη την πληροφορία που έχει καταχωρηθεί. Στο πρώτο στάδιο θα πρέπει να ανακτηθεί η γνώση που υπάρχει στη βάση των λέξεων-κλειδιών καθώς αυτή η πληροφορία θα πρέπει να συγκριθεί με το ερώτημα ως προς την ομοιότητά τους. Οπότε φορτώνεται το αρχείο με τη χρήση της συνάρτησης CasesLoad που παίρνει σαν όρισμα το μονοπάτι(path) που είναι αποθηκευμένο το αρχείο και επιστρέφει μία λίστα με Lucene Document που το καθένα από αυτά έχει πληροφορία για κάθε μία περίπτωση. Έτσι η γνώση της βάσης έχει ανακτηθεί και αποθηκευτεί σε Lucene Document. Στην περίπτωση της στατιστικής μεθοδολογίας το αρχείο που φορτώνεται είναι αυτό που έχει αποθηκευμένες της περιπτώσεις(cases), και η υπόλοιπη διαδικασία εκτελείται κανονικά.

Το επόμενο βήμα είναι η προ-επεξεργασία του κειμένου, τόσο του ερωτήματος όσο και των λέξεων-κλειδιών. Η προ-επεξεργασία του κειμένου περιλαμβάνει τα βήματα της Λεξικολογικής Ανάλυσης, τον Αποκλεισμό Λέξεων(stopwords) και τη στελέχωση κειμένου(stemming). Στη μεθοδολογία του ευρετηρίου με τη χρήση συνώνυμων όρων πριν από τη στελέχωση κειμένου του ερωτήματος μεσολαβεί η διαδικασία εντοπισμού όλων των συνώνυμων όρων του ερωτήματος όπως προκύπτουν μετά τη διαδικασία λεξικολογικής ανάλυσης και αποκλεισμού λέξεων, με τη χρήση του σημασιολογικού λεξικού WordNet. Μετά την προ-επεξεργασία το κείμενο είναι πολύ μικρότερο σε έκταση καθώς έχουν αφαιρεθεί οι μη-σημαντικές λέξεις όπως άρθρα, κτητικές αντωνυμίες, επίσης τα σημεία στίξης, παύλες, και τέλος όλες οι λέξεις μετατρέπονται στη ρίζα τους με τη διαδικασία του stemming. Έτσι η σύγκριση των λέξεων κλειδιών με το ερώτημα για τον εντοπισμό της ομοιότητάς του μπορεί να είναι αποτελεσματικότερη και πιο έγκυρη.

Προκειμένου να μη γίνεται η αναζήτηση των όρων του ερωτήματος σειριακά σε κάθε περίπτωση των λέξεων κλειδιών, δημιουργείται ένας ανεστραμμένος κατάλογος. Όπως περιγράφεται αναλυτικά και στο προηγούμενο κεφάλαιο, ο ανεστραμμένος κατάλογος δημιουργείται από τους όρους των λέξεων κλειδιών όπως αυτοί προκύπτουν από την προ-επεξεργασία. Με τον ανεστραμμένο κατάλογο αντί να αποθηκεύουμε πληροφορία για το ποιοι όροι περιέχονται σε ένα έγγραφο, καταγράφουμε το σύνολο των εγγράφων που περιέχουν ένα συγκεκριμένο όρο. Πέρα από τα έγγραφα που περιέχουν τον όρο, αποθηκεύεται και το πλήθος των εμφανίσεων στο καθένα, καθώς και το συνολικό πλήθος εμφανίσεων σε όλα τα έγγραφα. Με αυτό τον τρόπο η εφαρμογή των συναρτήσεων ομοιότητας του ερωτήματος με όλες τις λέξεις κλειδιά των περιπτώσεων εκτελείται πιο γρήγορα, εύκολα και αποδοτικά.



Εικόνα 7 - Αρχιτεκτονική Επεξεργασίας Ερωτήματος

Ο ανεστραμμένος κατάλογος και το ερώτημα αποτελούν τα βασικά ορίσματα για την εκτέλεση της συνάρτησης ομοιότητας. Η συνάρτηση ομοιότητας στόχο έχει να αποδώσει τιμές σε όλο το σύνολο των λέξεων κλειδιών σχετικά με την ομοιότητά τους με το ερώτημα. Η τιμή αυτή ορίζεται και ως βαθμολογία της κάθε περίπτωσης, και ανήκει στο κλειστό διάστημα  $[0,1]$ . Ο Τρόπος χρήσης του ανεστραμμένου καταλόγου από τη συνάρτηση εξαρτάται από το ερώτημα. Το πρώτο μοντέλο που χρησιμοποιείται είναι το Λογικό(Boolean), για να περιορίσει τις περιπτώσεις που πρέπει να βαθμολογηθούν.

Για παράδειγμα, αν το ερώτημα αποτελείται μόνο από έναν όρο, το μόνο που χρειάζεται είναι να εντοπιστεί ο όρος στον λεξικό του ανεστραμμένου καταλόγου, και έπειτα να διαβαστεί η αντίστοιχη λίστα εμφανίσεων για το συγκεκριμένο όρο. Όλες οι περιπτώσεις που δεν περιλαμβάνονται σε αυτή την λίστα βαθμολογούνται με την τιμή 0 και στη συνέχεια εφαρμόζεται το Διανυσματικό μοντέλο για να εντοπιστεί η τιμή ομοιότητας των περιπτώσεων που περιλαμβάνουν το όρο. Αν το ερώτημα περιέχει περισσότερους όρους τότε αναπαριστάται με τον τελεστή διάζευξης, που ενώνει όλους του όρους του.

Έστω ότι έχουμε το προηγούμενο παράδειγμα που χρησιμοποιήθηκε για τον ορισμό του ανεστραμμένου καταλόγου, και θεωρήσουμε ότι οι προτάσεις που ακολουθούν αποτελούν τη βάση Λέξεων Κλειδιών:

*d1* – How do I create class, method and static class

*d2* – How do I call method

*d3* – How can I organize class

Ο ανεστραμμένος κατάλογος της βάσης αυτής, θα δημιουργηθεί αφού πρώτα γίνει η προεπεξεργασία του κειμένου για κάθε ομάδα λέξεων κλειδιών. Το κείμενο μετά την προεπεξεργασία της παραπάνω βάσης θα είναι:

*d1* – How creat class method static class

*d2* – How call method

*d3* – How organ class

Έτσι ο ανεστραμμένος κατάλογος θα έχει την μορφή που ακολουθεί:

Λέξεις	Περιπτώσεις
How	3: (d1,1), (d2,1), (d3,1)
creat	1: (d1,1)
class	3: (d1,2), (d3,1)
method	2: (d1,1), (d2,1)
static	1: (d1,1)
call	1: (d2,1)
organ	1: (d3,1)

Πίνακας 1 - Μορφή Ανεστραμμένου Καταλόγου

Αν θεωρήσουμε ως αρχικό ερώτημα την πρόταση « What is a class?» , η προεπεξεργασία του θα το μετατρέψει στη μορφή «What class». Σύμφωνα με το Λογικό μοντέλο το ερώτημα αυτό μπορεί να εκφραστεί ως : **ερώτημα = (what) OR (class)**. Οι σχετικές ομάδες λέξεων κλειδιών ως προς το ερώτημα είναι αυτές που περιέχουν είτε τον όρο *what* είτε τον όρο *class*. Έτσι πρώτα εντοπίζονται οι λίστες εμφανίσεων των δύο όρων και στη συνέχεια υπολογίζεται η ένωση των ομάδων αυτών. Από τον

ανεστραμμένο κατάλογο γνωρίζουμε ότι οι ομάδες που περιέχουν τη λέξη *what* είναι καμία, καθώς δεν υπάρχει στο λεξιλόγιο του καταλόγου η συγκεκριμένη λέξη. Οι ομάδες που περιέχουν τον όρο *class* είναι οι *d1*, *d3*. Άρα οι σχετικές ως προς το ερώτημα ομάδες είναι η ένωση των αποτελεσμάτων για κάθε όρο. Έτσι οι ομάδες για τις οποίες θα εφαρμοστεί το διανυσματικό μοντέλο είναι :

$$A = \{\emptyset\} \cup \{d1, d3\} = \{d1, d3\}$$

Οι ομάδες που δεν περιλαμβάνονται στη παραπάνω λίστα βαθμολογούνται με μηδέν καθώς δεν περιέχουν κανένα όρο του ερωτήματος, και κατ' επέκταση καμία ομοιότητα.

Το επόμενο βήμα για την ολοκλήρωση της απονομής βαθμολογίας των ομάδων λέξεων κλειδιών που επιλέχτηκαν από το Λογικό μοντέλο είναι η εφαρμογή του Διανυσματικού μοντέλου. Το Διανυσματικό μοντέλο όπως έχει αναλυθεί βασίζεται κυρίως στη συχνότητα εμφάνισης των όρων. Για το παραπάνω παράδειγμα η βαθμολογία των τριών ομάδων σε σχέση με το ερώτημα είναι:

*d1* – 1.0

*d2* – 0

*d3* – 0.9428

Η βαθμολογία ομοιότητας βασίζεται στον τύπο που ακολουθεί και έχει αναλυθεί σε προηγούμενο κεφάλαιο:  $score(q,d) = coord(q,d) \cdot queryNorm(q) \cdot \sum(tf(t \text{ in } d) \cdot idf(t)^2 \cdot w_{t \text{ in } q} \cdot lengthNorm(t \text{ in } q))$

Αφού η βαθμολογημένη λίστα των λέξεων κλειδιών έχει δημιουργηθεί, ταξινομείται με βάση τη βαθμολογία ομοιότητας, και στη συνέχεια επιλέγονται οι ομάδες με βαθμολογία μεγαλύτερη ή ίση του 0,6. Για το προηγούμενο παράδειγμα η λίστα αυτή θα είναι:

*d1* – 1.0

*d3* – 0.9428

Το τελευταίο βήμα για την ολοκλήρωση των τεχνικών *case based reasoning* είναι η αντιστοίχιση των ομάδων λέξεων κλειδιών με τις περιπτώσεις από τη βάση περιπτώσεων.



