

# MetricMiner

Uma ferramenta web de apoio à mineração de repositórios de software

Francisco Sokol

Orientador: Marco Aurélio Gerosa  
Co-orientador: Mauricio Finavaro Aniche

12 de novembro de 2012

# Mineração de repositórios

- Estudo empírico da evolução de software
- Aplicação de técnicas de data mining aos dados do histórico de desenvolvimento de um software

# Mineração de repositórios

- Quais são as classes mais modificadas no projeto?
- Qual será a classe com mais bugs na próxima versão?
- Quais desenvolvedores devem trabalhar juntos?

# Motivação

Ferramentas atuais de mineração:

- Executam localmente
- Configuração complexa
- Recursos locais
- Pouca escalabilidade

# MetricMiner

Baseado no rEvolution <sup>1</sup>

Requisitos:

- Aplicação web
- Armazenamento de informações do sistema de controle de versão
- Cálculo de métricas de código
- Interface para consulta

---

<sup>1</sup><http://github.com/mauricioaniche/revolution>

# MetricMiner

## Tecnologias

- Java
  - VRaptor
  - Quartz Scheduler
- MySQL
  - Hibernate
- Infraestrutura de cloud da Locaweb

# MetricMiner

## Fila de execução

Tarefas de mineração são executadas *assíncronamente*

- Download do repositório (Git e SVN)
- Processamento do sistema de controle de versão
- Cálculo de métricas de código
- Métricas de projeto
- Consulta em SQL aos dados

# MetricMiner

Criação de novas tarefas de mineração:

```
public interface RunnableTaskFactory {  
    public RunnableTask build(Task task,  
        Session session,  
        StatelessSession statelessSession,  
        MetricMinerConfigs config);  
}
```

```
public interface RunnableTask {  
    public void run();  
}
```



# MetricMiner

Criação de novas métricas de código:

```
public interface MetricFactory {  
    public Metric build();  
}  
  
public interface Metric {  
    Collection<MetricResult> results(SourceCode source);  
    void calculate(InputStream is);  
    boolean matches(String name);  
    Class<?> getFactoryClass();  
}  
  
@MetricComponent(name="Cyclomatic Complexity")  
public class CCMetricFactory implements MetricFactory {  
    public Metric build() {  
        return new CCMetric();  
    }  
}
```

# MetricMiner

## Métricas implementadas

Implementação com a biblioteca javaparser<sup>2</sup>

- Complexidade Ciclomática
- LCOM
- Fan Out
- Número de linhas
- Número de métodos

---

<sup>2</sup><http://code.google.com/p/javaparser/>

# Avaliação da ferramenta

Mineração da Apache Software Foundation

Mineração de mais de 300 projetos da Apache<sup>3</sup>

- 90 horas de duração
- 800 mil commits
- 2 mil autores
- 1.5 milhões de artefatos
- 5 milhões de versões de código

---

<sup>3</sup><http://git.apache.org/>

# Avaliação da ferramenta

Reprodução de estudo da literatura

Expansão de estudo publicado na área  
[Soetens and Demeyer 2010]

*Qual o efeito de refatorações sobre a complexidade de um sistema?*

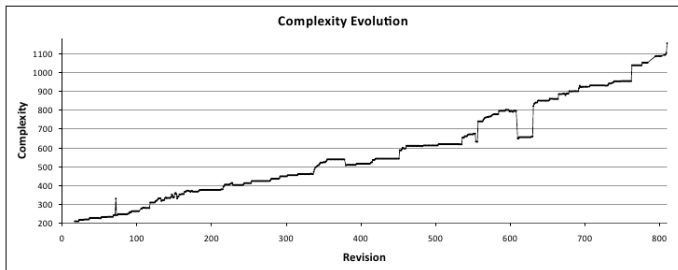
# Avaliação da ferramenta

Estudo original

- Complexidade Ciclomática
- 700 versões do projeto PMD
- SVNKit + Eclipse Metrics

# Avaliação da ferramenta

## Estudo original



# Avaliação da ferramenta

Estudo original

	Decrementaram CC	Não alteraram CC	Incrementaram CC
Refatorações documentadas	14	7	12
Sem refatoração documentada	27	580	136

