TWO WAY CLASSIFICATION APPROACHE OF STATISTICAL BUG REPORTS

Pankaj Kumar[1], Mr.Nishant Anand[2]

[1]Student, CSE Department, CBS Group Of Institutions, Jhajjar, Haryana

[2]Professor and Head of CSE Department, CBS Group Of Institutions, Jhajjar, Haryana

Abstract— It is very important to analyze the bugs that is collect by various sources (like from users ,developers that submit the bug report and testers that classified the bug reports)into right categories .In this paper we introduce a bug tracking system tools that analyze the bugs in two different ways. That help to classify bug reports .In first ways we introduce Naïve Bayes classification process by which we find probability of bugs categories on the basis of attributes of category of bug dataset .And in second ways we use natural language processing in summary attributes of bug dataset. By compare the results of two methods we able to classify more accurately and efficiently.

Keywords— Bug Tracking System, Natural language processing, Stemming, Naïve Bayes classifications,SBR,NSBR.

INTRODUCTION

A software bug is an error, flaw, mistake, failure, or fault in a computer program or system that produces an incorrect or unexpected result, or causes it to behave in unplanned ways. Most bugs arise from mistakes and errors made by people in either a program's source code or its design [1]. Bug Tracking System has life cycle as shown in following fig.

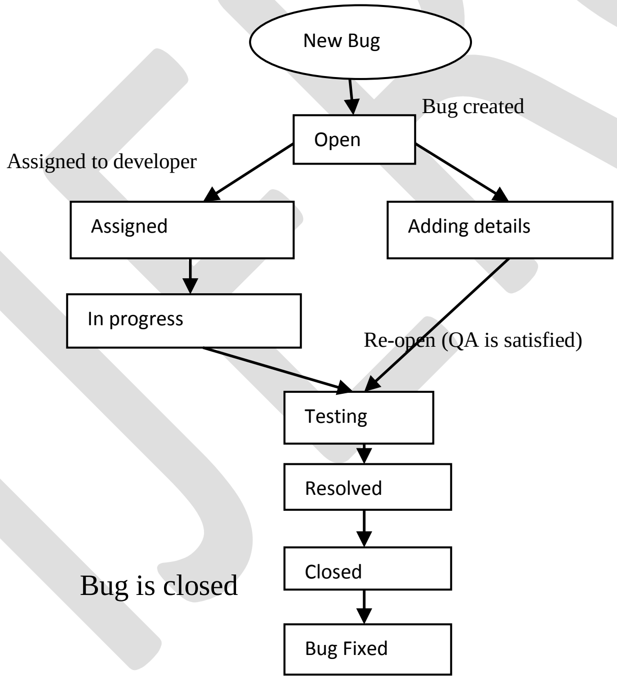


Figure 1: Bug Tracking Life Cycle [2]

Bugs move through a series of states in their lifetime. When a bug report is submit to the bug tracking system (BTS) its status is set to New Bug. Once a developer has been either assigned to or received responsibility for the report, the status is set to Assigned. When a report is closed its status is set to Resolved. It may further be marked as being verified or closed for fine. A report can be resolved in a number of ways; the resolution status in the bug report is used to record how the report was resolved. If the resolution resulted in a change to the code base, the bug is resolved as fixed. When a developer determines that the report is a copy of an existing report then it is marked as duplicate. If the developer was unable to reproduce the bug it is indicated by setting the resolution status to works-for-me. If the report describes a problem that will not be fixed, or is not an actual bug, the report is marked as wont-fix or invalid, respectively. A formerly resolved report may be reopened at a later date, and will have its status set to re-opened [3].

PROPOSED METHODOLOGY

In this methodology we load the data from Bug tracking system database that have different attributes such as product, component, status, resolution, Summary and its category state . After this we preprocess the data set by removing null values and redundant data. After that we apply bayes' rules for prediction .We considers only maximum probability of SBR(Security Bug Report) or NSBR(Non Security Bug Report). We out list minimum probability values. After gaining the probability we use natural language processing (NLP) on the summary attributes of data. We fetch the summary for surety of SBR we analyze bug report with natural language process on the summary of the reports. Now with the help of natural language processing we read line by line text and only consider security related terms like vulnerability, attack etc. We have to evaluate the bug related text and remove the stop words such as articles, prepositions, and conjunctions that are not use in text mining. For this we maintain stop words list. Get the maximum count for words occurrence. If the frequency of the words in whole dataset will give the maximum security related terms then it is SBRs and treats as SBRs. After this we Mach this result with bayes prediction if Mach then we confidentially say this reports is SBR. Following figure describes the proposed methodology.

