

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE COMPUTAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**TIAGO KENJI UMEMURA**

**UMA FERRAMENTA PARA MONITORAMENTO  
DA ENTROPIA DE MUDANÇA E SUA RELAÇÃO  
COM MÉTRICAS DE SOFTWARE**

MONOGRAFIA

**CAMPO MOURÃO**

**2016**

**TIAGO KENJI UMEMURA**

**UMA FERRAMENTA PARA MONITORAMENTO  
DA ENTROPIA DE MUDANÇA E SUA RELAÇÃO  
COM MÉTRICAS DE SOFTWARE**

Proposta de Trabalho de Conclusão de Curso de Graduação apresentado à disciplina de Trabalho de Conclusão de Curso 1, do Curso de Bacharelado em Ciência da Computação do Departamento Acadêmico de Computação da Universidade Tecnológica Federal do Paraná, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Igor Scaliante Wiese

**CAMPO MOURÃO**

**2016**

# Resumo

---

. Uma ferramenta para monitoramento da entropia de mudança e sua relação com métricas de software. 2016. 22. f. Monografia (Curso de Bacharelado em Ciência da Computação), Universidade Tecnológica Federal do Paraná. Campo Mourão, 2016.

**Contexto:** A entropia de mudança é uma medida para indicar o quanto um software sofre alterações em um determinado período de tempo. Estudos mostraram que o aumento da entropia pode causar desordem no processo de desenvolvimento podendo levar ao aumento no número de defeitos do software. Porém não foi analisado a relação da entropia com uma grande quantidade de métricas de software e além disso não há uma ferramenta que monitore a relação entre a entropia de mudança e métricas de softwares, como por exemplo, número de autores que modificaram um arquivo, número de *commits*, *authorship* e *ownership*

**Objetivo:** Implementar uma ferramenta que possibilite o monitoramento da entropia e das métricas de softwares de projetos armazenados no Github para ajudar os desenvolvedores no gerenciamento de projeto.

**Método:** A ferramenta é dividida em coleta de dados, cálculo da entropia, cálculo das métricas, análise estatística, visualização de dados e avaliação. Na coleta de dados os dados serão extraídos do GHTorrent e da API Github e após a coleta será realizado o cálculo da entropia e das métricas de software utilizando principalmente a ferramenta Change Metrics. Na etapa de análise estatística será feito a comparação e a correlação entre as métricas e em seguida será gerado um relatório sobre os resultados para o usuário. Na etapa de visualização de dados, a entropia de mudança será exibida utilizando o *Treemapping* e serão contruídos gráficos que mostram o comportamento das métricas ao longo do tempo. Para a avaliação da ferramenta serão convidados desenvolvedores para avaliar a ferramenta e depois serão feitos questionários sobre a usabilidade da ferramenta para esses desenvolvedores.

**Resultados esperados:** É esperado que a ferramenta ajude os desenvolvedores a gerenciar melhor os projetos oferecendo relatórios estatísticos e visualizações da relação da entropia com as métricas sociais, de processo e de autoria.

**Palavras-chaves:** Entropia. Fatores Sociais. Defeitos.

# Lista de figuras

---

|     |   |    |
|-----|---|----|
| 2.1 | Exemplo de cálculo da entropia de mudança por arquivo (Hassan, 2009). . . | 8  |
| 3.1 | Fluxograma do funcionamento da ferramenta . . . . .                       | 13 |
| 3.2 | Exemplo de visualização utilizando Treemapping. . . . .                   | 17 |
| 3.3 | Exemplo de Heat Map com matriz . . . . .                                  | 17 |
| 3.4 | Exemplo de gráfico de duas métricas ao longo do tempo . . . . .           | 18 |

# Lista de tabelas

---

|     |  |    |
|-----|--|----|
| 2.1 | Tabela de métricas sociais . . . . .     | 11 |
| 2.2 | Tabela de métricas de processo . . . . . | 12 |
| 3.1 | Tabela de cálculo das métricas . . . . . | 15 |
| 4.1 | Tabela do cronograma . . . . .           | 20 |

