

OpenRefine

Έγγραφο Ελέγχου Ποιότητας

Κωνσταντίνος Θωμασιάδης

Ευστάθιος Ιωσηφίδης

Αικατερίνη Κρότκα

Ημερομηνία: 20 Δεκεμβρίου 2024

Πίνακας Περιεχομένων

OpenRefine.....	1
Έγγραφο Ελέγχου Ποιότητας.....	1
1. Εισαγωγή.....	3
2. Αναφορά Εξέλιξης Ποιότητας.....	5
Μεθοδολογία εξέλιξης ποιότητας.....	5
1. NOC (Number of Classes).....	6
2. DIT (Depth of Inheritance Tree).....	7
3. NOCC (Number of Children).....	8
4. CBO (Coupling Between Objects).....	9
5. LCOM (Lack of Cohesion of Methods).....	10
6. WMC* (Weighted Methods per Class).....	11
7. SIZE1 (Μέγεθος Κώδικα).....	12
Ευρήματα.....	13
Συζήτηση για Ενδιαφέροντα Αποτελέσματα.....	13
3. Αναγνώριση Προβλημάτων Ποιότητας.....	14
Επεξήγηση Μεθοδολογίας Αναγνώρισης Προβλημάτων Ποιότητας.....	14
Ευρήματα.....	14
Συζήτηση Ευρημάτων.....	15
Συμπεράσματα και Προτάσεις.....	16

1. Εισαγωγή

1.1 Περιγραφή Συστήματος

Το **OpenRefine** είναι μια εφαρμογή επιφάνειας εργασίας ανοιχτού κώδικα που χρησιμοποιείται για την εκκαθάριση και τη μετατροπή δεδομένων σε άλλες μορφές, μια διαδικασία που συχνά ονομάζεται "**data wrangling**". Το OpenRefine λειτουργεί με σύνολα δεδομένων που είναι δομημένα σε σειρές και στήλες, όπως αυτά που συναντώνται σε υπολογιστικά φύλλα (π.χ., **CSV**) ή βάσεις δεδομένων.

Ωστόσο, σε αντίθεση με τα παραδοσιακά υπολογιστικά φύλλα, το OpenRefine προσφέρει λειτουργικότητα που θυμίζει περισσότερο τις σχεσιακές βάσεις δεδομένων, επιτρέποντας **φιλτράρισμα, ομαδοποίηση και καθαρισμό δεδομένων** με εξειδικευμένα εργαλεία και κριτήρια. Μερικές από τις κύριες δυνατότητες του περιλαμβάνουν:

- **Εκκαθάριση Δεδομένων:** Αναγνώριση και διόρθωση ασυνεπειών στα δεδομένα.
- **Μετατροπή Δεδομένων:** Μετασχηματισμός δεδομένων μεταξύ διαφορετικών μορφών, π.χ., από CSV σε JSON ή XML.
- **Εργαλεία Σύνδεσης (Linking):** Ενσωμάτωση δεδομένων από εξωτερικές πηγές - διασυνδέσεις με API.
- **Διαδραστικό Περιβάλλον:** Χρήση φιλικής διεπαφής που επιτρέπει τη γρήγορη διαχείριση μεγάλων συνόλων δεδομένων.

1.2 Αναφορές

Οι πληροφορίες σχετικά με τη λειτουργία του OpenRefine βασίστηκαν στους εξής πόρους:

- [Επίσημη ιστοσελίδα του OpenRefine](#)
- [GitHub Repository του OpenRefine](#)

1.3 Επισκόπηση

Το OpenRefine αναπτύσσεται συνεχώς, αυξάνοντας τόσο τη λειτουργικότητα όσο και τη δομή του κώδικα. Η παρούσα ανάλυση εστιάζει στη μελέτη βασικών μετρικών όπως **NOC** (Number of Classes), **DIT** (Depth of Inheritance Tree), **NOCC** (Number of Children), **CBO** (Coupling Between Objects), **LCOM** (Lack of Cohesion of Methods), **WMC*** (Weighted Methods per Class) και **SIZE1** (Μέγεθος Κώδικα). Στόχος είναι να εντοπιστούν προβλήματα ποιότητας και να προταθούν τρόποι βελτίωσης.

Στο παρόν έγγραφο, θα παρουσιαστούν:

1. Η **εξέλιξη της ποιότητας** του OpenRefine μέσα από διάφορες εκδόσεις (tags), εστιάζοντας σε μετρήσεις όπως η πολυπλοκότητα, η συνοχή και το μέγεθος.
2. Η **αναγνώριση προβλημάτων ποιότητας** στον κώδικα.
3. Η τελική ανάλυση θα περιλαμβάνει συζήτηση για τη βελτίωση της ποιότητας του λογισμικού και προτάσεις για περαιτέρω ανάπτυξη.

2. Αναφορά Εξέλιξης Ποιότητας

Μεθοδολογία εξέλιξης ποιότητας

Για την αναγνώριση των προβλημάτων ποιότητας στο λογισμικό, εφαρμόστηκε μια ανάλυση μετρικών με το εργαλείο MetricsCalculator.jar, που αφορούν βασικά χαρακτηριστικά της δομής και της πολυπλοκότητας του κώδικα. Οι μετρικές που εξετάστηκαν περιλαμβάνουν:

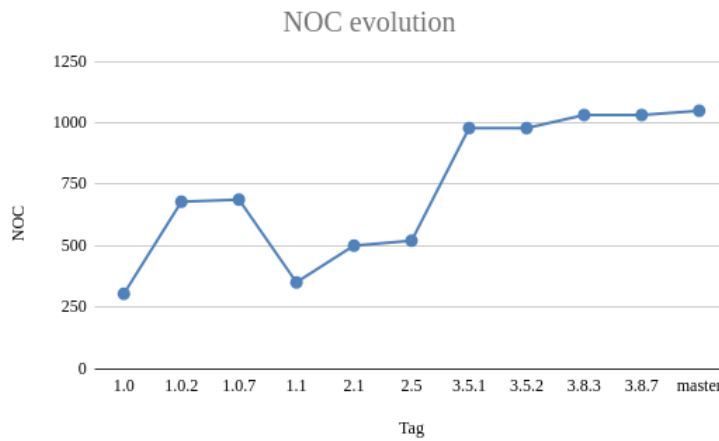
- **NOC (Number of Classes):** Εξετάζει τον αριθμό των κλάσεων, ο οποίος υποδεικνύει την αύξηση του μεγέθους του συστήματος.
- **WMC* (Weighted Methods per Class):** Αποκαλύπτει την πολυπλοκότητα των κλάσεων βάσει του αριθμού των μεθόδων και του "βάρους" τους.
- **CBO (Coupling Between Objects):** Καταγράφει τη σύζευξη μεταξύ των κλάσεων και εντοπίζει κλάσεις που εξαρτώνται έντονα από άλλες, αυξάνοντας τον κίνδυνο αλλαγών και δυσκολίας συντήρησης.
- **DIT (Depth of Inheritance Tree):** Εντοπίζει το βάθος της ιεραρχίας κληρονομικότητας, το οποίο επηρεάζει την κατανόηση του κώδικα.
- **LCOM (Lack of Cohesion in Methods):** Μετρά την συνοχή των μεθόδων μέσα σε μια κλάση. Μικρή συνοχή αποτελεί ένδειξη ότι η κλάση έχει πολλαπλές, άσχετες λειτουργίες.