## 4.12 Μετρήσεις - Αξιολόγηση

Προκειμένου να αξιολογηθούν οι τέσσερις μέθοδοι που υλοποιήθηκαν, πραγματοποιήθηκαν κάποιες μετρήσεις. Οι μετρήσεις στόχο έχουν να αναδείξουν την πιο κατάλληλη μεθοδολογία για να εφαρμοστεί στο σύστημα, δηλαδή τη μεθοδολογία που θα επιφέρει τα πιο ιδανικά αποτελέσματα. Ιδανικά αποτελέσματα θεωρούνται αυτά που για ένα σύνολο ερωτημάτων σε μία δεδομένη βάση επιστρέφουν την πιο κατάλληλη περίπτωση. Αν η πιο κατάλληλη περίπτωση δεν είναι αυτή με τη μεγαλύτερη βαθμολογία, τουλάχιστον το σύνολο της βαθμολογίας που θα έχει συγκεντρώσει να είναι αρκετά μεγάλο έτσι ώστε να ανήκει σε μία λίστα περιπτώσεων που θα επιστραφούν στον χρήστη ως παρεμφερείς ή προτεινόμενες περιπτώσεις. Η ιδέα της λίστας προέκυψε από το γεγονός ότι πολλές περιπτώσεις έχουν βαθμολογία μεγάλη, και πολύ κοντά στο 1(θεωρείται η μέγιστη βαθμολογία). Οπότε για να εξασφαλιστεί ότι μια περίπτωση που έχει ανακτηθεί και έχει μεγάλη βαθμολογία μπορεί να είναι αυτή που δίνει τη λύση στο πρόβλημα που τέθηκε, επιστρέφεται ως μια παρεμφερής λύση.

Για την αντικειμενικότερη επιλογή του κατάλληλου μοντέλου από αυτά που παρουσιάστηκαν, κατασκευάστηκαν δύο βάσεις δεδομένων, ως πηγή γνώσης από δύο μαθήματα προπτυχιακού περιεχομένου σε διαφορετικές θεματικές περιοχές. Η πρώτη συσχετίζεται με το μάθημα του HY-100 Εισαγωγή στην Επιστήμη των Υπολογιστών, που διεξάγεται από το τμήμα Επιστήμης Υπολογιστών. Το μάθημα πραγματεύεται τις βασικές έννοιες της γλώσσας προγραμματισμού Python. Για το υλικό που αποτέλεσε πηγή γνώσης και αποθηκεύτηκε στη βάση, χρησιμοποιήθηκε το εκπαιδευτικό υλικό που παρέχεται από την κεντρική σελίδα(<http://python.org/>) της Python.

Στον τομέα της πληροφορικής οι ορισμοί και οι λέξεις που περιγράφουν κάποιες λειτουργίες αναπαριστούνται με συγκεκριμένες λέξεις. Οι λέξεις αυτές είναι μοναδικές και η ερμηνεία τους μπορεί να αναπαρασταθεί μόνο από κάποιον ειδικό και γνώστη του αντικειμένου. Στην προσπάθεια ερμηνείας τους και την εύρεση συνώνυμων όρων από ένα λεξικό που δεν απευθύνεται αποκλειστικά στην ορολογία αυτή μπορεί να εξάγει λανθασμένα συμπεράσματα. Για παράδειγμα η ελληνική λέξη «αναδρομή» από ένα κοινό ελληνικό σημασιολογικό λεξικό μπορεί να θεωρηθεί ως «ανασκόπηση», «διήγηση», και «αναπόληση γεγονότων» και αυτές οι λέξεις αποτελούν συνώνυμους όρους καθώς έχουν την ίδια ερμηνεία στην ελληνική γλώσσα. Στην πραγματικότητα όμως ο όρος αναδρομή στο κλάδο της πληροφορικής αναφέρεται σε συναρτήσεις οι οποίες καλούν τον εαυτό τους και προκύπτει έτσι μια αναδρομική διαδικασία.

Υπάρχουν όμως και κάποιες λέξεις που μπορούν να χρησιμοποιηθούν στον τρόπο αναπαράστασης μίας ερώτησης που να μη συσχετίζονται με την ορολογία και οι συνώνυμες έννοιές τους να προσφέρουν μία σημαντική πληροφορία ως προς την έννοια που αντικατοπτρίζει το κείμενο.

Όπως έχει αναφερθεί οι δύο από τις τέσσερις μεθόδους που αναπτύχθηκαν κάνουν χρήση του σημασιολογικού λεξικού WordNet, το λεξικό αυτό αφορά την αγγλική γλώσσα και δεν συσχετίζεται με ορολογίες θεματικών ενοτήτων όπως στην προκειμένη περίπτωση της πληροφορικής. Από την αναζήτηση που πραγματοποιήθηκε δεν εντοπίστηκε κάποιο λεξικό που να είναι διαθέσιμο για την ορολογία της πληροφορικής. Η ιδέα της υλοποίησης του συστήματος που προτείνεται στα πλαίσια της παρούσας εργασίας δεν έχει στόχο τον περιορισμό της χρήσης μόνο σε ένα συγκεκριμένο αντικείμενο, αλλά να μπορεί να εφαρμοστεί σε οποιοδήποτε μάθημα ανεξαρτήτως αντικειμένου για την βελτίωση της επικοινωνίας του εκπαιδευτικού προσωπικού με τους εκπαιδευόμενους. Από την άλλη το σημασιολογικό λεξικό WordNet που είναι το πιο διαδεδομένο, μπορεί να χρησιμοποιηθεί για την εύρεση συνώνυμων όρων που δεν αποτελούν λέξεις ορολογίας, έτσι και αλλιώς οι όροι μια ερώτησης από κάποιον εκπαιδευόμενο δεν αποτελούνται μόνο από όρους που ανήκουν στην ορολογία του αντικειμένου. Φυσικά το κατά πόσο είναι αποδοτική η χρήση του σε ένα αντικείμενο που περιέχει δική του ορολογία θα αποδειχθεί από τις μετρήσεις που πραγματοποιήθηκαν.

Εφόσον η μία βάση που δημιουργήθηκε αποτελεί μέρος ενός αντικειμένου που εξαρτάται από την ορολογία του χώρου της πληροφορικής, αποφασίστηκε να δημιουργηθεί μία δεύτερη βάση που δεν θα συσχετίζεται με κάποια ειδική ορολογία του αντικειμένου, για να εξαχθούν αντικειμενικά συμπεράσματα σε σχέση με τις μεθοδολογίες που υλοποιήθηκαν. Έτσι η δεύτερη βάση που δημιουργήθηκε έχει να κάνει με μάθημα εισαγωγικού περιεχομένου του προπτυχιακού προγράμματος που σχετίζεται με τον τομέα της ψυχολογίας. Το υλικό που χρησιμοποιήθηκε για να δημιουργηθεί η βάση αυτή, είναι από ένα εισαγωγικό μάθημα στις βασικές έννοιες της ψυχολογίας.

Η αξιολόγηση των τεσσάρων μοντέλων πραγματοποιήθηκε χρησιμοποιώντας τις δύο βάσεις σε ένα πλήθος ερωτήσεων για να εντοπιστεί η αποτελεσματικότητα της κάθε μιας μεθοδολογίας, και εν τέλει να εντοπιστεί η πιο αποδοτική. Το μοντέλο που θα ικανοποιήσει το πλήθος ερωτήσεων που πραγματοποιήθηκε και στα δύο αντικείμενα, θα μπορέσει να χρησιμοποιηθεί σε όλα τα θεματικά πεδία διδασκαλίας. Επίσης να αναφέρουμε ότι το σύστημα σχεδιάστηκε με τέτοιο τρόπο έτσι ώστε να λειτουργεί αυτόνομα. Έτσι μπορεί να εφαρμοστεί εύκολα και σε άλλες πλατφόρμες, καθώς και σε άλλα είδη επικοινωνίας αρκεί να υπάρχει κάπου καταγεγραμμένη η πληροφορία (όπως σε μία βάση σε ένα mail Server κ.ά.) και να μπορούμε να έχουμε πρόσβαση από το σύστημα σε αυτή.

#### **4.13 Εφαρμογή στον κλάδο της Πληροφορικής**

Η πρώτη πηγή γνώσης που χρησιμοποιήθηκε είναι από τον κλάδο της πληροφορικής. Πιο συγκεκριμένα έχει να κάνει με τις βασικές έννοιες και δομές για τη γλώσσα προγραμματισμού Python. Επιλέχτηκε η συγκεκριμένη γλώσσα γιατί είναι αυτή που διδάσκεται στο μάθημα του HY-100 στην Επιστήμη των Υπολογιστών και επειδή υπάρχει ένα οργανωμένο υλικό στο διαδίκτυο και πιο συγκεκριμένα στη βασική σελίδα της Python για όλες τις βασικές έννοιες και δομές για τη διεξαγωγή ενός μαθήματος. Από το υλικό αυτό προέκυψε μία βάση με περίπου εκατό περιπτώσεις. Κάθε μία περίπτωση σύμφωνα με την αρχιτεκτονική του συστήματος αποτελείται από την ερώτηση(το πρόβλημα που τίθεται) και την απάντηση(τη λύση που έχει δοθεί). Οι περιπτώσεις που δημιουργήθηκαν έχουν ως πρώτο μέρος τον ορισμό του προβλήματος, και δεύτερο όλη την πληροφορία σχετικά με τον ορισμό. Η βάση βασίζεται σε εκπαιδευτικό υλικό και ουσιαστικά απεικονίζει ένα μεγάλο κομμάτι της θεωρίας με τη μορφή ερωτήσεων απαντήσεων.

Για το υλικό αυτό δημιουργήθηκαν είκοσι δύο ερωτήσεις που καλύπτουν όλα τα ενδεχόμενα. Δηλαδή για κάποιες από τις ερωτήσεις υπάρχει υλικό αποθηκευμένο για την ακριβή επίλυσή τους, για κάποιες άλλες μπορεί να λυθεί ένα μέρος του προβλήματος, σε σχέση πάντα με το υλικό που αποτελεί γνώση, και κάποιες άλλες που δεν υπάρχει γνώση που να συσχετίζεται με το ερώτημα. Για όλα αυτά τα ερωτήματα θα δούμε πώς ανταποκρίνεται η κάθε μεθοδολογία, για την εξαγωγή συμπερασμάτων για κάθε μία από αυτές.

