Πανεπιστήμιο Κρήτης

Σχολή Θετικών Επιστημών

Τμήμα Επιστήμης Υπολογιστών

Ένα Εργαλείο για Ηλεκτρονική Μάθηση Βασισμένο σε Τεχνικές Συλλογισμού.

Ειρήνη Α. Φιλιοπούλου

Μεταπτυχιακή Εργασία

Ηράκλειο, Νοέμβριος 2010

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Ένα Εργαλείο για Ηλεκτρονική Μάθηση Βασισμένο σε Τεχνικές Συλλογισμού

Μεταπτυχιακή Εργασία που υποβλήθηκε από την

Ειρήνη Α. Φιλιοπούλου

ως μερική εκπλήρωση των απαιτήσεων για την απόκτηση

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας: _____

Ειρήνη Φιλιοπούλου, Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Επόπτης: _____

Χρήστος Νικολάου, Καθηγητής

Επιβλέπων: _____

Δημήτριος Πλεξουσάκης, Καθηγητής

Μέλος: _____

Μιχάλης Καλογιαννάκης, Λέκτορας

Μέλος: _____

Νικόλαος Παπαδάκης, Καθηγητής

Έγκριση: _____

Άγγελος Μπίλας, Αναπληρωτής Καθηγητής,
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Ηράκλειο, Νοέμβριος 2010

Ένα Εργαλείο για Ηλεκτρονική Μάθηση Βασισμένο σε Τεχνικές Συλλογισμού.

ΕΙΡΗΝΗ Α. ΦΙΛΙΟΠΟΥΛΟΥ

Μεταπτυχιακή Εργασία

Πανεπιστήμιο Κρήτης

Τμήμα Επιστήμης Υπολογιστών

Περίληψη

Στην εποχή μας οι αλλαγές πραγματοποιούνται με ραγδαίους ρυθμούς. Η νέα οικονομία βασίζεται στη γνώση. Οι παραδοσιακές μορφές εκπαίδευσης, δεν επαρκούν πλέον για να καλύψουν τις ανάγκες της σύγχρονης οικονομίας, που ανανεώνονται σταδιακά με την υιοθέτηση καινοτόμων μεθόδων εκπαίδευσης. Στην εξ αποστάσεως εκπαίδευση σχεδόν επιβάλλεται η χρήση ηλεκτρονικής μάθησης λόγω των δυνατοτήτων που προσφέρει. Επιπλέον είναι εφικτή η διεξαγωγή ενός μαθήματος με το συνδυασμό της παραδοσιακής μάθησης με την ηλεκτρονική. Στην ηλεκτρονική εκπαίδευση ένα από τα πιο σημαντικά ζητήματα θεωρείται η ύπαρξη διαδικασιών που διευκολύνουν και αναπτύσσουν την αλληλεπίδραση μεταξύ εκπαιδευτή και εκπαιδευόμενου. Η ποιότητα της επικοινωνίας έχει καθοριστική σημασία τόσο στην εκπαιδευτική διεργασία όσο και την επιτυχία σπουδών των εκπαιδευόμενων.

Στην παρούσα εργασία αναπτύχθηκε ένα σύστημα που χρησιμοποιείται ως ενδιάμεσο της επικοινωνίας μεταξύ εκπαιδευτή και εκπαιδευόμενου που εφαρμόζει τεχνικές συλλογιστικής βασισμένες σε περιπτώσεις κειμένου, ώστε να αυτοματοποιηθεί η διαδικασία απάντησης των απορριών 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 εκπαιδευομένους ελαχιστοποιείται η αναμονή της απάντησής του. Το σύστημα βασίζεται στη χρήση της προϋπάρχουσας γνώσης για την επίλυση των απορριών 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 απορριών στο μέλλον. Η εφαρμογή του συστήματος μπορεί να πραγματοποιηθεί για τη διδασκαλία μαθημάτων όλων των θεματικών ενοτήτων.

A framework for electronic education based in reasoning mechanic

IRENE A. FILIOPOULOU

Master's Thesis

University of Crete

Computer Science Department

Abstract

Change today occurs at a very fast pace. New economies are based on knowledge. The traditional forms of education are no longer sufficient to address the needs of modern economy, which are gradually updated by adopting innovative educational methods. Nowadays distance learning almost requires the use of e-learning because of the possibilities it offers. Furthermore, it is possible to conduct a course combining traditional learning with e-learning. One of the most important aspects of e-learning is considered to be the existence of processes which facilitate and increase the interaction between tutor and student. The quality of communication is critical both for the educational process as well as for the success of the students' studies.

At the current master thesis we developed a system which is used as an intermediate step in the communication between tutor and student to automate the process of answering the student's questions that uses techniques on textual case based reasoning, and which benefits both the tutor and the student. The automation of this process significantly decreases the work load and involvement required by the tutor, whereas for the students it minimizes the time required to receive an answer. The system is based on the use of preexisting knowledge to answer questions asked by the students. For each new question, the system checks if there is knowledge stored for it by checking previous similar cases, which have been answered by the tutor in the past. If the system finds preexisting knowledge which can answer the question, it sends it to the student without requiring any action by the tutor. On the other hand, if no relevant information is found, the question will be answered by the tutor and it will be stored in the system, so similar questions can be automatically answered in the future. The system can be used for teaching courses in all subject fields.

Ευχαριστίες

Στο σημείο αυτό θα ήθελα να ευχαριστήσω, όλους όσους με βοήθησαν να φέρω σε πέρας την παρούσα εργασία. Αρχικά θα ήθελα να ευχαριστήσω τον επόπτη καθηγητή μου κ. Χρήστο Νικολάου για την βοήθεια και υποστήριξη του. Στη συνέχεια θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Δημήτρη Πλεξουσάκη για την πολύτιμη βοήθεια και καθοδήγηση που μου πρόσφερε συμβάλλοντας σημαντικά στην ολοκλήρωση της παρούσας εργασίας.

Θα ήθελα επίσης να ευχαριστήσω τον καθηγητή κ. Καλογιαννάκη Μιχάλη, τρίτο μέλος της επιτροπής γιατί με τίμησε με την παρουσία του ως μέλος αυτής. Θα ήθελα ακόμα να ευχαριστήσω τον καθηγητή Νικόλαο Παπαδάκη για τη βοήθεια που μου παρείχε και την άψογη συνεργασία μας.

Οφείλω ένα μεγάλο ευχαριστώ στους φίλους μου για τη συμπαράσταση και κατανόηση τους καθ' όλη τη διάρκεια ολοκλήρωσης της εργασίας. Ιδιαίτερα θα ήθελα να ευχαριστήσω τη Μαριάνα Καρμαζή για την υποστήριξη της ακόμα και τις πιο ακατάλληλες ώρες, το Νίκο Κόλλια, το Νίκο Γρίσπο και τη Μάρθα Τζήρου.

Θα ήθελα ακόμα να ευχαριστήσω τον αδερφό μου Γεώργιο που μου συμπαραστάθηκε ουσιαστικά σε δύσκολες στιγμές των μεταπτυχιακών μου σπουδών με την ανιδιοτελή αγάπη του, αλλά και τον αδερφό μου Ευθύμιο και τη νύφη μου Μαρία για την αγάπη και το ενδιαφέρον τους.

Ωστόσο το μεγαλύτερο ευχαριστώ το χρωστάω στους γονείς μου Αλέξανδρο και Χαρίκλεια για την αμέριστη αγάπη, υποστήριξη και εμπιστοσύνη που μου παρείχαν απλόχερα όλα τα χρόνια των σπουδών μου. Τους ευχαριστώ που ήταν πάντα δίπλα μου με υπομονή και κατανόηση και αποτέλεσαν για μένα πηγή για να συνεχίζω να αγωνίζομαι και να εκπληρώνω τους στόχους μου.

Πίνακας Περιεχομένων

Ευχαριστίες 7
Πίνακας Περιεχομένων 8
Πίνακας Εικόνων 11
Πίνακας Πινάκων 12
1 Εισαγωγή 13
1.1 Σκοπός 14
2 Εξ Αποστάσεως Εκπαίδευση 15
2.1 Ορισμός της ηλεκτρονικής μάθησης (e-Learning) 16
2.2 Εκπαιδευτικά Πρότυπα 17
2.3 Μορφές Ηλεκτρονικής Μάθησης 17
2.4 Συστήματα Διαχείρισης Μάθησης (LMS) 18
2.5 Πλατφόρμες e-Learning 20
2.6 Εργαλεία Σύγχρονης Επικοινωνίας και Ανάγκη που Δημιουργείται 25
3 Συστήματα Γνώσης 27
3.1 Λογικός Προγραμματισμός 28
3.2 Rule-Based Systems 29
3.3 Συστήματα βασισμένα στη Συλλογιστική των Περιπτώσεων (Case-Based Reaoning) 30
3.4 Πλεονεκτήματα και μειονεκτήματα 30
4 Θεωρητικό Υπόβαθρο 32
4.1 Συλλογιστική Βασισμένη σε Περιπτώσεις CBR (Case Based Reasoning) 32
4.2 Συλλογιστική Βασισμένη σε Περιπτώσεις Κειμένου TCBR(Textual Case Based Reasoning) 34
4.3 Συστήματα CBR 35
4.4 Σύστημα CBR - Jcolibri 36
4.4.1 Σημασιολογική (semantic) Μέθοδος 37
4.4.2 Στατιστική (statistical) Μέθοδος 38
5 Το Σύστημα μας 39
5.1 Δική μας Επέκταση 40
5.2 Στάδιο προ-επεξεργασίας 40
5.2.1 Γλωσσική προ-επεξεργασία (Λεξιλογική Ανάλυση) 41
5.2.2 Αποκλεισμός Λέξεων(stopwords) 41
5.2.3 Στελέχωση Κειμένου (Stemming) 42

5.3 Ανάκτηση Πληροφορίας ... 42
5.3.1 Lucene ... 43
5.3.1.1 Διαδικασία δημιουργίας Καταλόγου ... 43
5.3.2 Υπολογισμός Σημαντικότητας Όρων(weighting) ... 45
5.3.3 Συναρτήσεις ομοιότητας ... 46
5.4 Μέθοδοι Διαχείρισης Της Γνώσης ... 49
5.4.1 Στατιστική Μέθοδος ... 49
5.4.2 Μέθοδος Ευρετηρίου ... 50
5.4.3 Μέθοδος Ευρετηρίου με τη χρήση Συνώνυμων Όρων ... 52
5.4.3.1 Το Ηλεκτρονικό Λεξικό WordNet ... 52
5.4.4 Μέθοδος Ευρετηρίου με τη χρήση Συνώνυμων Όρων θέτοντας βάρη. ... 53
6 Υλοποίηση ... 56
6.1 Efront ... 57
6.2 Επεξεργασία και Μετατροπή του αρχείου xml ... 59
6.3 Επεξεργασία Ερωτήματος ... 61
7 Μετρήσεις - Αξιολόγηση ... 65
7.1 Εφαρμογή στον κλάδο της Πληροφορικής ... 66
7.1.1 Αξιολόγηση Μεθοδολογιών στον κλάδο της Πληροφορικής ... 67
7.2 Εφαρμογή στον κλάδο της Ψυχολογίας ... 70
7.2.1 Αξιολόγηση Μεθοδολογιών στον τομέα της Ψυχολογίας ... 71
7.3 Αποτελέσματα Αξιολόγησης ... 74
8 Επίλογος – Συμπεράσματα – Μελλοντική Δουλειά ... 80
8.1 Συμπεράσματα ... 80
8.2 Μελλοντική Δουλειά ... 81
8.2.1 Καταγραφή της επίδοσης των εκπαιδευόμενων ... 81
8.2.2 Προσαρμογή του συστήματος στην ελληνική γλώσσα ... 82
8.2.3 Προσαρμογή περιβάλλοντος Efront ... 82
8.2.4 Παροχή Εξατομικευμένης Πρόσβασης σε Υπηρησίες εξ Αποστάσεως Εκπαίδευσης(e-Learning Personalization) ... 82
Παράρτημα Α ... 84
Παράρτημα Β ... 84
Αναφορές ... 86

Πίνακας Εικόνων

Εικόνα 1 - Προβολή ATutor Συστήματος....21
Εικόνα 2 - Παρουσίαση του Συστήματος Claroline....22
Εικόνα 3 - Προβολή Συστήματος CoMPUs....23
Εικόνα 4 - Παρουσίαση Συστήματος Dokeos....23
Εικόνα 5 - Προβολή Κεντρικής Σελίδας του συστήματος Moodle....24
Εικόνα 6 - Ο κύκλος Ζωής του CBR....34
Εικόνα 7 - Στάδια Προ-επεξεργασίας....41
Εικόνα 8 - Παράδειγμα Λεξιλογικής Ανάλυσης....41
Εικόνα 9 - Αρχιτεκτονική Συστήματος....56
Εικόνα 10 - Στιγμιότυπο του Φόρουμ στο Efront....58
Εικόνα 11 - Αρχείο xml που εξάγεται από τη βάση του φόρουμ....59
Εικόνα 12 - Στιγμιότυπο Φόρουμ προσθέτοντας τις Λέξεις-Κλειδιά....61
Εικόνα 13 - Αρχιτεκτονική Επεξεργασίας Ερωτήματος....62
Εικόνα 14- Αποτελέσματα Αξιολόγησης για τη Βάση στην Επιστήμη Υπολογιστών....68
Εικόνα 15 - Αποτελέσματα Αξιολόγησης για τη Βάση στη Ψυχολογία....71
Εικόνα 16 - Συνολική Αξιολόγηση και από τις δύο Βάσεις....74
Εικόνα 17 - Συμπεριφορά Τελικών Αποτελεσμάτων από τις Μεταβολές του Συντελεστή Χ....77
Εικόνα 18 - Μέσος Όρος Βαθμολογίας όλων των Ιδανικών Περιπτώσεων Ανάλογα τον Συντελεστή Χ..78

Πίνακας Πινάκων

Πίνακας 1 - Παράδειγμα Ανεστραμμένου Καταλόγου.....45
Πίνακας 2 - Μορφή Ανεστραμμένου Καταλόγου 63

1 Εισαγωγή

Η ηλεκτρονική μάθηση μπορεί να εφαρμόζεται σε όλες τις βαθμίδες της εκπαίδευσης. Στην εξ αποστάσεως εκπαίδευση σχεδόν επιβάλλεται η χρήση ηλεκτρονικής μάθησης λόγω των δυνατοτήτων που προσφέρει. Επιπλέον είναι εφικτή η διεξαγωγή ενός μαθήματος με το συνδυασμό της παραδοσιακής μάθησης με την ηλεκτρονική. Ο συνδυασμός αυτός ονομάζεται μεικτή μάθηση ή υβριδική μάθηση. Σ' αυτό το εκπαιδευτικό μοντέλο οι εκπαιδευόμενοι ξεκινούν συνήθως την διαδικασία εκπαίδευσης σε συμβατικές εκπαιδευτικές δομές και στην συνέχεια έχουν πρόσβαση στο περιεχόμενο αλλά και επικοινωνούνε μεταξύ τους και με τον καθηγητή με την χρήση των τεχνολογιών του διαδικτύου. Στην ηλεκτρονική εκπαίδευση ένα από τα πιο σημαντικά ζητήματα θεωρείται η ύπαρξη διαδικασιών που διευκολύνουν και αναπτύσσουν την αλληλεπίδραση μεταξύ εκπαιδευτή και εκπαιδευόμενου. Η αλληλεπίδραση περιλαμβάνει την επικοινωνία και υποστήριξη του διδάσκοντα ως προς τον εκπαιδευόμενο. Η ποιότητα της επικοινωνίας έχει καθοριστική σημασία τόσο στην εκπαιδευτική διεργασία όσο και την επιτυχία σπουδών των εκπαιδευόμενων.

Στα πλαίσια των περισσοτέρων μαθημάτων που διδάσκονται στην τριτοβάθμια εκπαίδευση χρησιμοποιούνται εργαλεία για την επικοινωνία εκπαιδευτή με εκπαιδευόμενο. Στο τμήμα επιστήμης υπολογιστών του πανεπιστημίου Κρήτης σε κάποια μαθήματα χρησιμοποιούνται mailing list και σε κάποια άλλα διάφορα Συστήματα Διαχείρισης Μάθησης για την επικοινωνία αυτή. 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. Στο μάθημα του ΗΥ-100 Εισαγωγή στην Επιστήμη Υπολογιστών χρησιμοποιείται το Σύστημα Διαχείρισης Μάθησης e-Front. Η πλατφόρμα του e-Front μέσα σε ένα σύνολο υπηρεσιών που προσφέρει, παρέχει την υπηρεσία ενός φόρουμ μέσω του οποίου πραγματοποιείται η επικοινωνία εκπαιδευτή-εκπαιδευόμενου. Στο φόρουμ ο εκάστοτε εκπαιδευόμενος που είναι εγγεγραμμένος στο μάθημα μπορεί να δημοσιεύει τις ερωτήσεις – απορίες του και να περιμένει από τον εκπαιδευτή του μαθήματος να δώσει την απάντηση.

Το πιο σύνηθες φαινόμενο που παρατηρείται είναι το φόρουμ να γεμίζει από ερωτήσεις που οι περισσότερες από αυτές είναι είτε παρόμοιες είτε παρεμφερείς με άλλες που έχουν αναρτηθεί στο παρελθόν και έχει δοθεί απάντηση. Ένας εκπαιδευόμενος που έχει κάποια απορία σχετικά με την θεωρία ή με τις ασκήσεις προτιμάει να δημοσιεύσει κατευθείαν την ερώτηση του παρά να αναζητήσει όλο το φόρουμ για κάποια παρόμοια απορία, που μπορεί και να μην υπάρχει. Για το εκπαιδευτικό προσωπικό που ασχολείται με την διαχείριση του forum, ο φόρτος αυτός είναι πάρα πολύ μεγάλος, συνήθως απαιτούνται αρκετές ώρες ημερησίως κυρίως αν το πλήθος των εκπαιδευόμενων είναι μεγάλο. Από την άλλη μεριά για τους εκπαιδευόμενους η πρώτη περίπτωση δηλαδή να διαβάσει όλες τις αναρτήσεις που έχουν γίνει σε κάποιο θέμα και ενδεχομένως και σε παραπάνω για να δει αν έχει απαντηθεί η απορία του, απαιτεί αρκετό χρόνο, ενώ στην δεύτερη περίπτωση που αναρτήσει την ερώτηση του, περιμένει από τον εκπαιδευτή απάντηση, που και πάλι ο χρόνος αναμονής της απάντησης ενδέχεται να είναι μεγάλος.

Στην παρούσα εργασία 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). Η συλλογιστική βασισμένη σε Περιπτώσεις(CBR) ή Συλλογιστική των Περιπτώσεων όπως χρησιμοποιείται στην ελληνική βιβλιογραφία είναι ένας τομέας την Μηχανικής Μάθησης και των Συστημάτων Γνώσης. Η Βασισμένη σε Περιπτώσεις Συλλογιστική είναι μια μέθοδος για επίλυση προβλημάτων που η αποτελεσματικότητα της βασίζεται στην ικανότητα να μαθαίνει από κάθε καινούργια περίπτωση. Τα συστήματα Συλλογιστικής βασισμένα σε Περιπτώσεις έχουν σημειώσει ραγδαία ανάπτυξη την τελευταία δεκαπενταετία.

Ένα TCBR ουσιαστικά είναι ένα σύστημα συλλογιστικής βασισμένης σε περιπτώσεις(CBR), όπου όλο ή ένα μέρος της γνώσης που χρησιμοποιείται είναι με την μορφή κειμένου. Αυτού του είδους συστήματα είναι πολύ ενδιαφέροντα γιατί οι περισσότερες περιπτώσεις προβλημάτων με τις λύσεις τους είναι διαθέσιμες σε μορφές κειμένου. Η ενσωμάτωση τεχνικών Textual Cased Based Reasoning στο σύστημα, ουσιαστικά αξιοποιεί τη προϋπάρχουσα γνώση ανάλογων περιπτώσεων και αυτοματοποιεί τη διαδικασία απάντησης των απορριών. Αυτό θα έχει σαν αποτέλεσμα οι εκπαιδευόμενοι να μπορούν να επιλύσουν κάποια απορία τους άμεσα, χωρίς την απαραίτητη παρέμβαση του εκπαιδευτή.

1.1 Σκοπός

Σκοπός της εργασίας είναι η δημιουργία ενός συστήματος που θα αυτοματοποιήσει τη διαδικασία επίλυσης απορριών 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 επιστρέφετε 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 διάφορων θεματικών ενοτήτων πέραν του μαθήματος του ΗΥ-100 Εισαγωγή στην Επιστήμη των Υπολογιστών, 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.

2 Εξ Αποστάσεως Εκπαίδευση

Στην εποχή μας οι αλλαγές πραγματοποιούνται με ραγδαίους ρυθμούς. Η νέα οικονομία βασίζεται στη γνώση. Σε όλα τα επίπεδα της κοινωνίας γίνεται αισθητή όλο και περισσότερο η επίδραση της πληροφορίας και της επικοινωνίας. Οι σύγχρονες κοινωνίες γίνονται πολυσυλλεκτικές, και επιβάλλουν την ανάγκη για συνεχιζόμενη κατάρτιση και δια βίου μάθηση, εντείνοντας την αναζήτηση νέων και πιο ευέλικτων μοντέλων επιμόρφωσής και εκπαίδευσης. Οι παραδοσιακές μορφές εκπαίδευσης, δεν επαρκούν πλέον για να καλύψουν τις ανάγκες της σύγχρονης οικονομίας, ανανεώνονται σταδιακά με την υιοθέτηση καινοτόμων μεθόδων εκπαίδευσης.