Η μεθοδολογία περιελάμβανε τη συλλογή και ανάλυση των παραπάνω μετρικών σε διάφορες εκδόσεις του λογισμικού, εστιάζοντας στη συνολική τάση (SUM), τον μέσο όρο (AVG) και τη μέγιστη τιμή (MAX) για κάθε μετρική. Τα ευρήματα από τις διακυμάνσεις αυτών των μετρικών βοήθησαν στην ανάδειξη κλάσεων ή τμημάτων του κώδικα που χρήζουν περαιτέρω ανάλυσης ή refactoring.

Metric	Stat	1.0	1.0.2	1.0.7	1.1	2.1	2.5	3.5.1	3.5.2	3.8.3	3.8.7	master
NOC	SUM	305	679	688	351	501	520	978	978	1031	1031	1048
	AVG	20739	58372	58798	24537	40156	45886	67633	67633	75683	75683	79478
SIZE1	AVG	68	85,97	85,46	70,11	80,15	88,24	69,15	69,15	73,48	73,48	75,84
	MAX	1318	5351	5351	1318	5351	5351	5351	5351	5351	5351	5251
WMC*	SUM	650,74	1304,92	1314,46	708,73	981,79	1061,73	1409,77	1409,77	1537,72	1537,72	1556,33
	AVG	2,13	1,92	1,91	2,02	1,96	2,04	1,44	1,44	1,49	1,49	1,49
DIT	MAX	46	46	46	46	46	46	20	20	21,5	21,5	21,5
	SUM	498	997	1003	544	782	811	763	763	1079	1079	1143
NOCC	AVG	1,63	1,47	1,46	1,55	1,56	1,56	0,78	0,78	1,05	1,05	1,09
	MAX	19	19	19	19	19	19	5	5	5	5	5
CBO	SUM	145	336	340	249	249	259	484	484	524	524	537
	AVG	0,48	0,49	0,49	0,46	0,5	0,5	0,49	0,49	0,51	0,51	0,51
LCOM	MAX	43	80	81	45	63	65	111	111	124	124	81
	SUM	1276	2956	2988	1510	2121	2317	3468	3468	3927	3927	4029
CBO	AVG	4,18	4,35	4,34	4,31	4,23	4,46	3,55	3,55	3,81	3,81	3,84
	MAX	35	54	54	35	54	54	24	24	24	24	24
LCOM	SUM	3866	39603	39737	5197	33912	35634	46506	46506	49877	49877	52947
	AVG	12,68	58,33	57,76	14,85	67,69	68,53	47,55	47,55	48,42	48,42	50,52
LCOM	MAX	577	24976	24976	577	24976	24976	24976	24976	24976	24976	24753

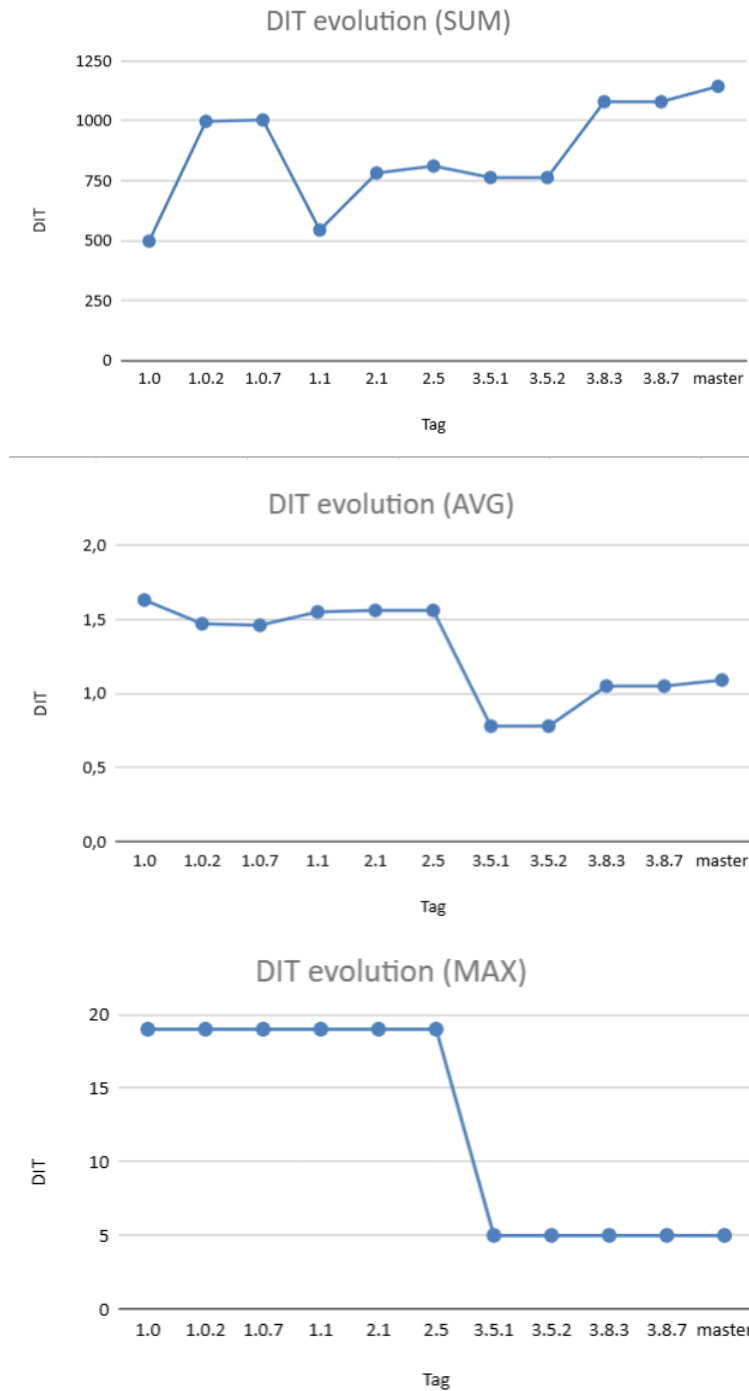
Ο συνολικός πίνακας με τις μετρικές ανά έκδοση είναι ο παραπάνω. Η ανάλυση των μετρικών για το OpenRefine πραγματοποιήθηκε στις παρακάτω εκδόσεις: **1.0, 1.0.2, 1.0.7, 1.1, 2.1, 2.5, 3.5.1, 3.5.2, 3.8.3, 3.8.7 και η master**. Ας σχολιάσουμε μερικά αποτελέσματα.

