Delivery 1:

Procedimento:

Spiegare procedimento, anno dei bug(parto dall anno primo bug fino a ultima commit. Linking basato su REGEX.

Classi:

E’ stata studiata la frequenza dei numeri per capire la distribuzione.

Dati:

E’ stato calcolato il numero di:

commit con id del bug

commit senza id

commit con altri id, ticket non bug e ticket di altri progetti(HIVE-5783,PARQUET-84,AVRO-1228 ecc)

Bug con commit associate

Bug senza commit associate.

ULTIMO AGGIORNAMENTO 07/05/2020

PROGETTI:

PARQUET-MR, ARROW, AVRO-TESTING

BRANCH PARQUET-MR: 35

BRANCH ARROW: 5

BRANCH ARROW-TESTING: 2

CASO 1:

repository parquet con branch di default

Size PARQUET commits: 248

Size WITHOUT TICKET commits: 1632

Size OTHER TICKET commits: 299

Size bugs with commit : 236

Size bugs without commit: 318

CASO 2:

tutti i repository con branch default

Size PARQUET commits: 461

Size WITHOUT TICKET commits: 1922

Size OTHER TICKET commits: 6417

Size bugs with commit : 415

Size bugs without commit: 139

Numero mesi 72

CASO 3:

tutti i repository con tutti i branch

Size PARQUET commits: 2037

Size WITHOUT TICKET commits: 41257

Size OTHER TICKET commits: 21805

Size bugs with commit : 418

Size bugs without commit: 136

Numero mesi 72

somma dei ticket presenti in ogni mese: 418

CASO 4:

REPOSITORY PARQUET CON TUTTI I BRANCH

Size PARQUET commits: 1157

Size WITHOUT TICKET commits: 40115

Size OTHER TICKET commits: 1313

Size bugs with commit : 239

Size bugs without commit: 315

Numero mesi 71

somma dei ticket presenti in ogni mese: 239

Ci chiediamo quindi quali sono queste 3 commit che sono negli altri branch ma non in quello di default.

DELIVERABLE 2

Classi java. Numero release.

Elimino ov=iv=fv

Fv=ov

Se OV sta prima dell iv quando AV è specificato, porto IV indietro e considero affette anche altre release. Li considero comunque nel proportion.

E’ GIUSTO METTERE AFFECTED VERSION DOVE NON CI STA?

Non esiste relase fixed RIMUOVO-bug: BOOKKEEPER-365 commit id: 2018-02-03T00:26:45

Release:

12317843 2011-12-07T00:00 true

12319145 2012-06-13T00:00 true

12321757 2013-01-16T00:00 true

12320244 2013-01-19T00:00 true

12323840 2013-03-27T00:00 true

12324601 2013-10-10T00:00 true

12323267 2014-02-02T00:00 true

12325569 2014-06-27T00:00 true

12328755 2015-05-24T00:00 true

12333529 2015-11-17T00:00 true

12325566 2016-05-16T00:00 true

12335340 2017-08-10T00:00 true

12341342 2017-09-10T00:00 false

12335749 2017-11-10T00:00 false

COMMIT FORK INIZIALI:

9ea37773fa07e8e1c16e654020ae34c3d6564963

7e7fcf6618417a0846baae4b6ee84778571a3556

5a71ed88116fcdffcccdb99eb79055912c46a85e

7e63cff1c110b7b9435a88ab8ab321659057c298

1e788b301f998a84d3a12a73b631473c82ce0697

26680c1507949bd8678af987a60fb3f5342fb1a6

Ignorare commenti e meta-cambiamenti.

Ignorare copie e rename.

After prototyping B-SZZ and AG-SZZ, we noticed that meta-changes (i.e., changes that are not related to source code modification) were being reported as bug-introducing. We observe three types of potential bug-introducing metachanges: (1) branch-changes, (2) merge-changes, and (3) property-changes. A branch change is a meta-change that copies the project state from one SCM branch to another (e.g., from the trunk to a feature branch). A merge-change is a meta-change that applies change activity from one branch to another (e.g., from a feature branch to the trunk). Finally, a property-change is a meta-change that only impacts file properties that are stored in the SCM (e.g., end of line property). Ideally, meta-changes should not be flagged as bugintroducing, since they do not change system behaviour, and hence, cannot introduce bugs.

Per la scelta del ramo master o ramo branch vengono prese in considerazione due soluzioni.

Una col master e una con tutti i branch saltando il nodo di merge. Quando prendo tutti i branch, devo prendere come insieme di file qello fornito dall’ultima commit sul master prima della versione. Questa cosa non è stata realizzata per mancanza di tempo, però vengono buttati i merge. Cosa cambia dal non considerare tutti i branch? Solo il numero di commit in piu effettuate nei vari branch.

Problema Bugginess: devo prendere le classi nelle commit buggy?

DATASET:

1)proportion increment

2)movin window

3)all branch

4)default ramo

MANCA DA CARCOLà SOLO LE METRICHE

Numero di commit: da contare anche copy rename e add?

Numero di fix su altri file e quello mio l’aggiungo?

OPERAZIONE DI RENAME: incremento NR ? NFIX?—NO CHANGE SET

OPERAZIONI DI COPY: copio solo la bugginess ?—NO CHANGE SET

OPERAZIONE DI ADD: size, size touched, NFIX? , NR ?

La bugginess la calcolo solo nelle MODIFY. Non ha senso calcolarla nella create.

Java approssima per difetto.

PROBLEMA: nell age pesato, vado avanti con age, ma resto fermo con loc added

WEKA:

modelli

- random forest

-ibk

-nayve bayes

-j48

-adaBoosting

-decision tree

-svm

Feature selection:

-wrapper

-pca

-filtro

Tuning parametri?

Bilanciamento

Walk forward

Cross fold validation

GRAFICI