# Sumário

---

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introdução</b>                                | <b>5</b>  |
| <b>2</b> | <b>Referencial Teórico</b>                       | <b>7</b>  |
| 2.1      | Entropia de mudança . . . . .                    | 7         |
| 2.1.1    | Trabalhos Relacionados . . . . .                 | 8         |
| 2.2      | Métricas . . . . .                               | 9         |
| 2.2.1    | Métricas de autoria . . . . .                    | 10        |
| 2.2.1.1  | <i>Authorship</i> . . . . .                      | 10        |
| 2.2.1.2  | <i>Ownership</i> . . . . .                       | 10        |
| 2.2.1.3  | Experiência . . . . .                            | 11        |
| 2.2.2    | Métricas Sociais . . . . .                       | 11        |
| 2.2.3    | Métricas de processo . . . . .                   | 11        |
| <b>3</b> | <b>Proposta</b>                                  | <b>13</b> |
| 3.1      | Coletar dados . . . . .                          | 13        |
| 3.2      | Calcular entropia . . . . .                      | 14        |
| 3.3      | Calcular métricas . . . . .                      | 14        |
| 3.4      | Gerar relatório de análise estatística . . . . . | 16        |
| 3.5      | Gerar Visualização . . . . .                     | 16        |
| 3.6      | Avaliar a ferramenta . . . . .                   | 18        |
| <b>4</b> | <b>Cronograma</b>                                | <b>20</b> |
|          | <b>Referências</b>                               | <b>21</b> |

---

# Introdução

---

Os artefatos de softwares são modificados ao longo do tempo e nesse processo a qualidade do software tende a piorar (Hassan, 2009), uma vez que existe a necessidade da mudança contínua de requisitos durante a evolução do software.

Para quantificar o impacto das mudanças contínuas os pesquisadores (Hassan, 2009) tem utilizado o conceito de entropia de mudança. Esta medida pode ser obtida a partir do número de mudanças que ocorrem em um projeto ou arquivo em um determinado período de tempo. Hassan (2009) observou que maior quantidade de mudanças está relacionada com o aumento do valor da entropia e consequentemente está relacionado com maior tendência do software apresentar defeitos.

O impacto da entropia na qualidade do projeto também é investigado na pesquisa de Canfora *et al.* (2014), que analisou a relação entre a entropia e atividades de desenvolvimento, como a refatoração, padrões de projetos e a quantidade de desenvolvedores que mudam um determinado arquivo.

Estes estudos não consideram uma grande quantidade de métricas de software, por exemplo, não foi analisado a comunicação dos desenvolvedores, número de *commits* e número de arquivos que são alterados. Portanto não existe uma ferramenta que possibilite os desenvolvedores e os gerentes monitorarem os efeitos da entropia e sua relação com métricas sociais, de autoria e de processo.

Diante desse contexto, sabe-se que o desenvolvimento de software é uma tarefa sóciotécnica, pois é essencial a comunicação entre os desenvolvedores para coordenar as atividades de desenvolvimento durante a evolução do software. Assim o objetivo deste trabalho é construir uma ferramenta que monitora os valores de entropia e a sua relação com as métricas sociais, de autoria de mudança (authorship, ownership e experience), métricas de processo (quantidade

de commits, quantidade de defeitos, quantidade de linhas removidas, quantidade de linhas adicionadas, Code Churn, quantidade de refatorações, max change set e average change set) e número de defeitos.

A ferramenta irá gerar relatórios estatísticos sobre a entropia e as métricas e fornecerá a visualização desses dados para auxiliar os desenvolvedores na tomada de decisões durante o desenvolvimento de um projeto. O relatório estatístico será feito utilizando as técnicas de *Wilcoxon*, *ANOVA* e *Cliffs's test* e irá sugerir quais arquivos do projeto possuem maior tendência para apresentar defeitos. Na visualização de dados será utilizado o *Treemapping* para mostrar quais arquivos do projeto possuem maior entropia e para as métricas de software serão utilizados gráficos que mostrem os valores das métricas ao longo do tempo.

A última etapa será a avaliação da ferramenta que será feita com desenvolvedores que utilizarão a ferramenta e depois responderão um questionário sobre as funcionalidades e usabilidade da ferramenta.

Esta proposta está organizada da seguinte forma. O capítulo 2 apresenta as definições das métricas sociais, de processos e os trabalhos relacionados. O capítulo 3 apresenta a proposta de construção da ferramenta, como a ferramenta será validada e uso da ferramenta. O capítulo 4 apresenta o cronograma.



---

## Referencial Teórico

---

Este capítulo apresenta os conceitos de entropia, métricas sociais, métricas de autoria e métricas de processo.