1. NOC (Number of Classes)



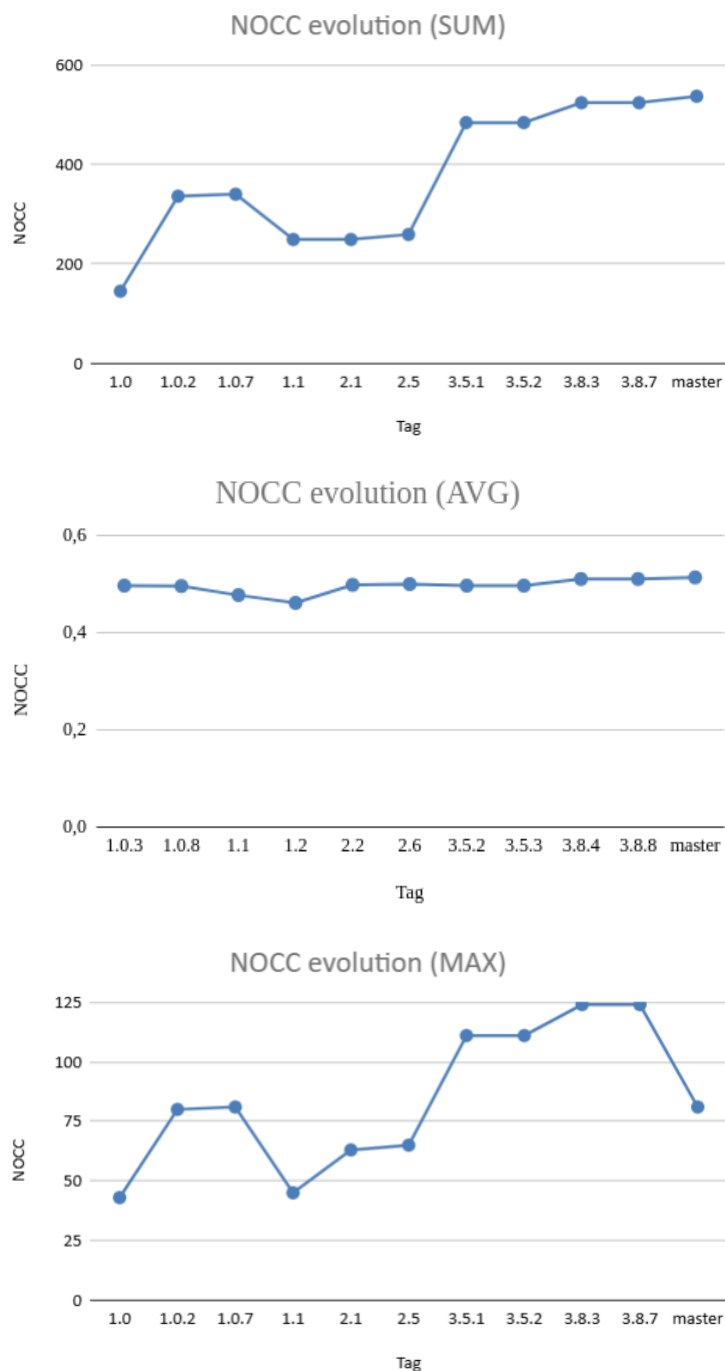
Η συνολική εξέλιξη του NOC δείχνει μια σταθερή αύξηση στον αριθμό των κλάσεων σε κάθε έκδοση. Η αύξηση αυτή είναι λογική, καθώς προκύπτει από τη σταδιακή ανάπτυξη του λογισμικού και την προσθήκη νέων λειτουργιών. Παράλληλα, η μετρική αποκαλύπτει ότι το σύστημα αυξήθηκε από 679 κλάσεις στην αρχή σε 1048 στο master (αύξηση περίπου 54%). Η ανάπτυξη αυτή είναι θετική, αλλά απαιτεί ανάλυση για την οργάνωση και την ποιότητα των νέων κλάσεων. Μεταξύ της έκδοσης 1.0.7 και 1.1 παρατηρούμε μείωση των κλάσεων από 688 σε 351. Πιθανό να έχει γίνει συγχώνευση κλάσεων (refactoring) ή διαγραφή μη χρήσιμων κλάσεων.

2. DIT (Depth of Inheritance Tree)



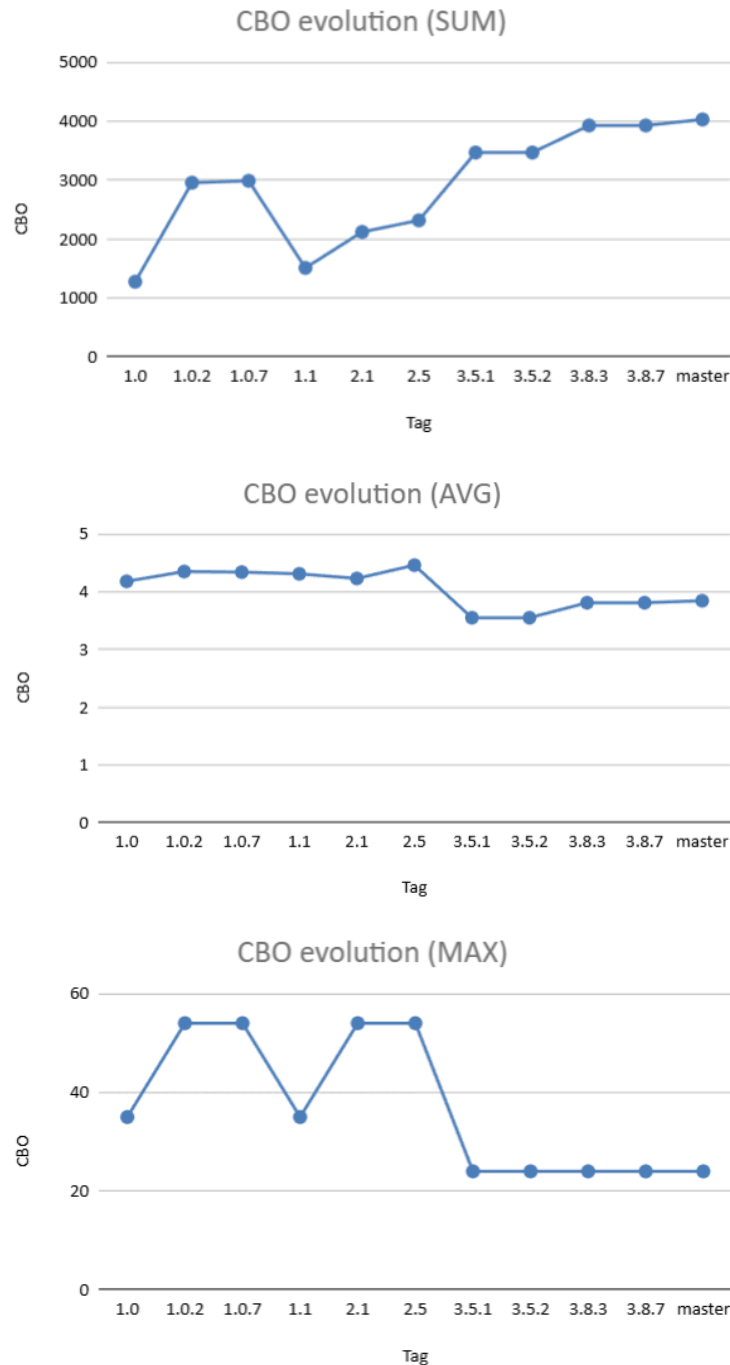
Η μείωση του AVG από 1.63 στην την έκδοση 1.0. σε 1.05 στην έκδοση 3.8.7 είναι ιδιαίτερα θετική, καθώς απλοποιεί την κληρονομικότητα και μειώνει την πολυπλοκότητα της ιεραρχίας. Η υπερβολικά βαθιά κληρονομικότητα δυσκολεύει την κατανόηση του κώδικα και αυξάνει τη σύζευξη. Ενδιαφέρον είναι ότι το MAX μειώθηκε από 19 στην έκδοση 1.0 σε 5 στην έκδοση 3.8.7, γεγονός που σηματοδοτεί εκκαθάριση βαθιών και περιττών ιεραρχιών. Αυτό ενισχύει τη συντηρησιμότητα του κώδικα. Σταθερές τιμές του MAX μέχρι την έκδοση 2.5 και μετά από την 3.5.1, μπορεί να σημαίνουν διατήρηση παλαιού κώδικα ή ότι η αρχιτεκτονική της κληρονομικότητας δεν έχει αναθεωρηθεί.