Το μοντέλο που αποτελεί μια αξιόπιστη λύση είναι αυτό της εξ αποστάσεως εκπαίδευσης. Η εκπαίδευση εξ αποστάσεως συνήθως απευθύνεται σε ενήλικα άτομα που έχουν την δυνατότητα της αυτοδιαχείρισης ενώ ταυτόχρονα δεν μπορούν να παρακολουθήσουν τον κλασικό τρόπο εκπαίδευσης που απαιτεί τη φυσική του παρουσία, λόγω του ότι υπόκεινται σε διάφορους περιορισμούς. Ένας από τους πιο κύριους περιορισμούς είναι ο χρόνος που μπορούν να διαθέσουν τα ενδιαφερόμενα άτομα για την εκπαίδευσή τους λόγω εργασιακών, κοινωνικών ή οικογενειακών υποχρεώσεών τους [12], ακόμα και ο τόπος διαμονής τους που ενδέχεται να είναι απομακρυσμένος από εκπαιδευτικά ιδρύματα. Μέσα στους περιορισμούς πρέπει να αναφερθούν και άλλα χαρακτηριστικά όπως δυνατότητες, προσόντα, αδυναμίες, ειδικές ανάγκες επιμόρφωσης κ.ά.

Ο αριθμός των ανθρώπων που επωφελούνται από την εξ αποστάσεως εκπαίδευση είναι συνεχώς αυξανόμενος, έτσι και η ανάπτυξη των συστημάτων που την υποστηρίζουν ήταν ραγδαία. Η εξέλιξη των συστημάτων αυτών επωφελήθηκε σημαντικά από την παράλληλη εξέλιξη των υπολογιστών, των επικοινωνιών και φυσικά του διαδικτύου. Αυτό είχε σαν αποτέλεσμα πολύ σύντομα τα συστήματα παροχής εξ αποστάσεως εκπαίδευσης να μετατραπούν σε συστήματα παροχής ηλεκτρονικής εκπαίδευσης εξ αποστάσεως(e-Learning Systems).

Στην εποχή μας ο όρος εξ αποστάσεως ηλεκτρονική εκπαίδευση [23] σημαίνει τη διαδικασία παροχής ή κατάρτισης από απόσταση μέσα από σύγχρονες τεχνολογίες που υποστηρίζονται ηλεκτρονικά, όπως π.χ. από προγράμματα υπολογιστών. Η εξ αποστάσεως εκπαίδευση αναδεικνύεται σε ενδιαφέρον συστατικό στοιχείο των εκπαιδευτικών συστημάτων, ιδιαίτερα για την ανώτερη/ανώτατη εκπαίδευση. Έτσι, όταν πλέον γίνεται αναφορά στην εκπαίδευση από απόσταση (ή τηλεκπαίδευση) εννοείται κυρίως η εκπαίδευση που υλοποιείται με τη χρήση των Τεχνολογιών της Πληροφορίας και των Επικοινωνιών.

Τα e-Learning συστήματα στηρίζονται σε μια ανοιχτή και ευέλικτη φιλοσοφία όπου ο σύγχρονος άνθρωπος πρέπει να έχει την δυνατότητα να μαθαίνει με πολλαπλούς τρόπους, να έχει ίσες ευκαιρίες για εκπαίδευση και κατάρτιση, απαλλαγμένες από δεσμεύσεις που αφορούν το χώρο και το χρόνο, να έχει επιλογές στο πώς και στο τι θα μαθαίνει και γενικότερα να αποτελεί το κέντρο της μαθησιακής διαδικασίας[43].

2.1 Ορισμός της ηλεκτρονικής μάθησης (e-Learning)

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.

Η ηλεκτρονική μάθηση βασίζεται στα εξής:

- Είναι δικτυακή και έτσι είναι εφικτή η άμεση ενημέρωση, αποθήκευση, ανάκτηση και η παροχή του εκπαιδευτικού υλικού.
- Παρέχεται στον εκπαιδευόμενο μέσω ηλεκτρονικού υπολογιστή με τη χρήση του διαδικτύου.
- Επικεντρώνεται στην ευρύτερη έννοια της μάθησης, πέρα από τα παραδοσιακά πρότυπα της εκπαίδευσης. Δεν περιορίζεται απλά σε εκπαίδευση που βασίζεται στον υπολογιστή, αλλά περιλαμβάνει και μεταφορά εργαλείων και πληροφοριών που βελτιώνουν την απόδοση.

Στην ηλεκτρονική εκπαίδευση ιδιαίτερα σημαντικό θεωρείται η ύπαρξη διαδικασιών που προωθούν και αναπτύσσουν την αλληλεπίδραση μεταξύ εκπαιδευτή και εκπαιδευόμενου ή την αλληλεπίδραση μεταξύ των εκπαιδευομένων, στοχεύοντας στην ατομική ή συνεργατική εκμάθηση (collaborative learning). Ο εκπαιδευόμενος έχει συνήθως πλήρη έλεγχο του ρυθμού προόδου, ενώ ταυτόχρονα υπάρχει διαθέσιμη, αν το επιθυμεί, υποστήριξη από τον εκπαιδευτή ή ειδικό του θέματος. Η υποστήριξη είναι απαραίτητη, αφού στην αντίθετη περίπτωση θα μιλούσαμε μόνο για *αυτοεκπαίδευση*, διαδικασία που μπορεί να γίνει και με άλλα μέσα, π.χ. βιβλίο ή CD-ROM.

Η συνεργατική εκμάθηση αναφέρεται σε μεθοδολογίες και περιβάλλοντα στα οποία οι εκπαιδευόμενοι αναλαμβάνουν εργασίες που αλληλοεξαρτώνται. Εκπαιδευόμενοι χωρισμένοι σε ομάδες δουλεύουν μαζί στην αναζήτηση πληροφορίας, στη κατανόηση του νοήματος ή στη δημιουργία ενός μαθησιακού αποτελέσματος. Η συνεργατική εκπαίδευση είναι σχετική και με τη συνεργαζόμενη εκπαίδευση (cooperative learning)[29]. Οι δραστηριότητες της συνεργαζόμενης μάθησης περιλαμβάνουν συνεργατικές δράσεις: ομαδικές εργασίες, κοινή παράδοση εργασιών, συγγραφή στο ίδιο μέσο κ.ά.

Για την υποστήριξη της ηλεκτρονικής εκπαίδευσης αναπτύχθηκαν συστήματα ευρέως γνωστά ως *Συστήματα Διαχείρισης Μάθησης* (*Learning Management Systems*, *LMSs*). Στη βιβλιογραφία συναντάμε και τους όρους Εικονικό Σύστημα Μάθησης (Virtual Learning System, VLE), Σύστημα Διαχείρισης Μαθημάτων (Course Management System, CMS) και Σύστημα Διαχείρισης Μαθησιακού Περιεχομένου (Learning Content Management System, LCMS). Τα LMSs είναι συστήματα που οργανώνουν και παρέχουν πρόσβαση σε online υπηρεσίες μάθησης για η-εκπαιδευόμενους (e-Learners), η-εκπαιδευτές(e-Tutors) αλλά και διαχειριστές. Οι υπηρεσίες αυτές αφορούν εργαλεία για υποστήριξη της ηλεκτρονικής

μάθησης και πιο συγκεκριμένα διαχείριση χρηστών, επικοινωνία μεταξύ εκπαιδευτικού και εκπαιδευομένων, παράδοση εκπαιδευτικού υλικού που αναφέρεται σε μία κατάλληλη σειρά μαθημάτων, αξιολόγηση και παρακολούθηση της πορείας των εκπαιδευομένων και παροχή υποστηρικτικών εργαλείων για τον εκπαιδευτικό.

Τα συστήματα αυτά λόγω του πλήθους των παροχών και της σταθερής συμπεριφοράς τους εφαρμόζονται πλέον σε ολοένα και περισσότερα εκπαιδευτικά ιδρύματα της χώρας στηρίζοντας ικανοποιητικά τα προγράμματα σπουδών τους.

2.2 Εκπαιδευτικά Πρότυπα

Τα εκπαιδευτικά πρότυπα (standards) έχουν στόχο να εξασφαλίσουν τη διαλειτουργικότητα ανάμεσα στα διάφορα συστήματα που αναπτύσσονται και εφαρμόζονται ώστε να συνδράμουν στην εξέλιξη και στην ευρεία αποδοχή των συστημάτων που εφαρμόζονται. Για τον λόγο αυτό δημιουργήθηκαν πρότυπα και στα συστήματα ηλεκτρονικής εκπαίδευσης. Γενικότερα, με τα πρότυπα:

- Παρέχεται η δυνατότητα επαναχρησιμοποίησης των εκπαιδευτικών αντικειμένων ανάμεσα σε διαφορετικές πλατφόρμες διαχείρισης μάθησης, μειώνοντας το κόστος κατασκευής τους,
- Είναι εφικτή η θέσπιση ενός ενιαίου τρόπου περιγραφής των εκπαιδευτικών αντικειμένων με τον οποίο τα συστήματα διαχείρισης εκπαιδευτικών αντικειμένων μπορούν να συνεργάζονται,
- Συντονίζονται όλες οι αλληλεπιδρώμενες συνιστώσες που αποτελούν τα θεμέλια των e-Learning συστημάτων: Εργαλεία, Περιεχόμενο, Διαχείριση, Εκπαιδευτές, Δημιουργοί Εφαρμογών.
- Καθορίζονται τα κριτήρια για μέτρηση της απόδοσής τους.

Πρότυπα έχουν οριστεί για την περιγραφή των εκπαιδευτικών αντικειμένων και για την περιγραφή των εκπαιδευόμενων.

2.3 Μορφές Ηλεκτρονικής Μάθησης

Η ηλεκτρονική μάθηση μπορεί να εφαρμόζεται σε όλες τις βαθμίδες της εκπαίδευσης. Στην εξ αποστάσεως εκπαίδευση σχεδόν επιβάλλεται η χρήση ηλεκτρονικής μάθησης λόγω των δυνατοτήτων που προσφέρει. Επιπλέον είναι εφικτή η διεξαγωγή ενός μαθήματος με τον συνδυασμό της παραδοσιακής μάθησης με την ηλεκτρονική. Ο συνδυασμός αυτός ονομάζεται μεικτή μάθηση ή υβριδική μάθηση. Έτσι η ηλεκτρονική μάθηση μπορεί να διακριθεί σε τρεις μορφές ανάλογα με τον τρόπο επικοινωνίας των συμμετεχόντων και τις μαθησιακές ανάγκες: στην σύγχρονη , στην ασύγχρονη και στην μικτή μάθηση.

Στην **σύγχρονη μάθηση** παρέχεται η δυνατότητα αλληλεπίδρασης σε πραγματικό χρόνο μεταξύ εκπαιδευομένου και εκπαιδευτή, μέσω της χρήσης τεχνολογιών. Η εκπαίδευση αυτού του τύπου έχει πολλά κοινά σημεία με την εκπαίδευση σε αίθουσα διδασκαλίας, με τη διαφορά ότι στην σύγχρονη μάθηση του e-learning, εκπαιδευτής και εκπαιδευόμενος βρίσκονται σε μια “εικονική” αίθουσα. Αυτού

του τύπου η εκπαίδευση μπορεί να καταγραφεί σε ηλεκτρονική μορφή και να ξαναχρησιμοποιηθεί αργότερα αν χρειαστεί. Το συγκεκριμένο είδος μάθησης θεωρείται απαγορευτικό γιατί το κόστος εξοπλισμού είναι μεγάλο, απαιτεί κατάλληλη υποδομή και προσωπικό υποστήριξης.

Ασύγχρονη μάθηση εννοείται η μορφή ηλεκτρονικής μάθησης κατά την οποία δεν υπάρχει άμεση, «σύγχρονη» επικοινωνία του εκπαιδευτή και του εκπαιδευόμενου κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Στην περίπτωση της ασύγχρονης εκπαίδευσης ο εκπαιδευτής μπορεί να προετοιμάσει το εκπαιδευτικό υλικό και ο εκπαιδευόμενος, μπορεί να πάρει αυτό το εκπαιδευτικό υλικό οπουδήποτε και αν βρίσκεται και σε όποια χρονική στιγμή επιλέξει ο ίδιος. Φαίνεται εκ πρώτης όψεως ότι η εκπαίδευση αυτού του τύπου είναι πιο ευέλικτη αλλά στην περίπτωση αυτή δεν υπάρχει επικοινωνία εκπαιδευτή – εκπαιδευόμενου σε πραγματικό χρόνο (real-time).

Η τελευταία μορφή ηλεκτρονικής μάθησης είναι η **Μικτή Μάθηση**. Μικτή μάθηση ονομάζεται το μαθησιακό μοντέλο εκπαίδευσης στο οποίο χρησιμοποιούνται για την υλοποίησή του και τεχνολογίες ασύγχρονης επικοινωνίας αλλά και συμβατικές δομές εκπαίδευσης. Σ' αυτό το εκπαιδευτικό μοντέλο οι εκπαιδευόμενοι ξεκινούν συνήθως την διαδικασία εκπαίδευσης σε συμβατικές εκπαιδευτικές δομές και στην συνέχεια έχουν πρόσβαση στο περιεχόμενο αλλά και επικοινωνούνε μεταξύ τους και με τον καθηγητή με την χρήση των τεχνολογιών του διαδικτύου.

2.4 Συστήματα Διαχείρισης Μάθησης (LMS)

Τα Συστήματα Διαχείρισης Μάθησης(ΣΔΜ) είναι λογισμικά και εμφανίστηκαν στο δεύτερο μισό της δεκαετίας του '90. Τα ΣΔΜ 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.

Κάποιες πλατφόρμες υποστηρίζουν τρεις τύπους μαθημάτων, ανάλογα με τη βούληση του διδάσκοντα: τα ανοικτά, τα ανοικτά σε εγγραφή και τα κλειστά.

1. Τα ανοικτά μαθήματα είναι τα μαθήματα ελεύθερης πρόσβασης για οποιονδήποτε χρήστη, ακόμα κι αν δεν έχει κάνει εγγραφή.
2. Τα ανοικτά σε εγγραφή είναι τα μαθήματα στα οποία μπορεί να έχει πρόσβαση ένας χρήστης μόνο εφόσον έχει κάνει εγγραφή στην πλατφόρμα και διαθέτει κάποιον κωδικό πρόσβασης.
3. Κλειστά θεωρούνται τα μαθήματα στα οποία μπορεί να εγγραφεί ένας χρήστης που διαθέτει κωδικό πρόσβασης στην πλατφόρμα, μόνο αν του το επιτρέψει ο υπεύθυνος καθηγητής.

Οι χρήστες ενός ΣΔΜ χωρίζονται στους εκπαιδευόμενους (learners), τους εκπαιδευτές (instructors) και τους διαχειριστές (administrators), η είσοδος στο σύστημα καθορίζεται από το διακριτό ρόλο που τους έχει ανατεθεί και που καθορίζει πλήρως τα δικαιώματα και τις δυνατότητές τους.

- Ο διαχειριστής έχει τη δυνατότητα να επέμβει σε κάθε διαδικασία του συστήματος (διαχειριστική ή και εκπαιδευτική).
- Ο εκπαιδευτής είναι ο υπεύθυνος του μαθήματος. Από αυτόν απαιτείται η δημιουργία και ενημέρωση των τάξεων και του υλικού, η προσωπική επικοινωνία με τους εκπαιδευόμενους, η διαχείριση των ομάδων συζήτησης και του πίνακα ανακοινώσεων και η παρακολούθηση της εργασίας και των επιδόσεων των εκπαιδευομένων.

Οι εκπαιδευόμενοι είναι οι εγγεγραμμένοι χρήστες, που έχουν τη δυνατότητα παρακολούθησης ενός μαθήματος, συμμετέχοντας στις δραστηριότητες που το συναποτελούν. Έχουν πρόσβαση στο εκπαιδευτικό υλικό του μαθήματος, δυνατότητα αυτοαξιολόγησης μέσα από τεστ αυτόματης διόρθωσης και δυνατότητες επικοινωνίας με τον εκπαιδευτή και άλλους εκπαιδευόμενους.

Στα ΣΔΜ οι λειτουργίες που διαθέτουν παρουσιάζουν κοινά χαρακτηριστικά, όπως ανοικτή αρχιτεκτονική, μαθησιακά και διαχειριστικά εργαλεία για τον εκπαιδευτή και τον εκπαιδευόμενο, προσφέροντας ενιαίο λογισμικό περιβάλλον για όλους (κατασκευαστές, εκπαιδευτικούς, μαθητές) και υποστηρίζοντας δραστηριότητες με σκοπό την ευελιξία τόπου και χρόνου. κ.ά.

Τα περισσότερα ΣΔΜ αποτελούνται από πολλά επιμέρους τμήματα. Τα πιο κοινά στα περισσότερα είναι τα εξής:

- Αναγνώριση χρήστη με μοναδικούς κωδικούς για λόγους ασφαλείας και οργάνωσης.
- Οργανόγραμμα μαθήματος, που περιλαμβάνει σημαντικά σημεία της διαδικασίας του μαθήματος.
- Ημερολόγιο για τις σημαντικές ημερομηνίες του προγράμματος.
- Λίστα τάξης και προσωπικές σελίδες συμμετεχόντων.
- Ηλεκτρονικό ταχυδρομείο.
- Πίνακες Ανακοινώσεων.
- Εργαλεία για σύγχρονη επικοινωνία (forum, chat, teleconference, whiteboards)
- Μεταδεδομένα, για λόγους κατηγοριοποίησης και ταξινόμησης, π.χ. λέξεις κλειδιά.
- Εργασίες, που προετοιμάζονται από τον διδάσκοντα με δυνατότητα συλλογής των αποτελεσμάτων.
- Αξιολόγηση – Βαθμολόγιο
- Περιοχή αποθήκευσης αρχείων
- Εργαλεία αναζήτησης
- Μοντέλο πλοήγησης στο περιβάλλον εργασίας
- Καταγραφή στατιστικών στοιχείων
- Γλωσσάρι
- Βοήθεια

Τα Συστήματα Διαχείρισης Μάθησης μπορεί να είναι είτε εμπορικά είτε κατασκευασμένα κατά παραγγελία για να υπηρετήσουν συγκεκριμένες ανάγκες είτε ανοιχτού κώδικα. Αποτελούν έναν ταχύτατα εξελισσόμενο χώρο, όπου αναφύονται καινούριοι κατασκευαστές. Η προσφερόμενη πληθώρα παρέχει μεγάλη δυνατότητα επιλογής, ανάλογα με τις ανάγκες αλλά και παράγοντες όπως το κόστος λειτουργίας και συντήρησης της υπηρεσίας ή των πόρων που αυτή απαιτεί για την ικανοποιητική και αποδοτική λειτουργία της. Ως προγράμματα ηλεκτρονικού υπολογιστή, τα ΣΔΜ θεωρούνται πνευματικά έργα και προστατεύονται από τη νομοθεσία περί προστασίας των πνευματικών δικαιωμάτων. Ειδικά για το χώρο των εκπαιδευτικών ιδρυμάτων, τα θέματα κόστους αποτελούν σημαντικό παράγοντα κατά την επιλογή μιας πλατφόρμας τηλεκπαίδευσης, ιδιαίτερα αν πέρα από την αρχική δαπάνη για την αγορά του προϊόντος υπάρχει επιπλέον κόστος για συγκεκριμένο αριθμό αδειών χρήσης (licences) ή/και ετήσια συνδρομή για ανανέωση της χρήσης. Γι' αυτό πολλά ιδρύματα και φορείς επιλέγουν να χρησιμοποιούν ελεύθερο λογισμικό ανοικτού κώδικα (open source), το οποίο συχνά συντηρούν, αναβαθμίζουν και επεκτείνουν. Έτσι συμβάλλουν στη μείωση του λειτουργικού κόστους, δεν εξαρτώνται από εταιρείες που αναπτύσσουν κλειστό, ιδιόκτητο λογισμικό, αποκτούν τεχνογνωσία και δημιουργούν το επιστημονικό υπόβαθρο για διάφορες αναπτυξιακές και ερευνητικές δραστηριότητες. Επιπλέον, έχει αποδειχθεί ότι:

• Το ανοικτό λογισμικό εξελίσσεται πιο γρήγορα και ισορροπημένα, διότι πολλά άτομα και ομάδες δουλεύουν παράλληλα.

• Οι ανάγκες των χρηστών καλύπτονται γρήγορα, καθώς το μοντέλο ανάπτυξης ανοικτού λογισμικού επιτρέπει την άμεση ενσωμάτωση της συλλογικής τεχνογνωσίας και συνεισφοράς.

• Νέες εκδόσεις του λογισμικού διατίθενται στους χρήστες του σε τακτά χρονικά διαστήματα.

•Η ομάδα ανάπτυξης του λογισμικού αποτελείται συνήθως από πολλούς γεωγραφικά διασκορπισμένους «εθελοντές» με εξειδίκευση σε διαφορετικούς τομείς και με διαφορετικές εμπειρίες.

• Δεν τίθενται σοβαρά θέματα ασφαλείας, επειδή ο κώδικας είναι ελεύθερα διαθέσιμος.

Το ανοικτό λογισμικό παρέχει ιδιαίτερα στα ιδρύματα της τριτοβάθμιας εκπαίδευσης έναν τρόπο να αξιοποιήσουν τις ιδέες της ακαδημαϊκής συνεργασίας και της κοινωνικής προσφοράς για να συνενώσουν τους πόρους και να εργαστούν για το κοινό καλό, αποτελώντας μια εναλλακτική λύση σε σχέση με το ιδιόκτητο λογισμικό.