### 2.1. Entropia de mudança

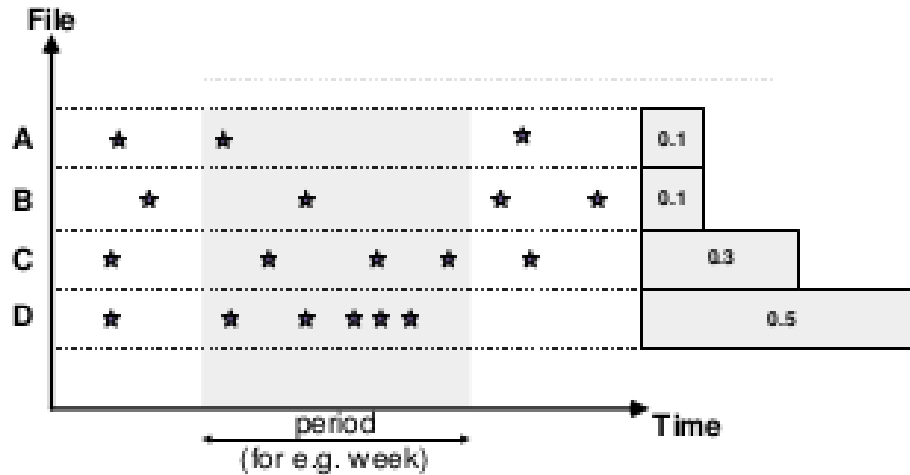
A entropia de Shannon (2001) é uma medida para mensurar a incerteza associada a uma variável que quantifica uma informação contida em uma mensagem produzida por um emissor de dados. A partir dessa definição foi criado o conceito de entropia de mudança que tem como objetivo indicar o quanto um código está mudando durante um determinado período de tempo.

Na entropia de mudança o software é considerado o emissor de dados e as modificações realizadas são os dados de entrada. A entropia é uma medida para mensurar a quantidade de mudanças que ocorreram em um determinado espaço de tempo em um arquivo de um projeto. As mudanças consideradas podem ser obtidas a partir da quantidade de linhas modificadas em um intervalo de tempo ou utilizando número de *commits*.

A entropia de mudança é definida como (Hassan, 2009):

$$H(S) = \sum_{n=1} \frac{chg(f_i)}{chg(S)} \log_2 \left( \frac{chg(f_i)}{chg(S)} \right) \quad (2.1)$$

A figura 2.1 mostra um exemplo do cálculo da entropia dos arquivos de um sistema (Hassan, 2009). Nesse exemplo é considerado os arquivos A, B, C e D de um sistema dado um período de tempo qualquer, na qual as estrelas indicam quando ocorreu uma mudança no arquivo.



**Figura 2.1.** Exemplo de cálculo da entropia de mudança por arquivo (Hassan, 2009).

Para cada arquivo no sistema, é feito a contagem de quantas vezes ele foi modificado em um período e depois é dividido pelo total de mudanças que ocorreram nesse mesmo período considerando todos os arquivos. Por exemplo, na figura 2.1 como ocorreram dez mudanças no período selecionado e uma mudança no arquivo A nesse mesmo período então a entropia de mudança para esse arquivo é 0,1.

### 2.1.1. Trabalhos Relacionados

No trabalho de Hassan (2009) foi introduzido o conceito de entropia de mudança do software e foi feito um modelo básico de mudança de código (BCC - *Basic Code Change Model*) para mensurar a complexidade de mudança.

No modelo básico de mudança de código são utilizados arquivos como unidade de código para mensurar a complexidade de código, o intervalo de tempo para o cálculo da entropia é fixo e é considerado que os arquivos do projeto sejam sempre os mesmos. Para eliminar a limitação do tempo e a limitação alteração no número arquivos é utilizado o modelo estendido de mudança de código (ECC).

No ECC o histórico de mudanças é dividido em períodos de tamanhos iguais. Essa divisão pode ser feita com base em diferentes critérios: com base no tempo, no número de modificações ou em períodos onde ocorrem mais modificações (*Burst Based Period*). A entropia será calculada com base no número de modificações que ocorreram no período definido.

Hassan (2009) aplicou o modelo ECC para 6 projetos: NetBSD, FreeBSD, OpenBSD, Postgres, KDE e KOffice. Nesse estudo foi comparado modelos de predição de defeitos

baseados na entropia, modelo baseados em defeitos anteriores e modelo baseado em modificações anteriores.

Os resultados indicam que modelos baseados na métrica de entropia tem o desempenho igual ou superior aos modelos baseados em modificações anteriores e baseados em defeitos anteriores.

A pesquisa de Canfora *et al.* (2014) relaciona a entropia de mudança com características do software e atividades de desenvolvimento. As características analisadas foram: refatoração, número de committers, padrões de projetos e nome de tópicos no projeto. Na pesquisa foram analisados os projetos ArgoUML, Eclipse-JDT, Mozilla e Samba em um período de cerca de 10 anos.