3. NOCC (Number of Children)



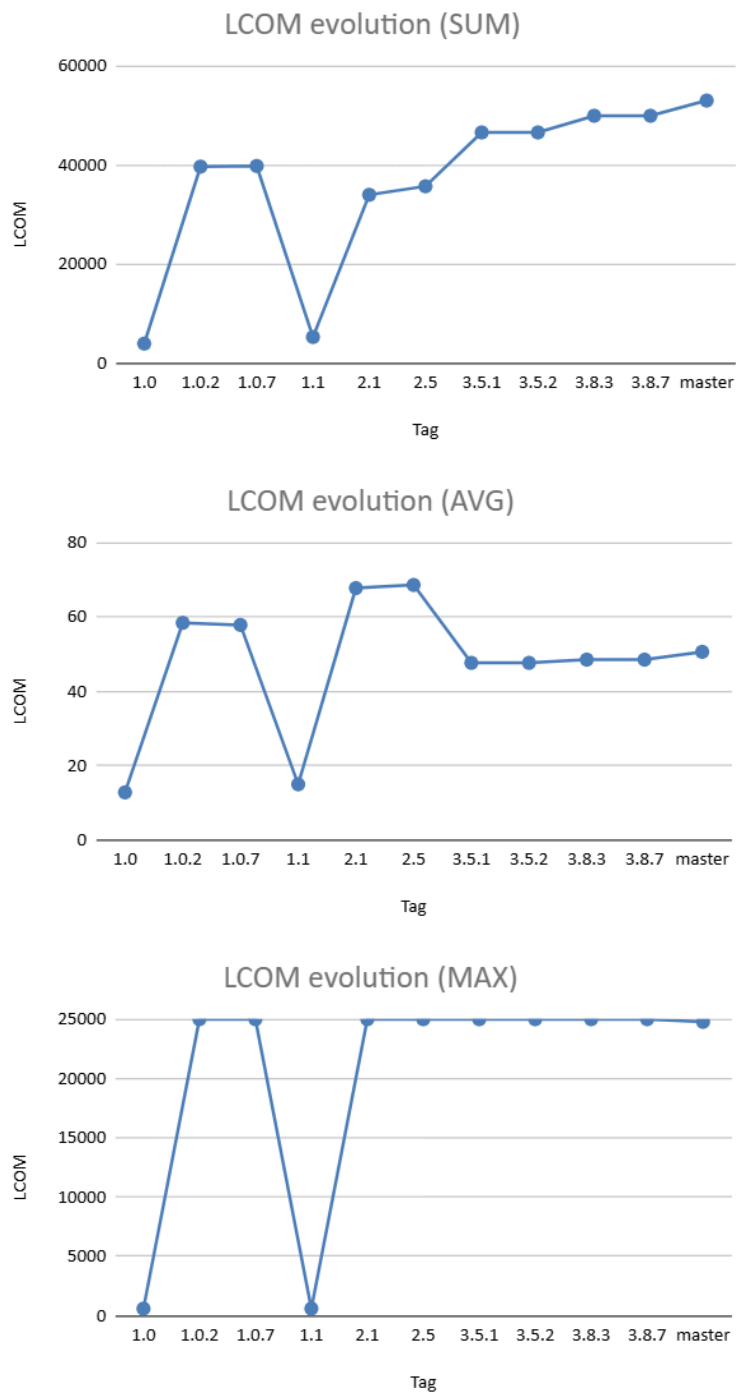
Το **AVG** παραμένει σχετικά σταθερό καθ' όλη τη διάρκεια της ανάπτυξης του έργου. Ωστόσο, το **MAX** ξεκινά από την τιμή 81 στην έκδοση 1.0.7, υποδεικνύοντας την ύπαρξη κλάσης με υπερβολικά πολλές υποκλάσεις, κάτι που ενδέχεται να προκαλέσει αυξημένη εξάρτηση και χαμηλή συντηρησιμότητα του κώδικα. Παρατηρούμε ότι το **MAX** μειώνεται στην έκδοση 1.1 στο 45, κάτι που υποδηλώνει ότι πραγματοποιήθηκε **refactoring**. Αυτή η μείωση αντικατοπτρίζεται και στη μετρική **NOCC**, όπου παρατηρείται μείωση του αριθμού των κλάσεων. Στην συνέχεια, από την έκδοση 1.1 και μετά, παρατηρείται σταδιακή αύξηση του **MAX**, φτάνοντας στην έκδοση 3.8.7 στην τιμή 124. Στην τρέχουσα έκδοση, το **MAX** μειώθηκε ξανά στο 81. Η κλάση με την υψηλότερη τιμή **MAX** πρέπει να αναλυθεί περαιτέρω και ενδεχομένως να ανασχεδιαστεί για να μειωθεί η εξάρτηση και να βελτιωθεί η συντηρησιμότητα του συστήματος. Η ισορροπία αυτή ενισχύει τη βιωσιμότητα του σχεδιασμού.