2.5 Πλατφόρμες e-Learning

Κάποια από αυτά τα συστήματα ενδεικτικά είναι τα εξής:

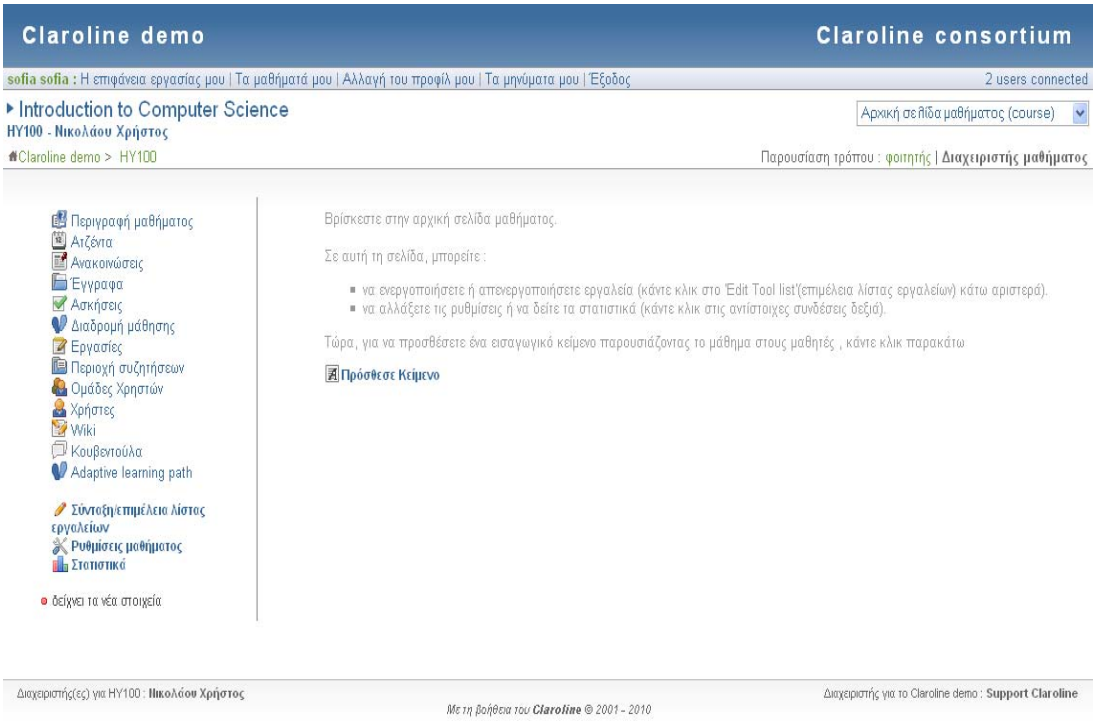
- ATutor (http://www.atutor.ca): πρόκειται για ένα ανοικτού κώδικα βασισμένο στον Παγκόσμιο Ιστό, σύστημα διαχείρισης μαθησιακού περιεχομένου (LCMS) σχεδιασμένο με βάση τις αρχές της προσβασιμότητας και της προσαρμοστικότητας. Οι εκπαιδευτικοί μπορούν να δημιουργήσουν, πακετάρουν και αναδιανείμουν εκπαιδευτικό υλικό που βασίζεται στον Παγκόσμιο Ιστό, να εισάγουν πακεταρισμένο υλικό και να διευθύνουν online τα μαθήματά τους. Οι σπουδαστές μαθαίνουν σε ένα προσαρμόσιμο περιβάλλον μάθησης. Αναπτύχθηκε το 2002

Adaptive Technology Resource Centre University of Toronto με τη γλώσσα προγραμματισμού PHP και χρησιμοποιεί ως σύστημα διαχείρισης βάσεων δεδομένων τη MySQL. Μετρά περισσότερους από 35 χιλιάδες εγγεγραμμένους χρήστες.

Εικόνα 1 - Προβολή ATutor Συστήματος

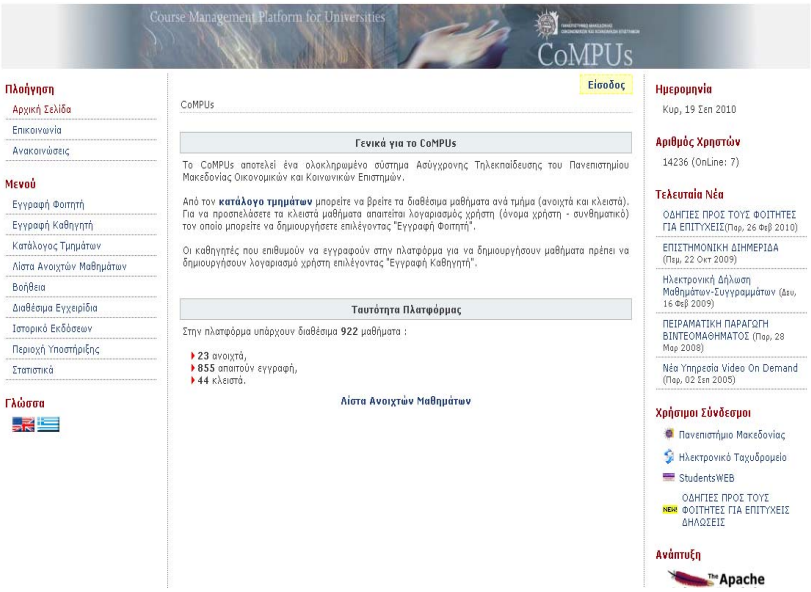
- Blackboard (http://www.blackboard.com): η εταιρεία Blackboard προσφέρει μία σειρά από εμπορικά εκπαιδευτικά πακέτα λογισμικού που βασίζονται σε τεχνολογίες Διαδικτύου για τη διασύνδεση σπουδαστών, καθηγητών, ερευνητών και της κοινωνίας σε ένα αναπτυσσόμενο δίκτυο από περιβάλλοντα εκπαίδευσης για καλύτερη επικοινωνία, συνεργασία και καλύτερο περιεχόμενο. Παρέχει δύο σουίτες λογισμικού: Blackboard Academic Suite και Blackboard Commerce Suite. Στην Blackboard Academic Suite περιέχεται το Blackboard Learning System, ένα σύστημα διαχείρισης μάθησης (CMS) για την online δημιουργία υλικού μαθημάτων και την διαδραστική επικοινωνία με τους σπουδαστές, το οποίο βασίστηκε στο WebCT. Το WebCT αρχικά αναπτύχθηκε στο University of British Columbia με σκοπό τη δημιουργία βασισμένων στον Παγκόσμιο Ιστό συστημάτων μάθησης και σήμερα κατέχεται από την Blackboard.
- Claroline (http://www.claroline.net): πρόκειται για ένα ανοικτού κώδικα σύστημα που ξεκίνησε από το Καθολικό Πανεπιστήμιο της Louvain(UCL) στο Βέλγιο το 2000 και σύμφωνα με τους κατασκευαστές της δημιουργήθηκε για να υποστηρίξει παραδοσιακή διδασκαλία με διαλέξεις, αυτόνομη μάθηση, μεικτή διδασκαλία ή εξ αποστάσεως εκπαίδευση. Βασίζεται στη

γλώσσα προγραμματισμού PHP και χρησιμοποιεί ως σύστημα διαχείρισης βάσεων δεδομένων τη MySQL. Μεταφρασμένη σε 35 γλώσσες και με χρήση σε 95 χώρες, έχει δημιουργήσει μια μεγάλη κοινότητα χρηστών και προγραμματιστών που ασχολούνται με την εξέλιξη ή προσαρμογή της πλατφόρμας.



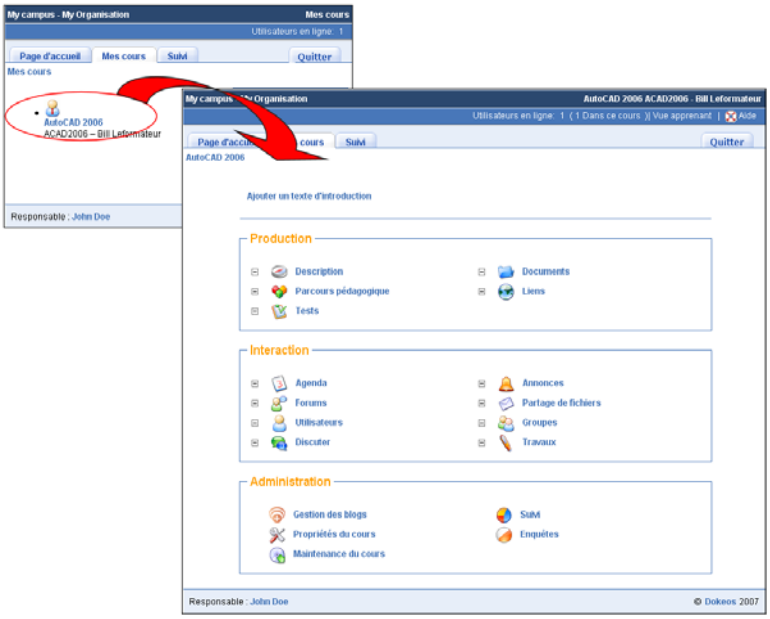
Εικόνα 2 - Παρουσίαση του Συστήματος Claroline

- CoMPUs (http://compus.uom.gr/index.php): πρόκειται για ένα σύστημα Ασύγχρονης Τηλεκπαίδευσης του Πανεπιστημίου Μακεδονίας ανοιχτού κώδικα που αναπτύχθηκε το 2004.



Εικόνα 3 - Προβολή Συστήματος CoMPUs

- Dokeos (http://www.dokeos.com): πρόκειται για μία ανοικτού κώδικα που υποστηρίζει τη δημιουργία ηλεκτρονικών μαθημάτων με βάση πρότυπα. Αναπτύχθηκε από την εταιρία Dokeos και αρχικά βασιζόταν στο Claroline. Υποστηρίζει περισσότερες από 30 γλώσσες και χρησιμοποιείται σε περίπου 60 χώρες.

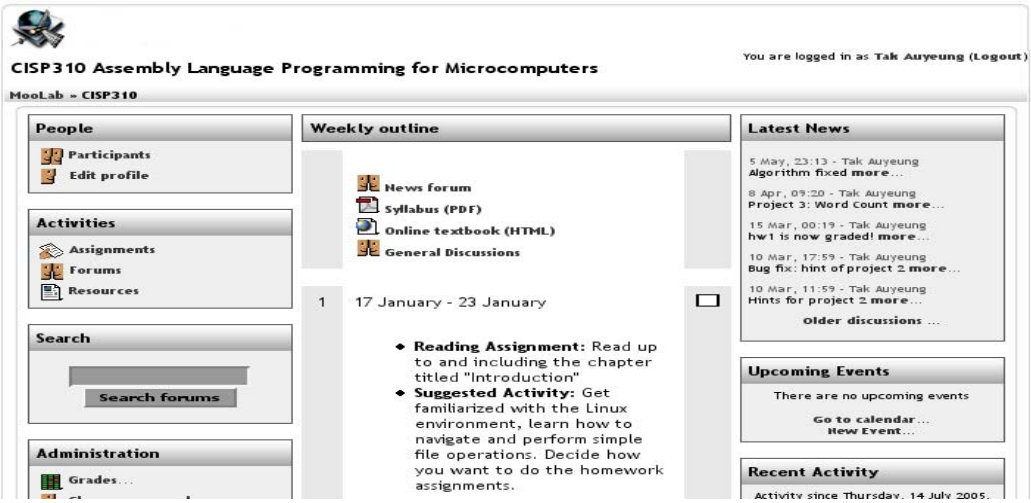


Εικόνα 4 - Παρουσίαση Συστήματος Dokeos

- e-Class: πλατφόρμα ασύγχρονης τηλε-εκπαίδευσης που βασίζεται στο ανοικτό σύστημα Claroline το οποίο εξελληνίστηκε και εμπλουτίστηκε από το Ελληνικό Ακαδημαϊκό Διαδίκτυο

GUNet (http://www.gunet.gr). Το Claroline είναι μία ανοικτού κώδικα (open source, GNU Public License (GPL)) πλατφόρμα ηλεκτρονικής μάθησης που προσφέρει στους εκπαιδευτικούς τη δυνατότητα δημιουργίας online μαθημάτων και διαχείρισης μαθησιακών και συνεργατικών δραστηριοτήτων στον Παγκόσμιο Ιστό. Διαθέτει μεγάλη κοινότητα χρηστών.

- eFront (http://www.efront.gr) : πρόκειται για ένα εύχρηστο ανοικτού κώδικα σύστημα διαχείρισης μάθησης (LMS) για εταιρική και εκπαιδευτική χρήση.
- ILIAS (http://www.ilias.de): πρόκειται για ένα ανοικτού κώδικα σύστημα διαχείρισης μάθησης (LMS) που προσφέρει εύκολη διαχείριση μαθησιακών πόρων σε ένα ολοκληρωμένο σύστημα.
- LRN (http://dotlrn.org): πρόκειται για ένα ανοικτού σύστημα για γρήγορη ανάπτυξη μαθησιακών κοινοτήτων που βασίζονται στον Παγκόσμιο Ιστό. Παρέχει μία μεγάλη σειρά από εργαλεία για τη διαχείριση «τάξεων» ή «κοινοτήτων».
- Moodle (http://moodle.org) : πρόκειται για ένα σύστημα διαχείρισης μαθημάτων, ελεύθερο, ανοικτού κώδικα πακέτο λογισμικού σχεδιασμένο με βάση παιδαγωγικές αρχές για την υποβοήθηση των εκπαιδευτικών στη δημιουργία αποτελεσματικών online κοινοτήτων μάθησης. Αναπτύχθηκε το 1999 σε γλώσσα προγραμματισμού PHP και χρησιμοποιεί ως σύστημα διαχείρισης βάσεων δεδομένων τη MySQL. Είναι μεταφρασμένο σε 78 γλώσσες και χρησιμοποιείται σε 210 χώρες.



Εικόνα 5 - Προβολή Κεντρικής Σελίδας του συστήματος Moodle

- OLAT (http://www.olat.org): πρόκειται για ένα ανοικτού κώδικα σύστημα διαχείρισης μάθησης (LMS) και παρέχει τα χαρακτηριστικά που υποστηρίζονται τυπικά από τις πλατφόρμες ηλεκτρονικής μάθησης.

2.6 Εργαλεία Σύγχρονης Επικοινωνίας και Ανάγκη που Δημιουργείται

Καθοριστική σημασία στην εκπαιδευτική διεργασία και την επιτυχία των σπουδών έχει η ποιότητα της επικοινωνίας μεταξύ εκπαιδευτή και εκπαιδευόμενων. Οι εκπαιδευόμενοι έχουν απόλυτη ανάγκη την επικοινωνία και υποστήριξη του διδάσκοντα διότι όπως φαίνεται από έρευνες που έχουν πραγματοποιηθεί [26][43][43], λίγοι μπορούν να μελετούν μόνοι τους χωρίς την ανάγκη ανθρώπινης υποστήριξης. Οι πιο πολλοί εκπαιδευόμενοι χρειάζονται υποστήριξη προκειμένου να αξιοποιήσουν το διδακτικό υλικό που τους παρέχεται.

Στην τριτοβάθμια εκπαίδευση χρησιμοποιούνται σε μεγάλο βαθμό συστήματα διαχείρισης μάθησης για την διεξαγωγή των μαθημάτων σε μεικτή μορφή. Σε όλα τα συστήματα διαχείρισης μάθησης παρέχονται εργαλεία για την εφικτή επικοινωνία ανάμεσα εκπαιδευτή και εκπαιδευόμενου και μεταξύ των εκπαιδευομένων. Κάποια από τα εργαλεία τα οποία χρησιμοποιούνται στα συστήματα διαχείρισης μάθησης είναι το ηλεκτρονικό ταχυδρομείο(email), forum, chat, πίνακας ανακοινώσεων, τηλεδιάσκεψη κ.ά. Η επικοινωνία μεταξύ των χρηστών σε ένα τέτοιο σύστημα είναι από τα πιο σημαντικά κομμάτια, γιατί με αυτό τον τρόπο παρέχεται η υποστήριξη από τον εκπαιδευτή. Αν δεν υπάρχει υποστήριξη ουσιαστικά μιλάμε για αυτοεκπαίδευση.

Συνήθως σένα σύστημα διαχείρισης μάθησης οι απορίες από τους εκπαιδευμένους και η επίλυσή τους από τον εκπαιδευτή γίνεται μέσω είτε του ηλεκτρονικού ταχυδρομείου είτε του φόρουμ. Η διαδικασία εκτελείται ως εξής: Γράφει ένα μήνυμα με το οποίο θέτει την απορία του ο εκπαιδευόμενος και είτε την στέλνει στον εκπαιδευτικό κατευθείαν χρησιμοποιώντας την χρήση του ηλεκτρονικού ταχυδρομείου, είτε τη δημοσιεύει σε μία θεματολογία στο φόρουμ όπου είναι ορατή από όλα τα μέλη της εκπαιδευτικής κοινότητας. Η αναμονή της απάντησης από τον εκπαιδευτικό μπορεί να είναι λεπτά, ώρες και σε κάποιες περιπτώσεις και μέρες.

Ένα σημαντικό πλεονέκτημα της χρήσης του φόρουμ ως μέσω επικοινωνίας είναι ότι αφενός ικανοποιείται η απορία του εκπαιδευόμενου που την έθεσε, αφετέρου επειδή είναι ορατή η ερώτηση - απάντηση σε όλους του εκπαιδευόμενους, μπορεί να ικανοποιήσει μια ανάλογη απορία οποιαδήποτε μετέπειτα στιγμή κάποιου εκπαιδευόμενου που θα τη διαβάσει. Αυτό το σημαντικό πλεονέκτημα όμως σε ένα μεγάλο βαθμό δεν είναι εφαρμόσιμο και ό λόγος είναι ότι η πληθώρα των ερωτήσεων-απαντήσεων που έχουν αναρτηθεί στο φόρουμ είναι συνεχώς αυξανόμενη και ανάλογη με το πλήθος των εκπαιδευόμενων, έτσι όταν κάποιος εκπαιδευόμενος έχει κάποια απορία προτιμάει να δημοσιεύσει κατευθείαν την ερώτησή του παρά να αναζητήσει όλο το φόρουμ για κάποια παρόμοια απορία, που μπορεί και να μην υπάρχει. Για τον εκπαιδευτή η απασχόλησή του με την απάντηση των αποριών είναι μεγάλος φόρτος και συνήθως απαιτούνται αρκετές ώρες ημερησίως κυρίως αν το πλήθος των εκπαιδευόμενων είναι μεγάλο.

Με αφορμή το πρόβλημα το οποίο δημιουργείται στην επικοινωνία εκπαιδευτή και εκπαιδευόμενου προτείνουμε ένα σύστημα που προσφέρει ένα αυτόματο τρόπο απαντήσεων χωρίς την εμπλοκή του εκπαιδευτή. Το σύστημα αυτό έχει σκοπό την ελαχιστοποίηση της ενασχόλησης του εκπαιδευτή με την διαδικασία απάντησης των αποριών που έχουν δημοσιευτεί, και την άμεση υποστήριξη των εκπαιδευόμενων. Για να αναπτυχθεί ένα τέτοιο σύστημα θα πρέπει να χρησιμοποιηθούν τεχνικές που θα βασίζονται στην αναπαράσταση γνώσης και στην λήψη αποφάσεων

3 Συστήματα Γνώσης

Τα ευφυή πληροφοριακά συστήματα επιδεικνύουν νοήμονα συμπεριφορά σε συγκεκριμένους τομείς και διαδικασίες, ανάλογη ενός ανθρώπου εμπειρογνώμονα με ειδικότητα σε κάποιο τομέα. Έχουν τη δυνατότητα να επεξεργάζονται αποθηκευμένη γνώση που αφορά την εμπειρία ενός ειδικού σε κάποιο τομέα και την εξασφάλιση της δυνατότητας του υπολογιστή να δρα πλέον ως ειδικός[45]. Κωδικοποιούν και χειρίζονται τη γνώση και τη συλλογιστική ενός ανθρώπου ειδικού σε έναν εξειδικευμένο τομέα, με σκοπό την επίλυση προβλημάτων ή την παροχή συμβουλών. [9], [4].Με τον τρόπο αυτό ένα ευφυή σύστημα μπορεί να δίδει προσαρμοσμένη συμβουλή σε ένα χρήστη. Τα ευφυή συστήματα είναι από τα πλέον χαρακτηριστικά παραδείγματα εφαρμογών της Τεχνητής Νοημοσύνης που τις περισσότερες φορές έχουν τη μορφή συστημάτων κανόνων, υλοποιούν δηλαδή ανίχνευση δια της συμπεριφοράς .

Τα ευφυή πληροφοριακά συστήματα διαχειρίζονται είτε σε μικρό είτε σε μεγάλο βαθμό κάποιο είδος γνώσης. Αυτό έχει σαν αποτέλεσμα οι τεχνικές αναπαράστασης γνώσης και ο τρόπος σκέψης (λογισμός) βασισμένος στη γνώση να ανήκουν σε ένα από τα πιο σημαντικά κομμάτια για την Τεχνητή Νοημοσύνη. Ο τρόπος που αναπαρίσταται η γνώση, προσδιορίζεται χρησιμοποιώντας μια *μέθοδο αναπαράστασης γνώσης*. Μια μέθοδος αποτελείται από ένα σύνολο *κατασκευασμάτων αναπαράστασης γνώσης* και ένα σύνολο λειτουργιών που ανακτούν ή αλλάζουν στιγμιότυπα των κατασκευασμάτων. Τα στιγμιότυπα κατασκευάσματος καλούνται *δομές αναπαράστασης*. Όλες οι δομές αναπαράστασης μαζί χτίζουν την *αναπαράσταση γνώσης (Knowledge Representation - KR)* του συστήματος.

Η διαθέσιμη γνώση η οποία διατίθεται σε οποιαδήποτε μορφή, ονομάζεται γνωστικό αντικείμενο και καθίσταται απαραίτητη η αναπαράστασή του μέσω κάποιων μεθόδων. Οι πιο βασικές μέθοδοι αναπαράστασης είναι: *βασισμένες στη λογική*(*Logic based*), σύστημα βασισμένο σε κανόνες αποτελείται από ένα σύνολο κανόνων αποκαλούμενων *βάση κανόνων (rule base)*, σημασιολογικά δίκτυα(*Semantic Nets*), συλλογιστική βασισμένη σε περιπτώσεις(Case Based Reasoning).

Το μεγαλύτερο και δυσκολότερο θέμα που προκύπτει είναι ότι το σύστημα πρέπει να επιλέξει το κατάλληλο για την περίπτωση πλάνο προκειμένου να πετύχει τα καλύτερα αποτελέσματα σε συγκεκριμένο χρόνο με δεδομένες τις τρέχουσες συνθήκες του περιβάλλοντος. Ένα ευφυές σύστημα μπορεί να έχει αυξημένες επιδόσεις και μεγαλύτερη ευελιξία, αν έχει δυνατότητες μάθησης και προσαρμογής σε νέες καταστάσεις. Η *μηχανική μάθηση* (*machine learning*) επικεντρώνεται κυρίως στην ανάπτυξη λογισμικού που βελτιώνεται αυτόματα μέσω εμπειρίας και στην εξαγωγή κανόνων από ένα σύνολο συγκεκριμένων δεδομένων.

