



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

---

Μέθοδος Αναζήτησης Σφαλμάτων  
Κώδικα με Χρήση Αναφορών  
Σφαλμάτων

---

Συγγραφέας:  
Μάριος Σταύρος ΓΡΗΓΟΡΙΟΥ

Επιβλέπων:  
Κωνσταντίνος ΚΟΝΤΟΓΙΑΝΝΗΣ

Αθήνα, Ιούνιος 2016



# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

---

## Μέθοδος Αναζήτησης Σφαλμάτων Κώδικα με Χρήση Αναφορών Σφαλμάτων

---

Συγγραφέας:

Μάριος Σταύρος ΓΡΗΓΟΡΙΟΥ

Επιβλέπων:

Κωνσταντίνος ΚΟΝΤΟΓΙΑΝΝΗΣ

*Εγκρίθηκε από την τριμελή επιτροπή την 23<sup>η</sup> Ιουνίου 2016*

---

Κωνσταντίνος Κοντογιάννης  
Αν. Καθηγητής ΕΜΠ

---

Ιωάννης Βασιλείου  
Καθηγητής ΕΜΠ

---

Νικόλαος Παπασπύρου  
Αν. Καθηγητής ΕΜΠ

Αθήνα, Ιούνιος 2016



Υπογραφή :

---

Ημερομηνία :

---

Μάριος Σταύρος ΓΡΗΓΟΡΙΟΥ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών

Copyright©, 2015, Μάριος Σταύρος ΓΡΗΓΟΡΙΟΥ

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.



*‘ΕΙ ΤΕ ΦΕΡΟΝ ΣΕ ΦΕΡΕΙ ΦΕΡΕ ΚΑΙ ΦΕΡΟΥ. ΕΙ Δ’ ΑΓΑΝΑΚΤΕΙΣ ΚΑΙ ΣΕΑΥΤΟΝ  
ΛΥΠΕΙΣ ΚΑΙ ΤΟ ΦΕΡΟΝ ΣΕ ΦΕΡΕΙ’*

Αρχαίο Γνωμικό





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

## Περίληψη

Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Διπλωματική Εργασία

### Μέθοδος Αναζήτησης Σφαλμάτων Κώδικα με Χρήση Αναφορών Σφαλμάτων

του Μάριου Σταύρου ΓΡΗΓΟΡΙΟΥ

Στις περασμένες δεκαετίες η παραγωγή Λογισμικού έχει φθάσει σε δυσθεώρητα ύψη. Νέα προϊόντα γίνονται διαθέσιμα κάθε μήνα, κάποια από αυτά μικρότερα αποτελούμενα μόνο από μερικές χιλιάδες γραμμές πηγαίου κώδικα, άλλα πολύ μεγαλύτερα με εκατομμύρια γραμμές κώδικα. Με δεδομένη αυτήν την καλπάζουσα αύξηση στην παραγωγή λογισμικού, είναι αναμενόμενο ότι η συντήρηση λογισμικού έχει αποκτήσει δικαίως μερίδιο της δόξας αυτής της βιομηχανίας για δύο πολύ βασικούς λόγους. Πρώτον τα παλιά συστήματα πρέπει να συντηρούνται γιατί το κόστος της αντικατάστασής του είναι συχνά πολύ μεγαλύτερο από το κόστος συντήρησης ενώ επιπλέον και τα νέα συστήματα πολύ γρήγορα (συχνά από την αρχή) αντιμετωπίζουν προβλήματα.

Ο τομέας της Τεχνολογίας Λογισμικού, αν και ενεργός ερευνητικά ήδη από τα πρώτα βήματα της οργανωμένης παραγωγής κώδικα, δυστυχώς μέχρι τώρα δεν έχει καταφέρει να προσφέρει αυτοματοποιημένες διαδικασίες με στόχο την αυτόματη επιδιόρθωση των σφαλμάτων Λογισμικού. Σε αυτή τη διπλωματική εργασία διερευνάται η πιθανότητα διεκπεραίωσης αυτοματοποιημένου εντοπισμού σφαλμάτων μέσα από αναφορές σφάλματος χρησιμοποιώντας κάποιες τεχνικές information retrieval και αξιοποιώντας τη δομή των συστημάτων. Η δουλειά αυτή γίνεται με την ελπίδα να βρεθεί ένας τρόπος για να διευκολυνθεί η δουλειά των προγραμματιστών στον αγώνα που δίνουν για πιο ασφαλή και ποιοτικότερα συστήματα.

Στην προσπάθεια για την επίτευξη αυτού του στόχου, δημιουργήθηκαν αρκετά εργαλεία τα οποία βοήθησαν στην συγκομιδή δεδομένων από πλατφόρμες διαφόρων οργανισμών, στην επεξεργασία των αναφορών, και στην τελική εύρεση και κατάταξή τους με βάση την 'απόστασή' τους από την πραγματική λύση.

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ :** Εντοπισμός Σφαλμάτων, Εξόρυξη Γνώσης, Τεχνολογία Λογισμικού, Επεξεργασία Φυσικής Γλώσσας, Clustering



NATIONAL TECHNICAL UNIVERSITY OF ATHENS

## *Abstract*

School of Electrical & Computer Engineering

Department of Computer Science

Graduate Thesis

### **Method for Bug Localization Using Bug Reports**

by Marios Stavros GRIGORIOU

In the past decades software production has reached unparalleled heights. New products are being launched on a monthly basis, some small consisting of only a few KLOC, whereas others much more sizeable, up to several MLOC. With this increase in software production, software maintenance is as important as ever, both because old systems must be maintained due to oftentimes unbearable migration costs, and because new systems rarely are bug free.

Though the field of Software maintenance has been active since the very beginning of organized software development, unfortunately until now automated processes for software maintenance and specifically bug fixing have proven to be elusive. In this diploma thesis we are exploring the possibility of automatic bug localization via the usage of a number of different algorithmic techniques, such as NLP, Latent Semantic Analysis, Comprehension Driven Clustering (ACDC), hoping to find such an approach as to enable the, now swamped with bugs, software community to become more productive by greatly diminishing the time between bug identification and bug resolution, the need for platform specialists, and to get rid of the dawdling through irrelevant code segments.

In our effort to achieve the above-mentioned goal, we created tools with which to, acquire a substantial variety of input data from several Bugzilla repositories, use apache openNLP to analyze our data, and extract, through programming, relations between different bug Reports.

**KEYWORDS :** Bug Localization, Information Retrieval, Software Engineering, Natural Language Processing, Clustering



# Ευχαριστίες

Θα ήθελα πρωτίστως να εκφράσω την βαθύτατη ευγνωμοσύνη μου στον επιβλέποντα καθηγητή κ. Κώστα Κοντογιάννη που ήταν πάντα πρόθυμος να με βοηθήσει, για τις εξαιρετικές ακαδημαϊκές συζητήσεις που είχαμε δουλεύοντας πάνω στην διπλωματική μου εργασία, για την ακεραιότητα του χαρακτήρα του και για τις πολύ χρήσιμες συμβουλές του σε ακαδημαϊκά και μη θέματα. Ο ανοιχτός χαρακτήρας του και η άμεση πίστη που έδειξε στις προοπτικές μου, μου έδωσαν την ώθηση να μετατρέψω τις προοπτικές αυτές σε προσόντα. Θα ήθελα τέλος να τον ευχαριστήσω για κάθε φορά που με νουθετούσε όταν ξόδευα χρόνο ασχολούμενος με ανώφελες δραστηριότητες.

Εκτός από τον κ. Κοντογιάννη, θέλω να ευχαριστήσω τους συναδέλφους μου από το εργαστήριο, οι οποίοι μου μετάγγισαν την μοναδικά χρήσιμη νοοτροπία με την οποία αντιμετωπίζουν τις ακαδημαϊκές προκλήσεις. Περισσότερο από όλους τους συναδέλφους μου θέλω να ευχαριστήσω τον φίλο μου Γιώργο Καραχάλια, που σε μία στιγμή μέγιστης ανάγκης μου πρόσφερε μία χείρα βοήθειας που έμελλε να με φέρει στο SoftLab όπου όπως είπε αστεειυόμενος 'πάντα θα δίνεται βοήθεια σε αυτούς που την ζητάνε'.

Θέλω να ευχαριστήσω και όλους μου τους φίλους που ευτυχώς είναι πάρα πολλοί για να αναφερθώ σε καθέναν ξεχωριστά, είμαι ευγνώμων για κάθε έναν από εσάς και σας ευχαριστώ που αντέχετε την κατά καιρούς ανυπόφορη συμπεριφορά μου.

Τελευταίους αλλά πιο σημαντικούς ευχαριστώ την οικογένειά μου και ειδικά τους γονείς μου Μάρκο και Αικατερίνη, που με έφεραν σε αυτόν τον κόσμο και που με έχουν φροντίσει και αναθρέψει με ιδανικά, ελευθερίες και ανέσεις που μακάρι όλα τα παιδιά να απολάμβαναν, που ποτέ δεν έφυγαν από το πλάι μου και πάντα με υποστήριζαν ακόμα κι όταν προσπαθούσα να τους απογοητεύσω, ένα ευχαριστώ δεν είναι αρκετό.



# Περιεχόμενα

|  |          |
|--|----------|
| Περίληψη   | vii      |
| Abstract   | ix       |
| Ευχαριστίες  | xi       |
| Περιεχόμενα  | xii      |
| <b>1 Εισαγωγή</b>  | <b>1</b> |
| 1.1 Λογισμικό και Σφάλματα Λογισμικού . . . . .              | 1        |
| 1.2 Στόχος . . . . .   | 4        |
| 1.3 Οργάνωση Κειμένου . . . . .                              | 5        |
| <b>2 Προαπαιτούμενες Γνώσεις</b>                             | <b>7</b> |
| 2.1 Software Clustering . . . . .                            | 7        |
| 2.1.1 ACDC . . . . .   | 9        |
| 2.1.1.1 Χρόνος Εκτέλεσης . . . . .                           | 12       |
| 2.2 Αναπαράσταση Αρχιτεκτονικής . . . . .                    | 14       |
| 2.2.1 Το πρότυπο Meta Object Facility (MOF) . . . . .        | 15       |
| 2.2.1.1 Τετραεπίπεδη Αρχιτεκτονική Μεταδεδομένων . . . . .   | 15       |
| 2.2.1.2 Δομή και Περιγραφή του MOF . . . . .                 | 17       |
| 2.2.2 Ανάλυση Πηγαίου Κώδικα . . . . .                       | 20       |
| 2.3 MongoDB . . . . .  | 27       |
| 2.3.1 JSON . . . . .   | 29       |
| 2.4 Συνοπτική Αναπαράσταση Κειμένων . . . . .                | 29       |
| 2.4.1 Διανύσματα TF-IDF . . . . .                            | 29       |
| 2.4.1.1 Ορισμός Συχνότητας Όρου . . . . .                    | 29       |
| 2.4.1.2 Αντίστροφη Συχνότητα Κειμένου . . . . .              | 29       |
| 2.4.1.3 Η τιμή TF-IDF . . . . .                              | 30       |
| 2.5 Meta Object Facility και EMF . . . . .                   | 30       |
| 2.6 Natural Language Processing . . . . .                    | 31       |
| 2.7 Αλγόριθμοι Εξόρυξης Δεδομένων - LSI, PLSI, LDA . . . . . | 31       |
| 2.7.1 Ομαδοποίηση Clustering . . . . .                       | 32       |
| 2.7.2 Κατηγοριοποίηση . . . . .                              | 32       |
| 2.7.2.1 LSI . . . . .  | 32       |
| 2.7.2.2 PLSA . . . . .                                       | 33       |

|          |  |           |
|----------|--|-----------|
| 2.7.2.3  | LDA . . . . .  | 34        |
| <b>3</b> | <b>Προτεινόμενη Μεθοδολογία και Υλοποιημένες Λειτουργίες</b> | <b>35</b> |
| 3.1      | Εργαλεία Συλλογής και Προεπεξεργασίας Δεδομένων . . . . .    | 35        |
| 3.1.1    | Bugzilla Extraction Tool . . . . .                           | 35        |
| 3.1.1.1  | XMLRPC API Client . . . . .                                  | 35        |
| 3.1.1.2  | Φορτωτής MongoDB . . . . .                                   | 36        |
| 3.1.2    | Εργαλεία Προεπεξεργασίας Αναφορών Σφάλματος . . . . .        | 37        |
| 3.1.2.1  | Αναζητητής Οντοτήτων Κώδικα . . . . .                        | 38        |
| 3.1.2.2  | NLP Tool . . . . .   | 38        |
| 3.1.3    | Μικρά Εργαλεία . . . . .                                     | 39        |
| 3.1.3.1  | Εγγραφή Source Code Entity από RSF . . . . .                 | 39        |
| 3.1.3.2  | MongoDB Query Client . . . . .                               | 39        |
| 3.1.3.3  | Source Code Entity Expander . . . . .                        | 40        |
| 3.1.3.4  | Λεξικό Όρων . . . . .  | 40        |
| 3.2      | Μεθοδολογία . . . . .  | 41        |
| 3.2.1    | Τελική Προεπεξεργασία . . . . .                              | 41        |
| 3.2.2    | Αρχικός Χώρος Αναζήτησης . . . . .                           | 41        |
| 3.2.3    | Διεύρυνση Χώρου Αναζήτησης . . . . .                         | 42        |
| 3.2.4    | Κατάταξη Οντοτήτων του Χώρου Αναζήτησης . . . . .            | 42        |
| 3.2.5    | Προκαταρκτικά Αποτελέσματα . . . . .                         | 43        |
| <b>4</b> | <b>Μοντέλο Δεδομένων</b>                                     | <b>45</b> |
| 4.1      | Αρχική Δομή Δεδομένων . . . . .                              | 45        |
| 4.1.1    | EMF-Model . . . . .  | 46        |
| 4.2      | Μετρική Ποιότητας Αναφορών . . . . .                         | 51        |
| <b>5</b> | <b>Εργαλεία Εξόρυξης Γνώσης</b>                              | <b>53</b> |
| 5.1      | Ομοιότητα Αναφορών Σφαλμάτων Λογισμικού . . . . .            | 53        |
| 5.2      | LSA . . . . .  | 53        |
| 5.2.1    | Επισκόπηση Τμημάτων της Τεχνικής . . . . .                   | 54        |
| 5.2.1.1  | Πίνακας Εμφάνισης . . . . .                                  | 54        |
| 5.2.1.2  | Υποβάθμιση Τάξης . . . . .                                   | 54        |
| 5.2.1.3  | Παραγωγή . . . . .   | 55        |
| 5.2.1.4  | Εφαρμογές . . . . .  | 58        |
| 5.2.2    | Υλοποίηση . . . . .  | 58        |
| 5.2.2.1  | Επιλογή Γλώσσας . . . . .                                    | 58        |
| 5.2.2.2  | Βασικές Βιβλιοθήκες . . . . .                                | 59        |
| 5.2.2.3  | Δομή του Εργαλείου LSA . . . . .                             | 59        |
| 5.3      | Natural Language Processing . . . . .                        | 61        |
| 5.3.1    | Ιστορική Αναδρομή του NLP . . . . .                          | 61        |
| 5.3.2    | Machine Learning και NLP . . . . .                           | 62        |
| 5.3.2.1  | Machine Learning . . . . .                                   | 62        |
| 5.3.3    | Βασικές Εφαρμογές της NLP . . . . .                          | 63        |
| 5.3.4    | Στατιστική Επεξεργασία Φυσικής Γλώσσας . . . . .             | 65        |
| <b>6</b> | <b>Αλγόριθμος Εντοπισμού Σφαλμάτων και Αποτελέσματα</b>      | <b>67</b> |
| 6.1      | Πειραματική Διαδικασία . . . . .                             | 67        |



|           |  |            |
|-----------|--|------------|
| 6.1.1     | Χαρακτηριστικά Υπολογιστικού Περιβάλλοντος . . . . .         | 67         |
| 6.1.2     | Περιβάλλον Ανάπτυξης . . . . .                               | 67         |
| 6.1.3     | Επιλογή Αναφορών Σφάλματος του Gold Standard . . . . .       | 68         |
| 6.1.4     | Δομή του Πειράματος . . . . .                                | 69         |
| 6.2       | Βαθμονόμηση και Κατάταξη Αποτελεσμάτων . . . . .             | 72         |
| 6.2.1     | Διαδικασία Βαθμονόμησης . . . . .                            | 73         |
| 6.2.2     | Παρουσίαση Αποτελεσμάτων . . . . .                           | 74         |
| 6.3       | Αποτελέσματα . . . . .                                       | 75         |
| 6.3.1     | Αναλυτικά Αποτελέσματα . . . . .                             | 76         |
| 6.3.2     | Μορφοποίηση Εξόδου . . . . .                                 | 81         |
| <b>7</b>  | <b>Συμπεράσματα και Μελλοντικές Ερευνητικές Κατευθύνσεις</b> | <b>105</b> |
| 7.1       | Συμπεράσματα . . . . .                                       | 105        |
| 7.2       | Σχετικές Εργασίες . . . . .                                  | 106        |
| 7.3       | Συνεισφορά . . . . .   | 107        |
| 7.4       | Απειλές ως προς την Ορθότητα . . . . .                       | 107        |
| 7.5       | Κατευθύνσεις για Μελλοντική Έρευνα . . . . .                 | 107        |
| <b>A'</b> | <b>Amarok Sample Results</b>                                 | <b>109</b> |
| <b>B'</b> | <b>Konqueror Sample Results</b>                              | <b>143</b> |
| <b>Γ'</b> | <b>Kopete Sample Results</b>                                 | <b>167</b> |
| <b>Δ'</b> | <b>K3b Sample Results</b>                                    | <b>187</b> |
|           | <b>Βιβλιογραφία</b>  | <b>194</b> |



# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Λογισμικό και Σφάλματα Λογισμικού

Η διαδικασία δημιουργίας μεγάλης κλίμακας πληροφοριακών συστημάτων αποτελείται από έναν αριθμό βημάτων. Τα βήματα αυτά ενώ σαν επιμέρους διαδικασίες είναι σαφώς διαχωρισμένες, στα πλαίσια ενός πληροφοριακού συστήματος μπορεί να έχουν μη μηδενικές συσχετίσεις. Η δημιουργική διαδικασία ξεκινάει με τη συλλογή απαιτήσεων και ακολουθείται από τον δημιουργία της αρχιτεκτονικής και σχεδιασμό σε μεγάλο επίπεδο αφαίρεσης. Μετά από αυτά το σύστημα αρχίζει να υλοποιείται. Συχνά κατά την υλοποίηση τα αποτελέσματα των προαναφερθέντων βημάτων μπορεί να χρειαστεί να αλλάξουν πλήρως ή να τροποποιηθούν μερικώς. Μετά την ολοκλήρωση της υλοποίησης το σύστημα αρχίζει να χρησιμοποιείται και ως εκ τούτου να συντηρείται. Κατά τη διάρκεια αυτής της περιόδου θα αρχίσει να εκδηλώνει κατά διαστήματα απρόβλεπτη συμπεριφορά. Αυτή η συμπεριφορά μπορεί να συμπεριλάβει είτε την εκτέλεση λειτουργιών με διαφορετικό τρόπο από αυτόν που έχει προβλεφθεί κατά την εξαγωγή των απαιτήσεων, είτε την ολοκληρωτική διακοπή εκτέλεσης του συστήματος.

Το παραπάνω αποτελεί έναν αδρό ορισμό του "σφάλματος λογισμικού" το οποίο στη αγγλική γλώσσα ονομάζεται "bug". Ένα Bug είναι ένα λάθος στον κώδικα, που οδηγεί σε εσφαλμένη συμπεριφορά, λάθος κατάσταση δεδομένων ή διακοπή εκτέλεσης του συστήματος. Τα Bugs μπορούν να εμφανιστούν σε κάθε είδος προϊόντων πληροφορικής. Η αντιμετώπιση τους γνωστή στους προγραμματιστές ως debugging ή bug squashing είναι ένα από τα σημαντικότερα τμήματα της παραγωγής και συντήρησης λογισμικού.

Η συντήρηση λογισμικού είναι ίσως το πιο δαπανηρό μέρος της ζωής ενός συστήματος λογισμικού. Συγκεκριμένα όπως έχει αναφερθεί και στο [1] το κόστος αυτής της διαδικασίας αποτελεί ένα πολύ σημαντικό ποσοστό των συνολικών κεφαλαίων που διακινούνται στην

βιομηχανία λογισμικού καθώς ήδη εν έτει 2002 αναφέρεται ότι το ετήσιο κόστος της αποσφαλμάτωσης και συντήρησης λογισμικού κόστιζε στις Ηνωμένες Πολιτείες της Αμερικής 60 δισεκατομμύρια δολάρια ετησίως. Δεδομένης της ολοένα αυξανόμενης διείσδυσης των συστημάτων πληροφορικής σε κάθε έκφανση της ανθρώπινης δραστηριότητας είναι σχετικά βέβαιο ότι το κόστος της συντήρησης συστημάτων λογισμικού έχει αυξηθεί δραστικά μέσα στα τελευταία 14 χρόνια.

Για να μπορέσει ο αναγνώστης να κατανοήσει τα μεγέθη καλύτερα αξίζει αναφοράς πως ήδη το 1996 είχε υπολογισθεί πως ένα τμήμα 80% του συνολικού χρόνου ενασχόλησης με ένα πρόγραμμα αφιερώνεται στην αποσφαλμάτωση και συντήρησή του και όπως είναι φυσικό ένα ανάλογο ποσοστό των συνολικών επενδύσεων πάνε τελικά στην ίδια κατεύθυνση [2]. Είναι ακόμα σημαντικό να σημειωθεί ότι η εταιρείες ανάπτυξης λογισμικού ενσωματώνουν αυτά τα κόστη στο ολικό κόστος ενός νέου συστήματος καθώς είναι αναπόσπαστο κομμάτι κάθε προγραμματιστικού πακέτου.

Τα bugs εισάγονται στον κώδικα κυρίως κατά τη διάρκεια της παραγωγής/υλοποίησής του οπότε ένα σημαντικό πλήθος λειτουργικότητας εισάγεται πολύ γρήγορα και πολύ συχνά χωρίς να πιστοποιείται ότι η νέα λειτουργικότητα είναι συμβατή με τις ήδη υπάρχουσες. Ωστόσο οι συντριπτική πλειοψηφία αυτών των bugs επιδιορθώνεται πριν το σύστημα γίνει διαθέσιμο εμπορικά ή τεθεί σε λειτουργία. Οπότε τα bugs που απασχολούν την κοινότητα δεν είναι αυτά που προκύπτουν κατά την παραγωγή αλλά αυτά που προκύπτουν κατά την συντήρηση/επέκταση του συστήματος. Ως αποτέλεσμα οι μορφή αυτών των bugs δεν επιτρέπει τη χρήση σχολίων χρηστών ή περιπτώσεις πραγματικής χρήσης.

Κάθε σύστημα λογισμικού μετά την πρώτη λειτουργική έκδοση συνεχίζει να εισάγει νέες λειτουργίες και να βελτιώνει ήδη παρεχόμενες, ενώ ταυτοχρόνως φροντίζει για τη συνεχή και αδιάλειπτη παροχή υπηρεσιών, καθώς κάτι τέτοιο θα εξασφαλίσει μεγαλύτερη αποδοχή στην αγορά και κατά συνέπεια μεγαλύτερα κέρδη για τον εκδότη ή χρήστη του συστήματος. Ωστόσο η διαδικασία αυτή δεν είναι τόσο δομημένη και αυστηρή όσο η διαδικασία της αρχικής υλοποίησης. Συχνά ανατίθεται σε μικρές ανεπαρκώς οργανωμένες ομάδες, χωρίς μία κεντρική επίβλεψη, που οδηγεί αναπόφευκτα στην καταστρατήγηση του αρχιτεκτονικού σχεδίου καθώς και των λοιπών σχεδιαστικών αποφάσεων που είχαν ληφθεί κατά το σχεδιασμό του συστήματος. Αυτή η ανοργάνωτη υφή του debugging μπορεί να οδηγήσει σε ακόμα μεγαλύτερα κόστη για των ιδιοκτήτη του συστήματος και μεγαλύτερο κόπο για τους προγραμματιστές. Στο τέλος της ζωής του συστήματος ο κώδικας στη γενική περίπτωση θα είναι εντελώς ασυγχρόνιστος με την αρχική αρχιτεκτονική και θα βρίσκεται σε μία κατάσταση όπου οι διαθέσιμες επιλογές είναι είτε η εκ βάθρων αναδιάρθρωση του κώδικα είτε ο παροπλισμός του συστήματος, οι οποίες έχουν εξαιρετικά υψηλό κόστος.

Η διαδικασία καθεαυτή του debugging είναι μία προγραμματιστική εργασία που απαιτεί από τον προγραμματιστή να αναπαράγει την ανώμαλη εκτέλεση του συστήματος, να αντιληφθεί

τί μπορεί να προκαλεί αυτή τη συμπεριφορά και να εντοπίσει μέσα στον κώδικα το σημείο που την εισάγει. Μετά τον εντοπισμού ο προγραμματιστής δοκιμάζει διαφορετικές προσεγγίσεις επίλυσης του προβλήματος, προσπαθώντας να ελαχιστοποιήσει κατά το δυνατόν τις επεμβάσεις που τελικά θα κάνει καθώς οι αλλαγές κάνουν τον κώδικα ολοένα και πιο δυσνόητο.

Έχει αναφερθεί ήδη πως το άμεσο κόστος του debugging είναι εξαιρετικά υψηλό. Ωστόσο δεν έχει γίνει καμία αναφορά στα έμμεσα κόστη που έχει για την επιχειρηματικά ενεργή κοινότητα προγραμματιστών. Όπως είπαμε είναι μία εξαιρετικά απαιτητική διαδικασία που απαιτεί ο προγραμματιστής να είναι γνώστης του συστήματος και να έχει μία υψηλού επιπέδου γνώση της γλώσσας προγραμματισμού στην οποία αυτό είναι γραμμένο. Από αυτά φαίνεται ότι ο ιδανικός προγραμματιστής για να ασχοληθεί με το debugging είναι εξαιρετικά χρήσιμος γενικότερα στην εταιρεία καθώς είναι ένας πολύ ικανός εργαζόμενος. Ωστόσο αυτός ο τύπος προγραμματιστή που κανονικά θα μπορούσε να χρησιμοποιηθεί για να εισάγει επιπλέον ανταγωνιστικά γνώρισμα σε κάποιο σύστημα είναι εξαναγκασμένος να αποστασιοποιηθεί από την παραγωγή και να αφιερώνει μεγάλο μέρος του χρόνου του στην συντήρηση.

Κατά καιρούς έχει προταθεί μία πλειάδα εργαλείων που αποσκοπούν στην απλοποίηση του έργου του debugger (προγραμματιστή που δουλεύει στην αποσφαλμάτωση κώδικα), τα οποία καλύπτουν κάθε εξελικτικό στάδιο της ύπαρξης ενός συστήματος λογισμικού, από τα πρώτα στάδια της παραγωγής μέχρι την διάθεση και τη συντήρηση. Κάποια από αυτά τα συστήματα είναι στατικά, και κάποια δυναμικά. Το ενδιαφέρον της κοινότητας για τέτοια εργαλεία αλλά και τεχνικές που υπόσχονται να διευκολύνουν το debugging μετά τη διάθεση ενός προϊόντος είναι αυξανόμενο, ειδικά όσον αφορά συστήματα τα οποία αξιοποιούν αναφορές σφάλματος χρηστών από κοινωνικά δίκτυα που χρησιμοποιούνται αποκλειστικά για την αναφορά σφαλμάτων. Κάποιες προσπάθειες σε αυτόν τον τομέα έχουν ήδη γίνει [3], [4], [5], [6], [7]. Δυστυχώς όμως τα προτεινόμενα εργαλεία υστερούν σε κάποια σημεία τα οποία είναι χαρακτηριστικά της μοντελοποίησης του προβλήματος αλλά και τις προτεινόμενης, κατά περίπτωση, λύσης. Ως εκ τούτου δεν υπάρχει ακόμα διαθέσιμο κάποιο εργαλείο το οποίο να αξιοποιεί στο μέγιστο δυνατό όλα τα διαθέσιμα δεδομένα για την αντιμετώπιση του προβλήματος.

Τέλος επειδή αυτή η ερευνητική περιοχή παρουσιάζει μεγάλο ενδιαφέρον για του επιστήμονες της τεχνολογίας λογισμικού, έχουν ήδη υπάρξει προτάσεις όχι απλά για τον αυτοματοποιημένο εντοπισμό σφαλμάτων αλλά ακόμα και για την αυτοματοποιημένη επίλυση αυτών [4], [8], [9], [10]. Αυτές η προτάσεις έχουν οδηγήσει με τη σειρά τους στην εκπόνηση ενός σημαντικού πλήθους δημοσιεύσεων που ασχολούνται με τα ποσοτικά χαρακτηριστικά αυτών των προσεγγίσεων σε πραγματικά συστήματα λογισμικού, με στόχο να εντοπιστούν τα δυνατά και τα αδύνατα σημεία τους[11].

## 1.2 Στόχος

Ο αντικειμενικός στόχος αυτής της διπλωματικής εργασίας είναι να δοκιμάσει μία ελαφρώς διαφοροποιημένη προσέγγιση στο πρόβλημα του εντοπισμού σφαλμάτων λογισμικού. Συγκεκριμένα επιδιώκει να ανακαλύψει εάν υπάρχει κάποιος υπονοούμενος συσχετισμός ανάμεσα στα σχόλια μίας αναφοράς σφάλματος (Bug Report) και του κώδικα ο οποίος ευθύνεται για το σφάλμα που την προκάλεσε, και να ελέγξει αν είναι δυνατή η δημιουργία μίας αυτοματοποιημένης διαδικασίας για την εκμετάλλευση αυτού.

Στα πλαίσια αυτής της εργασίας, υλοποιήθηκε μία πλατφόρμα η οποία δύναται να συγκεντρώσει όλες τις ελεύθερα διαθέσιμες αναφορές σφάλματος από οποιαδήποτε εγκατάσταση της υπηρεσίας δημοσίευσης αναφορών σφάλματος Bugzilla και τις οποίες μετά αποθηκεύει τοπικά σε μία MongoDB βάση δεδομένων σε JavaScript Object Notation (JSON) αναπαράσταση. Κατόπιν επιλέγονται συστήματα λογισμικού, για τα οποία υπάρχουν διαθέσιμες αναφορές σφάλματος, τα οποία είναι λογισμικά ελεύθερου λογισμικού των οποίων ο κώδικας είναι ελεύθερα διαθέσιμος. Ο κώδικας των επιλεγμένων συστημάτων αποθηκεύεται τοπικά και αναλύεται με στόχο να αποκτηθεί ένα λεξικό των όρων που ορίζονται μέσα στο πρόγραμμα το οποίο να αξιοποιηθεί για την εξαγωγή των αντίστοιχων όρων από τις αναφορές σφάλματος που έχουν γίνει για αυτό το σύστημα. Αυτό οδηγεί σε μία πιο συγκεντρωμένη απεικόνιση κάθε αναφοράς σφάλματος. Αυτά τα αποτελέσματα εν συνεχεία θα υποστούν περαιτέρω επεξεργασία και θα περαστούν σαν παράμετρος σε ένα σύνολο αλγορίθμων εξόρυξης δεδομένων σε μία προσπάθεια να αποκαλυφθεί αν υπάρχει όντως κάποια υφέρπουσα σχέση μεταξύ των αναφορών σφάλματος και του κώδικα που προκαλεί το σφάλμα. Τέλος θα γίνει προσπάθεια να διερευνηθεί η αξία αυτής της σχέσης, καθώς και η δυνατότητα εκμετάλλευσης της για τον αυτοματοποιημένο εντοπισμό του σημείου του κώδικα που προκαλεί το σφάλμα.

Τεχνικές που χρησιμοποιούνται εκτενώς στο πεδίο της εξόρυξης δεδομένων όπως οι LSA (Latent Semantic Indexing) και LDA (Latent Dirichlet Allocation) αξιοποιήθηκαν στα πλαίσια αυτής της εργασίας, δοκιμάζοντας διάφορες διαμορφώσεις με στόχο την βελτιστοποίηση της επίδοσής των.

Ο στόχος αυτής της διπλωματικής εργασίας είναι να δώσει στον αναγνώστη τη δυνατότητα να απαντήσει τις ακόλουθες ερωτήσεις:

- Η επιλογή οντοτήτων κώδικα από αναφορές σφάλματος ελεύθερου κειμένου βοηθάει στον ακριβή προσδιορισμό συσχετίσεων μεταξύ αναφορών σφάλματος;
- Μπορεί η συλλογή αλγορίθμων που επιλέχθηκαν να είναι αποτελεσματική μέσω λεπτομερούς προσδιορισμού των παραμέτρων τους για την ανακάλυψη χρήσιμων συσχετίσεων;
- Υπάρχει κάποια υπέρπουσα σημαντική σχέση ανάμεσα στις αναφορές σφάλματος ελεύθερου κειμένου η οποία μπορεί να διευκολύνει τη διαδικασία του εντοπισμού σφαλμάτων;
- Πόσο λεπτομερές πρέπει να είναι η ανάλυσή μας ώστε να είναι ταυτοχρόνως γρήγορη και να αποδίδει χρήσιμα αποτελέσματα;

### 1.3 Οργάνωση Κειμένου

Το υπόλοιπο αυτής της διπλωματικής εργασίας έχει οργανωθεί σε 6 κεφάλαια. Το δεύτερο κεφάλαιο συνοψίζει την ήδη υπάρχουσα γνώση στο πεδίο πάνω στην οποία βασίζεται όλη η υλοποίηση του εργαλείου που προτείνεται. Το τρίτο κεφάλαιο περιγράφει τη μεθοδολογία που ακολουθήθηκε καθώς και τα εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξη της πλατφόρμας. Το τέταρτο κεφάλαιο παρουσιάζει την συλλογή δεδομένων που χρησιμοποιήθηκε καθώς και το μοντέλο που επιλέχθηκε για την αναπαράστασή τους. Στο πέμπτο κεφάλαιο εξηγείται η διαδικασία της ανάλυσης αναφορών σφάλματος και στοιχεία επεξεργασίας φυσικής Γλώσσας NLP . Τέλος στο έκτο κεφάλαιο γίνεται η παρουσίαση της πειραματικής διαδικασίας, η μεθοδολογία που ακολουθήθηκε και τα αποτελέσματα. Στο έβδομο κεφάλαιο παρουσιάζονται τα συμπεράσματά της δουλειάς που έγινε, και παρουσιάζονται επιγραμματικά προτάσεις για μελλοντική δουλειά στον τομέα.

Συγκεκριμένα στο κεφάλαιο 2, παρουσιάζονται οι προαπαιτούμενες γνώσεις σχετικά με ανάλυση κώδικα και σχετικά εργαλεία που χρησιμοποιήθηκαν. Έπειτα γίνεται αναφορά στο εργαλεία ACDC που χρησιμοποιήθηκε για clustering και για εξόρυξη δεδομένων. Τέλος δίνονται σημαντικές πληροφορίες για τους αλγόριθμους που χρησιμοποιήθηκαν το LSI και το LDA .

Έχοντας εξετάσει εν συντομία τις πιο σημαντικές ιδέες, στο κεφάλαιο 3 θα εξηγηθεί η μεθοδολογία που ακολουθήθηκε, καθώς και επιπλέον μικρό εργαλεία που έπρεπε να υλοποιηθούν για τους σκοπούς της εργασίας. Πρώτα οι θα εξηγηθούν βασικές λειτουργίες της

πλατφόρμας και μετά θα γίνει το ίδιο για έτοιμο λογισμικό τρίτων το οποίο χρησιμοποιήθηκε ή ενσωματώθηκε στο σύστημα. Τέλος θα παρουσιαστεί η αρχιτεκτονική του συστήματος.

Στο κεφάλαιο 4, θα επεξηγηθεί η συλλογή δεδομένων και θα παρουσιαστεί το προτεινόμενο μοντέλο για την μοντελοποίηση των αναφορών σφάλματος. Επίσης θα οριστεί μία πρόχειρη μετρική για υπολογισμό τις ποιότητας μιας αναφοράς σφάλματος στα πλαίσια του συστήματος μας .

Στο κεφάλαιο 5, θα παρουσιασθούν τα βασικότερα εργαλεία που χρησιμοποιήθηκαν για την επεξεργασία των αναφορών σφάλματος. Αυτά είναι το εργαλείο επεξεργασίας φυσικής γλώσσας NLP και η τεχνική LSA που χρησιμοποιήθηκε για την εξαγωγή συσχετισμών ανάμεσα σε αναφορές σφάλματος.

Στο κεφάλαιο 6, αναλύονται τα πειραματικά αποτελέσματα. Παρουσιάζεται λεπτομερώς όλη η πειραματική διαδικασία που χρησιμοποιήθηκε και δίνονται επιπλέον πληροφορίες σχετικά με το υποσύνολο των αναφορών σφάλματος που αναλύθηκαν. Τέλος θα γίνει βήμα προς βήμα σχολιασμός των αποτελεσμάτων και τις σημασίας που αυτά έχουν.

Τέλος, στο κεφάλαιο 7, θα συζητηθούν τα συμπεράσματα, στα οποία καταλήξαμε μέσω πειραματισμού με διάφορες τεχνικές εξόρυξης δεδομένων για εντοπισμό σφαλμάτων. Επιπλέον θα αναφερθούν ζητήματα τα οποία μπορεί να αποτελούν απειλή για τα συμπεράσματα και θα παρουσιαστούν κάποιες ιδέες για μελλοντική έρευνα στο πεδίο.



## Κεφάλαιο 2

# Προαπαιτούμενες Γνώσεις

### 2.1 Software Clustering

Οι μεθοδολογίες του Software clustering (Ομαδοποίηση Λογισμικού), ομαδοποιούν σύνολα από οντότητες ενός συστήματος λογισμικού σε σημαντικά υποσυστήματα με σκοπό να βοηθήσουν στην διαδικασία κατανόησης της δομής στο υψηλότερο επίπεδο, πολύπλοκων συστημάτων λογισμικού μεγάλης κλίμακας. Μία προσέγγιση του Software Clustering η οποία θεωρείται επιτυχημένη στην διεκπεραίωση αυτής της εργασίας μπορεί να έχει σημαντικά πρακτικά οφέλη για τους μηχανικούς πληροφοριακών συστημάτων, ειδικά για αυτούς που εργάζονται πάνω σε μεγάλα legacy συστήματα με απαρχαιωμένη ή ανύπαρκτη τεκμηρίωση σε επίπεδο τεχνικού κειμένου.

Το Software Clustering έχει απασχολήσει την ερευνητική κοινότητα για πάνω από δύο δεκαετίες. Κατά τη διάρκεια αυτής της περιόδου έχουν προταθεί πολλών ειδών αλγόριθμοι στη βιβλιογραφία, οι περισσότεροι εκ των οποίων έχουν εφαρμοσθεί σε συγκεκριμένα συστήματα λογισμικού με μεγαλύτερη ή μικρότερη επιτυχία. Το κλειδί για να αξιολογηθεί αν ένας αλγόριθμος Software Clustering είναι επιτυχημένος, είναι η 'απόσταση' της λύσης που προτείνει για κάποιο σύστημα σε σχέση με την λύση που προτείνει κάποιος ειδήμων το συστήματος για το αυτό σύστημα. Προς αυτήν την κατεύθυνση η ερευνητική κοινότητα έχει αναπτύξει αρκετές μεθόδους για την αξιολόγηση της ποιότητας των προτεινόμενων αλγορίθμων για Software Clustering όπως η απόσταση MoJo [12] και το Craft framework [13] ενώ αρκετά πειράματα έχουν ήδη διεξαχθεί πάνω στο θέμα [14] και [15].

Το Software Clustering έχει μονοπωλήσει το ενδιαφέρον της επιστημονικής κοινότητας τόσο πολύ γιατί δύναται να επιλύσει προβλήματα που αντιμετωπίζονται στη τεχνολογία λογισμικού σε δύο καίριες περιοχές.

- Εξέλιξη Λογισμικού

- Ανάκτηση Πληροφορίας

*Εξέλιξη Λογισμικού:* Κάθε πληροφοριακό σύστημα εξελίσσεται με σκοπό να προστεθεί σε αυτό νέα λειτουργικότητα, να επιδιορθωθούν υπάρχοντα προβλήματα, και να ενισχυθεί η σταθερότητα λειτουργίας του. Τα εργαλεία για Software Clustering χρησιμοποιούνται για να βελτιώσουν τη δομή του λογισμικού software restructuring (αναδόμηση λογισμικού) ή να ελαττώσουν την πολυπλοκότητα μεγάλων λειτουργικών μονάδων (modules) source code decoupling (αποσύνδεση πηγαίου κώδικα).

Η Αναδόμηση Λογισμικού είναι ένας τρόπος βελτιωτικής συντήρησης ο οποίος τροποποιεί τη δομή του πηγαίου κώδικα ενός προγράμματος. Ο στόχος είναι να διευκολυνθούν οι εργασίες συντήρησης όπως είναι η προσθήκη λειτουργικότητας και η επιδιόρθωση, προηγουμένως άγνωστων σφαλμάτων μέσα σε ένα σύστημα λογισμικού [16]. Η αποσύνδεση πηγαίου κώδικα προσπαθεί να μειώσει την πολυπλοκότητα των σύνθετων υποσυστημάτων ή μεθόδων/συναρτήσεων. Στο [17] παρουσιάζεται μία μελέτη εργασίας όπου χρησιμοποιείται software clustering στον πηγαίο κώδικα για επίτευξη αποσύνδεση σε διαδικαστικό (procedural) επίπεδο. Το εργαλείο του software clustering ομαδοποιεί της εντολές για να δημιουργήσει μία κατάταξη της εξάρτησης ανάμεσα στις παραχθείσες ομάδες.

*Ανάκτηση Πληροφορίας:* Ο κύριος στόχος του reverse engineering (αποσύμπλιση) είναι η ανάκτηση των υποσυστημάτων ή η εξαγωγή αφαιρέσεων του συστήματος. Αρκετές τέτοιες προσεγγίσεις και τεχνικές έχουν προταθεί στην βιβλιογραφία. Ειδικότερα, μέθοδοι ανάκτησης αρχιτεκτονικής (architecture recovery) έχουν κάνει χρήση μεθόδων software clustering με ιδιαίτερη επιτυχία. Οι μέθοδοι για ανάκτηση αρχιτεκτονικής επικεντρώνονται στην ανακάλυψη της αρχιτεκτονικής του συστήματος μέσω της ανάλυσης των αφαιρέσεων που έχουν εξαχθεί από τον πηγαίο κώδικα, όπως τα υποτιμήματα, τα υποσυστήματα και τα σχεδιαστικά μοτίβα. Στο [18] γίνεται χρήση των σχεδιαστικών μοτίβων για ανάκτηση αρχιτεκτονικής.

Το Software Clustering χρησιμοποιείται επίσης για εντοπισμό διπλότυπου κώδικα [17] όπως επίσης και για πρόβλεψη της 'σφαλερότητας' κάποιων τμημάτων του λογισμικού [19]. Είναι προφανές ότι το software clustering, είναι μία πολύ σημαντική τεχνική η οποία μπορεί να χρησιμοποιηθεί για την επίλυση πολλών προβλημάτων στον τομέα της Τεχνολογίας Λογισμικού.

Παρακάτω παρουσιάζεται μία εις βάθος ανάλυση δύο εργαλείων τα οποία χρησιμοποιούνται για software clustering, του ACDC (Algorithm for Comprehension Driven Clustering) [20] και του Bunch [21] τα οποία χρησιμοποιούνται αρκετά συχνά σε σχετικές μελέτες που έχουν προστεθεί στη βιβλιογραφία.

### 2.1.1 ACDC

Το ACDC είναι ένας αλγόριθμος βασισμένος στον εντοπισμό συχνά παρατηρήσιμων μοτίβων στα υποσυστήματα πολύ μεγάλων συστημάτων λογισμικού. Όπως τονίζεται και στο [20], οι περισσότεροι αλγόριθμοι για Software Clustering στοχεύουν στην επίτευξη χαμηλής σύζευξης και υψηλής συνοχής, ελαχιστοποίηση των διαπροσωπιών, κοινών γειτόνων και τα λοιπά. Παρά όλα αυτά στην προσπάθεια για μεγιστοποίηση της απόδοσης ή της ακρίβειας των αλγορίθμων, οι ερευνητές έχουν παραγνωρίσει το κύριο μέλημά του του software clustering το οποίο είναι η κατανόηση του τελικού αποτελέσματος. Προς αυτήν την κατεύθυνση ο αλγόριθμος του ACDC χρησιμοποιεί αναγνώριση προτύπων για να βρίσκει συγκεκριμένες αποσυνθέσεις συστημάτων οι οποίες εμφανίζονται συχνά σε συστήματα λογισμικού. Επιπροσθέτως, δημιουργεί σημασιολογικά χρήσιμα ονόματα για τους clusters τα οποία δίνουν χρήσιμη πληροφορία για τον κάθε έναν και κάνουν την τελική αποσύνθεση πιο κατανοητή. Τέλος, τα μεγέθη των clusters περιορίζονται με στόχο να κάνουν αυτή την αποσύνθεση πιο διαχειρίσιμη και κατανοητή. Το μέγιστο μέγεθος κάθε ενός cluster ορίζεται εξάρχης σε 20 στοιχεία και αυτό είναι και το μέγεθος που θα χρησιμοποιηθεί και κατά τη χρήση του εργαλείου σε αυτήν την εργασία.

Το ACDC προσφέρει αρκετά χαρακτηριστικά τα οποία βοηθούν τον μηχανικό λογισμικού να καταλάβει την εξαχθείσα αποσύνθεση γρηγορότερα. Αυτά τα χαρακτηριστικά είναι :

1. *Αποδοτική ονοματοδοσία των clusters*: Αυτό το χαρακτηριστικό επιτρέπει στους μηχανικούς που έχουν μία εξοικείωση με το σύστημα λογισμικού, να καταλάβουν την αποσύνθεση πιο εύκολα και να αποκομίσουν οφέλη από αυτήν γρηγορότερα. Δυστυχώς, αυτή η παράμετρος της σωστής ονοματοδοσίας δεν έχει τραβήξει όσο ενδιαφέρον θα έπρεπε. Οι δημιουργεί του ACDC πιστεύουν ότι είναι σημαντικό για την αφύψηλού επίβλεψη ενός συστήματος λογισμικού να δίνονται ονόματα τα οποία συσχετίζονται κάπως με το σύστημα παρά να δίνονται ανούσια ονόματα όπως SS01 ή Subsystem01.
2. *Αποτελεσματική ονοματοδοσία των clusters*: Οι αποσυνθέσεις που περιέχουν μικρούς clusters από ένα ή δύο πηγές, παρά το ότι θεωρείται ότι επιτυγχάνουν χαμηλή σύζευξη, στην πράξη είναι αδύνατο να χρησιμοποιηθούν. Από την άλλη, υπάρχει ανάλογο πρόβλημα και με την ύπαρξη πολύ μεγάλων clusters. Για αυτόν τον λόγο το ACDC οριοθετεί το μέγιστο μέγεθος των παραγόμενων clusters σε ένα συγκεκριμένο πλήθος αντικειμένων ώστε να κάνει την αποσύνθεση πιο κατανοητή και χρήσιμη. Ωστόσο αυτός ο περιορισμός δεν επιβάλλεται σε βάρος της δομής του συστήματος και αν συντρέχει λόγος, μπορούν να δημιουργηθούν και μεγάλοι clusters η οποίοι απλά θα αποσυντεθούν περαιτέρω σε μικρότερους σε εσωτερικό επίπεδο.

3. *Προσέγγιση Καθοδηγούμενη από Μοτίβα*: Το ACDC είναι βασισμένο πάνω σε ένα σύνολο από μοτίβα τα οποία εμφανίζονται πολύ συχνά σε αποσυνθέσεις που διεξάγονται 'με το χέρι' από ειδικούς. Αυτά τα μοτίβα εμφανίζονται σε μεγάλα συστήματα με πάνω από 100 αρχεία πηγαίου κώδικα, και καλύπτουν μία ευρεία γκάμα από γνωστές δομές υποσυστημάτων οι οποίες εμφανίζονται σε αποσυνθέσεις 'με το χέρι' μεγάλων, βιομηχανικής κλίμακας, συστημάτων λογισμικού. Τα μοτίβα τα οποία εντοπίζονται από το ACDC παρουσιάζονται στον Πίνακα 2.1.

| <b>Pattern</b>              | <b>Description</b>   |
|-----------------------------|--|
| Source file pattern         | The set of procedures/functions/methods can be grouped together with the variables contained in the same source file.  |
| Directory structure pattern | The structure of the source code may correspond to a subsystem decomposition. However, this pattern is not always the case and must be used with caution because many systems contain folders with a collection of header files or libraries and such a cluster will not be useful from a comprehension point of view. |
| Body-header pattern         | The header file and the corresponding body file, such as .c and .h files in C, can create a cluster containing only these files. Even though this is a really small cluster, it can reduce the complexity of a system's structure.   |
| Leaf collection pattern     | A set of files which are not connected to each other, but offer similar services, such as a set of drivers.  |
| Support library pattern     | Procedures accessed by the majority of the system's subsystems [22] are grouped together into one cluster.   |
| Central dispatcher pattern  | This pattern is the dual of the support library pattern. Large systems commonly contain a small number of resources that depend on a large number of other resources. These are ignored initially, in order not to obscure other patterns and then are reconsidered once subsystems have already been formed.          |
| Subgraph dominator pattern  | This pattern searches for a subgraph in the system's graph which contains a set of nodes where in order to access each node of the set, one must go through one specific node, considered a "dominator node".  |

ΠΙΝΑΚΑΣ 2.1: Μοτίβα εντοπίσημα από το ACDC

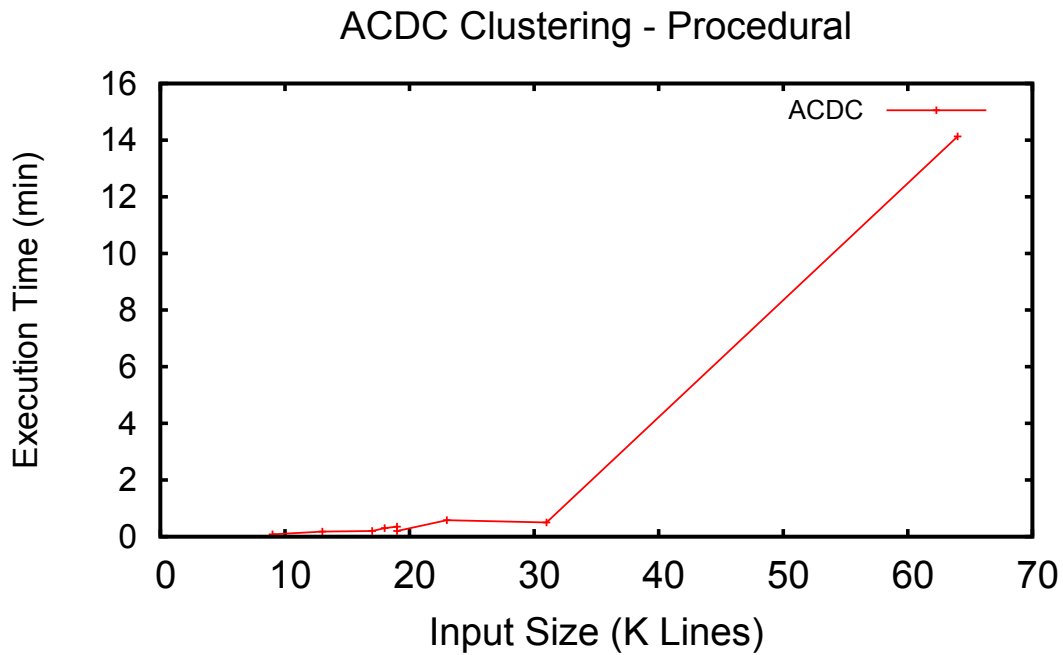
Για αυτό το λόγο το ACDC χρησιμοποιεί αυτή τη γνώση και αναγνωρίζει αυτά τα μοτίβα με στόχο να δημιουργήσει την αποσύνθεση ενός συστήματος.

Ο αλγόριθμος εκτελείται σε δύο βήματα. Στο πρώτο βήμα, βασιζόμενο στον εντοπισμό και την αναγνώριση των προαναφερθέντων μοτίβων, δημιουργεί μία αρχική αποσύνθεση και οι καινούριοι clusters λαμβάνουν κατάλληλα ονόματα. Πρώτα δημιουργούνται clusters οι οποίοι περιέχουν αρχεία πηγαίου κώδικα και από αυτό το σημείο κι έπειτα τα αρχεία και οι περιεχόμενες οντότητες αντιμετωπίζονται σαν μία αυτόνομη οντότητα. Μετά αναγνωρίζονται τα μοτίβα *body-header*, *leaf collection* και των *support libraries*. Για το τελευταίο δεν δημιουργούνται καινούριοι clusters αλλά απλά οι οντότητες αναγνωρίζονται σαν υποψήφιες οντότητες. Μετά οι clusters, *central dispatcher* και *subgraph dominator* δημιουργούνται και τέλος δημιουργείται και ο *supporting cluster* συμπεριλαμβάνοντας όλα τα εναπομείναντα συτοιχεία τα οποία έχουν αναγνωριστεί στο προηγούμενο βήμα ως υποψήφιες υποστηρικτικές οντότητες και ακόμα δεν έχουν ενσωματωθεί σε κανένα άλλο υποσύστημα. Κατά τη διάρκεια του δεύτερου βήματος τα εναπομείναντα στοιχεία τα οποία δεν έχουν ανατεθεί σε κανένα υποσύστημα συμπεριλαμβάνονται στην τρέχουσα αποσύνθεση χρησιμοποιώντας της τεχνική *orphan adoption* [23]. Με άλλα λόγια τα υπόλοιπα στοιχεία των συστημάτων τα οποία δεν ανατέθηκαν σε κάποιον cluster κατά το πρώτο βήμα, προστίθενται σε εκείνον με τον οποίον έχουν της περισσότερη αλληλεπίδραση. Αν υπάρχουν αρκετοί υποψήφιοι clusters οι οποίοι πληρούν της προϋποθέσεις, δημιουργείται ένας καινούριος και όλες οι οντότητες που δεν μπορούν να ανατεθούν σε κάποιον από τους υπάρχοντες με βεβαιότητα ανατίθενται σε αυτόν.

#### 2.1.1.1 Χρόνος Εκτέλεσης

Ο αλγόριθμος ACDC λαμβάνει σαν είσοδο ένα RSF αρχείο. Ο Χρόνος εκτέλεσης εξαρτάται από το μέγεθος του αρχείου εισόδου καθώς και από τις συνδέσεις μεταξύ των οντοτήτων του συστήματος. Διεξήχθησαν πειράματα για τον προσδιορισμό του χρόνου εκτέλεσης όπου τα αποτελέσματα κυμάνθηκαν από 10 δευτερόλεπτα για τα μικρότερα συστήματα μέχρι 15 λεπτά για κάποια από τα μεγαλύτερα. Οι γραφικές παραστάσεις 2.1 και 2.2 δείχνουν τον μέσο χρόνο που χρειάστηκε το clustering για αρκετά διαδικαστικά (procedural) και αντικειμενοστραφή συστήματα, σε σχέση με το πλήθος των σχέσεων που δόθηκαν σαν είσοδος. Για να γίνουν τα αποτελέσματα πιο ξεκάθαρα οι Πίνακες 2.2 και 2.3 δείχνουν αυτήν την πληροφορία και για τους δύο τύπους συστημάτων.

Σε αυτό το σημείο μπορεί να γίνει η παρατήρηση ότι τα αντικειμενοστραφή συστήματα χρειάζονται ελαφρώς περισσότερο από τα αντίστοιχου μεγέθους διαδικαστικά. Αυτό συμβαίνει γιατί παρόλο που αναφερόμαστε σε παρόμοιου μεγέθους αρχεία εισόδου, όπως μπορούμε να δούμε από τους Πίνακες 2.2 και 2.3 τα διαδικαστικά συστήματα έχουν σχεδόν τους διπλάσιους κόμβους από τα αντικειμενοστραφή και για αυτό το λόγο πρέπει να πρέπει να διευθετηθούν περισσότερες συγκρούσεις.



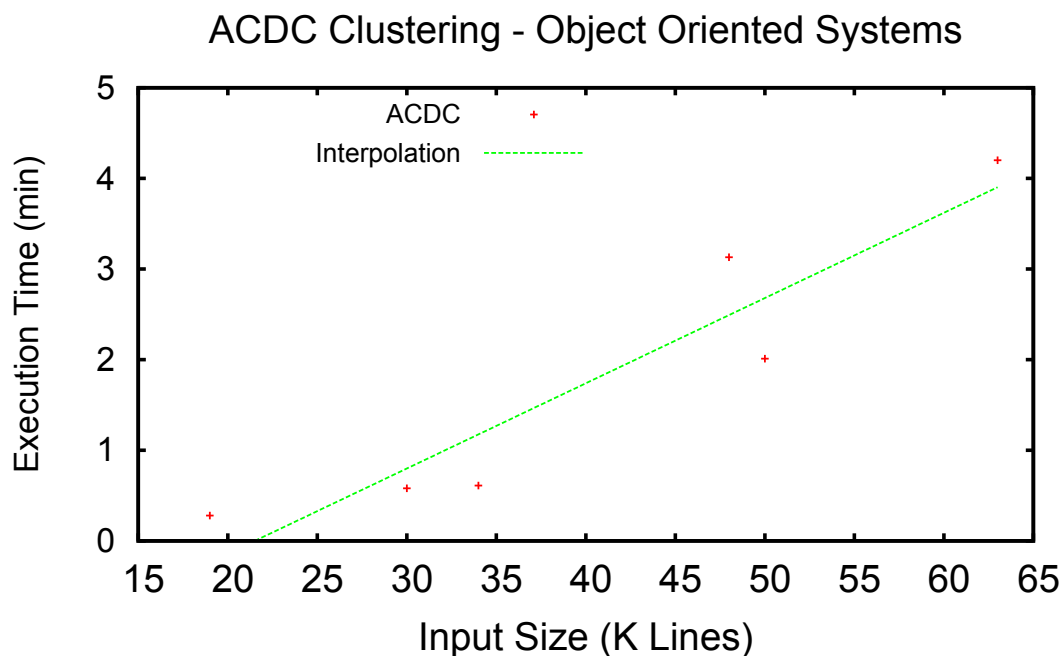
ΣΧΗΜΑ 2.1: Χρόνος εκτέλεσης του ACDC συγκριτικά με το πλήθος των σχέσεων της εισόδου για Procedural Systems

| System   | Input File Size | No of Nodes | Time (Min) |
|----------|-----------------|-------------|------------|
| Freeglut | 9275            | 4161        | 0.80       |
| Tcsh     | 12995           | 3138        | 0.18       |
| Zsh      | 16905           | 5309        | 0.20       |
| Putty    | 18439           | 6111        | 0.30       |
| OpenVPN  | 18876           | 4411        | 0.35       |
| OpenSSH  | 19370           | 4037        | 0.20       |
| Bash     | 22850           | 7246        | 0.58       |
| Clips    | 31217           | 6752        | 0.50       |
| OpenSSL  | 62283           | 23385       | 14.13      |

ΠΙΝΑΚΑΣ 2.2: Χρόνος εκτέλεσης του ACDC για procedural systems

| System       | Input File Size | No of Nodes | Time (Min) |
|--------------|-----------------|-------------|------------|
| TexMaker     | 18643           | 3970        | 0.28       |
| Apache Maven | 29827           | 11506       | 0.58       |
| jHotDraw     | 33797           | 10712       | 0.61       |
| Apache Ivy   | 48184           | 11280       | 3.13       |
| jEdit        | 49903           | 12875       | 2.01       |
| Apache Ant   | 63332           | 18149       | 4.20       |

ΠΙΝΑΚΑΣ 2.3: Χρόνος εκτέλεσης του ACDC για Object oriented systems



ΣΧΗΜΑ 2.2: Χρόνος εκτέλεσης του ACDC συγκριτικά με το πλήθος των σχέσεων της εισόδου για Object Oriented Systems

## 2.2 Αναπαράσταση Αρχιτεκτονικής

Σύμφωνα με το OMG, το μοντέλο ενός συστήματος αποτελεί περιγραφή ή αναπαράσταση αυτού του συστήματος και του περιβάλλοντός του για κάποιο συγκεκριμένο σκοπό. Ένα μοντέλο συχνά αναπαρίσταται ως ένας συνδυασμός σχεδίων και κειμένου, τα οποία μπορούν να ανήκουν είτε σε κάποια φυσική γλώσσα (αγγλικά, ελληνικά) είτε σε κάποια γλώσσα μοντελοποίησης [24]. Από τη στιγμή που τα μοντέλα χρησιμοποιούνται τόσο πολύ, όχι μόνο στον μοντελο-κεντρικό σχεδιασμό και την ανάστροφη μηχανική αλλά πάνω σε μία πλειάδα πεδίων της Τεχνολογίας Λογισμικού, είναι απαραίτητο τα μοντέλα να αναπαρίστανται με έναν ομοιογενή και εκφραστικό τρόπο. Για αυτόν τον λόγο έχουν προταθεί αρκετές γλώσσες μοντελοποίησης. Τέτοιες γλώσσες υπάρχουν και κειμενό-κεντρικές, και σχεδιαστικές. Οι γραφικές γλώσσες χρησιμοποιούν μία διαγραμματική τεχνική με επώνυμα σύμβολα τα οποία αναπαριστούν έννοιες, γραμμές που συνδέουν τα σύμβολα και αναπαριστούν σχέσεις και αρκετές άλλες γραφιστικές οντότητες για την έκφραση περιορισμών. Από την άλλη πλευρά οι κειμενό-κεντρικές γλώσσες μοντελοποίησης χρησιμοποιούν κανονικοποιημένες λέξεις κλειδιά συνοδευόμενες από παραμέτρους ή ακόμα και όρους φυσικής γλώσσας και φράσεις ώστε να δημιουργήσουν εκφράσεις κατανοητές από υπολογιστή. Η πιο διαδεδομένη και πολυχρησιμοποιημένη γλώσσα μοντελοποίησης είναι η UML (Unified Modeling Language) η οποία έχει καθιερωθεί πλέον σαν το προτιμώμενο εργαλείο στην βιομηχανία για ορισμό τεχνικών χαρακτηριστικών, σχεδιασμό και οπτικοποίηση καθώς και για γραπτή τεκμηρίωση συστημάτων λογισμικού. Με όρους της UML ένα μοντέλο είναι ένα στιγμιότυπο ενός UML



metamodel και ένα διάγραμμα περιγράφει την γραφική αναπαράσταση ενός μοντέλου [25]. Επιπλέον, κάθε στοιχείο ενός μοντέλου αναφέρεται σε ένα στιγμιότυπο του UML meta-class ενώ η ιδιότητες ενός στοιχείου ενός μοντέλου αναφέρονται σε ένα meta-attributes στιγμιότυπο που ανήκει στο συγκεκριμένο στοιχείο. Η κατάσταση του κάθε στοιχείου του μοντέλου καθορίζεται από την τιμή ιδιότητας.