O método para extração de dados foi dividido em 6 passos: extração das métricas de mudança do sistema de controle de versão, cálculo da entropia de mudança, identificação das mudanças relacionadas a refatoração, contagem número de autores que contribuem para o projeto, identificação dos padrões de projetos e por último foi necessário identificar os tópicos que são descritos na mensagem de cada *commit*.

Os resultados de Canfora indicam que a entropia de mudança diminui de forma significativa após uma atividade de refatoração, o valor da entropia é mais alto para arquivos com maior número de contribuidores, classes que participam de diferentes padrões de projetos exibem valores diferentes de entropia, mudanças classificadas em diferentes tópicos exibem valores de entropia diferentes e há maior relação entre o valor da entropia, tópicos de mudança e número de desenvolvedores que modificam um arquivo.

## 2.2. Métricas

Métricas de software são medidas para mensurar características presentes no desenvolvimento de software e servem para auxiliar na tomada de decisões durante o desenvolvimento do projeto (Koscianski; Soares, 2007). As métricas foram divididas em métricas de autoria, sociais e de processo.

As métricas de autoria medem a contribuição do desenvolvedor nos projetos, métricas sociais medem a interação entre os desenvolvedores e as métricas de processo mensuram as mudanças que ocorreram em relação aos artefatos criados durante o desenvolvimento do projeto.

### 2.2.1. Métricas de autoria

Esta seção apresenta métricas calculadas a partir das contribuições dos usuários em cada arquivo de um projeto.

#### 2.2.1.1. *Authorship*

Authorship é uma medida para mensurar o quanto um desenvolvedor contribuiu para um determinado módulo de software.

Essa medida pode ser obtida de várias formas: contando número de arquivos que o desenvolvedor modificou, número de commits e outra possibilidade é contar número de linhas modificadas pelo contribuidor, também chamada de *Code Churn* (Munson; Elbaum, 1998).

Na pesquisa realizada por Rahman e Devanbu (2011) o authorship é calculado utilizando o número de linhas modificadas no código pelo desenvolvedor dividido pelo número total de linhas do arquivo, e o autor com a maior contribuição é denominado *ownership*. Também é definido implicated code, que é o código modificado quando é corrigido um determinado erro no módulo de software. O trabalho de Rahman investiga a relação entre ownership, authorship e experience com implicated code. Para cada linha de código modificado é utilizado o comando blame para identificar o autor responsável por essa mudança. O resultado fornece evidências indicando que implicated code tende a ser mais frequentemente gerado por poucos autores, vários fragmentos de códigos modificados tem apenas um único autor.

#### 2.2.1.2. *Ownership*

No trabalho de Greiler e Herzig (2015) o Ownership é medido considerando número de contribuidores de um arquivo e também é verificado se existe um contribuidor principal, nesse caso a medida foi calculada com base no número de commits do autor em relação ao total de commits para aquele componente.

No artigo de Foucault *et al.* (2015) os contribuidores são classificados como owner, minor e major. Owner é o contribuidor com maior valor de contribuição, minor o desenvolvedor que contribuiu com menos de 5% e major contribuiu com mais de 5%.

A métrica social ownership será usada para se referir ao authorship do desenvolvedor que mais contribuiu com o projeto (Thongtanunam *et al.*, ).

### 2.2.1.3. Experiência

A experiência é a medida para calcular o nível de experiência do contribuidor, essa medida é computada analisando o número de linhas (Rahman; Devanbu, 2011) deltas comitadas pelo contribuidor em determinado espaço de tempo.

Na pesquisa de Rahman e Devanbu (2011) a experiência é dividida em dois tipos: a experiência especializada e experiência geral. A experiência especializada é medida considerando o quanto um indivíduo contribui em um determinado arquivo e a experiência geral é medida considerando um projeto inteiro.

### 2.2.2. Métricas Sociais

Esta seção apresenta métricas calculadas a partir da comunicação entre os desenvolvedores a partir das issues e de Pull Requests. As métricas são descritas na tabela 2.1.