4. CBO (Coupling Between Objects)



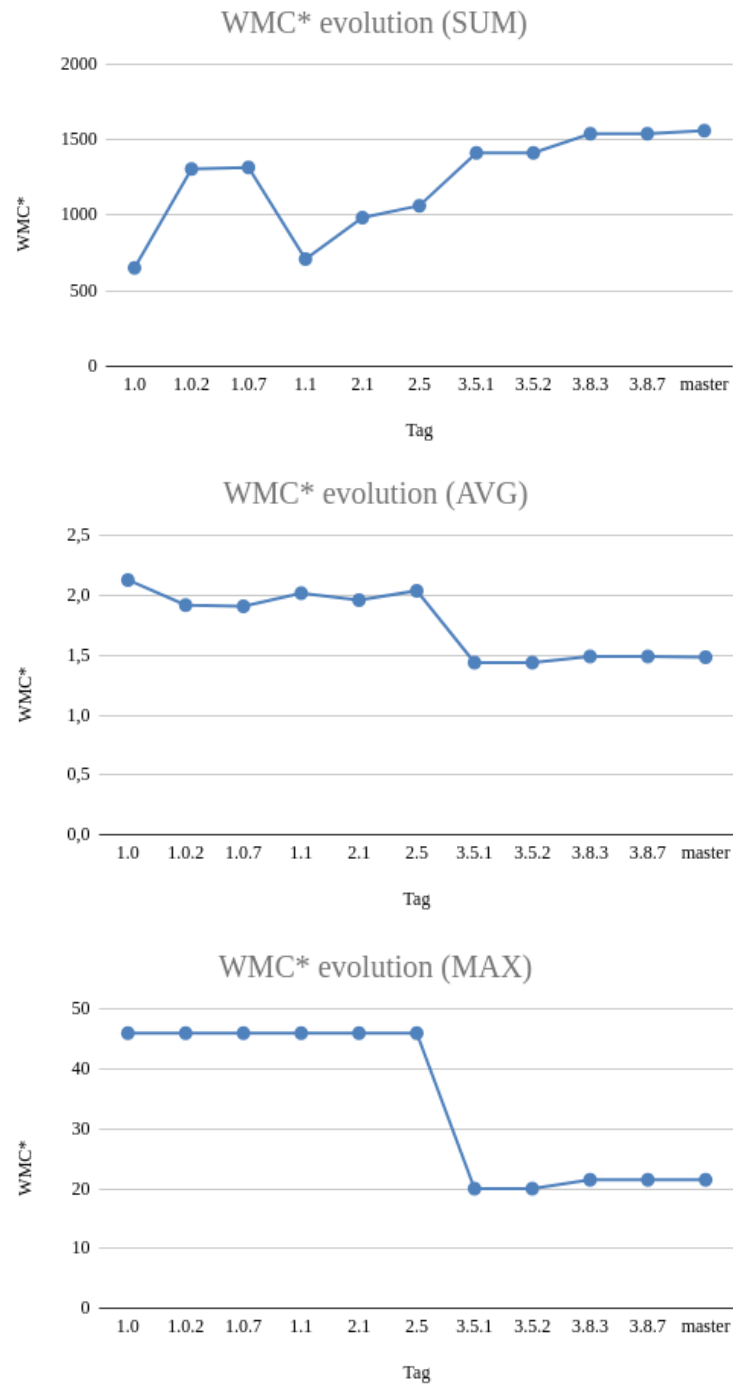
Παρατηρείται μείωση του **AVG** από 4.18 στην έκδοση 1.0 σε 3.81 στην έκδοση 3.8.7, κάτι το οποίο είναι ενθαρρυντικό, καθώς υποδηλώνει ότι η σύζευξη μεταξύ των κλάσεων μειώθηκε. Ωστόσο, το **MAX** παραμένει σταθερό στο 54 από την έκδοση 1.0.2 μέχρι την έκδοση 2.5. Στην έκδοση 1.1 παρατηρείται μείωση στο 35, αλλά στη συνέχεια αυξάνεται ξανά στις εκδόσεις 2.1 και 2.5 στο 54, όπως στην αρχική κατάσταση, υποδεικνύοντας υπερβολική σύζευξη σε συγκεκριμένες κλάσεις. Από την έκδοση 3.5.1 και μετά, το **MAX** μειώνεται στο 24. Παρά τη μείωση, η σταθερότητα του **MAX** δείχνει ότι ορισμένες κλάσεις παραμένουν έντονα συνδεδεμένες. Αν και η συνολική μείωση της σύζευξης είναι θετική, οι πιο συνδεδεμένες κλάσεις θα πρέπει να εξεταστούν για **refactoring**.

5. LCOM (Lack of Cohesion of Methods)



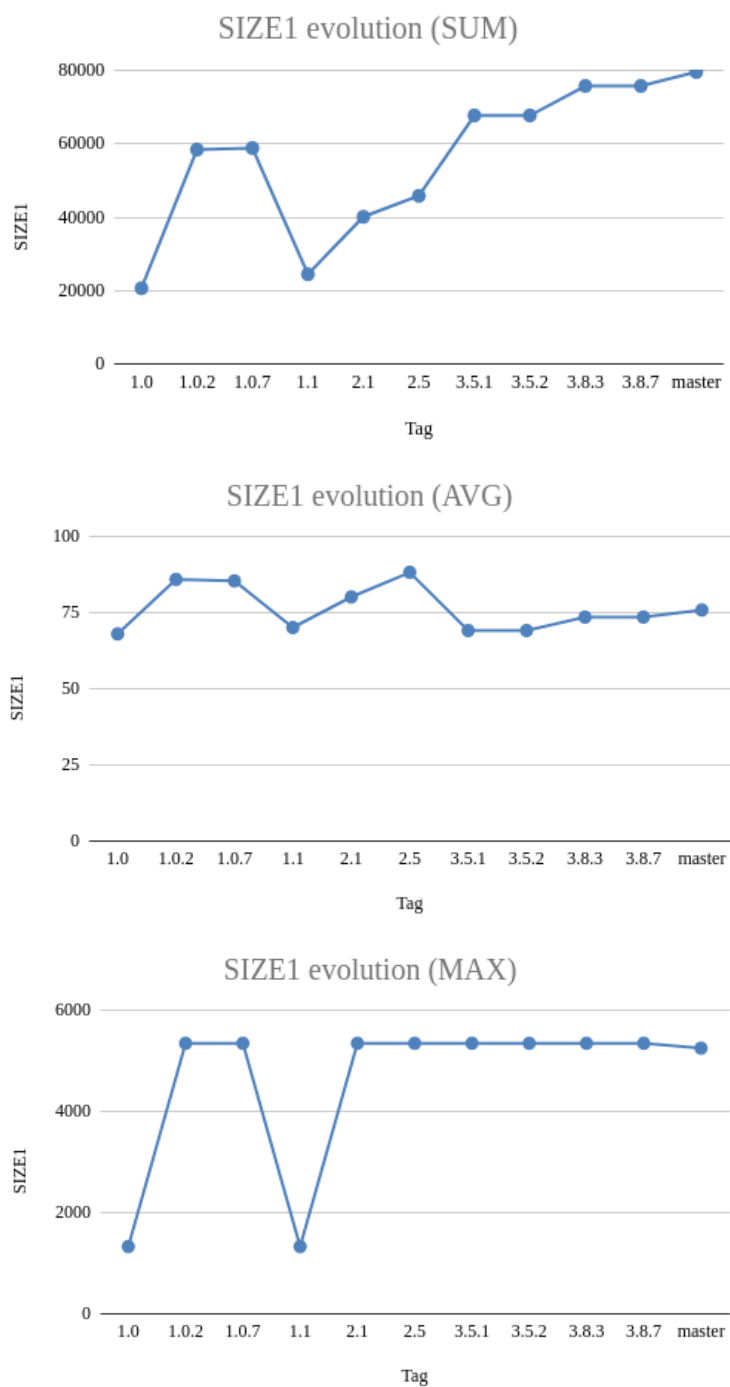
Το AVG παρουσιάζει μείωση από 58.33 στην έκδοση 1.0.2 σε 48.42 στην έκδοση 3.8.7, πράγμα που υποδηλώνει βελτίωση στη συνοχή των μεθόδων στις κλάσεις. Ωστόσο, το AVG παραμένει σχετικά υψηλό, γεγονός που ενδέχεται να υποδεικνύει ότι η συνοχή δεν έχει επιτευχθεί. Παράλληλα, η τιμή MAX παραμένει υψηλή, φτάνοντας τις 24976 στην τελική έκδοση, υποδεικνύοντας την ύπαρξη κλάσεων με χαμηλή συνοχή, γνωστών και ως "God Classes". Αυτές οι κλάσεις έχουν υπερβολικές ευθύνες και λειτουργίες που δεν συνδέονται μεταξύ τους, γεγονός που καθιστά τη συντήρησή τους δύσκολη για αυτό και είναι υποψήφιες για refactoring, ώστε να βελτιωθεί η συνοχή και η διαχείριση του κώδικα.