Στις τρεις από τις μεθοδολογίες όπως έχουμε δει χρησιμοποιείται η έννοια του ευρετηρίου, με τη χρήση λέξεων-κλειδιών τα οποία αποτελούν τα χαρακτηριστικά για κάθε περίπτωση και αποτελούν τη δεικτοδότηση του βασικού κειμένου. Έτσι έχει κατασκευαστεί μία παράλληλη βάση με την αρχική που περιέχει όλα τα χαρακτηριστικά(λέξεις – κλειδιά) για κάθε μια περίπτωση. Για το σύστημα τις λέξεις-κλειδιά θα τις ορίζει ο καθηγητής, ο οποίος θεωρείται ο καταλληλότερος για την απεικόνιση της κάθε περίπτωσης με τα χαρακτηριστικά τους.

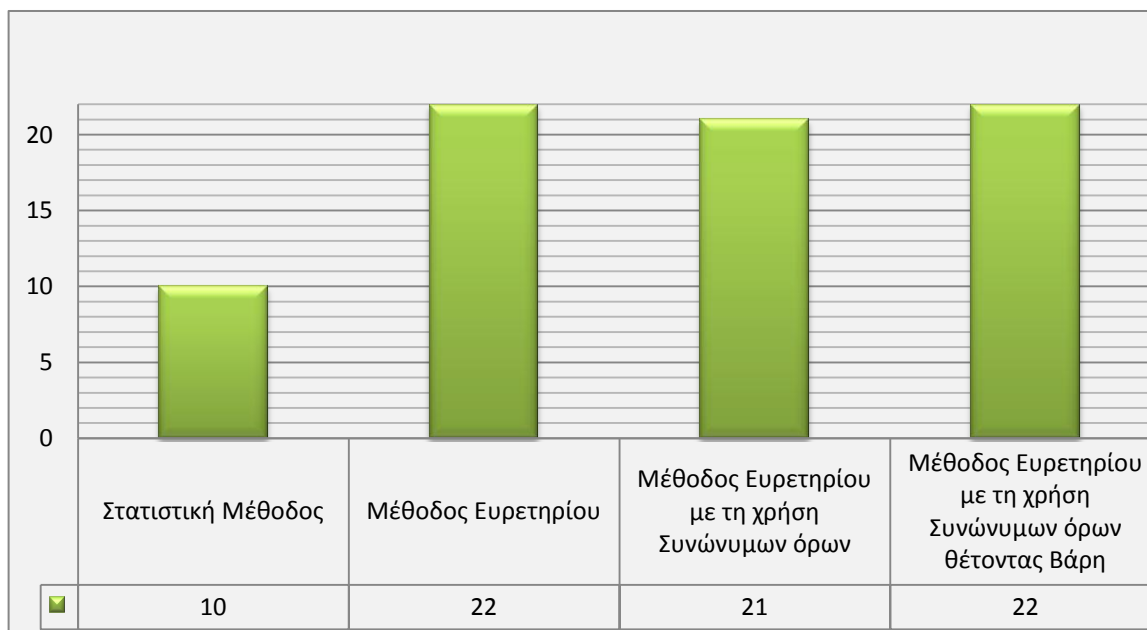
Δημιουργήθηκαν δύο βάσεις, η βάση περιπτώσεων και η βάση λέξεων κλειδιών, με το ίδιο πλήθος στοιχείων στην κάθε μία. Οι βάσεις αυτές χρησιμοποιήθηκαν και από τις τέσσερις μεθοδολογίες που δημιουργήθηκαν και σχεδιάστηκαν με αυτό το σκοπό. Για όλες τις μεθοδολογίες εξετάστηκαν οι ίδιες ερωτήσεις. Έτσι μπορούν να εξαχθούν αντικειμενικά και αντιπροσωπευτικά συμπεράσματα για κάθε μεθοδολογία σε σχέση με τη συμπεριφορά της στη συνολική εύρεση της πιο ιδανικής περίπτωσης.

#### **4.14 Αξιολόγηση Μεθοδολογιών στον κλάδο της Πληροφορικής**

Ο έλεγχος των μεθοδολογιών στη δεδομένη βάση που έχει κατασκευαστεί θα πραγματοποιηθεί με τη χρήση είκοσι δύο ερωτήσεων. Οι ερωτήσεις αυτές ανήκουν στην ίδια θεματολογία και καλύπτουν όλες τις πιθανές περιπτώσεις που μπορεί να προκύψουν. Για κάποιες από τις ερωτήσεις υπάρχει καταχωρημένη όλη η πληροφορία για την επίλυση τους, και το προσδοκώμενο αποτέλεσμα από όλες τις μεθοδολογίες είναι να εντοπιστεί και να επιστραφεί η γνώση αυτή για την επιτυχή λύση του προβλήματος.

Για ένα άλλο πλήθος ερωτήσεων δεν υπάρχει ολοκληρωμένη πληροφορία για αυτές αλλά για την επίλυση ενός μέρους του προβλήματος. Και σε αυτή την κατηγορία ερωτήσεων στόχος είναι ο έλεγχος συμπεριφοράς των μεθοδολογιών για να καλύψουν το μεγαλύτερο μέρος του προβλήματος με την γνώση που τους διατίθεται. Τέλος υπάρχει και ένα μικρό πλήθος ερωτήσεων για το οποίο δεν υπάρχει κάποια γνώση για να μπορέσει να πραγματοποιηθεί η επίλυση τους.

Με αυτά τα είδη των ερωτήσεων πραγματοποιήθηκε ο έλεγχος της συμπεριφοράς των μεθοδολογιών που έχουμε περιγράψει. Για κάθε μία μεθοδολογία ελέγχθηκαν όλες οι ερωτήσεις με την ίδια βάση. Τα συνολικά αποτελέσματα παρουσιάζονται στο διάγραμμα που ακολουθεί. Πιο συγκεκριμένα προβάλλεται το πλήθος των ερωτήσεων για τα οποία η απάντηση που επιστράφηκε από κάθε μεθοδολογία ήταν η πιο κατάλληλη από τη βάση γνώσης.



Εικόνα 8- Αποτελέσματα Αξιολόγησης για τη Βάση στην Επιστήμη Υπολογιστών

Στόχος από τη διαδικασία της αξιολόγησης των τεσσάρων αυτών μεθοδολογιών είναι ο εντοπισμός αυτής που θα ικανοποιήσει όλες τις ερωτήσεις. Ακόμα και η ιδανικότερη απάντηση

να μην είναι πάντα η πρώτη που επιστρέφεται αρκεί να είναι μέσα στο πλήθος των αποτελεσμάτων που θα επιστραφούν στο κάτοχο της ερώτησης, έτσι ώστε να εξασφαλιστεί ότι η χρήσιμη πληροφορία θα εντοπιστεί και θα επιστραφεί. Από τη βάση που χρησιμοποιήθηκε στα πλαίσια της θεματικής ενότητας της Επιστήμης των Υπολογιστών συμπεραίνουμε ότι μόνο η μέθοδος του ευρετηρίου και η μέθοδος του ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη κατάφεραν να ικανοποιήσουν όλες τις ερωτήσεις. Αυτό που απομένει να δούμε είναι τη συμπεριφορά κυρίως αυτών των δύο μεθοδολογιών στη δεύτερη βάση που έχει κατασκευαστεί στον τομέα της ψυχολογίας που είναι ένα αντικείμενο το οποίο δεν χρησιμοποιεί ορολογία, έτσι ώστε να εξαχθούν ολοκληρωμένα συμπεράσματα.

#### 4.15 Εφαρμογή στον κλάδο της Ψυχολογίας

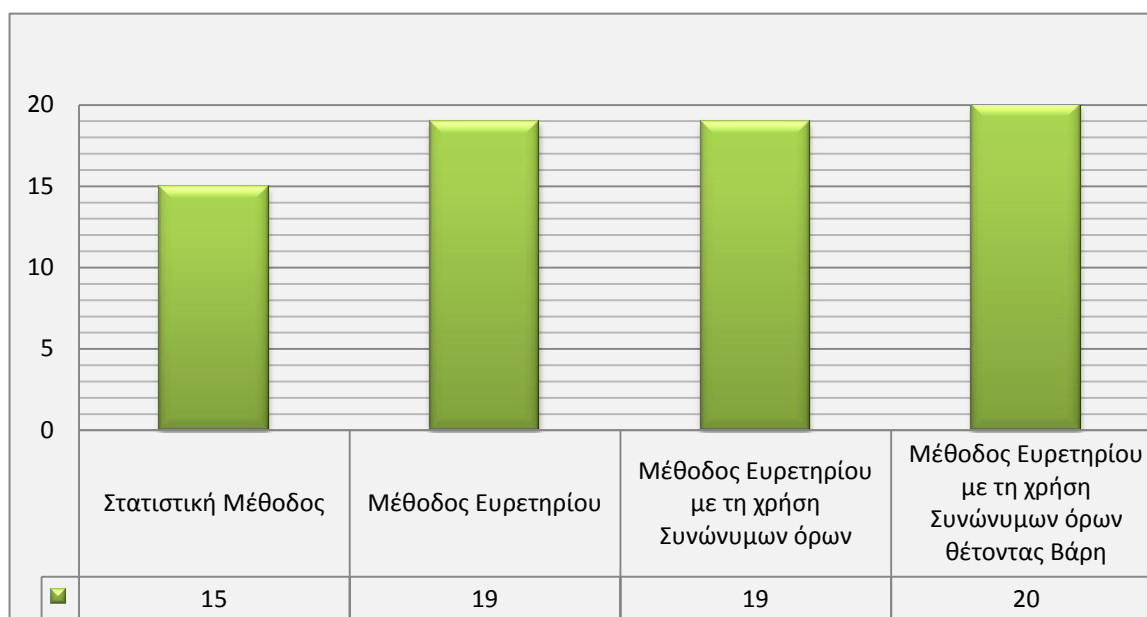
Η δεύτερη πηγή γνώσης που χρησιμοποιήθηκε είναι από τον τομέα της ψυχολογίας. Πιο συγκεκριμένα έχει να κάνει με τις βασικές έννοιες, και θεραπευτικές προσεγγίσεις της, όπως αυτές περιγράφονται από ένα ηλεκτρονικό μάθημα που έχει κατασκευαστεί μέσα σε ένα σύνολο πληροφοριών για την παροχή ολοκληρωμένης πληροφορίας στο τομέα αυτό. Επιλέχτηκε η συγκεκριμένη πηγή γνώσης καθώς προσφέρει ένα ολοκληρωμένο σύνολο πληροφοριών μέσα σε ένα οργανωμένο περιβάλλον που διαχειρίζονται εγκεκριμένοι ψυχολόγοι. Από το υλικό αυτό προέκυψε μία βάση με περίπου εκατό περιπτώσεις, που η κάθε μία αποτελείται από τον ορισμό του προβλήματος(ερώτηση) και τη λύση της(απάντηση).