**Tabela 2.1.** Tabela de métricas sociais

| Métrica                              | Descrição  |
|--------------------------------------|--|
| Pull Requests fechados por arquivo   | Representa o número total de Pull Request fechados onde determinado arquivo do projeto foi modificado.                                 |
| Pull Request abertos por arquivo     | Representa o número total de Pull Requests abertos onde determinado arquivo do projeto foi modificado.                                 |
| Comentários em Pull Request abertos  | Representa o número total de comentários considerando todos os Pull Request abertos até o momento para determinado arquivo do projeto  |
| Comentários em Pull Request fechados | Representa o número total de comentários considerando todos os Pull Request fechados até o momento para determinado arquivo do projeto |

### 2.2.3. Métricas de processo

As métricas de processo são métricas para mensurar as características relacionadas aos artefatos produzidos durante o desenvolvimento do projeto. As métricas de processo serão descrita na tabela 2.1 e serão calculadas utilizando a ferramenta *Change Metrics*:

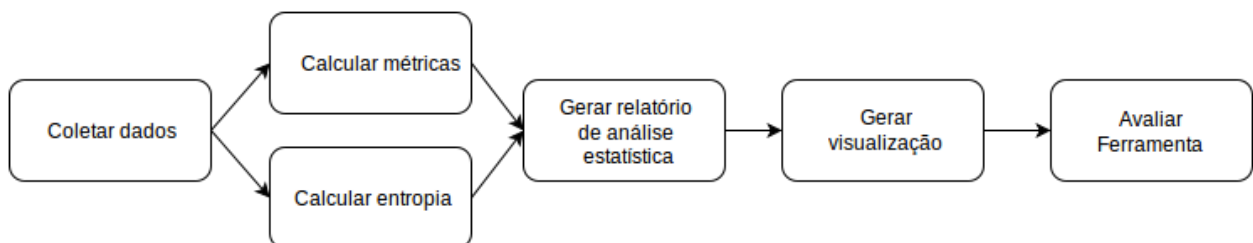
**Tabela 2.2.** Tabela de métricas de processo

| Nome da métrica                  | Descrição  |
|----------------------------------|--|
| Quantidade de commits            | A quantidade de commits representa o nível de atividade do projeto em termos de número de commits feitos. É calculado o número de commits do projeto em um certo período de tempo. |
| Quantidade de defeitos           | A quantidade de defeitos será calculado com o número de issues do projeto que foram criadas em uma determinada data.   |
| Idade do arquivo                 | A métrica idade do arquivo representa o tempo de existência do arquivo. O cálculo é utilizado medindo a diferença de tempo entre o primeiro e último commit.                       |
| Quantidade de linhas removidas   | Essa medida representa a quantidade de linhas removidas de um arquivo até o momento.   |
| Quantidade de linhas adicionadas | Essa medida representa a quantidade de linhas adicionadas de um arquivo até o momento.   |
| Code Churn                       | A métrica Code Churn representa a soma de todas as linhas de código removidas e adicionas no arquivo.  |
| Quantidade de refatorações       | Essa medida representa a quantidade de refatorações ocorridas até o momento, se a refatoração é citada no commit.  |
| Max Change Set                   | A métrica Max Change Set representa o número máximo de arquivos que foram alterados junto com o arquivo em questão.  |
| Average Change Set               | As métricas Average Change Set representa a média de número de arquivos alterados juntos.  |

## Proposta

Nos estudos realizados por Canfora *et al.* (2014) e Hassan (2009) foram analisados a relação da entropia de mudança com a qualidade do projeto e algumas métricas de software, como por exemplo, número de contribuidores, no entanto, várias métricas não foram analisadas nesses estudos e não há uma ferramenta que monitore a entropia, métricas sociais, de autoria e de processo.

Portanto o objetivo é implementar uma ferramenta que monitore o valor da entropia de mudança e sua relação com as métricas sociais, de autoria e de processo dos projetos. A ferramenta terá seis módulos que serão apresentados a seguir:



**Figura 3.1.** Fluxograma do funcionamento da ferramenta

### 3.1. Coletar dados

O módulo de coleta de dados é responsável pela extração de dados dos projetos do Github e Git que serão persistidos em um banco de dados local. Os dados serão obtidos utilizando o GHTorrent<sup>1</sup> e a API Github.

<sup>1</sup> GHTorrent é uma ferramenta que monitora os eventos públicos do Github, para cada evento é armazenado o seu conteúdo e as resposta JSON são armazenadas em um banco de dados MongoDB e também no MySQL

Utilizando a base de dados MySQL do GHTorrent será extraído dados sobre os *commits*, *issues*, *Pull Requests* e contribuidores do projeto. Além disso, será feito o clone das versões do projeto selecionado utilizando o comando *git log*, *git clone* e *git reset*.

## 3.2. Calcular entropia

Este módulo que realiza o cálculo da entropia de cada arquivo do projeto no período que o usuário informar. O valor da entropia será determinado utilizando como medida o número de commits que o arquivo teve durante um período de tempo selecionado pelo usuário. Para o cálculo da entropia é necessário medir a diferença da métrica Revisões(número total de commits) gerada pela ferramenta Change Metrics entre duas versões diferentes do projeto e assim obter o número de commits que cada arquivo teve entre uma versão e outra.

