#GIT数据中,考虑人员,模块的千行代码BUG率,单个特征或缺陷的提交次数分析.可以在一定程度上反应人员的开发水平,及时发现杜绝开发人员的一些不良习惯.

t.engine <- read.csv("E:/code/new/ds\_engine.txt",header = FALSE,sep = "|");

t.ds <- read.csv("E:/code/new/ds\_datastudio.txt",header = FALSE,sep = "|");

t <- rbind(t.engine,t.ds);

#处理LOG数据，得到修改信息。

t.id <- t[grep("^[a-z0-9]{7}$",t$V1),];

t.detail <- t[as.numeric(row.names(t.id))-1,];

t.detail[length(t.detail$V1) + 1,1] <- t[length(t),1];

t.1 <- regexpr(" file",t.detail$V1);

t.detail$filenum <- substr(t.detail$V1,2,t.1-1);

t.2 <- regexpr(" insertion",t.detail$V1);

t.3 <- regexpr(",",t.detail$V1);

t.detail$addlines <- substr(t.detail$V1,t.3+2,t.2-1);

t.detail[t.detail$addlines == "",6] <- 0;

t.detail$temp1 <- regexpr(", ",t.detail$V1);

t.detail$temp1 <- substr(t.detail$V1,start = t.detail$temp1 +2,stop = nchar(as.character(t.detail$V1)));

t.detail$temp2 <- substr(t.detail$temp1,start = 1,stop = regexpr(" ",t.detail$temp1)-1);

t.detail$dellines <- substr(t.detail$temp1,start = regexpr(",",t.detail$temp1) +2,stop = regexpr("deletion",t.detail$temp1)-2);

t.detail[t.detail$dellines == "",]$dellines <- 0;

t.result <- t.detail[,c(5,6,9)]

result <- data.frame("commitid"=t.id$V1,"date"=t.id$V2,"name"=t.id$V3)

result$jira <- t.id$V4

result$filenume <- t.result$filenum

result$addlines <- t.result$addlines

result$dellines <- t.result$dellines

#获得JIRA ID,得到提交次数排名

result$jiraid <- substr(result$jira,1,7);

t.jira <- aggregate(x = result$commitid,by = list("jira.id" = result$jiraid,"name"=result$name),FUN = length)

t.jira <- t.jira[order(t.jira$x,decreasing = TRUE),]

hist(t.jira$x);

colnames(t.jira)[3] <- "commit.number"

write.csv(t.jira,"e:/commit\_stata.csv")

#

jira.4 <- t.jira[t.jira$commit.number >= 2,]

jira.most <- aggregate(jira.4$jira.id,by = list(jira.4$name),formula = length)

jira.most <- jira.most[order(jira.most$x,decreasing = TRUE),]

colnames(x = jira.most) <- c("name","number(commit.number >=2)");

write.csv(jira.most,"e:/jira\_2.csv")

person <- person[(person$year == 2014 & person$month >= 11) | (person$year == 2015),]

bug <- aggregate(person$bug.newnumber,by = list(person$ASSIGNEE),sum)

result$addlines <- as.numeric(result$addlines)

result[is.na(result$addlines),]$addlines <- 0