6. WMC* (Weighted Methods per Class)



Στην έκδοση 3.5.1, το MAX μειώνεται στο 20, γεγονός που υποδεικνύει ότι υπήρξε βελτίωση στην αρχιτεκτονική του συστήματος και στην οργάνωση των κλάσεων πιθανών μέσω διαδικασίες refactoring. Στην τελική έκδοση παρατηρείται μια μικρή αύξηση του MAX στο 21,5. Αν και αυτή η αύξηση δεν είναι άμεσα ανησυχητική, εντούτοις απαιτεί προσοχή και παρακολούθηση για να διασφαλιστεί ότι δεν θα επαναληφθεί τάση υπερφόρτωσης των κλάσεων.

7. SIZE1 (Μέγεθος Κώδικα)



Το άθροισμα (SUM), παρουσιάζει σταδιακή αύξηση από 20.739 στην έκδοση 1.0 σε 75.683 στην έκδοση 3.8.7, κάτι που είναι αναμενόμενο λόγω της ενσωμάτωσης νέων λειτουργιών. Η μέγιστη τιμή (MAX) παραμένει σταθερή στις 5.351 γραμμές, γεγονός που υποδηλώνει την ύπαρξη ορισμένων πολύ μεγάλων κλάσεων, οι οποίες θα μπορούσαν να επηρεάζουν αρνητικά τη διαχειρισιμότητα του κώδικα και ίσως χρειάζονται αναδιάρθρωση. Παράλληλα, η μέση τιμή (AVG) ανά κλάση αυξάνεται ελαφρώς από 68 σε 73,48, δείχνοντας μια τάση προς μεγαλύτερες και πιθανώς πιο γενικευμένες κλάσεις, κάτι που μπορεί να δυσκολεύει τη συντήρηση αν δεν τηρούνται οι αρχές του αντικειμενοστραφούς σχεδιασμού. Η συνεχής παρακολούθηση είναι κρίσιμη για να διασφαλιστεί ότι η ανάπτυξη παραμένει ευέλικτη και ότι αποφεύγεται η συσσώρευση τεχνικού χρέους.

Ευρήματα

- **NOC (Number of Classes):**

Ο αριθμός των κλάσεων (NOC) αυξάνεται σταδιακά σε κάθε έκδοση, κάτι που είναι φυσιολογικό καθώς προστίθενται νέες λειτουργικότητες και επεκτείνεται το σύστημα. Ωστόσο, ο μέσος αριθμός γραμμών ανά κλάση (AVG SIZE1), δείχνει μικρές διακυμάνσεις, αλλά συνολικά παραμένει σε παρόμοιο εύρος (~68 έως ~75 γραμμές). Αυτό μπορεί να ερμηνευθεί ως θετικό, καθώς δείχνει ότι η αύξηση της λειτουργικότητας δεν οδηγεί σε υπερβολική πολυπλοκότητα στις κλάσεις.

- **WMC* (Weighted Methods per Class):**

Παρατηρείται μείωση στις μέσες τιμές (AVG WMC*) του WMC*, γεγονός που δείχνει ότι οι περισσότερες κλάσεις παραμένουν διαχειρίσιμες και δεν περιέχουν υπερβολικό αριθμό μεθόδων. Ωστόσο, η μέγιστη τιμή (MAX WMC*) διατηρείται σταθερή ή μειώνεται ελαφρώς, γεγονός που δείχνει την ύπαρξη ορισμένων πολύπλοκων κλάσεων. Αυτό μπορεί να επηρεάζει αρνητικά τη συντηρησιμότητα του κώδικα.

- **CBO (Coupling Between Objects):**

Η μέση τιμή (AVG CBO) του CBO εμφανίζει μικρή μείωση στις τελευταίες εκδόσεις, γεγονός που υποδεικνύει βελτιώσεις στη χαλάρωση της σύζευξης μεταξύ κλάσεων. Ωστόσο, η μέγιστη τιμή (MAX CBO) παραμένει σταθερή, δείχνοντας ότι ορισμένες κλάσεις εξακολουθούν να έχουν ισχυρή εξάρτηση από άλλες.

- **LCOM (Lack of Cohesion in Methods):**

Η συνολική και η μέση τιμή (AVG LCOM) του LCOM αυξάνονται σε ορισμένες εκδόσεις, ειδικά στις πιο πρόσφατες. Αυτό αποκαλύπτει την ύπαρξη ορισμένων κλάσεων με χαμηλή συνοχή, κάτι που μπορεί να υποδεικνύει την ανάγκη για αναδιάρθρωση αυτών των κλάσεων.

- **DIT (Depth of Inheritance Tree) και NOCC (Number of Children):**

Οι τιμές του DIT παραμένουν σταθερές και σχετικά χαμηλές, υποδεικνύοντας μια ισορροπημένη ιεραρχία κληρονομικότητας. Παράλληλα, η μικρή αύξηση στις μέγιστες τιμές του NOCC δείχνει ότι υπάρχουν κλάσεις με περισσότερα παιδιά, αλλά η αύξηση αυτή δεν είναι ανησυχητική.

Με βάση τα παραπάνω ευρήματα, οι περιοχές που χρήζουν προσοχής περιλαμβάνουν τη σύζευξη και τη συνοχή, ενώ η συνολική ποιότητα φαίνεται να βελτιώνεται με την πάροδο των εκδόσεων.

Συζήτηση για Ενδιαφέροντα Αποτελέσματα

Η εξέλιξη του OpenRefine αποτυπώνει μια συνεπή και ισορροπημένη ανάπτυξη λογισμικού, η οποία αντανακλάται σε πολλές από τις αναλυθείσες μετρικές. Η σταδιακή αύξηση του αριθμού των κλάσεων (NOC) και των γραμμών κώδικα (SIZE1) υποδεικνύει τη συνεχή προσθήκη νέας λειτουργικότητας και την επέκταση του συστήματος. Ωστόσο, η σταθερότητα στον μέσο αριθμό γραμμών ανά κλάση και η μείωση της πολυπλοκότητας των κλάσεων (WMC*) δείχνουν ότι η ανάπτυξη γίνεται με τρόπο που διατηρεί τη συνοχή και τη διαχειρισιμότητα του κώδικα.