## 3.3. Calcular métricas

O módulo de cálculo das métricas irá realizar o cálculo após a extração dos dados no módulo de coleta de dados. Nessa etapa o usuário deverá informar quais métricas ele deseja calcular e uma data anterior a data atual. O período definido será entre a data atual e a data definida pelo usuário e as métricas serão calculadas de 15 em 15 dias.

Para o cálculo das métrica de processo será utilizada a ferramenta Change Metrics e para executá-la é necessário passar três parâmetros: o projeto git, o arquivo csv onde serão armazenados os dados e tipo do projeto que pode ser *all* ou *single*. O comando para executar a Change Metrics é: `java -jar <tool.jar> projeto arquivo.csv single`.

A forma como as métricas serão calculadas estão descritas na tabela 3.1.



**Tabela 3.1.** Tabela de cálculo das métricas

| Métrica                             | Descrição do cálculo  |
|-------------------------------------|---|
| Authorship                          | Média da quantidade de commits considerando todos os contribuidores                         |
| Ownership                           | Maior valor da métrica de authorship  |
| Experiência                         | Média da quantidade de linhas modificadas no arquivo por todos os contribuidores            |
| Pull Request fechados por arquivo   | Contagem do número de Pull Request fechados relacionados a um determinado arquivo           |
| Pull Request abertos por arquivo    | Contagem do número de Pull Request abertos relacionados a um determinado arquivo            |
| Comentários em Pull Request aberto  | Contagem do número de comentários em todos os Pull Request abertos relacionados ao arquivo  |
| Comentários em Pull Request fechado | Contagem do número de comentários em todos os Pull Request fechados relacionados ao arquivo |
| Quantidade de commits               | Calculado utilizando a ferramenta Change Metrics  |
| Quantidade de defeitos              | Calculado utilizando a ferramenta Change Metrics  |
| Idade do arquivo                    | Calculado utilizando a ferramenta Change Metrics  |
| Quantidade de linhas removidas      | Calculado utilizando a ferramenta Change Metrics  |
| Quantidade de linhas adicionadas    | Calculado utilizando a ferramenta Change Metrics  |
| Code Churn                          | Calculado utilizando a ferramenta Change Metrics  |
| Quantidade de refatorações          | Calculado utilizando a ferramenta Change Metrics  |
| Max Change set                      | Calculado utilizando a ferramenta Change Metrics  |
| Average Change set                  | Calculado utilizando a ferramenta Change Metrics  |

### 3.4. Gerar relatório de análise estatística

Como foi feito na pesquisa de Canfora *et al.* (2014), para analisar a relação da entropia e as métricas de software serão utilizados os métodos estatísticos de *Wilcoxon*, *ANOVA* e *Cliff's Delta*, onde todos os testes serão feitos no ambiente estatístico R com nível de significância de 95%.

O teste de *Wilcoxon* pareado é um método não paramétrico para comparação de duas amostras com o objetivo de verificar se existem diferenças significativas entre elas. E para mensurar a diferença entre as amostras é utilizado o *Cliff's Delta*, uma medida de tamanho de efeito não paramétrico, por exemplo, pode ser utilizada para mensurar a diferença no valor da entropia de mudança antes e depois da refatoração.

O *ANOVA* (Análise de Variância) é usada para verificar se a média de duas ou mais populações são iguais e determina se as diferenças entre as médias amostrais sugerem diferenças significativas entre as médias das populações ou se essas diferenças decorrem apenas da variabilidade implícita de cada amostra. O método *ANOVA* irá ser utilizado para analisar a interação entre a entropia de mudança e outras métricas de software.

Após a análise estatísticas será gerado um relatório informando o usuário identificando quais métricas apresentam maior relação com a entropia e quais arquivos possuem maior entropia.

### 3.5. Gerar Visualização

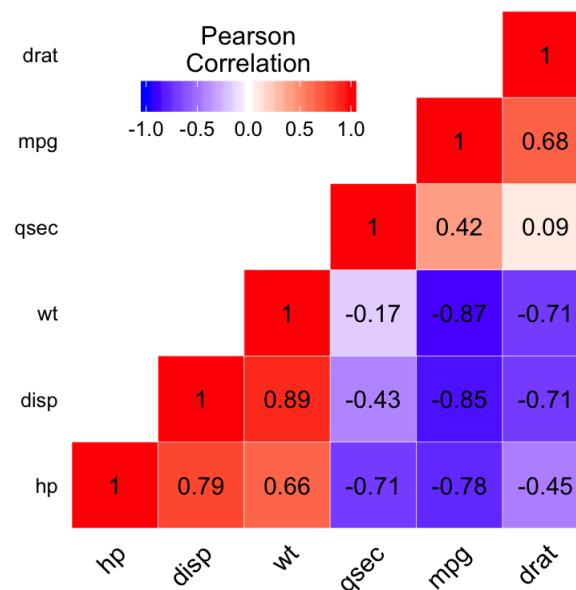
Este módulo permite o usuário utilizar filtros para escolher quais métricas ele deseja visualizar. Para visualização será utilizado o *Treemapping*, *Heat Map* e gráficos de duas dimensões.