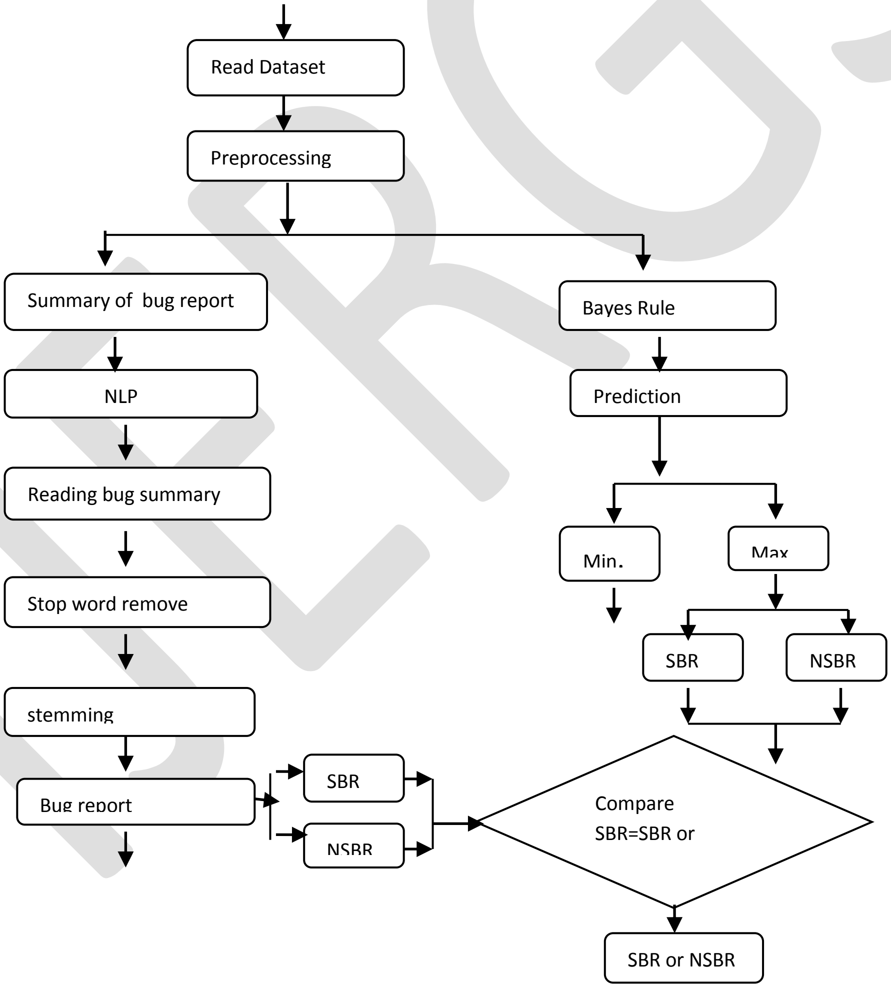
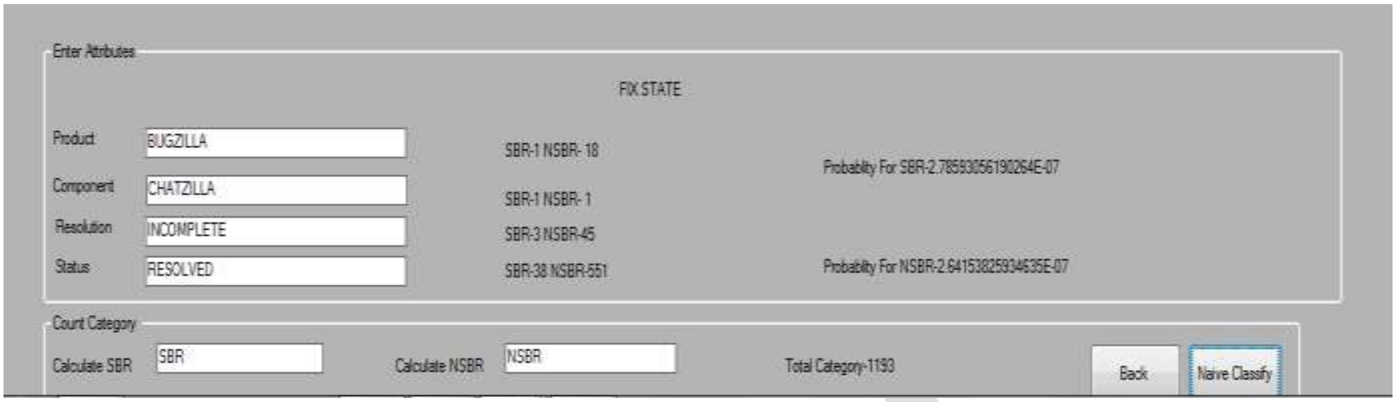