### 2.2.1 Το πρότυπο Meta Object Facility (MOF)

Η από μοντέλο οδηγούμενη αρχιτεκτονική (Model Driven Architecture) είναι ένας τρόπος σχεδιασμού λογισμικού για ανάπτυξη συστημάτων λογισμικού. Προσφέρει ένα σύνολο κανόνων για δόμηση των απαιτήσεων. Έγινε για πρώτη φορά διαθέσιμος από το Object Management Group(OMG) το 2001 και σχετίζεται με αρκετά πρότυπα συμπεριλαμβανομένης της Unified Modelling Language (UML) και του προτύπου ανταλλαγής δεδομένων XML Metadata Interchange (XMI).

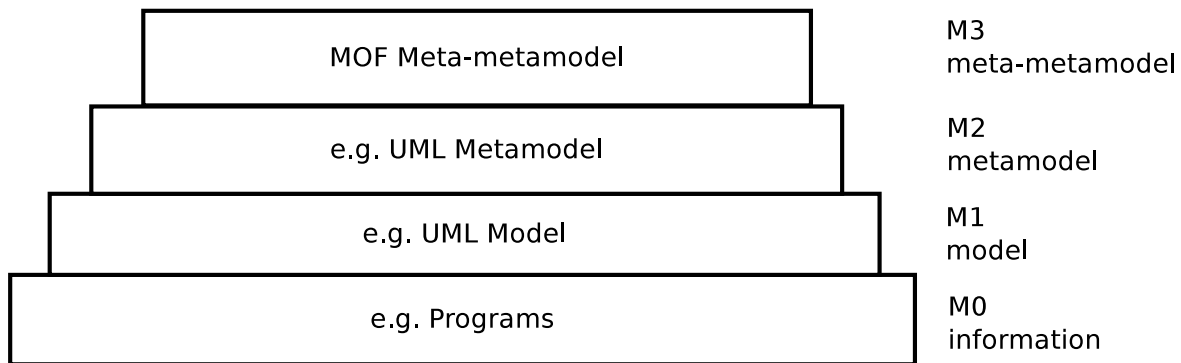
Όπως αναφέρεται στο [24], η προδιαγραφή του Meta Object Facility ορίζει μία αφηρημένη γλώσσα και ένα πλαίσιο εργασίας για τον προσδιορισμό, την κατασκευή και την διαχείριση, μεταμοντέλων ανεξαρτήτων τεχνολογίας. Ένα μεταμοντέλο είναι στην πράξη μία αφηρημένη γλώσσα για κάποιο τύπο μεταδεδομένων. Παραδείγματα μεταμοντέλων συμπεριλαμβάνουν, αλλά δεν περιορίζονται, την UML, το ίδιο το MOF και αρκετές από τις προδιαγραφές εργαλείων του OMG οι οποίες είναι υπό κατασκευή.

Επιπροσθέτως το MOF ορίζει ένα πλαίσιο εργασίας για την υλοποίηση συλλογών διάθεσης μεταδεδομένων (παραδείγματος χάριν μοντέλων) περιγραφόμενων από μεταμοντέλα. Αυτό το πλαίσιο εργασίας χρησιμοποιεί κοινά χρησιμοποιούμενη τεχνολογία για την αντιστοίχιση και τον μετασχηματισμό των MOF μεταμοντέλων σε API μεταδεδομένων. Αυτή η πρακτική προσφέρει συνεπή, εναλασσόμενης λειτουργικότητας συλλογές παροχής μεταδεδομένων, για διάφορες τεχνολογίες και προϊόντα.

#### 2.2.1.1 Τετραεπίπεδη Αρχιτεκτονική Μεταδεδομένων

Το κλασσικό πλαίσιο εργασία για την μετα-μοντελοποίηση βασίζεται σε μία αρχιτεκτονική τεσσάρων μετά-επιπέδων. Κάθε μετά-επίπεδο είναι ένα στιγμιότυπο του ακριβώς από πάνω επιπέδου. Το σχήμα 2.3 δείχνει γραφικά την κλασσική αρχιτεκτονική τεσσάρων μετά-επιπέδων.

Στην κορυφή της ιεραρχίας αυτή της αρχιτεκτονικής τεσσάρων επιπέδων είναι το επωνομαζόμενο M3 επίπεδο. Η κύρια υποχρέωση αυτού του επιπέδου είναι να ορίσει τη γλώσσα στην οποία θα οριστεί ένα μεταμοντέλο. Η MOF είναι ένα παράδειγμα ενός τέτοιου μετά-μετά-μοντέλου. Σύμφωνα με το [26], ένα μετά-μετά-μοντέλο είναι συνήθως πιο συνοπτικό



ΣΧΗΜΑ 2.3: Αρχιτεκτονική Τεσσάρων Μετά-Επιπέδων

από ένα μετά-μοντέλο το οποίο αυτό περιγράφει, ενώ συχνά περιγράφει περισσότερα από ένα μετά-μοντέλα. Είναι γενικά επιθυμητό, συσχετιζόμενα μετά-μοντέλα και μετά-μετά-μοντέλα να μοιράζονται κάποιες κοινές σχεδιαστικές αντιλήψεις και δομές. Άσχετα από τα προηγούμενα, κάθε επίπεδο μπορεί να εποπτευθεί εντελώς ανεξάρτητα από τα υπόλοιπα και είναι απαραίτητο να διατηρεί την δική του σχεδιαστική ορθότητα. Τα μετά-μετά-μοντέλα ορίζονται από τα ίδια τα μετά-μετά-μοντέλα και δεν χρειάζεται να ανέβει ο σχεδιαστής σε κάποιο επίπεδο, έστω, M4 για να τα σχεδιάσει. Αυτό συμβαίνει γιατί αυτά (τα μετά-μετά-μοντέλα) διατηρούν την ανακλαστική ιδιότητα, για παράδειγμα το MOF έχει οριστεί χρησιμοποιώντας το MOF.

Τα στοιχεία του επιπέδου του μετά-μοντέλου είναι στιγμιότυπα του επιπέδου του μετά-μετά-μοντέλου. Ο κύριος στόχος του επιπέδου του μετά-μοντέλου, του επονομαζόμενου και επιπέδου M2, είναι να ορίζει μία γλώσσα για τον ορισμό μοντέλων. Η γλώσσα UML και το Common Warehouse Metamodel (CWM) του OMG είναι παραδείγματα τέτοιων μετά-μοντέλων. Τα μετά-μοντέλα είναι συνήθως πιο περιγραφικά από τα μετά-μετά-μοντέλα που τα περιγράφουν. Πολύ περισσότερο μάλιστα αν ορίζουν και δυναμική σημασιολογία. Επί παραδείγματι η UML αποτελεί ένα μετά-μοντέλο στιγμιότυπο του μετά-μετά-μοντέλου MOF.

Τα μοντέλα ανήκουν στο επίπεδο των μοντέλων, το οποίο αναφέρεται επίσης ως επίπεδο M1. Ένα μοντέλο είναι ένα στιγμιότυπο ενός μετά-μοντέλου. Η κύρια λειτουργικότητα του επιπέδου του μοντέλου είναι να ορίσει μία γλώσσα η οποία να μπορεί να περιγράψει σημασιολογία συγκεκριμένων πεδίων εφαρμογών. Ένα μοντέλο χρήστη είναι ένα στιγμιότυπο, κατά συνέπεια, ενός UML μετά-μοντέλου. Στο πεδίο του λογισμικού, τα στοιχεία του επιπέδου M1 είναι τα μοντέλα των συστημάτων λογισμικού όπως αυτά είναι ορισμένα στα μοντέλα UML, δηλαδή διαγράμματα κλάσεων, ακολουθιακά διαγράμματα και τα λοιπά.

Στο χαμηλότερο σκαλοπάτι της ιεραρχία βρίσκεται το επίπεδο M0. Αυτό το επίπεδο περιέχει στιγμιότυπα χρόνου εκτέλεσης των στοιχείων του μοντέλου που ορίζονται σε ένα μοντέλο.

Αυτές οι απεικονίσεις οι οποία μοντελοποιούνται στο επίπεδο M1 είναι είναι 'περιορισμένες' εκδόσεις των στιγμιοτύπων χρόνου εκτέλεσης του επιπέδου M0.

Είναι σημαντικό να υποσημειωθεί ότι υπάρχουν και αρχιτεκτονικές μοντελοποίησης παραπάνω των τεσσάρων επιπέδων, η οποίες μοιράζονται ωστόσο όλες τα ίδια χαρακτηριστικά, συγκεκριμένα υπάρχουν ζεύγη κλάσεων και στιγμιοτύπων και ένας μηχανισμός για μετάβαση από το στιγμιότυπο στην κλάση.

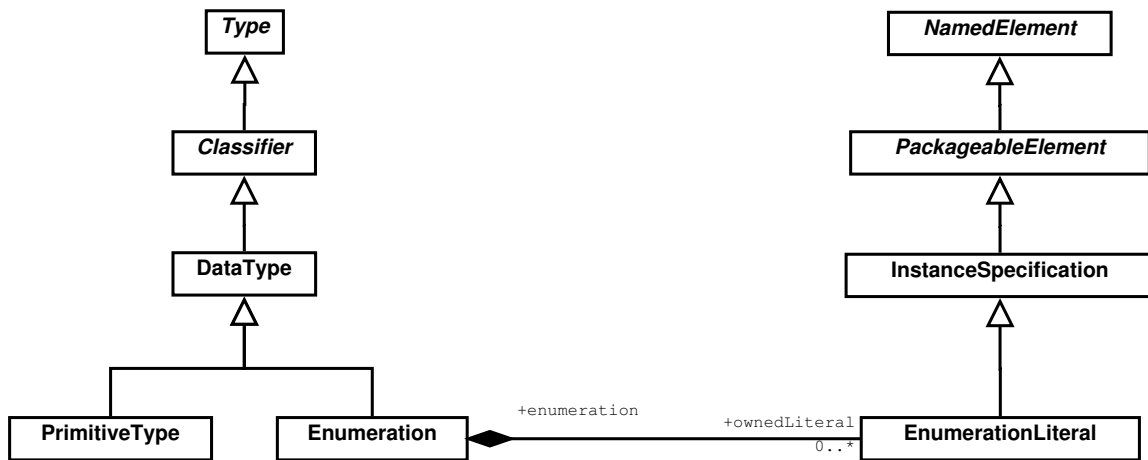
### 2.2.1.2 Δομή και Περιγραφή του MOF

Η πιο πρόσφατη έκδοση του MOF (2.4.2) έχει σχεδιαστεί και οριστεί με τέτοιο τρόπο ώστε να είναι σφικτά συνδεδεμένη με την UML v2.0. Επιπλέον, για τον προσδιορισμό του MOF χρησιμοποιεί συντακτικό και σημασιολογία που προκύπτουν από την UML. Στην πράξη για τον προσδιορισμό του MOF χρησιμοποιεί ένα υποσύνολο της UML (κυρίως τα διαγράμματα κλάσης), μία γλώσσα περιγραφής περιορισμών Object Constraint Language (OCL) και πολύ ακριβείς εκφράσεις φυσικής γλώσσας.

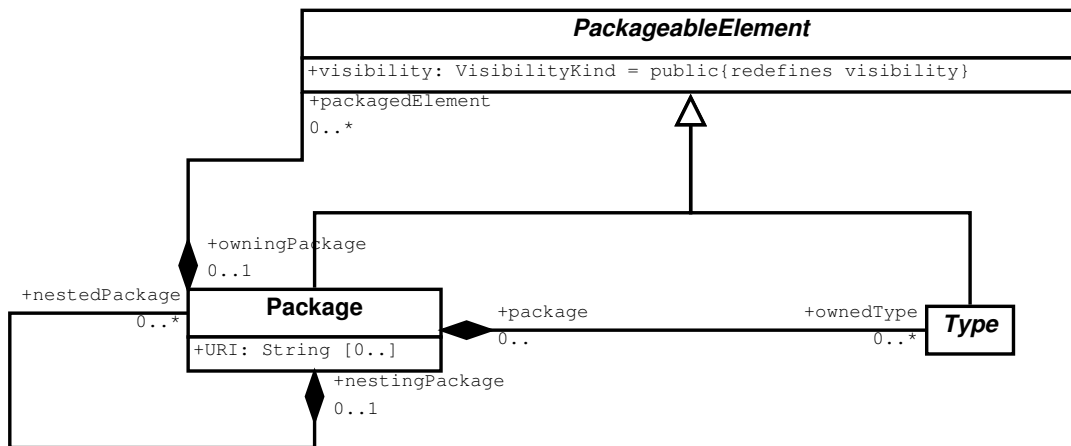
Ο ορισμός του MOF προσφέρει δύο παραλλαγές: το 'απαραίτητο' Essential MOF (EMOF) και το πλήρες MOF (CMOF). Τόσο το EMOF όσο και το CMOF ορίζονται χρησιμοποιώντας το CMOF, το οποίο χρησιμοποιείται επίσης για τον ορισμό της UML2. Το EMOF ορίζεται επίσης εξόλοκληρου στο EMOF εφαρμόζοντας εισαγωγή πακέτων, και σημασιολογία συνενώσεων από την CMOF περιγραφή του. Ως αποτέλεσμα, το EMOF και το CMOF χρησιμοποιούνται για να περιγράψουν τον εαυτό του, και το κάθε ένα προκύπτει και επαναχρησιμοποιεί τμήματα της δομικής βιβλιοθήκης (Infrastructure Library) της UML2.

Ενώ ο στόχος του CMOF είναι να παρέχει ένα γενικότερο πλαίσιο εργασίας για μετά-μοντελοποίηση, το EMOF είναι ένα υποσύνολο του MOF το οποίο ανταποκρίνεται καλύτερα στα εργαλεία που βρίσκει κανείς στις γλώσσες αντικειμενοστραφούς προγραμματισμού και την XML.

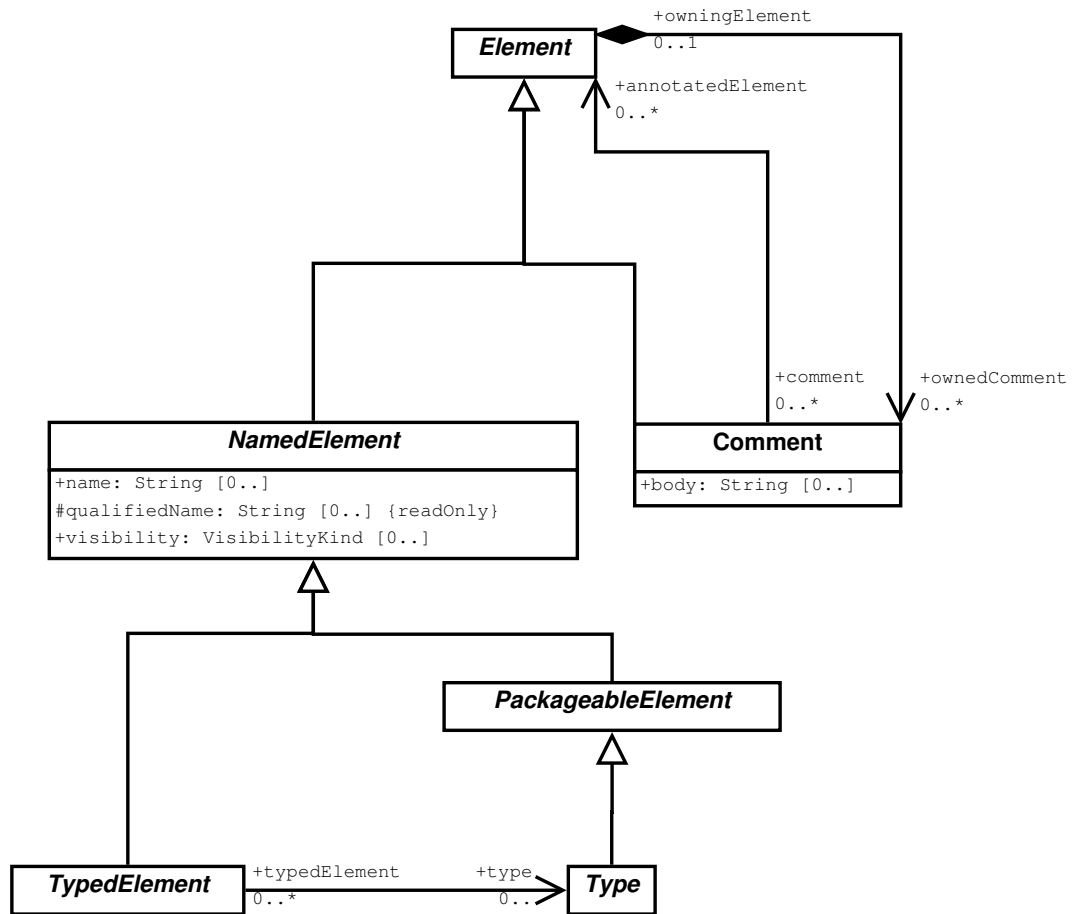
Καθώς η λειτουργικότητα που παρέχει το EMOF είναι αρκετή για τους σκοπούς αυτής της εργασίας, οι εικόνες 2.4 - 2.6 παρουσιάζουν τα MOF διαγράμματα τα οποία ορίζουν το EMOF όπως αυτά παρουσιάζονται στο κείμενο τεκμηρίωσης [27]. Σημειώνεται σε αυτό το σημείο ότι η πλήρης τεκμηρίωση του EMOF περιέχει αρκετούς περιορισμούς καθώς και κανόνες και ορισμούς οι οποίοι βρίσκονται στο κείμενο τεκμηρίωσης.



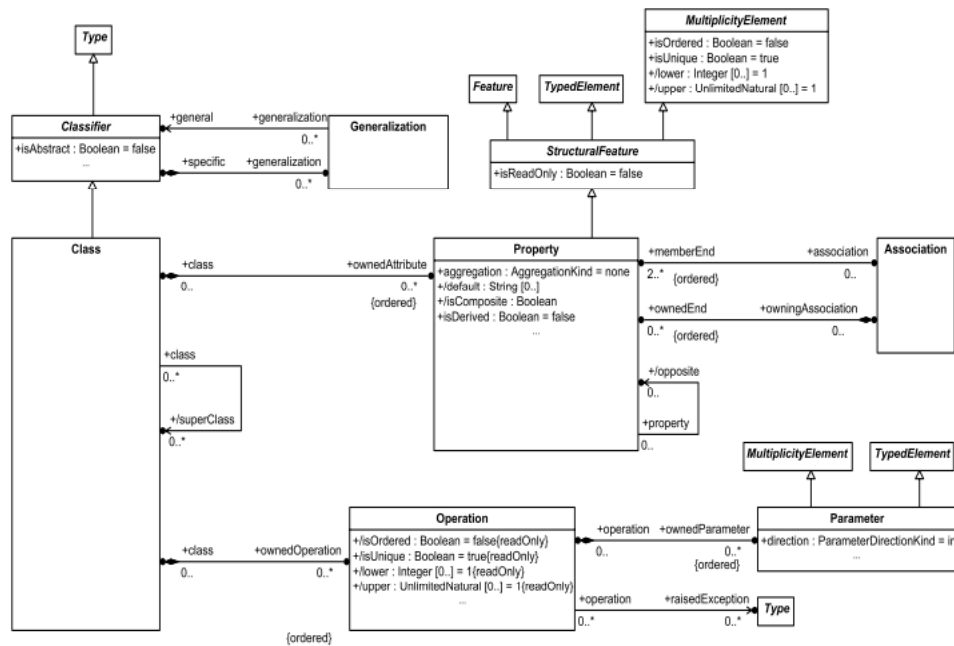
ΣΧΗΜΑ 2.4: EMOF Data Types



ΣΧΗΜΑ 2.5: EMOF Package



ΣΧΗΜΑ 2.6: EMOF Types



ΣΧΗΜΑ 2.7: EMOF Classes

### 2.2.2 Ανάλυση Πηγαίου Κώδικα

Για να αποκτηθεί πρόσβαση στις οντότητες του πηγαίου κώδικα οι οποίες είναι απαραίτητες για τους σκοπούς αυτής της εργασίας, έπρεπε να χρησιμοποιηθεί ένα εργαλείο το οποίο θα βοηθούσε στην εύρεση και συλλογή όλων των οντοτήτων αυτών καθώς και τον μεταξύ τους σχέσεων. Ευτυχώς κατά τη διάρκεια των περασμένων δεκαετιών η βιομηχανία της τεχνολογίας λογισμικού γνώρισε μία αλματώδη ανάπτυξη η οποία αυτή ανάπτυξη οδήγησε στη δημιουργία εργαλείων τα οποία προσφέρουν στους προγραμματιστές τη δυνατότητα να δημιουργούν μία αφαιρετική απεικόνιση των προγραμμάτων τους με στόχο να αποκτήσουν μία καλύτερη κατανόηση αυτών και να μπορούν έτσι πιο εύκολα να τα βελτιώνουν. Ένα από αυτά τα εργαλεία το οποίο χρησιμοποιήθηκε και στα πλαίσια αυτής της εργασίας είναι το Fact Extraction Tool-Chain [28].

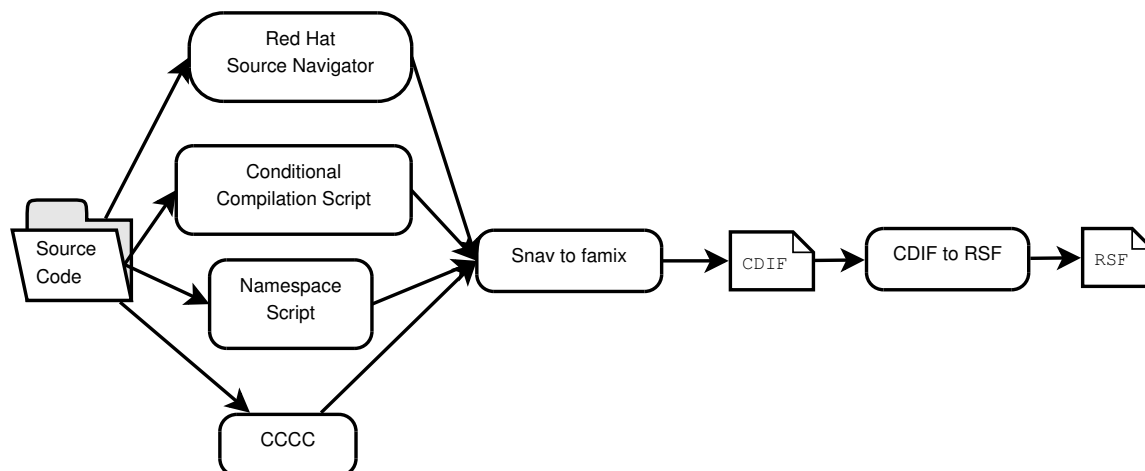
Αυτό το εργαλείο λαμβάνει ως είσοδο τον πηγαίο κώδικα ενός Java ή C/C++ προγράμματος και επιστρέφει κάποια αρχεία τα οποία επιτρέπουν :

- την εξαγωγή όλων των οντοτήτων του κώδικα.
- την εξαγωγή όλων των σχέσεων μεταξύ αυτών των οντοτήτων
- την απεικόνιση ενός προγράμματος με τη μορφή γράφου

**Fetch-** Το Fetch (Fact Extraction Tool CHain) [28] είναι μία σειρά εργαλείων για ανάλυση λογισμικού. Αποτελείται από αρκετά εργαλεία ανοιχτού κώδικα που στοχεύουν στην διερεύνηση μεγάλων C, C++ και Java συστημάτων λογισμικού. Είναι πρόγραμμα ανοιχτού κώδικα και προσφέρει αρκετά χρήσιμα εργαλεία. Το εργαλείο αυτό επιλέχθηκε επειδή προσφέρει εξαιρετικά κείμενα οδηγιών χρήσης τα οποία έκαναν πολύ εύκολη τόσο τη χρήση όσο και την επέκτασή του, και επίσης βοηθάει στην εύκολη και αυτοματοποιημένη εξαγωγή των 'δεδομένων' (Facts) από συστήματα C/C++, Java το οποίο επιτρέπει τη χρήση πιο ετερογενών συστημάτων για τα διεξαχθέντα πειράματα.

Η εικόνα 2.8 δείχνει πως το FETCH συνδυάζει διαφορετικά εργαλεία και εξάγει το επιθυμητό αποτέλεσμα. Οι λεπτομέρειες κάθε επιμέρους εργαλείου παρουσιάζονται παρακάτω.

Το FETCH παίρνει σαν είσοδο τον πηγαίο κώδικα του συστήματος που αναλύεται κάθε φορά. Ο πηγαίος κώδικας δεν χρειάζεται ούτε να είναι 'μεταφράσιμος' (compilable) ούτε να υποστεί κανενός είδους προεπεξεργασία. Το πρώτο εργαλείο που επεξεργάζεται τον πηγαίο κώδικα είναι το Source Navigator [29]. Αυτό το εργαλείο αναπτύχθηκε αρχικά από την Red Hat και είναι ένα εργαλείο για ανάλυση και κατανόηση κώδικα το οποίο προσφέρει επίσης και ένα Integrated Development Environment (IDE) για την κατανόηση και επανασχεδιασμό μεγάλων συστημάτων λογισμικού. Το Source Navigator πραγματοποιεί μία λεξικολογική



ΣΧΗΜΑ 2.8: Fact Extraction Tool Chain

ανάλυση στον κώδικα με την οποία εξάγει πληροφορία από C, C++, Java και άλλα συστήματα και μετά χρησιμοποιεί αυτήν την πληροφορία για να γεμίσει μία βάση δεδομένων για αυτό το σύστημα. Το εργαλείο προσφέρει και ένα γραφικό περιβάλλον που επιτρέπει την πλοήγηση σε αυτή τη βάση για την ανακάλυψη συγκεκριμένων συμβόλων όπως συναρτήσεις ή global μεταβλητές, καθώς και τις σχέσεις μεταξύ τους. Το Source Navigator δεν προσφέρει χαμηλού επιπέδου και αναλυτική περιγραφή του συστήματος υπό διερεύνηση το οποίο το κάνει ιδανικό για αρχιτεκτονική ανάλυση και το καθιστά εξαιρετικά γρήγορο.

Επιπλέον του Source Navigator, ένα άλλο εργαλείο, το C and C++ Code Counter, CCCC [30], χρησιμοποιείτε για να εξάγει πληροφορία που το Source Navigator δεν μπορεί να συμπεράνει από τη δομή του πηγαίου κώδικα όπως οι συνολικές γραμμές κώδικα, οι συνολικές γραμμές σχολίων και η κυκλωματική πολυπλοκότητα. Τέλος χρησιμοποιούνται δύο Bash Scripts φτιαγμένα από το εργαστήριο, το πρώτο για να εξάγει της εμβέλεια ορισμού μεταβλητών, και το άλλο για να εξάγει όλες τις συνθήκες 'κατά συνθήκη μεταγλώττισης'.

Τα δεδομένα τα οποία εξάγονται από αυτά τα εργαλεία συνδυάζονται και διασυνδέονται από το Snavtofamix, το οποίο πραγματοποιεί ερωτήσεις στη βάση που έχει δημιουργήσει το Source Navigator και επιλύει της αλληλοσυγκρουόμενες αναφορές όπως η κληρονομικότητα (inheritance), η κλήσεις (invocations), η προσβάσεις (accesses) κ.α. Το αποτέλεσμα του Snavtofamix είναι ένα αφηρημένο μοντέλο το οποίο είναι συμβατό με το Famix (FAMOOS [31] Information Exchange Model). Αυτή η αναπαράσταση του μοντέλου Famix χρησιμοποιεί την πρότυπο Case Data Interchange Format, CDIF [32]. Ο Πίνακας 2.4 απεικονίζει τα περιεχόμενα του μοντέλου Famix καθώς επίσης και το εργαλείο που εξάγει κάθε ένα από αυτά από τον πηγαίο κώδικα.

Το τελευταίο βήμα είναι η 'μετάφραση' του αφηρημένου μοντέλου Famix σε αναπαράσταση γράφου (RSF, Rigid Standard Format). Αυτός ο μετασχηματισμός υλοποιείται από το

| Entity                      | Origin                           |
|-----------------------------|----------------------------------|
| File, Include               | Source Navigator                 |
| Conditional Compilation     | Conditional Compilation Script   |
| Package                     | Namespace Script                 |
| Class, Inheritance, Typedef | Source Navigator                 |
| Method, Function            | Source Navigator                 |
| Attribute, Global Variable  | Source Navigator                 |
| Invocation, Access          | Source Navigator and Snavtofamix |
| Measurement                 | Source Navigator and CCCC        |

ΠΙΝΑΚΑΣ 2.4: Abstract model content

εργαλείο CDIF2RSF και είναι απαραίτητο για να διευκολυνθεί η διεξαγωγή ερωτήσεων στο αφηρημένο μοντέλο και για την εξαγωγή πολύ συγκεκριμένης πληροφορίας.

Το εργαλείο Fetch προσφέρει και κάποια ακόμα εργαλεία για οπτικοποίηση των αποτελεσμάτων και ερωτήσεις πάνω στον γράφο, όπως το Guess κάποια R scripts και το εργαλείο Crocopat, αλλά αυτά τα εργαλεία δεν απασχολούν αυτήν την εργασία και δεν θα παρατεθεί κάποια εις βάθος ανάλυση των.

Για να γίνουν λίγο πιο κατανοητά και απτά όσα παρουσιάστηκαν παραπάνω, παρουσιάζεται παρακάτω ένα παράδειγμα χρήσης του Fetch.

Στην παράθεση 2.1, δίνεται ένα μικρό C πρόγραμμα για το τρίγωνο Pascal.

```
#include <stdio.h>
long factorial(int);
int main(){
    int i, n, c;
    printf("Enter the number of rows to see in pascal triangle\n");
    scanf("%d",&n);

    for (i = 0; i < n; i++) {
        for (c = 0 ; c <= i; c++)
            printf("%ld ",factorial(i)/(factorial(c)*factorial(i-c)));
        printf("\n"); }
    return 0;
}
long factorial(int n){
    int c;
    long result = 1;

    for (c = 1; c <= n; c++)
        result = result*c;
    return result;
}
```

LISTING 2.1: Pascal Triangle

Αυτός ο πηγαίος κώδικας περνάει σαν είσοδος στο Fetch. Τα Source Navigator, CCCC και τα μικρά Bash scripts επεξεργάζονται την είσοδο και εισάγουν τα αποτελέσματα στη



βάση. Το εργαλείο Snavtofamix εκτελείται πάνω στα αποτελέσματα των προηγούμενων με αποτέλεσμα την δημιουργία του CDIF αρχείου, το οποίο παρουσιάζεται στη παράθεση 2.2.

Οι γραμμές 1–10 αντιπροσωπεύουν την επικεφαλίδα του μοντέλου και περιέχουν κάποιες πληροφορίες σχετικές με το πρόγραμμα όπως η γλώσσα στην οποία είναι γραμμένος ο πηγαίος κώδικας και το όνομα του συστήματος (το οποίο προκύπτει από το όνομα του top level φακέλου που περιέχει όλο το σύστημα και ο οποίος φάκελος είναι αυτός που περνάει σαν όρισμα στο Fetch). Αυτές τις πρώτες γραμμές ακολουθεί η δήλωση του αφηρημένου μοντέλου.

Οι γραμμές 12–16 αναπαριστούν ένα αρχείο κειμένου. Το πεδίο *uniqueName* δίνει το μοναδικό όνομα του αρχείου, το οποίο είναι διαφορετικό από το πεδίο *name* μόνο όταν υπάρχουν παραπάνω από ένα αρχεία με το ακριβώς ίδιο όνομα μέσα στο σύστημα, το οποίο εν προκειμένω δεν ισχύει. Αν υπήρχαν περισσότερα αρχεία κώδικα τότε αυτές οι γραμμές θα επαναλαμβάνονταν για κάθε ένα από αυτά.

Μετά στις γραμμές 18–22 δίνεται πληροφορία για τη βιβλιοθήκη που έχει γίνει include . Η πληροφορία δείχνει ποιο αρχείο συμπεριλαμβάνει ποιο αρχείο και σε ποια γραμμή του αρχικού αρχείου ορίζεται αυτό.

Οι γραμμές 24–30 και 39–45 σχετίζονται με τις δύο συναρτήσεις που έχουν οριστεί στον πηγαίο κώδικα, τις *main()* και *factorial()*, και δίνουν πληροφορίες για αυτές όπως το όνομα και την υπογραφή της συνάρτησης και τον τύπο επιστροφής της.

Επίσης επειδή η μέθοδος *factorial()* είναι ορισμένη μέσα στον φάκελο *pascal.c*, το CDIF περιέχει τις γραμμές 32–37 οι οποίες δηλώνουν σε ποιο σημείο του αρχείου έχει οριστεί αυτή η συνάρτηση.

Τέλος, η μέθοδος *main()* καλεί τη μέθοδο *factorial()* και για αυτό τον λόγο υπάρχει μία εγγραφή κλήσης, στις γραμμές 47–53. Η γραμμή 48 ορίζει μία συνάρτηση που προκαλεί την κλήση και η γραμμή 49 ορίζει μία συνάρτηση η οποία καλείται. Επιπροσθέτως, υπάρχει μία γραμμή για το που λαμβάνει χώρα αυτή η κλήση στη γραμμή 50, καθώς επίσης που ορίζονται αυτές οι δύο συναρτήσεις (γραμμές 51–52).

```
(:HEADER
  (:SUMMARY
    (ExporterName "snavtofamix")
    (ExporterVersion "1096")
    (ExporterDate "2015/05/27")
    (ExporterTime "14:44:24")
    (ParsedSystemName "pascalFolder")
    (SourceLanguage "C"))
)
(:MODEL
(SourceFile FM1
  (uniqueName "pascal.c")
  (name "pascal.c")
)
(Include FM2
  (includingFile "pascal.c")
  (includedFile "stdio.h")
  (sourceAnchor #[file "pascal.c" start 1 end 1]#)
)
(Function FM3
  (name "factorial")
  (signature "factorial(int)")
  (declaredReturnType "long")
  (declaredReturnClass "")
  (sourceAnchor #[file "pascal.c" start 3 end 3]#)
)
(FunctionDefinition FM4
  (name "factorial")
  (declaredBy "factorial(int)")
  (sourceAnchor #[file "pascal.c" start 26 end 26]#)
  (declSourceAnchor #[file "pascal.c" start 3 end 3]#)
)
(Function FM5
  (name "main")
  (signature "main()")
  (declaredReturnType "int")
  (declaredReturnClass "")
  (sourceAnchor #[file "pascal.c" start 5 end 5]#)
)
(Invocation FM6
  (invokedBy "main()")
  (invokes "factorial(int)")
  (sourceAnchor #[file "pascal.c" start 18 end 18]#)
  (sourceSourceAnchor #[file "pascal.c" start 5 end 5]#)
  (destinationSourceAnchor #[file "pascal.c" start 3 end 3]#)
)
)
```

LISTING 2.2: CDIF for Pascal Triangle

Όταν πλέον το CDIF έχει δημιουργηθεί μπορεί να μετατραπεί σε μορφή RSF ώστε να είναι πιο χρήσιμο. Τρέχοντας το CDIF2RSF script δημιουργείται ένα RSF αρχείο. Αυτό το αρχείο παρουσιάζεται στην παράθεση 2.3. Πρέπει να σημειωθεί ότι το CDIF περιέχει πληροφορία η οποία δεν περιέχεται στο RSF αρχείο, για παράδειγμα το γεγονός ότι η `stdlib.h` έχει συμπεριληφθεί `pascal.h`. Αυτό συμβαίνει επειδή οι στάνταρ βιβλιοθήκες συστήματος δεν συμπεριλαμβάνονται και δεν αναπαρίστανται στο RSF. Αν το `pascal.h` συμπεριλάμβανε κάποια άλλη βιβλιοθήκη ωστόσο, τότε αυτή θα είχε προκαλέσει τη δημιουργία μίας ακόμα γραμμής στο μοντέλο του συγκεκριμένου αρχείου.

|                     |                    |                    |
|---------------------|--------------------|--------------------|
| FileBelongsToModule | "pascal.c"#1       | "/"#M1             |
| DeclaredIn          | "factorial(int)"#3 | "pascal.c"#1       |
| DefinedIn           | "main()"#5         | "pascal.c"#1       |
| DeclaredIn          | "main()"#5         | "pascal.c"#1       |
| DefinedIn           | "factorial(int)"#3 | "pascal.c"#1       |
| Calls               | "main()"#5         | "factorial(int)"#3 |

LISTING 2.3: RSF for Pascal Triangle

Αν γίνει ένας προσεκτικός έλεγχος του αρχείου RSF, φαίνεται ότι η πληροφορία είναι δομημένη σε τριπλέτες της μορφής: *Relation Entity Entity*. Κάθε γραμμή αναπαριστά μία διαφορετική σχέση, δεν υπάρχουν διπλότυπες εγγραφές και οι Οντότητες ορίζονται μοναδικά τόσο από το όνομά τους όσο και από ένα μοναδικό αριθμητικό αναγνωριστικό.

Αυτό που παρουσιάστηκε είναι ένα πολύ απλό παράδειγμα με μόνο ένα αρχείο πηγαίου κώδικα το οποίο ορίζει μόλις δύο συναρτήσεις και έχει μόνο μία κλήση μεθόδου. Τα πραγματικά συστήματα λογισμικού περιέχουν εκατοντάδες ή χιλιάδες αρχεία πηγαίου κώδικα με χιλιάδες ή εκατομμύρια οντότητες και σχέσεις μεταξύ αυτών. Παρόλα αυτά η βασική ιδέα της ανάλυσης κώδικα παρουσιάζεται αρκετά παραστατικά με αυτό το παράδειγμα και είναι εύκολα κατανοητό πως αυτή επεκτείνεται για την παρουσίαση συστημάτων ανεξαρτήτως μεγέθους.

Τέλος, παρουσιάζονται στον Πίνακα 2.5 όλες οι σχέσεις η οποίες μπορούν να εξαχθούν από τον πηγαίο κώδικα χρησιμοποιώντας αυτά τα εργαλεία που παρουσιάστηκαν ως μέρη του εργαλείου Fetch. Κάθε σχέση παρουσιάζεται μαζί με μία σύντομη περιγραφή της σημασίας της. Υπάρχουν αρκετοί τύποι σχέσεων και όπως φάνηκε στο στάδιο των πειραμάτων δεν είναι όλες απαραίτητες για την προτεινόμενη μεθοδολογία εντοπισμού σφαλμάτων από αναφορές σφάλματος. Αυτές οι σχέσεις οι οποίες θεωρούνται πλεονασματικές για τους σκοπούς της παρούσας εργασίας θα παρουσιαστούν στο 6.

|                    | Relation                 | Description                                       |
|--------------------|--------------------------|---|
| <i>Containment</i> | Module Belongs to Module | States a containment relation between two modules |
|                    | File Belongs to Module   | States to which Module each File belongs          |
|                    | Class Belongs to File    | States to which File each Class or Struct belongs |

|                    |                                  |  |
|--------------------|----------------------------------|--|
|                    | Invocable Entity Belongs to File | States the Global Variables that belong to each File   |
|                    | Method Belongs to Class          | States the Methods that belong to each Class   |
|                    | Attribute Belongs to Method      | States the Attributes of each Method or Struct   |
| <i>Macros</i>      | Macro Definition                 | States the definition of a Macro as well as the file where it is defined   |
|                    | Macro Use                        | States in which File a Definition, that was previously defined, is used  |
| <i>Conditions</i>  | Conditional Compilation          | States under which conditions a block of code is compiled  |
| <i>Location</i>    | Entity Location                  | States in which File and from which to which line an Entity is located   |
|                    | Entity Belongs to Block          | States to which block of source code – which is conditionally compiled – an Entity belongs   |
| <i>Declaration</i> | Defined In                       | States in which file a function or a method is defined. This is the file where the actual body of the function or the method is located. We should underline that for every function and method there is only one <i>Defined In</i> relation in the output file of the extractor.              |
|                    | Declared In                      | This relation is similar to the one above, with one great difference. <i>Declared In</i> states in which file there is a declaration of the function or the method. As a result there can be more than one such relations for the same function or method in the output file of the extractor. |
| <i>Types</i>       | Type Definition                  | This relation states that a type name is associated with a class, a struct or another type name that was previously defined  |
|                    | Uses Type                        | This relation states which class or struct, not a primitive type, is used in a method or a function respectively   |

|                    |                     |   |
|--------------------|---------------------|---|
|                    | Has Type            | States the type of a function, method, attribute or global variable.  |
|                    | Has Type Definition | This relation combines the previous ones and associates a function, method, attribute or global variable with a type name |
| <i>Invocations</i> | Call                | Declares an invocation between two methods or functions   |
| <i>Information</i> | Signature           | States the signature of a method  |
|                    | Visibility          | States the visibility of a class, attribute or method when this is applicable   |
|                    | No of Lines         | States in which line of a File an Entity belong to  |
| <i>Other</i>       | Inherits From       | Represents the inheritance of classes in object-oriented programming  |
|                    | Include             | States which File is included in another File   |
|                    | Access              | States in which method or function, an attribute or a global variable is on the right side of an assignment statement     |
|                    | Set Variable        | States in which function or method an attribute or a global variable is on the left side of an assignment statement       |

ΠΙΝΑΚΑΣ 2.5: Relations that Fetch extracts

## 2.3 MongoDB

Οι σύγχρονες βάσεις δεδομένων απομακρύνονται σταδιακά από το κλασσικό σχεσιακό μοντέλο βάσης και προσανατολίζονται προς γρήγορες, κλιμακούμενου μεγέθους, λύσεις για διαχείριση δεδομένων. Αυτά τα νέα συστήματα βάσεων δεδομένων δεν χρησιμοποιούν πλέον ζεύγη τιμής - εγγραφής αλλά περιέχουν δομημένο κείμενο. Η χρήση της SQL αποφεύγεται για την πραγματοποίηση ερωτήσεων στα δεδομένα, από όπου προκύπτει και η ονομασία NoSQL βάσεις δεδομένων. Αυτές οι βάσεις προσφέρουν στους προγραμματιστές έναν πολύ πιο διασθητικά σωστό τρόπο για την αποθήκευση και πρόσβαση στα δεδομένα τους. Αυτό συμβαίνει γιατί:

- δεν υπάρχει πάντα ένας τρόπος μετασχηματισμού των δεδομένων σε ένα σχεσιακό σχήμα βάσης.

- οι βάσεις δεδομένων αποθήκευσης κειμένου μπορούν να αποθηκεύσουν στον ίδιο Πίνακα οποιαδήποτε μορφή και τύπο δεδομένων.

Για εκπλήρωση των στόχων της παρούσας εργασίας επιλέχθηκε η χρήση της βάσης MongoDB η οποία είναι μία ελαφριά, κλιμακούμενου μεγέθους υλοποίηση η οποία αποθηκεύει την πληροφορία σε μορφή JSON.

Η MongoDB (from huMongous) είναι μία κείμενο-κεντρική βάση δεδομένων η οποία είναι ανεξάρτητη πλατφόρμας και προσδιορίζεται ως μία NoSQL βάση. Η MongoDB απορρίπτει το παραδοσιακό σχεσιακό μοντέλο των βάσεων προς όφελος ενός προτύπου που μοιάζει με το JSON και χρησιμοποιεί δυναμικό σχήμα για τα αποθηκευόμενα δεδομένα. Είναι μία βάση διαθέσιμη δωρεάν και καλύπτεται από την GNU Affero GPL και την Apache License. Αρχικά δημοσιεύτηκε από την ομώνυμη εταιρεία τον Οκτώβριο του 2007 ως τμήμα μίας πλατφόρμας που σχεδίαζόταν τότε. Η εταιρεία άλλαξε το πρότυπο εργασίας της σε ανοιχτού λογισμικού το 2009, διαθέτοντας δωρεάν τα προϊόντα της και παρέχοντας υποστήριξη για αυτά και άλλες υπηρεσίες.

Τα κύρια χαρακτηριστικά του συστήματος αυτού που ώθησαν στην επιλογή του για την εργασία αυτή ως του ιδανικού εργαλείου για την διατήρηση των δεδομένων είναι τα κάτωθι:

- είναι κείμενο-κεντρική
- επιτρέπει την διεξαγωγή περίεργων ερωτήσεων
- υποστηρίζει Indexing
- υποστηρίζει Replication
- υποστηρίζει Load balancing
- επιτρέπει την αποθήκευση αρχείων
- υποστηρίζει Aggregation

Όλα αυτά τα χαρακτηριστικά ήταν σημαντικοί παράγοντες λόγω της φύσης και του μεγέθους του προβλήματος και των δεδομένων που σχετίζονται με αυτό. Επειδή αυτή η βάση χρησιμοποιείται για την αποθήκευση εξαιρετικά μεγάλων ποσοτήτων δεδομένων τα οποία προκύπτουν από διάφορες συλλογές αναφορών σφάλματος, και είναι επιθυμητό να υπάρχει η δυνατότητα να μπορεί η βάση να αποθηκεύσει οσοδήποτε πολλά δεδομένα διατηρώντας ταυτόχρονα την ταχύτητα της.

### 2.3.1 JSON

Η μορφοποίηση δεδομένων σύμφωνα με το JSON (JavaScript Object Notation) πρότυπο σημαίνει την μετατροπή των δεδομένων σε μία μορφή αναγνώσιμη από άνθρωπο, η οποία μπορεί ταυτοχρόνως είναι δυνατό να χρησιμοποιηθεί για ανταλλαγή δεδομένων μεταξύ απομακρυσμένων μηχανημάτων, όπως η XML (eXtended Markup Language) και ομοίως με την τελευταία αποτελείται από κείμενα που περιέχουν πληροφορία οργανωμένη σε ζεύγη κλειδιού-τιμής. Το πρότυπο αυτό χρησιμοποιείται πολύ για ασύγχρονη επικοινωνία του διαδικτυακού φυλλομετρητή και του διακομιστή μέσω της τεχνολογίας AJAX για τη μεταφορά αντικειμένων δεδομένων. Επιλέχθηκε γιατί υπάρχει διαθέσιμη μία πολύ καλή βιβλιοθήκη για την προετοιμασία, αποθήκευση και ανάκτηση των δεδομένων στη γλώσσα Java και επειδή είναι πιο ευανάγνωστη από τη μορφή XML. Επίσης ήταν αρκετά εύκολη η υλοποίηση κάποιων επιπλέον λειτουργικότητων οι οποίες μετέτρεπαν κάποιες εξωτικές δομές σε αυτό το πρότυπο αναπαράστασης.

## 2.4 Συνοπτική Αναπαράσταση Κειμένων

### 2.4.1 Διανύσματα TF-IDF

Ο δείκτης 'συχνότητα όρου - αντίστροφη συχνότητα κειμένου' είναι ένας αριθμός ο οποίος δείχνει πόσο σημαντικός είναι ένας όρος που βρίσκεται σε κάποιο αρχείο το οποίο ανήκει με τη σειρά του σε μία συλλογή αρχείων. Η τιμή TF-IDF αυξάνεται ανάλογα με των αριθμό των εμφανίσεων ενός όρου σε ένα αρχείο, αλλά η αύξηση αυτή περιορίζεται ή εξομαλύνεται από το πλήθος των αρχείων της συλλογής στα οποία εμφανίζεται ο όρος. Αυτό σημαίνει ότι αν μία λέξη-όρος είναι σχετικά σπάνιος στη συλλογή και βρεθεί σε ένα αρχείο, τότε αυτό το αρχείο μπορεί πιθανώς να χαρακτηριστεί από την παρουσία αυτού του όρου.

#### 2.4.1.1 Ορισμός Συχνότητας Όρου

Η συχνότητα όρου είναι ο αριθμός που ένας όρος εμφανίζεται σε ένα κείμενο διαιρεμένος με το σύνολο των όρων του κειμένου.

#### 2.4.1.2 Αντίστροφη Συχνότητα Κειμένου

Η Αντίστροφη Συχνότητα Κειμένου ενός όρου υπολογίζεται για την συλλογή αρχείων χρησιμοποιώντας τη σχέση:

$$\frac{\log(N_{docs})}{F_{doc} + 1} + 1$$

όπου  $N_{docs}$  είναι ο συνολικός αριθμός αρχείων/κειμένων και  $F_{doc}$  είναι το πλήθος των αρχείων που περιέχουν αυτόν τον όρο.

### 2.4.1.3 Η τιμή TF-IDF

Η τιμή της συχνότητας όρου πολλαπλασιάζεται με την αντίστροφη συχνότητα κειμένου και για κάθε όρο κάθε αρχείου αποδίδει την τιμή του δείκτη TF-IDF. Μετά για κάθε αρχείο δημιουργείται ένας πίνακας που περιέχει όλους τους όρους που υπάρχουν στη συλλογή αρχείων, αλλά για κάθε όρο που δεν εμφανίζεται στο αρχείο το αντίστοιχο κελί θα έχει τιμή 0. Με αυτόν τον τρόπο κάθε όρος μετατρέπεται σε 'διάσταση' ενός χώρου  $N$  διαστάσεων.

## 2.5 Meta Object Facility και EMF

Ένα εννοιολογικό μοντέλο είναι ένα μοντέλο το οποίο δομείται από την σύνθεση εννοιών, το οποίο χρησιμοποιείται για να βοηθήσει ανθρώπους να μάθουν, να καταλάβουν ή να προσομοιώσουν το θέμα το οποίο αναπαρίσταται από το μοντέλο. Ο όρος εννοιολογικό μοντέλο μπορεί να χρησιμοποιηθεί για να γίνει αναφορά σε μοντέλα τα οποία έχουν δημιουργηθεί από την εξαγωγή των εννοιών (Conceptualisation) ή τη γενίκευση (Generalisation) μίας διαδικασίας. Αυτά τα μοντέλα είναι συχνά αφαιρετικές αναπαραστάσεις αντικειμένων του πραγματικού κόσμου είτε αυτά είναι φυσικά αντικείμενα ή αφηρημένες έννοιες.

Αρκετά εργαλεία έχουν παρουσιαστεί τις τελευταίες δεκαετίες για τη δημιουργία τέτοιων μοντέλων. Παρόλο την ύπαρξη της UML, είναι δυστυχώς συχνά δύσκολο να χρησιμοποιηθεί ένα εργαλείο, κυρίως γιατί τα μοντέλα που δημιουργούνται χρησιμοποιώντας το είναι 'νεκρά' υπό την έννοια ότι δεν γίνεται ο δημιουργός να τα αξιοποιήσει ή να αλληλεπιδράσει με αυτά. Κάποιες από τις προσεγγίσεις που έχουν προταθεί για τη δημιουργία τέτοιων μοντέλων είναι το Telos [33] [34] τα οποία επιτρέπουν στον χρήστη να 'τρέξει' το μοντέλο που δημιουργεί αν φυσικά διαθέτει κάποιον μεταφραστή για τον κώδικα Telos που θα γράψει, και το KAOS [35] το οποίο αν και σχεδιάστηκε κυρίως για μοντελοποίηση απαιτήσεων, επιδεικνύει υψηλό δυναμικό στην μοντελοποίηση διεργασιών και όπως και το Telos διατίθενται κάποιες υλοποιήσεις του που επιτρέπουν την αλληλεπίδραση με το σχεδιασμένο μοντέλο.

Παρά την επιτυχία των δύο προαναφερθέντων εργαλείων μοντελοποίησης, το σύστημα που προτιμήθηκε για τη δόμηση του μοντέλου αναπαράστασης αναφορών σφάλματος, είναι το Eclipse Modelling Framework (EMF) το οποίο είναι συμβατό με το Meta Object Facility όπως αυτό παρουσιάστηκε στην σχετική παράγραφο. Το πιο ενδιαφέρον αυτού του συστήματος είναι ότι εκτός του γεγονότος ότι υλοποιεί ένα υπερσύνολο λειτουργιών σε σχέση με τις απαιτούμενες για συμβατότητα με το μοντέλο MOF, παρέχει στον χρήστη τη δυνατότητα να μεταφράσει compile το σχεδιασμένο μοντέλο και να δημιουργήσει για αυτό μέσω



του Eclipse κώδικα Java ο οποίος αντιστοιχεί απευθείας στο σχεδιασμό που έχει πραγματοποιηθεί και επιτρέπει έτσι στον προγραμματιστή/σχεδιαστή να χρησιμοποιήσει το μοντέλο που έχει φτιάξει, να δημιουργήσει πραγματικά αντικείμενα για να εισάγει λειτουργικότητα στο μοντέλο του και μετά να τα αποθηκεύσει κιάλας στον σκληρό του δίσκο. Όλες αυτές οι λειτουργίες το EMF παρέχονται μέσω της πλατφόρμας Eclipse με πολύ μικρό κόστος εκμάθησης σε ώρες.[27] [36].

Το Eclipse Modelling Framework κατέστησε εφικτή τη δημιουργία ενός αφηρημένου εννοιολογικού μοντέλου για αναφορές σφάλματος. Αυτό το μοντέλο δεν δημιουργήθηκε με στόχο να χρησιμοποιηθεί κάπως στα πλαίσια της παρούσας εργασίας, αλλά αποδείχθηκε πολύ χρήσιμο στην προσπάθεια για κατανόηση της δομής των αναφορών σφάλματος, και του πως αυτές σχετίζονται μεταξύ τους.

## 2.6 Natural Language Processing

Προκειμένου να μετατραπούν τα δεδομένα όπως αυτά εξάγονται από τα Bugzilla repositories έπρεπε να υποστούν κάποια επεξεργασία προκειμένου να παρακρατηθούν μόνο οι χρήσιμοι όροι, και να απορριφθούν οι άχρηστοι. Άχρηστοι θεωρήθηκαν οι όροι όπως stop-words και άρθρα καθώς επίσης και άλλες πολύ συχνές λέξεις που θεωρούνται θόρυβος. Για να επιτευχθεί αυτού του είδους η επεξεργασία αξιοποιήθηκε η βιβλιοθήκη για επεξεργασία φυσικής γλώσσας που είναι δωρεάν διαθέσιμη από τον οργανισμό Apache, OpenNLP. Αυτή η βιβλιοθήκη είναι φτιαγμένη στη γλώσσα Java και επιτρέπουν στον χρήστη να διεκπεραιώσει ένα πλήθος διεργασιών ανάμεσα στις οποίες συμπεριλαμβάνονται και κάποιες πολύ χρήσιμες για την επεξεργασία των αναφορών. Συγκεκριμένα δίνεται η δυνατότητα για POS(Part of Speech) Tagging,αφαίρεση stop-words, εντοπισμό προτάσεων, και εύρεση ρίζας λέξεων. Στην πράξη χρησιμοποιήθηκαν όλες αυτές οι λειτουργίες καθώς παρατηρήθηκε ότι οι αρχικές αναφορές περιείχαν κατά μεγάλο μέρος θόρυβο οποίος έπρεπε να φιλτραριστεί. Περισσότερα για το NLP θα παρουσιαστούν στο Κεφάλαιο 5.

## 2.7 Αλγόριθμοι Εξόρυξης Δεδομένων - LSI, PLSI, LDA

Τα τελευταία χρόνια ο όρος Εξόρυξη Δεδομένων (Data Mining) έχει περάσει από το στάδιο της συχνά χρησιμοποιούμενης λέξης μαγνήτη, σε πολλές τεχνικού περιεχομένου ομιλίες, ανασκοπήσεις και δημοσιεύσεις, στο στάδιο όπου έχει αρχίσει να δημιουργεί τη δική του ξεχωριστή κοινότητα μέσα στην μεγαλύτερη κοινότητα της πληροφορικής. Αυτό το τύπος νέο πεδίο της Εξόρυξης Δεδομένων δεν είναι και τόσο νέο. Όλες οι τεχνικές που χρησιμοποιούνται υπό τη σκέπη αυτού του όρου είναι αλγόριθμοι και τεχνικές οι οποίες ήταν

ήδη γνωστές στην κοινότητα της πληροφορικής για πολλά χρόνια. Αυτό που ουσιαστικά όμως ορίζει ο όρος Εξόρυξη Δεδομένων είναι το μέγεθος των δεδομένων στα οποία εφαρμόζονται αυτές οι τεχνικές πλέον. Αυτό συμβαίνει γιατί πλέον είναι διαθέσιμες συλλογές αποτελούμενες από δισεκατομμύρια και τρισεκατομμύρια Bytes που πρωτύτερα δεν υπήρχαν, καθώς επίσης υπάρχει η δυνατότητα εφαρμογής των τεχνικών πιο αποδοτικά αξιοποιώντας υπολογιστικά σχήματα όπως το MapReduce για την παραλληλοποίηση των υπολογισμών. Μπορεί κανείς να βρει αρκετούς διαφορετικούς ορισμούς για την *Εξόρυξη Δεδομένων*. Δυστυχώς δεδομένου ότι πρόκειται ακόμα για μία πολύ δημοφιλή λέξη στα τεχνικά κείμενα, και το πεδίο ακόμα ορίζεται, θα γίνει μια προσπάθεια ορισμού χωρίς τη χρήση κάποιου έτοιμου ορισμού. Η Εξόρυξη Δεδομένων είναι η διεργασία εξαγωγής μοτίβων από κρυμμένη γνώση που βρίσκεται 'θαμμένη' μέσα σε μεγάλες συλλογές δεδομένων, χρησιμοποιώντας αυτοματοποιημένους τρόπους. Οι Αλγόριθμοι εξόρυξης Δεδομένων μπορούν να διαχωριστούν σε τουλάχιστον δύο κατηγορίες, τους αλγόριθμους ομαδοποίησης και κατηγοριοποίησης.

