

How much and why do unlinked SATD happen?

Yasutaka Kamei

Nov 3rd, 2016

Versions:

- ver2: Fixed link_SATD_TO_UND.py, since top directory is moved from main/org/apache to src/main/org/apache

Data Load

```
library(knitr)
working_root <- "/Users/kamei/Research/techdebt/msr16_td_interest/"
setwd(working_root)
opts_knit$set(root.dir=working_root)
source("./r_scripts/data_read.r")
```

```
## Num Of Tech Debt : 393
##   Per of linking to v1 : 91.35% (=359/393)
##   Per of linking to v2 : 91.09% (=358/393)
##   Per of linking to all: 90.33% (=355/393)
```

Output 10 sampled unlinked SATD with metrics data

```
set.seed(1)
v1 <- data.s1[(data.s1$CountInput_v1 == -1 & data.s1$CountInput_v2 != -1),]
nrow(v1)
```

```
## [1] 3
```

```
samplesize <- ifelse(nrow(v1) >= 10, 10, nrow(v1))
sampled <- v1[sample(nrow(v1), samplesize), ]
```

```
v2 <- data.s1[(data.s1$CountInput_v1 != -1 & data.s1$CountInput_v2 == -1),]
nrow(v2)
```

```
## [1] 4
```

```
samplesize <- ifelse(nrow(v2) >= 10, 10, nrow(v2))
sampled <- rbind(sampled, v2[sample(nrow(v2), samplesize), ])
```

```
v1_v2 <- data.s1[(data.s1$CountInput_v1 == -1 & data.s1$CountInput_v2 == -1),]
nrow(v1_v2)
```

```
## [1] 31
```

```
samplesize <- ifelse(nrow(v1_v2) >= 10, 10, nrow(v1_v2))  
sampled <- rbind(sampled, v1_v2[sample(nrow(v1_v2), samplesize), ])  
write.table(sampled, file = "./datasets/CSV/sampled_unlinked_satd.ssv", sep="#")
```