Figure 2: Design Diagram of Proposed Methodologies

RESULTS

Bug tracking dataset have different attributes such as product, component, status, resolution, Summary and its category stage.

We enter the value for product is Bugzilla, for component is Chatzilla, for Resolution is Incomplete then status is Resolved.



Product: BUGZILLA
Component: CHATZILLA
Resolution: INCOMPLETE
Status: RESOLVED

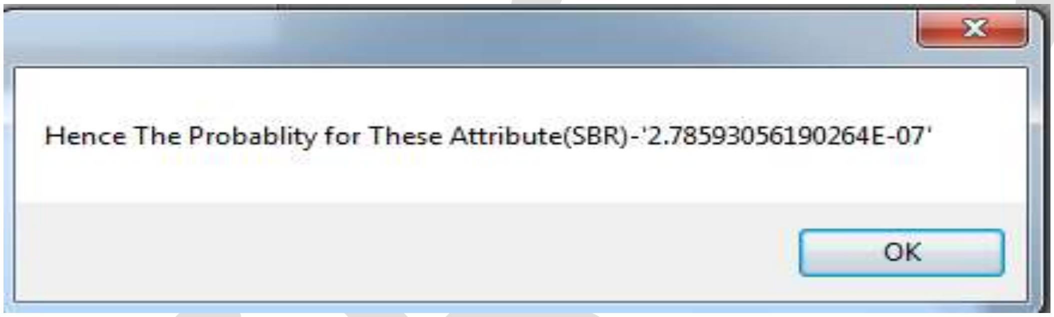
SBR-1 NSBR- 18
SBR-1 NSBR- 1
SBR-3 NSBR-45
SBR-38 NSBR-551

Probablity For SBR-2.78593056190264E-07
Probablity For NSBR-2.64153825934635E-07

Calculate SBR: SBR
Calculate NSBR: NSBR
Total Category-1193

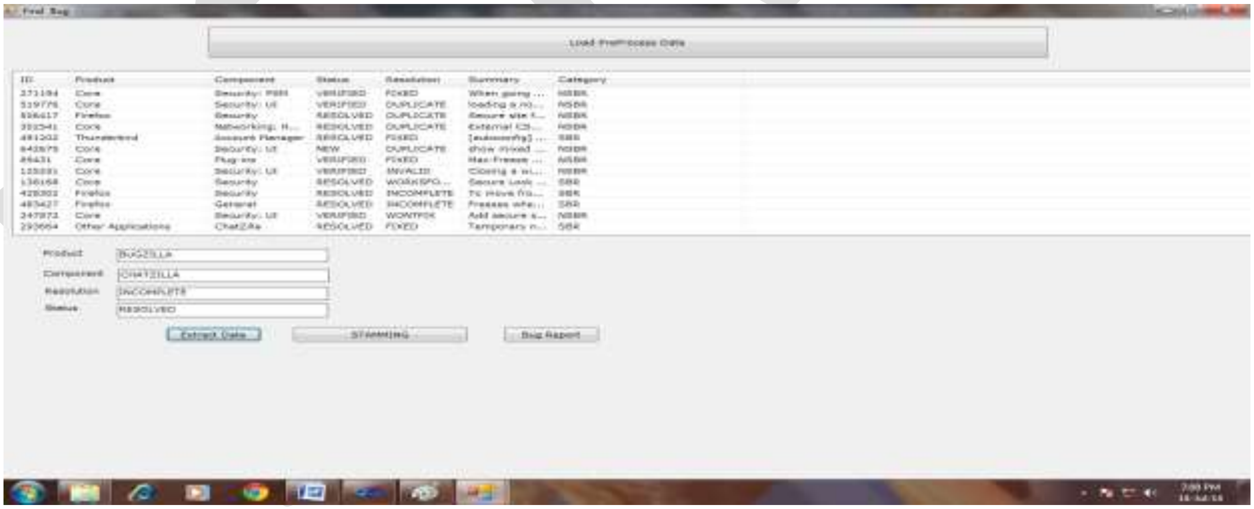
Screen shot 1: Maximum probability to whole data set

From above we get maximum probability for SBR is 2.79E-07, with respect to Bugzilla, chatzilla, Incomplete, and Resolved with respective to whole dataset.