### 2.7.1 Ομαδοποίηση Clustering

Η ομαδοποίηση(Clustering) είναι μία διεργασία εξόρυξης δεδομένων, η οποία δεδομένου ενός συνόλου δεδομένων, πραγματοποιεί κάποιες ενέργειες στοχεύοντας στην ομαδοποίηση των μελών της συλλογής, όπου τα μέλη κάθε ομάδας που δημιουργείται να έχουν μικρή απόσταση από ένα κοινό 'κέντρο', και εξαρτάται από τον τρόπο μέτρησης της απόστασης ο τρόπος με τον οποίο θα προσδιοριστεί ο τύπος της εγγύτητας που θα έχουν τα μέλη ενός cluster.

### 2.7.2 Κατηγοριοποίηση

Η Κατηγοριοποίηση (Classification) μοιάζει αρκετά με το clustering, με τη καίρια διαφορά ότι η κατηγοριοποίηση κατηγοριοποιεί νέες παρατηρήσεις στη σωστή, ήδη υπάρχουσα, κατηγορία (σύνολο. Κατά συνέπεια για τη χρήση αλγορίθμων κατηγοριοποίησης πρέπει να υπάρχει διαθέσιμη κάποια μετρική απόστασης και μία αρχική συλλογή δεδομένων τα οποία να έχουν ήδη ανατεθεί στη σωστή κατηγορία ώστε να εκπαιδευτεί ο αλγόριθμος κατηγοριοποίησης, και να χρησιμοποιηθεί μετέπειτα για την κατηγοριοποίηση νέων παρατηρήσεων.

Ακολουθεί μία εν τάχει παρουσίαση κάποιων αλγορίθμων κατηγοριοποίησης.

#### 2.7.2.1 LSI

Η τεχνική Latent Semantic Indexing γνωστή και σαν Latent Semantic Analysis είναι μία τεχνική για δεικτοδότηση και ανάκληση η οποία χρησιμοποιεί μία μέθοδο της γραμμικής

|               |  |
|---------------|--|
| Binary        | $l_{ij} = 1$ if the term exists in the document, or else 0                 |
| TermFrequency | $l_{ij} = tf_{ij}$ , the number of occurrences of term $i$ in document $j$ |
| Log           | $l_{ij} = \log(tf_{ij} + 1)$   |
| Augnorm       | $l_{ij} = \frac{\left(\frac{tf_{ij}}{\max_i(tf_{ij})}\right) + 1}{2}$      |

ΠΙΝΑΚΑΣ 2.6: LSI common local weighting functions

|         |  |
|---------|--|
| Binary  | $g_i = 1$  |
| Normal  | $g_i = \frac{1}{\sqrt{\sum_j tf_{ij}^2}}$  |
| GfIdf   | $g_i = gf_i/df_i$ , where $gf_i$ is the total number of times term $i$ occurs in the whole collection, and $df_i$ is the number of documents in which term $i$ occurs. |
| Idf     | $g_i = \log \frac{n}{1+df_i}$  |
| Entropy | $g_i = 1 + \sum_j \frac{p_{ij} \log p_{ij}}{\log n}$ where $p_{ij} = \frac{tf_{ij}}{gf_i}$   |

ΠΙΝΑΚΑΣ 2.7: LSI common global weighting functions

άλγεβρας, την αποσύνθεση μοναδικής τιμής για να εντοπίσει μοτίβα συσχέτισης ανάμεσα στα δεδομένα που παρέχονται σαν είσοδος. Αυτή η τεχνική βασίζεται στην αρχή ότι οι λέξεις που χρησιμοποιούνται στο ίδιο πλαίσιο αναφοράς τείνουν να έχουν παραπλήσιες σημασίες. Το κύριο προτέρημα της μεθόδου είναι η δυνατότητά της να εξάγει το εννοιολογικό περιεχόμενο από συλλογές ελευθέρου κειμένου, εντοπίζοντας τις συσχετίσεις μεταξύ αυτών των όρων που εμφανίζονται σε παρεμφερή εννοιολογικά πλαίσια.

Η LSI δημιουργεί πρώτα έναν πίνακα όρων-κειμένων,  $A$ , για να έχει πρόσβαση στις εμφανίσεις  $m$  των μοναδικών όρων μέσα σε μία συλλογή από  $n$  κείμενα. Στον πίνακα αυτό κάθε όρος αντιστοιχεί σε μία γραμμή, και κάθε κείμενο σε μία στήλη, όπου κάθε κελί αρχικά αναπαριστά το πλήθος των φορών που αυτός ο όρος εμφανίστηκε στο κάθε κείμενο. Αυτός ο πίνακας είναι συνήθως πολύ μεγάλος και πολύ αραιός.

Όταν ο πίνακας όρων-κειμένων κατασκευαστεί, εφαρμόζονται πάνω του local και global συναρτήσεις βαθμονόμησης για να εξομαλύνουν τα δεδομένα. Οι συναρτήσεις αυτές αλλάζουν το περιεχόμενο κάθε κελιού ώστε να είναι το γινόμενο του βάρους του όρου, ο οποίος περιγράφει την σχετική συχνότητα του όρου στο κείμενο, και ενός ολικού βάρους το οποίο περιγράφει τη σχετική συχνότητα του όρου μέσα σε όλη τη συλλογή των κειμένων. Συχνά χρησιμοποιούμενες συναρτήσεις είναι οι κάτωθι:

### 2.7.2.2 PLSA

Η Probabilistic latent semantic analysis (PLSA) επίσης γνωστή με την ονομασία probabilistic latent semantic indexing (PLSI, ειδικά στο περιβάλλον της ανάκτησης πληροφορίας) είναι μία στατιστική τεχνική για την ανάλυση two-mode και co-occurrence δεδομένων. Στην

πράξη, μπορεί κανείς να παράγει μία χαμηλότερης διάστασης αναπαράσταση των παρατηρηθέντων τιμών από την συσχέτισή τους με κάποιες κρυφές μεταβλητές όπως και στην LSA από την οποία έχει προκύψει η PLSA. Συγκρινόμενη με την απλή LSA η οποία προκύπτει από μεθόδους γραμμικής άλγεβρας και μειώνει το μέγεθος των Πινάκων εμφανίσεων μέσω SVD, η πιθανοτική παραλλαγή βασίζεται σε ένα μείγμα αποσύνθεσης που προκύπτει από ένα αφανές μοντέλο κλάσης.

### 2.7.2.3 LDA

Στην επεξεργασία φυσικής γλώσσας NLP, η τεχνική Latent Dirichlet Allocation (LDA) είναι ένα παραγωγικό μοντέλο το οποίο επιτρέπει σε σύνολα παρατηρήσεων να εξηγηθούν από μη παρατηρήσιμα σύνολα τα οποία εξηγούν κάποια σημεία των δεδομένων που είναι παρεμφερή. Για παράδειγμα, αν οι παρατηρήσεις είναι λέξεις μέσα σε κείμενα, ο αλγόριθμος προτείνει ότι κάθε κείμενο είναι ένα μείγμα ενός μικρού συνόλου θεμάτων και ότι ύπαρξη ή δημιουργία κάθε λέξεις οφείλεται και μπορεί να συνδεθεί με κάποιο από αυτά τα θέματα. Στην LDA κάθε κείμενο μπορεί να αναπαρασταθεί σαν μία συλλογή θεμάτων. Αυτό είναι παρεμφερές με αυτό που συμβαίνει στην probabilistic Latent Semantic Analysis (pLSA), με την εξαίρεση ότι στην LDA οι κατανομή των θεμάτων υποτίθεται ότι διαθέτει μία Dirichlet prior. Στην πράξη αυτό οδηγεί σε πιο 'λογικές' συλλογές θεμάτων για τα κείμενα.

## Κεφάλαιο 3

# Προτεινόμενη Μεθοδολογία και Υλοποιημένες Λειτουργίες

### 3.1 Εργαλεία Συλλογής και Προεπεξεργασίας Δεδομένων

#### 3.1.1 Bugzilla Extraction Tool

##### 3.1.1.1 XMLRPC API Client

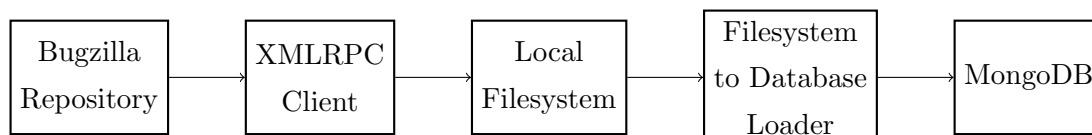
Για την επίτευξη των στόχων αυτής της εργασίας ήταν απαραίτητο να υλοποιηθούν κάποια βοηθητικά εργαλεία. Εδώ αξίζει να σημειωθεί ότι κατά την εκπόνηση αυτής της εργασίας έγινε συγγραφή περισσότερων των 40 χιλιάδων γραμμών κώδικα Java για την υλοποίηση των απαραίτητων εργαλείων και των προγραμμάτων που υλοποιούσαν διάφορες τεχνικές.

Το πρώτο πρόγραμμα που έπρεπε να συγγραφεί ήταν ένα πρόγραμμα για την ανάσυρση των αναφορών σφάλματος από οποιοδήποτε Bugzilla Repository. Για να γίνει αυτό επί της αρχής έλαβε χώρα μία προσπάθεια για χρήση του πιο σύγχρονου REST(Representational State Transfer) API το οποίο προστέθηκε στη βιβλιοθήκη του Bugzilla στην έκδοση 5.0. Δυστυχώς αυτό το σχέδιο απέτυχε (αν και η υλοποίηση ολοκληρώθηκε) διότι ο μόνος οργανισμός που είχε εγκατεστημένη την τελευταία έκδοση ήταν το Mozilla Foundation. Επόμενη επιλογή ήταν ανάμεσα στο JSONRPC API και το XMLRPC API. Αν και το JSONRPC API θα ήταν πιο εύχρηστο δεδομένης και της χρήσης μίας JSON βάσης δεδομένων, τελικά επιλέχθηκε η υλοποίηση ενός XMLRPC προγράμματος πελάτη. Ο λόγος για αυτήν την επιλογή ήταν ότι παρά το γεγονός ότι το JSONRPC API είναι εγκατεστημένο σε αρκετούς διακομιστές αρκετών οργανισμών, πολλές από τις διαθέσιμες μεθόδους είναι επισημειωμένες ως ασταθής ή πειραματικές από τους προγραμματιστές του Mozilla .

Συνέπεια αυτού το XMLRPC API, το οποίο είναι το παλιότερο διαθέσιμο πρότυπο και το οποίο είναι αρκετά σταθερό, επιλέχθηκε για χρήση στα πλαίσια αυτής της εργασίας και των πειραμάτων που διεξήχθησαν για την εξαγωγή αποτελεσμάτων, παρά τον επιπλέον κόπο για μετατροπή των δεδομένων σε μορφή JSON για χρήση με την βάση MongoDB.

Ο τρόπος που δουλεύει η υλοποίηση είναι ο ακόλουθος. Αρχικά δίνεται σαν είσοδος το URL μιας εγκατάστασης του συστήματος Bugzilla από τον οποίο πρέπει να τραβηχτούν οι αναφορές σφάλματος. Το επόμενο βήμα είναι η εισαγωγή του εύρος των μοναδικών αναγνωριστικών των αναφορών που θα πρέπει να κατεβούν.

Η διαδικασία για την πραγματοποίηση μίας απομακρυσμένης κλήσης συνάρτησης σε έναν φορτωμένο με κίνηση διακομιστή είναι μία αρκετά χρονοβόρος διαδικασία. Με στόχο την παράκαμψη αυτού του προβλήματος, επιλέχθηκε σαν λύση η τροποποίηση του πελάτη ώστε να υλοποιεί το Thread API της Java. Όπως έγινε αντιληπτό ωστόσο παρόλο που όπως είναι γνωστό η παράλληλοποίηση σαν έννοια και πρακτική όντως επιταχύνει όποια διαδικασία την υποστηρίζει, στην συγκεκριμένη περίπτωση αντιμετωπίστηκε το πρόβλημα ότι η παράλληλη υλοποίηση έστειλε πάρα πολλά αιτήματα ταυτόχρονα στον απομακρυσμένο διακομιστή. Το πρόβλημα με αυτή τη συμπεριφορά είναι ότι αυτή προσομοιάζει με επίθεση απάρνησης εξυπηρέτησης Denial Of Service Attack. Μέχρι να γίνει αυτό αντιληπτό είχε πραγματοποιηθεί ήδη μία επίθεση στους διακομιστές του Gnome με αποτέλεσμα να απαγορευτεί μόνιμα η πρόσβαση στους διακομιστές τους από τη διεύθυνση του εργαστηρίου. Για να αντιμετωπιστεί αυτή η απειλή προστέθηκε μία καθυστέρηση ανάμεσα σε κάθε αποστολή αιτήματος ούτως ώστε να φαίνεται το πρόγραμμα σαν ένας πολύ επίμονος αλλά φυσιολογικός (και όχι κακόβουλος) χρήστης. Τελικώς τα όποια οφέλη μπορούσαν να αποκομιστούν από την παράλληλη υλοποίηση εξανεμίστηκαν, ωστόσο η υλοποίηση είναι ακόμα παράλληλη οπότε σε περίπτωση που κάποιος οργανισμός θέλει να δοκιμάσει το πρόγραμμα αυτό σαν μέρος της πλατφόρμας μπορεί να επιτρέψει στους διακομιστές του την απάντηση οσονδήποτε αιτημάτων από το συγκεκριμένο πρόγραμμα χωρίς πρόβλημα.



ΣΧΗΜΑ 3.1: Bugzilla Extraction Tool Data Flow

### 3.1.1.2 Φορτωτής MongoDB

Οποιαδήποτε συλλογή δεδομένων είναι μάλλον άχρηστη στον ιδιοκτήτη της αν δεν είναι εξασφαλισμένη η εύκολη πρόσβαση σε αυτή. Για αυτόν τον λόγο έπρεπε να ληφθεί μία μάλλον περίπλοκη απόφαση. Να επιλεγθεί πως θα αποθηκευτούν οι αναφορές σφάλματος.

Αυτό το πρόβλημα δεν ήταν εύκολο να απαντηθεί με κάποιον άμεσο και μονοσήμαντα σωστό τρόπο. Ωστόσο, όπως έχει ήδη αναφερθεί επιλέχθηκε η χρήση της κείμενο-κεντρικής βάσης MongoDB. Η επιλογή δεν ήταν εύκολη όμως καθώς υπάρχουν αρκετά διαθέσιμα συστήματα διαχείρισης βάσεων δεδομένων και αρκετά από αυτά χρειάστηκαν αρκετή σκέψη ώστε να απορριφθούν τελικά. Βασικά η αντιμετώπιση μεγάλου όγκου δεδομένων τα οποία δεν έχουν καμία προφανή δομή είναι μία αρκετά δύσκολη δουλειά. Αρχικά για την αντιμετώπιση του ζητήματος υπήρξε η ιδέα χρήσης μίας παραδοσιακής SQL βάσης η οποία ευτυχώς απορρίφθηκε άμεσα. Οι λόγοι για αυτήν την άμεση απόρριψη ήταν:

- Απουσία ενός concrete schema.
- Απουσία σταθερού γνωστού εκ των προτέρων μήκους πεδίων.
- Ανομοιογένεια των διαθέσιμων πεδίων ανάμεσα σε διαφορετικές εγκαταστάσεις του Bugzilla .
- Ακόμα και με ένα ελάχιστο πλήθος από καθολικά χρησιμοποιούμενα πεδία, ο μετασχηματισμός αυτών σε ένα πρότυπο αποδεκτό από μία SQL Database θα ήταν μη αποδοτικός

Όταν η σχεσιακές βάσεις απορρίφθηκαν, αποφασίστηκε να γίνει έλεγχος για το κατά πόσο θα μπορούσε να χρησιμοποιηθεί κάποια NoSQL βάση. Για την ακρίβεια το ενδιαφέρον της ομάδας τράβηξαν δύο αρκετά παρεμφερείς βάσεις, η MongoDB και η CouchDB. Και η δύο έχουν τα ίδια προσόντα όπως αυτά που αναφέρονται στο MongoDB, το χαρακτηριστικό όμως το οποίο συνέβαλε στην απόρριψη της CouchDB και την χρήση τελικώς της MongoDB ήταν η ευκολία χρήσης της δεύτερης η οποία τουλάχιστον για τον προγραμματιστή αναφέρθηκε ότι ήταν εξαιρετικά πιο απλή από της CouchDB.

### 3.1.2 Εργαλεία Προεπεξεργασίας Αναφορών Σφάλματος

Έχοντας κάνει ήδη χρήση του Bugzilla Extraction Tool οι αναφορές σφάλματος θα πρέπει να περάσουν κάποια επεξεργασία καθώς στην αρχική τους μορφή περιέχουν μεγάλα ποσά θορύβου. Για να πραγματοποιηθεί αυτή η προεπεξεργασία υλοποιήθηκαν δύο ακόμα εργαλεία. Το πρώτο δημιουργήθηκε για να 'τραβάει' αναφορές από τη βάση, να διεξάγει κάποιες διεργασίες σχετικές με επεξεργασία φυσικής γλώσσας, και να προωθεί τα αποτελέσματα στο επόμενο εργαλείο. Το δεύτερο εργαλείο λαμβάνει σαν είσοδο την έξοδο του πρώτου και τα αποτελέσματα του Fact Extraction Tool Chain και εξάγει ένα σύνολο από σύνολα για κάθε αναφορά όπου το κάθε ένα περιέχει μόνο οντότητες του κώδικα που βρίσκονταν στην αναφορά, χωρισμένες σε δύο ομάδες, συναρτήσεις και μη-συναρτήσεις.

### 3.1.2.1 Αναζητητής Οντοτήτων Κώδικα

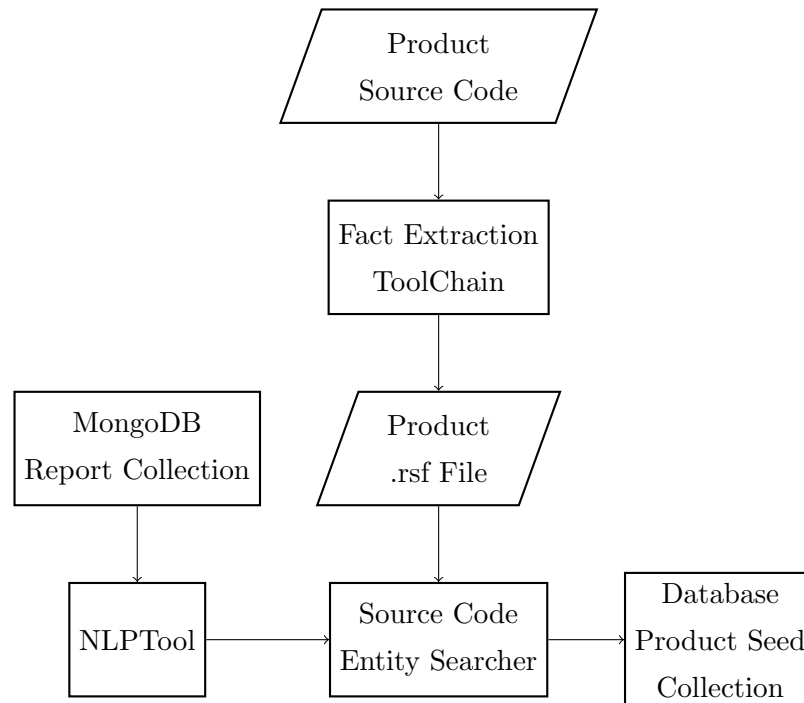
Για να γίνει εφικτό να διατηρηθούν από κάθε αναφορά σφάλματος μόνο οι όροι που αντιστοιχούν σε οντότητες του προγράμματος έπρεπε να υλοποιηθεί μία δομή αναζήτησης με πολύ χαμηλό χρόνο αναζήτησης η οποία να χρησιμοποιηθεί για την διεργασία αυτή. Αυτό έπρεπε να γίνει γιατί για κάθε έναν όρο από κάθε αναφορά θα γίνεται αναζήτηση σε αυτή τη δομή που θα περιέχει χιλιάδες οντότητες, και θα γίνουν χιλιάδες αναζητήσεις. Η δομή που φτιάχτηκε ήταν ένα AVL-tree το οποίο προσφέρει πολυπλοκότητα  $O(\log n)$ . Οι οντότητες με τις οποίες γέμισε η δομή βρίσκονται από το RSF αρχείο που εξάγεται από το εργαλείο Fetch.

### 3.1.2.2 NLP Tool

Για την εφαρμογή επεξεργασίας αδόμητου κειμένου στις αναφορές σφάλματος, όπως έχει ήδη αναφερθεί στο 2.6, χρησιμοποιήθηκαν οι βιβλιοθήκες OpenNLP. Ο λόγος ήταν ότι είναι δωρεάν διαθέσιμο, γρήγορο, και διατίθεται μαζί με εκπαιδευμένα μοντέλα στην αγγλική γλώσσα

Για τη δημιουργία του εργαλείου, το πρώτο βήμα ήταν η εξαγωγή από την MongoDB όλων των αναφορών σφάλματος για ένα συγκεκριμένο σύστημα του οποίου ο κώδικας ήταν διαθέσιμο. Μετά το σύνολο των αναφορών περνάνε μέσα από τον επεξεργαστή μία-μία με τον εξής τρόπο. Για κάθε αναφορά και για κάθε σχόλιο κάθε αναφοράς, επιλέγονται οι προτάσεις και αφαιρούνται οι άχρηστες λέξεις. Μετά χρησιμοποιώντας τη δομή για αναζήτηση των οντοτήτων του κώδικα διατηρούνται στη συλλογή μόνο οι οντότητες του κώδικα και αποθηκεύονται για μελλοντική χρήση σε ένα διαφορετικό Πίνακα της βάσης.





ΣΧΗΜΑ 3.2: Bug Report Preprocessing Tools Data Flow

### 3.1.3 Μικρά Εργαλεία

#### 3.1.3.1 Εγγραφή Source Code Entity από RSF

Αυτό είναι ένα πολύ απλό πρόγραμμα το οποίο διαβάζει το αποτέλεσμα του Fact Extraction Tool Chain και αποθηκεύει κάθε μοναδικό αναγνωριστικό ακριβώς μία φορά σε ένα αρχείο. Κάθε μοναδικό αναγνωριστικό είναι το όνομα μίας οντότητας και γράφεται με τέτοιο τρόπο που φορτώνεται εύκολα από τα υπόλοιπα προγράμματα που θα χρειαστεί να χρησιμοποιήσουν αυτήν την πληροφορία.

#### 3.1.3.2 MongoDB Query Client

Για να είναι εφικτό τόσο να διερευνηθεί η ορθότητα των εργαλείων που γεμίζουν τη βάση αλλά και την ορθότητα των δεδομένων στη βάση, δημιουργήθηκε ένα εργαλείο σε Java. Το εργαλείο αυτό δημιουργήθηκε επειδή η MongoDB έχει ένα περίεργο τρόπο για διεξαγωγή ερωτήσεων στα περιεχόμενα δεδομένα, και θεωρήθηκε χρήσιμο η δυσκολία αυτή να 'κρυφτεί' πίσω από ένα πολύ πιο απλό πρόγραμμα. Συγκεκριμένα το πρόγραμμα αυτό μπορεί να επιλέξει βάση στην οποία θα πραγματοποιηθεί η ερώτηση, σε ποιο πεδίο και για ποια τιμή και επιστρέφει τα αποτελέσματα σε μορφή αλφαριθμητικής ακολουθίας η οποία είναι συμβατή με το πρότυπο JSON και μπορεί να πραγματοποιηθεί καλύτερη εποπτεία με χρήση κάποιου online εργαλείου παρουσίασης JSON χειμένων.

### 3.1.3.3 Source Code Entity Expander

Αυτό το εργαλείο είναι λίγο πιο περίπλοκο από τα υπόλοιπα που παρουσιάζονται σε αυτήν την ενότητα, αλλά είναι και εξαιρετικά χρήσιμο για τη διεξαγωγή των πειραμάτων. Αφού έχουν εξαχθεί οι οντότητες του προγράμματος από τις αναφορές, υφίσταται μία σημαντική απειλή για την ποιότητα των δεδομένων. Συγκεκριμένα αρκετές αναφορές σφάλματος περιέχουν ένα πολύ μικρό σύνολο οντοτήτων και κατά συνέπεια είναι πολύ φτωχές σε σημασιολογία. Αυτό το πρόβλημα αντιμετωπίστηκε με τη δημιουργία ενός εργαλείου το οποίο προσθέτει σε κάθε αναφορά κάποιες επιπλέον οντότητες οι οποίες σχετίζονται στο RSF με τις ήδη υπάρχουσες στην αναφορά. Αυτό δεν μειώνει την ποιότητα των δεδομένων καθώς απορρέει κατευθείαν από το αποτέλεσμα του Fact Extraction Tool Chain. Το μόνο επιπλέον πρόβλημα που προκύπτει από αυτήν την πράξη της επέκτασης είναι η μείωση της ακρίβειας αλλά αυτό στα πλαίσια της εργασίας δεν απασχολεί καθώς δεν πρόκειται για ένα ολοκληρωμένο εργαλείο αλλά για μία απόδειξη ορθότητας της προσέγγισης που παρουσιάζεται στο Κεφάλαιο 6.

Η επέκταση γίνεται με διάφορους τρόπους ανάλογα με τον τύπο της εκάστοτε οντότητας. Με αυτόν τον τρόπο οντότητες που αντιστοιχούν σε συναρτήσεις επεκτείνονται με χρήση της σχέσης "call" η οποία είναι σημασιολογικά σημαντική σχέση για τις συναρτήσεις. Δυστυχώς όμως δεν υπάρχει κάποια αντίστοιχα σημαντική σχέση για οντότητες άλλου τύπου. Οι κλάσεις οι οποίες και αυτές επεκτείνονται με αυτόν τον τρόπο αξιοποιούν τη σχέση "inheritsFrom" η οποία αν και σημασιολογικά διαφέρει από την "call", υποδεικνύει ένα σημαντικό επίπεδο σύζευξης ανάμεσα σε δύο κλάσεις. Οι υπόλοιπες οντότητες αφήνονται ως έχουν και περνάνε ένα δεύτερο επίπεδο επέκτασης σε μετέπειτα φάση της όλης πειραματικής δομής όπως θα παρουσιαστεί αναλυτικά στο Κεφάλαιο 6.

### 3.1.3.4 Λεξικό Όρων

Πολλές φορές ενώ διεξαγόταν κάποια διεργασία που αξιοποιούσε τις οντότητες του πηγαίου κώδικα έγινε σαφές ότι υπήρχε χρησιμότητα για γνώση των σχέσεων με άλλες οντότητες και για τον τύπο της εκάστοτε οντότητας. Για αυτόν ακριβώς το λόγο της διευκόλυνσης αυτής της διαδικασίας, δημιουργήθηκε ένα λεξικό με περιεχόμενα αυτήν ακριβώς την πληροφορία. Αυτό αρχικά εξυπηρετούσε μόνο αυτόν τον στόχο αλλά αργότερα έγινε κατανοητό ότι μπορούσε να χρησιμοποιηθεί στη θέση του Source Code Entity Expander καθώς είναι γρηγορότερο και πολύ πιο επεκτάσιμο. Η πιο χρήσιμη από τις υλοποιηθείσες επεκτάσεις είναι η προσθήκη και των ανάστροφων σχέσεων οι οποίες επιτρέπουν πιο σύνθετες αναζητήσεις να γίνονται με έναν πιο εύκολο και διαισθητικά κατανοητό τρόπο.

## 3.2 Μεθοδολογία

Προκειμένου να επιτευχθεί το επιθυμητό αποτέλεσμα εντός του χρονοδιαγράμματος, αποφασίστηκε να χρησιμοποιηθεί ένα μικρό υποσύνολο των εργαλείων που υλοποιήθηκαν, και ακόμα και αυτά να χρησιμοποιηθούν με σημαντικές μετατροπές στην λειτουργία τους. Ο λόγος πίσω από αυτήν την απόφαση δεν είναι άλλος από την απλοποίηση της όλης διαδικασίας. Ο στόχος για ένα αρχικό ενδιάμεσο αποτέλεσμα κρίθηκε ότι είναι εφικτό να επιτευχθεί χρησιμοποιώντας ελάχιστα από τα προαναφερθέντα. Στην πράξη αν και δεν γίνεται άμεσα χρήση όλων των εργαλείων, αυτά παίζουν σημαντικό ρόλο κατά την υλοποίηση του προτεινόμενου μοντέλου για αναφορές σφάλματος όπως θα δειχθεί στη συνέχεια.

### 3.2.1 Τελική Προεπεξεργασία

Η προεπεξεργασία των δεδομένων που χρησιμοποιήθηκε για τη διεξαγωγή των πειραμάτων είναι μία απλοποιημένη έκδοση της προαναφερθείσας. Γίνεται χρήση όλων των εργαλείων μέχρι και τον επεξεργαστή φυσικής γλώσσας (NLP) ωστόσο ο επεξεργαστής χρησιμοποιείται με στόχο όχι τη δημιουργία ομαδοποιημένων οντοτήτων με βάση τα ποιοτικά χαρακτηριστικά τους (function ή non-function) αλλά απλά επιλέγονται και τοποθετούνται σε μία ομάδα όλες οι λέξεις που ταυτίζονται με κάποια οντότητα του κώδικα. Αυτό επιτυγχάνεται φορτώνοντας όλες τις αναφορές σφάλματος του συστήματος που εξετάζεται και χρησιμοποιώντας τη δομή HashMap της Java οι οποία αν και έχει μεγάλο memory overhead ακόμα καλύτερα από τη δομή AVL-tree επιτρέποντας τοποθέτηση και ανάκληση αντικειμένων σε  $O(1)$ . Η συγκεκριμένη δομή χρησιμοποιείται σαν λεξικό των όρων/οντοτήτων του προγράμματος και χρησιμοποιείται για να ελέγχεται κάθε λέξη που προκύπτει από την ανάγνωση των αναφορών σφάλματος.

### 3.2.2 Αρχικός Χώρος Αναζήτησης

Αφού κάθε αναφορά σφάλματος για κάποιο σύστημα έχει μετατραπεί σε μία συλλογή λέξεων που υπάρχουν στο αποτέλεσμα του εργαλείου Fact Extraction Tool Chain, χρησιμοποιείται η τεχνική LSA για να ομαδοποιηθούν οι αναφορές γύρω από αυτές που εκτελούν χρέη ερώτησης στα πλαίσια των δοκιμαστικών εκτελέσεων και για τις οποίες έχει βρεθεί η σωστή απάντηση. Με χρήση της τεχνικής αυτής, χωρίς κάποιας μορφής ειδική διαμόρφωσή της για να δουλεύει ειδικά για κάποιο συγκεκριμένο σύστημα, επιτυγχάνεται η ομαδοποίηση των αναφορών σφάλματος σε σύνολα που αποτελούνται από ομοειδείς αναφορές και αντιστοιχούν σε σχετικά σφάλματα. Με αυτόν τον τρόπο επιτυγχάνεται η δημιουργία ενός επεκταμένου αρχικού χώρου αναζήτησης στον οποίο ενσωματώνεται η πληροφορία, δηλαδή οι όροι, από όλες τις ομοειδείς αναφορές δίδοντας έτσι μία πολύ μεγάλη πιθανότητα να βρεθεί τελικά

κάποιος όρος ο οποίος να είναι 'αρκετά' κοντά στο σημείο του κώδικα που προκαλεί την αναφορά-ερώτηση.

### 3.2.3 Διεύρυνση Χώρου Αναζήτησης

Ο επεκταμένος χώρος αναζήτησης στην συντριπτική πλειοψηφία των συστημάτων που αναλύθηκαν έδειξε ότι δεν συμπεριλαμβάνει επακριβώς τις επιθυμητές οντότητες που μπορούν να κατευθύνουν τον αλγόριθμο στην επιθυμητή λύση. Με δεδομένο ότι στην περίπτωση εκλογής του απλού επεκταμένου χώρου η ανάκληση της μεθόδου θα έπεφτε αρκετά κάτω από 50% κρίθηκε σκόπιμο να εφαρμοσθεί ακόμα ένα βήμα επέκτασης στα στοιχεία του χώρου. Κατά αναλογία με τα προαναφερθέντα για επέκταση με χρήση των σχέσεων "calls" και "inherits\_from" προωθήθηκε η λύση της επέκτασης χρησιμοποιώντας το σύνολο των διαθέσιμων σχέσεων (οι οποίες εξαρτώνται από τη γλώσσα προγραμματισμού του υπό ανάλυση συστήματος).

### 3.2.4 Κατάταξη Οντοτήτων του Χώρου Αναζήτησης

Ο χώρος ο οποίος είναι διαθέσιμος σε αυτό το στάδιο παρατηρήθηκε ότι εμπεριέχει με μεγάλη πιθανότητα τμήμα ή και ολόκληρη τη λύση για τις διαθέσιμες αναφορές σφάλματος. Αυτό το γεγονός επιτρέπει στη μέθοδο ένα δείκτη ανάκλησης ο οποίος κυμαίνεται άνω του 75% ενώ για κάποια συστήματα ξεπερνάει το 90%. Αυτός ο υψηλός δείκτης αποτελεί μία απόδειξη της ορθότητας της μεθόδου που εφαρμόζεται, καθώς επιτυγχάνει ένα διπλό στόχο:

- τη δραστική μείωση του χώρου αναζήτησης κατά πάνω από 60% σταθερά.
- τη διατήρηση υψηλού δείκτη ανάκλησης άνω του 60%.

Παρά την αρχική επιτυχία όπως αυτή διαφαίνεται της μεθόδου, οι δείκτες αυτοί δεν είναι απαραίτητα καλοί. Συγκεκριμένα ο δείκτης μείωσης του χώρου αναζήτησης υπολογίζεται ως ποσοστό επί του συνόλου του συστήματος. Συνέπεια αυτού είναι ότι ένα μεγάλο σύστημα με 4.000 ή 5.000 αρχεία κώδικα και παραπάνω θα οδηγήσει στη δημιουργία ενός τελικού χώρου αναζήτησης που δύναται να περιέχει πάνω από 1.000 αρχεία στην οποία περίπτωση δεν διαφαίνεται ουσιαστική βελτίωση από την οπτική του προγραμματιστή που θα κληθεί να το αξιοποιήσει τα αποτελέσματα. Ο καλύτερος τρόπος για να αντιμετωπιστεί αυτό, μετά από αρκετούς πειραματισμούς, αποδείχτηκε πως είναι μία βαθμονόμηση κάθε οντότητας που επιστρέφεται εντός του διευρυμένου χώρου αναζήτησης, και κατόπιν η ταξινόμηση των βαθμονομημένων αποτελεσμάτων με βάση κάποια ποιοτικά χαρακτηριστικά που μπορούν στην πράξη να βοηθήσουν τον χρήστη.

### 3.2.5 Προκαταρκτικά Αποτελέσματα

Τα αποτελέσματα που αποκομίσθηκαν από την ως άνω περιγραφείσα διαδικασία όπως θα παρουσιαστούν εκτενώς τόσο στο σχετικό κεφάλαιο όσο και στο παράρτημα κρίθηκαν ως ικανοποιητικά. Δεδομένης της αρχικής θέσης η οποία ήταν η δημιουργία ενός υποβοηθητικού συστήματος μάλλον, παρά η δημιουργία ενός έμπειρου συστήματος εύρεσης σφαλμάτων, με ό,τι αυτό συνεπάγεται, τα κριτήρια για την αναγνώριση των αποτελεσμάτων ως θετικών ή αρνητικών είναι σαφώς διαφορετικά από αυτά που θα εφαρμόζονταν στην περίπτωση ενός ολοκληρωμένου έμπειρου συστήματος. Είναι επίσης σημαντικό να αναφερθεί ότι αυτά τα αποτελέσματα έχουν ουσία μόνο στα πλαίσια της απόδειξης της αρχικής υπόθεσης. Η εργασία αυτή εξάλλου δεν φιλοδοξεί να παρουσιάσει μία ολοκληρωμένη, λειτουργική υλοποίηση για το πρόβλημα του εντοπισμού σφαλμάτων αλλά να δείξει ότι κάτι τέτοιο είναι εφικτό να υποβοηθηθεί με τη χρήση του συνόλου τεχνικών που αξιοποιήθηκαν καθώς και της εξαχθείσας μέτα-γνώσης.



## Κεφάλαιο 4

# Μοντέλο Δεδομένων

### 4.1 Αρχική Δομή Δεδομένων

Τα δεδομένα προέρχονται από δύο βασικές πηγές. Από τη μία είναι οι αναφορές σφάλματος, Bug Reports οι οποίες περιέχουν διάφορες πληροφορίες σχετικές με κάποιο σφάλμα. Οι πληροφορίες αυτές χωρίζονται σε δύο κατηγορίες. Η πρώτη κατηγορία είναι η ημιδομημένες πληροφορίες που περιέχονται σε ξεχωριστά πεδία μέσα σε κάθε αναφορά, και ορίζουν το 'περιβάλλον' του σφάλματος. Σαν περιβάλλον ορίζεται η έκδοση του συστήματος στο οποίο εντοπίζεται το πρόβλημα, το λειτουργικό σύστημα που χρησιμοποιείται, το συγκεκριμένο υποσύστημα που παράγει το σφάλμα και άλλες τέτοιου είδους πληροφορίες. Η δεύτερη κατηγορία είναι τα σχόλια της αναφοράς. Τα σχόλια κάθε αναφοράς παρέχουν μία 'ελεύθερου κειμένου' πληροφορία η οποία στην μορφή που παρέχεται είναι άχρηστη για κάποιο υπολογιστικό σύστημα και ως εκ τούτου πρέπει να περάσει κάποια στάδια προεπεξεργασίας μέχρι να μετατραπεί σε κάποια μορφή που να μπορεί να χρησιμοποιηθεί από κάποιον αλγόριθμο. Η άλλη βασική πηγή πληροφορίας είναι οι οντότητες του κώδικα όπως αυτές προκύπτουν από το εργαλείο fetch.

Όπως ειπώθηκε ήδη τα δεδομένα που προκύπτουν από τις αναφορές σφάλματος βρίσκονται σε μία απεικόνιση τύπου bag of words. Αυτού του τύπου οι απεικόνιση, ενώ είναι φιλική προς κάποιες τεχνικές εξόρυξης δεδομένων, αποδείχτηκε ακατάλληλη για χρήση στην προτεινόμενη πλατφόρμα. Αυτή η συνειδητοποίηση σε συνδυασμό με το κενό που υπάρχει στην βιβλιογραφία για την αναπαράσταση αναφορών σφάλματος, οδήγησε στη δημιουργία ενός Domain Specific Model ειδικά για την κάλυψη αυτού του κενού.

Το μοντέλο που τελικώς δημιουργήθηκε για την αναπαράσταση των αναφορών σφάλματος υλοποιήθηκε στην πλατφόρμα EMF του Eclipse Framework και ως εκ τούτου έχει όλα τα πλεονεκτήματα που προκύπτουν από αυτό. Παρά την ορθότητα του μοντέλου και της

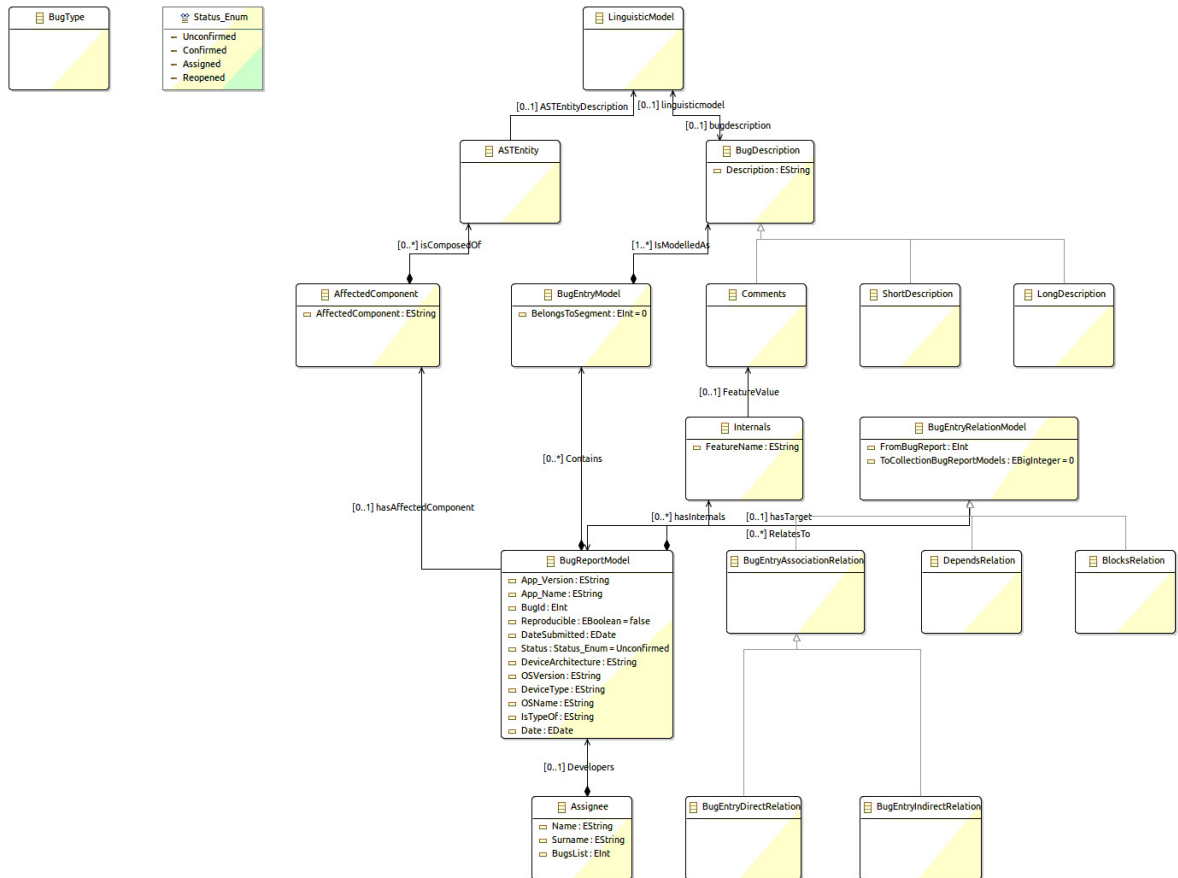
λειτουργικότητας που αυτό μπορεί να προσφέρει, τελικά αποφασίστηκε για λόγους χρονοδιαγράμματος να αναβληθεί η χρήση του σε μετέπειτα χρονική στιγμή, ωστόσο είναι χρήσιμο και για λόγους κατανόησης των υπολοίπων τμημάτων της εργασίας να γίνει μία εκτενής παρουσίαση.

Το προτεινόμενο μοντέλο για την αναπαράσταση των αναφορών σφάλματος, λαμβάνει υπόψιν ένα μεγάλο υποσύνολο της συνολικής πληροφορίας που είναι δυνατό να εξαχθεί από μία αναφορά. Συγκεκριμένα ενώ για την ολοκλήρωση του πειράματος αξιοποιήθηκαν μόνο οι όροι εκείνοι που εξήχθησαν από το κείμενο κάθε σχολίου της αναφοράς, στο μοντέλο διατηρείται πληροφορία, εφόσον υπάρχει, για την έκδοση του προγράμματος, το στοιχείο του προγράμματος στο οποίο παρουσιάζεται το σφάλμα, το όνομα του χρήστη που έκανε την αναφορά, τα διαγνωστικά μηνύματα που έβγαλε το σύστημα όταν απέτυχε η εκτέλεση και άλλα που ενώ εκ πρώτης όψεως φαίνονται επουσιώδη στην πράξη φάνηκε πως αποτελούν πηγή χρήσιμης πληροφορίας.

#### 4.1.1 EMF-Model

Παρακάτω παρουσιάζεται το σχεδιάγραμμα ολόκληρου του προτεινόμενου μοντέλου για αναπαράσταση αναφορών σφάλματος:





ΣΧΗΜΑ 4.1: Διάγραμμα κλάσης του Μοντέλου Αναφορών Σφάλματος

Η κεντρική οντότητα του προτεινόμενου μοντέλου είναι η οντότητα "BugReportModel". Αυτή η κλάση περιέχει πεδία για όλες τις βασικές πληροφορίες μίας αναφοράς σφάλματος. Αυτές οι πληροφορίες περιλαμβάνουν :

Όπως φαίνεται και στο Σχήμα 4.1 η οντότητα "BugReportModel" συνδέεται με μία σχέση "contains" με την "BugEntryModel" η οποία μοντελοποιεί την περιγραφή ενός Σφάλματος "BugDescription". Η κλάση "BugDescription" ουσιαστικά είναι αυτή που μοντελοποιεί ένα σφάλμα. Αυτή η κλάση κληρονομείται από την κλάση "Comments", "ShortDescription" και "LongDescription".

Κατά συνέπεια της παραπάνω περιγραφής και του σχήματος 4.1 προκύπτει ο παρακάτω πίνακας.

| Όνομα Πεδίου       | Περιγραφή Πεδίου   |
|--------------------|--|
| App_Version        | Η έκδοση του Συστήματος για την οποία έχει καταχωρηθεί η αναφορά σφάλματος.  |
| App_Name           | Το όνομα του Συστήματος για το οποίο έχει καταχωρηθεί η αναφορά σφάλματος.   |
| BugId              | Το μοναδικό αναγνωριστικό της συγκεκριμένης αναφοράς μέσα στην πλατφόρμα καταχώρησης αναφορών σφάλματος.   |
| Reproducible       | Δηλώνει αν είναι εφικτή η αναπαραγωγή της προβληματικής συμπεριφοράς από άλλον χρήστη.   |
| DateSubmitted      | Η ημερομηνία που έγινε η υποβολή της συγκεκριμένης αναφοράς σφάλματος.   |
| Status             | Πεδίο που μπορεί να πάρει μία από τέσσερις δυνατές τιμές, Unconfirmed, Confirmed, Assigned, Reopened , και δηλώνει την κατάσταση στην οποία βρίσκεται η αναφορά.                 |
| DeviceArchitecture | Η αρχιτεκτονική της συσκευής στην οποία έτρεχε το Σύστημα όταν παρουσίασε την προβληματική συμπεριφορά π.χ. intelx86, AMD64, ARM.  |
| OSVersion          | Η έκδοση του λειτουργικού συστήματος στην οποία έτρεχε το Σύστημα όταν παρουσίασε την προβληματική συμπεριφορά.  |
| DeviceType         | Ο τύπος της συσκευής στην οποία έτρεχε το Σύστημα όταν παρουσίασε την προβληματική συμπεριφορά. Π.χ. mobile, Personal Computer, Laptop, Table.                                   |
| OSName             | Το όνομα του λειτουργικού συστήματος στο οποίο έτρεχε το Σύστημα.  |
| IsTypeOf           | Ο τύπος του Bug παίρνει τιμές από μία δομή απαρίθμησης πιθανών τιμών. Δείχνει τί τύπου είναι το σφάλμα από προγραμματιστικής άποψης, π.χ. StackOverflow, FloatingPointErrorκ.τ.λ |
| Date               | Η ημερομηνία προσθήκης της αναφοράς στο μοντέλο.   |

ΠΙΝΑΚΑΣ 4.1: Περιγραφή Πεδίων της Κλάσης BugReportModel

| Όνομα Κλάσης     | Όνομα Κλάσης Γονέα | Πεδία   |
|------------------|--------------------|---|
| BugDescription   | Καμία              | Description. Τύπος : String. Κρατάει textual πληροφορία για την περιγραφή ενός σφάλματος. |
| Comments         | BugDescription     | Τα ίδια πεδία με την κλάση Γονέα.   |
| ShortDescription | BugDescription     | Τα ίδια πεδία με την κλάση Γονέα.   |
| LongDescription  | BugDescription     | Τα ίδια πεδία με την κλάση Γονέα.   |

ΠΙΝΑΚΑΣ 4.2: Περιγραφή της ιεραρχίας της κλάσης BugDescription

Οπότε όπως φαίνεται και στο σχήμα κάθε αναφορά σφάλματος έχει το πολύ ένα `shortDescription` και ένα `LongDescription` και τουλάχιστον ένα `Comments`.

Εξαιρετικό ενδιαφέρον παρουσιάζει η σχέση που υπάρχει ανάμεσα στην κλάση `Comments` και την κλάση `BugReportModel`, μέσω της κλάσης `Internals`. Αυτή η σχέση δείχνει πώς η κάθε αναφορά σφάλματος περιέχει περισσότερη πληροφορία από αυτή που απεικονίζεται άμεσα στην κλάση. Συγκεκριμένα φαίνεται πώς η `textual` πληροφορία κάθε αναφοράς μετατρέπεται σε ένα σύνολο από αντικείμενα που το καθένα περιέχει μία μεταβλητή τύπου `String` η οποία αναπαριστά πληροφορία σχετική με το σύστημα, π.χ. το όνομα κάποιας μεταβλητής ή κάποιας συνάρτησης του συστήματος.

Από την καθαρά δομική άποψη του μοντέλου δεν έχει ακόμα γίνει αναφορά στις συσχετίσεις ανάμεσα σε αναφορές σφάλματος. Αυτό το κομμάτι του μοντέλου είναι ίσως από τα πιο σημαντικά τμήματά του. Ο λόγος για τον οποίο το τμήμα που μοντελοποιεί τις σχέσεις ανάμεσα σε αναφορές σφάλματος είναι τόσο σημαντικό λόγω του τρόπου με τον οποίο αντιμετωπίζεται συνήθως μία νέα αναφορά σφάλματος. Οι προγραμματιστές και συντηρητές ενός συστήματος δεν αρχίζουν κατευθείαν να ψάχνουν να βρουν τί φταίει για το πρόβλημα στον πηγαίο κώδικα του συστήματος. Αντίθετα διερευνούν την πιθανότητα να είναι διπλότυπη αναφορά, ή να έχει αντιμετωπιστεί το ίδιο πρόβλημα στο παρελθόν για κάποια άλλη έκδοση ή για πολύ παρόμοιο περιβάλλον εκτέλεσης. Αυτή ακριβώς η πληροφορία είναι διαθέσιμη τόσο μέσω της πλατφόρμας `Bugzilla` όσο και μέσω της χρήσης τεχνικών όπως η `LSA`.

Η κάθε αναφορά σφάλματος όπως έχει δημιουργηθεί με βάση το μοντέλο, μπορεί να συσχετιστεί με άλλες αναφορές με τουλάχιστον τέσσερις τρόπους. Αυτός ο συσχετισμός επιτυγχάνεται μέσω της κλάσης `BugEntryRelationModel`. Αυτή η κλάση διατηρεί από ποια αναφορά προς ποιες αναφορές υπάρχουν σχέσεις. Αυτή η κλάση κληρονομείται από τρεις κλάσεις παιδιά, τις κλάσεις `BugEntryAssociationRelation`, την `DependsRelation` και την `BlocksRelation`. Οπότε κάθε σφάλμα μπορεί να συσχετίζεται με άλλα σφάλματα, είτε απαγορεύοντας την επίλυση αυτών μέχρι να επιλυθεί πρώτα το ίδιο, είτε αδυνατώντας να επιλυθεί μέχρι να επιλυθούν πρώτα τα άλλα. Τέλος η κλάση `BugEntryAssociationRelation` κληρονομείται από τις κλάσεις, `BugEntryDirectRelation` και την `BugEntryIndirectRelation`.

Και οι δύο αυτές κλάσεις χρησιμοποιούνται για την αναπαράσταση σχέσεων που έχουν προκύψει κατόπιν ανάλυσης για τις αναφορές, οπότε εμπλουτίζουν θεαματικά το βάθος της αναπαράστασης.

Όσα έχουν αναφερθεί έως εδώ αφορούν αποκλειστικά στο μοντέλο του σφάλματος και της αναφοράς σφάλματος. Ωστόσο όπως προέκυψε κατά τη διαδικασία της μοντελοποίησης αυτό δεν είναι αρκετό. Για αυτό το λόγο στην αριστερή μεριά του σχήματος 4.1 παρατηρείται η ύπαρξη δύο κλάσεων που δεν έχουν άμεση προφανή σχέση με την αναφορά σφάλματος. Η

κλάση `AffectedComponent`, η οποία ουσιαστικά είναι πεδίο της `BugReportModel`, διατηρεί πληροφορία για το όνομα υπό τη μορφή `String` για το υποσύστημα που πλήττεται από το συγκεκριμένο σφάλμα με το οποίο γεμίζει το μοντέλο κάθε φορά. Αυτή η πληροφορία όμως είναι αχρείαστη καθώς θα μπορούσε να είναι απλά ένα πεδίο στην κλάση `BugReportModel`. Στην πράξη η κλάση `AffectedComponent` περιέχει μία συλλογή από όρους του συστήματος, αυτό φαίνεται από τη σχέση `[0..*]isComposedOf` με την κλάση `ASTEntity`. Αυτοί οι όροι, που κάθε ένας αναπαρίσταται από ένα αντικείμενο της κλάσης `ASTEntity`, είναι οι όροι που έχουν εξαχθεί από το εργαλείο 2.8 και οι οποίοι σχετίζονται ακριβώς και μόνο με το προβληματικό υποσύστημα.

Με την πληροφορία όπως αυτή έχει μοντελοποιηθεί μέχρι τώρα, το μοντέλο αναφορών σφάλματος διατηρεί πολύ περισσότερη πληροφορία από όση παρέχεται μέσα από την πλατφόρμα `Bugzilla` καθώς η αρχική πληροφορία εμπλουτίζεται με `implied` γνώση και στοιχεία του κώδικα τα οποία απουσιάζουν εντελώς από τις αναφορές σφάλματος. Συνέπεια αυτού είναι να υπάρχει πλέον μία ολιστική αναπαράσταση του σφάλματος τόσο από τη μεριά που γίνεται άμεσα κατανοητή στον άνθρωπο όσο και από την μεριά που γίνεται πιο εύκολα κατανοητή στον υπολογιστή.

Για λόγους πληρότητας προστέθηκε στο μοντέλο και η ύπαρξη του προγραμματιστή που ασχολείται με την τρέχουσα αναφορά σφάλματος. Η κλάση `Assignee` πετυχαίνει ακριβώς αυτό. Περιέχοντας τρία πεδία για τον μοναδικό προσδιορισμό του κάθε προγραμματιστή. Επίσης το πεδίο `BugList` το οποίο είναι μία λίστα ακεραίων που περιέχει Μοναδικά Αναγνωριστικά Αναφορών Σφάλματος (`BugIds`) επιτρέπει την πρόσβαση σε όλες τις αναφορές με τις οποίες ασχολείται κάθε προγραμματιστής, και οι οποίες πιθανόν να έχουν κάποια κοινά χαρακτηριστικά.

Τέλος στην κορυφή του μοντέλου υπάρχει ο συνδυαστικός κρίκος ανάμεσα στην λεξικολογική αναπαράσταση και την αναπαράσταση σε όρους οντοτήτων κώδικα. Στην ουσία, από πλευράς περιεχόμενης πληροφορίας, η κλάση `LinguisticModel` συνδυάζει τους όρους που εξάγονται από την περιγραφή σε ανθρώπινη γλώσσα του σφάλματος, και τους συνδυάζει με τους σχετικότερους όρους που έχουν εξαχθεί από τον κώδικα. Για την σωστή πλήρωση με πληροφορία αυτής της κλάσης μπορεί να χρησιμοποιηθεί, είτε η προτεινόμενη μέθοδος, είτε κάποιας άλλης μορφής λογική που θα μπορούσε να χρησιμοποιηθεί για την αντιμετώπιση του `Bug Localization` προβλήματος.

Το Μοντέλο αυτό που σε πολύ αδρές γραμμές παρουσιάζεται, δυστυχώς δεν κατέστη εφικτό να χρησιμοποιηθεί στα πλαίσια της εργασίας που έγινε. Αυτό οφείλεται σε αρκετούς λόγους. Ο σημαντικότερο από αυτούς είναι ότι καθ'όλη τη διάρκεια του πειραματισμού και της προσαρμογής της εργαστηριακής ομάδας στην υφή των διαθέσιμων δεδομένων, έπρεπε να γίνονται συνεχώς τροποποιήσεις και επικαιροποιήσεις του μοντέλου.

Παρά το γεγονός ότι δεν χρησιμοποιήθηκε στα πλαίσια αυτής της εργασίας, η ύπαρξη πλέον ενός αρκετά αναλυτικού, πιστεύεται, μοντέλου θα βοηθήσει τόσο οποιονδήποτε από το εργαστήριο να εργαστεί στο μέλλον σε παρεμφερές θέμα, όσο και την ήδη υπάρχουσα ομάδα στην προσπάθειά της να οδηγήσει την έρευνα σε αυτόν τον τομέα σε νέες ανεξερεύνητες περιοχές εφαρμόζοντας πιο εξωτικές προσεγγίσεις για την επίλυση του προβλήματος.

Τέλος το μοντέλο δίνει πλέον τη δυνατότητα ολοκλήρωσης της γνώσης για κάθε μία αναφορά σφάλματος και της ενσωμάτωσης όλης της παράγωγης γνώσης σε ένα προγραμματιστικό περιβάλλον που μάλιστα παρέχεται 'δωρεάν' όπως έχει ήδη ειπωθεί στο Κεφάλαιο 2 από το EMF περιβάλλον του Eclipse.

## 4.2 Μετρική Ποιότητας Αναφορών

Η προσπάθεια για την δημιουργία ενός όσο το δυνατόν πιο αντιπροσωπευτικού μοντέλου για την αναπαράσταση την πληροφορίας που βρίσκεται μέσα στις αναφορές προϋπέθετε την ανάγνωση και την ανάλυση ενός σημαντικού πλήθους αναφορών από διάφορες συλλογές (KDE, Gnome, Mozilla) και για διάφορα συστήματα από κάθε συλλογή. Μέσω αυτής της προσπάθειας ανέκυψε η ανάγκη για κατηγοριοποίηση των αναφορών με βάση κάποια μετρική η οποία να αναδεικνύει την ποιότητα αυτών.

Η μετρική που επιλέχθηκε, και η οποία πιστεύεται ότι μπορεί να αναπαραστήσει με ακρίβεια τα ποιοτικά χαρακτηριστικά που ενδιαφέρουν για την εφαρμογή αυτόματων μεθόδων ανάλυσης αυτών, έχει να κάνει πρωτίστως με το πλήθος των οντοτήτων κώδικα οι οποίες βρίσκονται μέσα σε κάθε αναφορά ως προς το πλήθος των συνολικών λέξεων που περιέχει αυτή η αναφορά αφαιρώντας πρώτα όλους τους όρους (λέξεις) που για τους σκοπούς της εργασίας θεωρούνται κενό νοήματος και αφαιρούνται σε διάφορα στάδια της προεπεξεργασίας. Με αυτόν τον ορισμό προκύπτει η μετρική:

$$BugReport\_Quality = \frac{No\_Of\_Source\_Code\_Entities\_In\_Report}{No\_Of\_Non\_Stop\_Word\_Terms\_In\_Report}$$

Αν και τελικά δεν χρησιμοποιήθηκε αυτή η μετρική σε κανένα σημείο της επιλογής των Αναφορών του πειράματος ή της αξιολόγησης των αποτελεσμάτων αυτού, πιστεύεται ότι μπορεί να βοηθήσει στον εντοπισμό περιπτώσεων οι οποίες είναι εξαιρετικά δύσκολο να επιλυθούν μηχανιστικά λόγω απουσίας πληροφορίας. Ο λόγος για τον οποίο δεν χρησιμοποιήθηκε είναι ότι θα μπορούσε να εισάγει μία αχρείαστη απειλή προς την ορθότητα και αμεροληψία της μεθόδου.