Όταν αναφερόμαστε στη μάθηση για τους ανθρώπους συχνά εννοούμε την κατανόηση. Κάτι τέτοιο βέβαια δεν είναι πολύ ακριβές καθώς μπορεί να μας οδηγήσει στο συμπέρασμα ότι η κατανόηση είναι απαραίτητα για τη μάθηση. Εφόσον η αναφορά αφορά ανθρώπους η μάθηση συχνά μπορεί να προέρχεται από την εμπειρία και όχι απαραίτητα από την κατανόηση. Αντίστοιχα και στη μάθηση των μηχανών υπάρχουν δύο μεθοδολογίες, σε αυτές που στηρίζονται στην κατανόηση και σε αυτές που βασίζονται στην εμπειρία. Η πρώτη μεθοδολογία ονομάζεται Πρόθυμη Μάθηση(Eager Learning) και η δεύτερη Οκνηρή Μάθηση(Lazy Learning). Η Πρόθυμη Μάθηση αφορά μεθόδους που προσπαθούν εξαντλητικά

από τα δεδομένα να δημιουργήσουν δομές, δηλαδή δέντρα αποφάσεων, κανόνες παραγωγής κ.ά. ενώ η Οκνηρή Μάθηση συλλέγει εμπειρίες και τις αποθηκεύει ως δεδομένα για μελλοντική χρήση. Η επιλογή χρήσης την κατάλληλης μεθόδου με σκοπό την επίτευξη μάθησης του συστήματος εξαρτάται από το πρόβλημα. Στην υλοποίηση του συστήματος χρειαζόμαστε ένα σύστημα που να βασίζεται στην εμπειρία, δηλαδή να μπορέσει να εντοπίσει την απάντηση ενός ερωτήματος εφόσον υπάρχει γνώση για αυτό. Έτσι η επιλογή της μεθόδου θα πρέπει να περιλαμβάνεται στην κατηγορία της Οκνηρής Μάθησης. Επιπλέον πρέπει το σύστημα που θα επιλεγεί να περιλαμβάνει λήψη αποφάσεων με μία λογική για τον εντοπισμό της πιο ιδανικής περίπτωσης εφόσον αυτή υπάρχει.

Δύο από τις τεχνολογίες που έχουν αναπτυχθεί για την υποστήριξη της διαδικασίας συνεργατικής λήψης αποφάσεων είναι τα *βασισμένα σε κανόνες συστήματα (Rule-Based Systems)* και τα συστήματα βασισμένα στη συλλογιστική των περιπτώσεων *(Case-Based Reasoning).*Και τα δύο υπάγονται στην επιστημονική περιοχή της τεχνητής νοημοσύνης. Τα συστήματα βασισμένα σε κανόνες αναπαριστούν τη γνώση μέσα από ένα σύνολο κανόνων της μορφής "εάν ... τότε …" (“if-then” rules). Με τους κανόνες ορίζουν τι πρέπει να γίνει ή να αποφασιστεί από μία συγκεκριμένη θεώρηση του προβλήματος. Ένα σύστημα βασισμένο σε κανόνες εμπεριέχει ένα σύνολο από γνώσεις και ένα σύστημα συμπερασμάτων. Οι γνώσεις αφορούν ένα σύνολο από κανόνες παραγωγής και τα συστήματα συμπερασμάτων συνδυάζουν τους κανόνες με τις παρατηρήσεις για να καταλήξουν σε λύσεις και συμπεράσματα σχετικά με την κατάσταση του εκάστοτε προβλήματος καθώς και τις ενέργειες που πρέπει να πραγματοποιηθούν. Από την άλλη πλευρά, ακολουθώντας τη συλλογιστική βασισμένη σε περιπτώσεις, η γνώση και η εμπειρία των χρηστών κωδικοποιείται σε μια βιβλιοθήκη παρελθοντικών περιπτώσεων, που χρησιμοποιείται για τη λήψη αποφάσεων και δεν συσχετίζεται με κανόνες.

3.1 Λογικός Προγραμματισμός

Ο Λογικός Προγραμματισμός είναι μια υπολογιστική κωδικοποίηση που συνδυάζει δύο κύριες αρχές:

- χρησιμοποιεί *λογική* για να αναπαραστήσει τη γνώση,
- χρησιμοποιεί το μηχανισμό *των αποδείξεων* και των *συμπερασμάτων* για να επεξεργαστεί γνώση.

Στο πλαίσιο επίλυσης ενός προβλήματος, η πρώτη αρχή αφορά την αναπαράσταση υποθέσεων και συμπερασμάτων, ενώ η δεύτερη, αφορά τον ορισμό και την παραγωγή των λογικών σχέσεων μεταξύ υποθέσεων και συμπερασμάτων. Ο γενικός σκοπός σε κάθε παρόμοιο πλαίσιο είναι να *επιστρέφεται* το επιθυμητό αποτέλεσμα από τις δεδομένες υποθέσεις, με έναν τρόπο που είναι υπολογιστικά αποδεκτός.

Ένα πρόγραμμα λογικής αποτελείται από ένα σύνολο από κανόνες που ορίζουν σχέσεις μεταξύ των αντικειμένων. Οι κύριες δομές του λογικού προγραμματισμού είναι οι όροι και οι δηλώσεις. Οι δηλώσεις διακρίνονται σε γεγονότα, ερωτήματα και κανόνες. Η πιο ευρέως χρησιμοποιούμενη γλώσσα λογικού προγραμματισμού είναι η Prolog(programming in logic), η οποία είναι συμβολική γλώσσα που βασίζεται στην κατηγορηματική λογική πρώτης τάξης. Ένα πρόγραμμα στο λογικό προγραμματισμό περιλαμβάνει:

- Ένα σύνολο γνωστών γεγονότων (facts)

- Ένα σύνολο γνωστών κανόνων (rules)
- Μία πρόταση που πρέπει να αποδειχθεί ότι είναι αληθής με τη χρήση των παραπάνω.
- Δεν περιέχει τον ακριβή αλγόριθμο επίλυσης του προβλήματος

Στη γλώσσα Prolog τα διάφορα αποτελέσματα εξάγονται με τη διαδικασία της *εις άτοπον απαγωγής* σύμφωνα με την οποία η επαλήθευση ξεκινάει από την άρνηση της ερώτησης και συνεχίζει μέχρι να καταλήξει σε αντίφαση οπότε και αποδεικνύεται το αρχικό ζητούμενο. Στη συνέχεια, κατά τη διαδικασία ανάλυσης, η Prolog χρησιμοποιεί ένα μηχανισμό ενοποίησης ο οποίος προσπαθεί να ταιριάξει το ζητούμενο με ένα γεγονός. Επιπλέον χρησιμοποιούνται κάποιοι περιορισμοί οι οποίοι είναι εξισώσεις πάνω στην άλγεβρα των όρων, ενώ στη συνέχεια οι εξισώσεις χρησιμοποιούνται στην διαδικασία της ενοποίησης.

Η Prolog ήταν η πρώτη γλώσσα λογικής που περιγράφηκε εξ' ολοκλήρου με την χρήση περιορισμών και κάνει τους υπολογισμούς της στην γνωστή άλγεβρα Boole ενώ εκτελεί γραμμική αριθμητική πάνω στους λογικούς αριθμούς. Κατά καιρούς, πολλά προβλήματα όπως αναθέσεις στους τεχνικούς στόχων ή προγραμματισμός του ωραρίου εργασίας επιλύθηκαν με τη χρήση της γλώσσας Prolog. Επιπρόσθετα, χρησιμοποιήθηκε για την επίλυση του προβλήματος του προγραμματισμού παραγωγής που αποτελούσε ανέκαθεν ένα αρκετά μελετημένο πρόβλημα. Τα προβλήματα παραγωγής πολλές φορές δεν είναι παρά συνδυασμός από καλά αναγνωρισμένα επιμέρους προβλήματα τα οποία έχουν ήδη επιλυθεί.

3.2 Rule-Based Systems

Ένα σύστημα βασισμένο σε κανόνες αποτελείται από ένα σύνολο κανόνων αποκαλούμενων *βάση κανόνων (rule base)*, μια *βάση γεγονότων (fact base)*, και μια διαδικασία ερμηνείας που ελέγχει τους κανόνες για τη δυνατότητα εφαρμογής τους και εκτελεί τις εφαρμόσιμες ενέργειες. Ένας κανόνας είναι ένα ζευγάρι που αποτελείται από μια προϋπόθεση και μια ενέργεια. Η ενέργεια μπορεί να εκτελεσθεί εάν υπάρχει η προϋπόθεση. Η προϋπόθεση αναφέρεται στη βάση γεγονότος (που μπορεί να αναπαρασταθεί χρησιμοποιώντας οποιοδήποτε μέθοδο αναπαράστασης). Η ενέργεια τυπικά αλλάζει τη βάση γεγονότος, αλλά μπορεί επίσης να ρωτήσει το χρήστη για κάποια εισαγωγή ή εξαγωγή μερικών υπολογιστικών αποτελεσμάτων.

Πλεονεκτήματα των συστημάτων που βασίζονται σε κανόνες :

- Σαν κατασκεύασμα, οι κανόνες είναι επαρκείς για την αναπαράσταση των βασισμένων σε κανόνες εξαρτήσεων.
- Οι κανόνες υποστηρίζουν ανθρώπινη σκέψη (ειδικά για γενικευμένη γνώση), από άποψη κανόνων και heuristics .
- Οι κανόνες είναι τμηματικοί, δηλαδή απλοί κανόνες μπορούν να προστεθούν ή να αφαιρεθούν οποιαδήποτε στιγμή ανεξάρτητα από την ύπαρξη άλλων κανόνων. Αυτό είναι δυνατό, επειδή οι κανόνες δεν αναφέρονται ο ένας στον άλλο άμεσα, αλλά μόνο έμμεσα μέσω των αλλαγών της βάσης γεγονότος.

Μειονεκτήματα των συστημάτων που βασίζονται σε κανόνες :

- Οι πληροφορίες πρέπει να είναι σταθερές και καλά δομημένες.
- Είναι δύσκολη η επισκόπηση (διανοητικώς) της ροής ελέγχου ακόμη και για έναν μικρό αριθμό κανόνων. Αυτό σημαίνει ότι τα συστήματα αυτά είναι δύσκολο να κατανοηθούν, να διατηρηθούν και να επαληθευθούν.
- Δεδομένου ότι η ροή ελέγχου δεν προσδιορίζεται ρητά, μπορεί να εμφανιστούν προβλήματα απόδοσης.

3.3 Συστήματα βασισμένα στη Συλλογιστική των Περιπτώσεων (Case-Based Reaoning)

Η Βασισμένη σε Περιπτώσεις Συλλογιστική (Case-Based Reasoning) ή αλλιώς Συλλογιστική των Περιπτώσεων είναι μια μέθοδος για επίλυση προβλημάτων με την ανάμνηση των προηγούμενων παρόμοιων καταστάσεων και την επαναχρησιμοποίηση των πληροφοριών και της γνώσης για εκείνη την κατάσταση. Η βασισμένη σε περιπτώσεις συλλογιστική είναι στην ουσία «αναλογική συλλογιστική» όπου το σύστημα προσπαθεί να προσδιορίσει το αποτέλεσμα μιας δεδομένης κατάστασης σύμφωνα με το αποτέλεσμα της περισσότερο παρεμφερούς περίπτωσης (case), ή των περισσότερο παρεμφερών περιπτώσεων, από μια βάση γνώσεων που αποτελείται από αποθηκευμένες γνωστές περιπτώσεις (case-base). Η συλλογιστική των περιπτώσεων είναι μια κυκλική και ολοκληρωμένη διαδικασία για τη λύση ενός προβλήματος, η οποία παράλληλα μαθαίνει από αυτήν την εμπειρία. Τυπικά, κάθε περίπτωση αποτελείται από την περιγραφή ενός συγκεκριμένου στιγμιότυπου του προβλήματος, καθώς και της επίλυσής του. Για την επίλυση ενός νέου στιγμιότυπου, γίνεται αρχικά ένα ταίριασμα του έναντι παρελθοντικών περιπτώσεων (ακολουθώντας διάφορα μέτρα ομοιότητας), έχοντας ως στόχο την ανεύρεση παρόμοιων περιπτώσεων και την εκμετάλλευση των λύσεων που δόθηκαν σε αυτές. Το κατά πόσον η τεχνολογία αυτή είναι χρήσιμη σε ένα περιβάλλον συνεργατικής λήψης αποφάσεων εξαρτάται από τη διαθεσιμότητα παρελθοντικών περιπτώσεων, καθώς και το αν τέτοιες περιπτώσεις θεωρούνται πολύτιμες.

3.4 Πλεονεκτήματα και μειονεκτήματα

Συγκρίνοντας τις μεθόδους της συλλογιστικής βασισμένη σε περιπτώσεις, με τις συλλογιστικές των συστημάτων κανόνων όπως το σύστημα βασισμένο σε κανόνες και τη χρήση του λογικού προγραμματισμού καταλήγουμε στα ακόλουθα συμπεράσματα[46]:

Η συλλογιστική βασισμένη σε περιπτώσεις πλησιάζει τον τρόπο με τον οποίο σκέφτεται ο άνθρωπος. Τα συστήματα κανόνων χρησιμοποιούν στην εμπειρία που προκύπτει από κάποιον ειδικό, αλλά η αναπαράσταση της γνώσης του ειδικού στον υπολογιστή είναι πολύ διαφορετικά. Πιο συγκεκριμένα στα έμπειρα συστήματα η γνώση του ειδικού έχει υποστεί μετατροπές, είτε από τον ειδικό, που αναγκάζεται

να ομαδοποιήσει τις εμπειρίες του ή και από το μηχανικό της γνώσης που πρέπει αφαιρετικά να βρει κοινά στοιχεία στις εμπειρίες του ειδικού και να τις μετατρέψει σε κανόνες.

Στη συλλογιστική βασισμένη σε περιπτώσεις η εμπειρία του ειδικού μπορεί να αναπαρασταθεί άμεσα με τη μορφή περιπτώσεων που αντιμετώπισε στο παρελθόν και επαναφέρει από τη μνήμη του. Τα προβλήματα που αντιμετωπίστηκαν στο παρελθόν συγκρίνονται με τις τρέχουσες καταστάσεις και οι λύσεις που χρησιμοποιήθηκαν ξαναχρησιμοποιούνται. Η συλλογιστική βασισμένη σε περιπτώσεις οδηγεί στη συλλογιστική με αναλογίες, η οποία χρησιμοποιείται από τους περισσότερους ανθρώπους τόσο για επίλυση προβλημάτων όσο και για εκμάθηση νέας γνώσης. Η απόκτηση γνώσης είναι μια πιο απλή διαδικασία καθώς τις περισσότερες φορές είναι διαθέσιμη σε παλαιότερα έγγραφα καταχωρημένα σε βάσεις δεδομένων, και μπορεί ακόμα εύκολα να δημιουργηθεί ή να ενισχυθεί από τις εμπειρίες κάποιου ειδικού. Η περίπτωση αυτή βέβαια απαιτεί πολύ μεγαλύτερο κόπο εκ μέρους του ειδικού από μια απλή συμμετοχή του στη διαδικασία εκμαίευσης γνώσης των εμπείρων συστημάτων. Η διαχείριση της γνώσης περιλαμβάνει εργαλεία και μεθόδους για τη δημιουργία, ανάκτηση και επαναχρησιμοποίηση της γνώσης.

Το κυριότερο μειονέκτημα της συλλογιστικής βασισμένη σε περιπτώσεις είναι το υπολογιστικό κόστος της αναζήτησης στη βιβλιοθήκη των περιπτώσεων. Έτσι μια "έξυπνη" δεικτοδότηση αποτελεί ένα σημαντικό ζήτημα στη χρήση της, ώστε να επιταχύνεται αποδοτικότερα η αναζήτηση. Κάποια άλλα σημεία που επηρεάζουν την απόδοση και την ποιότητα των λύσεων που παρέχει ένα τέτοιο σύστημα είναι η δομή της βάσης καθώς και η ποιότητα και ποσότητα των περιπτώσεων που υπάρχουν αποθηκευμένες σε αυτή. Τα ζητήματα όμως αυτό δεν αποτελεί πρόβλημα μόνο για τα συστήματα που στηρίζονται στη συλλογιστική βασισμένη σε περιπτώσεις αλλά για οποιοδήποτε πληροφοριακό σύστημα.

Ο λογικός προγραμματισμός παρόλο που προσφέρει τη δυνατότητα σε κάποιον να εκφράσει τη γνώση με έναν τρόπο ανεξάρτητο της μηχανής, δημιουργώντας έτσι πιο συμπαγή ευέλικτα και κατανοητά προγράμματα, δεν ικανοποιεί την επίλυση του προβλήματος που καλούμαστε να επιλύσουμε στα πλαίσια της εργασίας. Όπως και τα συστήματα κανόνων έτσι και ο λογικός προγραμματισμός για την εφαρμογή τους στο σύστημα θα πρέπει να οριστεί ένα σύνολο κανόνων για την εφικτή εξαγωγή συμπερασμάτων. Το πρόβλημα που καλούμαστε να επιλύσουμε στα πλαίσια της μεταπτυχιακής εργασίας υπάγεται στον ορισμό της συλλογιστικής που βασίζεται σε περιπτώσεις, καθώς χρειάζεται να εντοπίζεται η απάντηση ενός ερωτήματος χρησιμοποιώντας την προηγούμενη γνώση. Η πηγή γνώσης είναι με μορφή κειμένου, και η συλλογιστική βασισμένη σε περιπτώσεις παρέχει τη δυνατότητα λήψης αποφάσεων και αναπαράστασης γνώσης σε μορφή κειμένου. Είναι η μόνη μέθοδος που μπορεί να καλύψει πλήρως τις απαιτήσεις του συστήματος με τον πιο αποδοτικό τρόπο. Συμπερασματικά από την έρευνα που πραγματοποιήθηκε σχετικά με τις τεχνικές που θα εφαρμοστούν στο σύστημα που σχεδιάστηκε επιλέχτηκε η συλλογιστική βασισμένη σε περιπτώσεις.

4 Θεωρητικό Υπόβαθρο

Με αφορμή το πρόβλημα το οποίο δημιουργείται στην επικοινωνία εκπαιδευτή και εκπαιδευόμενου προτείνουμε ένα σύστημα που προσφέρει έναν αυτόματο τρόπο απαντήσεων χωρίς την απαραίτητη εμπλοκή του εκπαιδευτή. Το σύστημα έχει στόχο την ελαχιστοποίηση της ενασχόλησης του εκπαιδευτή με τη διαδικασία απάντησης των απορριών που έχουν δημοσιευτεί, και την άμεση υποστήριξη των εκπαιδευόμενων.

Το σύστημα το οποίο προτείνουμε έχει την δυνατότητα να απαντάει σε κάποια καινούργια απορία που τίθεται, εφόσον έχει γνώση γι' αυτή. Τη γνώση το σύστημα την αποκτάει από τον εκπαιδευτή, με τις απαντήσεις που έχει δώσει σε προηγούμενες απορίες. Εφόσον δεν υπάρχει γνώση, το σύστημα δεν θα επιστρέφει κάποια απάντηση και ο εκπαιδευτής θα πρέπει να απαντήσει την τρέχουσα απορία. Ένα τέτοιο σύστημα θα επιφέρει σημαντική αντιμετώπιση στην επίλυση απορριών ως μία βοηθητική εκπαιδευτική υπηρεσία τόσο στον εκπαιδευτή όσο και στον εκπαιδευόμενο.

Οι απορίες που θα έχει την δυνατότητα να επιλύει το σύστημα μπορεί να ικανοποιούν τον εκπαιδευόμενο στο μέγιστο της απορίας του ή να λύνουν ένα μέρος αυτής. Οι απορίες που τίθενται από τους εκπαιδευόμενους σ' ένα μεγάλο βαθμό είναι παρόμοιες, παρατηρείται συχνά το φαινόμενο μια απορία να εμφανίζετε σ' ένα πλήθος των εκπαιδευόμενων και σπάνια μεμονωμένα. Μια απορία που έχουν πολλοί εκπαιδευόμενοι μπορεί να τεθεί από μία μερίδα αυτών χρησιμοποιώντας ένα σύνολο(μικρό ή μεγάλο) ίδιων λέξεων ή συνώνυμων για να την περιγράψουν. Καταγράφοντας όλες αυτές τις παραμέτρους το σύστημα που σχεδιάστηκε παρέχει τη λύση του συγκεκριμένου προβλήματος.

Το σύστημα που δημιουργήθηκε θα πρέπει να χρησιμοποιεί τεχνικές τόσο της Μηχανικής Μάθησης όσο και των Συστημάτων Γνώσης. Έτσι στο σύστημα εφαρμόστηκαν τεχνικές CBR(Case Based Reasoning), που στην ελληνική βιβλιογραφία αποκαλείται ως συλλογιστική βασισμένη σε περιπτώσεις. Η συλλογιστική βασισμένη στην περίπτωση (CBR) είναι μια μεθοδολογία που έχει εμπνευστεί από τον ανθρώπινο συλλογισμό. Οι άνθρωποι συχνά υιοθετούν μια προσέγγιση στην επίλυση κάποιου προβλήματος, με την οποία θυμούνται μια προηγούμενη περίπτωση ενός παρόμοιου προβλήματος και προσαρμόζουν εκείνη την προσέγγιση στο νέο πρόβλημα. Μια προσέγγιση CBR έχει ομοιότητα με των γνωστών προβλημάτων που αντιμετωπίζουμε. Όταν παρουσιάζεται ένα νέο πρόβλημα, τότε ανακτάται μια παρόμοια περίπτωση από τη βάση γνώσεων και χρησιμοποιείται ως η αφετηρία για την επίλυση του νέου προβλήματος. Τα συστήματα Συλλογιστικής των περιπτώσεων την τελευταία δεκαπενταετία έχουν σημειώσει ραγδαία ανάπτυξη και έχουν καταφέρει να καθιερωθούν ως ένας νέος κλάδος των συστημάτων γνώσης και της μηχανικής μάθησης.