Μια αξιοσημείωτη βελτίωση εντοπίζεται στη μείωση της σύζευξης μεταξύ κλάσεων (CBO). Η πτωτική τάση του CBO αποτελεί ένδειξη μιας πιο αρθρωτής και ανεξάρτητης αρχιτεκτονικής σχεδίασης, γεγονός που βελτιώνει τη συντηρησιμότητα και τη δυνατότητα επαναχρησιμοποίησης του κώδικα. Παράλληλα, η σταθερότητα στην ιεραρχία της κληρονομικότητας (DIT) και στον αριθμό των παιδιών ανά κλάση (NOCC) υποδηλώνει ότι η δομή του συστήματος παραμένει ισορροπημένη, αποφεύγοντας υπερβολική πολυπλοκότητα που θα μπορούσε να οδηγήσει σε δυσκολίες κατανόησης και συντήρησης.

Ένα ενδιαφέρον σημείο είναι η αύξηση της έλλειψης συνοχής των μεθόδων (LCOM) στις νεότερες εκδόσεις, με την υψηλότερη μέση τιμή να εμφανίζεται στην τρέχουσα έκδοση master. Η μειωμένη συνοχή σε κάποιες κλάσεις μπορεί να υποδηλώνει τάση προς υπερβολική γενίκευση. Αυτό αποτελεί έναν κρίσιμο τομέα που χρήζει περαιτέρω ανάλυσης και πιθανώς ανασχεδιασμού, ώστε να διασφαλιστεί ότι οι κλάσεις παραμένουν εστιασμένες στις ευθύνες τους.

Συνολικά, η ποιότητα του OpenRefine φαίνεται να εξελίσσεται θετικά, με βελτιώσεις στη διαχείριση πολυπλοκότητας, τη μείωση της σύζευξης και τη διατήρηση της ισορροπίας στην ιεραρχία του συστήματος. Ωστόσο, οι περιοχές που υποδεικνύουν αυξημένη ασυνέπεια στις μεθόδους απαιτούν προσοχή, καθώς η συνοχή είναι κρίσιμη για τη διατηρησιμότητα και την ευκολία κατανόησης του κώδικα. Η παρακολούθηση αυτών των μετρικών και η εφαρμογή κατάλληλων στρατηγικών βελτίωσης θα ενισχύσει περαιτέρω την ποιότητα του λογισμικού στο μέλλον.

3. Αναγνώριση Προβλημάτων Ποιότητας

Επεξήγηση Μεθοδολογίας Αναγνώρισης Προβλημάτων Ποιότητας

Η διαδικασία αναγνώρισης προβλημάτων ποιότητας βασίζεται στη χρήση της ανάλυσης Hotspot, η οποία επικεντρώνεται στον εντοπισμό κρίσιμων περιοχών κώδικα με βάση συγκεκριμένες μετρικές ποιότητας. Η μεθοδολογία περιλαμβάνει τα εξής βήματα:

- Συλλογή Μετρικών Κώδικα:** Οι μετρικές κώδικα συλλέγονται με την χρήση του λογισμικού MetricsCalculator.jar. Οι κύριες μετρικές που χρησιμοποιούνται είναι:
 - Σύζευξη Μεταξύ Αντικειμένων (CBO):** Καταγράφει τον βαθμό εξάρτησης μεταξύ διαφορετικών κλάσεων, επηρεάζοντας τη συντηρησιμότητα.
 - Έλλειψη Συνοχής (LCOM):** Μετρά τη συνοχή των μεθόδων μιας κλάσης, δείχνοντας την εσωτερική της συνοχή.
 - Σταθμισμένη Πολυπλοκότητα Μεθόδων (WMC*):** Υποδεικνύει την πολυπλοκότητα των μεθόδων.
 - Μέγεθος Κλάσης (SIZE1):** Εκτιμά το μέγεθος των κλάσεων σε γραμμές κώδικα ή άλλες μονάδες μέτρησης.
- Ανάλυση Hotspot:** Η ανάλυση εντοπίζει κλάσεις με ακραίες τιμές στις παραπάνω μετρικές. Αυτές οι τιμές συχνά υποδεικνύουν περιοχές που απαιτούν προσοχή λόγω πιθανών προβλημάτων ποιότητας.
- Καταγραφή και Αξιολόγηση Ευρημάτων:** Οι προβληματικές κλάσεις καταγράφονται και αναλύονται περαιτέρω για την εξαγωγή συμπερασμάτων και την πρόταση βελτιώσεων.

Ευρήματα

Η ανάλυση Hotspot μας έδωσε τον παρακάτω πίνακα:

Name	CBO	LCOM	WMC*	SIZE1	CATEGORY
main/src/com/google/refine/clustering/binning/Metaphone3.java	1	24753	3	5251	TOP
modules/core/src/main/java/com/google/refine/expr/util/CalendarParser.java	6	596	10,85	1238	TOP
main/src/com/google/refine/importers/WikitextImporter.java	15	745	10	1070	TOP
modules/core/src/main/java/com/google/refine/importing/ImportingUtilities.java	11	1083	3	987	TOP
modules/core/src/main/java/com/google/refine/model/recon/StandardReconConfig.java	19	139	7,07	641	TOP
main/tests/server/src/com/google/refine/importing/ImportingUtilitiesTests.java	13	282	2,24	588	TOP
main/tests/server/src/com/google/refine/importers/JsonImporterTests.java	9	595	1,39	564	TOP
modules/core/src/test/java/com/google/refine/model/recon/StandardReconConfigTests.java	9	191	1,95	548	TOP
modules/core/src/main/java/com/google/refine/exporters/CustomizableTabularExporterUtilities.java	18	9	13	473	TOP
extensions/wikibase/src/org/openrefine/wikibase/operations/PerformWikibaseEditsOperation.java	21	12	7	447	TOP
main/src/com/google/refine/importers/JsonImporter.java	9	93	8	445	TOP
main/src/com/google/refine/operations/column/ColumnAdditionByFetchingURLsOperation.java	19	65	3	444	TOP
main/tests/server/src/com/google/refine/importers/XmlImportUtilitiesTests.java	10	528	1	437	TOP
modules/core/src/main/java/com/google/refine/browsing/facets/ScatterplotFacet.java	16	456	3	422	TOP
modules/core/src/main/java/com/google/refine/model/changes/DataExtensionChange.java	11	10	7	398	TOP
main/src/com/google/refine/commands/row/GetRowsCommand.java	13	31	4	397	TOP
main/src/com/google/refine/importers/XmlImporter.java	9	76	6,58	381	TOP
main/tests/server/src/com/google/refine/operations/recon/ExtendDataOperationTests.java	14	120	2,06	361	TOP
main/src/com/google/refine/operations/recon/ExtendDataOperation.java	14	9	3	355	TOP
modules/core/src/main/java/com/google/refine/io/FileProjectManager.java	10	465	2,29	355	TOP
modules/core/src/main/java/com/google/refine/model/changes/ColumnSplitChange.java	10	6	7,75	353	TOP
main/src/com/google/refine/commands/recon/ReconJudgeOneCellCommand.java	13	0	21,5	343	TOP
extensions/database/src/com/google/refine/extension/database/DatabaseImportController.java	12	91	2,47	331	TOP
modules/core/src/main/java/com/google/refine/importers/tree/XmlImportUtilities.java	9	120	3	325	TOP
main/tests/server/src/com/google/refine/importers/XmlImporterTests.java	10	378	1	308	TOP
main/src/com/google/refine/operations/column/ColumnSplitOperation.java	16	70	2,7	295	TOP
modules/core/src/main/java/com/google/refine/ProjectManager.java	9	595	1	295	TOP
modules/core/src/main/java/com/google/refine/browsing/facets/RangeFacet.java	15	191	3	284	TOP
extensions/database/src/com/google/refine/extension/database/DatabaseUtils.java	5	78	4	283	TOP
modules/core/src/main/java/com/google/refine/clustering/knn/kNNClusterer.java	24	43	6	279	TOP
extensions/wikibase/src/org/openrefine/wikibase/editing/EditBatchProcessor.java	9	21	6,75	267	TOP
modules/core/src/main/java/com/google/refine/browsing/facets/ListFacet.java	18	142	1	260	TOP
modules/core/src/main/java/com/google/refine/commands/Command.java	10	630	2	259	TOP