Για το υλικό αυτό δημιουργήθηκαν είκοσι ερωτήσεις που καλύπτουν όλα τα πιθανά ενδεχόμενα που μπορεί να προκύψουν. Δηλαδή για ένα πλήθος ερωτήσεων υπάρχει όλη η πληροφορία για την ακριβή επίλυσή τους, για κάποιες άλλες μπορεί να λυθεί ένα μέρος του προβλήματος, και κάποιες άλλες που δεν υπάρχει γνώση που να συσχετίζεται με το ερώτημα. Για όλα αυτά τα ερωτήματα θα δούμε πώς ανταποκρίνεται η κάθε μεθοδολογία, για την εξαγωγή συμπερασμάτων για κάθε μία από αυτές.

Και στον κλάδο της ψυχολογίας όπως και τον τομέα της πληροφορικής δημιουργήθηκαν δύο βάσεις, η βάση περιπτώσεων και η βάση λέξεων κλειδιών, με το ίδιο πλήθος στοιχείων στην κάθε μία. Οι βάσεις αυτές χρησιμοποιήθηκαν από τις τέσσερις μεθοδολογίες που δημιουργήθηκαν και σχεδιάστηκαν γι' αυτό το σκοπό. Για όλες τις μεθοδολογίες εξετάστηκαν οι ίδιες ερωτήσεις. Στόχος είναι η εξαγωγή αντικειμενικών συμπερασμάτων ώστε να επιλεγεί η πιο κατάλληλη μέθοδος για την εφαρμογή της στο σύστημα.

#### 4.16 Αξιολόγηση Μεθοδολογιών στον τομέα της Ψυχολογίας

Στο παρακάτω διάγραμμα παρουσιάζονται τα αποτελέσματα σε σχέση με την ικανότητα της κάθε μεθοδολογίας στον εντοπισμό της πιο κατάλληλης περίπτωσης σε σχέση με όλη την πληροφορία που αποτελεί πηγή γνώσης. Πιο αναλυτικά προβάλλεται για κάθε μεθοδολογία το πλήθος των ερωτήσεων για τις οποίες η απάντηση/εις που επιλέχθηκαν ως κατάλληλες για την επίλυσή τους αντικατοπτρίζουν την πιο ιδανική περίπτωση, που ουσιαστικά αποτελεί την καταλληλότερη πληροφορία για την επίλυση όλου ή ενός μέρους του προβλήματος. Στο κάτω μέρος του διαγράμματος καταγράφεται και ο ακριβής αριθμός των ιδανικών απαντήσεων.



Εικόνα 9 - Αποτελέσματα Αξιολόγησης για τη Βάση στη Ψυχολογία

Στην τελευταία μέθοδο αυτή του *ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη* στη χρήση των είκοσι ερωτήσεων που διερευνήθηκαν το αποτέλεσμα ήταν η επιτυχής ανάδειξη της καταλληλότερης περίπτωσης σε όλες τις ερωτήσεις. Στη μέθοδο αυτή η τελική βαθμολογία για κάθε περίπτωση σε σχέση με το ερώτημα που διερευνάται προκύπτει από το άθροισμα του αποτελέσματος της μεθόδου του ευρετηρίου και της μεθόδου του ευρετηρίου με τη χρήση *μόνο* των συνώνυμων όρων, ορίζοντας κάποια βάρη στη βαθμολογία του καθενός. Τα βάρη αυτά προκειμένου το τελικό αποτέλεσμα της βαθμολογίας να είναι και αυτό κανονικοποιημένο στο διάστημα  $[0,1]$  για να υπάρχει ομοιομορφία στα συνολικά αποτελέσματα όλων των μεθόδων, το άθροισμα των συντελεστών θα έχει άθροισμα 1, δηλαδή αν θεωρήσουμε  $w_1$  την βαθμολογία

από τη μέθοδο του ευρετηρίου και  $w_2$  την βαθμολογία από τη μέθοδο του ευρετηρίου με τη χρήση μόνο συνώνυμων όρων, τότε η τελική βαθμολογία θα προκύψει ως:

$$x*w_1 + y*w_2 \quad \text{όπου } x+y = 1$$

Οι τιμές  $x$ ,  $y$  θα είναι τέτοιες ώστε να ικανοποιούν τα μεγαλύτερο δυνατό πλήθος των ερωτημάτων. Από τα είκοσι ερωτήματα που ελέγξαμε παρατηρήσαμε ότι αν η τιμή του  $x$  (συντελεστή του αποτελέσματος με τη μέθοδο του ευρετηρίου) είναι μικρότερη του 0.5 τότε σε όλες τις περιπτώσεις εντοπίζεται και επιστρέφεται το κατάλληλο αποτέλεσμα ως πρώτο δηλαδή αυτό με τη μεγαλύτερη τιμή και κατ' επέκταση αυτό που εντοπίζει το σύστημα με τη μεγαλύτερη ομοιότητα. Αν η τιμή του  $x$  κυμαίνεται μεταξύ του 0.5 και 0.7 τότε και πάλι ικανοποιούνται όλες οι ερωτήσεις δηλαδή εντοπίζεται η κατάλληλη περίπτωση αλλά σε μία από αυτές το αποτέλεσμα δεν είναι αυτό με τη μεγαλύτερη βαθμολογία αλλά με τιμή μεγαλύτερη του 0.6, και πιο συγκεκριμένα με τιμή 0,72344. Αυτό σημαίνει ότι σ' ένα εύρος τιμών του  $x$  από [0.2, 0.7] μπορούμε να έχουμε το μέγιστο της απόδοσης, και η επιτυχία για το συγκεκριμένο πλήθος ερωτήσεων είναι 100%. Όταν η τιμή αυτή κυμαίνεται από [0.2, 0.5) τότε η ιδανική περίπτωση επιστρέφεται πρώτη, ενώ από [0.5, 0.7] εντοπίζεται πάντα η κατάλληλη περίπτωση αλλά σε ποσοστό 5% η περίπτωση ενδέχεται να μην είναι αυτή με τη μεγαλύτερη βαθμολογία, αλλά σίγουρα θα επιστραφεί ως παρεμφερής περίπτωση, οπότε και με αυτές τις τιμές εξασφαλίζουμε ότι η κατάλληλη γνώση θα επιστρέφεται στον ενδιαφερόμενο.

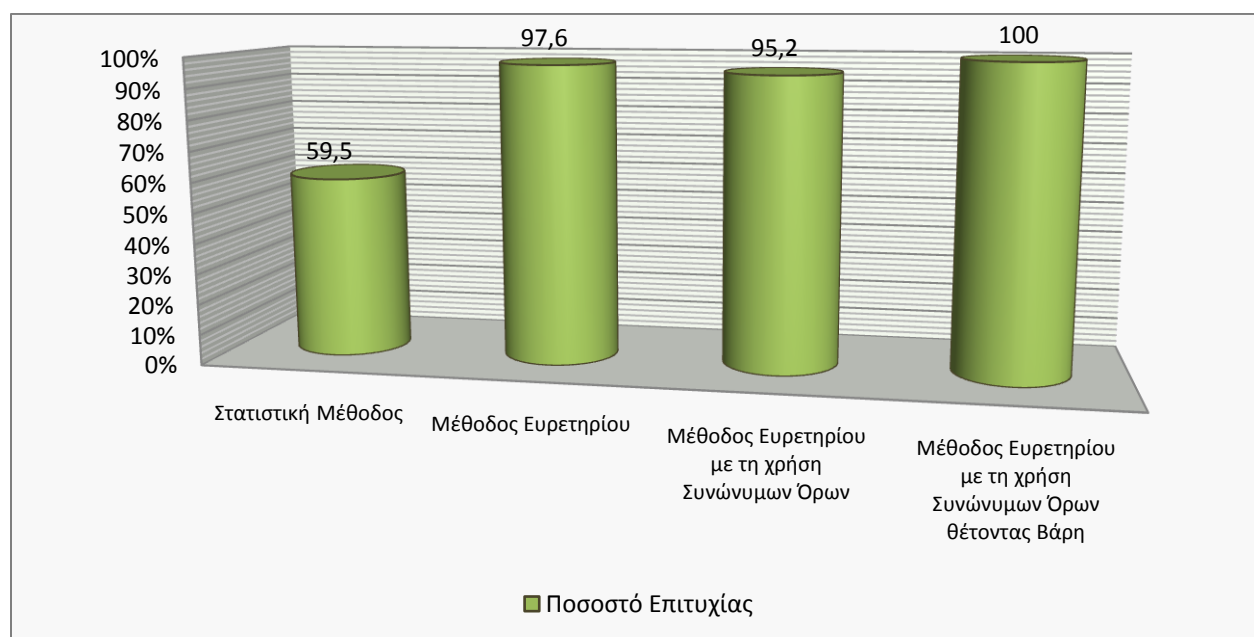
Ως αποτέλεσμα από την αξιολόγηση των τεσσάρων μεθόδων στο θεματικό αντικείμενο που εξετάσαμε ήταν ότι μόνο μία μέθοδος κατάφερε να έχει ιδανικό αποτέλεσμα για όλες τις ερωτήσεις που τέθηκαν. Εφόσον η μέθοδος του ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη ικανοποίησε ένα σύνολο ερωτήσεων που περιλάμβανε όλα τα πιθανά είδη ερωτήσεων που μπορούν να προκύψουν σε μια μεγάλη συλλογή πληροφοριών μπορούμε να είμαστε σίγουροι για την αποτελεσματικότητά της.

#### 4.17 Αποτελέσματα Αξιολόγησης

Εφόσον ολοκληρώθηκε η διαδικασία αξιολόγησης των τεσσάρων μεθοδολογιών για τις δύο διαφορετικές πηγές γνώσεων από ένα πλήθος ερωτήσεων για την κάθε μία, μπορούμε πλέον να εξάγουμε αντικειμενικά συμπεράσματα. Τα συνολικά συμπεράσματα έχουν να κάνουν με την απόδοση της κάθε μεθόδου στις μετρήσεις που πραγματοποιήθηκαν. Στο γράφημα που ακολουθεί παρουσιάζονται τα ποσοστιαία αποτελέσματα συγκεντρωτικά σε όλες τις ερωτήσεις που ελέγχθηκαν και στις δύο βάσεις.