O treemapping é uma forma de visualização de dados de forma hierárquica utilizando retângulos aninhados. Na ferramenta o *Treemapping* será utilizado para visualizar o valor da entropia dos arquivos do projeto, onde cada retângulo representará um arquivo e o tamanho desse retângulo representa o valor da entropia, quanto maior a entropia maior será o retângulo. O usuário poderá escolher qual período de tempo ele deseja visualizar o valor da entropia e também poderá comparar esse valor entre dois períodos diferentes, sendo que o valor da entropia é calculado de 15 em 15 dias.



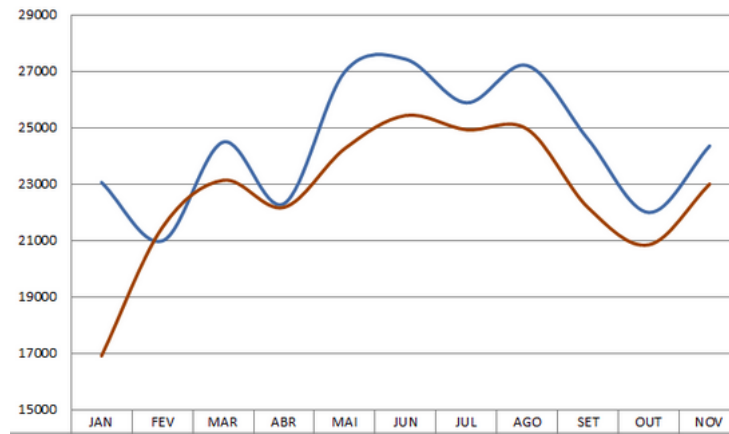
**Figura 3.2.** Exemplo de visualização utilizando Treemapping.

O *Heat Map* é uma representação gráfica que utiliza cores para mostrar as relações entre os valores dos dados. Essa visualização será utilizada em forma de matriz, onde as linhas e colunas representarão as métricas e as cores nas intersecções representarão o nível de correlação entre as métricas e entre as métricas e a entropia de mudança. A cada período de 15 dias será gerado um *Heat Map*.



**Figura 3.3.** Exemplo de Heat Map com matriz

Será utilizado um gráfico de duas dimensões para mostrar a evolução das métricas ao longo do tempo como mostra as figura 3.4. O eixo x do gráfico representa o tempo e o eixo y representa o valor da métricas, sendo que o tempo está dividido a cada 15 dias.



**Figura 3.4.** Exemplo de gráfico de duas métricas ao longo do tempo

### 3.6. Avaliar a ferramenta

Para a avaliação, a ferramenta será configurada para 10 projetos e serão convidados desenvolvedores para utilizá-la, onde os desenvolvedores deverão realizar tarefas sobre as funcionalidades da ferramenta. As tarefas são descritas nos cenários apresentados a seguir.

Cenário 1: O desenvolvedor poderá escolher o período de tempo em que ele deseja analisar o projeto e quais filtros (métricas) ele deseja calcular. Após definir o período e os filtros, será gerado um gráfico de duas dimensões mostrando o comportamento das métricas ao longo do tempo.

Cenário 2: O desenvolvedor poderá consultar o valor da entropia de cada arquivo após selecionar um período de 15 dias. O valor da entropia será mostrado através do *Treemap* do projeto, onde cada retângulo representará um arquivo e o tamanho do retângulo representará o valor da entropia nesse arquivo. Ao clicar no *Treemap* será gerado um relatório que irá conter informações sobre quais arquivo possuem maior entropia e portanto merecem maior atenção do desenvolvedor.

Cenário 3: O desenvolvedor poderá escolher um período (15 dias) e analisar a correlação entre as métricas e entre a entropia e as métricas utilizando o *Heat Map*. Clicando no *Heat Map* será gerado um relatório mostrando quais métricas possuem maior nível de correlação.