Hence The Probablity for These Attribute(SBR)-'2.78593056190264E-07'

Screen shot 2: Indicate maximum probability of SBR

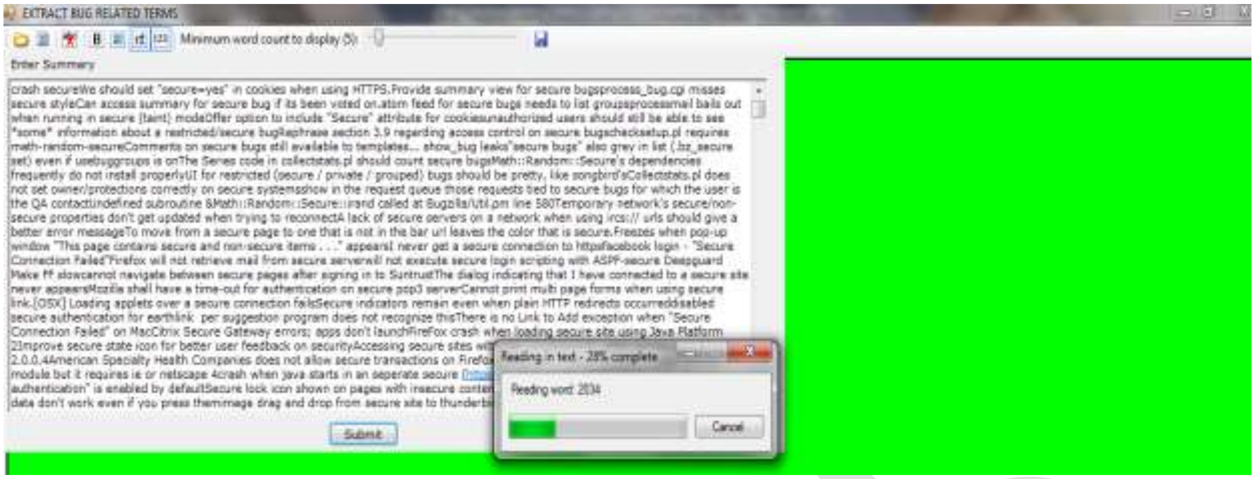


ID	Product	Component	Status	Resolution	Summary	Category
271184	Core	Security: PSM	VERIFIED	FIXED	When going ...	NSBR
519776	Core	Security: UI	VERIFIED	DUPLICATE	loading a no...	NSBR
526617	Firefox	Security	RESOLVED	DUPLICATE	Secure site f...	NSBR
392541	Core	Networking: H...	RESOLVED	DUPLICATE	External CS...	NSBR
481202	Thunderbird	Account Manager	RESOLVED	FIXED	[autoconfig] ...	SBR
642675	Core	Security: UI	NEW	DUPLICATE	show mixed ...	NSBR
86431	Core	Plug-ins	VERIFIED	FIXED	Mac Freeze ...	NSBR
125081	Core	Security: UI	VERIFIED	INVALID	Closing a wi...	NSBR
136168	Core	Security	RESOLVED	WORKSFO...	Secure Lock ...	SBR
428302	Firefox	Security	RESOLVED	INCOMPLETE	TC move fro...	SBR
483427	Firefox	General	RESOLVED	INCOMPLETE	Freezes whe...	SBR
247873	Core	Security: UI	VERIFIED	WONTFIX	Add secure s...	NSBR
293664	Other Applications	ChatZilla	RESOLVED	FIXED	Temporary n...	SBR