Η στατιστική μέθοδος στα δύο διαφορετικά αντικείμενα γνώσης που εξετάστηκε, δεν ικανοποίησε το πλήθος των ερωτήσεων που περιλάμβανε σε ένα μεγάλο ποσοστό. Όπως

φαίνεται και από το γράφημα στο 59,5% των ερωτημάτων η ιδανική περίπτωση επιστράφηκε στον ενδιαφερόμενο, ενώ στο υπόλοιπο ποσοστό οι περιπτώσεις που επιστράφηκαν ως προτεινόμενες λύσεις των ερωτήσεων περιείχαν πληροφορία που δεν συσχετιζόταν άμεσα με το νόημα του ερωτήματος. Εύκολα γίνεται αντιληπτό ότι εφόσον υπήρχε αποτυχία και στα δύο θεματικά αντικείμενα, η στατιστική μέθοδος δεν καλύπτει τις απαιτήσεις του συστήματος. Το φαινόμενο που παρατηρήθηκε με τη στατιστική μέθοδο μεταξύ των δύο θεματικών αντικειμένων είναι ότι στο αντικείμενο με θέμα την ψυχολογία, δηλαδή ένα αντικείμενο που δεν χρησιμοποιεί κάποιου είδους ορολογία το ποσοστό επιτυχίας αναλογικά με το αντικείμενο της πληροφορικής, που χρησιμοποιεί ορολογία ήταν μεγαλύτερο. Σε κανένα αντικείμενο δεν είναι επαρκής η στατιστική μέθοδος, αλλά εμφανώς μπορεί να φέρει μεγαλύτερα θετικά αποτελέσματα σε ένα αντικείμενο που δεν χρησιμοποιεί κάποιου είδους ορολογία.



Εικόνα 10 - Συνολική Αξιολόγηση και από τις δύο Βάσεις

Το αναμενόμενο ήταν η αντίθετη συμπεριφορά, καθώς σ' ένα αντικείμενο που χρησιμοποιείται ορολογία η αναζήτηση κάποιου όρου ορολογίας, δεν αφήνει περιθώρια παρερμηνείας του. Η κύρια αιτία αποτυχίας της στατιστικής μεθόδου ήταν ότι η εύρεση των όρων της ερώτησης πραγματοποιούνταν σε όλο το κείμενο της περίπτωσης. Το κείμενο της κάθε περίπτωσης όμως περιέχει και ένα μεγάλο πλήθος όρων που δεν προσδίδουν κάποια σημασία σε αυτό. Παρόλο που πραγματοποιείται η διαδικασία της προ-επεξεργασίας του κειμένου συνεχίζουν να υπάρχουν όροι που στην πραγματικότητα παραπλανούν τα αποτελέσματα. Σημαντικό επίσης είναι το γεγονός ότι κατά τη διαδικασία εκτέλεσης του Λογικού μοντέλου, που εντοπίζει όλες τις περιπτώσεις που περιέχουν όρους της ερώτησης που διερευνάται, για όλες τις ερωτήσεις η κατάλληλη απάντηση για αυτή επιλεγόταν. Οπότε στο πρώτο βήμα επιλογής περίπτωσης η στατιστική μέθοδο είχε 100% επιτυχία. Στο επόμενο βήμα εκτέλεσης της



διαδικασίας του Διανυσματικού μοντέλου, που πραγματοποιείται η απονομή βαθμολογίας σε όλες τις περιπτώσεις που έχουν επιλεχτεί από την προηγούμενη διαδικασία, η τιμή που καταλαμβάνει η κατάλληλη περίπτωση σε ένα ποσοστό 40,5% των ερωτήσεων είναι μικρότερη του 0.6, οπότε δεν επιστρέφονται ως προτεινόμενες απαντήσεις. Όλα αυτά έχουν ως αποτέλεσμα τον αποκλεισμό της συγκεκριμένης μεθοδολογίας ώστε να εφαρμοστεί στο σύστημά μας.

Στην μεθοδολογία του ευρετηρίου, τα αποτελέσματα είναι εμφανώς πιο ενθαρρυντικά, καθώς στο συνολικό πλήθος ερωτήσεων το ποσοστό που η ιδανική απάντηση εντοπίστηκε ήταν το 97,6%. Το ποσοστό αυτό απευθύνεται στο σύνολο των ερωτήσεων που τέθηκαν και στα δύο θεματικά αντικείμενα. Προφανώς η συμπεριφορά της μεθοδολογίας αυτής έναντι της στατιστικής είναι πολύ ικανοποιητική, παρόλο που κατέγραψε ένα μικρό ποσοστό αποτυχίας. Η διαφορά που παρατηρήθηκε στα δύο αντικείμενα σχετικά με τη μέθοδο του ευρετηρίου, είναι ότι το αντικείμενο στον κλάδο της πληροφορικής σημείωσε μια απόλυτη επιτυχία καθώς σε όλα τα ερωτήματα εντοπίστηκε η πιο κατάλληλη περίπτωση, ενώ στο αντικείμενο στον κλάδο της ψυχολογίας σ' ένα ερώτημα δεν κατάφερε να εντοπίσει την ιδανική λύση γι' αυτό. Η αποτυχία σε αυτό το ερώτημα χαρακτηρίζει τη μέθοδο αυτή μη ικανή για μια θεματολογία που δεν διαπραγματεύεται ορολογία, αντίθετα για ένα θεματικό αντικείμενο που χρησιμοποιεί ορολογία η μέθοδος του ευρετηρίου είναι απόλυτα ικανή ώστε να εντοπίσει την κατάλληλη απάντηση από ένα σύνολο ερωτημάτων διαφορετικών απαιτήσεων. Όπως έχει είδη αναφερθεί με τη χρήση ορολογίας δεν υπάρχει παρερμηνεία στους όρους τους. Έτσι σε ένα ερώτημα που περιλαμβάνεται κάποιος όρος με μια μοναδική ερμηνεία, ο εντοπισμός τους με τη μέθοδο του ευρετηρίου είναι ιδανικός αρκεί βέβαια να έχουν δηλωθεί με προσοχή οι όροι των λέξεων κλειδιών. Αν και η μέθοδος αυτή αποδείχτηκε σημαντική και ως προς ένα μεγάλο μέρος ικανοποιητική, δεν ανταποκρίνεται πλήρως στις απαιτήσεις του συστήματος, καθώς το σύστημα στοχεύει στην αντιμετώπιση και επίλυση των αποριών των εκπαιδευομένων με μια αυτοματοποιημένη διαδικασία ανεξάρτητα από τις θεματικές ενότητες που περιλαμβάνει.

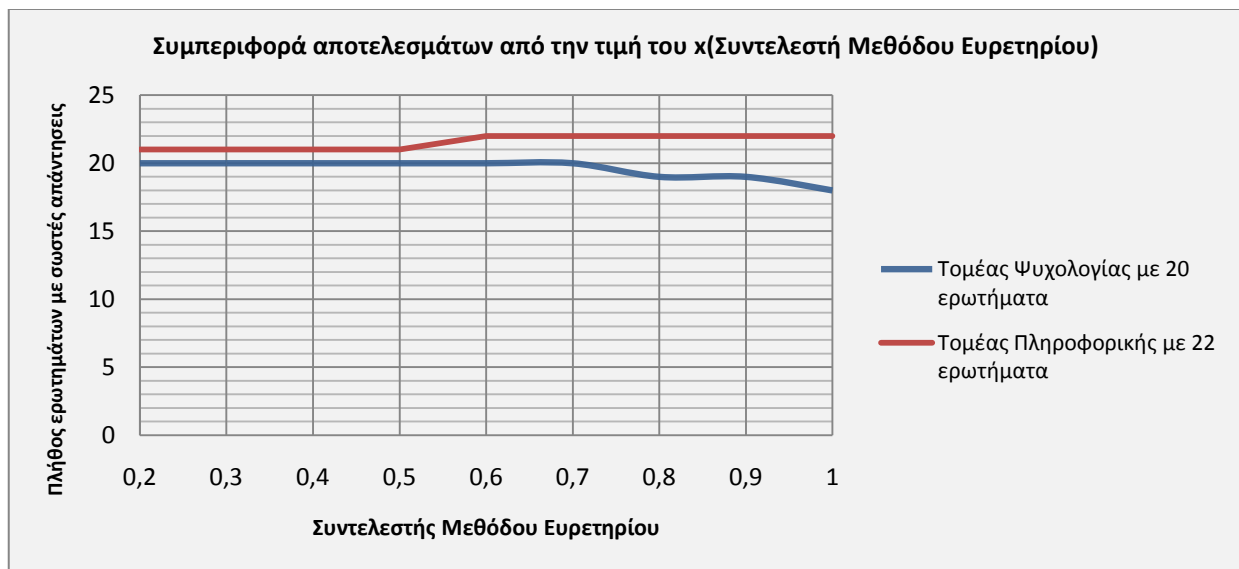
Πιο ξεκάθαρα παρουσιάζονται τα αποτελέσματα στη τρίτη μέθοδο του ευρετηρίου με τη χρήση συνώνυμων όρων, καθώς το συνολικό αποτέλεσμα είναι ικανοποιητικό αλλά υπάρχει ένα μικρό ποσοστό αποτυχίας και στα δύο θεματικά αντικείμενα. Και στις δύο διαφορετικές πηγές γνώσεων σε μία ερώτηση η κατάλληλη απάντηση δεν εντοπίστηκε με τη χρήση της μεθοδολογίας του ευρετηρίου με τη χρήση συνώνυμων όρων. Αυτό οδηγεί στο συμπέρασμα ότι η χρήση συνώνυμων όρων είναι σημαντική και φαίνεται από τα αξιόλογα αποτελέσματα της μεθόδου, αλλά η μεγάλη ισχύς που αποδίδεται στους συνώνυμους όρους σε αυτή τη μέθοδο σε κάποιες ερωτήσεις οδηγεί σε λανθασμένη επιλογή κατάλληλων απαντήσεων. Αυτός ήταν και ο βασικός λόγος για τη δημιουργία της επόμενης μεθόδου. Στη συγκεκριμένη μέθοδο οι συνώνυμοι όροι εντοπίζονται με τη χρήση του σημασιολογικού λεξικού wordnet. Η ιδέα της ένταξης του λεξικού προέκυψε από το γεγονός ότι όροι που περιλαμβάνονται στη ερώτηση, μπορούν να αντικατασταθούν με συνώνυμους και να έχουν ακριβώς την ίδια ερμηνεία. Έτσι