Name	CBO	LCOM	WMC*	SIZE1	CATEGORY
modules/core/src/main/java/com/google/refine/clustering/binning/BinningClusterer.java	18	33	4,93	254	TOP
main/src/com/google/refine/templating/Template.java	10	36	4	240	TOP
modules/core/src/test/java/com/google/refine/RefineTest.java	16	325	1	239	TOP
modules/core/src/main/java/com/google/refine/browsing/facets/TimeRangeFacet.java	16	174	2	231	TOP
modules/core/src/main/java/com/google/refine/model/RecordModel.java	11	45	4,36	229	TOP
main/src/com/google/refine/commands/expr/PreviewExpressionCommand.java	9	1	6	223	TOP
modules/core/src/main/java/com/google/refine/importing/ImportingManager.java	9	139	1,92	208	TOP
extensions/wikibase/tests/src/org/openrefine/wikibase/schema/WbStatementExprTest.java	13	137	1,94	207	TOP
main/src/com/google/refine/commands/browsing/GetScatterplotCommand.java	14	11	4,5	200	TOP
main/src/com/google/refine/commands/recon/ReconClearOneCellCommand.java	11	0	8	194	TOP
modules/core/src/main/java/com/google/refine/model/changes/ColumnReorderChange.java	9	6	5	188	TOP
main/src/com/google/refine/operations/cell/MassEditOperation.java	13	10	5	180	TOP
main/tests/server/src/com/google/refine/exporters/TemplatingExporterTests.java	9	78	1	178	TOP

Στον πίνακα αυτό παρατηρούμε τα παρακάτω προβλήματα ποιότητας στις κλάσεις του λογισμικού:

1. Υψηλή Σύζευξη (CBO):

- Η κλάση **PerformWikibaseEditsOperation** έχει CBO 21, υποδεικνύοντας ισχυρές εξαρτήσεις με άλλες κλάσεις.
- Παρόμοια, η **StandardReconConfig** παρουσιάζει υψηλή CBO (19), γεγονός που την καθιστά ευάλωτη σε αλλαγές.

2. Χαμηλή Συνοχή (LCOM):

- Η κλάση **Metaphone3** έχει εξαιρετικά υψηλό LCOM (24,753), δείχνοντας πολύ χαμηλή συνοχή.

3. Πολυπλοκότητα Μεθόδων (WMC*):

- Η κλάση **ScatterplotFacet** έχει WMC* 16, υποδεικνύοντας μεγάλη πολυπλοκότητα που δυσχεραίνει την κατανόηση και συντήρηση.

4. Μέγεθος Κλάσης (SIZE1):

- Η κλάση **Metaphone3** έχει SIZE1 5,251, υποδεικνύοντας υπερβολικό μέγεθος που πιθανώς απαιτεί διάσπαση.

Συζήτηση Ευρημάτων

Η υψηλή σύζευξη (CBO) που παρατηρείται σε κλάσεις όπως η **PerformWikibaseEditsOperation** και η **StandardReconConfig** αποτελεί ένδειξη υπερβολικής εξάρτησης από εξωτερικές οντότητες. Αυτός ο βαθμός σύζευξης όχι μόνο αυξάνει την πολυπλοκότητα του συστήματος αλλά και δυσχεραίνει τη συντήρηση του κώδικα, καθώς οποιαδήποτε αλλαγή σε μία εξαρτώμενη κλάση μπορεί να προκαλέσει αλυσιδωτές επιδράσεις. Η βελτιστοποίηση της σύζευξης θα μπορούσε να επιτευχθεί ίσως με την εισαγωγή σχεδιαστικών προτύπων.

Η χαμηλή συνοχή (LCOM) που χαρακτηρίζει την κλάση **Metaphone3** είναι αξιοσημείωτη. Το υψηλό LCOM υποδηλώνει ότι οι μέθοδοι της κλάσης είναι αποσυνδεδεμένες και πιθανώς εξυπηρετούν πολλαπλές λειτουργίες. Αυτή η κατάσταση συνιστά παραβίαση της αρχής της Μοναδικής Αρμοδιότητας και επιβαρύνει τόσο την κατανόηση όσο και τη συντηρησιμότητα του κώδικα. Η διάσπαση της κλάσης σε μικρότερες, περισσότερο στοχευμένες μονάδες είναι απαραίτητη για τη βελτίωση της συνοχής και της επαναχρησιμοποίησης του κώδικα.

Η αυξημένη πολυπλοκότητα μεθόδων (WMC*) στις κλάσεις όπως η **ScatterplotFacet** και το μεγάλο μέγεθος (SIZE1) στην **Metaphone3** υποδεικνύουν περαιτέρω περιοχές προβληματικής ποιότητας. Η υψηλή πολυπλοκότητα μεθόδων αυξάνει το φορτίο που απαιτείται για την κατανόηση του κώδικα. Η μείωση της

πολυπλοκότητας μέσω διάσπασης περίπλοκων μεθόδων και ο περιορισμός του μεγέθους των κλάσεων είναι κρίσιμα βήματα για τη βελτίωση της συνολικής ποιότητας και της επεκτασιμότητας του συστήματος.

Συμπεράσματα και Προτάσεις

Η ανάλυση δείχνει την ανάγκη για:

1. **Μείωση σύζευξης** μέσω αποσύνδεσης εξαρτώμενων κλάσεων.
2. **Αύξηση συνοχής** με διάσπαση μεγάλων κλάσεων σε μικρότερες.
3. **Απλοποίηση μεθόδων** για βελτίωση της κατανοησιμότητας.
4. **Refactoring μεγάλων κλάσεων** για μείωση του μεγέθους και της πολυπλοκότητας.

Η εφαρμογή αυτών των βελτιώσεων θα ενισχύσει τη συντηρησιμότητα και επεκτασιμότητα του λογισμικού.