## Κεφάλαιο 5

# Εργαλεία Εξόρυξης Γνώσης

### 5.1 Ομοιότητα Αναφορών Σφαλμάτων Λογισμικού

Τα σφάλματα λογισμικού μπορούν να προκληθούν, όπως έχει ήδη αναφερθεί, από μία πλειάδα λόγων. Ο στόχος της προτεινόμενης λύσης είναι να μπορέσει να βρει κάποιον κοινό τόπο ανάμεσα σε συλλογές τέτοιων αναφορών, με στόχο να μπορεί ο οποιοσδήποτε να αντλήσει πληροφορία σχετικά με μία καινούρια αναφορά σφάλματος από την ήδη υπάρχουσα πληροφορία που βρίσκεται σε όλες τις προηγούμενες αναφορές. Δεδομένου ότι το σύστημα λαμβάνει υπόψιν μόνο δεδομένα που σχετίζονται με το ίδιο σύστημα λογισμικού, δεν είναι παράλογο να υποθεθεί πως οι περιγραφές για το ίδιο σύστημα από διαφορετικούς χρήστες θα χρησιμοποιούν κάποιο κοινό σύνολο από λέξεις κλειδιά. Αυτές οι λέξεις κλειδιά θα πρέπει κατά συνέπεια να συσχετιστούν μεταξύ τους ώστε να γίνει εφικτός ο συνδυασμός πολλών διαφορετικών αναφορών σφάλματος ούτως ώστε να μπορέσει το σύστημα να επιτύχει μεγαλύτερη ανάκληση. Για να επιτευχθεί αυτός ο στόχος χρησιμοποιήθηκε ο αλγόριθμος LSA ο οποίος έχει αναφερθεί επιγραμματικά στις προαπαιτούμενες γνώσεις.

### 5.2 LSA

Η τεχνική LSA (Latent Semantic Analysis) σχετίζεται με το πεδίο της επεξεργασίας φυσικής γλώσσας NLP και συγχεκριμένα με τη σημασιολογία κατανομών (distributional semantics). Η τεχνική αυτή εφαρμόζεται στην ανάλυση σχέσεων μεταξύ συνόλων κειμένων και των όρων που αυτά περιέχουν με το να δημιουργεί ένα σύνολο αφηρημένων εννοιών σχετικών με τα κείμενα και τους όρους. Η βασική παραδοχή αυτής της τεχνικής είναι ότι όροι που έχουν παραπλήσια ερμηνεία θα εμφανίζονται σε όμοια κομμάτια κειμένου. Για την επίτευξη του στόχους της αυτή η μέθοδος βασίζεται σε μία απεικόνιση του χώρου το προβλήματος υπό τη μορφή πίνακα. Συγχεκριμένα ο πίνακας περιέχει το πλήθος εμφανίσεων κάθε λέξης

που ανήκει στο σύνολο των λέξεων, σε κάθε παράγραφο, οι γραμμές αντιπροσωπεύουν τις λέξεις και οι στήλες τις παραγράφους και ο πίνακας δημιουργείται από μία μεγάλη συλλογή κειμένων. Στο πρώτο στάδιο της επεξεργασίας αυτός ο πίνακας απλοποιείται χρησιμοποιώντας τη μέθοδο αποσύνθεσης μοναδικής τιμής SVD (Singular Value Decomposition) η οποία μειώνει το πλήθος των γραμμών διατηρώντας ωστόσο το πλήθος των στηλών αλλά και τη δομή ομοιότητας μεταξύ τους. Κατόπιν αυτού οι λέξεις συγκρίνονται ανά δύο μετρώντας το συνημίτονο της γωνίας μεταξύ των ανάλογων διανυσμάτων που αντιστοιχούν στις σχετικές γραμμές του πίνακα. Οι τιμές αυτού του υπολογισμού προφανώς περιορίζονται στο κλειστό διάστημα  $[0, 1]$  όπου τιμή 1 υποδηλώνει πολύ όμοιες λέξεις ενώ τιμή 0 υποδηλώνει πολύ ασυσχέτιστες λέξεις. Τέλος αξίζει αναφοράς το γεγονός ότι η εν λόγω τεχνική συχνά αποκαλείται και LSI (Latent Semantic Indexing) στα πλαίσια του πεδίου ανάκτησης πληροφορίας Information retrieval).

## 5.2.1 Επισκόπηση Τμημάτων της Τεχνικής

### 5.2.1.1 Πίνακας Εμφάνισης

Η τεχνική LSA χρησιμοποιεί έναν πίνακα όρων-κειμένων ο οποίος διατηρεί πληροφορία για την εμφάνιση κάθε όρου σε κάθε κείμενο. Αυτός είναι ένας αραιός πίνακας του οποίου οι γραμμές αντιστοιχούν σε όρους και οι στήλες σε κείμενα όπως έχει προαναφερθεί. Μία συνηθισμένη μέθοδος μέσω της οποίας προσδίδονται βάρη στους όρους του πίνακα είναι η χρήση του λόγου tf-idf ο οποίος όπως έχει αναλυθεί στο 2 είναι μία μετρική που υποδηλώνει την σημαντικότητα ενός όρου. Η μετρική αυτή είναι αντιστρόφως ανάλογη της συχνότητας εμφάνισης οπότε όσο πιο μεγάλη η συχνότητα εμφάνισης ενός όρου τόσο μικρότερη η σημασία του.

### 5.2.1.2 Υποβάθμιση Τάξης

Μετά τη δημιουργία του πίνακα εμφάνισης, η LSA βρίσκει μία προσέγγιση μικρότερου βαθμού για τον πίνακα όρων-κειμένων. Οι λόγοι για αυτή την προσέγγιση είναι οι κάτωθι:

1. Ο αρχικός πίνακας όρων-κειμένων έχει υποτεθεί ότι είναι εξαιρετικά μεγάλος για να χρησιμοποιηθεί αποδοτικά μέσα σε ένα υπολογιστικό σύστημα. Σε αυτή την περίπτωση ο πίνακας μειωμένου βαθμού είναι μία αναγκαία και απαραίτητη προσέγγιση, έστω κι αν οδηγεί σε μερική αλλοίωση των τελικών αποτελεσμάτων.
2. Ο αρχικός πίνακας όρων-κειμένων έχει υποτεθεί ότι είναι πολύ θορυβώδης: για παράδειγμα περιέχει όρους προϊόντα λεξιπλασίας ή διαφορετικής γλωσσικής προέλευσης



οι οποίοι πρέπει να αφαιρεθούν. Σε αυτήν την περίπτωση έχουμε να κάνουμε με έναν αποθορυβοποιημένο πίνακα σε σχέση με τον αρχικό.

3. Ο αρχικό πίνακας όρων-κειμένων έχει υποτεθεί ότι είναι εξαιρετικά αραιός σε σχέση με τον ανάλογο πίνακα 'πραγματικών' όρων της συλλογής κειμένων. Αυτό μπορεί να συμβεί γιατί λαμβάνονται υπόψιν κατά τη δημιουργία του πίνακα μόνο οι όροι που όντως βρίσκονται μέσα σε κάθε κείμενο ενώ στην πράξη μπορεί αυτοί οι όροι να αυξηθούν μέσω συμπερίληψης συνώνυμων.

Το αποτέλεσμα του υποβιβασμού τάξης στον πίνακα όρων-κειμένων είναι κάποιες από τις διαστάσεις συνδυάζονται με αποτέλεσμα να ενθυλακώνουν πλέον πληροφορία σχετικά με παραπάνω από έναν όρους.

Παράδειγμα :

(αυτοκίνητο), (φορτηγό), (λουλούδι)  $\rightarrow$  (1.3452 \*αυτοκίνητο + 0.2828 \* φορτηγό), (λουλούδι)

Αυτό αμβλύνει το πρόβλημα του εντοπισμού συνωνύμων, καθώς όπως ειπώθηκε ο υποβιβασμός τάξης του αναμένεται να συγχωνεύσει κάποιες από τις διαστάσεις σε μία διάσταση με σχετικό νόημα. Αυτό εξάλλου αμβλύνει και το πρόβλημα της αμφισημίας καθώς τέτοιες λέξεις που δείχνουν (δηλαδή το διάνυσμα τους) προς τη 'σωστή' κατεύθυνση συγχωνεύονται με άλλες σχετικές λέξεις, ενώ στην αντίθετη περίπτωση αλληλοαναιρούνται, τέλος στην χειρότερη περίπτωση είναι λιγότερο σημαντικές από άλλες λέξεις που δείχνουν στην ίδια κατεύθυνση αλλά είναι πιο σημαντικές για την δεδομένη σημασία.

### 5.2.1.3 Παραγωγήιση

Έστω  $X$  ένας πίνακας με στοιχεία  $(i, j)$  τα οποία δείχνουν την εμφάνιση του όρου  $i$  στο κείμενο  $j$ . Ο  $X$  έχει τότε την ακόλουθη μορφή:

$$\begin{bmatrix} x_{1,1} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,m} \end{bmatrix}$$

σε αυτήν την αναπαράσταση κάθε γραμμή αντιστοιχεί σε κάποιον όρο, και κάθε στήλη σε κάποιο κείμενο:  $t_i^T = [x_{1,1} \dots x_{1,n}]$

ομοίως κάθε στήλη αντιστοιχεί σε κάποιο κείμενο:  $d_j = \begin{bmatrix} x_{1,j} \\ \vdots \\ x_{m,j} \end{bmatrix}$

με αυτούς τους πίνακες το εσωτερικό γινόμενο δύο όρων θα είναι  $t_i^T * t_p$  και θα δίνει την ετεροσυσχέτιση μεταξύ τους πάνω σε όλα τα κείμενα ενώ το γινόμενο πίνακα  $XX^T$  θα περιέχει όλα τα εσωτερικά γινόμενα. Το στοιχείο  $(i, p)$  (το οποίο είναι ίσο με το στοιχείο  $(p, i)$ ) περιέχει το εσωτερικό γινόμενο  $t_i^T * t_p (= t_p^T * T_i)$ . Ομοίως ο πίνακας  $X^T X$  περιέχει όλα τα εσωτερικά γινόμενα όλων των κειμένων δίνοντας έτσι την συσχέτιση τους πάνω σε όλους τους όρους:  $d_j^T * d_q = d_q^T * d_j$ .

Πλέον χρησιμοποιώντας θεωρία γραμμικής άλγεβρας, είναι γνωστό ότι υπάρχει αποσύνθεση  $X$  τέτοια ώστε  $U$  και  $V$  να είναι δύο ορθογώνιοι πίνακες και  $\Sigma$  να είναι ένας διαγώνιος. Δηλαδή αποσύνθεση μοναδικής τιμής (SVD):

$$X = U\Sigma V^T$$

Τα γινόμενα των πινάκων που μας δίνουν τις συσχετίσεις όρων και κειμένων γίνονται με αυτόν τον τρόπο :

$$XX^T = (U\Sigma V^T)(U\Sigma V^T)^T = (U\Sigma V^T)(V^{TT}\Sigma^T U^T) = U\Sigma V^T V \Sigma^T U^T = U\Sigma \Sigma^T U^T$$

$$X^T X = (U\Sigma V^T)^T(U\Sigma V^T) = (V^{TT}\Sigma^T U^T)(U\Sigma V^T) = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$$

Δεδομένου τώρα ότι ο  $\Sigma \Sigma^T$  και ο  $\Sigma^T \Sigma$  είναι διαγώνιοι ο πίνακας  $U$  περιέχει τα ιδιοδιανύσματα του  $\Sigma \Sigma^T$  ενώ ο  $V$  τα ιδιοδιανύσματα του  $\Sigma^T \Sigma$ . Και τα δύο γινόμενα έχουν τις ίδιες μη μηδενικές ιδιοτιμές, οι οποίες δίνονται από τις μη μηδενικές τιμές του  $\Sigma \Sigma^T$  ή ομοίως του  $\Sigma^T \Sigma$ . Κατόπιν αυτών των βημάτων η σχετική αποσύνθεση είναι ως ακολούθως:

$$(\mathbf{t}_i^T) \rightarrow \begin{matrix} X \\ (\mathbf{d}_j) \\ \downarrow \\ \begin{bmatrix} x_{1,1} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,m} \end{bmatrix} \end{matrix} = (\hat{\mathbf{t}}_i^T) \rightarrow \begin{matrix} U \\ \downarrow \\ \left[ \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_l \end{bmatrix} \right] \end{matrix} \dots \begin{matrix} \Sigma \\ \downarrow \\ \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_l \end{bmatrix} \end{matrix} \cdot \begin{matrix} V^T \\ \hat{\mathbf{d}}_j \\ \downarrow \\ \left[ \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_l \end{bmatrix} \right] \end{matrix}$$

Οι τιμές  $\sigma_1, \dots, \sigma_l$  λέγονται μοναδικές τιμές και οι  $u_1, \dots, u_l$  και  $v_1, \dots, v_l$  τα εξ'αριστερών και εκ δεξιών μοναδικά ιδιοδιανύσματα. Παρατηρείται ότι το μοναδικό μέρος του  $U$  που συμμετέχει στο κάθε  $t_i$  είναι η  $i$ -οστή σειρά. Έστω  $\hat{t}_i$  αυτό το διάνυσμα. Ταυτοχρόνως το μόνο μέρος του  $V^T$  που συμμετέχει στο  $\mathbf{d}_j$  είναι η  $i$ -οστή στήλη, και έστω  $d_j$  το

διάνυσμα αυτό. Αυτά (τα  $\hat{\mathbf{t}}_i$  και  $\hat{\mathbf{d}}_j$ ) δεν είναι τα ιδιοδιανύσματα αλλά εξαρτώνται από όλα τα ιδιοδιανύσματα.

Τελικώς μπορεί ναδειχθεί ότι διαλέγοντας τις  $k$  μεγαλύτερες μοναδικές τιμές, και τα συσχετιζόμενα μοναδικά διανύσματα  $U$  και  $V$ , αποκομίζεται η τάξης  $k$  προσέγγιση του πίνακα  $X$  με το ελάχιστο σφάλμα (Μέτρο Frobenius). Αυτή η προσέγγιση έχει ένα ελάχιστο σφάλμα. Αλλά το πιο σημαντικό είναι ότι πλέον είναι δυνατόν να αξιοποιηθούν τα διανύσματα όρων και κειμένων ως ένας σημασιολογικός χώρος. Το διάνυσμα  $\hat{\mathbf{t}}_i$  έχει κατά συνέπεια  $k$  συνιστώσες που το ταιριάζουν με έναν μικρότερης διάστασης χώρο. Σχετικά με αυτές τις προσεγγίσεις είναι σημαντικό να σημειωθεί ότι δεν έχουν κάποια φυσική ερμηνεία αλλά είναι απλά η μικρότερης διάστασης προσέγγιση του υψηλότερης διάστασης χώρου. Τοιούτοτρόπως το διάνυσμα  $\hat{\mathbf{d}}_j$  είναι μία προσέγγιση σε αυτόν τον μικρότερης διάστασης χώρο. Τυπικά αυτή η προσέγγιση μπορεί να γραφεί ως:

$$X_k = U_k S_k V_k^T$$

Πλέον είναι δυνατό να γίνουν τα κάτωθι:

- Να διερευνηθεί πόσο όμοια είναι τα κείμενα  $j$  και  $q$  στον μικρής διάστασης χώρο μέσω της σύγκρισης των διανυσμάτων  $\sum_k \hat{\mathbf{d}}_j$  και  $\sum_k \hat{\mathbf{d}}_q$  (συνήθως μέσω χρήσης ομοιότητας συνημιτόνου).
- Να συγκριθούν οι όροι  $i$  και  $p$  συγκρίνοντας τα διανύσματα  $\sum_k \hat{\mathbf{t}}_i$  και  $\sum_k \hat{\mathbf{t}}_p$ .
- Να ομαδοποιηθούν οι αναπαραστάσεις των κειμένων και των διανυσμάτων των όρων χρησιμοποιώντας γνωστές μεθόδους (clustering) (συσταδοποίηση) όπως k-means και μέτρα ομοιότητας όπως η ομοιότητα συνημιτόνου.
- Να δοθούν ερωτήσεις ή οποίες θα αντιμετωπιστούν σαν μικρά κείμενα και θα συγκριθούν με τη λίστα κειμένων που υπάρχουν ήδη στον χώρο μικρότερης διάστασης. Για να επιτευχθεί αυτό πρέπει πρώτα να μεταφραστεί η ερώτηση στον μικρότερης διάστασης χώρο. Αυτό υποδεικνύει ότι είναι απαραίτητο να χρησιμοποιηθεί ο ίδιος μετασχηματισμός που χρησιμοποιήθηκε και για την αρχική συλλογή κειμένων:  $\hat{\mathbf{d}}_j = \sum_l^{-1} U_k^T \mathbf{d}_j$ . Σημειώνεται ότι ο αντίστροφος του διαγώνιου πίνακα  $\sum_k$  είναι δυνατό να βρεθεί αντιστρέφοντας κάθε μη μηδενική τιμή του (ιδιότητα που προκύπτει από γραμμική άλγεβρα). Αυτό σημαίνει ότι για κάποια ερώτηση  $q$  πρέπει να γίνει η μετάφραση  $\hat{\mathbf{q}} = \sum_k^{-1} U_k^T \mathbf{q}$  πριν γίνει εφικτό να συγκριθεί με τα υπόλοιπα διανύσματα κειμένων του μικρότερης διάστασης χώρου. Προφανώς ανάλογα μπορεί να γενικευθεί η διαδικασία και για ψευδή διανύσματα όρων ως εξής:

$$\mathbf{t}_i^T = \hat{\mathbf{t}}_i^T \sum_k \mathbf{V}_k^T \hat{\mathbf{t}}_i = \mathbf{t}_i^T \mathbf{V}_K^{-T} \sum_k^{-1} = \mathbf{t}_i^T \mathbf{V}_k \sum_k^{-1} \hat{\mathbf{t}}_i = \sum_k^{-1} \mathbf{V}_k^T \mathbf{t}_i$$

#### 5.2.1.4 Εφαρμογές

Η διαδικασία υποβιβασμού τάξης πίνακα που αναλύθηκε στην προηγούμενη ενότητα επιτρέπει μία πληθώρα εφαρμογών ανάμεσα στις οποίες είναι και οι ακόλουθες:

- Η Σύγκριση κειμένων στον μικρότερης διάστασης χώρο (χρησιμοποιείται σε αλγορίθμους συσταδοποίησης, και κατηγοριοποίησης).
- Η εύρεση παρεμφερών κειμένων ανάμεσα σε διαφορετικές γλώσσες, αφού πρώτα αναλυθεί κάποιο σύνολο μεταφρασμένων κειμένων (cross language retrieval).
- Η εύρεση συνωνύμων μεταξύ όρων, συνώνυμα ή πολυσημία.
- Δεδομένων των όρων μιας ερώτησης, να μεταφραστούν αυτοί στον μικρότερης διάστασης χώρο ώστε να συσχετιστούν με άλλα κείμενα της συλλογής (information retrieval).
- Η εύρεση της μέγιστης ομοιότητας μέσω μικρών ομάδων όρων, σύμφωνα με κάποιο σημασιολογικό κριτήριο, όπως αυτό συμβαίνει στην προσπάθεια επίλυσης ερωτήσεων πολλαπλών επιλογών χρησιμοποιώντας το μοντέλο MCQ για απάντηση ερωτήσεων

### 5.2.2 Υλοποίηση

#### 5.2.2.1 Επιλογή Γλώσσας

Για την υλοποίηση της τεχνικής LSA ώστε να χρησιμοποιηθεί για τους σκοπούς της εργασίας, επιλέχθηκε η γλώσσα Python. Οι λόγοι για αυτήν την επιλογή είναι:

- Η συγκεκριμένη γλώσσα προγραμματισμού είναι εύκολη στην εκμάθηση
- Δεν απαιτεί κάποια πολύπλοκη διαδικασία εγκατάστασης και είναι προεγκατεστημένη στη συντριπτική πλειοψηφία των υπολογιστών.
- Παρέχει στον προγραμματιστή μία εξαιρετικά πλούσια βιβλιοθήκη έτοιμων εργαλείων
- Εφαρμόζει βελτιστοποιήσεις το οποίο επιτρέπει η διαδικασία να επιτυγχάνει καλύτερους χρόνους εκτέλεσης.

### 5.2.2.2 Βασικές Βιβλιοθήκες

Για την υλοποίηση της τεχνικής LSA στάθηκε αναγκαίο να γίνει χρήση αρκετών βασικών βιβλιοθηκών στη γλώσσα Python. Συγκεκριμένα χρησιμοποιήθηκαν οι βιβλιοθήκες:

- Η βιβλιοθήκη `numpy` και η υποβιβλιοθήκη `numpy.linalg` που περιέχει εργαλεία για την εύρεση της αποσύνθεσης μοναδικής τιμής ενός πίνακα.
- Η βιβλιοθήκη `pickle` χρησιμοποιήθηκε για να πραγματοποιηθεί αποθήκευση των ενδιάμεσων αποτελεσμάτων κατά την επεξεργασία κάθε συστήματος ώστε να αποφευχθούν πλεονάζοντες και επαναλαμβανόμενοι εξαιρετικά πολύπλοκοι υπολογισμοί.
- Η βιβλιοθήκη `scipy.sparse.linalg` που περιέχει συναρτήσεις για εύρεση αποσύνθεσης μηδενικής τιμής σε αραιούς και εξαιρετικά αραιούς πίνακες όπως αυτοί που προέκυψαν κατά την προεπεξεργασία κάποιων συστημάτων.
- Η βιβλιοθήκη `bz2` για συμπίεση των αποθηκευόμενων ενδιάμεσων αποτελεσμάτων.
- Η βιβλιοθήκη `json` για επεξεργασία πληροφορίας σε αναπαράσταση τύπου JavaScript Object Notation
- Η βιβλιοθήκη `Pool` από τη βιβλιοθήκη `multiprocessing` ώστε να επιτευχθεί παράλληλη επεξεργασία αξιοποιώντας μεγαλύτερο μέρος της υπολογιστικής ισχύος του υπολογιστή που χρησιμοποιείτο.

### 5.2.2.3 Δομή του Εργαλείου LSA

Σε αυτήν την ενότητα αναλύονται λεπτομέρειες της υλοποίησης του εργαλείου LSA που τελικώς πραγματοποιήθηκε στη γλώσσα Python. Αρχικά φορτώνονται στο πρόγραμμα οι όροι οι οποίοι αποτελούν το λεξικό του χώρου ο οποίος ορίζεται από τις αναφορές σφάλματος. Ο χώρος δημιουργείται προσθέτοντας ένα καινούριο στοιχείο σε ένα διάνυσμα στο οποίο κάθε όρος υπάρχει ακριβώς μία φορά μετά το πέρας της δημιουργίας. Με αυτόν τον τρόπο έχει δημιουργηθεί ο πίνακας όρων του χώρου.

Σε αυτήν την ενότητα αναλύονται λεπτομέρειες της υλοποίησης του εργαλείου LSA που τελικώς πραγματοποιήθηκε στη γλώσσα Python. Αρχικά φορτώνονται στο πρόγραμμα οι όροι οι οποίοι αποτελούν το λεξικό του χώρου ο οποίος ορίζεται από τις αναφορές σφάλματος. Ο χώρος δημιουργείται προσθέτοντας ένα καινούριο στοιχείο σε ένα διάνυσμα στο οποίο κάθε όρος υπάρχει ακριβώς μία φορά μετά το πέρας της δημιουργίας. Με αυτόν τον τρόπο έχει δημιουργηθεί ο πίνακας όρων του χώρου. Αξίζει να σημειωθεί ότι κατά τη διάρκεια της πρώτης ανάγνωσης των κειμένων αγνοείται ένα σύνολο λέξεων οι οποίες έχουν θεωρηθεί

θόρυβος. Αυτές η λέξεις είναι κάποιες επιπλέον λέξεις σε σχέση με αυτές που θεωρούνται ως θόρυβος από το εργαλείο Natural Language Processing. Το σύνολο αυτό είναι το ακόλουθο:

```
["and" "edition" "for" "in" "little" "of" "the" "to"]
```

Η διαδικασία μετά την φόρτωση των κειμένων συνεχίζει με την δημιουργία ενός αντικειμένου DocProcess και την κλήση της μεθόδου

```
docs_process.initializeDocArr(dir_path, query_path)
```

η οποία επιστρέφει δύο πίνακες σχετικούς με τα κείμενα που έχουν οριστεί για επεξεργασία. Κατόπιν αυτού δημιουργείται ένα αντικείμενο τύπου LSA με παραμέτρους stopwords, ignorechars, query\_path) το οποίο εμπεριέχει τις μεθόδους:

- parse(my\_docs)
- build()
- buildIn(my\_indoc)
- calc()
- keep\_related\_bugs(strictness, docMap, namedoc, direction)

Οι μέθοδοι αυτές είναι οι βασικές μέθοδοι που πρέπει να χρησιμοποιηθούν κατά την διαδικασία ανάλυσης ενός συνόλου κειμένων (χώρος αναζήτησης) και της επερωτήσεως του χώρου με ένα κείμενο-ερώτηση.

Επιγραμματικά η μέθοδος parse απλά διαβάσει και διατηρεί μόνο τους αποδεκτούς όρους από κάθε κείμενο. Η build χρησιμοποιώντας το συσχετιστικό πίνακα που αντιστοιχεί τα ονόματα των αρχείων στους περιεχόμενους όρους τους, χτίζει τον πίνακα όρων-κειμένων. Η buildIn(my\_indoc) κάνει την ίδια επεξεργασία αλλά μόνο για το κείμενο ερώτηση. Η calc() υπολογίζει τους απαραίτητους πίνακες για την αρχική συλλογή κειμένων, δηλαδή τους πίνακες που σχετίζονται με την αποσύνθεση μοναδικής τιμής.

Τέλος η keep\_related\_bugs(strictness, docMap, namedoc, direction) ελέγχει την ομοιότητα μεταξύ του κειμένου-ερώτηση και ελέγχει για κάθε επίπεδο ομοιότητας πόσα κείμενα επιστρέφονται οπότε μέσω αυτού του ελέγχου περιορίζεται το πλήθος των κειμένων που μπορεί να επιστραφούν. Ο λόγος για την προσαρμοστική συμπεριφορά αυτής της μεθόδου είναι ότι ένα σταθερό κατώφλι ομοιότητας για όλα τα συστήματα κρίθηκε ακατάλληλο καθώς το πλήθος των επιστρεφόμενων αποτελεσμάτων κυμαινόταν σε τέτοιο βαθμό που ήταν δυνατό να αλλοιώσει σημαντικά τα τελικά αποτελέσματα.

## 5.3 Natural Language Processing

Η επεξεργασία φυσικής γλώσσας (NLP) είναι ένα πεδίο που βρίσκεται στην τομή της επιστήμης των υπολογιστών (computer science), της τεχνητής νοημοσύνης (Artificial Intelligence) και της υπολογιστικής γλωσσολογίας (computational linguistics). Το πεδίο εφαρμογής, τουλάχιστον επί της αρχής, ήταν η διευκόλυνση της αλληλεπίδρασης ανθρώπου-μηχανής με έναν φυσικό για τον άνθρωπο-χρήστη τρόπο. Κάτω από την ομπρέλα που καλύπτεται από τον όρο NLP συνυπάρχουν τεχνικές για κατανόηση φυσικής γλώσσας natural language understanding καθώς και για παραγωγή φυσικής γλώσσας natural language generation. Για τους σκοπούς της εργασίας αξιοποιήθηκαν ωστόσο μόνο εργαλεία που σχετίζονται με την κατανόηση, άρα και ανάλυση, φυσικής γλώσσας.

### 5.3.1 Ιστορική Αναδρομή του NLP

Ο όρος 'ανάλυση φυσικής γλώσσας' αν και υπήρχε και πριν το 1950 ουσιαστικά άρχισε από τότε να υπάρχει σαν πεδίο, το οποίο προέκυψε από τη δημοσίευση του Alan Turing, 'Computing Machinery and Intelligence' στην οποία προτάθηκε η διαδικασία Turing Test για την αξιολόγηση της νοημοσύνης.

Έκτοτε έγιναν προσπάθειες να αντιμετωπιστεί το γλωσσικό μέρος της ανθρώπινης επικοινωνίας ως ένα μηχανιστικά επιλύσιμο πρόβλημα το οποίο επιδεχόταν μία, προγραμματιστική, λύση. Ωστόσο όταν προσπάθειες που διήρκεσαν για πάνω από μία δεκαετία απέβησαν άκαρπες, τόσο η χρηματοδότηση όσο και το ενδιαφέρον για το συγκεκριμένο πρόβλημα γνώρισαν μία δραματική μείωση. Από τη δεκαετία του 1980 κι ενώ ακόμα τα περισσότερα NLP συστήματα ήταν ακόμα συλλογές κανόνων, εμφανίστηκαν τα πρώτα τέτοια συστήματα που ενσωμάτωναν τεχνικές μηχανικής μάθησης machine learning. Αυτό εν μέρει προκλήθηκε από τη σταθερή αύξηση της διαθέσιμης υπολογιστικής ισχύος όμως η ουσιαστική γενεσιουργός αιτία ήταν η απομάκρυνση της κοινότητας από τη ως τότε επικρατούσα γλωσσολογική προσέγγιση του Noam Chomsky που ακόμα και στο θεωρητικό επίπεδο αποθάρρυναν την γλωσσολογική προσέγγιση που απαιτείται για χρήση μηχανικής μάθησης.

Το επόμενο μεγάλο βήμα για τον τομέα με την εισαγωγή της τεχνικής της 'επισήμανσης των μερών του λόγου' για τη λειτουργία της οποία εφαρμόστηκαν στην ανάλυση φυσικής γλώσσας τα Hidden Markov Models και συνέπεια αυτού ήταν η ολοένα και μεγαλύτερη διείσδυση των στατιστικών μοντέλων, τα οποία υποκαθιστούν τους αυστηρούς ελέγχους με χρήση if-statements με εύκαμπτους ελέγχους που βασίζονται σε πιθανότητες που προκύπτουν από την βαθμονόμηση των δεδομένων εισόδου. Οι στατιστικές αυτές μέθοδοι βασίζονται σε επισημειωμένα δεδομένα εισόδου, και κατά συνέπεια απαιτείται σημαντική προετοιμασία του 'εκπαιδευτικού υλικού' για την εκπαίδευση των μοντέλων.

Πλέον το ενδιαφέρον συγκεντρώνεται στην δημιουργία ανεπίβλεπτων ή ημί-επιβλεπόμενων εργαλείων μηχανικής μάθησης που θα μπορούν να εκπαιδεύονται χωρίς την ύπαρξη κάποιας προεπεξεργασμένης μορφής δεδομένων εισόδου.

### 5.3.2 Machine Learning και NLP

Όπως αναφέρθηκε ήδη το NLP γίνεται πλέον σχεδόν αποκλειστικά με χρήση τεχνικών μηχανικής μάθησης και ιδιαιτέρως με χρήση στατιστικών μεθόδων μηχανικής μάθησης. Το πρότυπο της μηχανικής μάθησης ξεπερνάει για το σκοπό του NLP τις δυνατότητες των προγενέστερα χρησιμοποιούμενων μεθόδων τόσο σε επιδόσεις όσο και χρόνο εκπαίδευσης.

Στις παλαιότερες μεθόδους απαιτείτο η συγγραφή από προγραμματιστή του συνόλου των κανόνων συνεπαγωγής ενώ πλέον αυτό γίνεται με χρήση αλγορίθμων μάθησης γενικού σκοπού που ξεκινάνε από μία βάση και είναι θεμελιωμένοι σε στατιστική εξαγωγή γνώσης, ώστε να 'μάθουν' την πληροφορία χρησιμοποιώντας μεγάλα corpora συνηθισμένων καθημερινών φράσεων. Corpora (Corpus) είναι μία συλλογή εγγράφων, ή φράσεων που έχει επισημειωθεί με το χέρι με τις σωστές για εκμάθηση τιμές.

Στην πράξη υπάρχουν αρκετές διαφορετικές ομάδες αλγορίθμων που μπορούν να εφαρμοστούν σε NLP προβλήματα. Αυτοί οι αλγόριθμοι είνισται να λαμβάνουν σαν είσοδο ένα μεγάλο σύνολο από 'χαρακτηριστικά (features) τα οποία προκύπτουν από τα δεδομένα εισόδου. Αυτά τα χαρακτηριστικά ουσιαστικά βαθμονομούνται μέσω της χρήσης στατιστικών μοντέλων που έχουν δημιουργηθεί με αυτό το σκοπό. Αυτό επιτρέπει την αμφιμονοσήμαντη ερμηνεία ενός κειμένου εν αντιθέσει με τις προγενέστερες προσεγγίσεις που δεν επέτρεπαν αμφισημία στο αποτέλεσμα της τεχνικής.

#### 5.3.2.1 Machine Learning

Οι τεχνικές της Μηχανικής Μάθησης έχουν σημαντικά προτερήματα έναντι των συστημάτων κανόνων που δημιουργούνται 'με το χέρι'.

Παραθέτονται παρακάτω κάποια από τα σημαντικότερα αυτά προτερήματα:

- Οι εκπαιδευτικοί μηχανισμοί που χρησιμοποιούνται κατά τη Μηχανική Μάθηση εστιάζονται αυτόματα στις πιο συχνά εμφανιζόμενες περιπτώσεις. Αντιθέτως όταν λαμβάνει χώρα συγγραφή κανόνων 'με το χέρι' δεν είναι απαραίτητως προφανές το πού πρέπει να εστιάσει ο δημιουργός των κανόνων ή που πρέπει να αφιερωθεί περισσότερος χρόνος.



- Οι αυτοματοποιημένες εκπαιδευτικές διαδικασίες αξιοποιούν αλγορίθμους Στατιστικής Εξαγωγής Γνώσης για να παράξουν τα μοντέλα τα οποία εκδηλώνουν 'σθεναρή' (robust) συμπεριφορά απέναντι σε άγνωστη είσοδο, όπως άγνωστες (μη κατηγοριοποιημένες) λέξεις, αλλά ακόμα και λανθασμένη είσοδο, που περιέχει ατελής λέξεις ή στην οποία έχουν παραληφθεί λέξεις. Γενικά η αντιμετώπιση τέτοιου τύπου εισόδου με έναν κομψό και αποδοτικό τρόπο σε συστήματα κανόνων 'με το χέρι' είναι εξαιρετικά δύσκολο, χρονοβόρο και αποτελεί πρόσφορο έδαφος για εισαγωγή λαθών, όπως άλλωστε συμβαίνει και με την προσπάθεια για δημιουργία ενός συστήματος εύκαμπτων κανόνων.
- Τα συστήματα που βασίζονται στη λειτουργία τους στην αυτοματοποιημένη εκμάθηση κανόνων δύνανται να είναι σημαντικά πιο ακριβή με μόνο επιπλέον κόστος της παροχής ενός πιο επεκταμένου εκπαιδευτικού corpus. Αντιθέτως τα συστήματα κανόνων 'με το χέρι' δεν παρέχουν καμία τέτοια ευκολία καθώς οποιαδήποτε προσπάθεια για αύξηση της ακρίβειας συμπεριλαμβάνει την εισαγωγή στο σύστημα επιπλέον, δυνητικά πολύπλοκων, κανόνων για την αντιμετώπιση ολοένα και πιο ειδικών περιπτώσεων. Φυσικά η διαδικασία αυτή δεν μπορεί να εφαρμοστεί κατέξακολούθηση καθώς οι κανόνες γίνονται ολοένα και πιο περίπλοκοι κάτι που οδηγεί αναπόφευκτα στη δημιουργία ενός μη παρατηρήσιμου συστήματος το οποίο επιπλέον δεν είναι απαραίτητο ότι δουλεύει κατά τον επιθυμητό τρόπο. Από τη μεριά των συστημάτων Μηχανικής Μάθησης όμως η δημιουργία επιπλέον επισημειώσεων είναι αρκετά εύκολη διαδικασία που αν και απαιτεί κάποιον επιπλέον χρόνο προετοιμασίας, δεν αυξάνει ουδόλως την πολυπλοκότητα της διαδικασίας ούτε του κώδικα.

### 5.3.3 Βασικές Εφαρμογές της NLP

Η τεχνική NLP εφαρμόζεται σε μία πλειάδα προβλημάτων τα οποία απασχολούν σταθερά την επιστημονική κοινότητα τόσο στο πεδίο της μηχανικής μάθησης όσο και στο πεδίο της υπολογιστικής γλωσσολογίας. Αν και κάποιες από τις δυνητικές εφαρμογές αποτελούν ακόμα καθαρά ερευνητικά θέματα υπάρχει ένα υποσύνολο το οποίο έχει ήδη αποκτήσει εφαρμογή εκτός της ερευνητικής κοινότητας. Παρακάτω παρατίθεται μία λίστα όπου αναφέρονται επιγραμματικά κάποιες από αυτές:

- Αυτοματοποιημένη Περίληψη: Στόχος είναι η δημιουργία μία αναγνώσιμης περίληψης ενός κειμένου. Χρησιμοποιείται συχνά για την παροχή περιλήψεων σε γνωστά πεδία όπου υπάρχει διαθέσιμη γνώση όπως άρθρα οικονομικών σε εφημερίδες.
- Ανάλυση Αναφοράς: Στόχος είναι από ένα κείμενο να βρεθεί ποιες λέξεις αναφέρονται σε ποια τμήματα ή ποιες άλλες λέξεις του κειμένου.

- **Ανάλυση Αλληλεπίδρασης:** Ο Στόχος είναι, από τη μία να βρεθεί ο τύπος της αλληλεπίδρασης εξάγοντας των χαρακτήρα των προτάσεων (διευκρίνιση, επεξήγηση, αντίθεση) ενώ από την άλλη είναι να βρεθεί ο τύπος της πρότασης (ερώτηση, κατάφαση, επιβεβαίωση).
- **Μηχανική Μετάφραση:** Στόχος είναι να μεταφραστεί αυτόματα ένα κείμενο από μία ανθρώπινη γλώσσα σε άλλη. Πρόκειται για μία από τις πιο δύσκολες εφαρμογές του πεδίου καθώς απαιτεί εξαιρετικά πλήρη γνώση ώστε να δουλέψει κατά τον αναμενόμενο τρόπο.
- **Μορφολογική Διάσπαση:** Στόχος είναι η διαμέριση των λέξεων σε ξεχωριστά μορφήματα και η ταξινόμηση αυτών. Σε αυτή την εφαρμογή η δυσκολία έγκειται σε μεγάλο βαθμό αποκλειστικά στην πολυπλοκότητα της μορφολογίας της γλώσσας που έχει επιλεγθεί.
- **Αναγνώριση Ονομάτων:** Στόχος είναι από μία συλλογή κειμένων να αναγνωριστούν όλες οι λέξεις που αντιστοιχούν σε τοπωνύμια ή ονόματα ανθρώπων ή γενικότερα οντοτήτων, και η μετέπειτα κατηγοριοποίηση αυτών με βάση τον τύπο της οντότητας που ονοματίζουν.
- **Δημιουργία Φυσικής Γλώσσας:** Στόχος η μετάδοση πληροφορίας από τους υπολογιστές στους ανθρώπους σε μορφή κατανοητή από ανθρώπους.
- **Κατανόηση Φυσικής Γλώσσας:** Στόχος η τυποποίηση της πληροφορίας που ενυπάρχει σε μία συλλογή κειμένου φυσικής γλώσσας σε μορφή κατανοητή από υπολογιστή. Για να επιτευχθεί αυτό χρειάζεται να εξαχθεί και να αναπαρασταθεί η σημασιολογία της φυσικής γλώσσας.
- **Οπτική Αναγνώριση Χαρακτήρων:** Στόχος η κατανόηση ως κειμένου μίας συλλογής χαρακτήρων από κάποιο αλφάβητο με χρήση υπολογιστή.
- **Επισημείωση Μερών του Λόγου:** Ο στόχος είναι με δεδομένη μία πρόταση να εξαχθεί για κάθε λέξη αυτής τι μέρος του λόγου είναι και τη θέση κατά συνέπεια καταλαμβάνει στην πρόταση.
- **Γραμματική Ανάλυση:** Στόχος η γραμματική ανάλυση ενός κειμένου φυσικής γλώσσας και η εξαγωγή ενός συντακτικού δένδρου για αυτήν.
- **Διάσπαση Προτάσεων:** Στόχος να πραγματοποιηθεί διάσπαση ενός κειμένου σε προτάσεις χρησιμοποιώντας στοιχεία στίξης αλλά και άλλες μεθόδους που οδηγούν σε καλύτερα αποτελέσματα.
- **Ανάλυση Συναισθήματος:** Στόχος η εξαγωγή από ένα κείμενο φυσικής γλώσσας του συναισθήματος του συγγραφέα μόνο από τον τρόπο μετάδοσης της πληροφορίας.

### 5.3.4 Στατιστική Επεξεργασία Φυσικής Γλώσσας

Η στατιστική Επεξεργασία Φυσικής Γλώσσας αξιοποιεί στοχαστικές, πιθανοτικές και στατιστικές μεθόδους για να επιλύσει κάποια από τα προβλήματα που αντιμετωπίζονται στον χώρο της NLP. Έχει αποδειχθεί ότι βελτιώνει της επιδόσεις ειδικά όσο αυξάνεται το μήκος της αναλυόμενης πρότασης καθώς γίνεται πιο δύσκολο να γίνει επεξεργασία τέτοιων προτάσεων με της ρεαλιστικές ή συνήθεις γραμματικές, γιατί από αυτές προκύπτουν χιλιάδες ή εκατομμύρια πιθανές αναλύσεις. Οι μέθοδοι που αξιοποιούνται για αυτού του είδους την αποσαφήνιση μεγάλων προτάσεων συμπεριλαμβάνουν τη χρήση μεγάλων συλλογών κειμένων (corpora) και Markov Models. Επίσης η στατιστική NLP έχει εφαρμοστεί στην αναγνώριση φωνής διεξάγοντας συγκρίσεις μεταξύ έμπειρων συστημάτων και στατιστικών προσεγγίσεων, όπου το τελικώς επιλεγθέν αποτέλεσμα ήταν ένα πλήρως στατιστικό μοντέλο. Από της πρώτες δε εφαρμογές στατιστικών μοντέλων για Natural Language Understanding έγιναν από τους Roberto Pieraccini, Esther Levin, Chin-Hui Lee από τα Bell Laboratories [37]. Τελικώς καταλήγοντας η NLP συμπεριλαμβάνει όλες τις διαφορετικές ποσοτικές προσεγγίσεις για αυτοματοποιημένη επεξεργασία γλώσσας, χωρίς να εξαιρούνται η πιθανοτική μοντελοποίηση, η θεωρία πληροφορίας, και η γραμμική άλγεβρα. Η δε τεχνολογίες που χρησιμοποιούνται στα πλαίσια της προέρχονται από το πεδίο της Μηχανικής Μάθησης και της Εξόρυξης Δεδομένων, πεδία που αμφότερα άπτονται της Τεχνητής Νοημοσύνης και χρησιμοποιούνται για την εξαγωγή γνώσης από δεδομένα.



## Κεφάλαιο 6

# Αλγόριθμος Εντοπισμού Σφαλμάτων και Αποτελέσματα

### 6.1 Πειραματική Διαδικασία

Προκειμένου να αξιολογηθεί η προτεινόμενη μέθοδος έπρεπε να εκτελεστούν κάποια πειράματα χρησιμοποιώντας την πάνω σε πραγματικά δεδομένα από υπαρκτά συστήματα λογισμικού. Δεδομένου ότι σαν μέθοδος είναι, ακόμα και στην απλοποιημένη της μορφή, αρκετά πολύπλοκη, ήταν απαραίτητη η υλοποίηση όλης της απαραίτητης λογικής σε μία γλώσσα προγραμματισμού.

#### 6.1.1 Χαρακτηριστικά Υπολογιστικού Περιβάλλοντος

Τα πειράματα διεξήχθησαν σε περιβάλλον Linux και διανομή Ubuntu 14.04 το οποίο είναι εγκατεστημένο σε ένα φορητό υπολογιστή με CPU Intel i3 3327U@ 1.90GHz με δύο φυσικούς πυρήνες και υλοποίηση hyperThreading πάνω σε αυτούς για παροχή τεσσάρων λογικών υπολογιστικών πυρήνων στο λειτουργικό επίσης το μηχάνημα είχε τελικώς 16GB εγκατεστημένης μνήμης.

#### 6.1.2 Περιβάλλον Ανάπτυξης

Για την διεξαγωγή του πειράματος με βάση το οποίο επιτεύχθηκε η αξιολόγηση της προτεινόμενης μεθόδου εντοπισμού σφαλμάτων υλοποιήθηκε ένα πρόγραμμα με χρήση της γλώσσας προγραμματισμού Java και συγκεκριμένα της έκδοσης Java-7-oracle.

Οι λόγοι για τους οποίους επιλέχθηκε η συγκεκριμένη γλώσσα προγραμματισμού έχει να κάνει τόσο με την γνώση που προϋπήρχε εντός της ομάδος πάνω σε αυτήν την γλώσσα, όσο και με χαρακτηριστικά της γλώσσας και προκύπτουν από την μεγάλη διείσδυση αυτής στο χώρο τόσο της έρευνας όσο και της βιομηχανίας.

Συγκεκριμένα η Java προσφέρει στον προγραμματιστή μία από τις πιο πλούσιες συλλογές έτοιμων και δοκιμασμένων βιβλιοθηκών που καλύπτουν ένα ευρύτατο φάσμα λειτουργιών ενώ ταυτόχρονα διατηρούν την πολυπλοκότητα χρήσης των σε πολύ χαμηλό επίπεδο. Εξάλλου η γλώσσα αυτή έχει το πλεονέκτημα να είναι platform independent το οποίο είναι πολύ χρήσιμο για τη διεξαγωγή του πειράματος σε παραπάνω από ένα ετερογενή μηχανήματα, ώστε να διακριβωθεί η αποδοτικότητά του ανεξαρτήτως του υλισμικού (hardware).

Για τη γλώσσα προγραμματισμού Java υπάρχουν επίσης διαθέσιμα κάποια εξαιρετικά εύχρηστα IDE (Integrated Development Environments) (Ολοκληρωμένα Περιβάλλοντα Ανάπτυξης). Ένα από τα πλέον δημοφιλή είναι και το Eclipse IDE που χρησιμοποιήθηκε για τους σκοπούς της εργασίας. Το Eclipse προσφέρει στον προγραμματιστή μία εκτεταμένη συλλογή εργαλείων για κάθε προγραμματιστική εργασία που θα μπορούσε να προκύψει στα πλαίσια της πειραματικής διαδικασίας. Τα εργαλεία που προσφέρονται από αυτό το περιβάλλον ανάπτυξης υποβοήθησαν σημαντικά στη διαδικασία ανάπτυξης ιδιαίτερα κατά την επιδιόρθωση των αρκετών λογικών αλλά και σημασιολογικών σφαλμάτων.

Εν κατακλείδι η γλώσσα προγραμματισμού Java χάρη στον αντικειμενοστραφή σχεδιασμό της και την εξαιρετικά ισχυρή και ευέλικτη Java Virtual Machine πάνω στην οποία εκτελείται παρέιχε όλες τις δυνατότητες που ήταν απαραίτητες για την εύκολη, τεκμηριωμένη και γρήγορη υλοποίηση όλων των ιδεών κατευθείαν από το σχέδιο σε κώδικα.

### 6.1.3 Επιλογή Αναφορών Σφάλματος του Gold Standard

Κάθε πειραματική διαδικασία που βασίζεται για την αξιολόγηση της επιτυχίας της στη σύγκριση με κάποιο αρχικό σύνολο ορθών αποτελεσμάτων αντιμετωπίζει έναν πολύ βασικό κίνδυνο ως προς την ορθότητα του πειράματος. Ο κίνδυνος αυτός έχει να κάνει με τη διαδικασία με βάση την οποία επιλέγονται οι ορθές απαντήσεις που θα εισαχθούν στο σύνολο που θα χρησιμοποιηθεί για έλεγχο των αποτελεσμάτων. Πιστεύεται ότι ο κίνδυνος αυτός αν και όχι μηδενικός στην περίπτωση της πειραματικής διαδικασίας που παρουσιάζεται εδώ, είναι εξαιρετικά μικρός λόγω του τυχαίου τρόπου με τον οποίο επιλέχθηκαν οι αναφορές που χρησιμοποιήθηκαν για την πλήρωση του Gold Standard.

Για κάθε ένα από τα συστήματα που χρησιμοποιήθηκαν για την δοκιμή της προτεινόμενης διαδικασίας, επιλέχθηκε ένα σύνολο από τις πιο πρόσφατες διαθέσιμες αναφορές σφάλματος. Δεν υπήρξε μέριμνα ώστε να είναι αναφορές σφάλματος της ίδιας έκδοσης με αυτή που

χρησιμοποιήθηκε στο στάδιο εξαγωγής οντοτήτων από τον κώδικα με χρήση του Fetch ούτε να πληρούν κάποια άλλη προϋπόθεση όπως να αφορούν συγκεκριμένο υποσύστημα, να έχουν πολλές οντότητες κώδικα μέσα στα σχόλιά τους ή να είναι γραμμένες κάνοντας ορθή χρήση της αγγλικής γλώσσας.

Ο τρόπος επιλογής εξασφαλίζει ένα επίπεδο τυχαιότητας στη διαδικασία επιλογής ο οποίος δεν αφαιρεί αξία από τα αποτελέσματα τα οποία θα παρουσιαστούν παρακάτω. Αν μη τι άλλο μάλιστα, αυξάνει τη σημασία τους καθώς όπως παρατηρήθηκε μετά από εξέταση των επιλεχθέντων αναφορών, κάποιες από αυτές ήταν εξαιρετικά χαμηλής ποιότητας με βάση τη μετρική που ορίστηκε στο κεφάλαιο 4 και με πληθώρα τυπογραφικών λαθών τα οποία δεν αντιμετωπίζονται από κανένα στάδιο της μεθόδου που παρουσιάζεται.

#### 6.1.4 Δομή του Πειράματος

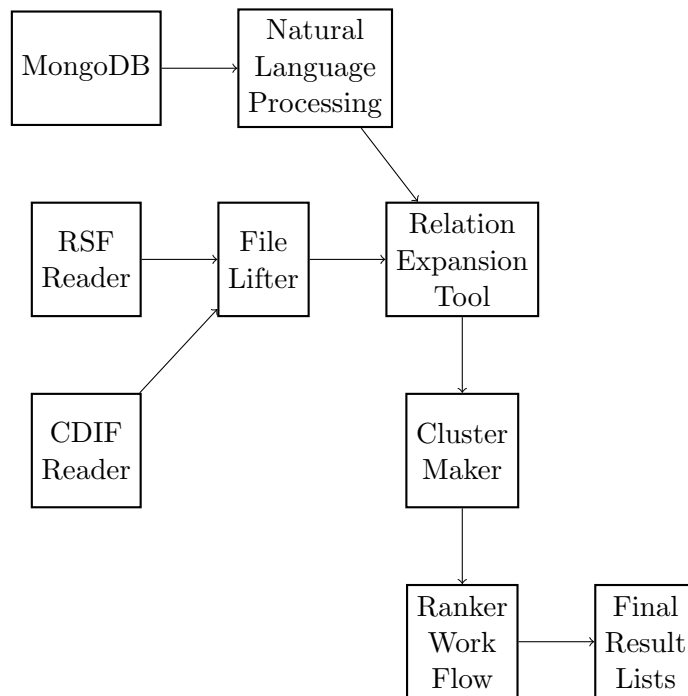
Για τη διεξαγωγή του πειράματος αξιοποιήθηκαν κάποιες από τις τεχνικές που έχουν ήδη αναλυθεί σε βάθος. Οι σημαντικότερες από αυτές είναι η τεχνικές LSA και κάποιες από τις εφαρμογές της NLP. Για τη διεξαγωγή του πειράματος οι Αναφορές Σφάλματος (τα δεδομένα) έπρεπε να είναι σε μία συγκεκριμένη μορφή, συμβατή τόσο με τα εργαλεία που κατασκευάστηκαν ειδικά για το σκοπό αυτό όσο και με κάποια εργαλεία που χρησιμοποιήθηκαν από έτοιμα πακέτα (βιβλιοθήκες).

Για την αξιολόγηση των τελικών αποτελεσμάτων αλλά και για να είναι παρατηρήσιμη η πορεία των πειραμάτων, δημιουργήθηκε μία συλλογή 'Gold Standard' για κάθε σύστημα που εξετάστηκε, η οποία περιλαμβάνει κάποιες αναφορές σφάλματος για τις οποίες είναι γνωστά τα σημεία του κώδικα που τελικώς προκάλεσαν το σφάλμα.

Αρχικά οι αναφορές σφάλματος εξάγονται από την βάση MongoDB. Επειδή όμως είναι σε μορφή free text πρέπει να περάσουν μία προεπεξεργασία προτού γίνει εφικτό να εξαχθούν οι όροι του συστήματος από κάθε μία.

Για να γίνει αυτό χρησιμοποιείται το εργαλείο openNLP το οποίο επιτρέπει να γίνει αυτόματα λεκτική ανάλυση και να σπάσουν οι αναφορές σε προτάσεις, και οι προτάσεις σε ξεχωριστούς όρους από τους οποίους απουσιάζουν όλα τα σημεία στίξης καθώς και όλες οι 'άχρηστες' λέξεις που δεν φέρουν ουσιαστική σημασιολογική αξία. Έπειτα κάνοντας χρήση της πληροφορίας που υπάρχει στα αποτελέσματα του 2.8 σχετικά με τα ονόματα των οντοτήτων του κώδικα, απορρίπτονται από τις αναφορές όλες οι συμβολοακολουθίες που δεν περιέχονται αυτούσιες στη συλλογή όρων του συστήματος.

Εν συνεχεία οι συλλογές ονομάτων όρων του συστήματος αποθηκεύονται σε αρχεία στον σκληρό δίσκο του συστήματος υπό τη μορφή απλών .txt αρχείων. Στο επόμενο βήμα χρησιμοποιείται η τεχνική LSA για να εξαχθούν οι τιμές συσχέτισης κάθε επιμέρους αναφοράς



ΣΧΗΜΑ 6.1: Μακροσκοπική Ροή Δεδομένων του Συστήματος

σφάλματος, με κάποιες από τις αναφορές του 'Gold Standard' οι οποίες ουσιαστικά αποτελούν τις ερωτήσεις στην πλατφόρμα. Αφού Ολοκληρωθεί και το LSA βήμα, καλείται ένα πρόγραμμα συσσωρευτής, το οποίο διαβάζει το αποτέλεσμα της LSA για κάθε μία από τις αναφορές του GS (Gold Standard). Το βήμα αυτό δημιουργεί τελικώς ένα καινούριο αρχείο χρησιμοποιώντας το όνομα του GS και προσθέτοντας την κατάληξη ".LSA.txt" το οποίο περιέχει όλους τους όρους από όλες τις αναφορές σφάλματος οι οποίες βρέθηκαν μέσω της LSA να έχουν συσχέτιση με το GS πάνω από κάποιο κατώφλι.

Μετά τη εφαρμογή της LSA τα αρχεία που έχουν προκύψει διαβάζονται από το τελικό στάδιο της τεχνικής που προτείνεται. Αφού διαβαστούν, φορτώνονται στο πρόγραμμα τα αρχεία [productName].rsf όσο και το [productName].cdif τα οποία χρησιμοποιούνται ως ακολούθως.

Το .cdif αρχεία περιέχει πληροφορία η οποία συσχετίζει κάθε οντότητα που ορίζεται μέσα στον κώδικα του υπό ανάλυση συστήματος με το αρχείο στο οποίο αυτός ο ορισμός λαμβάνει χώρα, αυτός ο συσχετισμός είναι απαραίτητος καθώς επιτρέπει να γίνει μία διαδικασία file lift η οποία 'μεταφράζει' κάθε οντότητα σε όνομα αρχείου. Το .rsf που περιέχει πληροφορία για τις σχέσεις μεταξύ οντοτήτων μεταφράζεται και αυτό χρησιμοποιώντας το λεξικό που προκύπτει από το .cdif σε επίπεδο αρχείου οπότε σχέσεις της μορφής:



$$RELATION \rightarrow ENTITY1 \rightarrow ENTITY2$$

γίνονται:

$$RELATION \rightarrow FILE\_OF\_ENTITY1 \rightarrow FILE\_OF\_ENTITY2$$

οπότε πλέον υπάρχει μέσα στο πρόγραμμα ένας τρόπος για να μετατραπούν όλες οι οντότητες σε ονόματα αρχείων. Αφού γίνει αυτή η μετατροπή και για τις οντότητες που έχουν προκύψει από την LSA πραγματοποιείται ένας αριθμός από επεκτάσεις του αρχικού LSA\_FILE\_NAME\_SET χρησιμοποιώντας ένα υποσύνολο των συνολικών file lifted σχέσεων. Συγκεκριμένα χρησιμοποιούνται μόνο οι σχέσεις:

- Include
- MacroUse
- Sets
- Calls
- Accesses

Το πλήθος αυτών των επεκτάσεων αρχικά είχε σχεδιαστεί με τέτοιο τρόπο ώστε να προσαρμόζεται στο μέγεθος τόσο του συστήματος σαν σύνολο, όσο και στο μέγεθος του αρχικού LSA\_FILE\_NAME\_SET. Κατόπιν πειραματισμού με τέσσερα διαφορετικά Projects εξεκριβώθηκε ότι μία προσαρμοστική προσέγγιση δεν προσφέρει κάποιο σημαντικό όφελος στις μετρικές των αποτελεσμάτων και ως εκ τούτου επιλέχθηκε η επέκταση να γίνεται για όλα τα συστήματα σε βάθος μίας σχέσης.

Το επόμενο βήμα της διαδικασίας είναι η επιλογή από το σύνολο των αρχείων που έχουν προκύψει, ενός υποσυνόλου του οποίου τα μέλη είναι τα πιο διασυνδεδεμένα αρχεία με βάση τις σχέσεις σε επίπεδο αρχείου που έχουν προκύψει από τον συνδυασμό .cdif και .rsf. Το σύνολο των επιλεχθέντων αρχείων ονομάζεται εφέξής, καταχρηστικά, Κλίκα και αξιοποιείται για την κατάταξη των αποτελεσμάτων με στόχο την ανάδειξη σε υψηλότερη θέση των αποτελεσμάτων εκείνων που έχουν καλύτερες πιθανότητες να αποτελούν όντως τη λύση του σφάλματος.

Εφόσον έχει δημιουργηθεί η Κλίμα το επεκταμένο σύνολο αρχείων γράφεται σε ένα .rsf αρχείο με τη μορφή που γίνεται δεκτή σαν είσοδος από το εργαλείο ACDC όπου σαν σχέσεις χρησιμοποιούνται όλες οι σχέσεις στις οποίες συμμετέχουν ανά δύο τα ονόματα που είναι διαθέσιμα. Με αυτόν τον τρόπο δημιουργείται ένα ξεχωριστό αρχείο για κάθε διαφορετική αναφορά σφάλματος. Μετά τη ολοκλήρωση της δημιουργίας του αρχείου αυτό περνάει σαν είσοδος στο εργαλείο ACDC το οποίο εκτελεί τον αλγόριθμο clustering πάνω στις οντότητες που περιέχονται στο αρχείο και έχει σαν έξοδο ένα αρχείο στο οποίο είναι ομαδοποιημένα τα αρχεία σε clusters. Αυτοί οι clusters φορτώνονται αμέσως μετά στο πρόγραμμα υπό τη μορφή ενός συσχετιστικού πίνακα με κλειδιά το όνομα του αρχείου και τιμές ένα σύνολο που περιέχει όλα τα περιεχόμενα ανά κλειδί.