όταν τίθεται μια ερώτηση στην ουσία γίνεται αναζήτηση των όρων αυτών μέσα στις περιπτώσεις, αν όμως ένας όρος με μια συγκεκριμένη έννοια μπορεί να αναπαρασταθεί και με άλλες λέξεις, πρέπει να εντοπιστούν έτσι ώστε αν περιέχονται στο σύνολο των λέξεων κλειδιών να μπορεί το σύστημα να τις αναγνωρίσει. Για παράδειγμα αν τεθεί μια ερώτηση της μορφής «ποια είναι τα είδη ασθενειών που υπάρχουν;» ο όρος ασθένεια που χρησιμοποιείται στην ερώτηση έχει το ίδιο νόημα με τη λέξη αρρώστια ή νόσο, και πιθανόν να υπάρχουν περιπτώσεις που ικανοποιούν την απάντηση από τη συλλογή χρησιμοποιώντας κάποιους από αυτούς τους όρους και ο εντοπισμός τους να μην είναι εφικτός καθώς δεν θα γνωρίζει τη σχέση τους. Όπως αποδείχθηκε και στην πράξη η εφαρμογή της μεθόδου ήταν καθοριστική καθώς ο εντοπισμός των συνώνυμων όρων είναι χρήσιμος κυρίως σε θεματικές ενότητες που δεν χρησιμοποιείται ορολογία. Παρόλα αυτά η μέθοδος του ευρετηρίου με τη χρήση συνώνυμων όρων δεν στάθηκε επαρκής για την εφαρμογή της στο σύστημα καθώς η αποτυχία που παρουσίασε αν και μικρή, δεν την καθιστά ικανή για εφαρμογή στην αυτοματοποιημένη διαδικασία απαντήσεων του συστήματος.

Η τελευταία και πιο ελπιδοφόρα μέθοδος είναι αυτή του ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη. Το διάγραμμα που αποτυπώνει τα αποτελέσματα των μεθόδων στις ερωτήσεις που ελέγχθηκαν σχετικά με το πόσο οι απαντήσεις που επιλέχθηκαν ως προτεινόμενες έλυναν το πρόβλημα παρουσιάζει τη μέθοδο αυτή ως την πιο κατάλληλη καθώς εντόπισε σε όλες τις ερωτήσεις την ιδανική απάντηση για αυτές. Η συμπεριφορά αυτή προκύπτει από τον έλεγχο και των δύο θεματικών αντικειμένων, που εξετάστηκαν. Το πρόβλημα όμως που προκύπτει είναι ότι δεν έχουν καθοριστεί τα βάρη των δύο ξεχωριστών μετρήσεων που πραγματοποιούνται έτσι ώστε να προκύψει η τελική βαθμολογία. Σε κάθε ένα αντικείμενο που εξετάστηκε ορίστηκε ένα εύρος τιμών για αυτά τα βάρη, έτσι ώστε να καλυφθούν όλες οι ερωτήσεις τους. Τα βάρη αυτά προκειμένου το τελικό αποτέλεσμα της βαθμολογίας να είναι και αυτό κανονικοποιημένο στο διάστημα  $[0,1]$  θα έχουν άθροισμα 1, δηλαδή αν θεωρήσουμε  $w_1$  την βαθμολογία από τη μέθοδο του ευρετηρίου και  $w_2$  τη βαθμολογία από τη μέθοδο του ευρετηρίου με τη χρήση μόνο συνώνυμων όρων, τότε η τελική βαθμολογία θα προκύψει ως:

$$x*w_1 + y*w_2 \quad \text{όπου } x+y = 1$$

Από την αξιολόγηση στον κλάδο της πληροφορικής προέκυψε ότι η τιμή του  $x$  για να ικανοποιεί όλα το πλήθος των ερωτήσεων πρέπει να κυμαίνεται στο διάστημα  $[0.6, 1.0]$ , ποιο συγκεκριμένα αν ανήκει στο διάστημα  $[0.6, 0.8)$  τότε εντοπίζεται σε όλες τις ερωτήσεις η ιδανική απάντηση αλλά σε μία ερώτηση δεν είναι αυτή με τη μεγαλύτερη τιμή, και αν ανήκει στο διάστημα  $[0.8, 1.0]$  τότε εντοπίζεται σε όλες τις ερωτήσεις η ιδανική περίπτωση και μάλιστα με τη μεγαλύτερη βαθμολογία σε όλες.



Εικόνα 11 - Συμπεριφορά Τελικών Αποτελεσμάτων από τις Μεταβολές του Συντελεστή  $X$

Για τον κλάδο της ψυχολογίας από την αξιολόγηση προκύπτει ότι αν το  $x$  ανήκει στο διάστημα  $[0.2, 0.5)$ , τότε σε όλες τις ερωτήσεις η ιδανική απάντηση εντοπίζεται και πιο συγκεκριμένα είναι πρώτη σε όλες, δηλαδή σε όλες τις ερωτήσεις η απάντηση που κατέλαβε τη μεγαλύτερη βαθμολογία είναι και η ιδανική. Σωστές απαντήσεις εξασφαλίζουν επίσης όλα τα ερωτήματα αν το  $x$  ανήκει στο διάστημα  $[0.5, 0.7]$ , με τη μικρή διαφορά ότι σ' ένα ερώτημα η κατάλληλη απάντηση γι' αυτή επιστρέφεται αλλά δεν είναι αυτή με την υψηλότερη βαθμολογία.

Από την αξιολόγηση θα πρέπει να επιλεγεί η τιμή του  $x$  και  $y$  ώστε να ικανοποιηθεί το μεγαλύτερο πλήθος ερωτήσεων και στην ιδανική εκδοχή όλες. Για να ικανοποιηθούν όλες οι ερωτήσεις θα πρέπει να υπάρχει ένα κοινό πεδίο στα διαστήματα των δύο αντικειμένων που μπορεί να καταλάβει το  $x$ , που όπως παρατηρούμε και από την Εικόνα 11 υπάρχει κοινό διάστημα και ορίζεται ως  $[0.6, 0.7]$ . Έτσι οποιαδήποτε τιμή και να πάρει το  $x$  από αυτό το διάστημα εξασφαλίζουμε ότι εφόσον υπάρχει γνώση σχετικά με μια ερώτηση που θα τεθεί, η γνώση αυτή θα εντοπιστεί και θα επιστραφεί στο χρήστη. Η τιμή που θα πάρει το  $x$  σε αυτό το διάστημα θα είναι αυτή που θα επιφέρει σε όλα τα ερωτήματα όσο το δυνατόν μεγαλύτερη βαθμολογία στη κατάλληλη περίπτωση για την κάθε μία. Στην Εικόνα 12 που ακολουθεί παρουσιάζεται ο μέσος όρος των ιδανικών απαντήσεων από όλα τα ερωτήματα σε σχέση με τις τιμές που μπορεί να πάρει το  $x$  στο διάστημα  $[0.6, 0.7]$ .



Εικόνα 12 - Μέσος Όρος Βαθμολογίας όλων των Ιδανικών Περιπτώσεων Ανάλογα τον Συντελεστή X

Ελέγχοντας όλα τα ερωτήματα ο αριθμός αυτός είναι το 0.6, έτσι ορίζουμε το  $x = 0.6$  οπότε το y θα ισούται με  $y = 0.4$ . Οπότε η τελική βαθμολογία(fg) στη μέθοδο του ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη θα ισούται με:

$$fg = 0.6*w1 + 0.4*w2$$

Με τη σχέση αυτή εξασφαλίζεται η επιτυχία σε όλα τα ερωτήματα ανεξάρτητα με τη θεματική ενότητα στην οποία απευθύνεται, και κατ' επέκταση τη χρήση ή μη ορολογίας.

Έχοντας ολοκληρωθεί η αξιολόγηση σε όλες τις μεθόδους που σχεδιάστηκαν καταλήγουμε στο συμπέρασμα ότι η μέθοδος που ικανοποίησε όλα τα ερωτήματα ανεξάρτητα του θεματικού αντικειμένου είναι αυτή του *ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη*. Γενικότερα να αναφέρουμε ότι πολύ σημαντική ήταν και η επιτυχία που σημείωσε και η μέθοδος του ευρετηρίου και έγινε η βάση για την ανάπτυξη και των άλλων δύο μεθόδων που χρησιμοποιήθηκε η ίδια τεχνική. Η στατιστική μέθοδος από την άλλη παρατηρήθηκε ότι δεν ήταν καθόλου αποτελεσματική καθώς η πιθανότητα αποτυχίας στον εντοπισμό της κατάλληλης απάντησης ήταν μεγάλη.

Από τα συνολικά αποτελέσματα της αξιολόγησης ορίστηκε και η μικρότερη βαθμολογία που θα καταλαμβάνουν οι περιπτώσεις που θα επιστρέφονται στον ενδιαφερόμενο ως παρεμφερείς ή προτεινόμενες, και η βαθμολογία αυτή είναι η τιμή 0.6. Η επιλογή της βαθμολογίας αυτής προέκυψε από τα συγκεντρωτικά αποτελέσματα καθώς οι περιπτώσεις που βαθμολογούνταν με βαθμολογία μεγαλύτερη ίση του 0.6 είχαν μεγάλη ομοιότητα με το ερώτημα και έτσι εξασφάλισαμε ότι:

- Θα εντοπίζεται με μεγαλύτερη πιθανότητα η περίπτωση που θα καλύπτει την ερώτηση.

- Δεν θα επιστρέφεται ένα μεγάλο πλήθος περιπτώσεων στο χρήστη που έθεσε την ερώτηση, για να μπορεί να εντοπίσει όσο τον δυνατόν ευκολότερα την περίπτωση που τον ικανοποιεί.