4.1 Συλλογιστική Βασισμένη σε Περιπτώσεις CBR (Case Based Reasoning)

Ένας ορισμός που μπορεί να δοθεί για τη συλλογιστική βασισμένη σε περιπτώσεις είναι ο εξής : Η λύση ενός νέου προβλήματος προκύπτει χρησιμοποιώντας τη λύση μιας προηγούμενης παρόμοιας κατάστασης καθώς και των πληροφοριών και της γνώσης εκείνης.

Η συλλογιστική βασισμένη σε περιπτώσεις αποθηκεύει ένα σύνολο προηγούμενων περιπτώσεων σε μία βάση μαζί με τις λύσεις τους. Οι περιπτώσεις (cases) που έχουν αποθηκευτεί χρησιμοποιούνται οποτεδήποτε σε μια νέα παρόμοια περίπτωση που χρήζει αντιμετώπισης. Στόχος της μεθόδου αυτής είναι η ανάδειξη των πιο σχετικών ιστορικών περιπτώσεων που ταιριάζουν με την τρέχουσα περίπτωση.

Η αποτελεσματικότητα ενός συστήματος συλλογιστικής βασισμένης σε περιπτώσεις βασίζεται στην ικανότητά του να μαθαίνει από κάθε καινούργια περίπτωση. Πιο συγκεκριμένα κάθε καινούργια περίπτωση πρέπει να αναλύεται, καταγράφεται και κατά συνέπεια να αποθηκεύεται ως γνώση στο σύστημα για μελλοντική χρήση.

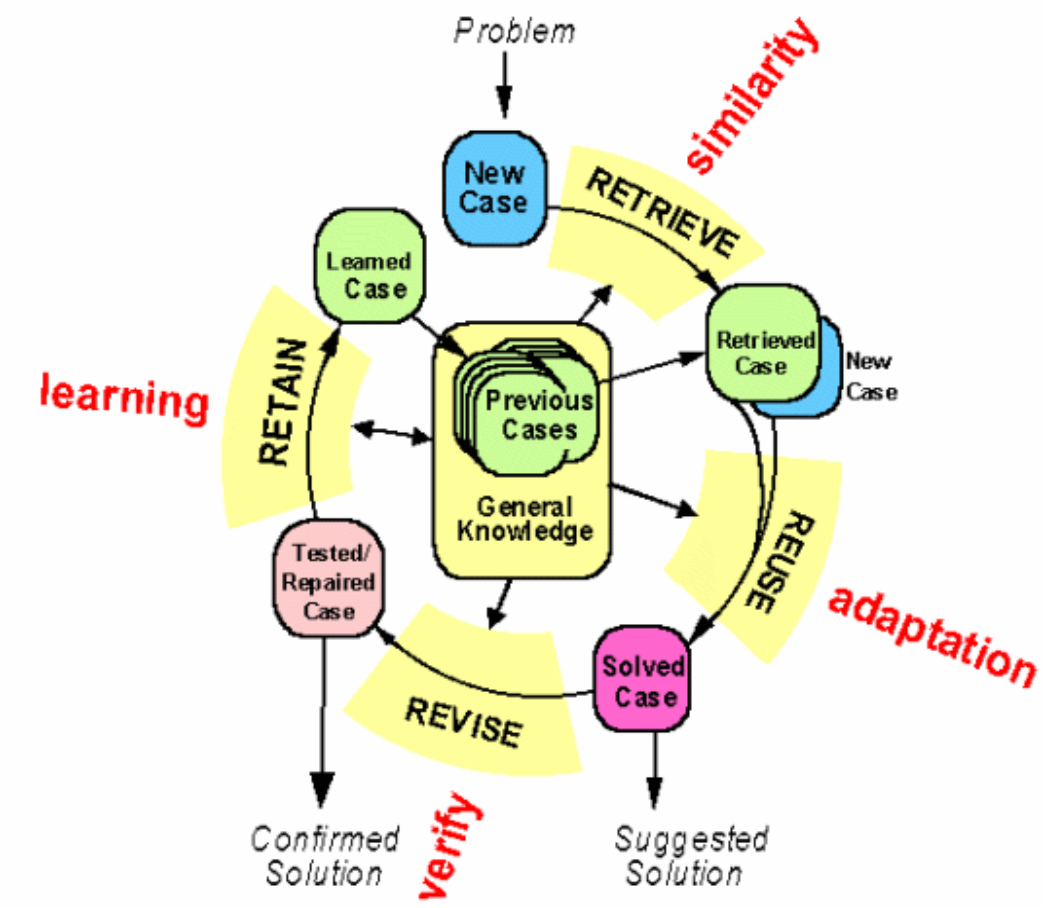
Στην ορολογία CBR μια **περίπτωση** δείχνει μια κατάσταση του προβλήματος. Μια προηγούμενη κατάσταση, που έχει καταγραφεί με τέτοιο τρόπο ώστε να μπορεί να επαναχρησιμοποιηθεί στην επίλυση μελλοντικών προβλημάτων, αναφέρεται ως προηγούμενη περίπτωση.

Η CBR μέθοδος είναι μια κυκλική και ολοκληρωμένη διαδικασία για τη λύση ενός προβλήματος, η οποία παράλληλα μαθαίνει από αυτήν την εμπειρία. Η επίλυση του νέου προβλήματος περιλαμβάνει τα εξής βήματα (Εικόνα 6):

1. Ανάκτηση (Retrieve). Η διαδικασία ανάκτησης πραγματοποιεί την εύρεση του καλύτερου ταιριάσματος παρόμοιας περίπτωσης ή παρόμοιων περιπτώσεων από τη βάση περιπτώσεων.
2. Επαναχρησιμοποίηση (Reuse). Η διαδικασία αυτή εντοπίζει τι μπορεί να επαναχρησιμοποιηθεί από παλαιές περιπτώσεις.
3. Αναθεώρηση (Revise). Στην συγκεκριμένη διαδικασία ελέγχεται εάν η προτεινόμενη λύση είναι σωστή.
4. Διατήρηση (Retain). Στην περίπτωση αυτή πραγματοποιείται η εκμάθηση από την εμπειρία επίλυσης του προβλήματος. Για να βελτιωθεί η μελλοντική συμπεριφορά του συστήματος είναι χρήσιμο να προστεθεί στη βάση περιπτώσεων το εκάστοτε πρόβλημα με την λύση του. Αυτό αποτελεί και το στάδιο της Μηχανικής Μάθησης.

Το πιο σημαντικό στάδιο στον κύκλο ζωής του CBR είναι η ανάκτηση, όπου το σύστημα καλείται να εντοπίσει τις πιο όμοιες περιπτώσεις από μία βάση περιπτώσεων έτσι ώστε να εξάγει το καλύτερο αποτέλεσμα για ένα συγκεκριμένο ερώτημα. Το CBR σύστημα βασίζεται στις περιπτώσεις, και οι περιπτώσεις λαμβάνονται κατά την διάρκεια της ανάκτησης. Η αποτελεσματικότητα του συστήματος εξαρτάται από το πόσο είναι ικανό να εντοπίσει τις κατάλληλες περιπτώσεις, αν δεν είναι ικανό να εντοπίσει την πιο ταιριαστή περίπτωση αυτή πιθανότατα να μην ανακληθεί παρόλο που θα ήταν χρήσιμη.

Σε πολλές εφαρμογές τα βήματα της επαναχρησιμοποίησης και της αναθεώρησης είναι δύσκολο να διακριθούν και κάποιες φορές χρησιμοποιούνται ως ένα ενιαίο στάδιο προσαρμογής που τα αντικαθιστά και τα συνδυάζει.



Εικόνα 6 - Ο κύκλος Ζωής του CBR

Οι περιπτώσεις σε ένα CBR μπορούν να αντιπροσωπεύουν πολλούς διαφορετικούς τύπους γνώσεων. Η γνώση αυτή μπορεί να αποθηκευτεί με πολλά αντιπροσωπευτικά σχήματα. Όταν η γνώση έχει την μορφή κειμένου (δομημένου ή αδόμητου) τότε συνίσταται η χρήση τεχνικών συλλογιστικής βασισμένες σε περιπτώσεις κειμένου(Textual Case Based Reasoning).

4.2 Συλλογιστική Βασισμένη σε Περιπτώσεις Κειμένου TCBR(Textual Case Based Reasoning)

Η συλλογιστική βασισμένη σε περιπτώσεις κειμένου(TCBR) είναι μια υποκατηγορία της συλλογιστικής βασισμένης σε περιπτώσεις, όπου όλο ή ένα μέρος της γνώσης που χρησιμοποιείται είναι με την μορφή κειμένου. Αυτού του είδους συστήματα είναι πολύ ενδιαφέροντα γιατί οι περισσότερες περιπτώσεις προβλημάτων με τις λύσεις τους είναι διαθέσιμες σε μορφές κειμένου. Για παράδειγμα: νομικά ζητήματα, φαρμακευτικές αγωγές, επικοινωνία μέσω ηλεκτρονικού ταχυδρομείου.

Στόχος των TCBR συστημάτων είναι να χρησιμοποιήσει αυτές τις πηγές γνώσεις σε μορφή κειμένου με έναν αυτοματοποιημένο ή ημι-αυτοματοποιημένο τρόπο για την υποστήριξη και επίλυση προβλημάτων

μέσω της σύγκρισης περιπτώσεων. Δεν υπάρχει κάποιο είδος πρότυπου για την δομή που θα πρέπει να ακολουθεί το κείμενο που θα αποτελέσει πηγή γνώσης του συστήματος, υπάρχουν όμως δύο κατηγορίες της μορφής του κειμένου, δομημένη και αδόμητη. Η δομημένη μορφή κειμένου αφορά κείμενα που περιέχουν συγκεκριμένη πληροφορία σε κάθε περίπτωση, και ακολουθούν συγκεκριμένη δομή. Η αδόμητη μορφή αφορά κείμενα που είναι γραμμένα σε ελεύθερη μορφή. Δεν υπάρχει κάτι κοινό σε όλες τις περιπτώσεις ως προς την πληροφορία που πρέπει να περιέχουν. Οι αδόμητη μορφή κειμένων είναι και η πιο δημοφιλής καθώς τα περισσότερα κείμενα διατίθενται σε αυτή τη μορφή.

Αν και είναι δύσκολο να καθοριστεί μια κοινή λειτουργία για τα συστήματα TCBR, αρκετοί ερευνητές έχουν ορίσει τις απαιτήσεις αυτών των συστημάτων. Κάποιες από αυτές τις απαιτήσεις είναι:

- Πως αξιολογείται η ομοιότητα μεταξύ περιπτώσεων σε μορφή κειμένου
- Πως πραγματοποιείται η χαρτογράφηση των περιπτώσεων σε μορφή δομημένου κειμένου.
- Πώς να προσαρμοστούν τα κείμενα σε περιπτώσεις

4.3 Συστήματα CBR

Έχουν αναπτυχθεί διάφορα συστήματα που παρέχουν τεχνικές συλλογιστικής βασισμένες σε περιπτώσεις (Case Based Reasoning). Υπάρχουν ολοκληρωμένες εφαρμογές που μπορούν να χρησιμοποιηθούν αυτούσιες και έχουν αναπτυχθεί η κάθε μία για να καλύπτει συγκεκριμένες ανάγκες είτε στον κλάδο της βιομηχανίας είτε σε ακαδημαϊκές κοινότητες. Κάποιες από αυτές είναι:

- Case-Based Poker Game (ομάδα AI, Πανεπιστήμιο του Auckland): Ένα ολοκληρωμένο παιχνίδι Poker που παίζεται σε τέσσερα βασικά στάδια.
- SMART: Αυτοματοποιημένη υποστήριξη σχετικά με τη διαχείριση τεχνολογίας για την εξυπηρέτηση των πελατών.
- FormTool: Επιλογή χρωμάτων για την εφαρμογή στα πλαστικά είδη.
- Vidur – Ένα έξυπνο σύστημα παροχής συμβουλών για τους αγρότες της Βορειοανατολικής Ινδίας.

Επίσης έχουν αναπτυχθεί και διάφορα εργαλεία τόσο από τη βιομηχανία όσο και από τον ακαδημαϊκό χώρο για την κατασκευή CBR συστημάτων. Κάποια από αυτά είναι ανοιχτού κώδικα και χρησιμοποιούνται ως βάση για την ανάπτυξη συστημάτων που εφαρμόζουν τεχνικές CBR. Τα εργαλεία αυτά παρέχουν όλες τις βασικές τεχνικές με αυτοτελή τρόπο ώστε να αναπτυχθεί στη συνέχεια ένα ολοκληρωμένο σύστημα για την κάλυψη συγκεκριμένων αναγκών. Δεν υπάρχει ένα πρότυπο εργαλείο που μπορεί να χρησιμοποιηθεί αυτούσιο σε όλα τα συστήματα. Στη συνέχεια ακολουθούν κάποια από τα βασικά εργαλεία που έχουν αναπτυχθεί μέχρι και σήμερα ώστε να επιλεγεί το πιο κατάλληλο για την ενσωμάτωση του στο σύστημα μας. Αυτά είναι:

- *FreeCBR*: Σύστημα ανοιχτού λογισμικού ανεπτυγμένο σε java που εφαρμόζει τεχνικές Case Based Reasoning. Είναι μια αυτόνομη εφαρμογή GUI σε γραμμή εντολών που παρέχει

διαδικτυακή εφαρμογή. Η χρήση του εργαλείου είναι αρκετά απλή αλλά οι δυνατότητες που παρέχει είναι πολύ περιορισμένες. Δεν παρέχει καμία δυνατότητα για ανάπτυξη συλλογισμού βασισμένη σε περιπτώσεις κειμένου, που είναι και το ζητούμενο για την ανάπτυξη της εργασίας.

- *jCOLIBRI* – Ένα πλαίσιο Case Based Reasoning που χρησιμοποιείται ως η βάση για την ανάπτυξη ενός ολοκληρωμένου συστήματος που θα ικανοποιεί τις ανάγκες για τις οποίες αναπτύχθηκε. Είναι ανοικτού λογισμικού και έχει αναπτυχθεί στη γλώσσα προγραμματισμού Java από το Πανεπιστήμιο Complutense της Μαδρίτης. Προσφέρει μια νέα αρχιτεκτονική που περιλαμβάνει επεκτάσεις για την κατασκευή: textual CBR, διαχείρισης οντολογιών, βάσεων δεδομένων, και διαδικτυακή απευθείας σύνδεση.
- *myCBR* – Ένα εργαλείο Case Based Reasoning ανοιχτού λογισμικού που αναπτύχθηκε από το Γερμανικό κέντρο ερευνών Τεχνητής Νοημοσύνης. Υποστηρίζει τη διδασκαλία και την έρευνα για την προσέγγιση των τεχνικών CBR. Είναι εύκολο στη χρήση, και επεκτάσιμο στις βασικές λειτουργίες που παρέχει. Το σημαντικότερο ως προς την έρευνα που πραγματοποιείται είναι ότι δεν μπορεί να εφαρμοστεί για συλλογιστική βασισμένη σε περιπτώσεις κειμένου(textual CBR).
- *NaCoDAE* (Navy Conversational Decision Aids Environment) - Ένα εργαλείο που χρησιμοποιεί τεχνικές Conversational case-based reasoning (CCBR), μια υποκατηγορία των συστημάτων CBR. Ο τρόπος λειτουργίας του βασίζεται στο να χρησιμοποιεί τον γραπτό διάλογο μεταξύ χρήστη και συστήματος καθορίζοντας την επόμενη ερώτηση βάσει του λεκτικού της απάντησης. Το εργαλείο αυτό περιορίζεται αποκλειστικά στην παροχή του συγκεκριμένου τύπου CBR χωρίς να προσφέρει τη δυνατότητα της επέκτασης.

4.4 Σύστημα CBR - Jcolibri

Στο προτεινόμενο πλαίσιο χρησιμοποιήθηκε το JColibri σαν πυρήνας του CBR. 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[18],[49].

1. Η πρώτη κατηγορία βασίζεται σε τεχνικές IE 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 αλγορίθμους 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.

4.4.1 Σημασιολογική (semantic) Μέθοδος

Η πρώτη κατηγορία ανήκει στη σημασιολογική ομάδα ανάκτησης, και βασίζεται στο μοντέλο στρωμάτων του Lenz (Lenz layered model). Το μοντέλο αυτό διαχωρίζει την επεξεργασία του κειμένου στα παρακάτω στάδια.

- Επίπεδο Λέξη-Κλειδί(Keyword Layer). Στο επίπεδο αυτό χωρίζεται το κείμενο σε όρους και στη συνέχεια διαγράφονται οι λέξεις αποκλεισμού(Stop Word). Οι λέξεις αποκλεισμού είναι αυτές που εμφανίζονται πολύ συχνά στο περιεχόμενο του εγγράφου, με αποτέλεσμα να περιέχουν πολύ λίγη πληροφορία για τα περιεχόμενα του εγγράφου όπου εμφανίζεται.
- Phrase Layer. Αναγνωρίζει συγκεκριμένους τομείς φράσεων με τη χρήση λεξικού.
- Treasures Layer. Σε αυτό το επίπεδο προσδιορίζονται συνώνυμα και σχετικοί όροι.
- Glossary Layer. Είναι μια συγκεκριμένη έκδοση του προηγούμενου επιπέδου που συνήθως καθορίζεται μια κοινή διεπαφή για τα δύο αυτά στρώματα.
- Feature Value Layer. Αφορά δομημένες και ημι-δομημένες περιπτώσεις, όπου εντοπίζονται τα χαρακτηριστικά τους και αποθηκεύονται σε ζευγάρια ως <χαρακτηριστικό, τιμή> για την αναπαράσταση αυτών των περιπτώσεων.
- Domain Structure Layer. Χρησιμοποιεί το προηγούμενο στρώμα για να κατηγοριοποιήσει τα κείμενα σε υψηλό επίπεδο. Δηλαδή ορίζει κάποιο χαρακτηριστικό ¨Θέμα¨ στην εκάστοτε περίπτωση για να μπορεί να χρησιμοποιηθεί για τη διαδικασία δημιουργίας ευρετηρίου.
- Information Extraction Layer. Κάποια σημεία του κειμένου μπορεί να εκφραστούν καλύτερα με μια δομημένη προσέγγιση. Το συγκεκριμένο επίπεδο μπορεί να καλυφθεί και από τα δύο προηγούμενα.

Στο τελευταίο στρώμα ορίζονται οι κανόνες για τους χρήστες έτσι ώστε να εξάγονται τα χαρακτηριστικά χρησιμοποιώντας πληροφορίες που συγκεντρώθηκαν από τα προηγούμενα επίπεδα. Έτσι οι προγραμματιστές μπορούν να λάβουν μια δομημένη απεικόνιση των περιπτώσεων και να τη διαχειρίζονται από κάποια πρότυπα σύγκρισης ομοιότητας που χρησιμοποιούνται από τα συστήματα CBR.

4.4.2 Στατιστική (statistical) Μέθοδος

Η Στατιστική μέθοδος δίνει εξίσου αξιόπιστα αποτελέσματα με την προηγούμενη μεθοδολογία. Βασίζεται σε διαδικασίες που χρησιμοποιούνται στην μηχανή αναζήτησης Apache Lucene[13]. 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 τρόπο 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.

5 Το Σύστημα μας

Στο μάθημα Εισαγωγή στην Επιστήμη Υπολογιστών του τμήματος Επιστήμης Υπολογιστών, για την επιτυχή διεξαγωγή του μαθήματος χρησιμοποιείται το σύστημα διαχείρισης μάθησης Efront. Η επικοινωνία εκπαιδευτή και εκπαιδευόμενου γίνετε μέσω αυτής της πλατφόρμας που προσφέρει κάποια εργαλεία επικοινωνίας. Το εργαλείο επικοινωνίας που χρησιμοποιείται στο συγκεκριμένο μάθημα για την επικοινωνία εκπαιδευτή – εκπαιδευόμενου είναι το φόρουμ. Εκεί ο κάθε εκπαιδευόμενος μπορεί να δημοσιεύει την ερώτηση/απορία του και να περιμένει την απάντηση του εκπαιδευτή.

Το διάστημα που χρησιμοποιήθηκε η επικοινωνία μέσω του φόρουμ, παρατηρήθηκε ότι οι εκπαιδευόμενοι σ' ένα μεγάλο βαθμό είχαν κοινές απορίες. Παρόλο που στο φόρουμ οι απορίες των εκπαιδευόμενων μαζί με τις λύσεις του εκπαιδευτή δημοσιεύονται και είναι ορατές σε όλους τους εγγεγραμμένους στο μάθημα, ο εκπαιδευόμενος προτιμάει να δημοσιεύει μια καινούργια απορία από το να ανατρέχει σε όλες αυτές που έχουν δοθεί στο παρελθόν. Ο κύριος λόγος είναι ότι η πληροφορία που υπάρχει στο φόρουμ με την πάροδο του χρόνου και σε αντιστοιχία με το πλήθος των φοιτητών αυξάνεται με ραγδαίους ρυθμούς. Αυτό έχει σαν αποτέλεσμα για τον εκπαιδευτή η επίλυση όλων των αποριών που προκύπτουν καθημερινά να είναι μια χρονοβόρα διαδικασία, και για τον εκπαιδευόμενο να υπάρχει κάποια καθυστέρηση της λύσης στην απορία του, που μπορεί να είναι λεπτά, ώρες και σε κάποιες περιπτώσεις και ημέρες.

Το σύστημα που αναπτύχθηκε στα πλαίσια της εργασίας αυτής 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 περιγράφτηκε παραπάνω, και πάνω σε αυτό ενσωματώθηκαν όλες οι λειτουργίες που υλοποιήθηκαν για την ολοκλήρωση της συγκεκριμένης εργασίας. Στο προηγούμενο κεφάλαιο περιγράφτηκαν 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 επιστρέφετε 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.

Στο συγκεκριμένο κεφάλαιο προσδιορίζουμε την δική μας επέκταση που αναπτύχθηκε πάνω στο προτεινόμενο σύστημα Jcolibri και συγκεκριμένα στη στατιστική μέθοδο ανάκτησης. Αναλύονται οι μεθοδολογίες που αναπτύχθηκαν για την εφαρμογή τους ως εναλλακτικές μεθόδους ανάκτησης. Πριν όμως παρουσιαστούν οι μεθοδολογίες που αναπτύχθηκαν περιγράφουμε την υπάρχουσα μέθοδο ανάκτησης καθώς και τις συναρτήσεις εντοπισμού ομοιότητας.