Product: BUGZILLA
Component: CHATZILLA
Resolution: INCOMPLETE
Status: RESOLVED

Screen shot 3: Find bug

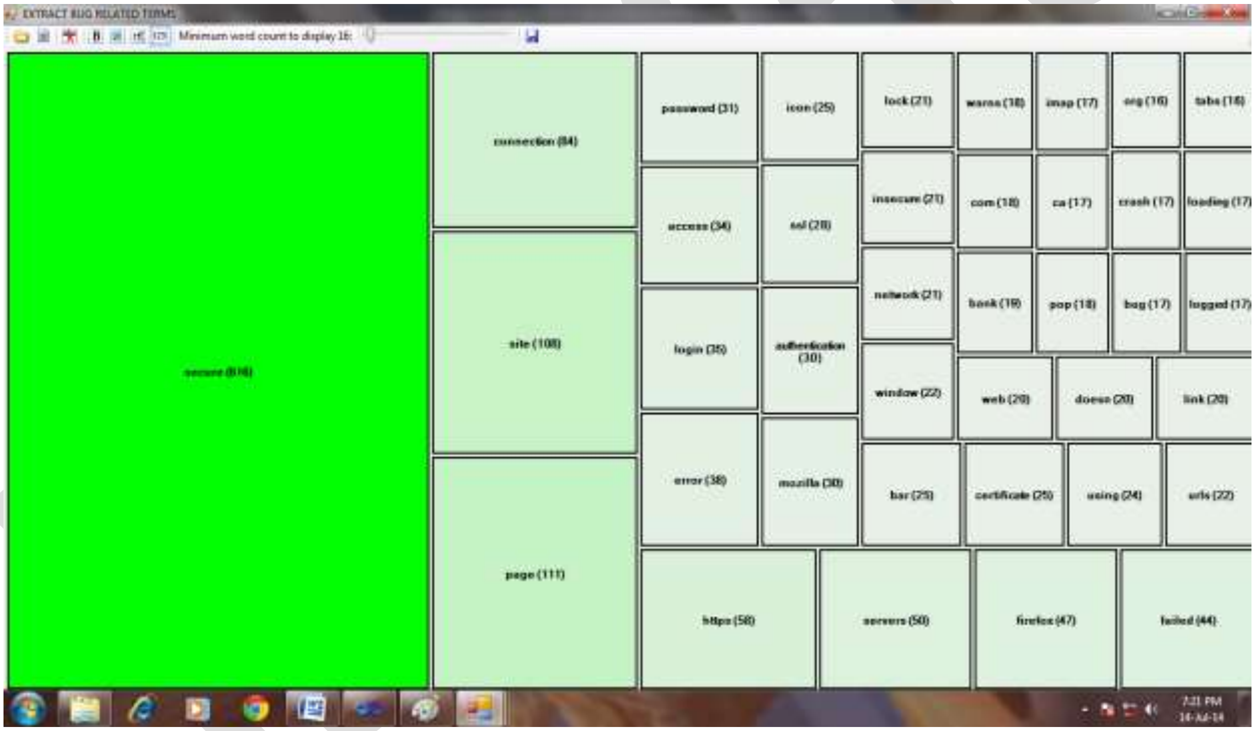
After naïve bayes classifications we use natural language processing in which we read line by line text and only consider security related terms like vulnerability, attack etc. After processing we get the terms with the frequency, in grid forms. If the terms are security related terms then the bug report for corresponding attributes are SBR.


EXTRACT BUG RELATED TERMS
Minimum word count to display (5):
Enter Summary
Submit
Reading in text - 28% complete
Reading word: 2034
Cancel