# Avaliação da ferramenta

## Análise dos autores

- Poucas refatorações com remoção de código duplicado
- Commits com refatoração + nova funcionalidade
- Pequenas mudanças como movimentação de métodos e variáveis



# Avaliação da ferramenta

## Reprodução do estudo

Consulta pela interface web<sup>4</sup> + programa auxiliar para processar csv

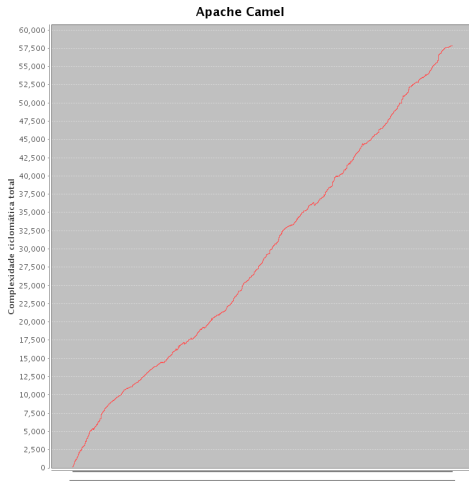
- 250 projetos java
- 500 mil commits processados

---

<sup>4</sup><http://metricminer.org.br/query/1>

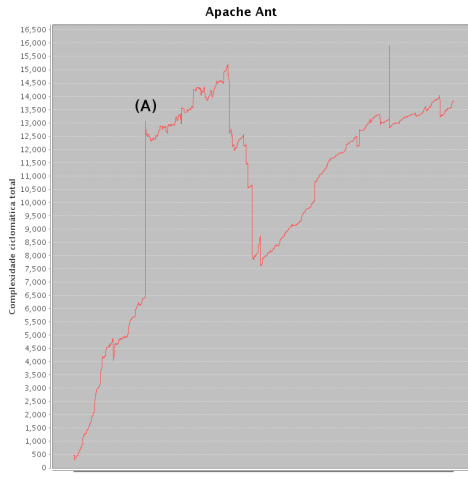
# Avaliação da ferramenta

## Reprodução do estudo



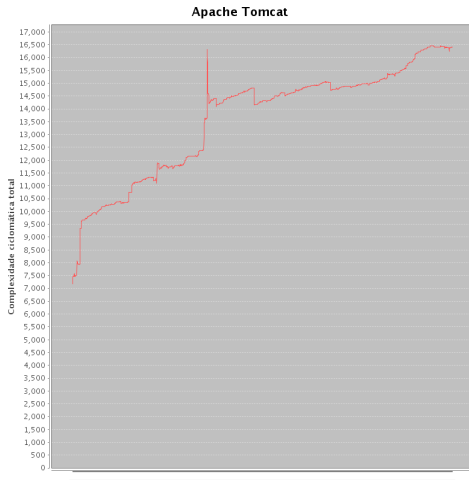
# Avaliação da ferramenta

## Reprodução do estudo



# Avaliação da ferramenta

## Reprodução do estudo



# Avaliação da ferramenta

## Reprodução do estudo

	Decrementaram CC	Não alteraram CC	Incrementaram CC
Refatorações documentadas	1504	1603	3230
Sem refatoração documentada	30145	99580	121239

# Avaliação da ferramenta

## Conclusão

- Maior quantidade e variedade de projetos analisados
- Resultados semelhantes ao do estudo original
- Processo de mineração mais simples e eficiente

# Trabalhos futuros

- Paralelizar
- Usabilidade da interface web
- Outras métricas e linguagens
- Sistemas de bug tracking, listas de email
- API

# Obrigado!

## Perguntas?

chico.sokol@gmail.com

<http://metricminer.org.br>

<http://github.com/metricminer/metricminer>





Soetens, Q. D. and Demeyer, S. (2010).

Studying the effect of refactorings: a complexity metrics perspective.

In *QUATIC 2010: The 7th International Conference on Quality in Information and Communications Technology*. IEEE Computer Society Press, IEEE Computer Society Press.