Αν ο αριθμός αυτός που χρησιμοποιείται ως φραγμός ήταν μικρότερος τότε υπήρχε πιθανότητα με την πάροδο του χρόνου και την αύξηση της πληροφορίας το πλήθος των περιπτώσεων να ήταν αρκετά μεγάλο. Με την επιλογή αυτής της τιμής ως φραγμό, παρατηρήσαμε και στα δύο θεματικά αντικείμενα που το πλήθος των περιπτώσεων ήταν περίπου εκατό, ότι οι πιο σχετικές περιπτώσεις έχουν βαθμολογία μεγαλύτερη από 0.6, και το πλήθος των περιπτώσεων με αυτή τη βαθμολογία δεν ήταν ποτέ μεγάλο, αλλά κυμαινόταν ανάμεσα στο διάστημα [1,4], με τον αριθμό 4 να το έχουν φτάσει μόλις δύο ερωτήματα. Ενώ αν η τιμή ήταν 0.2 παρόλο που εξασφαλίζουμε την επιτυχία επιστροφής της πιο κατάλληλης περίπτωσης το πλήθος των περιπτώσεων που θα επιστρεφόταν μπορεί να ήταν πολύ μεγάλο, για παράδειγμα έχουν καταγραφεί ερωτήματα που το πλήθος των περιπτώσεων με βαθμολογία μεγαλύτερη του 0.2 μπορεί να φτάνει μέχρι και τα είκοσι έξι ερωτήματα, ένας αριθμός αρκετά μεγάλος, που στην ουσία ο χρήστης πρέπει από μόνος του να εντοπίσει την περίπτωση που ταιριάζει κατάλληλα στο ερώτημα με μια μικρή συνδρομή του συστήματος. Στόχος του συστήματος είναι η ουσιαστική λύση στο ερώτημα του χρήστη, και αυτή να προέρχεται με την μικρότερη δυνατή συμμετοχή του. Το απόλυτα ιδανικό σενάριο θα ήταν να επιστρέφεται πάντα μία και μόνο απάντηση που θα ήταν και η καταλληλότερη, όμως αυτό όπως γίνεται αντιληπτό δεν μπορεί να εξασφαλιστεί από καμία μεθοδολογία με εκατό τις εκατό επιτυχία, οπότε επιλέξαμε το μεγαλύτερο βαθμό που θα εξασφάλιζε τη σίγουρη επιστροφή της κατάλληλης περίπτωσης με όσο το δυνατόν μικρότερο πλήθος επιστρεφόμενων περιπτώσεων.

## 5. Συμπεράσματα

Οι τεχνικές για την αποδοτικότερη επικοινωνία εκπαιδευτή-εκπαιδευόμενου και προσεγγίσεις για την ευκολότερη επίλυση των αποριών είναι ένα καίριο ζήτημα για τον τομέα της ηλεκτρονικής εκπαίδευσης. Στην παρούσα εργασία δημιουργήθηκε ένα σύστημα που χρησιμοποιείται ως ενδιάμεσο της επικοινωνίας μεταξύ εκπαιδευτή και εκπαιδευόμενου, ώστε να αυτοματοποιηθεί η διαδικασία απάντησης των αποριών του εκπαιδευόμενου προς όφελος τόσο του εκπαιδευτή όσο και του εκπαιδευόμενου. Για τον εκπαιδευτή η αυτοματοποίηση της διαδικασίας αυτής μειώνει σημαντικά το φόρτο εργασίας και ενασχόλησής του, και για τους εκπαιδευόμενους ελαχιστοποιείται η αναμονή της απάντησής του. Το σύστημα βασίζεται στη χρήση της προϋπάρχουσας γνώσης για την επίλυση των αποριών από τους εκπαιδευόμενους. Έτσι σε μια καινούργια απορία ελέγχεται αν υπάρχει γνώση για την επίλυσή της από προηγούμενες περιπτώσεις που είχαν επιλυθεί στο παρελθόν από τον εκπαιδευτή. Αν το σύστημα εντοπίσει προϋπάρχουσα γνώση για την επίλυση της απορίας την επιστρέφει στον εκπαιδευόμενο χωρίς την μεσολάβηση του εκπαιδευτή. Στην αντίθετη περίπτωση που δεν εντοπιστεί πληροφορία ώστε να επιλύει την απορία, την επιλύει ο εκπαιδευτής, και αυτή η γνώση αποθηκεύεται στο σύστημα για την επίλυση παρόμοιων αποριών στο μέλλον.

Οι τεχνικές που ενσωματώθηκαν για την πραγματοποίηση αυτής της ιδέας είναι αυτές του Textual Case Based Reasoning. Η συλλογιστική βασισμένη σε περιπτώσεις κειμένου (TCBR) όπως ορίζεται στην ελληνική βιβλιογραφία είναι μια υποκατηγορία της συλλογιστικής βασισμένη σε περιπτώσεις (CBR) όπου όλο ή ένα μέρος της γνώσης που χρησιμοποιείται είναι σε μορφή κειμένου. Ανάλογα με τη μορφή του κειμένου που αποτελεί πηγή γνώσης υπάρχουν δύο κατηγορίες η δομημένη και η αδόμητη. Από τα συστήματα που παρέχουν πλαίσια ενός πυρήνα CBR επιλέχτηκε το σύστημα Jcolibri που είναι και το μόνο ανοιχτού κώδικα που προσφέρει τις βασικές ενέργειες του κύκλου ζωής των CBR για γνώση σε μορφή κειμένου. Ο διαχωρισμός που πραγματοποιείται σχετικά με τη μορφή κειμένου που αποτελεί πηγή γνώσης για το σύστημα είναι στο βήμα της ανάκτησης και περιλαμβάνει τη σημασιολογική και τη στατιστική. Στο σύστημα που σχεδιάστηκε επιλέχτηκε η στατιστική καθώς το κείμενο που χρησιμοποιείται ως πηγή γνώσης είναι σε αδόμητη μορφή.

Πάνω σε αυτό το πλαίσιο υλοποιήθηκε ένα ολοκληρωμένο στάδιο προ-επεξεργασίας τόσο του ερωτήματος όσο και των πληροφοριών που είναι καταχωρημένες στη βάση του συστήματος, για την αποδοτικότερη σύγκριση των αποτελεσμάτων. Προκειμένου η εύρεση των όρων που αναζητούνται να μη γίνεται με μια σειριακή αναζήτηση η οποία σε μεγάλες συλλογές μπορεί να αποβεί τρομερά χρονοβόρα εφαρμόστηκε η ιδέα του ανεστραμμένου καταλόγου, με τον οποίο αντί να αποθηκεύουμε πληροφορία για το ποιοι όροι περιέχονται σε ένα έγγραφο, καταγράφουμε το σύνολο των εγγράφων που περιέχουν ένα συγκεκριμένο όρο. Ένας όρος όμως

που περιέχεται σε δύο κείμενα δεν έχει την ίδια σπουδαιότητα και για τα δύο έτσι χρησιμοποιήθηκε μια διαδικασία υπολογισμού σημαντικότητας όρων που βασίζεται στο σχήμα TF-IDF (term frequency – inverse document frequency). Με το συγκεκριμένο σχήμα ένας όρος θεωρείται σημαντικός εφόσον η συχνότητα εμφάνισής του σε ένα κείμενο είναι μεγάλη, λαμβάνοντας όμως υπόψη του το πλήθος των εμφανίσεών του σε όλα τα κείμενα.

Για τον εντοπισμό της ομοιότητας χρησιμοποιήθηκε η βιβλιοθήκη της Apache Lucene που εφαρμόζει ένα συνδυασμό του Μοντέλου Διανυσματικού Χώρου και του Λογικού Μοντέλου, και αποδίδει βαθμολογία σε κάθε έγγραφο(περίπτωση) σχετικά με την ομοιότητά του με την τρέχουσα απορία. Η βαθμολογία αυτή είναι κανονικοποιημένη στο διάστημα  $[0,1]$  με 1 τη μέγιστη ομοιότητα.

Η πιο σημαντική λειτουργία του συστήματος είναι το κομμάτι της ανάκτησης που ουσιαστικά πραγματοποιείται η διαχείριση της γνώσης. Η διαδικασία που περιείχε το πλαίσιο του Jcolibri δεν ήταν αρκετή για να καλύψει τις ανάγκες του συστήματος που υλοποιήθηκε, έτσι αναπτύχθηκαν τέσσερις ολοκληρωμένες μέθοδοι διαχείρισης γνώσης για την εφαρμογή τους στο βήμα της ανάκτησης. Η κάθε μία από αυτές αναπτύχθηκε για να καλύψει ένα φάσμα απαιτήσεων. Για να αξιολογήσουμε τις μεθόδους δημιουργήθηκαν δύο βάσεις δεδομένων από διαφορετικά θεματικά αντικείμενα και ένα πλήθος ερωτήσεων για την κάθε μια που περιλαμβάνει όλα τα πιθανά σενάρια που μπορούν να εκδηλωθούν. Από την αξιολόγηση καταλήξαμε ότι μέθοδος που θα επιλεχτεί για την εφαρμογή της στο σύστημα είναι αυτή του ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη. Ήταν η μόνη μέθοδος έναντι των υπολοίπων που συγκέντρωσε τα πιο ικανοποιητικά αποτελέσματα σε όλο το πλήθος των ερωτήσεων και για τις δύο διαφορετικές θεματικές ενότητες.

Έτσι καταλήγουμε στο συμπέρασμα ότι δημιουργήθηκε ένα σύστημα που αυτοματοποιεί τη διαδικασία απάντησης των αποριών αξιοποιώντας την προϋπάρχουσα γνώση. Το σύστημα αυτό αποδεδειγμένα είναι σε θέση να ικανοποιήσει την απορία του εκπαιδευόμενου εφόσον υπάρχει γνώση για αυτή. Επίσης μπορεί να εφαρμοστεί για τη διδασκαλία μαθημάτων που ανήκουν σε οποιοδήποτε θεματική ενότητα.

## 5.1 Καταγραφή της επίδοσης των εκπαιδευόμενων

Στα πλαίσια της μεταπτυχιακής εργασίας δημιουργήθηκε ένα ολοκληρωμένο σύστημα που χρησιμοποιείται ως ενδιάμεσο της επικοινωνίας μεταξύ εκπαιδευτή και εκπαιδευόμενου με σκοπό την αυτοματοποίηση της διαδικασίας απάντησης αποριών. Οι τεχνικές που εφαρμόστηκαν για την ανάπτυξη του συστήματος είναι αυτές του case based reasoning (συλλογιστική βασισμένη σε περιπτώσεις), οι οποίες μπορούν να χρησιμοποιηθούν και για άλλους λόγους. Μία από τις επεκτάσεις που θα μπορούσε να πραγματοποιηθεί είναι η εξαγωγή συμπερασμάτων σχετικά με τις επιδόσεις των εκπαιδευόμενων σε διάφορες θεματικές ενότητες. Εφόσον οι απορίες των εκπαιδευόμενων εισέρχονται ως είσοδος στο σύστημα μπορούν να καταγράφονται και να εντοπίζονται τα σημεία εκείνα στα οποία οι εκπαιδευόμενοι παρουσίασαν προβλήματα επίδοσης και κατανόησης. Η χαρτογράφηση όλων αυτών των συμπερασμάτων ουσιαστικά μπορεί να αποτελέσει την καλύτερη ενημέρωση ως προς τον εκπαιδευόμενο και της πραγματικής κατάστασης σχετικά με την επίδοση των εκπαιδευόμενων. Ο εκπαιδευτής έχοντας τον έλεγχο της επίδοσης μπορεί έγκαιρα να βρει άμεσες λύσεις στα προβλήματα που μπορεί να δημιουργηθούν αναθεωρώντας ενδεχομένως το εκπαιδευτικό υλικό ή άλλες διδακτικές αδυναμίες που οδήγησαν στο φαινόμενο αυτό.