crash secureWe should set "secure=yes" in cookies when using HTTPS.Provide summary view for secure bugsprocess_bug.cgi misses secure styleCan access summary for secure bug if its been voted on.atom feed for secure bugs needs to list groupsprocessmail bails out when running in secure (taint) modeOffer option to include "Secure" attribute for cookiesunauthorized users should still be able to see "some" information about a restricted/secure bugRephrase section 3.9 regarding access control on secure bugscheckssetup.pl requires math-random-secureComments on secure bugs still available to templates... show_bug leaks"secure bugs" also grey in list (.bz_secure set) even if usebuggroups is onThe Series code in collectstats.pl should count secure bugsMath::Random::Secure's dependencies frequently do not install properlyUI for restricted (secure / private / grouped) bugs should be pretty, like songbird'sCollectstats.pl does not set owner/protections correctly on secure systemsshow in the request queue those requests tied to secure bugs for which the user is the QA contactUndefined subroutine &Math::Random::Secure::irand called at Bugzilla/Util.pm line 580Temporary network's secure/non-secure properties don't get updated when trying to reconnectA lack of secure servers on a network when using ircs:// urls should give a better error messageTo move from a secure page to one that is not in the bar url leaves the color that is secure.Freezes when pop-up window "This page contains secure and non-secure items . . ." appearsI never get a secure connection to httpsfacebook login - "Secure Connection Failed"Firefox will not retrieve mail from secure serverwill not execute secure login scripting with ASPF-secure Deepguard Make ff slowcannot navigate between secure pages after signing in to SuntrustThe dialog indicating that I have connected to a secure site never appearsMozilla shall have a time-out for authentication on secure pop3 serverCannot print multi page forms when using secure link.[OSX] Loading applets over a secure connection failsSecure indicators remain even when plain HTTP redirects occurreddisabled secure authentication for earthlink per suggestion program does not recognize thisThere is no Link to Add exception when "Secure Connection Failed" on MacCitrix Secure Gateway errors; apps don't launchFireFox crash when loading secure site using Java Platform 2Improve secure state icon for better user feedback on securityAccessing secure sites wi
2.0.0.4American Specialty Health Companies does not allow secure transactions on Firefo
module but it requires ie or netscape 4crash when java starts in an seperate secure http
authentication" is enabled by defaultSecure lock icon shown on pages with insecure conten
data don't work even if you press themimage drag and drop from secure site to thunderb

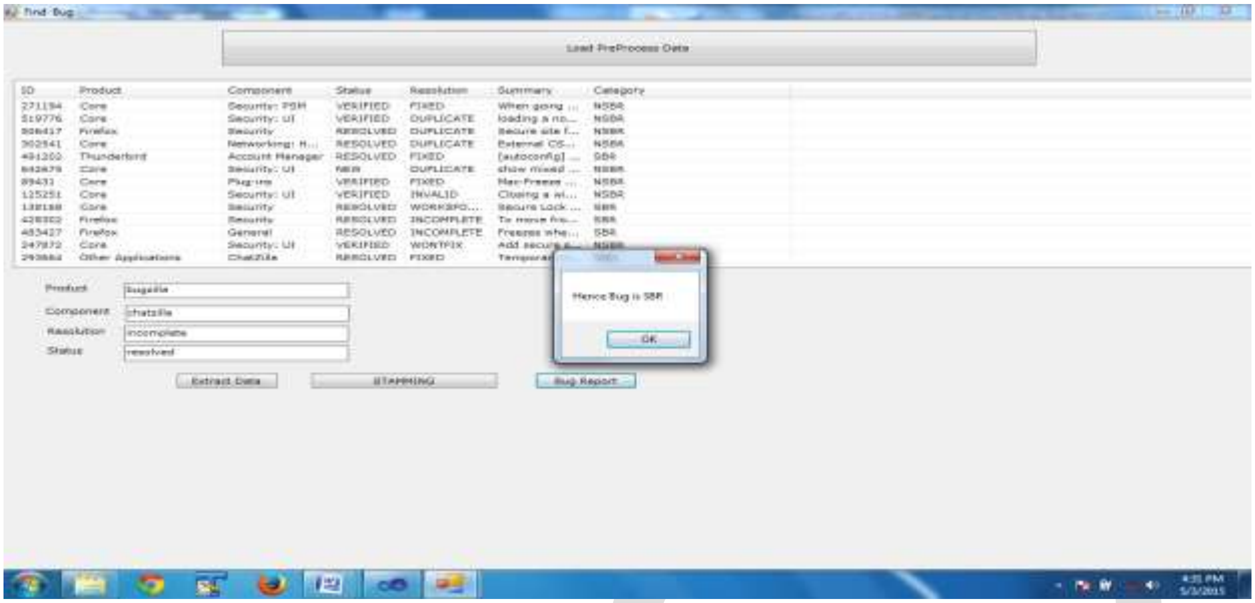
Screen shot 4: Reading Text

After natural language processing we get security related word in grid forms. We get the terms with the frequency, in grid forms. If the frequency of terms is security related terms then the bug report for corresponding attributes are SBR.



Screen shot 5: Bug related terms

Following window analyze that our bug report is SBR or NSBR. But we get SBR its means that our bug category is SBR.



