

Guilty: Statistics Proposal

Sergio Arroutbi Braojos

February 23, 2013

1 Introduction

This document is an approachment to the different statistics that should be dumped associated to the different data which can be dumped from a project via the Guilty Tool.

Guilty is a tool written by Carlos Garcia Campos and hosted in Libre-Soft. This tool allows dumping, in different output formats, the result of a typical "blame" control version subcommand to dump for a particular file the author, date and commit of the different lines contained in itself.

2 Guilty

From MetricsGrimoire perspective, although some extensions analyse the lines of code (LOC) being changed on a particular commit, there is not a strict control regarding the LOC hosted on a particular repository.

Taking this lack into account, there is an opportunity to take profit from existing tools that allow a deep inspection of the LOC changed on a particular file, the commit associated to it, the user who performed it and the date where it was completed.

As stated previously, Guilty tool, hosted on Libresoft Git repository, allows performing this type of analysis, LOC oriented, and providing information for each line of code having to do with SCM.

In this aspect, the main purpose will be, **in a first stage, to introduce Guilty application in MetricsGrimoire forge**, allowing to provided

additional information around metrics, including to dump information to a database that will keep different entries, one per line of code, containing the author, the date, the file, the line number, the last revision.

Regarding metrics to be extracted from Guilty, next list shows some examples of the different possibilities:

- LOC changes by file and/or commit in a certain period
- LOC changes by file and/or user in a certain period
- LOC changes on a certain preconfigured period
- LOC changes multiplied per user in a certain period (in order to prioritize activity from different developers)
- Proportion of the source code (%) unchanged in a certain period

In order to enumerate the different kind of statistics that can be generated with Guilty, first some categories will be identified, in order to differentiate the different kind of statistics.

3 Alive vs. Dead Code

This subcategory will enable knowledge about how antique the code is. Statistics would work by dumping in total lines of code and in proportion statistics, code unchanged in a certain period of time.

- Statistic Input: Period of time
- Statistic Output: LOC unchanged (total) / LOC unchanged (%) / Unchanged files (total) / Unchanged files (%)
- Statistic Output: LOC changed (total) / LOC changed (%) / Changed files (total) / Changed files (%)

4 Changes on a file or group of files

This subcategory will allow to dump specific reports for a particular file, group of file, directory and group of directories:

- Statistic Input: Period of time
- Statistic Output: LOC unchanged (total) / LOC unchanged (%)
- Statistic Output: LOC changed (total) / LOC changed (%)

5 Changes associated to users

This subcategory will allow to dump specific reports associated to the LOC changes of particular user:

- Statistic Input: Period of time (and optionally, user or group of users)
- Statistic Output: LOC changed per user (lines / user)
- Statistic Output: LOC changed multiplied per user (lines * user)

6 Conclusion

Guilty tool provides some enhancements that can help on contributing to Metrics Grimoire, in order to achieve a wider collection of metrics data to be dumped and later analyzed.

Previous statistics will help on clarifying the state of the activity that a particular project is experiencing, in order to complete the statistics dumped for the different reports available in VizGrimoire.