## 5.2 Προσαρμογή του συστήματος στην ελληνική γλώσσα

Στο σύστημα χρησιμοποιήθηκαν τεχνικές για την προ-επεξεργασία κειμένου που αφορούν την αγγλική γλώσσα. Την περίοδο υλοποίησης του συστήματος δεν υπήρχαν υλοποιήσεις στο κομμάτι την προ-επεξεργασίας ελληνικού κειμένου που να επιφέρουν αξιόπιστα αποτελέσματα. Από το Ινστιτούτο Επεξεργασίας του Λόγου είναι υπο-κατασκευή μια ολοκληρωμένη εφαρμογή προ-επεξεργασίας ελληνικού κειμένου στη γλώσσα προγραμματισμού Java. Οι συναρτήσεις που εκτελούν τα στάδια της προ-επεξεργασίας έχουν υλοποιηθεί σε ένα interface, έτσι ώστε η δημιουργία ενός αντίστοιχου για την ελληνική γλώσσα και η χρήση τους στο κυρίως σύστημα να πραγματοποιηθεί πάρα πολύ εύκολα όταν αυτές είναι διαθέσιμες από το Ινστιτούτο Επεξεργασίας του Λόγου.

## 5.3 Προσαρμογή περιβάλλοντος Efront

Το σύστημα για να εφαρμοστεί πλήρως στην πλατφόρμα του efront θα πρέπει να υπάρξουν κάποιες αλλαγές στη διεπαφή. Αρχικά να υπάρξει μια διαφορετική διεπαφή στα δύο είδη χρηστών εκπαιδευτή και εκπαιδευόμενου. Για τον εκπαιδευτή να του δίνεται η δυνατότητα στο φόρουμ σε ένα ξεχωριστό πεδίο να εισάγει όλες τις λέξεις κλειδιά για το πρόβλημα που επιλύει. Για τον εκπαιδευόμενο να υπάρχει μια επιλογή αν τον ικανοποιεί η απάντηση ή όχι, έτσι ώστε από την πληροφορία αυτή να ολοκληρωθεί και το βήμα της



αναθεώρησης του κύκλου ζωής του case based reasoning. Επίσης για την καλύτερη παρουσίαση των επιστρεφόμενων αποτελεσμάτων στο περιβάλλον του φόρουμ θα ήταν καλό να υπάρχει ένα πλαίσιο με τίτλο «Μήπως εννοείτε:» και στη συνέχεια να ακολουθούν όλοι οι τίτλοι των προτινόμενων απαντήσεων ως σύνδεσμοι, και να σε μεταφέρουν σε ολόκληρη την απάντηση. Τέλος για την καλύτερη εφαρμογή του θα μπορούσε να ενταχθεί πλήρως στο σύστημα, και να εφαρμοστεί και για την επικοινωνία εκπαιδευτή-εκπαιδευόμενου που πραγματοποιείται μέσω προσωπικών μηνύματων ή να συνδυαστεί αν χρησιμοποιούνται και οι δύο μέθοδοι επικοινωνίας.

#### **5.4 Παροχή Εξατομικευμένης Πρόσβασης σε Υπηρεσίες εξ Αποστάσεως Εκπαίδευσης(e-Learning Personalization)**

Μια από τις ενδιαφέρουσες τεχνικές για τη μεγιστοποίηση της αποτελεσματικότητας της εκπαιδευτικής μαθησιακής διαδικασίας είναι η εξατομικευμένη πρόσβαση σε υπηρεσίες. Η παροχή εξατομικευμένων υπηρεσιών αναφέρεται στην διαδικασία της προσαρμογής περιεχομένου ώστε να απευθύνεται αποκλειστικά σε ένα χρήστη ανάλογα με τις προτιμήσεις και το προφίλ του. Η εφαρμογή μιας τέτοιας υπηρεσίας απαιτεί παρακολούθηση της συμπεριφοράς του χρήστη, και απόφαση για τις υπηρεσίες που θα του παρέχει σύμφωνα με την προηγούμενη συμπεριφορά του. Μια από τις σύνηθες τεχνικές που χρησιμοποιούνται για την παροχή υπηρεσιών εξατομικευμένης πρόσβασης είναι η συλλογιστική βασισμένη σε περιπτώσεις(Case Based Reasoning). Το σύστημα που αναπτύχθηκε στα πλαίσια της εργασίας μπορεί να επεκταθεί ώστε να παρέχει και εξατομικευμένες υπηρεσίες, καθώς περιέχει όλο το βασικό πλαίσιο για την ανάπτυξη του.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

Δημητρακοπούλου Α, Γρηγοριάδου Μ, Καρδάκη Μ, Καμέας Α. (2004) Τεχνολογίες των επικοινωνιών στην Εκπαίδευση Ενηλίκων: Διδακτικές προσεγγίσεις και εργαλεία, Πρακτικά 4<sup>ου</sup> Συνεδρίου ΕΤΠΕ, β' τόμος, Σεπτέμβριος 2004, Αθήνα

Σολομωνίδου Χ.(2003), Σύγχρονη Εκπαιδευτική Τεχνολογία, Υπολογιστές και Μάθηση στην Κοινωνία της Γνώσης, Β' έκδοση, Θεσσαλονίκη, Εκδόσεις Κώδικας

Ράπτης, Α. και Ράπτη, Α. (2001), Μάθηση και διδασκαλία στην εποχή της πληροφορίας, Αθήνα

<http://e-learning-education.blogspot.gr/2012/02/m.html>

<http://spapac7.wordpress.com/2011/06/10/%CF%80%CE%BB%CE%B5%CE%BF%CE%BD%CE%B5%CE%BA%CF%84%CE%AE%CE%BC%CE%B1%CF%84%CE%B1-%CF%84%CE%BF%CF%85-e-learning/>

## **ΠΑΡΑΡΤΗΜΑ**

### **ΛΙΣΤΑ ΣΥΝΤΜΗΣΕΩΝ**

**E.R.P:** Enterprise Resource Planning

**W.H.M:** Warehouse Management

**T.R.M:** Transportation Management

**E.O.Q:** Economic Order Quantity

**S.S:** Safety Stock

**B.O.M.P:** Bill Of Material Processing

**W.O.M:** Work Order Management

**M.R.P:** Material Requirements Planning

**R.O.I:** Return On Investment

**M.I.S:** Management Information System

**C.R.M:** Customer Relationship Management

**S.C.M:** Supply Chain Management

**E.A.I:** Enterprise Application Integration

**R.O.A.D.S:** Reusable Objects For Application Development Systems

**B.I.T.s:** Business Intelligence Toolkit

**MS-SQL:** Microsoft SQL Server

## **ΑΤΥΠΟ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ**

- 1. ΟΝΟΜΑ ΕΤΑΙΡΕΙΑΣ (ΠΕΛΑΤΗ) :**
- 2. ΟΝΟΜΑ ΧΡΗΣΤΗ ΠΡΟΓΡΑΜΜΑΤΟΣ:**
- 3. ΘΕΣΗ ΧΡΗΣΤΗ ΠΡΟΓΡΑΜΜΑΤΟΣ :**
- 4. ΕΤΑΙΡΕΙΑ ΠΡΟΓΡΑΜΜΑΤΟΣ (ΠΩΛΗΤΗΣ) :**
- 5. ΟΝΟΜΑ ΟΛΟΚΛΗΡΩΜΕΝΟΥ ΠΛΗΡΟΦΟΡΙΑΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ-ΕΚΔΟΣΗ (ERP):**
- 6. ΧΡΟΝΟΣ ΕΓΚΑΤΑΣΤΑΣΗΣ ΠΡΟΓΡΑΜΜΑΤΟΣ:**
- 7. ΧΡΟΝΟΣ ΕΚΠΑΙΔΕΥΣΗΣ ΧΡΗΣΤΩΝ:**
- 8. ΕΑΝ ΧΡΗΣΙΜΟΠΟΙΟΥΣΑΝ ΚΑΠΟΙΟ ΑΛΛΟ ERP ΠΡΟΓΡΑΜΜΑ :**
  - ❖ ΠΟΙΟ ΗΤΑΝ ΑΥΤΟ:**
  - ❖ ΠΟΙΕΣ ΕΙΝΑΙ ΟΙ ΠΑΡΑΤΗΡΗΣΕΙΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟ ΜΕ ΤΟ ΠΑΡΟΝ ΠΡΟΓΡΑΜΜΑ:**
  - ❖ ΠΟΙΕΣ ΕΙΝΑΙ ΟΙ ΣΗΜΑΝΤΙΚΕΣ ΑΛΛΑΓΕΣ/ ΔΙΑΦΟΡΕΣ ΕΝΑΝΤΙ ΤΟΥ ΠΡΟΗΓΟΥΜΕΝΟΥ:**
- 9. ΠΛΗΡΟΥΝΤΑΙ ΟΙ ΑΠΑΙΤΗΣΕΙΣ ΤΗΣ ΕΠΙΧΕΙΡΗΣΗΣ ΜΕ ΤΟ ΠΑΡΟΝ ΠΡΟΓΡΑΜΜΑ:**
- 10. ΚΟΣΤΟΣ ΕΓΚΑΤΑΣΤΑΣΗΣ:**
- 11. ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ ΜΕ ΤΙΣ ΟΠΟΙΕΣ ΣΥΝΕΡΓΑΖΕΤΑΙ ΤΟ ΠΡΟΓΡΑΜΜΑ:**
- 12. ΠΡΟΣΑΡΜΟΣΤΙΚΟΤΗΤΑ ΧΡΗΣΤΗ:**
- 13. ΑΛΛΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ ΤΟΥ ΧΡΗΣΤΗ:**