Σε αυτό το σημείο και ενώ μέχρι τώρα έχουν χρησιμοποιηθεί σε κάθε βήμα ως ισότιμα αρχεία, αφαιρούνται από όλες τις δομές τα αρχεία βιβλιοθηκών, δηλαδή τα '.h' αρχεία. Ο λόγος για την αφαίρεση των βιβλιοθηκών είναι η παρατήρηση που έγινε κατόπιν πειραματισμού πως τα συγκεκριμένα αρχεία ξεπερνούν κατά πολύ σε πλήθος τα αρχεία κώδικα, τόσο σε πλήθος όσο και σε συνδεσιμότητα με άλλα αρχεία. Αυτές οι δύο παρατηρήσεις μαζί με το γεγονός ότι κάθε επιδιόρθωση που χρησιμοποιήθηκε για την αξιολόγηση του συστήματος περιείχε πάντα για κάθε αρχείο κώδικα και τη σχετική βιβλιοθήκη οδήγησαν στην απόφαση για την αφαίρεση αυτών. Έτσι ενώ χρησιμοποιούνται και για το στάδιο του clustering μετά απορρίπτονται για να μην αποσπάται ο αλγόριθμος από την θορυβώδη και εξαιρετικά διασυνδεδεμένη φύση τους.

Σε αυτό το σημείο είναι διαθέσιμα στο πρόγραμμα όλα τα δεδομένα που είναι απαραίτητα για τη διεξαγωγή της βαθμονόμησης των αρχείων με χρήση της μεθόδου που προτείνεται.

## 6.2 Βαθμονόμηση και Κατάταξη Αποτελεσμάτων

Για τη βαθμονόμηση, κατάταξη και ομαδοποίηση των αποτελεσμάτων χρησιμοποιείται μία αλληλουχία βημάτων η οποία δεν εντάσσεται σε κάποια από της χρησιμοποιούμενες μεθόδους που υπάρχουν διαθέσιμες για αυτήν την διαδικασία, αλλά είναι περισσότερο εμπειρική και είναι αποτέλεσμα εντατικών πειραματισμών πάνω στο σύνολο των δεδομένων που υπάρχουν διαθέσιμα. Στόχος είναι η, κατά περίπτωση, παρουσίαση μετά το πέρας αυτού του βήματος είτε μίας λίστας αρχείων ομαδοποιημένων κατά φακέλους, είτε η παρουσίαση τριών διαφορετικών επιπέδων προτεινόμενων λύσεων όπου το πρώτο επίπεδο θα είναι τα πλέον πιθανά αποτελέσματα, δεύτερο θα είναι ελαφρώς λιγότερο πιθανά, και το τρίτο θα περιέχει όλα υπόλοιπα αρχεία ταξινομημένα σύμφωνα με τη βαθμονόμηση και ανά επίπεδο της κλίμακας βαθμονόμησης ομαδοποιημένα κατά φακέλους.

### 6.2.1 Διαδικασία Βαθμονόμησης

Για να επιτευχθεί η βαθμονόμηση των αποτελεσμάτων γίνεται αξιοποίηση της Κλίκας αρχείων η οποία έχει δημιουργηθεί σε προγενέστερο βήμα. Ο τρόπος με τον οποίο χρησιμοποιείτε βασίζεται στην υπόθεση ότι τα αρχεία που επιλέγονται για συμμετοχή στην κλίκα αναπαριστούν κάποιας μορφής πυρήνα για μία συγκεκριμένη αναφορά σφάλματος. Αν και τελικώς φαίνεται ότι κάτι τέτοιο δεν ισχύει απόλυτα, ωστόσο μπορεί να χρησιμοποιηθεί για τη βαθμονόμηση δίνοντας ικανοποιητικά αποτελέσματα.

Μετά τη διεξαγωγή του clustering δημιουργείται ένας πίνακας από τα ονόματα των αρχείων σε ακέραιες τιμές, ο οποίος μετράει πόσες συνδέσεις, μέσω σχέσεων, υπάρχουν ανάμεσα σε οποιονδήποτε από τους clusters (δηλαδή ανάμεσα σε αρχεία που περιέχονται σε αυτούς) που έχουν δημιουργηθεί και τους clusters οι οποίοι περιέχουν μέρος της Κλίκας. Με αυτόν τον τρόπο επιτυγχάνεται η ανάθεση μία τιμής σε κάθε cluster η οποία, σύμφωνα πάντα με την υπόθεση, δείχνει την 'απόσταση' από την Κλίκα (η οποία Κλίκα πλέον θεωρείται ότι έχει μεταφραστεί σε clusters), η οποία λειτουργεί σαν ένα άτυπο Centroid κάποιου data clustering αλγορίθμου. Πλέον, διαθέτοντας αυτές τις τιμές, υπάρχει το πρόβλημα ότι δεν είναι εφικτό να ορίσουμε μία διάταξη η οποία να έχει νόημα και επίσης να 'ξεχωρίζει' τα αρχεία που μας ενδιαφέρουν. Για το λόγο αυτό οι τιμές ομαλοποιούνται χρησιμοποιώντας μία εμπειρική σχέση η οποία προέκυψε κατόπιν πειραματισμού με διάφορες τροποποιήσεις σχέσεων οι οποίες θα μπορούσαν να χρησιμοποιηθούν για την επίτευξη πιο κατανοητών αποτελεσμάτων. Η σχέση που τελικά χρησιμοποιήθηκε είναι η:

$$cluster\_score = \frac{cluster\_connectivity}{cluster\_To\_Clique\_Connectivity} * cluster\_size \quad (6.1)$$

Η ολοκλήρωση της εφαρμογής και αυτής της σχέσης επιτρέπει πλέον την κατάταξη των αρχείων με βάση την επιλεγείσα μετρική. Δυστυχώς παρατηρήθηκε ότι δεν είναι εφικτό να καταταχθούν όλα τα αρχεία χρησιμοποιώντας αποκλειστικά αυτήν την μετρική. Ο λόγος είναι ο ίδιος για τον οποίο δεν ισχύει απόλυτα η υπόθεση ότι η Κλίκα αποτελεί κάποιον ουσιαστικό πυρήνα για τις οντότητες (αρχεία) της εκάστοτε αναφοράς σφάλματος. Συγκεκριμένα προκύπτει ότι υπάρχει ένα μη αμελητέο πλήθος αρχείων το οποίο δε συμμετέχει σε σχέση με κανένα αρχείο είτε από την Κλίκα είτε με αρχεία τα οποία ανήκουν σε ίδιο cluster με στοιχεία αυτής, τουλάχιστον άμεσα. Αυτό οδηγεί στην παρουσία μηδενικών τιμών στην βαθμονόμηση των αρχείων μέσα σε κάποιους από τους δημιουργηθέντες clusters. Αγνωώντας, για λόγους που θα εξηγηθούν παρακάτω, το συγκεκριμένο ζήτημα γίνεται παρουσίαση των clusters που έχουν επιτύχει το μεγαλύτερο σκορ, σε φθίνουσα σειρά, παραθέτοντας το όνομα κάθε cluster μαζί με το αντίστοιχο σκορ.

Σε αυτό το στάδιο παρατηρήθηκε ότι σε κάθε σύστημα που αναλύθηκε ο αλγόριθμος του ACDC δημιούργησε έναν τουλάχιστον cluster ο οποίος είχε μέγεθος κάποιες φορές μεγαλύτερο από τους υπόλοιπους. Αυτή η παρατήρηση οδήγησε στη λήψη της απόφασης για ξεχωριστή παρουσίαση χρησιμοποιώντας κάποιο μέτρο που να συσχετίζει το μέγεθος κάθε cluster με το σκορ αυτού ούτως ώστε να εξαχθεί κάποια τιμή 'μαθηματικής ελπίδας'. Αυτή η τιμή χρησιμοποιήθηκε για να διαχωριστούν οι clusters σε δύο ομάδες. Η πρώτη ομάδα περιέχει τους clusters οι οποίοι έχουν τιμή 'μαθηματικής ελπίδας' μικρότερη από κάποιο κατώφλι, ενώ η δεύτερη αυτούς που έχουν τιμή μεγαλύτερη από το κατώφλι. Η τιμή κατωφλίου, όπως είναι φυσικό δεδομένης της ετερογένειας των αναλυθέντων συστημάτων, ορίζεται προσαρμοστικά και συσχετίζεται με τη μέση τιμή της 'μαθηματικής ελπίδας' για κάθε cluster.

Αφού έχει ολοκληρωθεί και ο διαχωρισμός στις δύο ομάδες με χρήση της 'μαθηματικής ελπίδας' το σύνολο με τους clusters με μεγάλη τιμή διατηρείται για το τελευταίο επίπεδο της τελικής αναφοράς. Για το άλλο χρησιμοποιείται η ήδη υπολογισθείσα συνδεσιμότητα κάθε αρχείου με τα αρχεία που ανήκουν είτε στην Κλίκα είτε σε κάποιον cluster στον οποίο ανήκει τουλάχιστον ένα στοιχείο της Κλίκας για την ταξινόμησή τους.

### 6.2.2 Παρουσίαση Αποτελεσμάτων

Όταν έχει ολοκληρωθεί και το στάδιο της βαθμονόμησης, κατάταξης και ταξινόμησης των αρχείων, πρέπει επιπλέον να παρουσιαστούν με κάποια μορφή στο χρήστη η οποία να βοηθάει τον χρήστη στον εντοπισμό του σφάλματος. Για αυτό το στάδιο επιλέχθηκε η παρουσίαση των αποτελεσμάτων σε τρία διαφορετικά επίπεδα. Το πρώτο επίπεδο παρουσιάζει τα υποψήφια αρχεία τα οποία ανήκουν στους clusters οι οποίοι συμπεριλαμβάνονται στο ανώτερο 50% των αποτελεσμάτων και τα οποία έχουν μη μηδενική συνδεσιμότητα με την Κλίκα ή αρχεία των clusters στους οποίους υπάρχει η Κλίκα. Για αυτό το πρώτο επίπεδο τυπώνεται για κάθε έναν cluster το όνομά του, και ακολουθούν τα ονόματα των περιεχόμενων αρχείων, ένα ανά γραμμή, ταξινομημένα από τα λιγότερο πιθανά προς τα πιο πιθανά, με βάση τη συνδεσιμότητα. Στα αποτελέσματα αυτά το πιο πιθανό αρχείο να περιέχει κώδικα σχετικό με το σφάλμα, σύμφωνα με την προτεινόμενη θεώρηση, βρίσκεται στην τελευταία γραμμή.

Μετά την παρουσίαση του πρώτου επιπέδου, γίνεται μία επακριβώς ίδια παρουσίαση η οποία παρουσιάζει τα αρχεία χωρίς συνδεσιμότητα προς την Κλίκα, αλλά η ομαδοποίηση γίνεται με βάση πρώτα τον cluster στον οποίο είναι και εσωτερικά με το φάκελο στον οποίο αυτά ανήκουν καθώς δεν υπάρχει κάποια διαθέσιμη μη μηδενική τιμή για να πραγματοποιηθεί κάποιου άλλου είδους σύγκριση. Οι clusters παρουσιάζονται όπως και στην πρώτο επίπεδο από τον λιγότερο πιθανό προς τον περισσότερο.

Τέλος το τρίτο και τελευταίο επίπεδο αναφοράς των αποτελεσμάτων διεξάγεται για τους clusters που είχαν χαρακτηριστεί από υψηλή τιμή 'μαθηματικής ελπίδας' για τους οποίους γίνεται μία παρουσίαση όπως ακριβώς για το πρώτο επίπεδο. Συγκεκριμένα οι clusters είναι ταξινομημένη από λιγότερο πιθανοί προς περισσότερο πιθανοί και εσωτερικά τα αρχεία με βάση τη συνδεσιμότητά τους, συμπεριλαμβανοντας και τις μηδενικές τιμές.

Καθώς παρατηρήθηκε ότι για κάποια συστήματα δεν είναι εφικτό να δημιουργηθεί ένα ικανού μεγέθους διευρυμένο σύνολο αρχείων με χρήση της LSA και της επέκτασης μέσω σχέσεων, η τεχνική βαθμονόμησης δεν χρησιμοποιείται σε περίπτωση που το σύνολο αρχείων περιέχει λιγότερα αρχεία από ένα κατώτατο κατώφλι το οποίο είναι ανάλογο του συνόλου των αρχείων που είναι διαθέσιμα στο σύστημα που αναλύεται κάθε φορά. Σε αυτήν την περίπτωση απλά παρουσιάζονται τα αρχεία ομαδοποιημένα κατά συνδεσιμότητα και ταξινομημένα κατά περιέχοντα φάκελο.

### 6.3 Αποτελέσματα

Για τη αξιολόγηση της προτεινόμενης μεθοδολογίας επιλέχθηκαν τέσσερα συστήματα για τα οποία υπάρχουν αναφορές σφάλματος στο KDE Bugzilla Repository για τα οποία παρτίθενται στον πίνακα 6.1 κάποια γενικά χαρακτηριστικά τους.

| System Name | SLOC    | # Files | # RSF Lines | # CDIF Lines | # total Reports |
|-------------|---------|---------|-------------|--------------|-----------------|
| amarok      | 265,116 | 2,089   | 120,591     | 1,007,278    | 17,976          |
| k3b         | 105,775 | 817     | 30,252      | 265,250      | 3,994           |
| konqueror   | 112,436 | 6,248   | 438,868     | 4,038,100    | 38,135          |
| kopete      | 387,753 | 2,720   | 117,076     | 1,024,230    | 9,765           |

ΠΙΝΑΚΑΣ 6.1: Πληροφορίες Συστημάτων που Χρησιμοποιήθηκαν για την Αξιολόγηση

Για την παρουσίαση των αποτελεσμάτων επιλέχθηκε η εκτύπωση τους σε .txt αρχεία. Ανάλογα με την περίπτωση τα αποτελέσματα αποθηκεύονται σε διαφορετικούς φακέλους. Η δομή του συστήματος αναφορών είναι:

Τύπος αναφοράς → Όνομα Συστήματος → Όνομα αναφοράς σφάλματος

Ο λόγος για τον οποίο επιλέχθηκε αυτή η δομή είναι ότι θεωρήθηκε πως είναι εύκολα κατανοητή, και διαισθητικά πιο κοντά σε αυτό που θα έβλεπε ένας προγραμματιστής καθώς θα πλοηγούταν στο σύστημα αρχείων ενός συστήματος με στόχο τον προσδιορισμό της θέσης ενός σφάλματος.

| System Name | Average Recall over all Reports |
|-------------|---------------------------------|
| amarok      | 0.8                             |
| k3b         | 0.86                            |
| konqueror   | 0.625                           |
| kopete      | 0.75                            |

ΠΙΝΑΚΑΣ 6.2: Report Level Recall

Οι τεχνική εφαρμοσμένη σε τέσσερα διαφορετικά συστήματα δίνει αρκετά καλά αποτελέσματα, συγκεκριμένα παρακάτω παρουσιάζεται για πόσες από τις διαθέσιμες αναφορές σφαλμάτων για κάθε σύστημα κατάφερε το σύστημα να παρουσιάσει τουλάχιστον ένα από τα αρχεία που ανήκουν στο σύνολο των προβληματικών αρχείων που προκάλεσαν εξάρχης το σφάλμα.

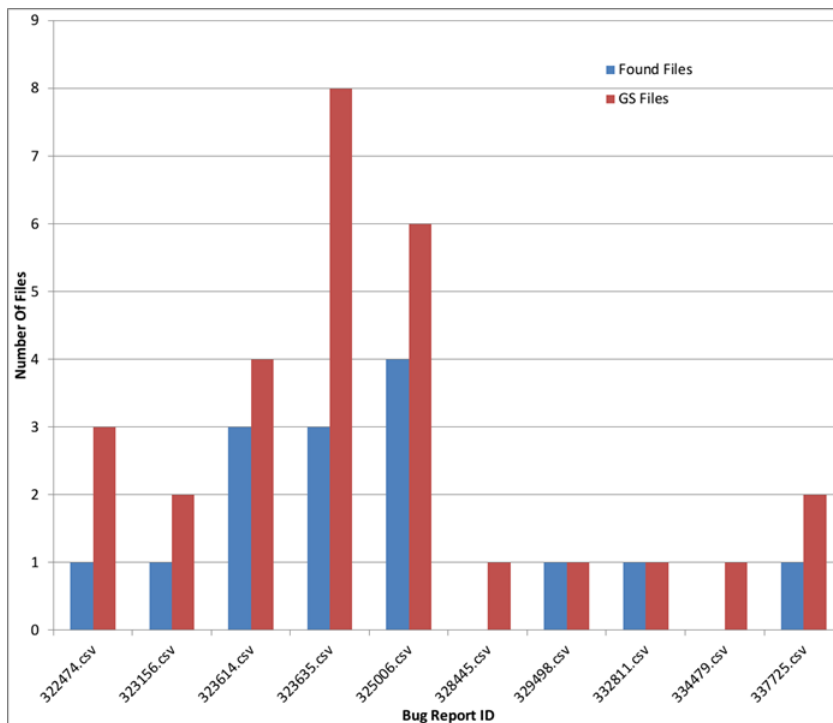
Από τον πίνακα 6.2 είναι εμφανές ότι η μέθοδος δύναται να παράξει εξαιρετικά αποτελέσματα, αν ενσωματωθούν κάποια ακόμα βήματα φιλτραρίσματος των προκαταρκτικών αυτών αποτελεσμάτων. Ο συγκεκριμένος τρόπος μέτρησης μπορεί εύκολα να θεωρηθεί ότι είναι ένας τρόπος να παρουσιαστούν κάποια κακά αποτελέσματα με όμορφο τρόπο. Η αλήθεια είναι ότι αυτός είναι ένας πολύ πραγματιστικός τρόπος παρουσίασης. Ο λόγος πίσω από αυτή την αντιμετώπιση είναι ότι στην πραγματικότητα ένας προγραμματιστής που έχει γνώση του συστήματος για το οποίο θα χρησιμοποιήσει την προτεινόμενη μεθοδολογία, θα εντοπίσει από αυτό το ένα αρχείο που σχετίζεται με το σφάλμα και όλα τα υπόλοιπα σχετικά αρχεία που πρέπει να τροποποιηθούν και προφανώς και όλα τα αρχεία βιβλιοθηκών που έχουν αφαιρεθεί νωρίτερα από το σύνολο των προτεινόμενων αρχείων.

Παρά το γεγονός ότι η μιμηλιστική αυτή παρουσίαση δίνει μία εικόνα της επιτυχίας του συστήματος με έναν, ακραία ίσως, συνοπτικό τρόπο, στην ουσία είναι ο πλέον κατάλληλος τρόπος παρουσίασης για να φανεί ακριβώς η ορθότητα της προσέγγισης παρά η απόλυτη επιτυχία αυτής. Ωστόσο για λόγους πληρότητας τα αποτελέσματα σε μία πλήρη μορφή τους παρατίθενται στα παραρτήματα Α, Β, Γ και Δ όπου στο Α παρουσιάζονται τα αποτελέσματα για το σύστημα amarok ενώ ακολουθούν στα επόμενα παραρτήματα κατά σειρά για τα συστήματα konqueror, k3b και kopete.

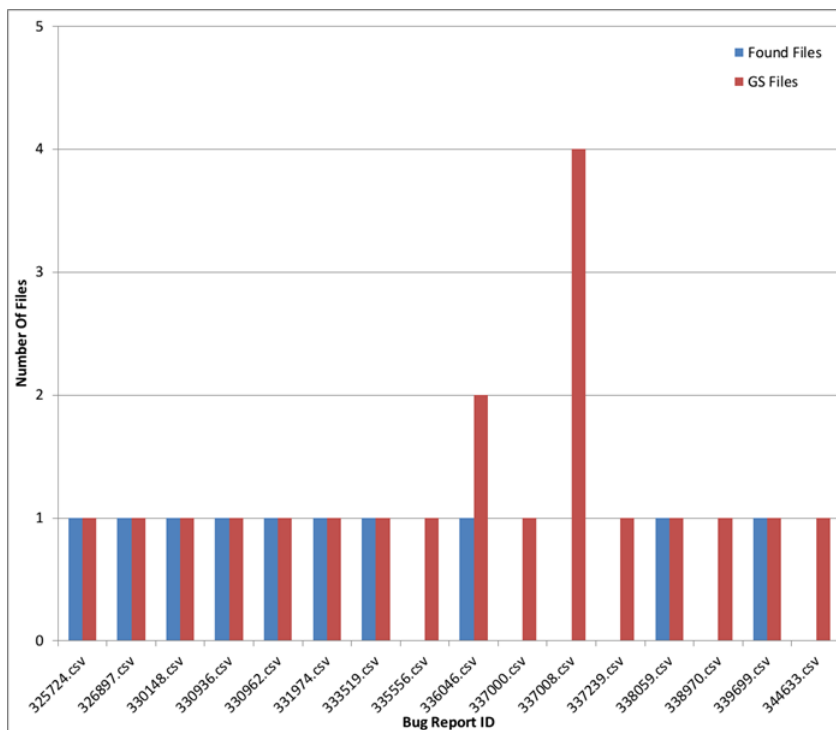
### 6.3.1 Αναλυτικά Αποτελέσματα

Στην παράγραφο αυτή γίνεται εκτενής παρουσίαση των χαρακτηριστικών των αποτελεσμάτων σε σχέση με το πόσες από τις σωστές απαντήσεις βρίσκει η μέθοδος σχετικά με της συνολικές σωστές απαντήσεις όπως αυτές βρέθηκαν στην επίσημη επίλυση για κάθε αναφορά.

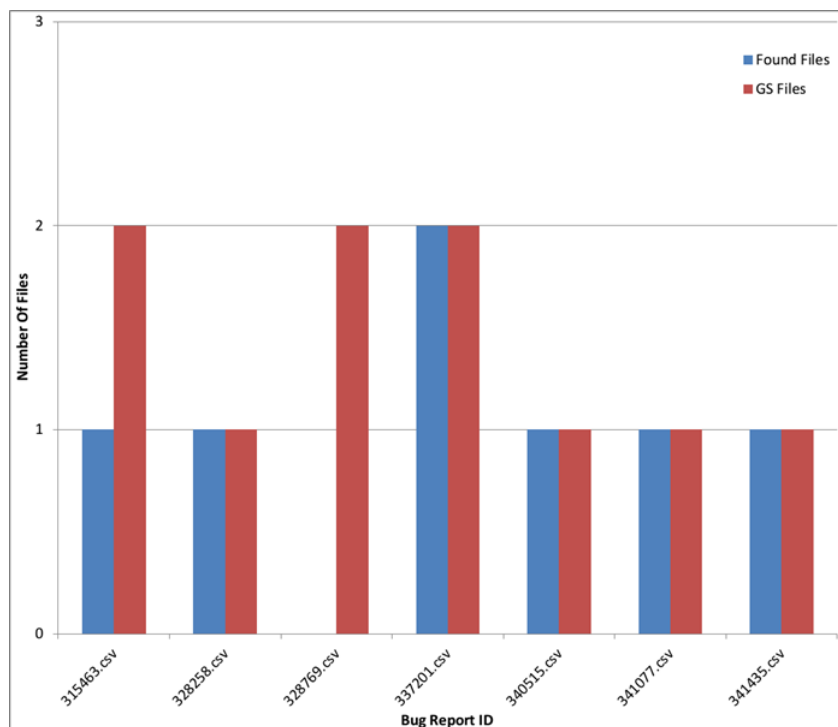
Στους κάτωθι πίνακες φαίνεται τη ποσοστό της λύσης βρίσκει η προτεινόμενη μέθοδος σε κάθε επίπεδο παρουσίασης όπως αυτά έχουν οριστεί παραπάνω.



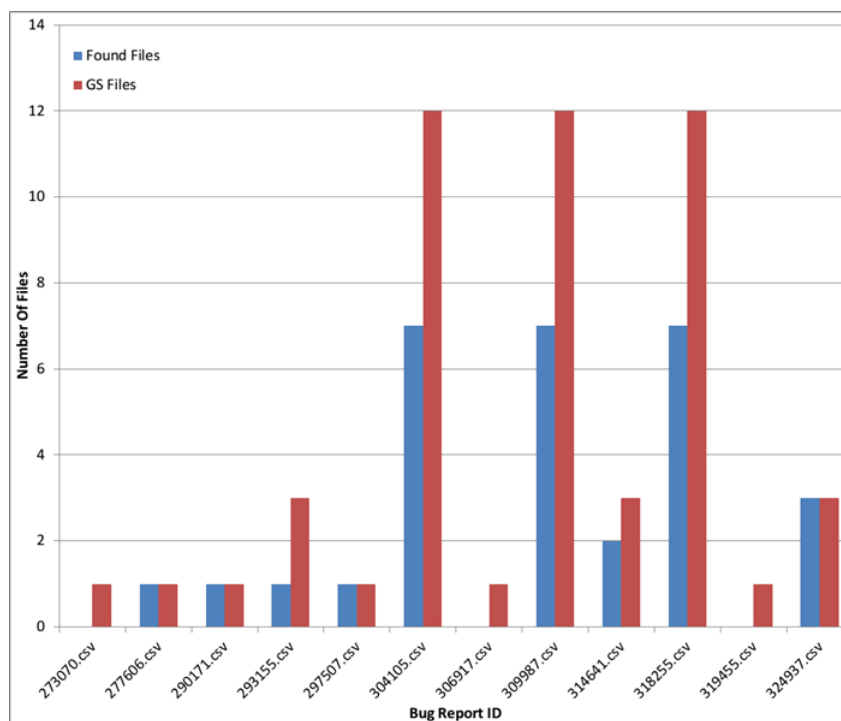
ΣΧΗΜΑ 6.2: Συγκεντρωτικά αποτελέσματα για το σύστημα Amarok



ΣΧΗΜΑ 6.3: Συγκεντρωτικά αποτελέσματα για το σύστημα konqueror



ΣΧΗΜΑ 6.4: Συγκεντρωτικά αποτελέσματα για το σύστημα k3b



ΣΧΗΜΑ 6.5: Συγκεντρωτικά αποτελέσματα για το σύστημα kopete

Επίσης παρατίθενται αναλυτικά για κάθε σύστημα και κάθε αναφορά που χρησιμοποιήθηκε για την αξιολόγηση της μεθόδου τα αποτελέσματα που σωστά αποτελέσματα που εμφανίζονται σε κάθε επίπεδο αναφοράς.



| A/A Αναφοράς Σφάλματος | 1 <sup>ο</sup> επίπεδο | 0 <sup>ο</sup> επίπεδο | Μεγάλοι clusters |
|------------------------|------------------------|------------------------|------------------|
| 322474                 | 0                      | 0                      | 1                |
| 323156                 | 1                      | 0                      | 0                |
| 323614                 | 3                      | 0                      | 0                |
| 323635                 | 2                      | 0                      | 1                |
| 325006                 | 2                      | 1                      | 1                |
| 328445                 | 0                      | 0                      | 0                |
| 329498                 | 0                      | 1                      | 0                |
| 332811                 | 1                      | 0                      | 0                |
| 334479                 | 0                      | 0                      | 0                |
| 337725                 | 0                      | 0                      | 1                |

ΠΙΝΑΚΑΣ 6.3: Επιτυχίες Ανά Επίπεδο για Amarok

| A/A Αναφοράς Σφάλματος | 1 <sup>ο</sup> επίπεδο | 0 <sup>ο</sup> επίπεδο | Μεγάλοι clusters |
|------------------------|------------------------|------------------------|------------------|
| 325724                 | 1                      | 0                      | 0                |
| 326897                 | 0                      | 1                      | 0                |
| 330148                 | 0                      | 1                      | 0                |
| 330936                 | 0                      | 1                      | 0                |
| 330962                 | 0                      | 1                      | 0                |
| 331974                 | 0                      | 1                      | 0                |
| 333519                 | 0                      | 1                      | 0                |
| 335556                 | 0                      | 0                      | 0                |
| 336046                 | 0                      | 1                      | 0                |
| 337000                 | 0                      | 0                      | 0                |
| 337008                 | 0                      | 0                      | 0                |
| 337239                 | 0                      | 0                      | 0                |
| 338059                 | 1                      | 0                      | 0                |
| 338970                 | 0                      | 0                      | 0                |
| 339699                 | 1                      | 0                      | 0                |
| 344633                 | 0                      | 0                      | 0                |

ΠΙΝΑΚΑΣ 6.4: Επιτυχίες Ανά Επίπεδο για konqueror

| A/A Αναφοράς Σφάλματος | 1 <sup>ο</sup> επίπεδο |
|------------------------|------------------------|
| 315463                 | 1                      |
| 328258                 | 0                      |
| 328769                 | 2                      |
| 337201                 | 1                      |
| 340515                 | 1                      |
| 341077                 | 1                      |
| 341435                 | 1                      |

ΠΙΝΑΚΑΣ 6.5: Επιτυχίες για το k3b

| A/A Αναφοράς Σφάλματος | 1 <sup>ο</sup> επίπεδο |
|------------------------|------------------------|
| 273070                 | 0                      |
| 277606                 | 1                      |
| 290171                 | 1                      |
| 293155                 | 1                      |
| 297507                 | 1                      |
| 304105                 | 7                      |
| 306917                 | 0                      |
| 309987                 | 7                      |
| 314641                 | 2                      |
| 318255                 | 7                      |
| 319455                 | 0                      |
| 324937                 | 3                      |

ΠΙΝΑΚΑΣ 6.6: Επιτυχίες για το kopete

Όπως φαίνεται από τους πιο αναλυτικούς πίνακες που παρατέθηκαν παραπάνω η μέθοδος βγάζει πολύ καλά αποτελέσματα για όλα τα συστήματα χάρη στον τρόπο με τον οποίο σχεδιάστηκε το σύστημα που προσαρμόζεται τόσο στο μέγεθος του συστήματος όσο και στον τρόπο με τον οποίο το σύστημα είναι δομημένο μέσω του εργαλείου ACDC, καθώς και στον ιδιαίτερο τρόπο με τον οποίο γίνονται οι αναφορές για κάθε σύστημα μέσω της χρήσης της τεχνικής LSA.

### 6.3.2 Μορφοποίηση Εξόδου

Παρατίθενται δειγματοληπτικά κάποιες περιπτώσεις καλών και κακών αποτελεσμάτων για τα συστήματα amarok και kopete. Στην παράθεση 6.7 φαίνεται το πρώτο επίπεδο αναφορών για την αναφορά υπ' αριθμόν 323614 όπως αυτό προκύπτει μετά την εφαρμογή της προτεινόμενης μεθοδολογίας. Παρατηρείται η παρουσίαση των ταξινομημένων με βάση την σχέση 6.1 clusters και πέρα από την πληροφορία που παρουσιάζεται επισημειώνεται και κάθε cluster ο οποίος περιέχει τουλάχιστον ένα τμήμα της λύσης. Η Διάταξη είναι ανάστροφη οπότε τα καλύτερα αποτελέσματα είναι στο κάτω μέρος και τα χειρότερα προς τα επάνω. Σε αυτήν την περίπτωση είναι εύκολο να εξακριβωθεί ότι ένας προγραμματιστής που θα αξιοποιούσε αυτή τη μέθοδο θα χρειαζόταν να ψάξει ανάμεσα σε λιγότερα από 50 αρχεία μέχρι να βρει το πρώτο αρχείο που να σχετίζεται με το σφάλμα, άρα υπάρχει μείωση του χώρου αναζήτησης κατά πάνω από 97.5%. Κάτω από την παρουσίαση των clusters παρατίθενται και τα περιεχόμενα αρχεία του κάθε ενός τα οποία έχουν μη μηδενικές σχέσεις με τα αρχεία της Κλίκας, όπως φαίνεται και στην παράθεση 6.8.

| ClusterName   | Score           | Expectation     | Connections      | ToClique         | Size     |
|---|-----------------|-----------------|------------------|------------------|----------|
| src/scriptmanager.cpp   | 6.285714        | 0.074387        | 11.000000        | 7.000000         | 4        |
| src/svghandler.cpp  | 6.285714        | 0.074387        | 11.000000        | 7.000000         | 4        |
| src/services/gpodder/gpodderservice.cpp                           | 6.300000        | 0.055917        | 21.000000        | 10.000000        | 3        |
| src/playlist/layouts/layouteditdialog.cpp                         | 6.333333        | 0.056213        | 19.000000        | 9.000000         | 3        |
| <b>src/playlist/playlistsortwidget.cpp</b>                        | <b>6.666667</b> | <b>0.039448</b> | <b>10.000000</b> | <b>3.000000</b>  | <b>2</b> |
| src/amarokurls/amarokurlhandler.cpp                               | 6.851852        | 0.101359        | 37.000000        | 27.000000        | 5        |
| src/services/lastfm/lastfmservice.cpp                             | 7.076923        | 0.083751        | 23.000000        | 13.000000        | 4        |
| tests/core-impl/collections/db/sql/testsqlscanmanager.cpp         | 7.500000        | 0.110947        | 9.000000         | 6.000000         | 5        |
| src/core-impl/collections/playdarcollection/playdarquerymaker.cpp | 8.000000        | 0.047337        | 4.000000         | 1.000000         | 2        |
| <b>src/mainwindow.cpp</b>   | <b>8.880000</b> | <b>0.157633</b> | <b>37.000000</b> | <b>25.000000</b> | <b>6</b> |
| src/aboutdialog/extendedaboutdialog.cpp                           | 9.000000        | 0.079882        | 9.000000         | 3.000000         | 3        |
| src/playlist/proxymodels/sortfilterproxy.cpp                      | 10.500000       | 0.093195        | 7.000000         | 2.000000         | 3        |

|   |                  |                 |                  |                  |           |
|---|------------------|-----------------|------------------|------------------|-----------|
| src/covermanager/covermanager.cpp                                 | 11.368421        | 0.269075        | 54.000000        | 38.000000        | 8         |
| src/services/amazon/amazonstore.cpp                               | 11.666667        | 0.172584        | 14.000000        | 6.000000         | 5         |
| src/services/magnatune/magnatunestore.cpp                         | 12.600000        | 0.335503        | 21.000000        | 15.000000        | 9         |
| src/services/scriptable/scriptableservicemanager.cpp              | 12.800000        | 0.151479        | 16.000000        | 5.000000         | 4         |
| src/context/applets/albums/albums.cpp                             | 14.000000        | 0.289941        | 44.000000        | 22.000000        | 7         |
| src/context/engines/wikipedia/wikipediaengine.cpp                 | 15.157895        | 0.538150        | 96.000000        | 76.000000        | 12        |
| src/services/amazon/amazonshoppingcartview.h                      | 16.000000        | 0.378698        | 12.000000        | 6.000000         | 8         |
| src/dialogs/tagguesserdialog.cpp                                  | 20.571429        | 0.365173        | 24.000000        | 7.000000         | 6         |
| src/context/applets/upcomingevents/upcomingeventsapplet.cpp       | 21.083333        | 0.686144        | 46.000000        | 24.000000        | 11        |
| src/core-impl/collections/ipodcollection/ipodplaylistprovider.cpp | 22.368421        | 1.125039        | 25.000000        | 19.000000        | 17        |
| src/core-impl/collections/db/sql/sqlscanresultprocessor.h         | 24.888889        | 1.030901        | 16.000000        | 9.000000         | 14        |
| src/core-impl/collections/umscollection/umscollectionlocation.cpp | 31.214286        | 1.754649        | 23.000000        | 14.000000        | 19        |
| <b>tests/synchronization/testonewaysynchronizationjob.cpp</b>     | <b>47.971429</b> | <b>3.264328</b> | <b>73.000000</b> | <b>35.000000</b> | <b>23</b> |
| src/browsers/browserbreadcrumbwidget.cpp                          | 50.666667        | 2.848126        | 88.000000        | 33.000000        | 19        |
| src/configdialog/dialogs/scriptsconfig.cpp                        | Infinity         | Infinity        | 4.000000         | 0.000000         | 3         |
| src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp | Infinity         | Infinity        | 1.000000         | 0.000000         | 1         |

ΠΙΝΑΚΑΣ 6.7: Αποτελέσματα Cluster Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 323614

| File Name   | Connectivity To Clique |
|---|------------------------|
| <b>Cluster Name : src/covermanager/covermanager.cpp</b> |                        |
| src/configdialog/dialogs/excludedlabelsdialog.cpp       | 2.000000               |
| src/covermanager/coverfetchunit.cpp                     | 2.000000               |

|  |          |
|--|----------|
| src/statusbar/compoundprogressbar.cpp  | 2.000000 |
| src/covermanager/coverfounddialog.cpp  | 3.000000 |
| src/amarokurls/bookmarktreeview.cpp  | 8.000000 |
| src/covermanager/covermanager.cpp  | 9.000000 |
| <b>Cluster Name</b> :src/configdialog/dialogs/scriptsconfig.cpp              |          |
|  |          |
| <b>Cluster Name</b> : src/services/amazon/amazonshoppingcartview.h           |          |
| src/playlist/proxymodels/proxybase.cpp                                       | 1.000000 |
| src/browsers/playlistbrowser/playlistsbyproviderproxy.cpp                    | 1.000000 |
| src/playlist/playlistqueueeditor.cpp   | 1.000000 |
| tests/dynamic/testdynamicmodel.cpp   | 1.000000 |
| src/browsers/playlistbrowser/qtgroupingproxy.cpp                             | 2.000000 |
| <b>Cluster Name</b> : src/context/applets/albums/albums.cpp                  |          |
| src/context/applets/albums/trackitem.cpp                                     | 1.000000 |
| src/context/applets/albums/albums.cpp  | 1.000000 |
| src/context/applets/albums/albumsmodel.cpp                                   | 2.000000 |
| src/context/applets/wikipedia/wikipediaapplet.cpp                            | 2.000000 |
| src/context/applets/albums/albumitem.cpp                                     | 2.000000 |
| src/context/applets/albums/albumsview.cpp                                    | 6.000000 |
| src/context/applets/upcomingevents/upcomingeventsstack.cpp                   | 8.000000 |
| <b>Cluster Name</b> : tests/synchronization/testonewaysynchronizationjob.cpp |          |

|  |          |
|--|----------|
| textitsrc/playlist/playlistviewurlrunner.cpp                                     | 1.000000 |
| shared/filetype.cpp  | 1.000000 |
| src/browsers/singlecollectiontreeitemmodel.cpp                                   | 1.000000 |
| tests/testtrackorganizer.cpp   | 2.000000 |
| src/services/mp3tunes/mp3tunesservice.cpp  | 2.000000 |
| tests/browsers/testsinglecollectiontreeitemmodel.cpp                             | 3.000000 |
| src/musicbrainz/musicdnssaudiodecoder.cpp  | 3.000000 |
| src/browsers/collectiontreeitemmodelbase.cpp                                     | 4.000000 |
| src/browsers/collectiontreeitemmodel.cpp   | 4.000000 |
| tests/synchronization/testonewaysynchronizationjob.cpp                           | 6.000000 |
| src/core-impl/collections/nepomukcollection/nepomukcache.cpp                     | 8.000000 |
| <b>Cluster Name</b> : src/services/lastfm/lastfm.service.cpp                     |          |
| src/services/lastfm/lastfmservicesettings.cpp                                    | 2.000000 |
| src/services/lastfm/lastfmtreemodel.cpp  | 4.000000 |
| src/services/lastfm/lastfm.service.cpp   | 4.000000 |
| <b>Cluster Name</b> : src/core-impl/collections/db/sql/sqlscanresultprocessor.h  |          |
| src/core-impl/collections/db/mountpointmanager.cpp                               | 1.000000 |
| src/dialogs/deletedialog.cpp   | 1.000000 |
| src/core-impl/collections/db/sql/device/massstorage/massstoragedevicehandler.cpp | 1.000000 |
| tests/core-impl/collections/db/sql/testsqlquerymaker.cpp                         | 1.000000 |
| src/scanner/abstractscanresultprocessor.cpp                                      | 2.000000 |
| src/core-impl/collections/db/sql/sqlscanresultprocessor.cpp                      | 3.000000 |

| <b>Cluster Name : src/core-impl/collections/ipodcollection/ipodplaylistprovider.cpp</b> |                 |
|---|-----------------|
| src/opmlwriter.cpp  | 1.000000        |
| src/core-impl/support/urlstatisticsstore.cpp  | 1.000000        |
| src/widgets/horizontaldivider.cpp   | 1.000000        |
| src/services/lastfm/meta/lastfmmultiplayablecapability.cpp                              | 1.000000        |
| tests/core-impl/meta/file/testmetafiletrack.cpp   | 1.000000        |
| src/context/widgets/recentlyplayedlistwidget.cpp  | 1.000000        |
| src/transcoding/transcodingpropertysliderwidget.cpp                                     | 1.000000        |
| src/services/lastfm/scrobbleradapter.cpp  | 1.000000        |
| src/synchronization/synchronizationbasejob.cpp  | 1.000000        |
| src/services/opmldirectory/opmldirectorydatabasehandler.cpp                             | 1.000000        |
| src/amarokmimedata.cpp  | 2.000000        |
| src/core-impl/collections/ipodcollection/ipodplaylist.cpp                               | 2.000000        |
| src/core-impl/collections/ipodcollection/ipodplaylistprovider.cpp                       | 5.000000        |
| <b>Cluster Name : src/playlist/playlistsortwidget.cpp</b>                               |                 |
| <b>textitsrc/playlist/playlistsortwidget.cpp</b>  | <b>1.000000</b> |
| src/playlist/playlistmodelstack.cpp   | 2.000000        |
| <b>Cluster Name : src/aboutdialog/extendedaboutdialog.cpp</b>                           |                 |
| src/aboutdialog/extendedaboutdialog.cpp   | 1.000000        |
| src/aboutdialog/ocspersonitem.cpp   | 1.000000        |
| src/aboutdialog/animatedwidget.cpp  | 1.000000        |
| <b>Cluster Name : src/context/engines/wikipedia/wikipediaengine.cpp</b>                 |                 |

|   |          |
|---|----------|
| src/services/mp3tunes/mp3tunesworkers.cpp   | 1.000000 |
| src/context/engines/wikipedia/wikipediaengine.cpp                                       | 2.000000 |
| <b>Cluster Name : src/services/scriptable/scriptableservicemanager.cpp</b>              |          |
| src/services/scriptable/scriptableservice.cpp   | 1.000000 |
| src/services/scriptable/scriptableservicemeta.cpp                                       | 1.000000 |
| src/browsers/browsercategorylistmodel.cpp   | 1.000000 |
| src/services/scriptable/scriptableservicemanager.cpp                                    | 2.000000 |
| <b>Cluster Name : tests/core-impl/collections/db/sql/testsqlscanmanager.cpp</b>         |          |
| tests/core-impl/collections/db/sql/testdatabaseupdater.cpp                              | 1.000000 |
| src/core-impl/collections/db/sql/mysqlcollection/mysqlembeddedstorage.cpp               | 1.000000 |
| src/core-impl/collections/db/sql/mysqlservercollection/mysqlservercollection.cpp        | 1.000000 |
| tests/core-impl/collections/db/sql/testsqlscanmanager.cpp                               | 3.000000 |
| <b>Cluster Name : src/services/magnatune/magnatunestore.cpp</b>                         |          |
| src/services/lastfm/lovetrackaction.cpp   | 1.000000 |
| src/core-impl/support/tagstatisticsstore.cpp  | 1.000000 |
| src/services/magnatune/magnatunedatabaseworker.cpp                                      | 2.000000 |
| src/dialogs/collectionsetup.cpp   | 3.000000 |
| src/services/magnatune/magnatunestore.cpp   | 4.000000 |
| <b>Cluster Name : src/core-impl/collections/umscollection/umscollectionlocation.cpp</b> |          |
| src/core-impl/collections/umscollection/umscollection.cpp                               | 1.000000 |
| src/services/mp3tunes/libmp3tunes/harmony.c   | 1.000000 |



|   |          |
|---|----------|
| tests/testsmartpointerlist.cpp  | 1.000000 |
| src/aboutdialog/ocspersonlistwidget.cpp                               | 1.000000 |
| src/musicbrainz/musicdnsfinder.cpp                                    | 1.000000 |
| src/dialogs/trackorganizer.cpp  | 3.000000 |
| src/core-impl/collections/umscollection/umscollectionlocation.cpp     | 3.000000 |
| src/covermanager/coverfetchingactions.cpp                             | 3.000000 |
| <b>Cluster Name : src/services/amazon/amazonstore.cpp</b>             |          |
| src/services/amazon/amazoninfoparser.cpp                              | 1.000000 |
| src/services/amazon/amazonitemtreemodel.cpp                           | 1.000000 |
| src/services/amazon/amazonstore.cpp                                   | 3.000000 |
| <b>Cluster Name : src/svghandler.cpp</b>                              |          |
| src/svgtinter.cpp   | 1.000000 |
| src/svghandler.cpp  | 3.000000 |
| src/moodbar/moodbarmanager.cpp  | 3.000000 |
| <b>Cluster Name : src/browsers/browserbreadcrumbwidget.cpp</b>        |          |
| src/browsers/browsermessagearea.cpp                                   | 1.000000 |
| src/services/lastfm/synchronizationadapter.cpp                        | 1.000000 |
| src/core-impl/collections/db/sql/device/nfs/nfsdevicehandler.cpp      | 1.000000 |
| tests/qt-modeltest/modeltest.cpp                                      | 1.000000 |
| src/browsers/browserbreadcrumbitem.cpp                                | 1.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipoddeletetracksjob.cpp | 1.000000 |
| src/core-impl/collections/db/sql/device/smb/smbdevicehandler.cpp      | 1.000000 |

|  |                 |
|--|-----------------|
| src/amarokurls/navigationurlrunner.cpp                           | 2.000000        |
| src/services/amazon/amazonitemtreeview.cpp                       | 2.000000        |
| src/browsers/infoproxy.cpp                                       | 2.000000        |
| tests/core-impl/collections/db/sql/testsqlcollectionlocation.cpp | 2.000000        |
| src/browsers/servicebrowser/servicebrowser.cpp                   | 2.000000        |
| src/browsers/browsercategory.cpp                                 | 3.000000        |
| src/core/support/amarok.cpp                                      | 4.000000        |
| src/browsers/browserbreadcrumbwidget.cpp                         | 4.000000        |
| src/browsers/browsercategorylist.cpp                             | 5.000000        |
| <b>Cluster Name : src/services/gpodder/gpodderservice.cpp</b>    |                 |
| src/services/gpodder/gpodderservice.cpp                          | 1.000000        |
| src/services/servicecollectiontreeview.cpp                       | 4.000000        |
| src/services/servicebase.cpp                                     | 5.000000        |
| <b>Cluster Name : src/mainwindow.cpp</b>                         |                 |
| src/browsers/browserdock.cpp                                     | 1.000000        |
| src/amarokurls/bookmarkmanager.cpp                               | 3.000000        |
| src/dialogs/equalizerdialog.cpp                                  | 4.000000        |
| <b>textitsrc/mainwindow.cpp</b>                                  | <b>6.000000</b> |
| <b>Cluster Name : src/dialogs/tagguesserdialog.cpp</b>           |                 |
| src/widgets/token.cpp  | 1.000000        |
| src/widgets/tokenpool.cpp  | 2.000000        |
| src/widgets/filenamelayoutwidget.cpp                             | 4.000000        |

|   |          |
|---|----------|
| <b>Cluster Name</b> : src/core-impl/collections/playdarcollection/playdarquerymaker.cpp |          |
| src/core-impl/collections/playdarcollection/playdarcollection.cpp                       | 1.000000 |
| <b>Cluster Name</b> : src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp |          |
|   |          |
| <b>Cluster Name</b> : src/playlist/proxymodels/sortfilterproxy.cpp                      |          |
| src/playlist/proxymodels/sortalgorithms.cpp   | 1.000000 |
| src/playlist/proxymodels/sortscheme.cpp   | 1.000000 |
| <b>Cluster Name</b> : src/context/applets/upcomingevents/upcomingeventsapplet.cpp       |          |
| src/context/engines/similarartists/similarartistsengine.cpp                             | 1.000000 |
| src/scriptengine/amaroknetworkscript.cpp  | 2.000000 |
| src/network/networkaccessviewer.cpp   | 2.000000 |
| src/context/applets/upcomingevents/upcomingeventsmapwidget.cpp                          | 2.000000 |
| src/context/engines/labels/labelsengine.cpp   | 2.000000 |
| src/covermanager/coverfetcher.cpp   | 4.000000 |
| src/context/applets/upcomingevents/upcomingeventsapplet.cpp                             | 4.000000 |
| src/network/networkaccessmanagerproxy.cpp   | 7.000000 |
| <b>Cluster Name</b> : src/scriptmanager.cpp   |          |
| src/scriptmanager.cpp   | 4.000000 |
| <b>Cluster Name</b> : src/amarokurls/amarokurlhandler.cpp                               |          |
| src/amarokurls/bookmarkcurrentbutton.cpp  | 1.000000 |
| src/amarokurls/contexturlrunner.cpp   | 1.000000 |

|   |          |
|---|----------|
| src/amarokurls/bookmarkgroup.cpp                                | 2.000000 |
| src/amarokurls/amarokurlhandler.cpp                             | 3.000000 |
| src/amarokurls/bookmarkmodel.cpp                                | 5.000000 |
| <b>Cluster Name : src/playlist/layouts/layouteditdialog.cpp</b> |          |
| src/playlist/layouts/layouteditdialog.cpp                       | 1.000000 |
| src/widgets/tokendroptarget.cpp                                 | 3.000000 |
| src/widgets/tokenwithlayout.cpp                                 | 5.000000 |

ΠΙΝΑΚΑΣ 6.8: Αποτελέσματα File Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 323614

Στην παράθεση 6.9 παρουσιάζεται η αναπαράσταση που επιλέγεται για την παρουσίαση των υποψήφιων προβληματικών αρχείων όπως έχει προκύψει για την αναφορά σφάλματος υπ' αριθμόν 323635 τα οποία κατά τη διαδικασία της ομαδοποίησης, ομαδοποιούνται σε έναν πολύ μεγάλο cluster. Καθώς υπάρχουν πολλά αρχεία και δεν υπάρχει αξιόπιστος τρόπος για την σωστή διάταξη όλων, επιλέγεται η παρουσίαση των αποτελεσμάτων κατά ομάδες συνδεσιμότητας προς την Κλίμα, στην παράθεση 6.10 φαίνεται εντός κάθε ομάδας η ομαδοποίηση με βάση την δομή των περιεχόντων φακέλων. Είναι φανερό ότι με αυτόν τον τρόπο παρουσίασης ένας ειδικευμένος προγραμματιστής θα αναγνωρίσει εύκολα τα αρχεία που μπορεί να τον ενδιαφέρουν παρά την απουσία απόλυτα ταξινομημένων αποτελεσμάτων.

| ClusterName                              | Score             | expectation       | Connections       | ToClique          | ClusterSize |
|--|-------------------|-------------------|-------------------|-------------------|-------------|
| <i>src/covermanager/covermanager.cpp</i> | <i>390.017341</i> | <i>168.130643</i> | <i>918.000000</i> | <i>346.000000</i> | <i>147</i>  |

ΠΙΝΑΚΑΣ 6.9: Αποτελέσματα Cluster Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 323635

|   |
|---|
| <b>Cluster Name : src/covermanager/covermanager.cpp</b> |
|---|

| <b>File Name</b>                                    | <b>Connectivity To Clique</b> |
|---|-------------------------------|
| src/app.cpp   | 4.000000                      |
| src/toolbar/maintoolbar.cpp                         | 2.000000                      |
| src/amarokurls/amarokurl.cpp                        | 1.000000                      |
| src/amarokurls/bookmarkmodel.cpp                    | 1.000000                      |
| src/amarokurls/bookmarktreeview.cpp                 | 1.000000                      |
| src/amarokurls/playurlrunner.cpp                    | 1.000000                      |
| src/configdialog/dialogs/notificationsconfig.cpp    | 1.000000                      |
| src/context/applets/info/infoapplet.cpp             | 1.000000                      |
| src/context/applets/labels/labelsapplet.cpp         | 1.000000                      |
| src/context/applets/songkick/songkickapplet.cpp     | 1.000000                      |
| src/context/applets/tabs/tabsview.cpp               | 1.000000                      |
| src/context/applets/wikipedia/wikipediaapplet.cpp   | 1.000000                      |
| src/context/engines/labels/labelsengine.cpp         | 1.000000                      |
| src/context/widgets/containmentselectionlayer.cpp   | 1.000000                      |
| src/covermanager/coverfetcher.cpp                   | 1.000000                      |
| src/covermanager/coverfounddialog.cpp               | 1.000000                      |
| src/enginecontroller.cpp                            | 1.000000                      |
| src/services/amazon/amazonitemtreeview.cpp          | 1.000000                      |
| src/services/opmldirectory/opmldirectoryservice.cpp | 1.000000                      |
| src/statusbar/longmessagewidget.cpp                 | 1.000000                      |
| src/widgets/bookmarkpopup.cpp                       | 1.000000                      |

|  |                 |
|--|-----------------|
| src/widgets/bookmarktriangle.cpp                               | 1.000000        |
| src/widgets/editdeletedelegate.cpp                             | 1.000000        |
| src/widgets/flowlayout.cpp                                     | 1.000000        |
| src/widgets/osd.cpp  | 1.000000        |
| src/widgets/prettytreedelegate.cpp                             | 1.000000        |
| src/widgets/volumedial.cpp                                     | 1.000000        |
| shared/filetype.cpp  | 0.000000        |
| src/aboutdialog/ocspersonlistwidget.cpp                        | 0.000000        |
| src/amarokmimedata.cpp   | 0.000000        |
| src/amarokurls/bookmarkcurrentbutton.cpp                       | 0.000000        |
| src/amarokurls/bookmarkgroup.cpp                               | 0.000000        |
| src/browsers/browsermessagearea.cpp                            | 0.000000        |
| src/browsers/collectiontreeitemmodel.cpp                       | 0.000000        |
| src/browsers/collectiontreeview.cpp                            | 0.000000        |
| src/browsers/filebrowser/fileview.cpp                          | 0.000000        |
| src/configdialog/dialogs/excludedlabelsdialog.cpp              | 0.000000        |
| src/context/applets/analyzer/analyzerapplet.cpp                | 0.000000        |
| <b>src/context/applets/analyzer/discoanalyzer.cpp</b>          | <b>0.000000</b> |
| src/context/applets/upcomingevents/upcomingeventsstack.cpp     | 0.000000        |
| src/context/applets/upcomingevents/upcomingeventsstackitem.cpp | 0.000000        |
| src/context/engines/lyrics/lyricsengine.cpp                    | 0.000000        |
| src/context/engines/photos/photosengine.cpp                    | 0.000000        |

|  |          |
|--|----------|
| src/context/engines/similarartists/similarartistsengine.cpp                      | 0.000000 |
| src/context/engines/tabs/tabsengine.cpp  | 0.000000 |
| src/context/engines/wikipedia/wikipediaengine.cpp                                | 0.000000 |
| src/context/widgets/recentlyplayedlistwidget.cpp                                 | 0.000000 |
| src/context/widgets/textscrollingwidget.cpp                                      | 0.000000 |
| src/core-impl/collections/audiocd/audiocdcollection.cpp                          | 0.000000 |
| src/core-impl/collections/db/sql/device/massstorage/massstoragedevicehandler.cpp | 0.000000 |
| src/core-impl/collections/db/sql/device/nfs/nfsdevicehandler.cpp                 | 0.000000 |
| src/core-impl/collections/db/sql/device/smb/smbdevicehandler.cpp                 | 0.000000 |
| src/core-impl/collections/db/sql/mysql-shared/mysqlstorage.cpp                   | 0.000000 |
| src/core-impl/collections/db/sql/mysqlcollection/mysqlembeddedstorage.cpp        | 0.000000 |
| src/core-impl/collections/db/sql/mysqlservercollection/mysqlservercollection.cpp | 0.000000 |
| src/core-impl/collections/db/sql/sqlcapabilities.cpp                             | 0.000000 |
| src/core-impl/collections/db/sql/sqlregistry_p.cpp                               | 0.000000 |
| src/core-impl/collections/ipodcollection/ipodcollection.cpp                      | 0.000000 |
| src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp               | 0.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipodcopytracksjob.cpp              | 0.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipoddeletetracksjob.cpp            | 0.000000 |
| src/core-impl/collections/support/collectionmanager.cpp                          | 0.000000 |
| src/core-impl/collections/support/memorycustomvalue.cpp                          | 0.000000 |
| src/core-impl/collections/support/memorymatcher.cpp                              | 0.000000 |
| src/core-impl/meta/multi/multitrack.cpp  | 0.000000 |

|  |          |
|--|----------|
| src/core-impl/meta/timecode/timecodeobserver.cpp     | 0.000000 |
| src/core-impl/support/tagstatisticsstore.cpp         | 0.000000 |
| src/core-impl/support/urlstatisticsstore.cpp         | 0.000000 |
| src/core/podcasts/podcastimagefetcher.cpp            | 0.000000 |
| src/covermanager/covercache.cpp                      | 0.000000 |
| src/covermanager/coverfetchingactions.cpp            | 0.000000 |
| src/covermanager/coverfetchqueue.cpp                 | 0.000000 |
| src/covermanager/coverfetchunit.cpp                  | 0.000000 |
| src/covermanager/covermanager.cpp                    | 0.000000 |
| src/databaseimporter/itunes/itunesimporterworker.cpp | 0.000000 |
| src/dbus/dbusqueryhelper.cpp                         | 0.000000 |
| src/dialogs/collectionsetup.cpp                      | 0.000000 |
| src/dialogs/databaseimporterdialog.cpp               | 0.000000 |
| src/dialogs/deletedialog.cpp                         | 0.000000 |
| src/dialogs/editfilterdialog.cpp                     | 0.000000 |
| src/dialogs/organizecollectiondialog.cpp             | 0.000000 |
| src/dialogs/transferdialog.cpp                       | 0.000000 |
| src/dynamic/dynamicmodel.cpp                         | 0.000000 |
| src/mac/growlinterface.cpp                           | 0.000000 |
| src/main.cpp   | 0.000000 |
| src/musicbrainz/musicbrainztagsitem.cpp              | 0.000000 |
| src/musicbrainz/musicdnssaudiodecoder.cpp            | 0.000000 |



|   |          |
|---|----------|
| src/musicbrainz/musicdnsfinder.cpp                          | 0.000000 |
| src/opmlwriter.cpp  | 0.000000 |
| src/playback/fadeouter.cpp                                  | 0.000000 |
| src/playlist/layouts/layoutitemconfig.cpp                   | 0.000000 |
| src/playlist/playlistmodel.cpp                              | 0.000000 |
| src/playlist/playlisttoolbar.cpp                            | 0.000000 |
| src/playlist/playlistviewurlrunner.cpp                      | 0.000000 |
| src/playlist/view/listview/inlineditorwidget.cpp            | 0.000000 |
| src/playlistmanager/sql/sqlplaylist.cpp                     | 0.000000 |
| src/scanner/abstractscanresultprocessor.cpp                 | 0.000000 |
| src/scanner/genericscannerjob.cpp                           | 0.000000 |
| src/services/jamendo/jamendoservice.cpp                     | 0.000000 |
| src/services/lastfm/lastfmtreeview.cpp                      | 0.000000 |
| src/services/lastfm/lovetrackaction.cpp                     | 0.000000 |
| src/services/lastfm/meta/lastfmmultiplayablecapability.cpp  | 0.000000 |
| src/services/lastfm/scrobbleradapter.cpp                    | 0.000000 |
| src/services/lastfm/synchronizationadapter.cpp              | 0.000000 |
| src/services/mp3tunes/harmonydaemon/main.cpp                | 0.000000 |
| src/services/mp3tunes/libmp3tunes/locker.c                  | 0.000000 |
| src/services/mp3tunes/mp3tunesservice.cpp                   | 0.000000 |
| src/services/mp3tunes/mp3tunesworkers.cpp                   | 0.000000 |
| src/services/opmldirectory/opmldirectorydatabasehandler.cpp | 0.000000 |

|  |          |
|--|----------|
| src/services/servicesqlquerymaker.cpp                            | 0.000000 |
| src/services/servicesqlregistry.cpp                              | 0.000000 |
| src/statsyncing/jobs/matchtracksjob.cpp                          | 0.000000 |
| src/statusbar/compoundprogressbar.cpp                            | 0.000000 |
| src/statusbar/kjobprogressbar.cpp                                | 0.000000 |
| src/statusbar/progressbar.cpp                                    | 0.000000 |
| src/synchronization/synchronizationbasejob.cpp                   | 0.000000 |
| src/transcoding/transcodingpropertysliderwidget.cpp              | 0.000000 |
| src/widgets/horizontaldivider.cpp                                | 0.000000 |
| src/widgets/iconbutton.cpp                                       | 0.000000 |
| src/widgets/progresswidget.cpp                                   | 0.000000 |
| src/widgets/trackactionbutton.cpp                                | 0.000000 |
| src/widgets/trackselectwidget.cpp                                | 0.000000 |
| tests/core-impl/collections/db/sql/testdatabaseupdater.cpp       | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlalbum.cpp              | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlcollectionlocation.cpp | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlquerymaker.cpp         | 0.000000 |
| tests/core-impl/collections/db/sql/testsqltrack.cpp              | 0.000000 |
| tests/core-impl/meta/file/testmetafiletrack.cpp                  | 0.000000 |
| tests/core-impl/meta/multi/testmetamultitrack.cpp                | 0.000000 |
| tests/core-impl/playlists/types/file/asx/testasxplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/m3u/testm3uplaylist.cpp     | 0.000000 |

|  |          |
|--|----------|
| tests/core-impl/playlists/types/file/pls/testplsplaylist.cpp   | 0.000000 |
| tests/core-impl/playlists/types/file/xspf/testxspfplaylist.cpp | 0.000000 |
| tests/core/collections/collectionlocationtest.cpp              | 0.000000 |
| tests/core/collections/support/testtrackforurlworker.cpp       | 0.000000 |
| tests/core/interfaces/testmetacapability.cpp                   | 0.000000 |
| tests/core/meta/support/testmetaalbumkey.cpp                   | 0.000000 |
| tests/core/meta/support/testmetatrackkey.cpp                   | 0.000000 |
| tests/core/meta/support/testprivatemetaregistry.cpp            | 0.000000 |
| tests/core/meta/testmetatrack.cpp                              | 0.000000 |
| tests/core/playlists/testplaylistobserver.cpp                  | 0.000000 |
| tests/mocks/mocktrackforurlworker.cpp                          | 0.000000 |
| tests/qt-modeltest/modeltest.cpp                               | 0.000000 |
| tests/synchronization/testmasterslavesynchronizationjob.cpp    | 0.000000 |
| tests/testsmartpointerlist.cpp                                 | 0.000000 |
| tests/testtrackorganizer.cpp                                   | 0.000000 |
| utilities/afttagger/afttagger.cpp                              | 0.000000 |