5.1 Δική μας Επέκταση

Στο δικό μας σύστημα επεκτείνουμε την αρχική στατιστική μέθοδο που προσφέρει το Jcolibri 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 που θα περιγραφτούν πιο αναλυτικά στη συνέχεια. 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 τόσου 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.

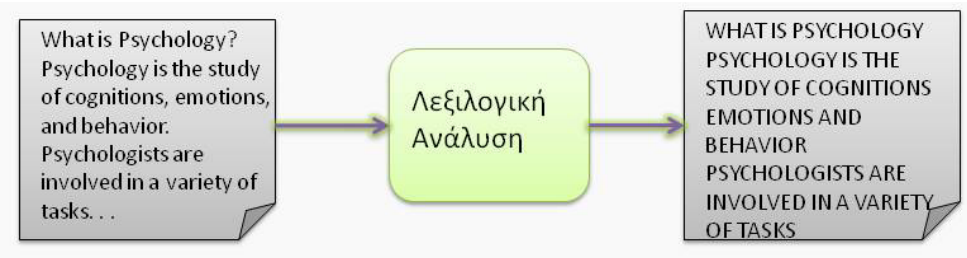
5.2 Στάδιο προ-επεξεργασίας

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 7 και περιγράφονται αναλυτικά στη συνέχεια.

Εικόνα 7 - Στάδια Προ-επεξεργασίας

5.2.1 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 8 - Παράδειγμα Λεξιλογικής Ανάλυσης

5.2.2 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.

5.2.3 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 [24]. 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).

5.3 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(qi, dj)] 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(qi, dj) είναι μια συνάρτηση κατάταξης, η οποία συνδέει έναν πραγματικό αριθμό με ένα ερώτημα qi ∈ Q και μια αναπαράσταση κειμένου dj ∈ D. Μια τέτοια κατάταξη ορίζει μια διάταξη πάνω στα κείμενα πάντα με βάση το ερώτημα qi.

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.

5.3.1 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.

5.3.1.1 Διαδικασία δημιουργίας Καταλόγου

Για κάθε μία από τις μεθοδολογίες που υλοποιήθηκαν χρησιμοποιείται η βιβλιοθήκη Lucene για την ανάκτηση πληροφοριών και την εφαρμογή των συναρτήσεων ομοιότητας. Στην ανάκτηση της πληροφορίας όμως είναι απαραίτητη η χρήση ενός καταλόγου ώστε να αποφευχθεί η εξέταση όλων των εγγράφων. Στην αντίθετη περίπτωση, το μόνο που θα μπορούσε να εφαρμοστεί είναι η σειριακή αναζήτηση. Η σειριακή αναζήτηση σε μεγάλες συλλογές κειμένων είναι μια χρονοβόρα διαδικασία που

πρέπει να αποφεύγεται. Ο ανεστραμμένος κατάλογος είναι η πιο συχνά χρησιμοποιούμενη μορφή καταλόγου λόγω της απόδοσής του, και της απλής χρήσης του. Ο όρος ανεστραμμένος εκφράζει ότι αντί να αποθηκεύουμε πληροφορία για το ποιοι όροι περιέχονται σε ένα έγγραφο, καταγράφουμε το σύνολο των εγγράφων που περιέχουν ένα συγκεκριμένο όρο.

Ένας ανεστραμμένος κατάλογος απαρτίζεται από δύο μέρη, τους όρους(λέξεις)και τις λίστες εμφανίσεων. Έτσι κάθε όρος αντιστοιχεί με μία λίστα που περιγράφει πληροφορίες σχετικά με την εμφάνιση του όρου μέσα στα έγγραφα. Στο σύστημά μας η πληροφορία αυτή αφορά το πλήθος των εμφανίσεων του όρου στη συλλογή εγγράφων, τον κωδικό του εγγράφου που εμφανίζεται ο όρος μαζί με το πλήθος των εμφανίσεων του σε αυτό.

Η λειτουργία ξεκινάει διαβάζοντας το αρχείο κειμένου που αποτελεί τη βάση του συστήματος, και δημιουργεί ένα Lucene Document για κάθε μία περίπτωση που διαβάζει. Οι περιπτώσεις στο αρχείο χωρίζονται μεταξύ τους με τρεις κενές γραμμές. Το Lucene Document αποτελεί ένα τμήμα του καταλόγου και της αναζήτησης.

Εφόσον έχουν δημιουργηθεί τα Lucene Document μπορεί να εφαρμοστεί το βήμα της προ-επεξεργασίας που περιγράψαμε παραπάνω. Στη συνέχεια τα δεδομένα αποθηκεύονται σε μία δομή ανεστραμμένου καταλόγου. Ουσιαστικά με την χρήση του ανεστραμμένου καταλόγου κάθε λέξη που περιέχεται στην περίπτωση (μετά την προ-επεξεργασία) τη θεωρούμε ως ένα όρο που περιέχει κάποια πληροφορία για αυτή. Έτσι αντί να προσπαθήσουμε να αναζητήσουμε τις λέξεις που περιέχονται σε κάθε μία περίπτωση μπορούμε να εντοπίζουμε πολύ εύκολα ποιες περιπτώσεις περιέχουν την λέξη που αναζητούμε.

Στη συνέχεια δίνεται ένα μικρό παράδειγμα ανεστραμμένου καταλόγου. Έστω ότι στη βάση υπάρχουν τρεις περιπτώσεις:

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 που δημιουργείται παρουσιάζεται στο πίνακα που ακολουθεί

Λέξεις	**Περιπτώσεις**
How	3: (d1,1), (d2,1), (d3,1)
do	2: (d1,1), (d2,1)
I	3: (d1,1), (d2,1), (d3,1)
Create	1: (d1,1)
class	3: (d1,2), (d2,1)

method	2: (d1,1), (d2,1)
and	1: (d1,1)
static	1: (d1,1)
call	1: (d2,1)
can	1: (d3,1)
organize	1: (d3,1)

Πίνακας 1 - Παράδειγμα Ανεστραμμένου Καταλόγου

Στη δεύτερη στήλη του πίνακα εμφανίζεται αρχικά ένας αριθμός που αναπαριστά το πλήθος τον εμφανίσεων της λέξης μέσα σε όλα τα έγγραφα, και στη συνέχεια ένα ζεύγος με τον κωδικό του εγγράφου που περιέχεται μαζί με το πλήθος εμφανίσεων του σε αυτό.

5.3.2 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 εγγραφών 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 $\mathbf{w_{t,d}}$ και ισούται:

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

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:

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

Όπου *N* το πλήθος των εγγράφων, και f_t 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:

$$\boldsymbol{wt,d} = \frac{\boldsymbol{tf_{\mathrm{t,d}}} * \boldsymbol{idf_{\mathrm{d}}}}{\sqrt{\Sigma_d}(w_{t,d}^2)}$$

5.3.3 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 t_i μπορούμε να δώσουμε μία τιμή σε κάθε περίπτωση d_j , 0 ή 1. Η τιμή 0 σημαίνει ότι ο όρος t_i 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 περιγράφτηκαν, 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 επιλέχτηκαν 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:

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

Όπου

- *tf(t 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:

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

- *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:

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

- *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:

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

- *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 ^ 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.

5.4 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.

5.4.1 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 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 παρουσιάζοντα στη συνέχεια.

5.4.2 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 γίνετε 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 μικρότερου μεγέθους. 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.

5.4.3 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 λόγο 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 ίση του 0,6 με την ένδειξη «Μήπως εννοείτε: ».

5.4.3.1 Το Ηλεκτρονικό Λεξικό WordNet

Το WordNet είναι ένα ηλεκτρονικό λεξικό βασισμένο στις έννοιες των λέξεων και στις σημασιολογικές σχέσεις που ισχύουν μεταξύ τους. Δημιουργήθηκε βασισμένο στον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου, με τον οποίο ο άνθρωπος μαθαίνει τα νοήματα των λέξεων. Με τη βοήθεια των σημασιολογικών σχέσεων, οι έννοιες των λέξεων σε μια γλώσσα μπορούν να διασυνδεθούν, αποτελώντας ένα δίκτυο εννοιών.

Τα σημασιολογικά δίκτυα αποτελούν μια μορφή ηλεκτρονικών λεξικών, στα οποία καταχωρείται σημασιολογική πληροφορία για τα λήμματα που περιλαμβάνουν. Οι έννοιες των λέξεων σε μια γλώσσα, μπορούν να διασυνδεθούν με τη βοήθεια των σημασιολογικών σχέσεων, αποτελώντας ένα δίκτυο εννοιών, όπως το WordNet.

Τα δίκτυα σημασιολογικών συσχετίσεων παρουσιάζουν αρκετά κοινά χαρακτηριστικά γνωρίσματα με τα παραδοσιακά ερμηνευτικά λεξικά και τους θησαυρούς συνωνύμων, καθώς παρέχουν πληροφορία όχι μόνο για τις ερμηνείες των λημμάτων που περιέχουν αλλά και για τις σημασιολογικές τους σχέσεις με άλλα λήμματα του δικτύου. Ωστόσο, παρά τις ομοιότητές τους, εμφανίζουν και σημαντικές διαφοροποιήσεις με τα υφιστάμενα λεξικά εφόσον η πληροφορία που περιλαμβάνουν στηρίζεται αποκλειστικά σε σημασιολογικά και όχι λεξιλογικά κριτήρια.

Ένα σύνολο συνωνύμων (synset) είναι ένα σύνολο λέξεων οι οποίες σε ένα συγκεκριμένο περιβάλλον μπορούν να χρησιμοποιηθούν η μία στη θέση της άλλης. Η βασική διαχώριση που πραγματοποιεί το WordNet στις λέξεις είναι στα τέσσερα μέρη του λόγου: Ουσιαστικά (nouns), ρήματα (verbs), επίθετα (adjectives) και επιρρήματα(adverbs). Ο διαχωρισμός πραγματοποιείται επειδή το κάθε ένα είδος ακολουθεί διαφορετικούς γραμματικούς κανόνες. Τα περισσότερα συνώνυμα (synset) συνδέονται με άλλα μέσω ενός αριθμού σημασιολογικών σχέσεων. Κάποιες από τις πιο κοινές σχέσεις που καθορίζονται είναι:

- *Υπερώνυμο /* Υπώνυμο *(hypernym/* hyponyms*):* Η λέξη έχει γενικότερη έννοια από αυτήν της οποίας είναι υπερώνυμο. Το Α είναι υπερώνυμο του Β (ή με άλλα λόγια το Β είναι υπώνυμο του Α), σημαίνει ότι το Β είναι ένα είδος Α. Π.χ. Το {ασθενοφόρο, νοσοκομειακό} είναι ένα υπώνυμο του {αυτοκίνητο} (δηλ. το ασθενοφόρο είναι ένα είδος αυτοκινήτου.
- *Μερώνυμο / Ολώνυμο(meronym / holonym):* Το Α είναι ολώνυμο του Β (ή αλλιώς το Β είναι μερώνυμο του Α), σημαίνει ότι το Β είναι μέρος του Α. Π.χ., Ο {προφυλακτήρας} είναι μερώνυμο του {αυτοκίνητο} (δηλ. ο προφυλακτήρας είναι μέρος του αυτοκινήτου)

5.4.4 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 περιγράφτηκε 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:

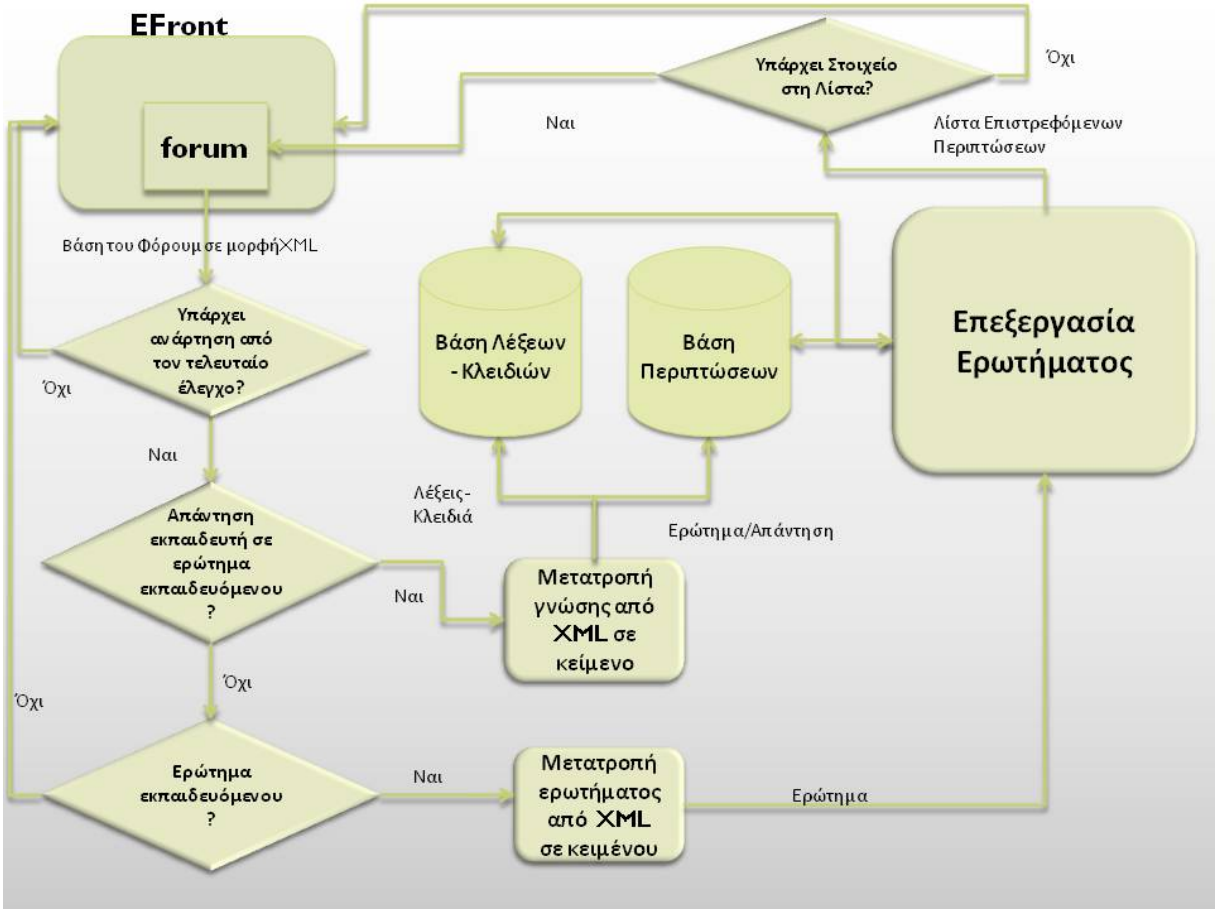
$$\mathbf{v = x * w_1 + y * w_2}$$

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.

6 Υλοποίηση

Στο κεφάλαιο αυτό θα περιγράψουμε την αρχιτεκτονική του συστήματος που υλοποιήθηκε, και το κάθε ένα επιμέρους μέρος της. Στην Εικόνα 9 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 9 - 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.

6.1 Efront

Το efront αποτελεί μία μοντέρνα πλατφόρμα ηλεκτρονικής μάθησης που χρησιμοποιείται για την αποδοτικότερη διεξαγωγή του μαθήματος ΗΥ-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 9.

Για να γίνει πιο σαφές στην Εικόνα 10 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 10 - Στιγμιότυπο του Φόρουμ στο Efront

Στο φόρουμ του μαθήματος μπορεί να υπάρχει 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 11. 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="n">a</span> <span class="o">=</span> <span class="p">[</span><span class="s">'spam'</span><span class="p">,</span> <span
      class="s">'eggs'</span><span class="p">,</span> <span class="mi">100</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 11 - Αρχείο xml που εξάγεται από τη βάση του φόρουμ

6.2 *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:

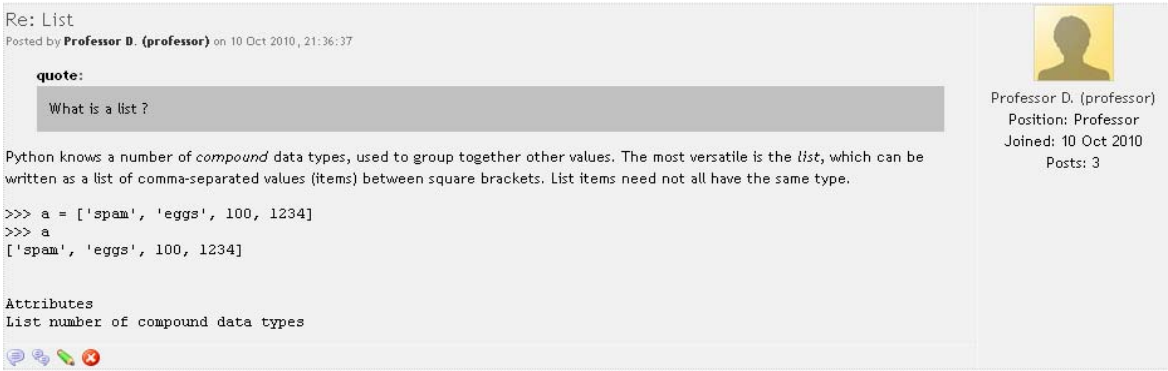
- Αν υπάρχει ανάρτηση από κάποιον εκπαιδευόμενο χωρίς να έχει δοθεί απάντηση από τον εκπαιδευτή.
- Αν υπάρχει απάντηση(ανάρτηση) από τον καθηγητή σε κάποια απορία του εκπαιδευόμενου.

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 12 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 12 - Στιγμιότυπο Φόρουμ προσθέτοντας τις Λέξεις-Κλειδιά

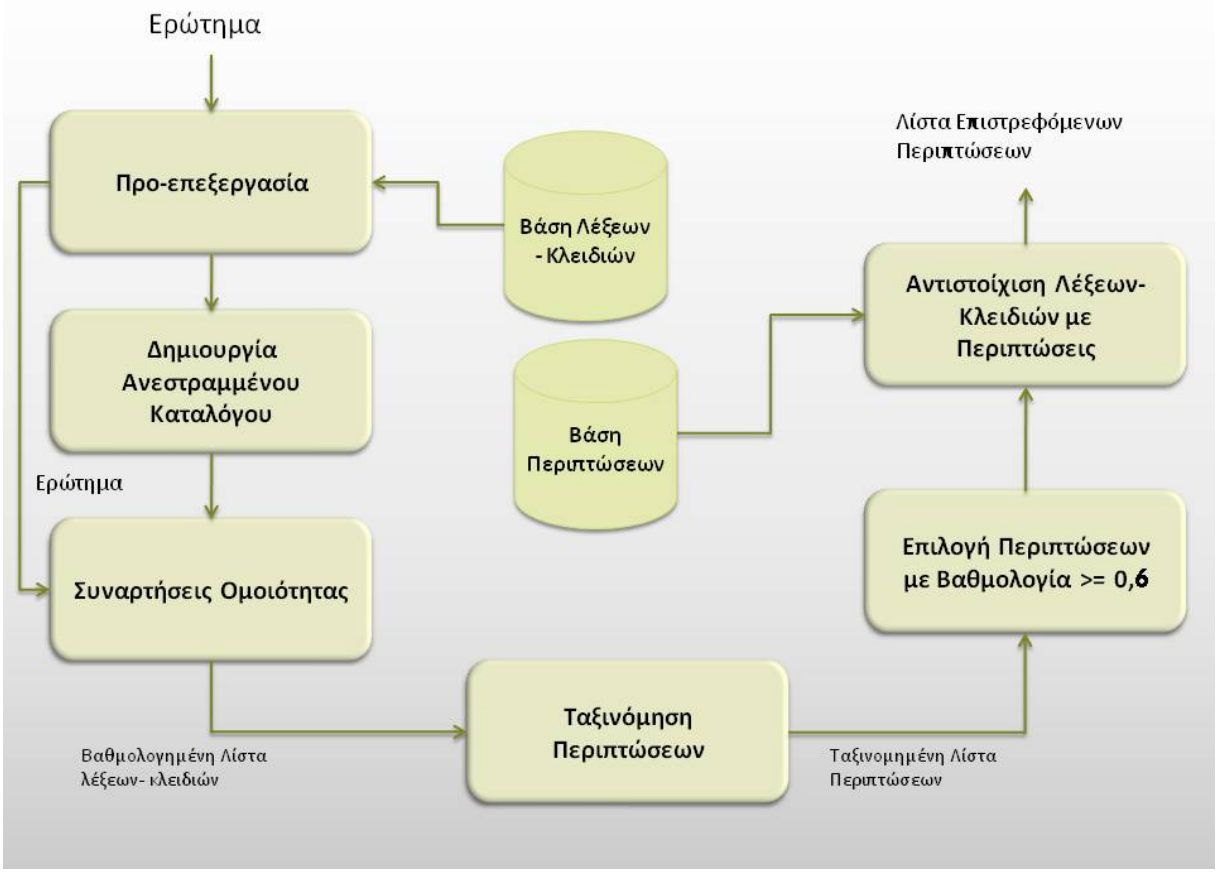
6.3 Επεξεργασία Ερωτήματος

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 13 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 13 - 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 προ-επεξεργασία του κειμένου για κάθε ομάδα λέξεων κλειδιών. Το κείμενο μετά την προ-επεξεργασία 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 2 - Μορφή Ανεστραμμένου Καταλόγου

Αν θεωρήσουμε ως αρχικό ερώτημα την πρόταση « What is a class?» , η προ-επεξεργασία 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 :

$$A = \{O\} U \{d1, d3\} = \{d1, d3\}$$

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: $\text{score}(q,d) = \text{coord}(q,d) \cdot \text{queryNorm}(q) \cdot \sum(\text{tf}(t \text{ in } d) \cdot \text{idf}(t)^2 \cdot w_{t\ in\ q} \cdot \text{lengthNorm}(t \text{ in } q)$

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 ιση 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.

