

MAC0431 EP2

Fernando Omar Aluani
Jefferson Serafim Ascano

18 de outubro de 2012

Introdução

Propósito:

Propósito:

Análise da participação de desenvolvedores em um repositório *Git*, extraindo estatísticas dos commits no projeto.

Fonte dos Dados:

Fonte dos Dados:

Relatório dos commits (de qualquer intervalo de tempo) de um repositório. Gerado pelo próprio Git pelo comando

git log --numstat

Problema

Entrada:

Entrada:

Log dos commits do Git. O log é composto pelas informações de cada commit, do mais recente ao mais antigo, de forma sequencial. O “bloco” de cada commit contém as seguintes informações:

Entrada:

Log dos commits do Git. O log é composto pelas informações de cada commit, do mais recente ao mais antigo, de forma sequencial. O “bloco” de cada commit contém as seguintes informações:

- Hashcode que identifica unicamente o commit.
- Nome e email do autor do commit.
- Data e hora em que o commit foi feito.
- Texto descrevendo o commit (escrito pelo autor).
- Lista de triplas contendo: número de linhas adicionadas, número de linhas removidas, nome do arquivo alterado.

Saída:

Saída:

Para cada desenvolvedor que fez algum commit:

- Número de commits feitos pelo desenvolvedor;
- Média, desvio padrão e total de linhas modificadas pelo desenvolvedor;

Saída:

Para cada desenvolvedor que fez algum commit:

- Número de commits feitos pelo desenvolvedor;
- Média, desvio padrão e total de linhas modificadas pelo desenvolvedor;

Para os commits no geral:

- Média e desvio padrão de linhas modificadas por commit;

Saída:

Para cada desenvolvedor que fez algum commit:

- Número de commits feitos pelo desenvolvedor;
- Média, desvio padrão e total de linhas modificadas pelo desenvolvedor;

Para os commits no geral:

- Média e desvio padrão de linhas modificadas por commit;

Para cada arquivo alterado (criado, modificado, removido, etc) no repositório:

- Média e desvio padrão de linhas modificadas no arquivo.

Abordagem

Map:

Entrada Par (Chave, Texto).

Saida Coleção de pares (Atributo, Valor).

Map:

Entrada Par (Chave, Texto).

Saida Coleção de pares (Atributo, Valor).

Onde:

- Chave é o hash identificando o commit;
- Texto é o bloco de texto descrevendo o commit;
- Atributo é o nome de algum atributo extraído do commit, como por exemplo `numcommits_NomeDoDesenvolvedor`;
- Valor é o valor desse atributo extraído do commit;

Reduce:

Entrada Pares produzidos pelo **Map**.

Saida Coleção de pares (Estatística, Valor).

Reduce:

Entrada Pares produzidos pelo **Map**.

Saida Coleção de pares (Estatística, Valor).

Onde:

- Estatística é o nome de alguma estatística calculada a partir dos atributos do commit, como por exemplo `NomeDoAtributo_media`;
- Valor é o valor da estatística;