ΠΙΝΑΚΑΣ 6.10: Αποτελέσματα File Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 323635

Στον πίνακα 6.11 παρουσιάζονται αποτελέσματα για την περίπτωση του συστήματος korpete και την αναφορά υπ' αριθμόν 304105. Το σύστημα δεν συγκεντρώνει ικανοποιητικό πλήθος αρχείων ώστε να αξιοποιηθεί ο αλγόριθμος ταξινόμησης των αποτελεσμάτων. Κατά συνέπεια παρουσιάζονται

τα αρχεία ομαδοποιημένα σύμφωνα με τους φακέλους στους οποίους ανήκουν. Αυτή η πρακτική μπορεί να βοηθήσει πολύ ένα προγραμματιστή στην προσπάθειά του να εντοπίσει γρήγορα και αποτελεσματικά την παρουσία ή την απουσία κάποιου πιθανώς χρήσιμου αρχείου. Όπως φαίνεται αυτό είναι αρκετά εύκολο να συμβεί λόγω του τρόπου με τον οποίο έχει επιλεγθεί να γίνει η παρουσίαση, καθώς το πρόβλημα εν προκειμένω συγκεντρώνεται σε μία ομάδα προβληματικών αρχείων που βρίσκονται όλα στον ίδιο υποφάκελο.

| <b>File Name</b>                                  | <b>Connectivity to Clique</b> |
|---|-------------------------------|
| protocols/yahoo/yahooverifyaccount.cpp            | 1                             |
| protocols/yahoo/yahoocontact.cpp                  | 1                             |
| protocols/yahoo/yahooconferencemessagemanager.cpp | 1                             |
| protocols/yahoo/yahooaccount.cpp                  | 0                             |
| protocols/yahoo/ui/yahoouserinfoDialog.cpp        | 2                             |
| protocols/yahoo/ui/yahooinvitelistimpl.cpp        | 2                             |
| protocols/wlm/wlmtransfermanager.cpp              | 2                             |
| protocols/wlm/wlmsocket.cpp                       | 1                             |
| protocols/wlm/wlmprotocol.cpp                     | 6                             |
| protocols/wlm/wlmcontact.cpp                      | 4                             |
| protocols/wlm/wlmchatsession.cpp                  | 4                             |
| protocols/wlm/wlmchatmanager.cpp                  | 4                             |
| protocols/wlm/wlmaccount.cpp                      | 10                            |
| protocols/wlm/ui/wlmeditaccountwidget.cpp         | 1                             |
| protocols/oscar/oscarcontact.cpp                  | 10                            |
| protocols/oscar/oscaraccount.cpp                  | 14                            |
| protocols/oscar/liboscar/tests/buffertest.cpp     | 1                             |

|   |    |
|---|----|
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp                    | 5  |
| protocols/oscar/liboscar/tasks/profiletask.cpp                      | 2  |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp              | 4  |
| protocols/oscar/liboscar/tasks/logintask.cpp                        | 2  |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp            | 3  |
| protocols/oscar/liboscar/tasks/icqtask.cpp                          | 1  |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp                 | 2  |
| protocols/oscar/liboscar/flapprotocol.cpp                           | 1  |
| protocols/oscar/liboscar/connection.cpp                             | 1  |
| protocols/oscar/liboscar/buffer.cpp                                 | 50 |
| protocols/oscar/icq/ui/icquserinfowidget.cpp                        | 1  |
| protocols/oscar/icq/ui/icqsearchdialog.cpp                          | 3  |
| protocols/jabber/ui/dlgjabberchatroomslist.cpp                      | 0  |
| protocols/jabber/ui/dlgjabberchatjoin.cpp                           | 0  |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc | 1  |
| protocols/jabber/libjingle/talk/session/phone/mediasessionclient.cc | 5  |
| protocols/jabber/libjingle/talk/session/phone/mediasession.cc       | 3  |
| protocols/jabber/libjingle/talk/session/phone/channel.cc            | 4  |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc         | 2  |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc     | 1  |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc   | 5  |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc    | 6  |

|  |   |
|--|---|
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc             | 1 |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                      | 3 |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc            | 2 |
| protocols/jabber/libjingle/talk/p2p/base/portallocatorproxy.cc             | 2 |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc                  | 8 |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                           | 4 |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc            | 4 |
| protocols/jabber/libjingle/talk/examples/peerconnection/client/main_wnd.cc | 2 |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                        | 0 |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc             | 2 |
| protocols/jabber/libjingle/talk/base/taskrunner.cc                         | 1 |
| protocols/jabber/libjingle/talk/base/taskparent.cc                         | 2 |
| protocols/jabber/libjingle/talk/base/flags.cc                              | 2 |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc                    | 1 |
| protocols/jabber/libjingle/talk/app/webrtc/webrtcSDP.cc                    | 4 |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp                   | 5 |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp                | 2 |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                          | 0 |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp             | 3 |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp                   | 0 |
| protocols/jabber/libiris/src/irisnet/appledns/appledns.cpp                 | 0 |
| protocols/irc/ui/ircdialogs.cpp  | 2 |

|  |          |
|--|----------|
| <b>plugins/otr/privkeypopup.cpp</b>                | <b>2</b> |
| <b>plugins/otr/otrpreferences.cpp</b>              | <b>0</b> |
| <b>plugins/otr/otrplugin.cpp</b>                   | <b>0</b> |
| <b>plugins/otr/otrlconfinterface.cpp</b>           | <b>0</b> |
| <b>plugins/otr/otrlchatinterface.cpp</b>           | <b>5</b> |
| <b>plugins/otr/otrquiclient.cpp</b>                | <b>0</b> |
| <b>plugins/otr/authenticationwizard.cpp</b>        | <b>2</b> |
| libkopete/ui/kopetelistviewsearchline.cpp          | 2        |
| libkopete/private/kopeteviewmanager.cpp            | 0        |
| libkopete/kopetetransfermanager.cpp                | 0        |
| libkopete/kopeteonlinestatus.cpp                   | 18       |
| libkopete/kopetemetaccontact.cpp                   | 0        |
| kopete/identity/identitystatuswidget.cpp           | 0        |
| kopete/config/avdevice/avdeviceconfig.cpp          | 1        |
| kopete/chatwindow/kopetechatwindowstylemanager.cpp | 4        |
| kopete/chatwindow/kopetechatwindow.cpp             | 3        |
| kopete/chatwindow/chatview.cpp                     | 2        |
| kopete/chatwindow/chatmessagepart.cpp              | 1        |

ΠΙΝΑΚΑΣ 6.11: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 304105

Τέλος ο πίνακας 6.12 παρουσιάζει μία περίπτωση κακών αποτελεσμάτων για το kopete και συγκεκριμένα για την αναφορά 273070 φαίνεται η αδυναμία της μεθόδου να εντοπίσει κάποιο σωστό αρχείο το οποίο όμως λόγω του μεγέθους του αποτελέσματος δεν προκαλεί σημαντική επιπλέον καθυστέρηση

στην προσπάθεια του προγραμματιστή να εντοπίσει το σφάλμα.

| File Name   | Connectivity To Clique |
|---|------------------------|
| protocols/oscar/liboscar/tests/buffertest.cpp                         | 1                      |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp                      | 5                      |
| protocols/oscar/liboscar/tasks/profiletask.cpp                        | 2                      |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp                | 4                      |
| protocols/oscar/liboscar/tasks/logintask.cpp                          | 2                      |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp              | 3                      |
| protocols/oscar/liboscar/tasks/icqtask.cpp                            | 1                      |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp                   | 2                      |
| protocols/oscar/liboscar/flapprotocol.cpp                             | 1                      |
| protocols/oscar/liboscar/connection.cpp                               | 1                      |
| protocols/oscar/liboscar/buffer.cpp                                   | 50                     |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc   | 1                      |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc           | 2                      |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc       | 1                      |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc     | 5                      |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc      | 6                      |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc        | 1                      |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                 | 3                      |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc       | 2                      |
| protocols/jabber/libjingle/talk/p2p/base/portallocatorsessionproxy.cc | 2                      |



|   |   |
|---|---|
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc       | 8 |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                | 4 |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc | 4 |
| protocols/jabber/libjingle/talk/base/winfirewall.cc             | 0 |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc  | 2 |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc         | 1 |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp        | 5 |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp     | 2 |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c               | 0 |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp  | 3 |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp        | 0 |
| plugins/otr/privkeypopup.cpp                                    | 2 |
| plugins/otr/otrpreferences.cpp                                  | 0 |
| plugins/otr/otrplugin.cpp                                       | 0 |
| plugins/otr/otrconfinterface.cpp                                | 0 |
| plugins/otr/otrchatinterface.cpp                                | 5 |
| plugins/otr/otrquiclient.cpp                                    | 0 |
| plugins/otr/authenticationwizard.cpp                            | 2 |
| libkopete/ui/kopetelistviewsearchline.cpp                       | 2 |
| kopete/config/avdevice/avdeviceconfig.cpp                       | 1 |

ΠΙΝΑΚΑΣ 6.12: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 273070



## Κεφάλαιο 7

# Συμπεράσματα και Μελλοντικές Ερευνητικές Κατευθύνσεις

### 7.1 Συμπεράσματα

Μέσα από τις τεχνικές, τους αλγορίθμους και τα δεδομένα τα οποία αξιοποιήθηκαν για τους σκοπούς αυτής της διπλωματικής εργασίας, αποδείχθηκε ότι είναι εφικτό χρησιμοποιώντας μία μηχανιστική, δηλαδή αυτόματη, διαδικασία να προσδιοριστεί η γενεσιουργός αιτία μίας αναφοράς σφάλματος μέσα στον κώδικα του συστήματος για το οποίο αυτή έχει δημιουργηθεί.

Συγκεκριμένα αποδείχθηκε μέσω της προτεινόμενης διαδικασίας ότι οι αναφορές σφάλματος περιέχουν πληροφορία η οποία σχετίζεται με οντότητες του προγράμματος οι οποίες έχουν μη μηδενική και μη τυχαία συσχέτιση με τα προβληματικά σημεία του κώδικα. Αυτό το συμπέρασμα δύναται να βοηθήσει εξαιρετικά, όταν προχωρήσει η έρευνα πάνω σε αυτό το θέμα, τη βιομηχανία παραγωγής και συντήρησης συστημάτων λογισμικού η οποία ξοδεύει όπως αναφέρθηκε και στην εισαγωγή υπέρογκα ποσά για την αποσφαλμάτωση προβληματικών συστημάτων.

Δυστυχώς δεν κατέστη εφικτός ο εντοπισμός επακριβώς των σχέσεων αυτών που μπορούν να βοηθήσουν ώστε να μπορεί ταυτοχρόνως και να ανακαλύπτεται περιέχον αρχείο του σφάλματος, και να περιορίζεται ο θόρυβος που συλλέγεται. Αυτό ωστόσο αποτελεί πεδίο μελλοντικής εργασίας και όπως ειπώθηκε και στο 6 ήδη υπάρχουν δείκτες ότι συγκεκριμένες σχέσεις μπορούν με ασφάλεια να αποκλειστούν από το σύνολο των διαθέσιμων σχέσεων.

## 7.2 Σχετικές Εργασίες

Η εργασία αυτή έρχεται να προστεθεί στο σύνολο των εργασιών που έχουν καταπιαστεί με το θέμα του Bug Localization. Μέχρι τη στιγμή της συγγραφής αυτής της εργασίας έχουν δημοσιευθεί αρκετές εργασίες στις οποίες παρουσιάζονται μέθοδοι που συνεισφέρουν στον τομέα και μάλιστα κάποιες έχουν επιτύχει, με άλλες μεθόδους, εξαιρετικά αποτελέσματα.

Στη σχετική βιβλιογραφία στην οποία έρχεται να προστεθεί αυτό το κείμενο είναι η πρώτη φορά που γίνεται μία καθαρά software engineering προσέγγιση η οποία αποπειράται να εφαρμόσει μία προσέγγιση που εκμεταλλεύεται την δομή του εκάστοτε συστήματος και δεν εξαρτάται από περιστασιακές πληροφορίες οι οποίες υπάρχουν μέσα σε αναφορά, ούτε εξαρτάται από τη σωστή 'ονοματοδοσία' που μπορεί να έχει ή να μην έχει εφαρμοστεί κατά τη συγγραφή μίας αναφοράς σφάλματος ή του κώδικα του συστήματος.

Άλλες εργασίες οι οποίες έχουν ασχοληθεί με το ίδιο πρόβλημα είναι η εργασία των Andrian Marcus, Andrey Sergeev, Vaclav Rajlich και Jonathan I. Maletic [10] του 2004 η οποία παρουσιάζει μία προσέγγιση η οποία είναι αποκλειστικά βασισμένη στην τεχνική LSI (Latent Semantic Indexing), η οποία είναι μία άλλη ονομασία για την μέθοδο που στο παρόν σύγγραμμα αναφέρεται σαν LSA. Σε αυτήν την εργασία τόσο η αναφορά σφάλματος όσο και τα αρχεία κώδικα αντιμετωπίζονται σαν κείμενα. Ο κώδικας μάλιστα διασπάται σε μικρότερα τμήματα, μεθόδους, κλάσεις, μπλοκς και μετά πάνω σε αυτά τα 'παράγωγα' κείμενα εφαρμόζεται η τεχνική LSI με corpus το σύνολο του κώδικα, και ερώτηση την αναφορά σφάλματος. Η εργασία έχει επιτύχει ικανοποιητικά αποτελέσματα αλλά δυστυχώς σε ένα σύστημα που είναι εξαιρετικά πεπαιωμένο για την εφαρμογή αυτής της μεθόδου.

Στο θέμα έχει δημοσιευθεί επίσης η εργασία των Anh Tuan Nguyen, Tung Thanh et al. [4] οι οποίοι εφαρμόζουν μία τεχνική πολύ παρεμφερή με την LSA την τεχνική LDA η οποία έχει αναφερθεί στο 2. Ωστόσο δεν μένουν μόνο στην εφαρμογή της LDA αλλά προχωρούν στη χρήση επιπλέον πληροφορίας για τα αρχεία, κατατάσσοντάς τα με βάση την 'σφαλερότητα' Buginess των αρχείων αυτών με βάση στατιστικά στοιχεία σχετικά με το πόσο συχνά τροποποιούνται τα αρχεία κατά την εφαρμογή κάποιας επιδιόρθωσης.

Τέλος μία άλλη ενδιαφέρουσα δουλειά είναι αυτή των Shivani Rao και Avinash Kak [9] οι οποίοι πραγματοποιούν συγκριτική μελέτη πάνω σε ένα σύνολο από information retrieval αλγόριθμους και συνδυασμούς αυτών.

Παρατηρείται λοιπόν ότι η ερευνητική κοινότητα κινείται έντονα προς την χρήση information retrieval αλγορίθμων και μεθόδων για την αντιμετώπιση αυτού του προβλήματος.

### 7.3 Συνεισφορά

Η συνεισφορά αυτής της εργασίας είναι η εισαγωγή ανάμεσα σε πολλές εργασίες που έχουν ασχοληθεί με το θέμα μία προσέγγισης η οποία διαφοροποιείται δραστικά από την προσέγγιση την κοινότητας. Η προσέγγιση αντιλαμβάνεται το πρόβλημα του bug localization ως ένα καθαρά software engineering πρόβλημα στο οποίο μπορούν να εφαρμοστούν μέθοδοι και τεχνικές που χρησιμοποιούνται επί δεκαετίες στον τομέα και έχουν επανειλημμένα αποδειχτεί σωστές. Ως εκ τούτου πλέον είναι εφικτό να οι ερευνητές της τεχνολογίας λογισμικού να πλησιάσουν ακόμα περισσότερο στη σωστή λύση ίσως απλά εφαρμόζοντας τις information retrieval προσεγγίσεις τους στο μειωμένο χώρο που προκύπτει από την μέθοδο αυτής της εργασίας.

### 7.4 Απειλές ως προς την Ορθότητα

Η εργασία αυτή μπορεί να ειπωθεί ότι πάσχει ως προς το πλήθος των αναφορών που χρησιμοποιήθηκαν για την αξιολόγηση της τεχνικής. Ωστόσο αυτό είναι ένα εύκολα επιλύσιμο πρόβλημα καθώς υπάρχει διαθέσιμη όλη η λειτουργικότητα ώστε και το πλήθος των αναφορών ανά σύστημα να αυξηθεί αλλά να εισαχθούν και άλλα συστήματα και από άλλες συλλογές.

Επίσης είναι σημαντικό να αναφερθεί ότι δεν χρησιμοποιήθηκε κανένα από τα συνήθως χρησιμοποιούμενα συστήματα για την αξιολόγηση αλγορίθμων και τεχνικών επίλυσης του bug localization προβλήματος. Χαρακτηριστικά δεν χρησιμοποιήθηκε το σύστημα AspectJ, γραμμένο στη γλώσσα Java το οποίο στη βιβλιογραφία χρησιμοποιείται εξαιρετικά συχνά.

Τέλος η τελευταία και ίσως σημαντικότερη απειλή προς την ορθότητα αυτής της εργασίας είναι το γεγονός ότι όλα τα συστήματα, τόσο αυτά που παρουσιάστηκαν όσο και αυτά τα οποία χρησιμοποιήθηκαν κατά τον σχεδιασμό, την υλοποίηση και τον προκαταρκτικό έλεγχο του συστήματος είναι όλα εκτός από ένα γραμμένα στη γλώσσα προγραμματισμού C/C++ το οποίο αφήνει πολύ μεγάλα περιθώρια αμφισβήτησης των αποτελεσμάτων και της μεθόδου καθεαυτής.

### 7.5 Κατευθύνσεις για Μελλοντική Έρευνα

Με την ολοκλήρωση αυτής της εργασίας ολοκληρώθηκε μία ενδιαφέρουσα περιπλάνηση στον χώρο του προβλήματος του bug localization. Μετά από αυτή τη μικρή οδύσσεια, υπάρχει πλέον πολύ συσσωρευμένη γνώση η οποία θα είναι πολύ χρήσιμο να αξιοποιηθεί προς την κατεύθυνση της δημιουργίας μία μεθόδου η οποία θα συνδυάζει τα καλύτερα αποτελέσματα

τόσο της μεθόδου που παρουσιάστηκε εδώ όσο και των μεθόδων που ήδη έχουν δημοσιευτεί με πυρήνες από το πεδίο του information retrieval.

Μία πολύ ενδιαφέρουσα ιδέα που αξίζει περαιτέρω διερεύνησης είναι η δημιουργία μία μετρικής για την ποιότητα των αναφορών σφάλματος, με βάση τους λέξεις που χρησιμοποιούνται, τους όρους προγράμματος που χρησιμοποιούνται, την ύπαρξη ή απουσία αναλυτικής μηνυμάτων σφάλματος από το πρόγραμμα, ακόμα και η σωστή χρήση της φυσικής γλώσσας.

Επίσης μία άλλη ενδιαφέρουσα κατεύθυνση προς την οποία έχει ήδη αρχίσει να δουλεύει το εργαστήριο είναι η υλοποίηση ενός πιο τυπικού αλγόριθμου για την ομαδοποίηση και κατάταξη των αποτελεσμάτων αυτής της μεθόδου.

## Παράρτημα Α΄

# Amarok Sample Results

109

Στους Επόμενους Πίνακες Παρουσιάζονται τα αποτελέσματα για το σύστημα Amarok.

| ClusterName  | Score    | expectation | Connections | ToClique  | ClusterSize |
|--|----------|-------------|-------------|-----------|-------------|
| src/services/magnatune/magnatunestore.cpp                    | 4.000000 | 0.039344    | 16.000000   | 12.000000 | 3           |
| src/context/popupdropper/libpud/popupdropperview.cpp         | 4.000000 | 0.013115    | 4.000000    | 1.000000  | 1           |
| src/playlist/proxymodels/sortalgorithms.cpp                  | 4.000000 | 0.026230    | 4.000000    | 2.000000  | 2           |
| src/services/magnatune/magnatuneredownloadhandler.cpp        | 4.000000 | 0.026230    | 4.000000    | 2.000000  | 2           |
| tests/core-impl/collections/support/testmemoryquerymaker.cpp | 4.000000 | 0.026230    | 4.000000    | 2.000000  | 2           |
| tests/services/amazon/testamazonparser.cpp                   | 4.000000 | 0.026230    | 4.000000    | 2.000000  | 2           |
| src/mediadevicemonitor.cpp                                   | 4.000000 | 0.026230    | 4.000000    | 2.000000  | 2           |
| src/core-impl/collections/db/sql/sqlcollection.cpp           | 4.333333 | 0.042623    | 13.000000   | 9.000000  | 3           |
| src/app.cpp  | 4.615385 | 0.060530    | 15.000000   | 13.000000 | 4           |

|  |           |          |           |           |    |
|--|-----------|----------|-----------|-----------|----|
| src/dialogs/tagguesserdialog.cpp                                   | 5.000000  | 0.049180 | 10.000000 | 6.000000  | 3  |
| src/services/mp3tunes/mp3tuneslocker.cpp                           | 5.000000  | 0.032787 | 5.000000  | 2.000000  | 2  |
| src/aboutdialog/extendedaboutdialog.cpp                            | 5.400000  | 0.053115 | 9.000000  | 5.000000  | 3  |
| src/context/applets/albums/albums.cpp                              | 5.846154  | 0.076671 | 19.000000 | 13.000000 | 4  |
| src/services/jamendo/jamendoxmlparser.cpp                          | 6.000000  | 0.039344 | 3.000000  | 1.000000  | 2  |
| src/services/scriptable/scriptableservice.cpp                      | 6.000000  | 0.039344 | 6.000000  | 2.000000  | 2  |
| src/svghandler.cpp   | 6.000000  | 0.059016 | 8.000000  | 4.000000  | 3  |
| src/services/amazon/amazonstore.cpp                                | 7.428571  | 0.097424 | 13.000000 | 7.000000  | 4  |
| src/scriptmanager.cpp  | 7.500000  | 0.073770 | 15.000000 | 6.000000  | 3  |
| src/services/lastfm/lastfmservice.cpp                              | 9.230769  | 0.181589 | 40.000000 | 26.000000 | 6  |
| src/covermanager/covermanager.cpp                                  | 10.062500 | 0.230943 | 46.000000 | 32.000000 | 7  |
| src/amarokurls/amarokurlhandler.cpp                                | 13.500000 | 0.398361 | 96.000000 | 64.000000 | 9  |
| src/browsers/browserbreadcrumbwidget.cpp                           | 14.285714 | 0.281030 | 50.000000 | 21.000000 | 6  |
| src/mainwindow.cpp   | 20.385965 | 0.935749 | 83.000000 | 57.000000 | 14 |
| src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp  | Infinity  | Infinity | 1.000000  | 0.000000  | 1  |
| src/services/amazon/amazonshoppingcartview.h                       | NaN       | NaN      | 0.000000  | 0.000000  | 1  |
| src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp | NaN       | NaN      | 0.000000  | 0.000000  | 1  |

ΠΙΝΑΚΑΣ Α'.1: Αποτελέσματα First Cluster Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 322474

| File Name   | Connectivity To Clique |
|---|------------------------|
| <b>Cluster Name : src/playlist/proxymodels/sortalgorithms.cpp</b> |                        |
| src/playlist/proxymodels/sortalgorithms.cpp                       | 1.000000               |
| src/playlist/proxymodels/sortscheme.cpp                           | 1.000000               |
| <b>Cluster Name : src/covermanager/covermanager.cpp</b>           |                        |



|  |           |
|--|-----------|
| src/context/applets/albums/albumitem.cpp                                 | 1.000000  |
| src/configdialog/dialogs/excludedlabelsdialog.cpp                        | 2.000000  |
| src/covermanager/coverfetchunit.cpp                                      | 2.000000  |
| src/statusbar/compoundprogressbar.cpp                                    | 3.000000  |
| src/covermanager/coverfounddialog.cpp                                    | 5.000000  |
| src/covermanager/covermanager.cpp  | 11.000000 |
| <b>Cluster Name : src/services/amazon/amazonshoppingcartview.h</b>       |           |
| <b>Cluster Name : src/context/applets/albums/albums.cpp</b>              |           |
| src/context/applets/albums/trackitem.cpp                                 | 1.000000  |
| src/context/applets/albums/albumsmodel.cpp                               | 2.000000  |
| src/context/applets/albums/albums.cpp                                    | 3.000000  |
| src/context/applets/albums/albumsview.cpp                                | 7.000000  |
| <b>Cluster Name : src/services/lastfm/lastfmservice.cpp</b>              |           |
| src/scriptengine/amaroknetworkscript.cpp                                 | 2.000000  |
| src/services/lastfm/lastfmservicesettings.cpp                            | 2.000000  |
| src/network/networkaccessmanagerproxy.cpp                                | 3.000000  |
| src/services/lastfm/lastfmtreemodel.cpp                                  | 4.000000  |
| src/services/lastfm/lastfmservice.cpp                                    | 5.000000  |
| <b>Cluster Name : src/core-impl/collections/db/sql/sqlcollection.cpp</b> |           |
| src/core-impl/collections/db/mountpointmanager.cpp                       | 3.000000  |
| src/widgets/tokenwithlayout.cpp  | 6.000000  |
| <b>Cluster Name : src/aboutdialog/extendedaboutdialog.cpp</b>            |           |
| src/aboutdialog/extendedaboutdialog.cpp                                  | 1.000000  |
| src/aboutdialog/animatedwidget.cpp                                       | 1.000000  |

|   |          |
|---|----------|
| src/aboutdialog/ocspersonitem.cpp   | 3.000000 |
| <b>Cluster Name : src/services/magnatune/magnatunestore.cpp</b>             |          |
| src/browsers/infoproxy.cpp  | 1.000000 |
| src/services/magnatune/magnatunestore.cpp                                   | 6.000000 |
| <b>Cluster Name : src/services/amazon/amazonstore.cpp</b>                   |          |
| src/services/amazon/amazoninfoparser.cpp                                    | 1.000000 |
| src/services/amazon/amazonitemtreemodel.cpp                                 | 1.000000 |
| src/services/amazon/amazonstore.cpp   | 3.000000 |
| <b>Cluster Name : src/services/jamendo/jamendoxmlparser.cpp</b>             |          |
| src/services/jamendo/jamendodatabasehandler.cpp                             | 1.000000 |
| <b>Cluster Name : src/svghandler.cpp</b>                                    |          |
| src/svghandler.cpp  | 2.000000 |
| src/moodbar/moodbarmanager.cpp  | 2.000000 |
| <b>Cluster Name : src/browsers/browserbreadcrumbwidget.cpp</b>              |          |
| src/browsers/filebrowser/filebrowser.cpp                                    | 1.000000 |
| src/browsers/browserbreadcrumbitem.cpp                                      | 1.000000 |
| src/amarokurls/navigationurlrunner.cpp                                      | 2.000000 |
| src/browsers/browserbreadcrumbwidget.cpp                                    | 4.000000 |
| src/browsers/browsercategory.cpp  | 6.000000 |
| src/browsers/browsercategorylist.cpp  | 7.000000 |
| <b>Cluster Name : src/services/scriptable/scriptableservice.cpp</b>         |          |
| src/services/scriptable/scriptableservice.cpp                               | 1.000000 |
| src/services/scriptable/scriptableservicemeta.cpp                           | 1.000000 |
| <b>Cluster Name : src/services/magnatune/magnatuneredownloadhandler.cpp</b> |          |

|  |          |
|--|----------|
| src/services/magnatune/magnatuneredownloadhandler.cpp                              | 2.000000 |
| <b>Cluster Name : src/mainwindow.cpp</b>   |          |
| src/browsers/browserdock.cpp   | 1.000000 |
| src/network/networkaccessviewer.cpp  | 1.000000 |
| src/palettehandler.cpp   | 1.000000 |
| src/main.cpp   | 2.000000 |
| src/browsers/servicebrowser/servicebrowser.cpp                                     | 2.000000 |
| src/amarokurls/bookmarkmanager.cpp   | 3.000000 |
| src/browsers/collectiontreeitemmodelbase.cpp                                       | 3.000000 |
| src/browsers/collectiontreeitem.cpp  | 4.000000 |
| src/dialogs/equalizerdialog.cpp  | 4.000000 |
| src/browsers/collectionbrowser/collectionwidget.cpp                                | 7.000000 |
| src/mainwindow.cpp   | 9.000000 |
| <b>Cluster Name : tests/core-impl/collections/support/testmemoryquerymaker.cpp</b> |          |
| tests/core-impl/collections/support/testmemoryquerymaker.cpp                       | 2.000000 |
| <b>Cluster Name : src/dialogs/tagguesserdialog.cpp</b>                             |          |
| src/dialogs/tagguesserdialog.cpp   | 3.000000 |
| src/widgets/token.cpp  | 3.000000 |
| <b>Cluster Name : src/app.cpp</b>  |          |
| src/covermanager/covercache.cpp  | 2.000000 |
| src/app.cpp  | 6.000000 |
| <b>Cluster Name : tests/services/amazon/testamazonparser.cpp</b>                   |          |
| src/services/amazon/amazonparser.cpp   | 1.000000 |
| tests/services/amazon/testamazonparser.cpp   | 1.000000 |

|  |          |
|--|----------|
| <b>Cluster Name : src/core-impl/collections/nepomukcollection/nepomukquermaker.cpp</b>   |          |
| <b>Cluster Name : src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp</b> |          |
| <b>Cluster Name : src/mediadevicemonitor.cpp</b>   |          |
| src/mediadevicecache.cpp   | 2.000000 |
| <b>Cluster Name : src/services/mp3tunes/mp3tuneslocker.cpp</b>                           |          |
| src/services/mp3tunes/mp3tuneslocker.cpp   | 2.000000 |
| <b>Cluster Name : src/scriptmanager.cpp</b>  |          |
| src/scriptmanager.cpp  | 6.000000 |
| <b>Cluster Name : src/amarokurls/amarokurlhandler.cpp</b>                                |          |
| src/services/servicecapabilities.cpp   | 1.000000 |
| src/amarokurls/amarokurlhandler.cpp  | 2.000000 |
| src/amarokurls/amarokurl.cpp   | 2.000000 |
| src/context/applets/upcomingevents/upcomingeventsstackitem.cpp                           | 4.000000 |
| src/amarokurls/navigationurlgenerator.cpp  | 4.000000 |
| src/amarokurls/bookmarkmodel.cpp   | 5.000000 |
| <b>Cluster Name : src/context/popupdropper/libpud/popupdropperview.cpp</b>               |          |
| src/context/popupdropper/libpud/popupdropperview.cpp                                     | 1.000000 |

ΠΙΝΑΚΑΣ Α'.2: Αποτελέσματα First File Level για το Amarok για την Αναφορά Σφάλματος υπέρριθμόν 322474

| File Name   | Connectivity To Clique |
|---|------------------------|
| <b>Cluster Name : src/playlist/proxymodels/sortalgorithms.cpp</b> |                        |
|   |                        |
| <b>Cluster Name : src/covermanager/covermanager.cpp</b>           |                        |

|  |          |
|--|----------|
| src/statusbar/progressbar.cpp  | 0.000000 |
| <b>Cluster Name : src/services/amazon/amazonshoppingcartview.h</b>       |          |
| src/services/amazon/amazonshoppingcartview.cpp                           | 0.000000 |
| <b>Cluster Name : src/context/applets/albums/albums.cpp</b>              |          |
| <b>Cluster Name : src/services/lastfm/lastfmservice.cpp</b>              |          |
| src/services/lastfm/avatardownloader.cpp                                 | 0.000000 |
| <b>Cluster Name : src/core-impl/collections/db/sql/sqlcollection.cpp</b> |          |
| src/core-impl/collections/db/sql/sqlcollection.cpp                       | 0.000000 |
| <b>Cluster Name : src/aboutdialog/extendedaboutdialog.cpp</b>            |          |
| <b>Cluster Name : src/services/magnatune/magnatunestore.cpp</b>          |          |
| src/services/magnatune/magnatunemeta.cpp                                 | 0.000000 |
| <b>Cluster Name : src/services/amazon/amazonstore.cpp</b>                |          |
| src/services/amazon/amazonwantcountrywidget.cpp                          | 0.000000 |
| <b>Cluster Name : src/services/jamendo/jamendoxmlparser.cpp</b>          |          |
| src/services/jamendo/jamendoxmlparser.cpp                                | 0.000000 |
| <b>Cluster Name : src/svghandler.cpp</b>                                 |          |
| src/svgtinter.cpp  | 0.000000 |
| <b>Cluster Name : src/browsers/browserbreadcrumbwidget.cpp</b>           |          |
| <b>Cluster Name : src/services/scriptable/scriptableservice.cpp</b>      |          |

|  |          |
|--|----------|
| <b>Cluster Name : src/services/magnatune/magnatuneredownloadhandler.cpp</b>              |          |
| src/services/magnatune/magnatunedownloaddialog.cpp                                       | 0.000000 |
| <b>Cluster Name : src/mainwindow.cpp</b>   |          |
| src/context/contextdock.cpp  | 0.000000 |
| src/services/scriptable/scriptableservicecollectiontreemodel.cpp                         | 0.000000 |
| src/widgets/amarokdockwidget.cpp   | 0.000000 |
| <b>Cluster Name : tests/core-impl/collections/support/testmemoryquerymaker.cpp</b>       |          |
| src/core-impl/collections/support/memoryfilter.cpp                                       | 0.000000 |
| <b>Cluster Name : src/dialogs/tagguesserdialog.cpp</b>                                   |          |
| src/dialogs/tagguesser.cpp   | 0.000000 |
| <b>Cluster Name : src/app.cpp</b>  |          |
| src/core-impl/collections/support/jobs/writetagsjob.cpp                                  | 0.000000 |
| src/playback/fadeouter.cpp   | 0.000000 |
| <b>Cluster Name : tests/services/amazon/testamazonparser.cpp</b>                         |          |
|  |          |
| <b>Cluster Name : src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp</b>  |          |
| src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp                        | 0.000000 |
| <b>Cluster Name : src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp</b> |          |
| src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp                       | 0.000000 |
| <b>Cluster Name : src/mediadevicemonitor.cpp</b>   |          |
| src/mediadevicemonitor.cpp   | 0.000000 |
| <b>Cluster Name : src/services/mp3tunes/mp3tuneslocker.cpp</b>                           |          |
| src/services/mp3tunes/mp3tuneslockermeta.cpp   | 0.000000 |

| Cluster Name : src/scriptmanager.cpp                                |          |
|---|----------|
| src/services/scriptable/scriptableservicemanager.cpp                | 0.000000 |
| src/scriptengine/amarokscriptableservicescript.cpp                  | 0.000000 |
| Cluster Name : src/amarokurls/amarokurlhandler.cpp                  |          |
| src/amarokurls/bookmarkcurrentbutton.cpp                            | 0.000000 |
| src/amarokurls/contexturlrunner.cpp                                 | 0.000000 |
| src/services/opmldirectory/opmldirectoryservice.cpp                 | 0.000000 |
| Cluster Name : src/context/popupdropper/libpud/popupdropperview.cpp |          |
|   |          |

ΠΙΝΑΚΑΣ Α'.3: Αποτελέσματα Zero File Level για το Amarok για την Αναφορά Σφάλματος υπάριθμόν 322474

| ClusterName                     | Score             | expectation       | Connections       | ToClique          | ClusterSize |
|---------------------------------|-------------------|-------------------|-------------------|-------------------|-------------|
| <i>src/core/support/debug.h</i> | <i>384.872727</i> | <i>211.995469</i> | <i>756.000000</i> | <i>330.000000</i> | <i>168</i>  |

ΠΙΝΑΚΑΣ Α'.4: Αποτελέσματα Big Cluster Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 322474

| File Name   | Connectivity To Clique |
|---|------------------------|
| Cluster Name : src/core/support/debug.h                     |                        |
| src/browsers/collectiontreeview.cpp                         | 4.000000               |
| src/covermanager/coverfetcher.cpp                           | 4.000000               |
| src/amarokurls/bookmarktreeview.cpp                         | 3.000000               |
| src/browsers/collectiontreeitemmodel.cpp                    | 3.000000               |
| src/context/applets/similarartists/artistwidget.cpp         | 3.000000               |
| src/context/applets/upcomingevents/upcomingeventsapplet.cpp | 3.000000               |

|   |          |
|---|----------|
| src/widgets/trackselectwidget.cpp                                   | 3.000000 |
| tests/synchronization/testmasterslavesynchronizationjob.cpp         | 3.000000 |
| tests/synchronization/testunionjob.cpp                              | 3.000000 |
| src/amarokurls/playurlrunner.cpp                                    | 2.000000 |
| src/context/engines/labels/labelsengine.cpp                         | 2.000000 |
| src/services/servicebase.cpp  | 2.000000 |
| tests/core-impl/collections/db/sql/testsqlalbum.cpp                 | 2.000000 |
| src/browsers/collectionsortfilterproxymodel.cpp                     | 1.000000 |
| src/context/applets/labels/labelsapplet.cpp                         | 1.000000 |
| src/context/widgets/droppixmapitem.cpp                              | 1.000000 |
| src/core-impl/collections/db/sql/sqlcapabilities.cpp                | 1.000000 |
| src/core-impl/collections/db/sql/sqlregistry_p.cpp                  | 1.000000 |
| src/core-impl/collections/db/sql/sqlscanresultprocessor.cpp         | 1.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipodcopytracksjob.cpp | 1.000000 |
| src/core-impl/collections/playdarcollection/playdarcollection.cpp   | 1.000000 |
| src/core-impl/collections/support/memorycustomvalue.cpp             | 1.000000 |
| src/core-impl/collections/support/memorymatcher.cpp                 | 1.000000 |
| src/core-impl/meta/multi/multitrack.cpp                             | 1.000000 |
| src/core-impl/support/trackloader.cpp                               | 1.000000 |
| src/core/support/amarok.cpp   | 1.000000 |
| src/covermanager/coverfetchingactions.cpp                           | 1.000000 |
| src/dbus/dbusqueryhelper.cpp  | 1.000000 |
| src/dialogs/databaseimporterdialog.cpp                              | 1.000000 |
| src/dialogs/musicbrainztagger.cpp                                   | 1.000000 |
| src/dialogs/tagdialog.cpp   | 1.000000 |



|  |          |
|--|----------|
| src/dialogs/trackorganizer.cpp                             | 1.000000 |
| src/enginecontroller.cpp                                   | 1.000000 |
| src/mac/growlinterface.cpp                                 | 1.000000 |
| src/musicbrainz/musicbrainzfinder.cpp                      | 1.000000 |
| src/musicbrainz/musicbrainztagsitem.cpp                    | 1.000000 |
| src/musicbrainz/musicdnssaudiodecoder.cpp                  | 1.000000 |
| src/musicbrainz/musicdnsfinder.cpp                         | 1.000000 |
| src/playlist/layouts/layoutitemconfig.cpp                  | 1.000000 |
| src/playlist/playlisttoolbar.cpp                           | 1.000000 |
| src/playlist/proxymodels/proxybase.cpp                     | 1.000000 |
| src/playlistmanager/sql/sqlplaylist.cpp                    | 1.000000 |
| src/playlistmanager/syncedplaylist.cpp                     | 1.000000 |
| src/scanner/abstractscanresultprocessor.cpp                | 1.000000 |
| src/scriptengine/metatypeexporter.cpp                      | 1.000000 |
| src/services/amazon/amazonitemtreeview.cpp                 | 1.000000 |
| src/services/amazon/amazonshoppingcartdialog.cpp           | 1.000000 |
| src/services/lastfm/lovetrackaction.cpp                    | 1.000000 |
| src/services/lastfm/meta/lastfmmultiplayablecapability.cpp | 1.000000 |
| src/services/lastfm/scrobbleradapter.cpp                   | 1.000000 |
| src/services/magnatune/magnatunedatabaseworker.cpp         | 1.000000 |
| src/services/mp3tunes/mp3tunesworkers.cpp                  | 1.000000 |
| src/statsyncing/jobs/matchtracksjob.cpp                    | 1.000000 |
| src/synchronization/synchronizationbasejob.cpp             | 1.000000 |
| src/widgets/filenamelayouwidget.cpp                        | 1.000000 |
| tests/core-impl/collections/db/sql/testsqlscanmanager.cpp  | 1.000000 |

|  |          |
|--|----------|
| tests/core-impl/collections/db/sql/testsqltrack.cpp            | 1.000000 |
| tests/core-impl/meta/file/testmetafiletrack.cpp                | 1.000000 |
| tests/core-impl/support/testtrackloader.cpp                    | 1.000000 |
| tests/core/collections/support/testtrackforurlworker.cpp       | 1.000000 |
| tests/core/meta/support/testmetaalbumkey.cpp                   | 1.000000 |
| tests/core/meta/support/testmetatrackkey.cpp                   | 1.000000 |
| tests/core/meta/support/testprivatemetaregistry.cpp            | 1.000000 |
| tests/core/meta/testmetatrack.cpp                              | 1.000000 |
| tests/testtrackorganizer.cpp                                   | 1.000000 |
| shared/filetype.cpp  | 0.000000 |
| src/aboutdialog/ocspersonlistwidget.cpp                        | 0.000000 |
| src/amarokurls/bookmarkgroup.cpp                               | 0.000000 |
| src/browsers/playlistbrowser/playlistsinfoldersproxy.cpp       | 0.000000 |
| src/configdialog/dialogs/notificationsconfig.cpp               | 0.000000 |
| src/configdialog/dialogs/pluginsconfig.cpp                     | 0.000000 |
| src/configdialog/dialogs/scriptsconfig.cpp                     | 0.000000 |
| src/context/applets/analyzer/analyzerapplet.cpp                | 0.000000 |
| src/context/applets/photos/photoscrollwidget.cpp               | 0.000000 |
| src/context/applets/tabs/tabsview.cpp                          | 0.000000 |
| src/context/applets/upcomingevents/upcomingeventsmapwidget.cpp | 0.000000 |
| src/context/applets/upcomingevents/upcomingeventsstack.cpp     | 0.000000 |
| src/context/applets/wikipedia/wikipediaapplet.cpp              | 0.000000 |
| src/context/contextobserver.cpp                                | 0.000000 |
| src/context/engines/lyrics/lyricsengine.cpp                    | 0.000000 |
| src/context/engines/photos/photosengine.cpp                    | 0.000000 |

|  |                 |
|--|-----------------|
| src/context/engines/tabs/tabseengine.cpp   | 0.000000        |
| src/context/widgets/recentlyplayedlistwidget.cpp                                 | 0.000000        |
| src/context/widgets/textscrollingwidget.cpp                                      | 0.000000        |
| src/core-impl/collections/aggregate/aggregatemetata.cpp                          | 0.000000        |
| src/core-impl/collections/db/sql/device/massstorage/massstoragedevicehandler.cpp | 0.000000        |
| src/core-impl/collections/db/sql/device/nfs/nfsdevicehandler.cpp                 | 0.000000        |
| src/core-impl/collections/db/sql/device/smb/smbdevicehandler.cpp                 | 0.000000        |
| <b>src/core-impl/collections/db/sql/sqlcollectionlocation.cpp</b>                | <b>0.000000</b> |
| src/core-impl/collections/ipodcollection/ipodcollection.cpp                      | 0.000000        |
| src/core-impl/collections/ipodcollection/jobs/ipoddeletetracksjob.cpp            | 0.000000        |
| src/core-impl/collections/ipodcollection/jobs/ipodparsetracksjob.cpp             | 0.000000        |
| src/core-impl/collections/ipodcollection/support/iphonemountpoint.cpp            | 0.000000        |
| src/core-impl/collections/nepomukcollection/nepomukcache.cpp                     | 0.000000        |
| src/core-impl/collections/playdarcollection/playdarquerymaker.cpp                | 0.000000        |
| src/core-impl/collections/support/collectionmanager.cpp                          | 0.000000        |
| src/core-impl/logger/proxylogger.cpp   | 0.000000        |
| src/core-impl/meta/timecode/timecodeobserver.cpp                                 | 0.000000        |
| src/core-impl/support/tagstatisticsstore.cpp                                     | 0.000000        |
| src/core-impl/support/urlstatisticsstore.cpp                                     | 0.000000        |
| src/core/playlists/playlistformat.cpp  | 0.000000        |
| src/core/podcasts/podcastimagefetcher.cpp  | 0.000000        |
| src/core/support/debug.cpp   | 0.000000        |
| src/databaseimporter/amarok14/fastforwardworker.cpp                              | 0.000000        |
| src/databaseimporter/itunes/itunesimporterworker.cpp                             | 0.000000        |
| src/dbus/mpris1/playerhandler.cpp  | 0.000000        |

|  |          |
|--|----------|
| src/dialogs/collectionsetup.cpp                      | 0.000000 |
| src/dialogs/deletedialog.cpp                         | 0.000000 |
| src/dialogs/deviceconfiguredialog.cpp                | 0.000000 |
| src/dialogs/organizecollectiondialog.cpp             | 0.000000 |
| src/dialogs/transferdialog.cpp                       | 0.000000 |
| src/musicbrainz/musicbrainztagsmodeldelegate.cpp     | 0.000000 |
| src/opmlparser.cpp                                   | 0.000000 |
| src/opmlwriter.cpp                                   | 0.000000 |
| src/playlist/playlistmodel.cpp                       | 0.000000 |
| src/playlist/playlistqueueeditor.cpp                 | 0.000000 |
| src/playlist/view/listview/inlineeditorwidget.cpp    | 0.000000 |
| src/playlistmanager/file/kconfigsyncrelstore.cpp     | 0.000000 |
| src/playlistmanager/syncrelationstorage.cpp          | 0.000000 |
| src/scanner/abstractdirectorywatcher.cpp             | 0.000000 |
| src/scanner/genericscannerjob.cpp                    | 0.000000 |
| src/scriptupdater.cpp                                | 0.000000 |
| src/services/ampache/ampacheaccountlogin.cpp         | 0.000000 |
| src/services/infoparserbase.cpp                      | 0.000000 |
| src/services/lastfm/lastfmtreeview.cpp               | 0.000000 |
| src/services/lastfm/synchronizationadapter.cpp       | 0.000000 |
| src/services/magnatune/magnatunealbumdownloader.cpp  | 0.000000 |
| src/services/magnatune/magnatuneneedupdatewidget.cpp | 0.000000 |
| src/services/magnatune/magnatuneredownloaddialog.cpp | 0.000000 |
| src/services/mp3tunes/harmonydaemon/main.cpp         | 0.000000 |
| src/services/mp3tunes/libmp3tunes/harmony.c          | 0.000000 |

|  |          |
|--|----------|
| src/services/mp3tunes/libmp3tunes/locker.c                       | 0.000000 |
| src/services/opmldirectory/opmldirectorydatabasehandler.cpp      | 0.000000 |
| src/services/servicesqlquerymaker.cpp                            | 0.000000 |
| src/services/servicesqlregistry.cpp                              | 0.000000 |
| src/transcoding/transcodingjob.cpp                               | 0.000000 |
| src/transcoding/transcodingpropertysliderwidget.cpp              | 0.000000 |
| src/widgets/editdeletedelegate.cpp                               | 0.000000 |
| src/widgets/horizontaldivider.cpp                                | 0.000000 |
| src/widgets/iconbutton.cpp                                       | 0.000000 |
| src/widgets/osd.cpp  | 0.000000 |
| src/widgets/prettytreedelegate.cpp                               | 0.000000 |
| src/widgets/tokendroptarget.cpp                                  | 0.000000 |
| src/widgets/tokenpool.cpp  | 0.000000 |
| src/widgets/volumedial.cpp                                       | 0.000000 |
| tests/context/engines/testdataengineblackbox.cpp                 | 0.000000 |
| tests/core-impl/collections/db/sql/testdatabaseupdater.cpp       | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlcollectionlocation.cpp | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlquerymaker.cpp         | 0.000000 |
| tests/core-impl/meta/multi/testmetamultitrack.cpp                | 0.000000 |
| tests/core-impl/playlists/types/file/asx/testasxplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/m3u/testm3uplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/pls/testplsplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/testplaylistfilesupport.cpp | 0.000000 |
| tests/core-impl/playlists/types/file/xspf/testxspfplaylist.cpp   | 0.000000 |
| tests/core/collections/collectionlocationtest.cpp                | 0.000000 |

|   |          |
|---|----------|
| tests/core/interfaces/testmetacapability.cpp    | 0.000000 |
| tests/core/playlists/testplaylistformat.cpp     | 0.000000 |
| tests/core/playlists/testplaylistobserver.cpp   | 0.000000 |
| tests/mocks/sqlstoragemock.cpp                  | 0.000000 |
| tests/playlist/testplaylistmodels.cpp           | 0.000000 |
| tests/qt-modeltest/modeltest.cpp                | 0.000000 |
| tests/services/amazon/testamazonalbum.cpp       | 0.000000 |
| tests/services/amazon/testamazonmetafactory.cpp | 0.000000 |
| tests/services/amazon/testamazontrack.cpp       | 0.000000 |
| tests/testsmartpointerlist.cpp                  | 0.000000 |
| utilities/afttagger/afttagger.cpp               | 0.000000 |
| utilities/updatesigner/signer.cpp               | 0.000000 |

ΠΙΝΑΚΑΣ Α'5: Αποτελέσματα Big File Level για το Amarok για την Αναφορά Σφάλματος υπάριθμόν 322474

| ClusterName   | Score    | expectation | Connections | ToClique  | ClusterSize |
|---|----------|-------------|-------------|-----------|-------------|
| src/core-impl/collections/db/sql/sqlregistry.cpp            | 4.727273 | 0.050831    | 13.000000   | 11.000000 | 4           |
| src/context/popupdropper/libpud/popupdropper.cpp            | 4.764706 | 0.038425    | 27.000000   | 17.000000 | 3           |
| src/scriptmanager.cpp                                       | 4.800000 | 0.038710    | 16.000000   | 10.000000 | 3           |
| src/services/mp3tunes/mp3tuneslocker.cpp                    | 5.000000 | 0.026882    | 5.000000    | 2.000000  | 2           |
| src/playlist/playlistbreadcrumbitem.cpp                     | 5.142857 | 0.041475    | 12.000000   | 7.000000  | 3           |
| src/context/applets/upcomingevents/upcomingeventsapplet.cpp | 5.181818 | 0.041789    | 19.000000   | 11.000000 | 3           |
| src/aboutdialog/extendedaboutdialog.cpp                     | 5.400000 | 0.043548    | 9.000000    | 5.000000  | 3           |
| src/services/magnatune/magnatunestore.cpp                   | 5.600000 | 0.060215    | 21.000000   | 15.000000 | 4           |
| src/services/jamendo/jamendoxmlparser.cpp                   | 6.000000 | 0.032258    | 3.000000    | 1.000000  | 2           |

|  |                 |                 |                 |                 |          |
|--|-----------------|-----------------|-----------------|-----------------|----------|
| src/app.cpp  | 6.481481        | 0.087117        | 35.000000       | 27.000000       | 5        |
| src/services/lastfm/lastfm.service.cpp                             | 6.857143        | 0.073733        | 24.000000       | 14.000000       | 4        |
| src/context/applets/albums/albums.cpp                              | 7.857143        | 0.105607        | 22.000000       | 14.000000       | 5        |
| src/services/magnatune/magnatuneredownloadhandler.cpp              | 9.000000        | 0.072581        | 9.000000        | 3.000000        | 3        |
| src/playlist/playlistsortwidget.cpp                                | 9.333333        | 0.100358        | 14.000000       | 6.000000        | 4        |
| tests/synchronization/testonewaysynchronizationjob.cpp             | 9.483871        | 0.152966        | 49.000000       | 31.000000       | 6        |
| src/covermanager/covermanager.cpp                                  | 10.161290       | 0.191207        | 45.000000       | 31.000000       | 7        |
| src/playlist/proxymodels/sortfilterproxy.cpp                       | 10.500000       | 0.084677        | 7.000000        | 2.000000        | 3        |
| src/services/amazon/amazonstore.cpp                                | 10.714286       | 0.144009        | 15.000000       | 7.000000        | 5        |
| src/context/engines/wikipedia/wikipediaengine.cpp                  | 10.894737       | 0.263582        | 92.000000       | 76.000000       | 9        |
| src/mainwindow.cpp   | 10.962963       | 0.235763        | 74.000000       | 54.000000       | 8        |
| src/services/mp3tunes/libmp3tunes/locker.h                         | 13.000000       | 0.454301        | 4.000000        | 4.000000        | 13       |
| src/context/applets/analyzer/blockanalyzer.cpp                     | 15.000000       | 0.120968        | 10.000000       | 2.000000        | 3        |
| src/configdialog/dialogs/scriptsconfig.cpp                         | Infinity        | Infinity        | 4.000000        | 0.000000        | 3        |
| <b>src/transcoding/transcodingoptionsstackedwidget.cpp</b>         | <b>Infinity</b> | <b>Infinity</b> | <b>2.000000</b> | <b>0.000000</b> | <b>2</b> |
| src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp  | Infinity        | Infinity        | 1.000000        | 0.000000        | 1        |
| src/services/amazon/amazonshoppingcartview.cpp                     | NaN             | NaN             | 0.000000        | 0.000000        | 1        |
| src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp | NaN             | NaN             | 0.000000        | 0.000000        | 1        |

ΠΙΝΑΚΑΣ Α'.6: Αποτελέσματα First Cluster Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 329498

| File Name   | Connectivity To Clique |
|---|------------------------|
| <b>Cluster Name : src/covermanager/covermanager.cpp</b> |                        |
| src/configdialog/dialogs/excludedlabelsdialog.cpp       | 2.000000               |
| src/covermanager/coverfetchunit.cpp                     | 2.000000               |

|  |           |
|--|-----------|
| src/statusbar/compoundprogressbar.cpp  | 3.000000  |
| src/covermanager/coverfounddialog.cpp  | 5.000000  |
| src/covermanager/covermanager.cpp  | 10.000000 |
| <b>Cluster Name : src/configdialog/dialogs/scriptsconfig.cpp</b>             |           |
|  |           |
| <b>Cluster Name : src/context/popupdropper/libpud/popupdropper.cpp</b>       |           |
| src/context/popupdropper/libpud/popupdropperview.cpp                         | 1.000000  |
| src/context/popupdropper/libpud/popupdropperitem.cpp                         | 5.000000  |
| src/context/popupdropper/libpud/popupdropper.cpp                             | 11.000000 |
| <b>Cluster Name : src/playlist/playlistbreadcrumbitem.cpp</b>                |           |
| src/playlist/playlistbreadcrumbitem.cpp                                      | 2.000000  |
| src/playlist/playlistbreadcrumbitemsortbutton.cpp                            | 2.000000  |
| src/widgets/breadcrumbitembutton.cpp   | 3.000000  |
| <b>Cluster Name : src/services/amazon/amazonshoppingcartview.cpp</b>         |           |
|  |           |
| <b>Cluster Name : src/context/applets/albums/albums.cpp</b>                  |           |
| src/context/applets/albums/trackitem.cpp                                     | 1.000000  |
| src/context/applets/albums/albumsmodel.cpp                                   | 2.000000  |
| src/context/applets/albums/albumitem.cpp                                     | 2.000000  |
| src/context/applets/albums/albums.cpp  | 2.000000  |
| src/context/applets/albums/albumsview.cpp                                    | 7.000000  |
| <b>Cluster Name : tests/synchronization/testonewaysynchronizationjob.cpp</b> |           |
| src/browsers/singlecollectiontreeitemmodel.cpp                               | 1.000000  |
| tests/browsers/testsinglecollectiontreeitemmodel.cpp                         | 4.000000  |



|   |           |
|---|-----------|
| src/browsers/collectiontreeitemmodelbase.cpp                            | 6.000000  |
| tests/synchronization/testonewaysynchronizationjob.cpp                  | 7.000000  |
| src/browsers/collectiontreeitem.cpp                                     | 13.000000 |
| <b>Cluster Name : src/services/lastfm/lastfm.service.cpp</b>            |           |
| src/services/lastfm/avatar downloader.cpp                               | 1.000000  |
| src/services/lastfm/lastfm.service.settings.cpp                         | 2.000000  |
| src/services/lastfm/lastfm.treemodel.cpp                                | 4.000000  |
| src/services/lastfm/lastfm.service.cpp                                  | 6.000000  |
| <b>Cluster Name : src/playlist/playlistsortwidget.cpp</b>               |           |
| src/playlist/playlistmodelstack.cpp                                     | 2.000000  |
| src/playlist/playlistviewurlgenerator.cpp                               | 2.000000  |
| src/playlist/playlistsortwidget.cpp                                     | 2.000000  |
| <b>Cluster Name : src/aboutdialog/extendedaboutdialog.cpp</b>           |           |
| src/aboutdialog/extendedaboutdialog.cpp                                 | 1.000000  |
| src/aboutdialog/animatedwidget.cpp                                      | 1.000000  |
| src/aboutdialog/ocspersonitem.cpp                                       | 3.000000  |
| <b>Cluster Name : src/context/engines/wikipedia/wikipediaengine.cpp</b> |           |
| src/context/engines/wikipedia/wikipediaengine.cpp                       | 2.000000  |
| <b>Cluster Name : src/services/magnatune/magnatunestore.cpp</b>         |           |
| src/browsers/infoproxy.cpp  | 2.000000  |
| src/services/magnatune/magnatunestore.cpp                               | 5.000000  |
| <b>Cluster Name : src/context/applets/analyzer/blockanalyzer.cpp</b>    |           |
| src/context/applets/analyzer/blockanalyzer.cpp                          | 2.000000  |
| <b>Cluster Name : src/core-impl/collections/db/sql/sqlregistry.cpp</b>  |           |

|   |           |
|---|-----------|
| src/core-impl/collections/db/sql/databaseupdater.cpp                        | 2.000000  |
| src/core-impl/collections/db/mountpointmanager.cpp                          | 3.000000  |
| src/core-impl/collections/db/sql/sqlregistry.cpp                            | 6.000000  |
| <b>Cluster Name : src/services/amazon/amazonstore.cpp</b>                   |           |
| src/services/amazon/amazoninfoparser.cpp                                    | 1.000000  |
| src/services/amazon/amazonitemtreemodel.cpp                                 | 1.000000  |
| src/services/amazon/amazonstore.cpp   | 3.000000  |
| <b>Cluster Name : src/services/jamendo/jamendoxmlparser.cpp</b>             |           |
| src/services/jamendo/jamendodatabasehandler.cpp                             | 1.000000  |
| <b>Cluster Name : src/transcoding/transcodingoptionsstackedwidget.cpp</b>   |           |
| <b>Cluster Name : src/services/magnatune/magnatuneredownloadhandler.cpp</b> |           |
| src/services/magnatune/magnatuneredownloadinfo.cpp                          | 1.000000  |
| src/services/magnatune/magnatuneredownloadhandler.cpp                       | 2.000000  |
| <b>Cluster Name : src/mainwindow.cpp</b>                                    |           |
| src/palettehandler.cpp  | 1.000000  |
| src/browsers/browserdock.cpp  | 2.000000  |
| src/amarokurls/bookmarkmanager.cpp  | 3.000000  |
| src/context/applets/upcomingevents/upcomingeventsstackitem.cpp              | 4.000000  |
| src/dialogs/equalizerdialog.cpp   | 4.000000  |
| src/browsers/collectionbrowser/collectionwidget.cpp                         | 6.000000  |
| src/mainwindow.cpp  | 11.000000 |
| <b>Cluster Name : src/app.cpp</b>   |           |
| src/network/networkaccessviewer.cpp   | 1.000000  |

|  |          |
|--|----------|
| src/covermanager/covercache.cpp  | 2.000000 |
| src/network/networkaccessmanagerproxy.cpp  | 2.000000 |
| src/app.cpp  | 7.000000 |
| <b>Cluster Name : src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp</b>  |          |
| <b>Cluster Name : src/playlist/proxymodels/sortfilterproxy.cpp</b>                       |          |
| src/playlist/proxymodels/sortalgorithms.cpp  | 1.000000 |
| src/playlist/proxymodels/sortscheme.cpp  | 1.000000 |
| <b>Cluster Name : src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp</b> |          |
| <b>Cluster Name : src/services/mp3tunes/mp3tuneslocker.cpp</b>                           |          |
| src/services/mp3tunes/mp3tuneslocker.cpp   | 2.000000 |
| <b>Cluster Name : src/context/applets/upcomingevents/upcomingeventsapplet.cpp</b>        |          |
| src/context/applets/upcomingevents/lastfmevent.cpp                                       | 1.000000 |
| src/context/applets/upcomingevents/upcomingeventsmapwidget.cpp                           | 3.000000 |
| src/context/applets/upcomingevents/upcomingeventsapplet.cpp                              | 7.000000 |
| <b>Cluster Name : src/scriptmanager.cpp</b>  |          |
| src/services/scriptable/scriptableservicemanager.cpp                                     | 2.000000 |
| src/scriptmanager.cpp  | 5.000000 |
| <b>Cluster Name : src/services/mp3tunes/libmp3tunes/locker.h</b>                         |          |
| src/playlist/playlistqueueeditor.cpp   | 1.000000 |
| src/services/mp3tunes/libmp3tunes/harmony.c  | 1.000000 |
| tests/core-impl/collections/db/sql/testdatabaseupdater.cpp                               | 1.000000 |

|                                |          |
|--------------------------------|----------|
| tests/testsmartpointerlist.cpp | 1.000000 |
|--------------------------------|----------|