7 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 ΗΥ-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 εδική 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.

7.1 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 ΗΥ-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 χαρακτηρισικά 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 απιεκόνιση της κάθε περίπτωσης με τα χαρακτηριστικά τους.

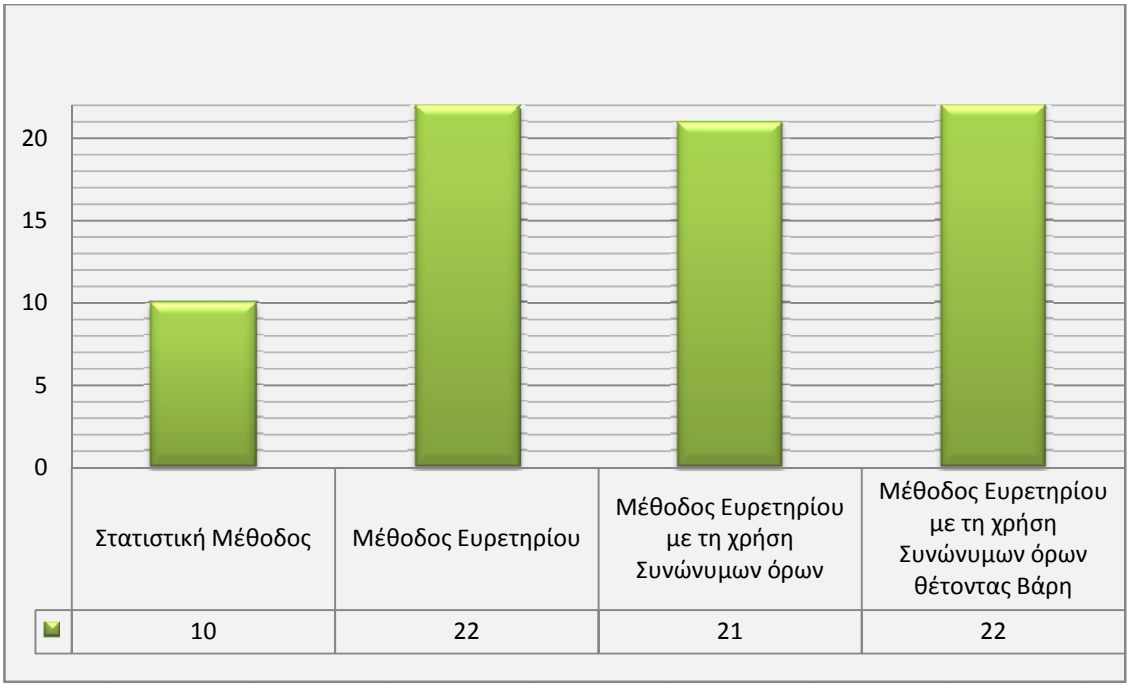
Δημιουργήθηκαν δύο βάσεις, η βάση περιπτώσεων και η βάση λέξεων κλειδιών, με το ίδιο πλήθος στοιχείων στην κάθε μία. Οι βάσεις αυτές χρησιμοποιήθηκαν και από τις τέσσερις μεθοδολογίες που δημιουργήθηκαν και σχεδιάστηκαν με αυτό το σκοπό. Για όλες τις μεθοδολογίες εξετάστηκαν οι ίδιες ερωτήσεις. Έτσι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

7.1.1 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 ελέγχτηκαν 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 14- Αποτελέσματα Αξιολόγησης για τη Βάση στην Επιστήμη Υπολογιστών

Όπως παρατηρούμε και από το διάγραμμα η *στατιστική μέθοδος* στις είκοσι δύο ερωτήσεις που τέθηκαν μόνο στις 10 από αυτές εντόπισε την ιδανική απάντηση. Ιδανική θεωρείται η απάντηση η οποία θα καλύπτει το μεγαλύτερο δυνατό της ερώτησης που τέθηκε, σε σχέση με την διαθέσιμη γνώση που υπάρχει αποθηκευμένη. Κατά τη διάρκεια εκτέλεσης της πρώτης διαδικασίας ανάκτησης, που περιλαμβάνει το Λογικό μοντέλο, για την επιλογή των περιπτώσεων που περιέχουν έστω και ένα όρο της ερώτησης, η ιδανική περίπτωση για κάθε μία ερώτηση επιλέχτηκε. Στην εφαρμογή του διανυσματικού μοντέλου για τον υπολογισμό της βαθμολογίας σχετικά με την ομοιότητα του ερωτήματος με τις περιπτώσεις που επιλέχτηκαν από την προηγούμενη διαδικασία, η ιδανική περίπτωση στις 11 από τις 22 περιπτώσεις βαθμολογήθηκε με τιμή κάτω του 0,6, στη μία βαθμολογήθηκε με τιμή 0,9(δεν αποτελούσε την επιστρεφόμενη περίπτωση αλλά ήταν μία από τις προτεινόμενες που θα επιστρεφόταν), ενώ στις 10 από αυτές με την μεγαλύτερη βαθμολογία. Οι ερωτήσεις για τις οποίες η μεθοδολογία απέτυχε να εντοπίσει την καταλληλότερη να αναφέρουμε ότι άνηκαν σε όλα τα είδη των ερωτήσεων σχετικά με την πληροφορία που υπάρχει για να τις καλύψει. Εύκολα μπορούμε να συμπεράνουμε ότι μία τέτοια μεθοδολογία που η αποτυχία στην εύρεση της καταλληλότερης απόφασης ήταν περισσότερες από τις μισές, δεν αρκεί για την εφαρμογή της στο σύστημα για την ολοκλήρωση του. Παρόλα αυτά, θα ελέγξουμε τη συμπεριφορά της και σε μια βάση που δεν χρησιμοποιεί όρους κάποιας ορολογίας, όπως είναι η δεύτερη βάση που έχει κατασκευαστεί.

Σχετικά με τη *μέθοδο του ευρετηρίου*, όπως φαίνεται και στο γράφημα, σε ένα αντικείμενο γνώσης που χρησιμοποιείται ορολογία και ο ορισμός των λέξεων κλειδιών έχουν πραγματοποιηθεί με προσοχή μπορεί να φέρει τα μέγιστα δυνατά αποτελέσματα. Στον έλεγχο που πραγματοποιήθηκε η μεθοδολογία

του ευρετηρίου κατάφερε να ικανοποιήσει όλες τις ερωτήσεις εντοπίζοντας στη κάθε μία την ιδανική περίπτωση. Οι ερωτήσεις που δοκιμάστηκαν ήταν ίδιες με την προηγούμενη μεθοδολογία ώστε να παρατηρήσουμε τη συμπεριφορά τους στις ίδιες ερωτήσεις για μία συγκεκριμένη βάση γνώσεων. Με τη δεδομένη βάση σε όλες τις ερωτήσεις η μέθοδος του ευρετηρίου εντόπισε τη πιο κατάλληλη απάντηση για κάθε δεδομένη ερώτηση, ακόμη και σε μία ερώτηση που δεν υπήρχε κάποια απάντηση από τη βάση που να την ικανοποιεί δεν επέστρεψε αποτέλεσμα, σε αντίθεση με τη στατιστική μέθοδο που για την συγκεκριμένη ερώτηση επέστρεψε απάντηση, που ήταν προφανώς λανθασμένη. Να αναφέρουμε επίσης ότι για όλες τις ερωτήσεις η ιδανική απάντηση ήταν πάντα η πρώτη που επιστρεφόταν, δηλαδή με βαθμολογία 1,0 , και όχι κάποια από τις προτεινόμενες. Με το όρο προτεινόμενες εννοούμε τις απαντήσεις που έχουν βαθμολογία μεγαλύτερη του 0,6 και εμφανίζονται σε ένα μέρος τις απάντησης ως πιθανές απαντήσεις με την ένδειξη «Μήπως εννοείτε:».

Η επόμενη μέθοδος που δοκιμάστηκε είναι αυτή του *ευρετηρίου με τη χρήση συνώνυμων όρων*. Στη συγκεκριμένη μέθοδο ελέγχουμε και όλα τα συνώνυμα της ερώτησης για τον εντοπισμό της πιο κατάλληλης απάντησης. Στις είκοσι δύο ερωτήσεις όπως είναι εύκολα αντιληπτό και από το διάγραμμα ικανοποίησε τις είκοσι μία ερωτήσεις. Σε μία ερώτηση ανάκτησε την πιο κατάλληλη για εκείνη, αλλά η βαθμολογία της ήταν μικρότερη του 0,6 έτσι ώστε να μπορέσει να την επιστρέψει ως προτεινόμενη. Ως γνωστό στη συγκεκριμένη μεθοδολογία εντοπίζουμε και τους συνώνυμους όρους της ερώτησης όπως αυτοί προκύπτουν από τη μέθοδο της προ-επεξεργασίας. Η τελική ερώτηση προκύπτει από τους συνώνυμους όρους μαζί με την ερώτηση. 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, έτσι έχουμε ένα πλήθος συνώνυμων που μπορεί να φτάνει και το δεκαπλάσιο, από τους όρους του ερωτήματος. Καταγράφοντας όλους τους συνώνυμους όρους για να παρακολουθήσουμε τη συμπεριφορά του συστήματος, παρατηρούμε ότι οι συνώνυμοι όροι του ερωτήματος σε μία θεματολογία που χρησιμοποιείται ορολογία δεν έχουν και τόσο μεγάλη ισχύ. Συνήθως σε μία ορολογία ένας συνώνυμος όρος έχει τελείως διαφορετική ερμηνεία από τον πρωτότυπο όρο. Παρόλα αυτά μία ερώτηση δεν περιλαμβάνει μόνο όρους της ορολογίας, έτσι η εύρεση συνωνύμων αυτών των όρων παρατηρήσαμε ότι έχει αξία και σημασία. Γενικότερα η μέθοδος αυτή έχει σημασία ακόμα και σε μια θεματολογία με χρήση συνώνυμων όρων, αρκεί να έχουν μικρότερη ισχύ από τους όρους που υποβάλει ο χρήστης στο ερώτημα του. Η συμπεριφορά της μεθοδολογίας αυτής στις ερωτήσεις που υποβλήθηκαν ήταν αρκούντως ικανοποιητική, αν παρατηρήσουμε ότι σε μία απάντηση από τις είκοσι δύο δεν κατάφερε να εντοπίσει την ιδανική απάντηση, αλλά μη ικανοποιητική γιατί στην ερώτηση αυτή δεν κατάφερε να εντοπίσει ως μία προτεινόμενη την πιο κατάλληλη απάντηση, δηλαδή να την βαθμολογήσει με τιμή μεγαλύτερη ίση του 0,6. Η ερώτηση αυτή να αναφέρουμε ότι ανήκει στην κατηγορία των ερωτήσεων που υπάρχει ολοκληρωμένη πληροφορία για την επίλυσή της.

Η τελευταία μεθοδολογία του *ευρετηρίου με τη χρήση συνώνυμων όρων θέτοντας βάρη* όπως γίνεται εύκολα αντιληπτό από το διάγραμμα και στις είκοσι δύο ερωτήσεις εντόπισε την ιδανικότερη περίπτωση. Όπως έχει ήδη αναφερθεί στη συγκεκριμένη μεθοδολογία, η τελική βαθμολογία 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 στο αποτέλεσμα του καθενός. Προκειμένου το

αποτέλεσμα να είναι κανονικοποιημένο στο διάστημα [0,1] τα βάρη των αποτελεσμάτων πρέπει να έχουν άθροισμα 1.0. Τα βάρη θα πάρουν τέτοιες τιμές ώστε να ικανοποιούν το μέγιστο δυνατό τον αποτελεσμάτων. Για τη βάση στη θεματολογία της πληροφορικής για οποιαδήποτε ερώτηση όπου ο συντελεστής του αποτελέσματος της μεθόδου του ευρετηρίου είναι μεγαλύτερος ίσος του 0,6 ικανοποιούνται όλες οι ερωτήσεις. Στην περίπτωση που η τιμή ανήκει στο διάστημα [0.6, 0.8) στις εικοσιένα ερωτήσεις η ιδανική απάντηση έχει τη μέγιστη βαθμολογία δηλαδή 1.0. Αλλά στη μία η κατάλληλη απάντηση έχει βαθμολογία μεγαλύτερη του 0,6 έτσι ώστε να επιστραφεί ως προτεινόμενη, αλλά δεν είναι αυτή με μέγιστη βαθμολογία. Στην περίπτωση που ο συντελεστής του αποτελέσματος του ευρετηρίου είναι μεγαλύτερος ή ίσος του 0,8 τότε σε όλες τις ερωτήσεις επιστρέφεται η πιο κατάλληλη περίπτωση έχοντας καταλάβει και τη μεγαλύτερη βαθμολογία.

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.

7.2 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).

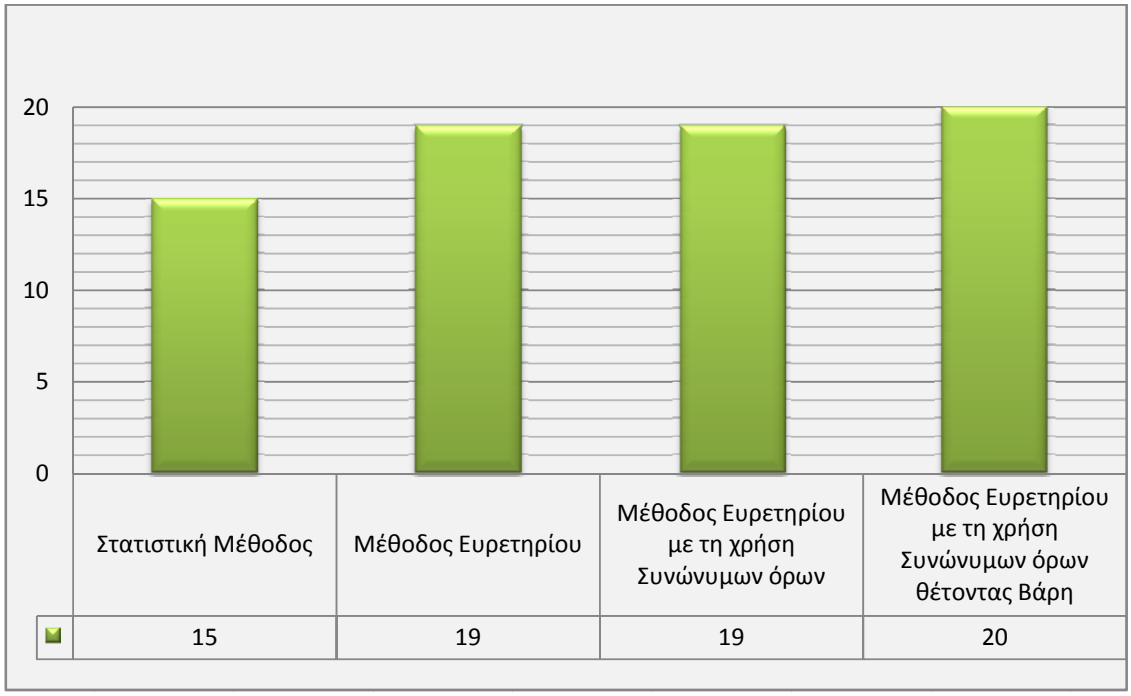
Για το υλικό αυτό δημιουργήθηκαν είκοσι ερωτήσεις που καλύπτουν όλα τα πιθανά ενδεχόμενα που μπορεί να προκύψουν. Δηλαδή για ένα πλήθος ερωτήσεων υπάρχει όλη η πληροφορία για την ακριβή επίλυσή τους, για κάποιες άλλες μπορεί να λυθεί ένα μέρος του προβλήματος, και κάποιες άλλες που δεν υπάρχει γνώση 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 τηνς 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 επιλεχτεί η πιο κατάλληλη μέθοδος για την εφαρμογή της στο σύστημα.

7.2.1 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 επιλέχτηκαν 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 15 - Αποτελέσματα Αξιολόγησης για τη Βάση στη Ψυχολογία

Από τις είκοσι ερωτήσεις που εξετάστηκαν στο τομέα της ψυχολογίας με την εφαρμογή της στατιστικής μεθόδου στις δεκαπέντε από αυτές εντοπίστηκε η καταλληλότερη απάντηση. Στις δύο από τις δεκαπέντε αυτές ερωτήσεις η ιδανική απάντηση δεν ήταν αυτή με τη μεγαλύτερη βαθμολογία αλλά επιστρέφεται ως προτεινόμενη με βαθμολογία μεγαλύτερη του 0,6. Στη λεπτομερή παρακολούθηση της εκτέλεσης της στατιστικής μεθοδολογίας, παρατηρούμε ότι σε όλες τις ερωτήσεις που τέθηκαν η ιδανικότερη

περίπτωση ανακτήθηκε από το Λογικό μοντέλο ως μία περίπτωση που περιέχει κάποιους όρους της ερώτησης. Στην εφαρμογή του Διανυσματικού μοντέλου η κατανομή της βαθμολογίας στις επιμέρους περιπτώσεις στις πέντε από τις δεκαπέντε ερωτήσεις δεν ήταν μεγαλύτερη του 0,6, και συγκεκριμένα ήταν μικρότερη του 0,4. Το συμπέρασμα από τη λεπτομερή παρατήρηση της εκτέλεσης αυτού του φαινομένου είναι το γεγονός ότι σε όλο το κείμενο της εκάστοτε περίπτωσης υπάρχει πληροφορία που δεν σχετίζεται άμεσα με το θέμα της πληροφορίας που απευθύνεται με αποτέλεσμα την παραπλάνηση των αποτελεσμάτων και κατ' επέκταση την αποτυχία εύρεσης της κατάλληλης περίπτωσης.

Η επόμενη μεθοδολογία που εξετάστηκε είναι αυτή του *ευρετηρίου* με τη χρήση των είκοσι ερωτήσεων που δημιουργήθηκαν για τη βάση αυτή. Στις ερωτήσεις αυτές η μέθοδος του ευρετηρίου αντεπεξήλθε στις δεκαεννέα από αυτές, εκ των οποίων η καταλληλότερη απάντηση ήταν και αυτή με τη μεγαλύτερη βαθμολογία. Στη μία όμως ερώτηση η απάντηση στην οποία δόθηκε η πιο μεγάλη βαθμολογία δεν ήταν η πιο κατάλληλη, και η βαθμολόγια που της αποδόθηκε ήταν 0.4375 μικρότερη του 0.6 έτσι ώστε να εμφανιστεί στη λίστα παρόμοιων απαντήσεων. Το αποτέλεσμα αυτό προέκυψε σε μία ερώτηση που είχε την ίδια έννοια με την ακριβώς προηγούμενη που είχε τεθεί, πιο συγκεκριμένα και οι δυο ερωτήσεις είχαν το ίδιο νόημα σχετικά με την κατηγοριοποίηση για τις ψυχικές ασθένειες. Να αναφέρουμε ότι στη βάση υπήρχαν πάνω από τέσσερις περιπτώσεις που χρησιμοποιούσαν στις λέξεις κλειδιά τους όρους ψυχικές ασθένειες(mental illness),οπότε η διαφορά στο θέμα ήταν στη χρήση της λέξης κατηγοριοποίηση. Η κατάλληλη περίπτωση στις λέξεις κλειδιά χρησιμοποιούσε τον όρο "classification" για να προσδιορίσει την έννοια την κατηγοριοποίησης, έτσι δημιουργήθηκαν δύο ερωτήσεις η μία της μορφής "How to classify mental illness?" και η άλλη της μορφής "How to sort mental illness?". Οι δύο αυτές ερωτήσεις ουσιαστικά έχουν αρχικά το ίδιο ακριβώς νόημα, και στην ουσία έχουν ακριβώς την ίδια πιθανότητα να επιλεχτούν για να εκφράσουν την ίδια απορία από κάποιον χρήστη. Έτσι θέτοντας αυτές τις δύο ερωτήσεις στην μέθοδο του ευρετηρίου, στη μεν πρώτη που υπάρχει η ίδια ρίζα του όρου με την έννοια κατηγοριοποίησης η βαθμολογία της είναι 1.0 με διαφορά από τις υπόλοιπες και στη δεύτερη η βαθμολογία της ήταν 0.4375. Ο λόγος της απόδοσης αυτής ήταν σε μία ακραία περίπτωση καθώς υπήρχαν πολλές ομάδες λέξεων κλειδιών που χρησιμοποιούσαν τους άλλους δύο όρους(mental illness) και σε μικρότερης έκτασης από αυτή. Το γενικό συμπέρασμα όμως από τη μία είναι ενθαρρυντικό καθώς από τις είκοσι ερωτήσεις εντόπισε την ιδανική περίπτωση στις δεκαεννέα και μάλιστα με τη μεγαλύτερη βαθμολογία, αλλά από την άλλη δεν είναι και τόσο ικανοποιητικό καθώς όσο η πληροφορία που υπάρχει καταχωρημένη αυξάνεται και υπάρχουν όροι που επαναλαμβάνονται στις ομάδες των λέξεων κλειδιών η συγκεκριμένη μεθοδολογία σε ένα αντικείμενο που δεν επεξεργάζεται κάποιου είδους ορολογία θα αδυνατεί να ανταπεξέλθει.

Στη συνέχεια εξετάστηκε η μέθοδος *του ευρετηρίου με τη χρήση συνώνυμων όρων*. Από τις είκοσι ερωτήσεις που ερευνήθηκαν στις δεκαεννέα από αυτές η ιδανική απάντηση εντοπίστηκε και επιστράφηκε και μάλιστα με την καλύτερη βαθμολογία. Σε μία ερώτηση όμως το αποτέλεσμα δεν ήταν το αναμενόμενο, καθώς ανακτήθηκε η ιδανική απάντηση αλλά η βαθμολογία που κατέλαβε ήταν μικρότερη του 0.5. Η αιτία παρουσίασης αυτού του φαινομένου οφείλεται στο γεγονός ότι συνώνυμοι όροι της ερώτησης περιλαμβάνονταν σε ένα πλήθος λέξεων κλειδιών για διαφορετικές περιπτώσεις. Καθώς στη συγκεκριμένη μέθοδο οι συνώνυμοι όροι είναι περισσότεροι από τους όρους της ερώτησης, αυτό έχει σαν αποτέλεσμα στη διαδικασία απονομής βαθμολογίας, επειδή οι συνώνυμοι όροι είναι περισσότεροι να

υποκινούν κάποιες φορές το αποτέλεσμα. Έτσι και στην ερώτηση που το αποτέλεσμα από τη μεθοδολογία ήταν λανθασμένο, ένας όρος της «Notice» να επιστρέψει ένα μεγάλο πλήθος συνώνυμων όρων, που κάποιοι από αυτούς περιλαμβάνονταν σε διαφορετικές ομάδες λέξεων κλειδιών, με αποτέλεσμα τη λανθασμένη απόδοση βαθμολογίας και την αποτυχία εύρεσης της κατάλληλης περίπτωσης.