ID	Product	Component	Status	Resolution	Summary	Category
271194	Core	Security: PSM	VERIFIED	FIXED	When going ...	NSBR
519776	Core	Security: UI	VERIFIED	DUPLICATE	loading a no...	NSBR
308417	Firefox	Security	RESOLVED	DUPLICATE	Secure site f...	NSBR
302541	Core	Networking: H...	RESOLVED	DUPLICATE	External CS...	NSBR
491202	Thunderbird	Account Manager	RESOLVED	FIXED	[autoconfig] ...	SBR
642679	Core	Security: UI	NEW	DUPLICATE	show mixed ...	NSBR
89431	Core	Plug-ins	VERIFIED	FIXED	Mac Freeze ...	NSBR
125251	Core	Security: UI	VERIFIED	INVALID	Closing a wi...	NSBR
138188	Core	Security	RESOLVED	WORKSFO...	Secure Lock ...	SBR
428302	Firefox	Security	RESOLVED	INCOMPLETE	To move fro...	SBR
483427	Firefox	General	RESOLVED	INCOMPLETE	Freezes whe...	SBR
247872	Core	Security: UI	VERIFIED	WONTFIX	Add secure ...	NSBR
293884	Other Applications	ChatZilla	RESOLVED	FIXED	Temporar...	

Product: bugzilla
Component: chatzilla
Resolution: incomplete
Status: resolved

Hence Bug is SBR

CONCLUSION

Current bug tracking systems do not effectively produce all of the information needed by developers. Without this information developers cannot resolve bugs in a timely method. We analyze the same bug report in summary attributes in different ways. First we analyze the bug with help of naïve bayes classification which is efficient to other classification algorithm. Natural language processing enables us to implements a more automated and more efficient bug triaging process. By use of this automated system the security engineers independently reviews each BR. If two security engineers disagreed on their evaluations of manually labeled BRs, then they discussed their differences and reached an agreeable consensus (result). By analyzing same report with different methods it increase the efficiency of the system. While implementing a range of improvements from these areas may be ideal, bug tracking systems may instead prefer to specialize, thus providing a rich set of choices. Finally we consider two mode of bug classification such as semantic based and attribute too, typically essential requirement from summary/comment point of view get fetch bug logs as needed to resolve for estimated product type in efficient time. Identify information needs in a large sample of bug reports through manual inspection.

REFERENCES

[1] Prof Dr. Muhammad Younus Javed and Hufsa Mohsin,2012 on "An Automated Approach for Software Bug Classification" in Complex, Intelligent and Software Intensive Systems (CISIS) 2012 page 414-419.

[2] ZatulAmilahShaffiei, MudianaMokhsin, SaidatulRahahHamidi (2010). Change and Bug Tracking System: AnjungPenchalaSdn. Bhd. International Journal of Computer Applications (0975 – 8887) Volume 10– No.3

[3] John Anvik, Lyndon Hiew and Gail C. Murphy Department of Computer Science University of British Columbia "Who Should Fix This Bug?"

[4] Saniat Javid Sohrawardi, Iftekhar Azam, Shazzad Hosain,"A Comparative Study of Text Classification Algorithms on User Submitted Bug Reports", 2014 IEEE.

[5] H. Witten, E. Frank, M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann; 3rd edition, 2011.

[6] E. Giger, M. Pinzger, and H. Gall. "Predicting the fix time of bugs", Proc. 2nd International Workshop on Recommendation Systems for Software Engineering (RSSE'10), May 4, 2010, Cape Town, South Africa ,ACM.

[7] R. L. Scheaffer, M. Mulekar, J. T. McClave, Probability and Statistics for Engineers, Duxbury Press, 5th edition, 2010.

[8] M.Gegick, P.Rtella,T.Xie,1012 on "Identifying security bug report via Text Mining:An Industrial Case Study"

[9] A. Podgurski, D. Leon, P. Francis, W. Masri, M. Minch, J. Sun, and B. Wang, "Automated Support for Classifying Software Failure Reports" Proc of the ICSE, pp. 465-475, 2003.

[10] Saniat Javid Sohrawardi, Iftekhar Azam, Shazzad Hosain,"A Comparative Study of Text Classification Algorithms on User Submitted Bug Reports", 2014 IEEE.

[11] I. H. Witten, E. Frank, M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann; 3rd edition, 2011.

[12] Noraida Haji Ali, Noor Syakirah Ibrahim " Porter Stemming Algorithm for Semantic Checking" IEEE Press