ΠΙΝΑΚΑΣ Α.7: Αποτελέσματα First File Level για το Amarok για την Αναφορά Σφάλματος υπέρριθμόν 329498

| File Name  | Connectivity To Clique |
|--|------------------------|
| <b>Cluster Name : src/covermanager/covermanager.cpp</b>                      |                        |
| src/statusbar/progressbar.cpp  | 0.000000               |
| src/statusbar/kjobprogressbar.cpp  | 0.000000               |
| <b>Cluster Name : src/configdialog/dialogs/scriptsconfig.cpp</b>             |                        |
| src/dialogs/scriptselector.cpp   | 0.000000               |
| src/configdialog/dialogs/scriptsconfig.cpp                                   | 0.000000               |
| src/configdialog/dialogs/scriptselector.cpp                                  | 0.000000               |
| <b>Cluster Name : src/context/popupdropper/libpud/popupdropper.cpp</b>       |                        |
|  |                        |
| <b>Cluster Name : src/playlist/playlistbreadcrumbitem.cpp</b>                |                        |
|  |                        |
| <b>Cluster Name : src/services/amazon/amazonshoppingcartview.cpp</b>         |                        |
| src/services/amazon/amazonshoppingcartview.cpp                               | 0.000000               |
| <b>Cluster Name : src/context/applets/albums/albums.cpp</b>                  |                        |
|  |                        |
| <b>Cluster Name : tests/synchronization/testonewaysynchronizationjob.cpp</b> |                        |
| src/services/scriptable/scriptableservicecollectiontreemodel.cpp             | 0.000000               |
| <b>Cluster Name : src/services/lastfm/lastfmservice.cpp</b>                  |                        |
|  |                        |

|   |          |
|---|----------|
| <b>Cluster Name : src/playlist/playlistsortwidget.cpp</b>               |          |
| tests/playlist/testplaylistmodels.cpp                                   | 0.000000 |
| <b>Cluster Name : src/aboutdialog/extendedaboutdialog.cpp</b>           |          |
|   |          |
| <b>Cluster Name : src/context/engines/wikipedia/wikipediaengine.cpp</b> |          |
| src/playlist/proxymodels/proxybase.cpp                                  | 0.000000 |
| tests/mocks/mocktrackforurlworker.cpp                                   | 0.000000 |
| src/dialogs/databaseimporterdialog.cpp                                  | 0.000000 |
| src/scriptengine/metatypeexporter.cpp                                   | 0.000000 |
| src/dbus/dbusqueryhelper.cpp  | 0.000000 |
| src/core-impl/meta/multi/multitrack.cpp                                 | 0.000000 |
| src/covermanager/coverfetchqueue.cpp                                    | 0.000000 |
| src/playlist/layouts/layoutitemconfig.cpp                               | 0.000000 |
| <b>Cluster Name : src/services/magnatune/magnatunestore.cpp</b>         |          |
| src/services/magnatune/magnatuneconfig.cpp                              | 0.000000 |
| src/services/magnatune/magnatunemeta.cpp                                | 0.000000 |
| <b>Cluster Name : src/context/applets/analyzer/blockanalyzer.cpp</b>    |          |
| src/context/applets/analyzer/fht.cpp                                    | 0.000000 |
| src/fht.cpp   | 0.000000 |
| <b>Cluster Name : src/core-impl/collections/db/sql/sqlregistry.cpp</b>  |          |
| src/core-impl/collections/db/sql/sqlmeta.cpp                            | 0.000000 |
| <b>Cluster Name : src/services/amazon/amazonstore.cpp</b>               |          |
| src/services/amazon/amazonshoppingcartdialog.cpp                        | 0.000000 |
| src/services/amazon/amazonwantcountrywidget.cpp                         | 0.000000 |

|  |                 |
|--|-----------------|
| <b>Cluster Name : src/services/jamendo/jamendoxmlparser.cpp</b>                          |                 |
| src/services/jamendo/jamendoxmlparser.cpp  | 0.000000        |
| <b>Cluster Name : src/transcoding/transcodingoptionsstackedwidget.cpp</b>                |                 |
| src/transcoding/transcodingoptionsstackedwidget.cpp                                      | 0.000000        |
| <b>src/core/transcoding/transcodingconfiguration.cpp</b>                                 | <b>0.000000</b> |
| <b>Cluster Name : src/services/magnatune/magnatuneredownloadhandler.cpp</b>              |                 |
| src/services/magnatune/magnatuneredownloadhandler.cpp                                    | 0.000000        |
| <b>Cluster Name : src/mainwindow.cpp</b>   |                 |
| src/widgets/amarokdockwidget.cpp   | 0.000000        |
| <b>Cluster Name : src/app.cpp</b>  |                 |
| src/mysqlservertester.cpp  | 0.000000        |
| <b>Cluster Name : src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp</b>  |                 |
| src/core-impl/collections/nepomukcollection/nepomukquerymaker.cpp                        | 0.000000        |
| <b>Cluster Name : src/playlist/proxymodels/sortfilterproxy.cpp</b>                       |                 |
| src/playlist/proxymodels/sortfilterproxy.cpp   | 0.000000        |
| <b>Cluster Name : src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp</b> |                 |
| src/core-impl/collections/ipodcollection/ipodcollectionfactory.cpp                       | 0.000000        |
| <b>Cluster Name : src/services/mp3tunes/mp3tuneslocker.cpp</b>                           |                 |
| src/services/mp3tunes/mp3tuneslockermeta.cpp   | 0.000000        |
| <b>Cluster Name : src/context/applets/upcomingevents/upcomingeventsapplet.cpp</b>        |                 |
|  |                 |
| <b>Cluster Name : src/scriptmanager.cpp</b>  |                 |
| src/scriptengine/amarokscriptableservicescript.cpp                                       | 0.000000        |

| <b>Cluster Name</b> : src/services/mp3tunes/libmp3tunes/locker.h |          |
|--|----------|
| tests/context/engines/testdataengineblackbox.cpp                 | 0.000000 |
| tests/services/amazon/testamazonmetafactory.cpp                  | 0.000000 |
| src/dialogs/deviceconfiguredialog.cpp                            | 0.000000 |
| tests/mocks/sqlstoragemock.cpp                                   | 0.000000 |
| tests/services/amazon/testamazontrack.cpp                        | 0.000000 |
| src/core/playlists/playlistformat.cpp                            | 0.000000 |
| src/services/magnatune/magnatuneredownloaddialog.cpp             | 0.000000 |
| tests/services/amazon/testamazonalbum.cpp                        | 0.000000 |
| utilities/updatesigner/signer.cpp                                | 0.000000 |

ΠΙΝΑΚΑΣ Α'.8: Αποτελέσματα Zero File Level για το Amarok για την Αναφορά Σφάλματος υπάριθμόν 329498

| <b>ClusterName</b>       | <b>Score</b> | <b>expectation</b> | <b>Connections</b> | <b>ToClique</b> | <b>ClusterSize</b> |
|--------------------------|--------------|--------------------|--------------------|-----------------|--------------------|
| src/core/support/debug.h | 509.317136   | 277.933813         | 981.000000         | 391.000000      | 203                |

ΠΙΝΑΚΑΣ Α'.9: Αποτελέσματα Big Cluster Level για το Amarok για την Αναφορά Σφάλματος υπ' αριθμόν 329498

| <b>File Name</b>                               | <b>Connectivity To Clique</b> |
|--|-------------------------------|
| <b>Cluster Name</b> : src/core/support/debug.h |                               |
| src/covermanager/coverfetcher.cpp              | 4.000000                      |
| src/toolbar/maintoolbar.cpp                    | 4.000000                      |
| src/browsers/collectiontreeview.cpp            | 2.000000                      |
| src/context/engines/current/currentengine.cpp  | 2.000000                      |
| src/context/engines/labels/labelsengine.cpp    | 2.000000                      |

|   |          |
|---|----------|
| src/widgets/trackselectwidget.cpp                                   | 2.000000 |
| tests/core-impl/collections/db/sql/testsqlalbum.cpp                 | 2.000000 |
| src/amarokmimedata.cpp  | 1.000000 |
| src/amarokurls/amarokurl.cpp  | 1.000000 |
| src/amarokurls/bookmarkmodel.cpp                                    | 1.000000 |
| src/amarokurls/bookmarktreeview.cpp                                 | 1.000000 |
| src/amarokurls/contexturlrunner.cpp                                 | 1.000000 |
| src/amarokurls/navigationurlrunner.cpp                              | 1.000000 |
| src/amarokurls/playurlrunner.cpp                                    | 1.000000 |
| src/browsers/browsercategorylist.cpp                                | 1.000000 |
| src/browsers/collectiontreitemmodel.cpp                             | 1.000000 |
| src/browsers/filebrowser/fileview.cpp                               | 1.000000 |
| src/context/engines/upcomingevents/upcomingeventsengine.cpp         | 1.000000 |
| src/context/widgets/droppixmapitem.cpp                              | 1.000000 |
| src/core-impl/collections/db/sql/sqlcapabilities.cpp                | 1.000000 |
| src/core-impl/collections/db/sql/sqlregistry_p.cpp                  | 1.000000 |
| src/core-impl/collections/db/sql/sqlscanresultprocessor.cpp         | 1.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipodcopytracksjob.cpp | 1.000000 |
| src/core-impl/collections/playdarcollection/playdarcollection.cpp   | 1.000000 |
| src/core-impl/collections/support/memorycustomvalue.cpp             | 1.000000 |
| src/core-impl/collections/support/memorymatcher.cpp                 | 1.000000 |
| src/core-impl/support/trackloader.cpp                               | 1.000000 |
| src/core/support/amarok.cpp   | 1.000000 |
| src/covermanager/coverfetchingactions.cpp                           | 1.000000 |
| src/dialogs/musicbrainztagger.cpp                                   | 1.000000 |



|  |          |
|--|----------|
| src/dialogs/trackorganizer.cpp                       | 1.000000 |
| src/enginecontroller.cpp                             | 1.000000 |
| src/mac/growlinterface.cpp                           | 1.000000 |
| src/musicbrainz/musicbrainzfinder.cpp                | 1.000000 |
| src/musicbrainz/musicbrainztagsitem.cpp              | 1.000000 |
| src/musicbrainz/musicdnssaudiodecoder.cpp            | 1.000000 |
| src/musicbrainz/musicdnsfinder.cpp                   | 1.000000 |
| src/playlist/playlisttoolbar.cpp                     | 1.000000 |
| src/playlistmanager/sql/sqlplaylist.cpp              | 1.000000 |
| src/playlistmanager/syncedplaylist.cpp               | 1.000000 |
| src/scanner/abstractscanresultprocessor.cpp          | 1.000000 |
| src/scriptengine/amaroknetworkscript.cpp             | 1.000000 |
| src/services/amazon/amazonitemtreeview.cpp           | 1.000000 |
| src/services/jamendo/jamendoservice.cpp              | 1.000000 |
| src/services/lastfm/lovetrackaction.cpp              | 1.000000 |
| src/services/lastfm/scrobbleradapter.cpp             | 1.000000 |
| src/services/magnatune/magnatunedatabaseworker.cpp   | 1.000000 |
| src/services/magnatune/magnatuneneedupdatewidget.cpp | 1.000000 |
| src/services/mp3tunes/mp3tunesworkers.cpp            | 1.000000 |
| src/services/opmldirectory/opmldirectoryservice.cpp  | 1.000000 |
| src/services/servicealbumcoverdownloader.cpp         | 1.000000 |
| src/services/servicebase.cpp                         | 1.000000 |
| src/services/servicecapabilities.cpp                 | 1.000000 |
| src/statsyncing/jobs/matchtracksjob.cpp              | 1.000000 |
| src/synchronization/synchronizationbasejob.cpp       | 1.000000 |

|   |          |
|---|----------|
| src/widgets/filenamelayoutwidget.cpp                        | 1.000000 |
| tests/core-impl/collections/db/sql/testsqlscanmanager.cpp   | 1.000000 |
| tests/core-impl/collections/db/sql/testsqltrack.cpp         | 1.000000 |
| tests/core-impl/meta/file/testmetafiletrack.cpp             | 1.000000 |
| tests/core-impl/support/testtrackloader.cpp                 | 1.000000 |
| tests/core/collections/support/testtrackforurlworker.cpp    | 1.000000 |
| tests/core/meta/support/testmetaalbumkey.cpp                | 1.000000 |
| tests/core/meta/support/testmetatrackkey.cpp                | 1.000000 |
| tests/core/meta/support/testprivatemetaregistry.cpp         | 1.000000 |
| tests/core/meta/testmetatrack.cpp                           | 1.000000 |
| tests/synchronization/testmasterslavesynchronizationjob.cpp | 1.000000 |
| tests/testtrackorganizer.cpp                                | 1.000000 |
| shared/filetype.cpp   | 0.000000 |
| src/aboutdialog/ocspersonlistwidget.cpp                     | 0.000000 |
| src/amarokurls/bookmarkcurrentbutton.cpp                    | 0.000000 |
| src/amarokurls/bookmarkgroup.cpp                            | 0.000000 |
| src/browsers/browsercategory.cpp                            | 0.000000 |
| src/browsers/browsercategorylistmodel.cpp                   | 0.000000 |
| src/browsers/browsermessagearea.cpp                         | 0.000000 |
| src/browsers/collectionsortfilterproxymodel.cpp             | 0.000000 |
| src/browsers/filebrowser/filebrowser.cpp                    | 0.000000 |
| src/browsers/playlistbrowser/playlistbrowsermodel.cpp       | 0.000000 |
| src/browsers/playlistbrowser/playlistbrowserview.cpp        | 0.000000 |
| src/browsers/playlistbrowser/playlistsbyproviderproxy.cpp   | 0.000000 |
| src/browsers/playlistbrowser/playlistsinfoldersproxy.cpp    | 0.000000 |

|   |          |
|---|----------|
| src/browsers/playlistbrowser/qtgroupingproxy.cpp            | 0.000000 |
| src/browsers/servicebrowser/servicebrowser.cpp              | 0.000000 |
| src/configdialog/configdialog.cpp                           | 0.000000 |
| src/configdialog/dialogs/notificationsconfig.cpp            | 0.000000 |
| src/configdialog/dialogs/pluginsconfig.cpp                  | 0.000000 |
| src/context/applets/analyzer/analyzerapplet.cpp             | 0.000000 |
| src/context/applets/analyzer/ballsanalyzer.cpp              | 0.000000 |
| src/context/applets/analyzer/discoanalyzer.cpp              | 0.000000 |
| src/context/applets/info/infoapplet.cpp                     | 0.000000 |
| src/context/applets/labels/labelsapplet.cpp                 | 0.000000 |
| src/context/applets/songkick/songkickapplet.cpp             | 0.000000 |
| src/context/applets/tabs/tabsview.cpp                       | 0.000000 |
| src/context/applets/upcomingevents/upcomingeventsstack.cpp  | 0.000000 |
| src/context/applets/wikipedia/wikipediaapplet.cpp           | 0.000000 |
| src/context/contextdock.cpp                                 | 0.000000 |
| src/context/contextobserver.cpp                             | 0.000000 |
| src/context/engines/lyrics/lyricsengine.cpp                 | 0.000000 |
| src/context/engines/photos/photosengine.cpp                 | 0.000000 |
| src/context/engines/similarartists/similarartistsengine.cpp | 0.000000 |
| src/context/engines/tabs/tabsengine.cpp                     | 0.000000 |
| src/context/widgets/containmentselectionlayer.cpp           | 0.000000 |
| src/context/widgets/recentlyplayedlistwidget.cpp            | 0.000000 |
| src/context/widgets/textscrollingwidget.cpp                 | 0.000000 |
| src/context/widgets/toolboxicon.cpp                         | 0.000000 |
| src/core-impl/collections/aggregate/aggreatemeta.cpp        | 0.000000 |

|  |          |
|--|----------|
| src/core-impl/collections/audiocd/audiocdcollection.cpp                          | 0.000000 |
| src/core-impl/collections/db/sql/device/massstorage/massstoragedevicehandler.cpp | 0.000000 |
| src/core-impl/collections/db/sql/device/nfs/nfsdevicehandler.cpp                 | 0.000000 |
| src/core-impl/collections/db/sql/device/smb/smbdevicehandler.cpp                 | 0.000000 |
| src/core-impl/collections/db/sql/mysql-shared/mysqlstorage.cpp                   | 0.000000 |
| src/core-impl/collections/db/sql/mysqlcollection/mysqlembeddedstorage.cpp        | 0.000000 |
| src/core-impl/collections/db/sql/mysqlservercollection/mysqlservercollection.cpp | 0.000000 |
| src/core-impl/collections/db/sql/sqlcollection.cpp                               | 0.000000 |
| src/core-impl/collections/db/sql/sqlcollectionlocation.cpp                       | 0.000000 |
| src/core-impl/collections/db/sql/sqlquerymaker.cpp                               | 0.000000 |
| src/core-impl/collections/ipodcollection/ipodcollection.cpp                      | 0.000000 |
| src/core-impl/collections/ipodcollection/ipodcollectionlocation.cpp              | 0.000000 |
| src/core-impl/collections/ipodcollection/ipodmeta.cpp                            | 0.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipoddeletetracksjob.cpp            | 0.000000 |
| src/core-impl/collections/ipodcollection/jobs/ipodparsetracksjob.cpp             | 0.000000 |
| src/core-impl/collections/ipodcollection/support/iphonemountpoint.cpp            | 0.000000 |
| src/core-impl/collections/nepomukcollection/nepomukcache.cpp                     | 0.000000 |
| src/core-impl/collections/playdarcollection/playdarquerymaker.cpp                | 0.000000 |
| src/core-impl/collections/support/collectionmanager.cpp                          | 0.000000 |
| src/core-impl/collections/support/filecollectionlocation.cpp                     | 0.000000 |
| src/core-impl/collections/support/jobs/writetagsjob.cpp                          | 0.000000 |
| src/core-impl/collections/upnpcollection/dbuscodec.cpp                           | 0.000000 |
| src/core-impl/logger/proxylogger.cpp   | 0.000000 |
| src/core-impl/meta/timecode/timecodeobserver.cpp                                 | 0.000000 |
| src/core-impl/support/tagstatisticsstore.cpp                                     | 0.000000 |

|  |          |
|--|----------|
| src/core-impl/support/urlstatisticsstore.cpp         | 0.000000 |
| src/core/collections/collectionlocation.cpp          | 0.000000 |
| src/core/podcasts/podcastimagefetcher.cpp            | 0.000000 |
| src/core/support/debug.cpp                           | 0.000000 |
| src/databaseimporter/amarok14/fastforwardworker.cpp  | 0.000000 |
| src/databaseimporter/itunes/itunesimporterworker.cpp | 0.000000 |
| src/dbus/mpris1/playerhandler.cpp                    | 0.000000 |
| src/dialogs/deletedialog.cpp                         | 0.000000 |
| src/dialogs/editfilterdialog.cpp                     | 0.000000 |
| src/dialogs/organizecollectiondialog.cpp             | 0.000000 |
| src/dialogs/transferdialog.cpp                       | 0.000000 |
| src/dynamic/dynamicmodel.cpp                         | 0.000000 |
| src/main.cpp   | 0.000000 |
| src/musicbrainz/musicbrainztagsmodeldelegate.cpp     | 0.000000 |
| src/opmlparser.cpp                                   | 0.000000 |
| src/opmlwriter.cpp                                   | 0.000000 |
| src/playback/fadeouter.cpp                           | 0.000000 |
| src/playlist/playlistmodel.cpp                       | 0.000000 |
| src/playlist/playlistviewurlrunner.cpp               | 0.000000 |
| src/playlist/view/listview/inlineditorwidget.cpp     | 0.000000 |
| src/playlistmanager/file/kconfigsyncrelstore.cpp     | 0.000000 |
| src/playlistmanager/syncrelationstorage.cpp          | 0.000000 |
| src/scanner/abstractdirectorywatcher.cpp             | 0.000000 |
| src/scanner/genericscanmanager.cpp                   | 0.000000 |
| src/scanner/genericscannerjob.cpp                    | 0.000000 |

|   |          |
|---|----------|
| src/scriptengine/scriptimporter.cpp                         | 0.000000 |
| src/scriptupdater.cpp                                       | 0.000000 |
| src/services/ampache/ampacheaccountlogin.cpp                | 0.000000 |
| src/services/ampache/ampacheservice.cpp                     | 0.000000 |
| src/services/gpodder/gpodderserviceconfig.cpp               | 0.000000 |
| src/services/infoparserbase.cpp                             | 0.000000 |
| src/services/lastfm/lastfmserviceconfig.cpp                 | 0.000000 |
| src/services/lastfm/lastfmtreeview.cpp                      | 0.000000 |
| src/services/lastfm/meta/lastfmplayablecapability.cpp       | 0.000000 |
| src/services/magnatune/magnatunealbumdownloader.cpp         | 0.000000 |
| src/services/mp3tunes/harmonydaemon/main.cpp                | 0.000000 |
| src/services/mp3tunes/libmp3tunes/locker.c                  | 0.000000 |
| src/services/opmldirectory/opmldirectorydatabasehandler.cpp | 0.000000 |
| src/services/scriptable/scriptableservicequerymaker.cpp     | 0.000000 |
| src/services/servicepluginmanager.cpp                       | 0.000000 |
| src/services/servicesqlquerymaker.cpp                       | 0.000000 |
| src/services/servicesqlregistry.cpp                         | 0.000000 |
| src/statsyncing/ui/matchedtrackspage.cpp                    | 0.000000 |
| src/statusbar/longmessagewidget.cpp                         | 0.000000 |
| src/transcoding/transcodingjob.cpp                          | 0.000000 |
| src/transcoding/transcodingpropertysliderwidget.cpp         | 0.000000 |
| src/widgets/bookmarkpopup.cpp                               | 0.000000 |
| src/widgets/bookmarktriangle.cpp                            | 0.000000 |
| src/widgets/flowlayout.cpp                                  | 0.000000 |
| src/widgets/horizontaldivider.cpp                           | 0.000000 |

|  |          |
|--|----------|
| src/widgets/iconbutton.cpp                                       | 0.000000 |
| src/widgets/osd.cpp  | 0.000000 |
| src/widgets/prettytreedelegate.cpp                               | 0.000000 |
| src/widgets/token.cpp  | 0.000000 |
| src/widgets/tokenpool.cpp  | 0.000000 |
| src/widgets/trackactionbutton.cpp                                | 0.000000 |
| src/widgets/volumedial.cpp                                       | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlcollectionlocation.cpp | 0.000000 |
| tests/core-impl/collections/db/sql/testsqlquerymaker.cpp         | 0.000000 |
| tests/core-impl/meta/multi/testmetamultitrack.cpp                | 0.000000 |
| tests/core-impl/playlists/types/file/asx/testasxplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/m3u/testm3uplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/pls/testplsplaylist.cpp     | 0.000000 |
| tests/core-impl/playlists/types/file/testplaylistfilesupport.cpp | 0.000000 |
| tests/core-impl/playlists/types/file/xspf/testxspfplaylist.cpp   | 0.000000 |
| tests/core/collections/collectionlocationtest.cpp                | 0.000000 |
| tests/core/collections/testcollection.cpp                        | 0.000000 |
| tests/core/interfaces/testmetacapability.cpp                     | 0.000000 |
| tests/core/playlists/testplaylistformat.cpp                      | 0.000000 |
| tests/core/playlists/testplaylistobserver.cpp                    | 0.000000 |
| tests/dynamic/testdynamicmodel.cpp                               | 0.000000 |
| tests/qt-modeltest/modeltest.cpp                                 | 0.000000 |
| utilities/afttagger/afttagger.cpp                                | 0.000000 |

ΠΙΝΑΚΑΣ Α'.10: Αποτελέσματα Big File Level για το Amarok για την Αναφορά Σφάλματος υπάριθμόν 329498





## Παράρτημα Β'

# Konqueror Sample Results

143

Στους Επόμενους Πίνακες Παρουσιάζονται τα αποτελέσματα για το σύστημα Konqueror.

| ClusterName                              | Score     | expectation | Connections | ToClique   | ClusterSize |
|--|-----------|-------------|-------------|------------|-------------|
| kioslave/file/file_unix.cpp              | 4.736842  | 0.026661    | 60.000000   | 38.000000  | 3           |
| kjsembed/examples/kjsconsole/console.cpp | 6.000000  | 0.022514    | 3.000000    | 1.000000   | 2           |
| khtml/test_regression_fontoverload.cpp   | 6.206897  | 0.046581    | 45.000000   | 29.000000  | 4           |
| kfile/kurlnavigator.cpp                  | 6.500000  | 0.024390    | 13.000000   | 4.000000   | 2           |
| plasma/applets/folderview/iconview.cpp   | 6.755556  | 0.050698    | 76.000000   | 45.000000  | 4           |
| kdeui/widgets/kmainwindow.cpp            | 6.761062  | 0.076110    | 382.000000  | 339.000000 | 6           |
| kjs/grammar.cpp                          | 14.745763 | 0.248990    | 290.000000  | 177.000000 | 9           |
| security/kcert/kcertpart.cc              | 19.650000 | 0.442402    | 131.000000  | 80.000000  | 12          |
| kded/kbuildsycoca.cpp                    | 21.252525 | 0.637974    | 263.000000  | 198.000000 | 16          |

|  |                  |                 |                  |                  |           |
|--|------------------|-----------------|------------------|------------------|-----------|
| kdesu/kdesu_stub.c                           | 22.000000        | 0.165103        | 22.000000        | 4.000000         | 4         |
| <b>khtml/misc/htmlnames.h</b>                | <b>24.700000</b> | <b>0.602439</b> | <b>76.000000</b> | <b>40.000000</b> | <b>13</b> |
| kdoctools/meinproc.cpp                       | 27.690852        | 0.987104        | 924.000000       | 634.000000       | 19        |
| dolphin/src/dolphinpart_ext.cpp              | 35.894737        | 0.740792        | 62.000000        | 19.000000        | 11        |
| konqueror/src/konqmainwindow.cpp             | 38.057471        | 1.570852        | 301.000000       | 174.000000       | 22        |
| khtml/css/css_stylesheetimpl.cpp             | Infinity         | Infinity        | 1.000000         | 0.000000         | 1         |
| khtml/java/netscape/javascript/jsobject.java | Infinity         | Infinity        | 2.000000         | 0.000000         | 2         |
| khtml/css/cssparser.h                        | NaN              | NaN             | 0.000000         | 0.000000         | 1         |

ΠΙΝΑΚΑΣ Β'.1: Αποτελέσματα First Cluster Level για το Konqueror για την Αναφορά Σφάλματος υπ' αριθμόν 325724

| File Name   | Connectivity To Clique |
|---|------------------------|
| <b>Cluster Name : kded/kbuildsycoca.cpp</b>           |                        |
| kded/kbuildservicetypefactory.cpp                     | 1.000000               |
| kdecore/sycoca/kmemfile.cpp                           | 1.000000               |
| kdecore/services/kservicetypefactory.cpp              | 1.000000               |
| kdecore/network/k3socketdevice.cpp                    | 1.000000               |
| kdecore/sycoca/ksycocafactory.cpp                     | 3.000000               |
| kded/kbuildservicefactory.cpp                         | 3.000000               |
| kjsembed/kjsembed/settings.cpp                        | 6.000000               |
| kded/vfolder_menu.cpp                                 | 7.000000               |
| kdecore/services/kservicetype.cpp                     | 53.000000              |
| kdecore/sycoca/ksycocaentry.cpp                       | 60.000000              |
| kded/kbuildsycoca.cpp                                 | 62.000000              |
| <b>Cluster Name : dolphin/src/dolphinpart_ext.cpp</b> |                        |

|  |            |
|--|------------|
| dolphin/src/dolphinpart_ext.cpp                                | 1.000000   |
| dolphin/src/kitemviews/kfileitemlistview.cpp                   | 2.000000   |
| dolphin/src/kitemviews/kitemlistviewaccessible.cpp             | 2.000000   |
| dolphin/src/kitemviews/kitemlistcontroller.cpp                 | 3.000000   |
| dolphin/src/views/dolphinview.cpp                              | 5.000000   |
| dolphin/src/kitemviews/kfileitemmodel.cpp                      | 6.000000   |
| <b>Cluster Name : kdeui/widgets/kmainwindow.cpp</b>            |            |
| dolphin/src/views/tooltips/tooltipmanager.cpp                  | 2.000000   |
| kdeui/widgets/ktoolbar.cpp                                     | 5.000000   |
| kdeui/widgets/kmenubar.cpp                                     | 7.000000   |
| kio/kio/kurifilter.cpp   | 21.000000  |
| kdeui/widgets/kmainwindow.cpp                                  | 69.000000  |
| kdeui/icons/kiconloader.cpp                                    | 235.000000 |
| <b>Cluster Name : khtml/css/css_stylesheetimpl.cpp</b>         |            |
|  |            |
| <b>Cluster Name : khtml/css/cssparser.h</b>                    |            |
|  |            |
| <b>Cluster Name : kjsembed/examples/kjsconsole/console.cpp</b> |            |
| kjsembed/examples/kjsconsole/console.cpp                       | 1.000000   |
| <b>Cluster Name : plasma/applets/folderview/iconview.cpp</b>   |            |
| plasma/applets/folderview/iconview.cpp                         | 8.000000   |
| kio/kio/kfileitemactions.cpp                                   | 9.000000   |
| plasma/applets/folderview/folderview.cpp                       | 11.000000  |
| kfile/knewfilemenu.cpp   | 17.000000  |

| <b>Cluster Name : security/kcert/kcertpart.cc</b> |           |
|---|-----------|
| kio/kssl/kopenssl.cpp                             | 1.000000  |
| kio/kssl/ksslpemcallback.cpp                      | 2.000000  |
| kingio/dds.cpp                                    | 2.000000  |
| kio/kssl/ksslcertificate.cpp                      | 4.000000  |
| kdecore/date/ktimezone.cpp                        | 13.000000 |
| security/kcert/kcertpart.cc                       | 58.000000 |
| <b>Cluster Name : kdesu/kdesu_stub.c</b>          |           |
| kdeui/tests/kxerrorhandlertest.cpp                | 1.000000  |
| kdecore/tests/kcharsetstest.cpp                   | 1.000000  |
| kio/misc/kpac/kpac_dhcp_helper.c                  | 1.000000  |
| kdesu/kdesu_stub.c                                | 1.000000  |
| <b>Cluster Name : kdoctools/meinproc.cpp</b>      |           |
| kded/kctimefactory.cpp                            | 1.000000  |
| kdecore/kernel/kkernel_mac.cpp                    | 2.000000  |
| konqueror/src/konqfactory.cpp                     | 4.000000  |
| konqueror/src/konqsessionmanager.cpp              | 6.000000  |
| keditbookmarks/importers.cpp                      | 6.000000  |
| dolphin/src/main.cpp                              | 13.000000 |
| kinit/klauncher_main.cpp                          | 13.000000 |
| kdoctools/main_ghelp.cpp                          | 13.000000 |
| konqueror/src/konqmain.cpp                        | 13.000000 |
| kdoctools/main.cpp                                | 13.000000 |
| kioslave/metainfo/metainfo.cpp                    | 14.000000 |
| kconf_update/kconf_update.cpp                     | 14.000000 |

|  |                  |
|--|------------------|
| kioslave/http/http_cache_cleaner.cpp                               | 14.000000        |
| konqueror/client/kfmclient.cpp                                     | 16.000000        |
| keditbookmarks/main.cpp  | 18.000000        |
| kdoctools/meinproc.cpp   | 90.000000        |
| <b>Cluster Name : kioslave/file/file_unix.cpp</b>                  |                  |
| kioslave/file/file_unix.cpp  | 3.000000         |
| kioslave/file/file_win.cpp   | 3.000000         |
| kioslave/file/file.cpp   | 32.000000        |
| <b>Cluster Name : kfile/kurlnavigator.cpp</b>                      |                  |
| lib/konq/konq_historyprovider.cpp                                  | 2.000000         |
| kfile/kurlnavigator.cpp  | 2.000000         |
| <b>Cluster Name : khtml/misc/htmlnames.h</b>                       |                  |
| khtml/html/html_formimpl.cpp                                       | 1.000000         |
| khtml/ecma/kjs_dom.cpp   | 1.000000         |
| konqueror/src/konqhistoryproxymodel.cpp                            | 1.000000         |
| khtml/ecma/kjs_navigator.cpp                                       | 1.000000         |
| khtml/java/kjavaapplet.cpp   | 4.000000         |
| kdeui/util/kxmessages.cpp  | 5.000000         |
| kdecore/io/ksavefile.cpp   | 7.000000         |
| khtml/khtmlview.cpp  | 9.000000         |
| <b>textitkhtml/khtml_ext.cpp</b>                                   | <b>11.000000</b> |
| <b>Cluster Name : khtml/java/netscape/javascript/jsobject.java</b> |                  |
|  |                  |
| <b>Cluster Name : kjs/grammar.cpp</b>                              |                  |

|  |            |
|--|------------|
| khtml/css/parser.cpp   | 1.000000   |
| khtml/dom/dom_string.cpp                                     | 1.000000   |
| kdecore/services/yacc.c                                      | 1.000000   |
| solid/solid/predicate_parser.c                               | 1.000000   |
| khtml/xpath/parser.cpp                                       | 1.000000   |
| khtml/css/css_mediaquery.cpp                                 | 1.000000   |
| kjs/grammar.cpp  | 171.000000 |
| <b>Cluster Name : khtml/test_regression_fontoverload.cpp</b> |            |
| kdeui/widgets/kdeuiwidgetsproxystyle.cpp                     | 1.000000   |
| kdeui/widgets/kcharselect.cpp                                | 3.000000   |
| plasma/private/tooltip.cpp                                   | 4.000000   |
| khtml/test_regression_fontoverload.cpp                       | 21.000000  |
| <b>Cluster Name : konqueror/src/konqmainwindow.cpp</b>       |            |
| kdeui/shortcuts/kshortcut.cpp                                | 1.000000   |
| kdeui/tests/klineedit_uitest.cpp                             | 1.000000   |
| konqueror/src/konqmainwindowadaptor.cpp                      | 1.000000   |
| kdeui/paged/kpagewidgetmodel.cpp                             | 1.000000   |
| kio/kio/kurlcompletion.cpp                                   | 2.000000   |
| kdeui/widgets/kcompletionbox.cpp                             | 3.000000   |
| konqueror/src/konqsettings.cpp                               | 3.000000   |
| kdeui/util/kcompletionbase.cpp                               | 4.000000   |
| kdeui/widgets/kmenu.cpp                                      | 5.000000   |
| kdeui/paged/kpageview.cpp                                    | 5.000000   |
| kdeui/util/kcompletion.cpp                                   | 5.000000   |
| kdeui/widgets/khistorycombobox.cpp                           | 6.000000   |

|   |           |
|---|-----------|
| konqueror/sidebar/trees/bookmark_module/bookmark_module.cpp | 8.000000  |
| lib/konq/konq_popupmenu.cpp                                 | 9.000000  |
| kutils/kcmultidialog.cpp                                    | 9.000000  |
| konqueror/src/konqcombo.cpp                                 | 11.000000 |
| kio/kio/kprotocolmanager.cpp                                | 12.000000 |
| kdecore/services/kmimetypetrader.cpp                        | 12.000000 |
| konqueror/src/konqmainwindow.cpp                            | 76.000000 |

ΠΙΝΑΚΑΣ Β'.2: Αποτελέσματα First File Level για το Konqueror για την Αναφορά Σφάλματος υπάριθμόν 325724

| File Name   | Connectivity To Clique |
|---|------------------------|
| <b>Cluster Name : kded/kbuildsycoca.cpp</b>           |                        |
| kjs/bytocode/generator/parser.cpp                     | 0.000000               |
| kded/kbuildservicegroupfactory.cpp                    | 0.000000               |
| kded/kbuildmimetypefactory.cpp                        | 0.000000               |
| kded/kbuildprotocolinfofactory.cpp                    | 0.000000               |
| kjs/bytocode/generator/lexer.cpp                      | 0.000000               |
| <b>Cluster Name : dolphin/src/dolphinpart_ext.cpp</b> |                        |
| dolphin/src/kitemviews/kitemset.cpp                   | 0.000000               |
| dolphin/src/tests/kitemlistselectionmanagertest.cpp   | 0.000000               |
| khtml/svg/svgtextcontentelement.cpp                   | 0.000000               |
| dolphin/src/tests/kitemsettest.cpp                    | 0.000000               |
| dolphin/src/kitemviews/kitemlistselectionmanager.cpp  | 0.000000               |
| <b>Cluster Name : kdeui/widgets/kmainwindow.cpp</b>   |                        |
|   |                        |

|  |          |
|--|----------|
| <b>Cluster Name : khtml/css/css_stylesheetimpl.cpp</b>         |          |
| khtml/css/css_stylesheetimpl.cpp                               | 0.000000 |
| <b>Cluster Name : khtml/css/cssparser.h</b>                    |          |
| khtml/css/cssparser.cpp  | 0.000000 |
| <b>Cluster Name : kjsembed/examples/kjsconsole/console.cpp</b> |          |
| kjsembed/examples/kjsconsole/kjs_object_model.cpp              | 0.000000 |
| <b>Cluster Name : plasma/applets/folderview/iconview.cpp</b>   |          |
| <b>Cluster Name : security/kcert/kcertpart.cc</b>              |          |
| kio/kssl/ksslx509map.cpp                                       | 0.000000 |
| kio/kssl/ksslcertchain.cpp                                     | 0.000000 |
| kio/kssl/ksslsigners.cpp                                       | 0.000000 |
| kio/kssl/ksslpkcs7.cpp   | 0.000000 |
| kio/kssl/ksslx509v3.cpp  | 0.000000 |
| konqueror/src/konqframestatusbar.cpp                           | 0.000000 |
| <b>Cluster Name : kdesu/kdesu_stub.c</b>                       |          |
| <b>Cluster Name : kdoctools/meinproc.cpp</b>                   |          |
| dolphin/src/dolphinapplication.cpp                             | 0.000000 |
| kde3support/kdeui/k3buttonbox.cpp                              | 0.000000 |
| keeditbookmarks/globalbookmarkmanager.cpp                      | 0.000000 |
| <b>Cluster Name : kioslave/file/file_unix.cpp</b>              |          |



|  |          |
|--|----------|
| <b>Cluster Name : kfile/kurlnavigator.cpp</b>                      |          |
|  |          |
| <b>Cluster Name : khtml/misc/htmlnames.h</b>                       |          |
| khtml/ecma/kjs_window.cpp  | 0.000000 |
| khtml/xml/dom_nodeimpl.cpp   | 0.000000 |
| khtml/xml/dom_textimpl.cpp   | 0.000000 |
| khtml/rendering/render_form.cpp                                    | 0.000000 |
| <b>Cluster Name : khtml/java/netscape/javascript/jsobject.java</b> |          |
| khtml/java/org/kde/javascript/jsobject.java                        | 0.000000 |
| khtml/java/netscape/javascript/jsobject.java                       | 0.000000 |
| <b>Cluster Name : kjs/grammar.cpp</b>                              |          |
| khtml/ecma/xmlhttprequest.cpp                                      | 0.000000 |
| khtml/xml/dom_stringimpl.cpp                                       | 0.000000 |
| <b>Cluster Name : khtml/test_regression_fontoverload.cpp</b>       |          |
|  |          |
| <b>Cluster Name : konqueror/src/konqmainwindow.cpp</b>             |          |
| konqueror/src/konqndomanager.cpp                                   | 0.000000 |
| konqueror/src/konqapplication.cpp                                  | 0.000000 |
| kdeui/actions/ktogglefullscreenaction.cpp                          | 0.000000 |

ΠΙΝΑΚΑΣ Β'.3: Αποτελέσματα Zero File Level για το Konqueror για την Αναφορά Σφάλματος υπάριθμόν 325724

| ClusterName              | Score       | expectation | Connections  | ToClique    | ClusterSize |
|--------------------------|-------------|-------------|--------------|-------------|-------------|
| kdecore/sycoca/ksycoca.h | 5440.444706 | 3705.218439 | 38218.000000 | 2550.000000 | 363         |

ΠΙΝΑΚΑΣ Β'.4: Αποτελέσματα Big Cluster Level για το Konqueror για την Αναφορά Σφάλματος υπ' αριθμόν 325724

| τεξτλατινΦιλε Ναμε<br>Cluster Name : kdecore/sycoca/ksycoca.h | Connectivity To Clique |
|---|------------------------|
| kio/kio/dataprotocol.cpp                                      | 15.000000              |
| kdeui/tests/kcomboboxtest.cpp                                 | 3.000000               |
| kdeui/kernel/kglobalsettings.cpp                              | 2.000000               |
| kdeui/tests/kjobtrackerstest.cpp                              | 2.000000               |
| kdeui/tests/kledtest.cpp                                      | 2.000000               |
| kdeui/tests/kpixmapsequenceoverlaypaintertest.cpp             | 2.000000               |
| kinit/kinit.win.cpp   | 2.000000               |
| kinit/klauncher.cpp   | 2.000000               |
| kjs/tests/testkjs.cpp   | 2.000000               |
| konq-plugins/rellinks/plugin_rellinks.cpp                     | 2.000000               |
| kross/test/main.cpp   | 2.000000               |
| nsplugins/viewer/viewer.cpp                                   | 2.000000               |
| dolphin/src/views/versioncontrol/versioncontrolobserver.cpp   | 1.000000               |
| kde3support/kdeui/k3listview.cpp                              | 1.000000               |
| kde3support/kunittest/modrunner.cpp                           | 1.000000               |
| kde3support/tests/itemcontainertest.cpp                       | 1.000000               |
| kde3support/tests/k3listviewtest.cpp                          | 1.000000               |
| kde3support/tests/k3wizardtest.cpp                            | 1.000000               |
| kde3support/tests/kaboutdialogtest.cpp                        | 1.000000               |
| kde3support/tests/kactivelabeltest.cpp                        | 1.000000               |

|  |          |
|--|----------|
| kde3support/tests/kdocktest.cpp                  | 1.000000 |
| kde3support/tests/kdockwidgetdemo.cpp            | 1.000000 |
| kdecore/auth/policy-gen/policy-gen.cpp           | 1.000000 |
| kdecore/kconfig_compiler/kconfig_compiler.cpp    | 1.000000 |
| kdecore/kde-config.cpp                           | 1.000000 |
| kdecore/kernel/kstandarddirs.cpp                 | 1.000000 |
| kdecore/sonnet/tests/test.cpp                    | 1.000000 |
| kdecore/sonnet/tests/test_config.cpp             | 1.000000 |
| kdecore/sonnet/tests/test_dialog.cpp             | 1.000000 |
| kdecore/tests/dbuscalltest.cpp                   | 1.000000 |
| kdecore/tests/globalcleantest.cpp                | 1.000000 |
| kdecore/tests/kconfigafterkglobaltest1.cpp       | 1.000000 |
| kdecore/tests/kconfigafterkglobaltest2.cpp       | 1.000000 |
| kdecore/tests/kdebug_qcoreapptest.cpp            | 1.000000 |
| kdecore/tests/kdirwatchtest.cpp                  | 1.000000 |
| kdecore/tests/kdirwatchtest_gui.cpp              | 1.000000 |
| kdecore/tests/kglobalstatictest.cpp              | 1.000000 |
| kdecore/tests/kmimetype_nomimetypes.cpp          | 1.000000 |
| kdecore/tests/ksharedconfig_in_global_object.cpp | 1.000000 |
| kdecore/tests/ksortablelisttest.cpp              | 1.000000 |
| kdecore/tests/ktartest.cpp                       | 1.000000 |
| kdecore/tests/kziptest.cpp                       | 1.000000 |
| kded/kded.cpp                                    | 1.000000 |
| kdepasswd/kdepasswd.cpp                          | 1.000000 |
| kdeui/dialogs/kbugreport.cpp                     | 1.000000 |

|  |          |
|--|----------|
| kdeui/kernel/kuniqueapplication.cpp            | 1.000000 |
| kdeui/sonnet/tests/backgroundtest.cpp          | 1.000000 |
| kdeui/sonnet/tests/test_config.cpp             | 1.000000 |
| kdeui/sonnet/tests/test_configdialog.cpp       | 1.000000 |
| kdeui/sonnet/tests/test_dialog.cpp             | 1.000000 |
| kdeui/sonnet/tests/test_dictionarycombobox.cpp | 1.000000 |
| kdeui/sonnet/tests/test_highlighter.cpp        | 1.000000 |
| kdeui/tests/bihash/functionalitytest.cpp       | 1.000000 |
| kdeui/tests/fixx11h_test.cpp                   | 1.000000 |
| kdeui/tests/fixx11h_test2.cpp                  | 1.000000 |
| kdeui/tests/kanimatedbuttontest.cpp            | 1.000000 |
| kdeui/tests/kassistantdialogtest.cpp           | 1.000000 |
| kdeui/tests/kbugreporttest.cpp                 | 1.000000 |
| kdeui/tests/kcategorizedviewtest.cpp           | 1.000000 |
| kdeui/tests/kcharselecttest.cpp                | 1.000000 |
| kdeui/tests/kcodeactiontest.cpp                | 1.000000 |
| kdeui/tests/kcolorcollectiontest.cpp           | 1.000000 |
| kdeui/tests/kcolorcombotest.cpp                | 1.000000 |
| kdeui/tests/kcompletionuitest.cpp              | 1.000000 |
| kdeui/tests/kconfig_compiler/test9main.cpp     | 1.000000 |
| kdeui/tests/kdatepicktest.cpp                  | 1.000000 |
| kdeui/tests/kdatetabletest.cpp                 | 1.000000 |
| kdeui/tests/kdatetimewidgettest.cpp            | 1.000000 |
| kdeui/tests/kdatewidgettest.cpp                | 1.000000 |
| kdeui/tests/kdebugtest_gui.cpp                 | 1.000000 |

|   |          |
|---|----------|
| kdeui/tests/kdialogbuttonboxtest.cpp            | 1.000000 |
| kdeui/tests/kdialogtest.cpp                     | 1.000000 |
| kdeui/tests/keditlistboxtest.cpp                | 1.000000 |
| kdeui/tests/kfontdialogtest.cpp                 | 1.000000 |
| kdeui/tests/kglobalsettingsclient.cpp           | 1.000000 |
| kdeui/tests/khboxtest.cpp                       | 1.000000 |
| kdeui/tests/kiconeffecttest.cpp                 | 1.000000 |
| kdeui/tests/klanguagebuttontest.cpp             | 1.000000 |
| kdeui/tests/kmainwindowtest.cpp                 | 1.000000 |
| kdeui/tests/kmessagetest.cpp                    | 1.000000 |
| kdeui/tests/kmodifierkeyinfotest.cpp            | 1.000000 |
| kdeui/tests/knuminputtest.cpp                   | 1.000000 |
| kdeui/tests/kpagedialogtest.cpp                 | 1.000000 |
| kdeui/tests/kpagewidgettest.cpp                 | 1.000000 |
| kdeui/tests/kpassivepoptest.cpp                 | 1.000000 |
| kdeui/tests/kpixmapregionselectordialogtest.cpp | 1.000000 |
| kdeui/tests/kpoptest.cpp                        | 1.000000 |
| kdeui/tests/kreplacetest.cpp                    | 1.000000 |
| kdeui/tests/krichtexteditor/main.cpp            | 1.000000 |
| kdeui/tests/krulertest.cpp                      | 1.000000 |
| kdeui/tests/kselectactiontest.cpp               | 1.000000 |
| kdeui/tests/kseparatorstest.cpp                 | 1.000000 |
| kdeui/tests/ksqueezedtextlabeltest.cpp          | 1.000000 |
| kdeui/tests/kstatusnotifieritemtest.cpp         | 1.000000 |
| kdeui/tests/kstraytest.cpp                      | 1.000000 |

|  |          |
|--|----------|
| kdeui/tests/ktabwidgettest.cpp                     | 1.000000 |
| kdeui/tests/ktextbrowsertest.cpp                   | 1.000000 |
| kdeui/tests/ktextedittest.cpp                      | 1.000000 |
| kdeui/tests/ktitlewidgettest.cpp                   | 1.000000 |
| kdeui/tests/ktoolbarlabelactiontest.cpp            | 1.000000 |
| kdeui/tests/ktoolbarwidgettest.cpp                 | 1.000000 |
| kdeui/tests/kwidgetitemdelegatetest.cpp            | 1.000000 |
| kdeui/tests/kwindowtest.cpp                        | 1.000000 |
| kdeui/tests/kxmlguitest.cpp                        | 1.000000 |
| kdeui/tests/kxmlguiwindowtest.cpp                  | 1.000000 |
| kdeui/tests/proxymodeltestapp/main.cpp             | 1.000000 |
| kdeui/tests/proxymodeltestsuite/templates/main.cpp | 1.000000 |
| kdeui/util/kcursor.cpp                             | 1.000000 |
| kdewidgets/makekdewidgets.cpp                      | 1.000000 |
| kdialoq/kdialoq.cpp                                | 1.000000 |
| kdoctools/docbook110nhelper.cpp                    | 1.000000 |
| kdoctools/genshortcutents.cpp                      | 1.000000 |
| kdoctools/meinproc_simple.cpp                      | 1.000000 |
| keditbookmarks/kbookmarkmerger.cpp                 | 1.000000 |
| kfile/tests/kdirselectdialogtest.cpp               | 1.000000 |
| kfile/tests/kfiletreeviewtest.cpp                  | 1.000000 |
| kfind/kftabdlg.cpp                                 | 1.000000 |
| kfind/main.cpp                                     | 1.000000 |
| khtml/ecma/testecma.cpp                            | 1.000000 |
| khtml/java/tests/testkjavaappletserver.cpp         | 1.000000 |

|  |          |
|--|----------|
| khtml/khtmlimage.cpp                     | 1.000000 |
| khtml/kmultipart/kmultipart.cpp          | 1.000000 |
| khtml/xpath/interpreter_tester.cpp       | 1.000000 |
| kinit/kinit.cpp                          | 1.000000 |
| kinit/kioslave.cpp                       | 1.000000 |
| kinit/kwrapper_win.cpp                   | 1.000000 |
| kinit/lnusertemp.c                       | 1.000000 |
| kinit/start_kdeinit.c                    | 1.000000 |
| kinit/start_kdeinit_wrapper.c            | 1.000000 |
| kinit/wrapper.c                          | 1.000000 |
| gio/kio/kbuildsycocapprogressdialog.cpp  | 1.000000 |
| gio/kio/kdirlister.cpp                   | 1.000000 |
| gio/kssl/kssl/cert_extract.c             | 1.000000 |
| gio/misc/kmailservice.cpp                | 1.000000 |
| gio/misc/ksendbugmail/main.cpp           | 1.000000 |
| gio/misc/ktelnet-service.cpp             | 1.000000 |
| gio/tests/kdbus-service-starter-test.cpp | 1.000000 |
| gio/tests/kdirmodel-test-gui.cpp         | 1.000000 |
| gio/tests/kfiledialog-test.cpp           | 1.000000 |
| gio/tests/kfstest.cpp                    | 1.000000 |
| gio/tests/kicondialog-test.cpp           | 1.000000 |
| gio/tests/kionetrctest.cpp               | 1.000000 |
| gio/tests/kiopassdlg-test.cpp            | 1.000000 |
| gio/tests/kioslavetest.cpp               | 1.000000 |
| gio/tests/kmfitest.cpp                   | 1.000000 |

|  |          |
|--|----------|
| kio/tests/kmimetypechoosertest_gui.cpp                 | 1.000000 |
| kio/tests/kopenwithtest.cpp                            | 1.000000 |
| kio/tests/kpropsdlgtest.cpp                            | 1.000000 |
| kio/tests/kprotocolinfotest.cpp                        | 1.000000 |
| kio/tests/kruntest.cpp                                 | 1.000000 |
| kio/tests/kscantest.cpp                                | 1.000000 |
| kio/tests/ksslkeygentest.cpp                           | 1.000000 |
| kio/tests/kurlcompletiontest.cpp                       | 1.000000 |
| kio/tests/kurlrequestertest.cpp                        | 1.000000 |
| kio/tests/netaccesstest.cpp                            | 1.000000 |
| kio/tests/previewtest.cpp                              | 1.000000 |
| kio/tests/speed.cpp                                    | 1.000000 |
| kjs/kjs.cpp  | 1.000000 |
| kjsebed/examples/kjsconsole/main.cpp                   | 1.000000 |
| kjsebed/kjscmd/kjscmd.cpp                              | 1.000000 |
| knewstuff/tests/kdxspreview.cpp                        | 1.000000 |
| knewstuff/tests/knewstuff2_cache.cpp                   | 1.000000 |
| knewstuff/tests/knewstuff2_download.cpp                | 1.000000 |
| knewstuff/tests/knewstuff2_standard.cpp                | 1.000000 |
| knewstuff/tests/knewstuff2_test.cpp                    | 1.000000 |
| knotify/tests/main.cpp                                 | 1.000000 |
| konq-plugins/searchbar/searchbar.cpp                   | 1.000000 |
| konq-plugins/shellcmdplugin/kshellcmdexecutor.cpp      | 1.000000 |
| konqueror/settings/konq/tests/kcustommeneditortest.cpp | 1.000000 |
| konqueror/src/konqhistoryview.cpp                      | 1.000000 |



|  |          |
|--|----------|
| konqueror/src/konqview.cpp                                 | 1.000000 |
| konqueror/src/tests/centralwidget.cpp                      | 1.000000 |
| konqueror/src/tests/konqviewmgrtest.cpp                    | 1.000000 |
| kparts/tests/normalktm.cpp                                 | 1.000000 |
| kparts/tests/openorsavequestion.cpp                        | 1.000000 |
| kparts/tests/partviewer.cpp                                | 1.000000 |
| kparts/tests/testmainwindow.cpp                            | 1.000000 |
| kross/console/main.cpp                                     | 1.000000 |
| kross/qts/main.cpp   | 1.000000 |
| kutils/kemoticons/tests/main.cpp                           | 1.000000 |
| nepomuk/rcgen/rcgen.cpp                                    | 1.000000 |
| nepomuk/test/ratingpaintertestapp.cpp                      | 1.000000 |
| nepomuk/test/tagwidgetapp.cpp                              | 1.000000 |
| nepomuk/ui/kedittagsdialog.cpp                             | 1.000000 |
| nsplugins/pluginscan.cpp                                   | 1.000000 |
| plasma/private/qtjolie-branch/tests/calculatorservice.cpp  | 1.000000 |
| plasma/private/qtjolie-branch/tests/testprinter.cpp        | 1.000000 |
| plasma/private/qtjolie-branch/tests/trivialyahooclient.cpp | 1.000000 |
| plasma/private/qtjolie-branch/tests/visiondriver.cpp       | 1.000000 |
| plasma/widgets/scrollwidget.cpp                            | 1.000000 |
| solid/tests/networkingclient.cpp                           | 1.000000 |
| solid/tests/solidnettestdbusservice.cpp                    | 1.000000 |
| threadweaver/examples/jobs/jobs.cpp                        | 1.000000 |
| threadweaver/examples/qprocessjob/main.cpp                 | 1.000000 |
| threadweaver/examples/smiv/smiv.cpp                        | 1.000000 |

|   |          |
|---|----------|
| threadweaver/tests/deletetest.cpp             | 1.000000 |
| dolphin/src/tests/kfileitemmodelbenchmark.cpp | 0.000000 |
| dolphin/src/tests/kitemlistcontrollertest.cpp | 0.000000 |
| experimental/libkdeclarative/test/test.cpp    | 0.000000 |
| kde3support/kdeui/k3iconviewsearchline.cpp    | 0.000000 |
| kde3support/kdeui/k3listviewsearchline.cpp    | 0.000000 |
| kde3support/kdeui/k3passworddialog.cpp        | 0.000000 |
| kde3support/kio/kfile/k3fileview.cpp          | 0.000000 |
| kde3support/tests/k3processtest.cpp           | 0.000000 |
| kde3support/tests/k3prociotest.cpp            | 0.000000 |
| kde3support/tests/kdockwidgettest.cpp         | 0.000000 |
| kdecore/config/kconfig.cpp                    | 0.000000 |
| kdecore/config/kdesktopfile.cpp               | 0.000000 |
| kdecore/io/kdebug.cpp                         | 0.000000 |
| kdecore/io/kmountpoint.cpp                    | 0.000000 |
| kdecore/io/kurl.cpp                           | 0.000000 |
| kdecore/kernel/kaboutdata.cpp                 | 0.000000 |
| kdecore/kernel/kcmdlineargs.cpp               | 0.000000 |
| kdecore/kernel/kcomponentdata.cpp             | 0.000000 |
| kdecore/kernel/ktoolinvocation.cpp            | 0.000000 |
| kdecore/kernel/ktoolinvocation_win.cpp        | 0.000000 |
| kdecore/kernel/ktoolinvocation_x11.cpp        | 0.000000 |
| kdecore/localization/klocale.cpp              | 0.000000 |
| kdecore/localization/klocalizedstring.cpp     | 0.000000 |
| kdecore/network/ktcpsocket.cpp                | 0.000000 |