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 w1 την βαθμολογία από τη μέθοδο του ευρετηρίου και w2 την βαθμολογία από τη μέθοδο του ευρετηρίου με τη χρήση μόνο συνώνυμων όρων, τότε η τελική βαθμολογία θα προκύψει ως:

$$x*w1 + y*w2 \quad \text{όπου } x+y = 1$$

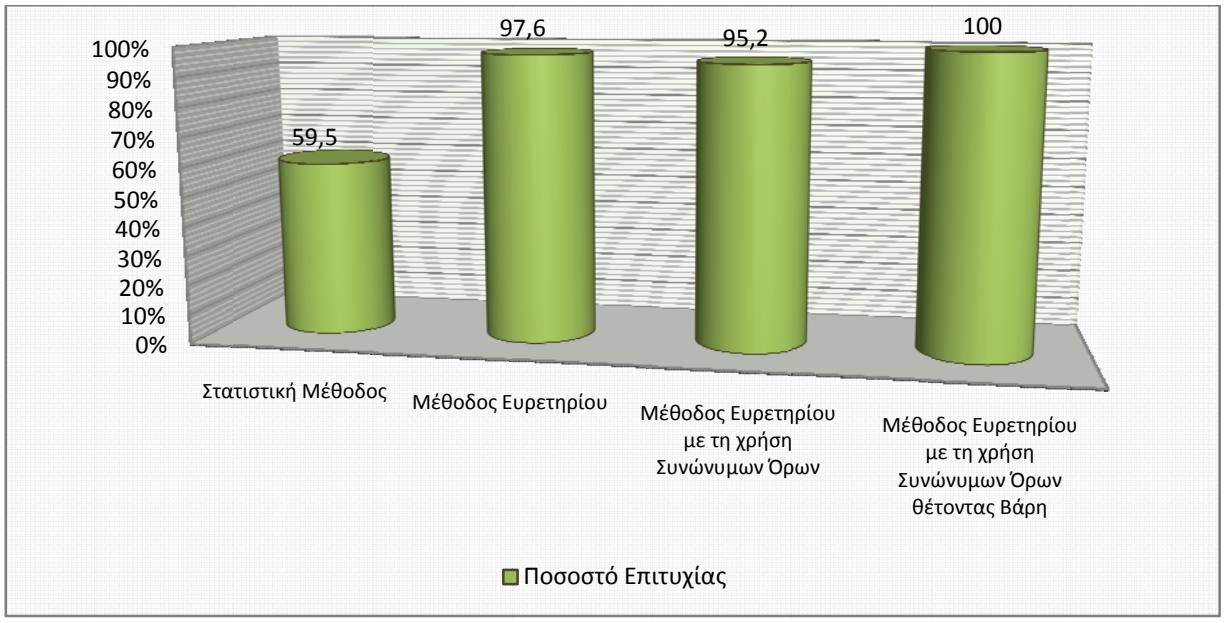
Οι τιμές x, y θα είναι τέτοιες ώστε να ικανοποιούν τα μεγαλύτερο δυνατό πλήθος των ερωτημάτων. Από τα είκοσι ερωτήματα που ελέγξαμε παρατηρήσαμε ότι αν η τιμή του x (συντελεστή του αποτελέσματος με τη μέθοδο του ευρετηρίου) είναι μικρότερη του 0.5 τότε σε όλες τις περιπτώσεις εντοπίζεται και επιστρέφεται 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.

7.3 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 ελέγχτηκαν 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 16 - 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 απορριών των εκπαιδευομένων με μια αυτοματοποιημένη διαδικασία ανεξάρτητα από τις θεματικές ενότητες που περιλαμβάνει.

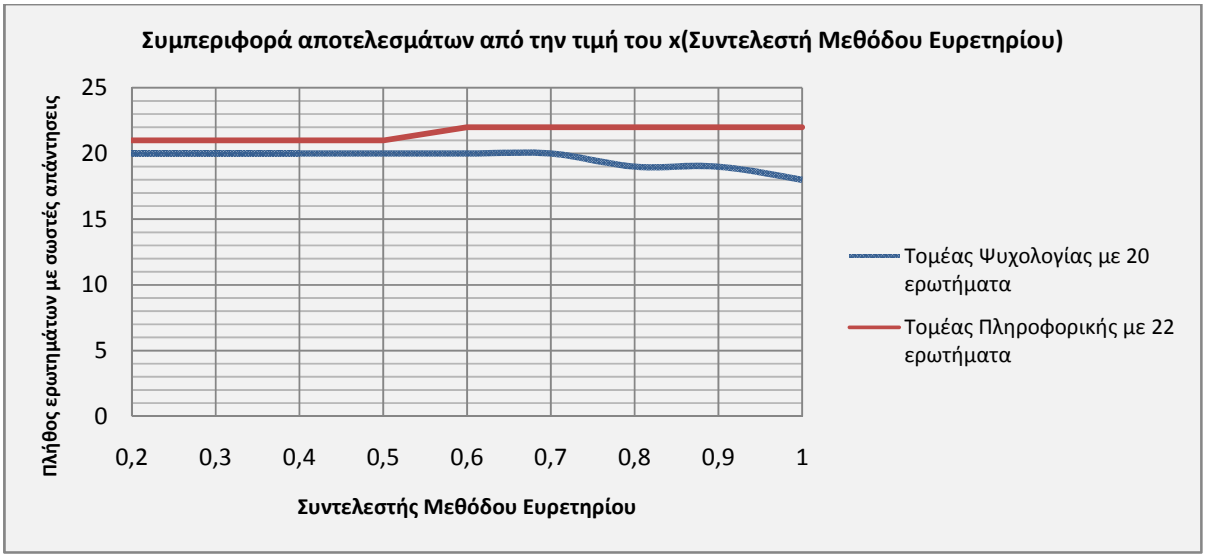
Πιο ξεκάθαρα παρουσιάζονται τα αποτελέσματα στη τρίτη μέθοδο του ευρετηρίου με τη χρήση συνώνυμων όρων, καθώς το συνολικό αποτέλεσμα είναι ικανοποιητικό αλλά υπάρχει ένα μικρό ποσοστό αποτυχίας και στα δύο θεματικά αντικείμενα. Και στις δύο διαφορετικές πηγές γνώσεων σε μία ερώτηση η κατάλληλη απάντηση 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 αντικατασταθούνε 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 ελέγχτηκαν σχετικά με το πόσο οι απαντήσεις που επιλέχτηκαν 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 w1 την βαθμολογία από τη μέθοδο του ευρετηρίου και w2 τη βαθμολογία από τη μέθοδο του ευρετηρίου με τη χρήση μόνο συνώνυμων όρων, τότε η τελική βαθμολογία θα προκύψει ως:

$$x*w1 + y*w2 \quad \text{όπου } x+y = 1$$

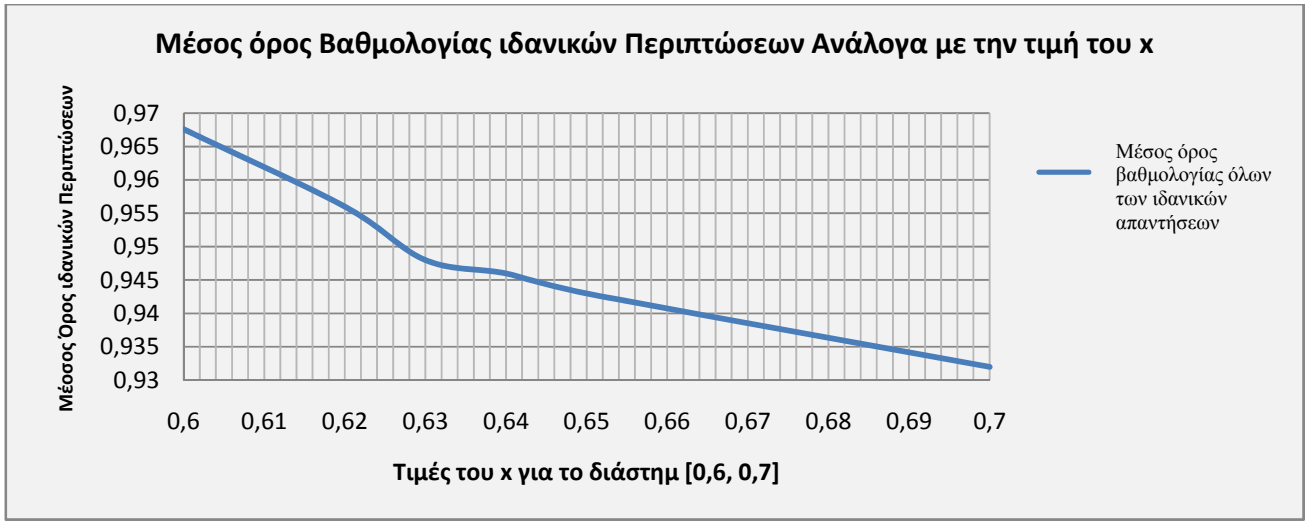
Από την αξιολόγηση στον κλάδο της πληροφορικής προέκυψε ότι η τιμή του x για να ικανοποιεί όλα το πλήθος των ερωτήσεων πρέπει να κυμαίνεται στο διάστημα [0.6, 1.0], ποιο συγκεκριμένα αν ανήκει στο διάστημα [0.6, 0.8) τότε εντοπίζεται σε όλες τις ερωτήσεις η ιδανική απάντηση αλλά σε μία ερώτηση δεν είναι 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 17 - 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 17 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 18 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 18 - 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 εξασφαλίσαμε 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.

8 Επίλογος – Συμπεράσματα – Μελλοντική Δουλειά

8.1 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 εκπαιδευομένους 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 ερωτή-ματος 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 απορριών 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.

8.2 Μελλοντική Δουλειά

8.2.1 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 απορριών. 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.

8.2.2 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 ύπηρχαν 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.

8.2.3 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 δυνατότηταστο 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 προτινόμεων 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.

8.2.4 Παροχή Εξατομικευμένης Πρόσβασης σε Υπηρησίες 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 χρήση 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.

Παράρτημα Α

Ερωτήσεις που ελέγχτηκαν στη βάση της πληροφορικής

1. What is import?
2. How to use Multiple Directories?
3. Why my signal handler doesn't work?
4. How can I program using threads?
5. Can I send email's with Python?
6. How to use conditional statement?
7. How to catch error type exception?
8. How to install python?
9. How to install python in UNIX or Linux?
10. How do I generate random numbers in Python?
11. Can I have internet in Python?
12. How to work with arrays in python?
13. How to use lists?
14. *How to use time* function in *python*?
15. What operation can I use?
16. How to use dictionary in python?
17. What Binary operation can I use?
18. How to define function?
19. How to use break in my function?
20. How to write text to a file?
21. How to read data from user?
22. How to print something?

Παράρτημα Β

Ερωτήσεις που ελέγχτηκαν στη βάση της Ψυχολογίας

1. What is Perception?
2. What is mental ill?
3. How to classify mental illness?
4. How to sort mental illness?
5. What is Psychotherapy?
6. Who needs treatment?
7. What is sleepwalking?
8. How to Recognize Cyclothymic Disorder?
9. How to Recognize Agoraphobia Disorder?
10. What is Parapsychology?

11. What Are the Signs and Symptoms of Dementia?
12. How to Recognize Symptoms of and Treat Cyclical Depression?
13. How to Recognize Symptoms of Bipolar Depression?
14. How to Diagnose Bipolar in a Child?
15. How Does Eating Disorder Affect Families?
16. How to Notice the Signs of Sleep Disorder?
17. What Are the Signs and Symptoms of Motor Development in Infants?
18. What is the Humanistic Theory?
19. Why do I Sometimes Remember My Dreams?

Αναφορές

1. Aono M. and Kobayashi M., *Vector space models for search and cluster,* In M. Berry (ed.), A Comprehensive Survey of Text Mining. Springer, to be published 2002.
2. Athanasoula-Reppa A., "*The role of tutor-mentor in Open and Distant Education*", Theory and Practice, Athens, Propompos, 2006.
3. Bello Juan Jose et al "*JColibri: An object oriented Framework for building CBR Systems*" book chapter in Advances in Case-Based-Reasoning, SpringerLink, ISBN-978-3-5400-22882-0, November 2004.
4. Bramer, M., "*Expert systems: where are we and where are we going?*", *EXPERT SYSTEMS: Principles and case studies*, 2nd Ed, Chapman and Hall Ltd, pp. 31-53, 1989.
5. Brüninghaus, S. and Ashley, K. D., "Reasoning with Textual Cases," in *Proceedings of the Sixth International Conference on Case-Based Reasoning (ICCBR 2005); LNCS 3620*, H. Muñoz-Avila and F. Ricci, Eds. Berlin: Springer, 2005.
6. C.D. Manning, P. Raghavan, and H. Schutze. An Introduction to Information Retrieval (draft version). Cambridge University Press, 2007.
7. Cunningham, C., Weber, R., Proctor, J. M., Fowler, C., and Murphy, M., "Investigating Graphs in Textual Case-Based Reasoning," in *Advances in Case-Based Reasoning: 7th European Conference (ECCBR 2004); LNAI 3155*, P. Funk and P. A. González Calero, Eds. Berlin: Springer, 2004, pp. 573–586.
8. David A. Grossman and Ophir Frieder, "*Information Retrieval: Algorithms and Heuristics*", Springer, 2nd edition, 2004.
9. Durkin, J., *Expert systems: Design and development*, New York: Macmillan, 1994
10. Ellis EA, Bentrup G, Schoeneberger MM, "Computer-based tools for decision support in agroforestry: current state and future needs", Agrofor Syst 61:401–421, 2004
11. Eyke Hullermeier, "*Case based approximate reasoning*", Springer, pp 30-32, 2007.
12. Hamburg, D., Lindecke, C. and tenThij, H, *Social aspects of e-Learning and blending Learning methods*, 4th European Conference E-Commerce, Bucharest , 2003.
13. Hatcher E., Gospodnetic O., *Lucene in Action (In Action series)*, Manning Publications, Greenwich, CT, USA, 2004.
14. J. A. Recio, B. Dıaz-Agudo, M. G. Martin, and N. Wiratunga, "*Extending jCOLIBRI for Textual CBR*", Springer-Verlag, 2005.
15. J. A. Recio-García, B. Díaz-Agudo, and P. A. González-Calero, "*Textual CBR in jCOLIBRI: From retrieval to reuse*", Proceedings of the ICCBR 2007 Workshop on Textual Case-Based Reasoning: Beyond Retrieval, pages 217–226, 2007.
16. J. J. Bello-Tomas, P. A. Gonzalez-Calero, and B. Dıaz-Agudo, "*jCOLIBRI: An object-oriented framework for building cbr systems. In Proceedings of Advances in Case-Based Reasoning*", 7th European Conference, ECCBR-04, *Madrid, Spain*, pages 32–46, 2004.

17. J.Ramos, “Using TF-IDF to Determine Word Relevance in Document Queries”, Instructional Conference on Machine Learning Web Site and Instructions, 2003.

18. K. M. Gupta and D. W. Aha, *Towards acquiring case indexing taxonomies from text*, in Proceedings of the 17th Int. FLAIRS Conference, pages 307–315, Miami Beach, FL, 2004.

19. Kolodner, J., *Case-Based Reasoning*. San Francisco CA: Morgan Kaufmann, 1993.

20. Lamontagne, L. and Lapalme, G., "Textual Reuse for Email Response," in *Advances in Case-Based Reasoning: 7th European Conference, ECCBR 2004; LNAI 3155*, P. Funk and P. A.González Calero, Eds. Berlin: Springer, 2004, pp. 242–256.

21. Lenz, M. and Ashley, K. D., "Textual Case-Based Reasoning: Papers from the AAAI-98 Workshop (Technical Report WS-98-12)." Menlo Park CA: AAAI Press, 1998.

22. Lenz M., *Defining knowledge layers for textual case-based reasoning*, In Proceedings of the 4th European Workshop on Advances in Case-Based Reasoning, EWCBR-98, pages 298–309, Springer-Verlag, 1998.

23. Lenz, M., Hübner, A., and Kunze, M., "Textual CBR," in *Case-Based Reasoning Technology - From foundations to Applications ,* Springer Verlag, pp. 115–137, 1998.

24. Lovins J.B., "Development of a stemming algorithm", Mechanical Translation and Computational Linguistics 11, 22-31, 1968.

25. Marchioro, T. and Landau, R., "*Web-based Education in Computational Science and Engineering*", Computational Science & Engineering, IEEE, pp. 19-26, 1997.

26. Mills, R. & Tait, A. "*Supporting the learner in open and Distance Learning*", London: Pitman Publishing, 1996.

27. M.Kantrowitz, B.Mohit, V.Mittal, “Stemming and its effects on TF-IDF Ranking”, Annual ACM Conference on Research and Development in Information Retrieval, 2000.

28. M. Lenz, "*Defining knowledge layers for textual case-based reasoning*", in Smyth and Cunningham pages 298–309, 1998.

29. Paleologou, A.-M. and Kerkiri, T., "*Telementoring: an alternative capacity of LMSs*", 5th International Conference on Open and Distance Learning, Open and Distance Education for Global Collaboration & Educational Development, 2009.

30. Partridge D., Hussain K., "*Knowledge based information systems*", McGraw Hill, New York, 1995.

31. Porter, M., "An Algorithm for Suffix Stripping, *Program*", vol. 14, pp. 130–137, 1980.

32. F P. Willett, "*The Porter stemming algorithm: then and now*," Program: electronic library and information systems, vol. 40, no. 219-223, 2006.

33. Riesbeck, C. K. and Schank, R. C., "*Inside Case-Based Reasoning*". Mahwah NJ: Lawrence Erlbaum Associates, 1989.

34. Rosina Weber, K.A., Br¨uninghaus, S., "*Textual case-based reasoning. The Knowledge Engineering Review*", 20 255–260, 2006.

35. Weber, D. W. Aha, N. Sandhu, and H. Munoz-Avila, "*A textual case-based reasoning framework for knowledge management applications*", in Proceedings of the 9th German Workshop on Case-Based Reasoning. Shaker Verlag., 2001.

36. Salton, G. and Buckley, C., "*Term-Weighting Approaches in Automatic Retrieval*, *Information Processing & Management*", vol. 24, pp. 513–523, 1988.

37. S. Bruninghaus and K. D. Ashley. "The role of information extraction for textual CBR", pages 74–89. Springer-Verlag, 2001.

38. P. Willett, "*The Porter stemming algorithm: then and now*," Program: electronic library and information systems, vol. 40, no. 219-223, 2006.

39. Shaw, W. M., Jr., Burgin, R., and Howell, P., "*Performance Standards and Evaluations in IR Test Collections: Cluster-Based Retrieval Models*", Information Processing & Management, vol. 33, pp. 1–14, 1997.

40. Simic, D., Kurbalija, V., Budimac, Z., An Application of Case-Based Reasoning in Multidimensional Database Architecture. In: Proc. Of 5th DaWaK, LNCS, Vol. 2737. Springer-Verlag 66-75,2003

41. S. J. Delany and D. Bridge, *Feature-based and feature-free textual CBR: A comparison in spam filtering,* in Procs. of the 17th Irish Conference on Artificial Intelligence and Cognitive Science, pages 244–253, Belfast, Northern Ireland, 2006.

42. Stevenson, K. & Sander, P., How to do Open University students expect to be Taught at Tutorials. *Open Learning* v.13 (2), p.p 42-46, 1998.

43. Tait, A., Planning Student Support for Open and Distance Learning, *Open Learning* v.15, 2000.

44. Vivitsou, M., Lambropoulos, N., Konetas, D., Paraskevas, M. and Grigoropoulos, E., *The Project Method e-course: the use of tools towards the evolution of the Greek teachers' online community*, International Journal of Continuing Engineering education and lifelong learning, Special Issue on Reference Models for Forming Organisational or Collaborative Pedagogical Best Practices, 2007.

45. Vlaxavas I, Kefalas P, Vasileiadis N, Kokkoras F, Sakelariou I, "Artificial intelligence", Giourdas Publications, 2006.

46. Vlaxavas I, Kefalas P, Vasileiadis N, Refanidis I., Kokkoras F, Sakelariou I, "Artificial intelligence", Gartagani Publications, 2002.

47. Watson I., *Applying case-based reasoning: techniques for enterprise systems*, morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.

48. W.B. Frakes and Baeza-Yates (eds). Information Retrieval: Data Structures and Algorithms. Prentice Hall, Englewood Cli_s, NJ, 1992.

49. Weber R., Aha D.W., Sandhu N., and Munoz-Avila H., *A textual case-based reasoning framework for knowledge management applications*, in Proceedings of the 9th German Workshop on Case-Based Reasoning, 2001.

50. Weber, R., Waldstein, I., Deshpande A., and Proctor J. M., "Integrated Approach to Detect Inconspicuous Content," in *Professional Knowledge Management*, K.-D. Althoff, A. Dengel, R. Bergmann, M. Nick, and T. Roth-Berghofer, Eds. Berlin: Springer (in press), 2005.

51. William B. Frakes and Christopher J. Fox, " *Strength and similarity of affix removal stemming algorithms*", 2003.

52. Wilson, D. C. and Bradshaw, S., "CBR Textuality," *Expert Update*, vol. 3, pp. 28–37, 2000.

53. Wiratunga, N., Koychev, I., and Massie, S., "Feature Selection and Generalisation for Retrieval of Textual Cases," in *Advances in Case-Based Reasoning: 7th European Conference LNAI 3155*, P. Funk and P. A. González Calero, Eds. Berlin: Springer, pp. 806–820, 2004.