**Bug ID**: 001  
**Title**: System crashed when input file contain character \_ within the keyword   
**Reported Time**: 2013-03-02 12:18 EDT **Reported User**: LEI Xihan **Vender**: HW **Product**: Code-Similarity **Versions**: 0.2  
**Platform**: PC Microsoft Windows 7 x86  
**Severity**: Critical **Updated**: 2013-03-10 19:25 HKT  
**Status**: Closed **Assigned to**: YUAN Siyan

**Problem:**Instructions:  
When I followed the instructions in the readme file: input two paths of the file into the program. The program simply crashed.>java -jar cs.jar

Please enter original name: H:\original.c

Please enter target name: H:\target.c

Unexpected Result:  
LineByLine Comparar : 100.00% Exception in thread "main" Exception in thread "main" java.util.regex.PatternSyntaxException: Unmatched closing ')' near index 16

void (\w\*?)\(\w\*?)\(\w\*?)\(char\\* (\w\*?)\, int (\w\*?)\, int (\w\*?)\, (\w\*?)\(\w\*?) (\w\*?)\) \{(?:.\*?)? printf\(\"\[\%(\w\*?)\] \%\10 \(\%(\w\*?)\) (\w\*?)\. (\w\*?) (\w\*?) (\w\*?) \%\11\.\"\)\;(?:.\*?)?\}

^

at java.util.regex.Pattern.error(Pattern.java:1924)

at java.util.regex.Pattern.compile(Pattern.java:1669)

at java.util.regex.Pattern.<init>(Pattern.java:1337)

at java.util.regex.Pattern.compile(Pattern.java:1022)

at edu.hkcity.cs.Utility.replace(Utility.java:248)

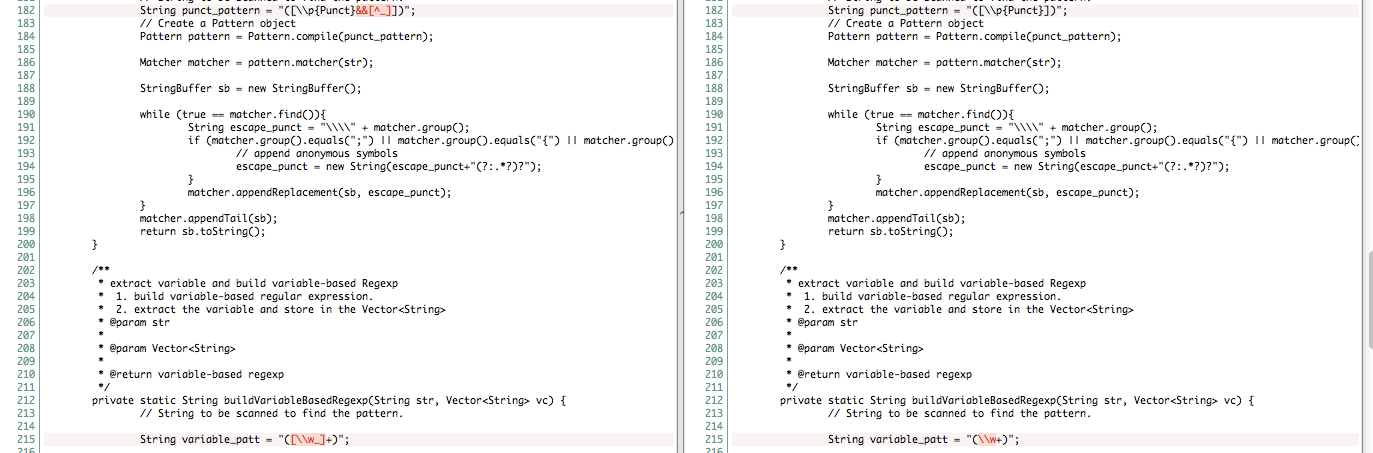
at edu.hkcity.cs.FuncByFuncComparar.compare(FuncByFuncComparar.java:64)

at edu.hkcity.cs.Input.getInput(Input.java:92)

at edu.hkcity.cs.Main.main(Main.java:13)

**Bug description after analysis or fixed:**LEI Xihan 2013-03-02 12:18 EDT  
I think the trigger of the bug may be the underscore character since the error occurs when I input a file with underscore within the variable keywords.  
  
[Solved]YUAN Siyan 2013-03-10 19:25 EDT

Yes. Xihan, you are right. The bug occurs when calling the Utilty::replace() function. From the error message, the error comes from the regular expression. It replaces the “\_” character to “\\_” which cause the exception. I have revised the code handling generating regular expression to ignore the character “\_” when escaping the punt and include the character “\_” when extracting the variable. Problem solved.

**Diff:**

**Appendix:**original.c  
<<< START

void procs\_create\_msg(char\* who, int pid, int ppid, time\_t timer) {

printf("[%s] %s (%d) created. Its parent is %d.");

}

int main(){

procs\_create\_msg(“child”, getpid(), getppid(), time(NULL));

}

<<<EOF

target.c  
<<< START

void procs\_create\_msg(char\* who, int pid, int ppid, time\_t timer) {

printf("[%s] %s (%d) created. Its parent is %d.");

}

int main(){

procs\_create\_msg(“child”, getpid(), getppid(), time(NULL));

}

<<<EOF