|  |          |
|--|----------|
| kdecore/services/kmimetype.cpp           | 0.000000 |
| kdecore/services/kmimetypefactory.cpp    | 0.000000 |
| kdecore/services/kmimetyperepository.cpp | 0.000000 |
| kdecore/services/kplugininfo.cpp         | 0.000000 |
| kdecore/services/kservice.cpp            | 0.000000 |
| kdecore/services/kservicefactory.cpp     | 0.000000 |
| kdecore/services/kserVICETypetrader.cpp  | 0.000000 |
| kdecore/sycoca/kprotocolinfo.cpp         | 0.000000 |
| kdecore/sycoca/kprotocolinfofactory.cpp  | 0.000000 |
| kdecore/sycoca/ksycoca.cpp               | 0.000000 |
| kdecore/sycoca/ksycocadiCT.cpp           | 0.000000 |
| kdecore/tests/cplusplustest.cpp          | 0.000000 |
| kdecore/tests/kcmdlineargstest.cpp       | 0.000000 |
| kdecore/tests/kconfigtest.cpp            | 0.000000 |
| kdecore/tests/klockfile_testlock.cpp     | 0.000000 |
| kdecore/tests/kmdcodectest.cpp           | 0.000000 |
| kdecore/tests/kmemtest.cpp               | 0.000000 |
| kdecore/tests/kmimetyPETest.cpp          | 0.000000 |
| kdecore/tests/kserVICETest.cpp           | 0.000000 |
| kdecore/tests/ksycocadiCTtest.cpp        | 0.000000 |
| kdecore/tests/kurlmimetest.cpp           | 0.000000 |
| kdecore/tests/startserviceby.cpp         | 0.000000 |
| kdecore/util/klibloader.cpp              | 0.000000 |
| kdecore/util/kuser_unix.cpp              | 0.000000 |
| kdecore/util/kuser_win.cpp               | 0.000000 |

|   |          |
|---|----------|
| kdecore/util/kuser_wince.cpp                  | 0.000000 |
| kdepasswd/passwd.cpp                          | 0.000000 |
| kdeui/actions/kactioncollection.cpp           | 0.000000 |
| kdeui/actions/kcodeaction.cpp                 | 0.000000 |
| kdeui/dialogs/kdialog.cpp                     | 0.000000 |
| kdeui/dialogs/klinkdialog.cpp                 | 0.000000 |
| kdeui/dialogs/kmessagebox.cpp                 | 0.000000 |
| kdeui/dialogs/kpassworddialog.cpp             | 0.000000 |
| kdeui/icons/kicontheme.cpp                    | 0.000000 |
| kdeui/itemviews/kfilterproxysearchline.cpp    | 0.000000 |
| kdeui/itemviews/klistwidgetsearchline.cpp     | 0.000000 |
| kdeui/notifications/kssystemtrayicon.cpp      | 0.000000 |
| kdeui/shortcuts/kacceleratomanager.cpp        | 0.000000 |
| kdeui/tests/kactionselectortest.cpp           | 0.000000 |
| kdeui/tests/kapptest.cpp                      | 0.000000 |
| kdeui/tests/kcolordlgttest.cpp                | 0.000000 |
| kdeui/tests/kcolorutilsdemo.cpp               | 0.000000 |
| kdeui/tests/kiconloader_unittest.cpp          | 0.000000 |
| kdeui/tests/kiconloadertest.cpp               | 0.000000 |
| kdeui/tests/kinputdialogtest.cpp              | 0.000000 |
| kdeui/tests/kmainwindowrestorettest.cpp       | 0.000000 |
| kdeui/tests/kmessageboxtest.cpp               | 0.000000 |
| kdeui/tests/kmessagewidgettest.cpp            | 0.000000 |
| kdeui/tests/knewpassworddialogtest.cpp        | 0.000000 |
| kdeui/tests/knotificationrestrictionstest.cpp | 0.000000 |

|   |          |
|---|----------|
| kdeui/tests/kpassworddialogtest.cpp           | 0.000000 |
| kdeui/tests/kprogressdialogtest.cpp           | 0.000000 |
| kdeui/tests/kstatusbartest.cpp                | 0.000000 |
| kdeui/tests/ktabwidget_unittest.cpp           | 0.000000 |
| kdeui/tests/kuniqueapptest.cpp                | 0.000000 |
| kdeui/tests/testplot_main.cpp                 | 0.000000 |
| kdeui/tests/testqtargs.cpp                    | 0.000000 |
| kdeui/util/kcrash.cpp                         | 0.000000 |
| kdeui/util/kguiitem.cpp                       | 0.000000 |
| kdeui/util/kpassivepopup.cpp                  | 0.000000 |
| kdeui/widgets/kcombobox.cpp                   | 0.000000 |
| kdeui/widgets/kdatetable.cpp                  | 0.000000 |
| kdeui/widgets/keditlistwidget.cpp             | 0.000000 |
| kdeui/widgets/klanguagebutton.cpp             | 0.000000 |
| kdeui/widgets/klineedit.cpp                   | 0.000000 |
| kdeui/widgets/kpixmapregionselectorwidget.cpp | 0.000000 |
| kdeui/widgets/ktextedit.cpp                   | 0.000000 |
| kdeui/widgets/ktimezonewidget.cpp             | 0.000000 |
| kdeui/windowmanagement/kwindowinfo_mac.cpp    | 0.000000 |
| kdeui/windowmanagement/kwindowinfo_win.cpp    | 0.000000 |
| kdeui/windowmanagement/kwindowinfo_x11.cpp    | 0.000000 |
| kdeui/windowmanagement/kwindowsystem_mac.cpp  | 0.000000 |
| kdeui/windowmanagement/kwindowsystem_qws.cpp  | 0.000000 |
| kdeui/windowmanagement/kwindowsystem_win.cpp  | 0.000000 |
| kdeui/windowmanagement/kwindowsystem_x11.cpp  | 0.000000 |

|   |          |
|---|----------|
| kdeui/xmlgui/kxmlguiwindow.cpp            | 0.000000 |
| kdoctools/xslt_kde.cpp                    | 0.000000 |
| keditbookmarks/kbookmarkmodel/treitem.cpp | 0.000000 |
| keditbookmarks/kebsearchline.cpp          | 0.000000 |
| khtml/java/kjavadoanloader.cpp            | 0.000000 |
| khtml/test_regression.cpp                 | 0.000000 |
| khtml/test_regression_gui_main.cpp        | 0.000000 |
| khtml/testkhtml.cpp                       | 0.000000 |
| khtml/tests/testkhtmlnew.cpp              | 0.000000 |
| khtml/ui/findbar/khtmlfind.cpp            | 0.000000 |
| kinit/klauncher_cmds.cpp                  | 0.000000 |
| kinit/proctitle.cpp                       | 0.000000 |
| kinit/tests/klaunchertest.cpp             | 0.000000 |
| kio/bookmarks/kbookmark.cc                | 0.000000 |
| kio/bookmarks/kbookmarkmanager.cc         | 0.000000 |
| kio/kfile/kdiskfreospaceinfo.cpp          | 0.000000 |
| kio/kfile/kfiledialog.cpp                 | 0.000000 |
| kio/kfile/kfilemetadaprovider.cpp         | 0.000000 |
| kio/kfile/kfilemetadareaderprocess.cpp    | 0.000000 |
| kio/kfile/kopenwithdialog.cpp             | 0.000000 |
| kio/kfile/kurlrequester.cpp               | 0.000000 |
| kio/kfile/kurlrequesterdialog.cpp         | 0.000000 |
| kio/kio/kautomount.cpp                    | 0.000000 |
| kio/kio/kdesktopfileactions.cpp           | 0.000000 |
| kio/kio/kfileitem.cpp                     | 0.000000 |

|  |          |
|--|----------|
| kio/kio/kmimetypechooser.cpp                             | 0.000000 |
| kio/kio/krun.cpp   | 0.000000 |
| kio/tests/getalltest.cpp                                 | 0.000000 |
| kio/tests/kacctest.cpp                                   | 0.000000 |
| kio/tests/kdcopcheck.cpp                                 | 0.000000 |
| kio/tests/kdirlistertest_gui.cpp                         | 0.000000 |
| kio/tests/ksycocatest.cpp                                | 0.000000 |
| kio/tests/ksycocaupdatetest.cpp                          | 0.000000 |
| kioslave/http/kcookiejar/main.cpp                        | 0.000000 |
| kjs/array_instance.cpp                                   | 0.000000 |
| kjs/bytecode/generator/driver.cpp                        | 0.000000 |
| kjs/list.cpp   | 0.000000 |
| kjsembed/kjsembed/dom.cpp                                | 0.000000 |
| kjsembed/kjsembed/qobject_binding.cpp                    | 0.000000 |
| kjsembed/kjsembed/svg_binding.cpp                        | 0.000000 |
| knewstuff/tests/knewstuffentrytest.cpp                   | 0.000000 |
| konq-plugins/adblock/adblock.cpp                         | 0.000000 |
| konq-plugins/dirfilter/dirfilterplugin.cpp               | 0.000000 |
| konq-plugins/fsview/main.cpp                             | 0.000000 |
| konq-plugins/validators/plugin_validators.cpp            | 0.000000 |
| konqueror/about/konq_aboutpage.cc                        | 0.000000 |
| konqueror/settings/ebrowsing/main.cpp                    | 0.000000 |
| konqueror/settings/kio/kproxydlg.cpp                     | 0.000000 |
| konqueror/sidebar/sidebar_part.cpp                       | 0.000000 |
| konqueror/sidebar/trees/konq_sidebartreetoplevelitem.cpp | 0.000000 |

|   |          |
|---|----------|
| konqueror/sidebar/web_module/web_module.cpp | 0.000000 |
| konqueror/src/konqextensionmanager.cpp      | 0.000000 |
| konqueror/src/konqframe.cpp                 | 0.000000 |
| konqueror/src/konqframevisitor.cpp          | 0.000000 |
| konqueror/src/konqguiclients.cpp            | 0.000000 |
| konqueror/src/konqpixmapprovider.cpp        | 0.000000 |
| konqueror/src/konqrmbeventfilter.cpp        | 0.000000 |
| konqueror/src/konqrun.cpp                   | 0.000000 |
| konqueror/src/konqtabs.cpp                  | 0.000000 |
| konqueror/src/konqviewadaptor.cpp           | 0.000000 |
| konqueror/src/tests/konqviewtest.cpp        | 0.000000 |
| kross/ui/plugin.cpp                         | 0.000000 |
| kutils/kpluginsselector.cpp                 | 0.000000 |
| nsplugins/test/testnsplugin.cpp             | 0.000000 |
| nsplugins/viewer/qxteventloop.cpp           | 0.000000 |
| plasma/packagemetadata.cpp                  | 0.000000 |
| plasma/packagestructure.cpp                 | 0.000000 |
| plasma/widgets/tabbar.cpp                   | 0.000000 |

ΠΙΝΑΚΑΣ Β'.5: Αποτελέσματα Big File Level για το Konqueror για την Αναφορά Σφάλματος υπάριθμόν 325724



## Παράρτημα Γ'

# Kopete Sample Results

167

Στους Επόμενους Πίνακες Παρουσιάζονται τα αποτελέσματα για το σύστημα Kopete.

| File Name  | Connectivity To Clique |
|--|------------------------|
| protocols/oscar/liboscar/tests/buffertest.cpp            | 1                      |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp         | 5                      |
| protocols/oscar/liboscar/tasks/profiletask.cpp           | 2                      |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp   | 4                      |
| protocols/oscar/liboscar/tasks/logintask.cpp             | 2                      |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp | 3                      |
| protocols/oscar/liboscar/tasks/icqtask.cpp               | 1                      |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp      | 2                      |
| protocols/oscar/liboscar/flapprotocol.cpp                | 1                      |

|   |    |
|---|----|
| protocols/oscar/liboscar/connection.cpp                               | 1  |
| protocols/oscar/liboscar/buffer.cpp                                   | 50 |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc   | 1  |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc           | 2  |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc       | 1  |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc     | 5  |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc      | 6  |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc        | 1  |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                 | 3  |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc       | 2  |
| protocols/jabber/libjingle/talk/p2p/base/portallocatorsessionproxy.cc | 2  |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc             | 8  |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                      | 4  |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc       | 4  |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                   | 0  |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc        | 2  |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc               | 1  |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp              | 5  |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp           | 2  |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                     | 0  |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp        | 3  |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp              | 0  |
| plugins/otr/privkeypopup.cpp  | 2  |
| plugins/otr/otrpreferences.cpp  | 0  |
| plugins/otr/otrplugin.cpp   | 0  |

|   |   |
|---|---|
| plugins/otr/otrlconfinterface.cpp         | 0 |
| plugins/otr/otrlchatinterface.cpp         | 5 |
| plugins/otr/otrguiclient.cpp              | 0 |
| plugins/otr/authenticationwizard.cpp      | 2 |
| libkopete/ui/kopetelistviewsearchline.cpp | 2 |
| kopete/config/avdevice/avdeviceconfig.cpp | 1 |

ΠΙΝΑΚΑΣ Γ'.1: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 273070

| File Name   | Connectivity To Clique |
|---|------------------------|
| protocols/yahoo/yahooverifyaccount.cpp            | 1                      |
| protocols/yahoo/yahoocontact.cpp                  | 1                      |
| protocols/yahoo/yahooconferencemessagemanager.cpp | 1                      |
| protocols/yahoo/yahooaccount.cpp                  | 0                      |
| protocols/yahoo/ui/yahoouserinfodialog.cpp        | 2                      |
| protocols/yahoo/ui/yahooinvitelistimpl.cpp        | 2                      |
| protocols/wlm/wlmtransfermanager.cpp              | 2                      |
| protocols/wlm/wlmsocket.cpp                       | 1                      |
| protocols/wlm/wlmprotocol.cpp                     | 6                      |
| protocols/wlm/wlmcontact.cpp                      | 4                      |
| protocols/wlm/wlmchatsession.cpp                  | 4                      |
| protocols/wlm/wlmchatmanager.cpp                  | 4                      |
| protocols/wlm/wlmaccount.cpp                      | 10                     |
| protocols/wlm/ui/wlmeditaccountwidget.cpp         | 1                      |
| protocols/oscar/oscarcontact.cpp                  | 10                     |

|   |    |
|---|----|
| protocols/oscar/oscaraccount.cpp                                    | 14 |
| protocols/oscar/liboscar/tests/buffertest.cpp                       | 1  |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp                    | 5  |
| protocols/oscar/liboscar/tasks/profiletask.cpp                      | 2  |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp              | 4  |
| protocols/oscar/liboscar/tasks/logintask.cpp                        | 2  |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp            | 3  |
| protocols/oscar/liboscar/tasks/icqtask.cpp                          | 1  |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp                 | 2  |
| protocols/oscar/liboscar/flapprotocol.cpp                           | 1  |
| protocols/oscar/liboscar/connection.cpp                             | 1  |
| protocols/oscar/liboscar/buffer.cpp                                 | 50 |
| protocols/jabber/ui/dlgjabberchatroomslist.cpp                      | 0  |
| protocols/jabber/ui/dlgjabberchatjoin.cpp                           | 0  |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc | 1  |
| protocols/jabber/libjingle/talk/session/phone/mediasessionclient.cc | 5  |
| protocols/jabber/libjingle/talk/session/phone/mediasession.cc       | 3  |
| protocols/jabber/libjingle/talk/session/phone/channel.cc            | 4  |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc         | 2  |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc     | 1  |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc   | 5  |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc    | 6  |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc      | 1  |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc               | 3  |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc     | 2  |

|  |   |
|--|---|
| protocols/jabber/libjingle/talk/p2p/base/portallocatorsessionproxy.cc      | 2 |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc                  | 8 |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                           | 4 |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc            | 4 |
| protocols/jabber/libjingle/talk/examples/peerconnection/client/main_wnd.cc | 2 |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                        | 0 |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc             | 2 |
| protocols/jabber/libjingle/talk/base/taskrunner.cc                         | 1 |
| protocols/jabber/libjingle/talk/base/taskparent.cc                         | 2 |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc                    | 1 |
| protocols/jabber/libjingle/talk/app/webrtc/webrtcspd.cc                    | 4 |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp                   | 5 |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp                | 2 |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                          | 0 |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp             | 3 |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp                   | 0 |
| protocols/jabber/libiris/src/irisnet/appledns/appledns.cpp                 | 0 |
| plugins/otr/privkeypopup.cpp   | 2 |
| plugins/otr/otrpreferences.cpp   | 0 |
| plugins/otr/otrplugin.cpp  | 0 |
| plugins/otr/otrlconfinterface.cpp  | 0 |
| plugins/otr/otrlchatinterface.cpp  | 5 |
| plugins/otr/otrguiclient.cpp   | 0 |
| plugins/otr/authenticationwizard.cpp                                       | 2 |
| libkopete/ui/kopetelistviewsearchline.cpp                                  | 2 |

|   |          |
|---|----------|
| libkopete/private/kopeteviewmanager.cpp   | 0        |
| libkopete/kopeteonlinestatus.cpp          | 18       |
| <b>libkopete/kopetemetact.cpp</b>         | <b>0</b> |
| kopete/config/avdevice/avdeviceconfig.cpp | 1        |
| kopete/chatwindow/kopetechatwindow.cpp    | 3        |
| kopete/chatwindow/chatview.cpp            | 4        |

ΠΙΝΑΚΑΣ Γ'.2: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 277606

| File Name   | Connectivity To Clique |
|---|------------------------|
| protocols/yahoo/yahooverifyaccount.cpp            | 1                      |
| protocols/yahoo/yahoocontact.cpp                  | 1                      |
| protocols/yahoo/yahooconferencemessagemanager.cpp | 1                      |
| protocols/yahoo/yahooaccount.cpp                  | 0                      |
| protocols/yahoo/ui/yahoouserinfodialog.cpp        | 2                      |
| protocols/yahoo/ui/yahooinvitelistimpl.cpp        | 2                      |
| protocols/wlm/wlmtransfermanager.cpp              | 2                      |
| protocols/wlm/wlmsocket.cpp                       | 1                      |
| protocols/wlm/wlmprotocol.cpp                     | 6                      |
| protocols/wlm/wlmcontact.cpp                      | 4                      |
| protocols/wlm/wlmchatsession.cpp                  | 4                      |
| protocols/wlm/wlmchatmanager.cpp                  | 4                      |
| protocols/wlm/wlmaccount.cpp                      | 10                     |
| protocols/wlm/ui/wlmeditaccountwidget.cpp         | 1                      |
| protocols/oscar/oscarcontact.cpp                  | 10                     |

|   |          |
|---|----------|
| protocols/oscar/oscaraccount.cpp                                    | 14       |
| protocols/oscar/liboscar/tests/buffertest.cpp                       | 1        |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp                    | 5        |
| protocols/oscar/liboscar/tasks/profiletask.cpp                      | 2        |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp              | 4        |
| protocols/oscar/liboscar/tasks/logintask.cpp                        | 2        |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp            | 3        |
| protocols/oscar/liboscar/tasks/icqtask.cpp                          | 1        |
| <b>protocols/oscar/liboscar/tasks/filetransfertask.cpp</b>          | <b>2</b> |
| protocols/oscar/liboscar/flapprotocol.cpp                           | 1        |
| protocols/oscar/liboscar/connection.cpp                             | 1        |
| protocols/oscar/liboscar/buffer.cpp                                 | 50       |
| protocols/oscar/icq/ui/icquserinfowidget.cpp                        | 1        |
| protocols/oscar/icq/ui/icqsearchdialog.cpp                          | 3        |
| protocols/jabber/ui/dlgjabberchatroomslist.cpp                      | 0        |
| protocols/jabber/ui/dlgjabberchatjoin.cpp                           | 0        |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc | 1        |
| protocols/jabber/libjingle/talk/session/phone/mediasessionclient.cc | 5        |
| protocols/jabber/libjingle/talk/session/phone/mediasession.cc       | 3        |
| protocols/jabber/libjingle/talk/session/phone/channel.cc            | 4        |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc         | 2        |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc     | 1        |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc   | 5        |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc    | 6        |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc      | 1        |

|  |   |
|--|---|
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                      | 3 |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc            | 2 |
| protocols/jabber/libjingle/talk/p2p/base/portallocatorproxy.cc             | 2 |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc                  | 8 |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                           | 4 |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc            | 4 |
| protocols/jabber/libjingle/talk/examples/peerconnection/client/main_wnd.cc | 2 |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                        | 0 |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc             | 2 |
| protocols/jabber/libjingle/talk/base/taskrunner.cc                         | 1 |
| protocols/jabber/libjingle/talk/base/taskparent.cc                         | 2 |
| protocols/jabber/libjingle/talk/base/flags.cc                              | 2 |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc                    | 1 |
| protocols/jabber/libjingle/talk/app/webrtc/webrtcscdp.cc                   | 4 |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp                   | 5 |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp                | 2 |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                          | 0 |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp             | 3 |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp                   | 0 |
| protocols/jabber/libiris/src/irisnet/appledns/appledns.cpp                 | 0 |
| protocols/irc/ui/ircdialogs.cpp  | 2 |
| plugins/otr/privkeypopup.cpp   | 2 |
| plugins/otr/otrpreferences.cpp   | 0 |
| plugins/otr/otrplugin.cpp  | 0 |
| plugins/otr/otrlconfinterface.cpp  | 0 |



|   |    |
|---|----|
| plugins/otr/otrchatinterface.cpp          | 5  |
| plugins/otr/otrclient.cpp                 | 0  |
| plugins/otr/authenticationwizard.cpp      | 2  |
| libkopete/ui/kopetelistviewsearchline.cpp | 2  |
| libkopete/private/kopeteviewmanager.cpp   | 0  |
| libkopete/kopeteonlinestatus.cpp          | 18 |
| libkopete/kopetemetaccontact.cpp          | 0  |
| kopete/config/avdevice/avdeviceconfig.cpp | 1  |
| kopete/chatwindow/kopetechatwindow.cpp    | 3  |
| kopete/chatwindow/chatview.cpp            | 4  |

ΠΙΝΑΚΑΣ Γ'.3: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 293155

| File Name   | Connectivity To Clique |
|---|------------------------|
| protocols/yahoo/yahooverifyaccount.cpp            | 1                      |
| protocols/yahoo/yahoocontact.cpp                  | 1                      |
| protocols/yahoo/yahooconferencemessagemanager.cpp | 1                      |
| protocols/yahoo/yahooaccount.cpp                  | 0                      |
| protocols/yahoo/ui/yahouserinfodialog.cpp         | 2                      |
| protocols/yahoo/ui/yahoinvitelistimpl.cpp         | 2                      |
| protocols/wlm/wlmtransfermanager.cpp              | 2                      |
| protocols/wlm/wlmsocket.cpp                       | 1                      |
| protocols/wlm/wlmprotocol.cpp                     | 6                      |
| protocols/wlm/wlmcontact.cpp                      | 4                      |
| protocols/wlm/wlmchatsession.cpp                  | 4                      |

|   |    |
|---|----|
| protocols/wlm/wlmchatmanager.cpp                                    | 4  |
| protocols/wlm/wlmaccount.cpp  | 10 |
| protocols/wlm/ui/wlmeditaccountwidget.cpp                           | 1  |
| protocols/oscar/oscarcontact.cpp                                    | 10 |
| protocols/oscar/oscaraccount.cpp                                    | 14 |
| protocols/oscar/liboscar/tests/buffertest.cpp                       | 1  |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp                    | 5  |
| protocols/oscar/liboscar/tasks/profiletask.cpp                      | 2  |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp              | 4  |
| protocols/oscar/liboscar/tasks/logintask.cpp                        | 2  |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp            | 3  |
| protocols/oscar/liboscar/tasks/icqtask.cpp                          | 1  |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp                 | 2  |
| protocols/oscar/liboscar/flapprotocol.cpp                           | 1  |
| protocols/oscar/liboscar/connection.cpp                             | 1  |
| protocols/oscar/liboscar/buffer.cpp                                 | 50 |
| protocols/oscar/icq/ui/icquserinfowidget.cpp                        | 1  |
| protocols/oscar/icq/ui/icqsearchdialog.cpp                          | 3  |
| protocols/jabber/ui/dlgjabberchatroomslist.cpp                      | 0  |
| protocols/jabber/ui/dlgjabberchatjoin.cpp                           | 0  |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc | 1  |
| protocols/jabber/libjingle/talk/session/phone/mediasessionclient.cc | 5  |
| protocols/jabber/libjingle/talk/session/phone/mediasession.cc       | 3  |
| protocols/jabber/libjingle/talk/session/phone/channel.cc            | 4  |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc         | 2  |

|  |   |
|--|---|
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc            | 1 |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc          | 5 |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc           | 6 |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc             | 1 |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                      | 3 |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc            | 2 |
| protocols/jabber/libjingle/talk/p2p/base/portallocatorsessionproxy.cc      | 2 |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc                  | 8 |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                           | 4 |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc            | 4 |
| protocols/jabber/libjingle/talk/examples/peerconnection/client/main_wnd.cc | 2 |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                        | 0 |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc             | 2 |
| protocols/jabber/libjingle/talk/base/taskrunner.cc                         | 1 |
| protocols/jabber/libjingle/talk/base/taskparent.cc                         | 2 |
| protocols/jabber/libjingle/talk/base/flags.cc                              | 2 |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc                    | 1 |
| protocols/jabber/libjingle/talk/app/webrtc/webrtcspd.cc                    | 4 |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp                   | 5 |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp                | 2 |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                          | 0 |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp             | 3 |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp                   | 0 |
| protocols/jabber/libiris/src/irisnet/appledns/appledns.cpp                 | 0 |
| protocols/irc/ui/ircdialogs.cpp  | 2 |

|  |          |
|--|----------|
| plugins/otr/privkeypopup.cpp                       | 2        |
| plugins/otr/otrpreferences.cpp                     | 0        |
| plugins/otr/otrplugin.cpp                          | 0        |
| plugins/otr/otrlconfinterface.cpp                  | 0        |
| plugins/otr/otrlchatinterface.cpp                  | 5        |
| plugins/otr/otrguiclient.cpp                       | 0        |
| plugins/otr/authenticationwizard.cpp               | 2        |
| libkopete/ui/kopetelistviewsearchline.cpp          | 2        |
| libkopete/private/kopeteviewmanager.cpp            | 0        |
| libkopete/kopetetransfermanager.cpp                | 0        |
| libkopete/kopeteonlinestatus.cpp                   | 18       |
| libkopete/kopetemetacontact.cpp                    | 0        |
| <b>kopete/identity/identitystatuswidget.cpp</b>    | <b>0</b> |
| kopete/config/avdevice/avdeviceconfig.cpp          | 1        |
| kopete/chatwindow/kopetechatwindowstylemanager.cpp | 4        |
| kopete/chatwindow/kopetechatwindow.cpp             | 3        |
| kopete/chatwindow/chatview.cpp                     | 2        |
| kopete/chatwindow/chatmessagepart.cpp              | 1        |

ΠΙΝΑΚΑΣ Γ'.4: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 297507

| File Name                                      | Connectivity To Clique |
|--|------------------------|
| protocols/yahoo/yahooverifyaccount.cpp         | 1                      |
| protocols/yahoo/yahoocontact.cpp               | 1                      |
| protocols/yahoo/yahooconferencemessagepart.cpp | 1                      |

|  |    |
|--|----|
| protocols/yahoo/yahooaccount.cpp                         | 0  |
| protocols/yahoo/ui/yahouserinfodialog.cpp                | 2  |
| protocols/yahoo/ui/yahooinvitelistimpl.cpp               | 2  |
| protocols/wlm/wlmtransfermanager.cpp                     | 2  |
| protocols/wlm/wlmsocket.cpp                              | 1  |
| protocols/wlm/wlmprotocol.cpp                            | 6  |
| protocols/wlm/wlmcontact.cpp                             | 4  |
| protocols/wlm/wlmchatsession.cpp                         | 4  |
| protocols/wlm/wlmchatmanager.cpp                         | 4  |
| protocols/wlm/wlmaccount.cpp                             | 10 |
| protocols/wlm/ui/wlmeditaccountwidget.cpp                | 1  |
| protocols/oscar/oscarstatusmanager.cpp                   | 5  |
| protocols/oscar/oscarcontact.cpp                         | 5  |
| protocols/oscar/oscaraccount.cpp                         | 10 |
| protocols/oscar/liboscar/userdetails.cpp                 | 9  |
| protocols/oscar/liboscar/tests/buffertest.cpp            | 1  |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp         | 4  |
| protocols/oscar/liboscar/tasks/profiletask.cpp           | 2  |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp   | 4  |
| protocols/oscar/liboscar/tasks/logintask.cpp             | 2  |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp | 3  |
| protocols/oscar/liboscar/tasks/icqttask.cpp              | 1  |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp      | 2  |
| protocols/oscar/liboscar/flaprotocol.cpp                 | 1  |
| protocols/oscar/liboscar/filetransferhandler.cpp         | 2  |

|   |    |
|---|----|
| protocols/oscar/liboscar/contactmanager.cpp                             | 5  |
| protocols/oscar/liboscar/contact.cpp                                    | 11 |
| protocols/oscar/liboscar/connection.cpp                                 | 1  |
| protocols/oscar/liboscar/buffer.cpp                                     | 50 |
| protocols/oscar/icq/ui/icquserinfowidget.cpp                            | 1  |
| protocols/oscar/icq/ui/icqsearchdialog.cpp                              | 3  |
| protocols/oscar/icq/icqaccount.cpp                                      | 1  |
| protocols/oscar/icq/aimcontact.cpp                                      | 3  |
| protocols/oscar/aim/icqcontact.cpp                                      | 5  |
| protocols/jabber/ui/dlgjabberchatroomslist.cpp                          | 0  |
| protocols/jabber/ui/dlgjabberchatjoin.cpp                               | 0  |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc     | 1  |
| protocols/jabber/libjingle/talk/session/phone/mediasessionclient.cc     | 5  |
| protocols/jabber/libjingle/talk/session/phone/mediasession.cc           | 3  |
| protocols/jabber/libjingle/talk/session/phone/channel.cc                | 4  |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc             | 2  |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc         | 1  |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc       | 5  |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc        | 6  |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc          | 1  |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                   | 3  |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc         | 2  |
| protocols/jabber/libjingle/talk/p2p/base/portallocatororsessionproxy.cc | 2  |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc               | 8  |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                        | 4  |

|  |   |
|--|---|
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc            | 4 |
| protocols/jabber/libjingle/talk/examples/peerconnection/client/main_wnd.cc | 2 |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                        | 0 |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc             | 2 |
| protocols/jabber/libjingle/talk/base/taskrunner.cc                         | 1 |
| protocols/jabber/libjingle/talk/base/taskparent.cc                         | 2 |
| protocols/jabber/libjingle/talk/base/flags.cc                              | 2 |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc                    | 1 |
| protocols/jabber/libjingle/talk/app/webrtc/webrtcspd.cc                    | 4 |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp                   | 5 |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp                | 2 |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                          | 0 |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp             | 3 |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp                   | 0 |
| protocols/jabber/libiris/src/irisnet/appledns/appledns.cpp                 | 0 |
| protocols/irc/ui/ircdialogs.cpp  | 2 |
| plugins/otr/privkeypopup.cpp   | 2 |
| plugins/otr/otrpreferences.cpp   | 0 |
| plugins/otr/otrplugin.cpp  | 0 |
| plugins/otr/otrlconfinterface.cpp  | 0 |
| plugins/otr/otrlchatinterface.cpp  | 5 |
| plugins/otr/otrquiclient.cpp   | 0 |
| plugins/otr/authenticationwizard.cpp                                       | 2 |
| libkopete/ui/kopetelistviewsearchline.cpp                                  | 2 |
| libkopete/private/kopeteviewmanager.cpp                                    | 0 |

|  |    |
|--|----|
| libkopete/kopetetransfermanager.cpp                | 0  |
| libkopete/kopeteonlinestatus.cpp                   | 18 |
| libkopete/kopetemetacontact.cpp                    | 0  |
| kopete/identity/identitystatuswidget.cpp           | 0  |
| kopete/config/avdevice/avdeviceconfig.cpp          | 1  |
| kopete/chatwindow/kopetechatwindowstylemanager.cpp | 4  |
| kopete/chatwindow/kopetechatwindow.cpp             | 3  |
| kopete/chatwindow/chatview.cpp                     | 2  |
| kopete/chatwindow/chatmessagepart.cpp              | 1  |

ΠΙΝΑΚΑΣ Γ'.5: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 306917

| <b>File Name</b>                                  | <b>Connectivity To Clique</b> |
|---|-------------------------------|
| protocols/yahoo/yahooverifyaccount.cpp            | 1                             |
| protocols/yahoo/yahoocontact.cpp                  | 1                             |
| protocols/yahoo/yahooconferencemessagemanager.cpp | 1                             |
| protocols/yahoo/yahooaccount.cpp                  | 0                             |
| protocols/yahoo/ui/yahoouserinfodialog.cpp        | 2                             |
| protocols/yahoo/ui/yahooinvitelistimpl.cpp        | 2                             |
| protocols/wlm/wlmtransfermanager.cpp              | 2                             |
| protocols/wlm/wlmsocket.cpp                       | 1                             |
| protocols/wlm/wlmprotocol.cpp                     | 6                             |
| protocols/wlm/wlmcontact.cpp                      | 4                             |
| protocols/wlm/wlmchatsession.cpp                  | 4                             |
| protocols/wlm/wlmchatmanager.cpp                  | 4                             |



|  |    |
|--|----|
| protocols/wlm/wlmaccount.cpp                             | 10 |
| protocols/wlm/ui/wlmeditaccountwidget.cpp                | 1  |
| protocols/oscar/oscarstatusmanager.cpp                   | 5  |
| protocols/oscar/oscarcontact.cpp                         | 5  |
| protocols/oscar/oscaraccount.cpp                         | 10 |
| protocols/oscar/liboscar/userdetails.cpp                 | 9  |
| protocols/oscar/liboscar/tests/buffertest.cpp            | 1  |
| protocols/oscar/liboscar/tasks/ssimodifytask.cpp         | 4  |
| protocols/oscar/liboscar/tasks/profiletask.cpp           | 2  |
| protocols/oscar/liboscar/tasks/messagereceivertask.cpp   | 4  |
| protocols/oscar/liboscar/tasks/logintask.cpp             | 2  |
| protocols/oscar/liboscar/tasks/icquserinfoupdatetask.cpp | 3  |
| protocols/oscar/liboscar/tasks/icqtask.cpp               | 1  |
| protocols/oscar/liboscar/tasks/filetransfertask.cpp      | 2  |
| protocols/oscar/liboscar/flapprotocol.cpp                | 1  |
| protocols/oscar/liboscar/filetransferhandler.cpp         | 2  |
| protocols/oscar/liboscar/contactmanager.cpp              | 5  |
| protocols/oscar/liboscar/contact.cpp                     | 11 |
| protocols/oscar/liboscar/connection.cpp                  | 1  |
| protocols/oscar/liboscar/buffer.cpp                      | 50 |
| protocols/oscar/icq/ui/icquserinfowidget.cpp             | 1  |
| protocols/oscar/icq/ui/icqsearchdialog.cpp               | 3  |
| protocols/oscar/icq/icqaccount.cpp                       | 1  |
| protocols/oscar/icq/aimcontact.cpp                       | 3  |
| protocols/oscar/aim/icqcontact.cpp                       | 5  |

|  |          |
|--|----------|
| <b>protocols/jabber/ui/jabberregisteraccount.cpp</b>                       | <b>0</b> |
| protocols/jabber/ui/dlgjabberchatroomslist.cpp                             | 0        |
| protocols/jabber/ui/dlgjabberchatjoin.cpp                                  | 0        |
| protocols/jabber/libjingle/talk/session/phone/win32devicemanager.cc        | 1        |
| protocols/jabber/libjingle/talk/session/phone/mediasessionclient.cc        | 5        |
| protocols/jabber/libjingle/talk/session/phone/mediasession.cc              | 3        |
| protocols/jabber/libjingle/talk/session/phone/channel.cc                   | 4        |
| protocols/jabber/libjingle/talk/p2p/client/socketmonitor.cc                | 2        |
| protocols/jabber/libjingle/talk/p2p/client/httpportallocator.cc            | 1        |
| protocols/jabber/libjingle/talk/p2p/client/connectivitychecker.cc          | 5        |
| protocols/jabber/libjingle/talk/p2p/client/basicportallocator.cc           | 6        |
| protocols/jabber/libjingle/talk/p2p/base/relayport_unittest.cc             | 1        |
| protocols/jabber/libjingle/talk/p2p/base/relayport.cc                      | 3        |
| protocols/jabber/libjingle/talk/p2p/base/rawtransportchannel.cc            | 2        |
| protocols/jabber/libjingle/talk/p2p/base/portallocatorsessionproxy.cc      | 2        |
| protocols/jabber/libjingle/talk/p2p/base/port_unittest.cc                  | 8        |
| protocols/jabber/libjingle/talk/p2p/base/port.cc                           | 4        |
| protocols/jabber/libjingle/talk/p2p/base/p2ptransportchannel.cc            | 4        |
| protocols/jabber/libjingle/talk/examples/peerconnection/client/main_wnd.cc | 2        |
| protocols/jabber/libjingle/talk/base/winfirewall.cc                        | 0        |
| protocols/jabber/libjingle/talk/base/virtualsocket_unittest.cc             | 2        |
| protocols/jabber/libjingle/talk/base/taskrunner.cc                         | 1        |
| protocols/jabber/libjingle/talk/base/taskparent.cc                         | 2        |
| protocols/jabber/libjingle/talk/base/flags.cc                              | 2        |
| protocols/jabber/libjingle/talk/base/buffer_unittest.cc                    | 1        |

|  |          |
|--|----------|
| protocols/jabber/libjingle/talk/app/webrtc/webrtcscdp.cc           | 4        |
| protocols/jabber/libiris/src/xmpp/xmpp-im/xmpp_xdata.cpp           | 4        |
| protocols/jabber/libiris/src/xmpp/jid/jid.cpp                      | 28       |
| protocols/jabber/libiris/src/jdns/src/qjdns/qjdnsshared.cpp        | 2        |
| protocols/jabber/libiris/src/jdns/src/jdns/jdns.c                  | 0        |
| protocols/jabber/libiris/src/irisnet/noncore/cutestuff/bsocket.cpp | 4        |
| protocols/jabber/libiris/src/irisnet/corelib/netnames_jdns.cpp     | 3        |
| protocols/jabber/libiris/src/irisnet/appledns/sdtest.cpp           | 0        |
| protocols/jabber/libiris/src/irisnet/appledns/appledns.cpp         | 0        |
| <b>protocols/jabber/jabberclient.cpp</b>                           | <b>0</b> |
| protocols/irc/ui/ircdialogs.cpp                                    | 2        |
| plugins/otr/privkeypopup.cpp                                       | 2        |
| plugins/otr/otrpreferences.cpp                                     | 0        |
| plugins/otr/otrplugin.cpp  | 0        |
| plugins/otr/otrlconfinterface.cpp                                  | 0        |
| plugins/otr/otrlchatinterface.cpp                                  | 5        |
| plugins/otr/otrclient.cpp  | 0        |
| plugins/otr/authenticationwizard.cpp                               | 2        |
| plugins/history2/history2dialog.cpp                                | 0        |
| plugins/history/historyclient.cpp                                  | 0        |
| plugins/history/historydialog.cpp                                  | 1        |
| libkopete/ui/kopetelistviewsearchline.cpp                          | 2        |
| libkopete/private/kopeteviewmanager.cpp                            | 0        |
| libkopete/kopetetransfermanager.cpp                                | 0        |
| libkopete/kopeteonlinestatus.cpp                                   | 18       |

|  |   |
|--|---|
| libkopete/kopetemetaccontact.cpp                   | 0 |
| kopete/identity/identitystatuswidget.cpp           | 0 |
| kopete/config/avdevice/avdeviceconfig.cpp          | 1 |
| kopete/chatwindow/kopetechatwindowstylemanager.cpp | 4 |
| kopete/chatwindow/kopetechatwindow.cpp             | 3 |
| kopete/chatwindow/chatview.cpp                     | 2 |
| kopete/chatwindow/chattexteditpart.cpp             | 0 |
| kopete/chatwindow/chatmessagepart.cpp              | 1 |

ΠΙΝΑΚΑΣ Γ'.6: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 314641

## Παράρτημα Δ΄

# K3b Sample Results

187

Στους Επόμενους Πίνακες Παρουσιάζονται τα αποτελέσματα για το σύστημα K3b.

| File Name                                  | Connectivity To Clique |
|--|------------------------|
| src/projects/k3bdataprojectmodel.cpp       | 3                      |
| src/main.cpp                               | 3                      |
| <b>src/k3bjobprogressdialog.cpp</b>        | <b>8</b>               |
| libk3b/projects/videodvd/k3bvideodvdoc.cpp | 2                      |

ΠΙΝΑΚΑΣ Δ΄.1: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ΄ αριθμόν 315463

| File Name                            | Connectivity To Clique |
|--------------------------------------|------------------------|
| src/projects/k3bdataprojectmodel.cpp | 3                      |

|   |    |
|---|----|
| src/projects/k3baudiotrackwidget.cpp          | 2  |
| src/projects/k3baudiotracksplitdialog.cpp     | 2  |
| src/projects/k3baudiodatasourceeditwidget.cpp | 2  |
| src/main.cpp                                  | 3  |
| src/k3bjobprogressdialog.cpp                  | 8  |
| libk3bdevice/k3bdeviceglobals.cpp             | 3  |
| libk3bdevice/k3bdevice_mmc.cpp                | 3  |
| libk3b/tools/k3bmsfedit.cpp                   | 1  |
| libk3b/projects/videodvd/k3bvideodvdoc.cpp    | 2  |
| libk3b/projects/audiocd/k3baudiojob.cpp       | 11 |

ΠΙΝΑΚΑΣ Δ'.2: Αποτελέσματα για το κομμάτι για την Αναφορά Σφάλματος υπ' αριθμόν 328258

| File Name                                     | Connectivity To Clique |
|---|------------------------|
| src/projects/k3bdatapjectmodel.cpp            | 3                      |
| src/projects/k3baudiotrackwidget.cpp          | 2                      |
| src/projects/k3baudiotracksplitdialog.cpp     | 2                      |
| src/projects/k3baudiodatasourceeditwidget.cpp | 2                      |
| src/main.cpp                                  | 3                      |
| src/k3bjobprogressdialog.cpp                  | 8                      |
| libk3bdevice/k3bdeviceglobals.cpp             | 3                      |
| libk3bdevice/k3bdevice_mmc.cpp                | 3                      |
| libk3b/tools/k3bmsfedit.cpp                   | 1                      |
| <b>libk3b/tools/k3blibdvdcss.cpp</b>          | <b>3</b>               |
| <b>libk3b/tools/k3bcdparanoialib.cpp</b>      | <b>1</b>               |

|  |    |
|--|----|
| libk3b/projects/videodvd/k3bvideodvdoc.cpp | 2  |
| libk3b/projects/audiocd/k3baudiojob.cpp    | 11 |

ΠΙΝΑΚΑΣ Δ'.3: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 328769

| File Name                                     | Connectivity To Clique |
|---|------------------------|
| src/rip/k3baudioconvertoptionwidget.cpp       | 2                      |
| src/projects/k3bdataprojectmodel.cpp          | 3                      |
| src/projects/k3baudiotrackwidget.cpp          | 2                      |
| src/projects/k3baudiotracksplitdialog.cpp     | 2                      |
| src/projects/k3baudiodatasourceeditwidget.cpp | 2                      |
| src/main.cpp                                  | 3                      |
| src/k3bjobprogressdialog.cpp                  | 8                      |
| libk3bdevice/k3bdeviceglobals.cpp             | 3                      |
| libk3bdevice/k3bdevice_mmc.cpp                | 3                      |
| libk3b/tools/k3bmsfedit.cpp                   | 1                      |
| libk3b/tools/k3blibdvdcss.cpp                 | 3                      |
| libk3b/tools/k3bcdparanoialib.cpp             | 1                      |
| libk3b/projects/videodvd/k3bvideodvdoc.cpp    | 2                      |
| <b>libk3b/projects/k3bcuefileparser.cpp</b>   | <b>2</b>               |
| libk3b/projects/dataacd/k3bisoptions.cpp      | 1                      |
| libk3b/projects/dataacd/k3bisoimager.cpp      | 6                      |
| libk3b/projects/audiocd/k3baudiojob.cpp       | 11                     |

ΠΙΝΑΚΑΣ Δ'.4: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 337201

| File Name                                     | Connectivity To Clique |
|---|------------------------|
| src/rip/k3baudioconvertingoptionwidget.cpp    | 2                      |
| src/projects/k3bdataprojectmodel.cpp          | 3                      |
| src/projects/k3baudiotrackwidget.cpp          | 2                      |
| src/projects/k3baudiotracksplitdialog.cpp     | 2                      |
| src/projects/k3baudiodatasourceeditwidget.cpp | 2                      |
| src/main.cpp                                  | 3                      |
| <b>src/k3blsofwrapper.cpp</b>                 | <b>3</b>               |
| src/k3bjobprogressdialog.cpp                  | 8                      |
| libk3bdevice/k3bdeviceglobals.cpp             | 3                      |
| libk3bdevice/k3bdevice_mmc.cpp                | 3                      |
| libk3b/tools/k3bmsfedit.cpp                   | 1                      |
| libk3b/tools/k3blibdvdcss.cpp                 | 3                      |
| libk3b/tools/k3bcdparanoialib.cpp             | 1                      |
| libk3b/projects/videodvd/k3bvideodvdoc.cpp    | 2                      |
| libk3b/projects/k3bcuefileparser.cpp          | 2                      |
| libk3b/projects/dataacd/k3bisoptions.cpp      | 1                      |
| libk3b/projects/dataacd/k3bisoimager.cpp      | 6                      |
| libk3b/projects/audiocd/k3baudiojob.cpp       | 11                     |

ΠΙΝΑΚΑΣ Δ'.5: Αποτελέσματα για το κοπετε για την Αναφορά Σφάλματος υπ' αριθμόν 340515

| File Name                                  | Connectivity To Clique |
|--|------------------------|
| src/rip/k3baudioconvertingoptionwidget.cpp | 2                      |
| src/projects/k3bdataprojectmodel.cpp       | 3                      |



|  |          |
|--|----------|
| src/projects/k3baudiotrackwidget.cpp           | 2        |
| src/projects/k3baudiotracksplitdialog.cpp      | 2        |
| src/projects/k3baudiodatasourceeditwidget.cpp  | 2        |
| src/main.cpp                                   | 3        |
| src/k3blsofwrapper.cpp                         | 3        |
| src/k3bjobprogressdialog.cpp                   | 8        |
| <b>plugins/decoder/flac/k3bflacdecoder.cpp</b> | <b>2</b> |
| libk3bdevice/k3bdeviceglobals.cpp              | 3        |
| libk3bdevice/k3bdevice_mmc.cpp                 | 3        |
| libk3b/tools/k3bmsfedit.cpp                    | 1        |
| libk3b/tools/k3blibdvdcss.cpp                  | 3        |
| libk3b/tools/k3bcdparanoialib.cpp              | 1        |
| libk3b/projects/videodvd/k3bvideodvdoc.cpp     | 2        |
| libk3b/projects/k3bcuefileparser.cpp           | 2        |
| libk3b/projects/dataacd/k3bisoptions.cpp       | 1        |
| libk3b/projects/dataacd/k3bisoimager.cpp       | 6        |
| libk3b/projects/audioacd/k3baudiojob.cpp       | 11       |

ΠΙΝΑΚΑΣ Δ'.6: Αποτελέσματα για το kopete για την Αναφορά Σφάλματος υπ' αριθμόν 341077

| File Name                                  | Connectivity To Clique |
|--|------------------------|
| src/rip/k3baudioconvertingoptionwidget.cpp | 2                      |
| <b>src/rip/k3baudioacdview.cpp</b>         | <b>5</b>               |
| src/projects/k3bdataurladdingdialog.cpp    | 2                      |
| src/projects/k3bdataprotectmodel.cpp       | 3                      |

|  |    |
|--|----|
| src/projects/k3baudiotrackwidget.cpp                                   | 2  |
| src/projects/k3baudiotracksplitdialog.cpp                              | 2  |
| src/projects/k3baudiotrackaddingdialog.cpp                             | 1  |
| src/projects/k3baudiodatasourceeditwidget.cpp                          | 2  |
| src/main.cpp   | 3  |
| src/k3bwelcomewidget.cpp   | 5  |
| src/k3blsofwrapper.cpp   | 3  |
| src/k3bjobprogressdialog.cpp   | 8  |
| src/k3bapplication.cpp   | 10 |
| plugins/project/audiometainforenamer/k3baudiometainforenamerplugin.cpp | 5  |
| plugins/decoder/mp3/k3bmaddecoder.cpp                                  | 10 |
| plugins/decoder/flac/k3bflacdecoder.cpp                                | 2  |
| libk3bdevice/k3bdeviceglobals.cpp                                      | 3  |
| libk3bdevice/k3bdevice_mmc.cpp   | 3  |
| libk3b/tools/k3bmultichoicedialog.cpp                                  | 3  |
| libk3b/tools/k3bmsfedit.cpp  | 1  |
| libk3b/tools/k3blibdvdcss.cpp  | 3  |
| libk3b/tools/k3bcdparanoialib.cpp                                      | 1  |
| libk3b/projects/videodvd/k3bvideodvddoc.cpp                            | 2  |
| libk3b/projects/k3bcuefileparser.cpp                                   | 2  |
| libk3b/projects/k3bcdrecordwriter.cpp                                  | 1  |
| libk3b/projects/k3bcdrdaowriter.cpp                                    | 2  |
| libk3b/projects/datacd/k3bisooptions.cpp                               | 1  |
| libk3b/projects/datacd/k3bisoiager.cpp                                 | 6  |
| libk3b/projects/datacd/k3bdatapreparationjob.cpp                       | 5  |

|   |    |
|---|----|
| libk3b/projects/audiocd/k3baudiojob.cpp | 11 |
| libk3b/core/k3bthreadjob.cpp            | 4  |

ΠΙΝΑΚΑΣ Δ'.7: Αποτελέσματα για το κοπετε για την Αναφορά Σφάλματος υπ' αριθμόν 341435



# Bibliography

- [1] Gregory Tassef. The economic impacts of inadequate infrastructure for software testing. *National Institute of Standards and Technology, RTI Project, 7007(011)*, 2002.
- [2] Thomas M. Pigoski. *Practical Software Maintenance: Best Practices for Managing Your Software Investment*. John Wiley & Sons, Inc., New York, NY, USA, 1996. ISBN 0471170011.
- [3] Quinn Hanam, Lin Tan, Reid Holmes, and Patrick Lam. Finding patterns in static analysis alerts: Improving actionable alert ranking. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, MSR 2014, pages 152–161, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2863-0. doi: 10.1145/2597073.2597100. URL <http://doi.acm.org/10.1145/2597073.2597100>.
- [4] Anh Tuan Nguyen, Tung Thanh Nguyen, J. Al-Kofahi, Hung Viet Nguyen, and T.N. Nguyen. A topic-based approach for narrowing the search space of buggy files from a bug report. In *Automated Software Engineering (ASE), 2011 26th IEEE/ACM International Conference on*, pages 263–272, Nov 2011. doi: 10.1109/ASE.2011.6100062.
- [5] Tian Jiang, Lin Tan, and Sunghun Kim. Personalized defect prediction. In *Automated Software Engineering (ASE), 2013 IEEE/ACM 28th International Conference on*, pages 279–289, Nov 2013. doi: 10.1109/ASE.2013.6693087.
- [6] Jon Eyolfson, Lin Tan, and Patrick Lam. Correlations between bugginess and time-based commit characteristics. *Springer Empirical Software Engineering*, 2013.
- [7] Rafael Lotufo, Zeeshan Malik, and Krzysztof Czarnecki. Modelling the ‘hurried’ bug report reading process to summarize bug reports. In *International Conference on Software Maintenance*, Trento, Italy, 09/2012 2012. IEEE, IEEE. Best paper award at ICSM 2012.
- [8] Stacy K. Lukins, Nicholas A. Kraft, and Letha H. Etzkorn. Bug localization using latent dirichlet allocation. *Information and Software Technology*, 52(9):972 – 990,

2010. ISSN 0950-5849. doi: <http://dx.doi.org/10.1016/j.infsof.2010.04.002>. URL <http://www.sciencedirect.com/science/article/pii/S0950584910000650>.
- [9] Shivani Rao and Avinash Kak. Retrieval from software libraries for bug localization: A comparative study of generic and composite text models. In *Proceedings of the 8th Working Conference on Mining Software Repositories, MSR '11*, pages 43–52, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0574-7. doi: 10.1145/1985441.1985451. URL <http://doi.acm.org/10.1145/1985441.1985451>.
- [10] A. Marcus, A. Sergeyev, V. Rajlich, and J.I. Maletic. An information retrieval approach to concept location in source code. In *Reverse Engineering, 2004. Proceedings. 11th Working Conference on*, pages 214–223, Nov 2004. doi: 10.1109/WCRE.2004.10.
- [11] S.W. Thomas, M. Nagappan, D. Blostein, and A.E. Hassan. The impact of classifier configuration and classifier combination on bug localization. *Software Engineering, IEEE Transactions on*, 39(10):1427–1443, Oct 2013. ISSN 0098-5589. doi: 10.1109/TSE.2013.27.
- [12] Vassilios Tzerpos and Richard C. Holt. Mojo: A distance metric for software clusterings. In *Sixth Working Conference on Reverse Engineering, WCRE '99, Atlanta, Georgia, USA, October 6-8, 1999*, page 187, 1999. doi: 10.1109/WCRE.1999.806959. URL <http://dx.doi.org/10.1109/WCRE.1999.806959>.
- [13] Brian S. Mitchell and Spiros Mancoridis. CRAFT: A Framework for Evaluating Software Clustering Results in the Absence of Benchmark Decompositions. In *WCRE*, pages 93–102, 2001.
- [14] Nicolas Anquetil, Cédric Fourier, and Timothy C. Lethbridge. Experiments with clustering as a software remodularization method. In *Proceedings of the Sixth Working Conference on Reverse Engineering, WCRE '99*, pages 235–, Washington, DC, USA, 1999. IEEE Computer Society. ISBN 0-7695-0303-9. URL <http://dl.acm.org/citation.cfm?id=832306.837051>.
- [15] Rainer Koschke and Thomas Eisenbarth. A Framework for Experimental Evaluation of Clustering Techniques. In *8th International Workshop on Program Comprehension (IWPC 2000), 10-11 June 2000, Limerick, Ireland*, pages 201–210. IEEE Computer Society, 2000. ISBN 0-7695-0656-9. doi: 10.1109/WPC.2000.852494.
- [16] Chung-Horng Lung, Marzia Zaman, and Amit Nandi. Applications of clustering techniques to software partitioning, recovery and restructuring. *J. Syst. Softw.*, 73(2):227–244, October 2004. ISSN 0164-1212. doi: 10.1016/S0164-1212(03)00234-6. URL [http://dx.doi.org/10.1016/S0164-1212\(03\)00234-6](http://dx.doi.org/10.1016/S0164-1212(03)00234-6).

- [17] Chung-Horng Lung, Xia Xu, Marzia Zaman, and Anand Srinivasan. Program restructuring using clustering techniques. *Journal of Systems and Software*, 79(9):1261 – 1279, 2006. ISSN 0164-1212. doi: <http://dx.doi.org/10.1016/j.jss.2006.02.037>. URL <http://www.sciencedirect.com/science/article/pii/S0164121206000604>. Selected papers from the fourth Source Code Analysis and Manipulation (SCAM 2004) Workshop Fourth Source Code Analysis and Manipulation Workshop.
- [18] Markus Bauer and Mircea Trifu. Architecture-aware adaptive clustering of OO systems. In *8th European Conference on Software Maintenance and Reengineering (CSMR 2004), 24-26 March 2004, Tampere, Finland, Proceedings*, pages 3–14, 2004. doi: 10.1109/CSMR.2004.1281401. URL <http://doi.ieeecomputersociety.org/10.1109/CSMR.2004.1281401>.
- [19] Shi Zhong, Taghi M. Khoshgoftaar, and Naeem Seliya. Analyzing software measurement data with clustering techniques. *IEEE Intelligent Systems*, 19(2):20–27, March 2004. ISSN 1541-1672. doi: 10.1109/MIS.2004.1274907. URL <http://dx.doi.org/10.1109/MIS.2004.1274907>.
- [20] Vassilios Tzerpos and R. C. Holt. Acdc : An algorithm for comprehension-driven clustering. In *In Proceedings of the Seventh Working Conference on Reverse Engineering*, pages 258–267. IEEE, 2000.
- [21] S. Mancoridis, B. S. Mitchell, Y. Chen, and E. R. Gansner. Bunch: A clustering tool for the recovery and maintenance of software system structures. In *In Proceedings? IEEE International Conference on Software Maintenance*, page pages. IEEE Computer Society Press, 1999.
- [22] Hausi A. Müller, Mehmet A. Orgun, Scott R. Tilley, and James S. Uhl. A reverse engineering approach to subsystem structure identification, 1993.
- [23] V. Tzerpos and R. C. Holt. The orphan adoption problem in architecture maintenance. In *Working Conference on Reverse Engineering (WCRE 1997)*, page 76, Amsterdam, The Netherlands, Oktober 1997.
- [24] Inc. Object Management Group. Object management group, inc. [www.omg.org/mda](http://www.omg.org/mda), .
- [25] Petri Selonen. Set operations for unified modeling language,. In *Proceedings of the Eight Symposium on Programming Languages and Tools, SPLST'2003*, pages 70–81, 2003.
- [26] Inc. Object Management Group. Uml infrastructure specification. <http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF/>, .

- [27] Inc. Object Management Group. Mof core specification version 2.4.1. <http://www.omg.org/spec/MOF/2.4.1/>, 2013.
- [28] Bart Du Bois, Bart Van Rompaey, Karel Meijfroidt, and Eric Suijs. Supporting reengineering scenarios with FETCH: an experience report. *ECEASST*, 8, 2007. URL <http://eceasst.cs.tu-berlin.de/index.php/eceasst/article/view/116>.
- [29] The source navigator ide homepage. <http://sourcnav.sourceforge.net/>.
- [30] Tim Littlefair B. Sc. An investigation into the use of software code metrics in the industrial software development environment, 2001.
- [31] Famoos. <http://scg.unibe.ch/archive/famoos/>.
- [32] Mike Imber. Software engineering environments, 1991.
- [33] John Mylopoulos, Alexander Borgida, Matthias Jarke, and Manolis Koubarakis. Telos: Representing knowledge about information systems. *ACM Trans. Inf. Syst.*, 8(4):325–362, 1990. doi: 10.1145/102675.102676. URL <http://doi.acm.org/10.1145/102675.102676>.
- [34] University of Toronto. Dept. of Computer Science, M. Koubarakis, J. Mylopoulos, M. Stanley, and A. Bordiga. *Telos: features and formalization*. FORTH CSI TR: Computer Science Institute. Department of Computer Science, University of Toronto, 1989. URL [https://books.google.gr/books?id=Ke0\\_PgAACAAJ](https://books.google.gr/books?id=Ke0_PgAACAAJ).
- [35] Axel Van Lamsweerde and Emmanuel Letier. From object orientation to goal orientation: A paradigm shift for requirements engineering. In *Radical Innovations of Software & System Engineering, Monterey'02 Workshop, Venice(Italy), LNCS*, pages 4–8. Springer-Verlag, 2003.
- [36] URL <https://eclipse.org/modeling/emf/>.
- [37] Lee C.H. Pieraccini R., Levin R. Stochastic representation of conceptual structure in the atis task, Feb 1991.