Cenário 4: O desenvolvedor poderá comparar os valores da entropia e das métricas em dois períodos diferentes. Dois *Treemaps* serão apresentados, um para cada período selecionado, para exibir os valores da entropia de cada arquivo do projeto e será apresentado uma tabela com os valores das métricas para cada período selecionado.

Cenário 5: O desenvolvedor ao gerar o *Treemap* de um projeto durante um período poderá

ter acesso a mais informações de cada arquivo do projeto. Ao clicar em um dos retângulos que representa um arquivo, o desenvolvedor terá acesso a um relatório com informações sobre quais métricas possuem maior correlação com a entropia daquele arquivo.

Após a realização das tarefas, os desenvolvedores responderão um questionário sobre a usabilidade da ferramenta. O questionário irá conter questões sobre a visualização de dados e o relatório estatístico para avaliar se a ferramenta auxilia o desenvolvedor na tomada de decisões do projeto. O questionário é composto pelas seguintes perguntas:

1. Houve dificuldades para gerar o *Treemap* do valor da entropia de um determinado período?
2. Houve dificuldade para gerar o *Heat Map*?
3. Houve dificuldades para gerar os gráficos?
4. O relatório estatístico sobre o valor da entropia nos arquivos é relevante para o desenvolvimento do projeto?
5. A visualização dos valores da entropia de mudança utilizando *Treemapping* é de fácil compreensão?
6. A visualização de *Heat Map* utilizada para analisar a correlação é de fácil compreensão?
7. A correlação entre o valor da entropia e os valores das métricas é relevante para o desenvolvimento do projeto?
8. Os gráficos de duas dimensões para analisar o histórico dos valores da métrica são importantes para tomar decisões no desenvolvimento do projeto?

## Cronograma

**Tabela 4.1.** Tabela do cronograma

| Módulos Ferramenta                     | Jan | Fev | Mar | Abr | Jun |
|--|-----|-----|-----|-----|-----|
| Coletar dados                          | X   |     |     |     |     |
| Calcular entropia                      |     | X   |     |     |     |
| Calcular métricas                      |     | X   |     |     |     |
| Gerar Relatório de Análise Estatística |     |     | X   |     |     |
| Visualizar Dados                       |     |     |     | X   |     |
| Avaliar Ferramenta                     |     |     |     |     | X   |

# Referências

---

CANFORA, Gerardo; CERULO, Luigi; CIMITILE, Marta; PENTA, Massimiliano Di. How changes affect software entropy: an empirical study. *Empirical Software Engineering*, v. 19, n. 1, p. 1–38, 2014. ISSN 1573-7616. Disponível em: <http://dx.doi.org/10.1007/s10664-012-9214-z>.

FOUCAULT, Matthieu; TEYTON, Cédric; LO, David; BLANC, Xavier; FALLERI, Jean-rémy. On the usefulness of ownership metrics in open-source software projects. *Information and Software Technology*, Elsevier B.V., v. 64, p. 102–112, 2015. ISSN 0950-5849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2015.01.013>.

GREILER, Michaela; HERZIG, Kim. Code Ownership and Software Quality : A Replication Study. 2015.

HASSAN, Ahmed E. Predicting faults using the complexity of code changes. In: *Proceedings of the 31st International Conference on Software Engineering*, 2009. (ICSE '09), p. 78–88. ISBN 978-1-4244-3453-4. Disponível em: <http://dx.doi.org/10.1109/ICSE.2009.5070510>.

KOSCIANSKI, A.; SOARES, M. dos Santos. *Qualidade de Software - 2ª Edição: Aprenda as metodologias e técnicas mais modernas para o desenvolvimento de software*. Novatec, 2007. ISBN 9788575221129. Disponível em: <https://books.google.com.br/books?id=09aWoUq6L88C>.

MUNSON, J. C.; ELBAUM, S. G. Code churn: A measure for estimating the impact of code change. In: *Proceedings of the International Conference on Software Maintenance*, 1998. (ICSM '98), p. 24–. ISBN 0-8186-8779-7. Disponível em: <http://dl.acm.org/citation.cfm?id=850947.853326>.

RAHMAN, Foyzur; DEVANBU, Premkumar. Ownership , Experience and Defects :. 2011.

SHANNON, C. E. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 5, n. 1, p. 3–55, jan. 2001. ISSN 1559-1662. Disponível em: <http://doi.acm.org/10.1145/584091.584093>.

THONGTANUNAM, Patanamon; MCINTOSH, Shane; HASSAN, Ahmed E; IIDA, Hajimu. Revisiting Code Ownership and its Relationship with Software Quality in the Scope of Modern Code Review. n. 1, [s.d.].