



IME - Instituto de  
Matemática e Estatística



# MetricMiner: uma ferramenta web de apoio à mineração de repositórios de software

Apoio



**Aluno:** Francisco Zigmund Sokol  
**Orientador:** Marco Aurélio Gerosa  
**Co-orientador:** Mauricio Finavaro Aniche

## 1. Introdução

A Mineração de Repositórios de Software estuda o processo de evolução de um software de forma empírica, por meio da análise dos artefatos envolvidos no seu desenvolvimento como código fonte, dados do sistema de controle de versão e sistemas de rastreamento de bugs [1]. Por meio da mineração desses dados, podem ser extraídas diversas informações úteis ao desenvolvimento de um software, como por exemplo a identificação de classes que são modificadas constantemente, classes mais propensas a falhas, entre outras. Com essas informações, equipes de desenvolvimento podem tomar ações para aprimorar o processo de desenvolvimento do sistema.

## 2. MetricMiner

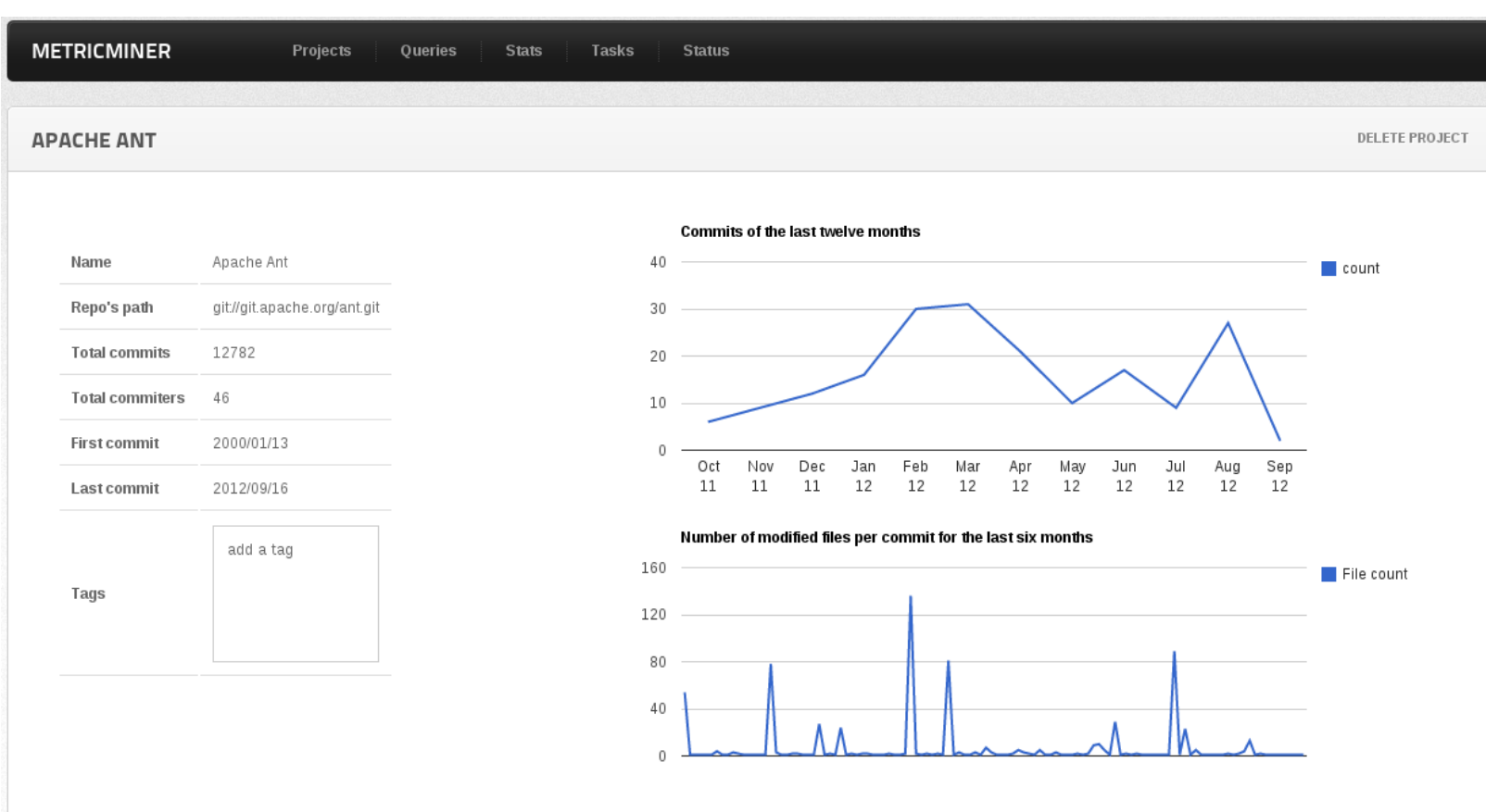
O MetricMiner possui uma interface web pela qual os usuários podem cadastrar projetos de software para serem minerados. Para esse cadastro, devem ser fornecidos um nome e a url pública para o repositório de código do projeto. Até o momento, os sistemas de controle de versão suportados são o Git e o SVN.

As seguintes tarefas de mineração são executadas assíncronamente no MetricMiner:

- Download do repositório de código
- Processamento das versões do código fonte e dos metadados do sistema de controle de versão
- Cálculo de métricas de código
- Execução de consultas SQL dos usuários

Cada tarefa é armazenada em uma fila de execução e processada por um componente que constrói as dependências de cada tarefa e as executa na ordem em que foram cadastradas.

Ao final do processo de mineração, todas as informações do repositório do projeto estão persistidas no banco de dados e disponíveis para serem extraídas pelo pesquisador. A Figura 1 exibe a tela de um sistema minerado no MetricMiner, com algumas estatísticas simples do projeto.



**Figura 1:** Tela de um projeto minerado no MetricMiner

## 3. Mineração dos projetos da Apache

Os projetos de código aberto da Apache foram minerados no MetricMiner para avaliar a capacidade da ferramenta. Ao final da mineração, foram obtidos os seguintes resultados:

- 307 projetos de código aberto minerados
- Processamento de 800 mil commits de 2 mil autores
- 1.5 milhões de artefatos e 5 milhões de versões de arquivos de código fonte
- 180 GB de dados armazenados no banco de dados
- 90 horas de processamento

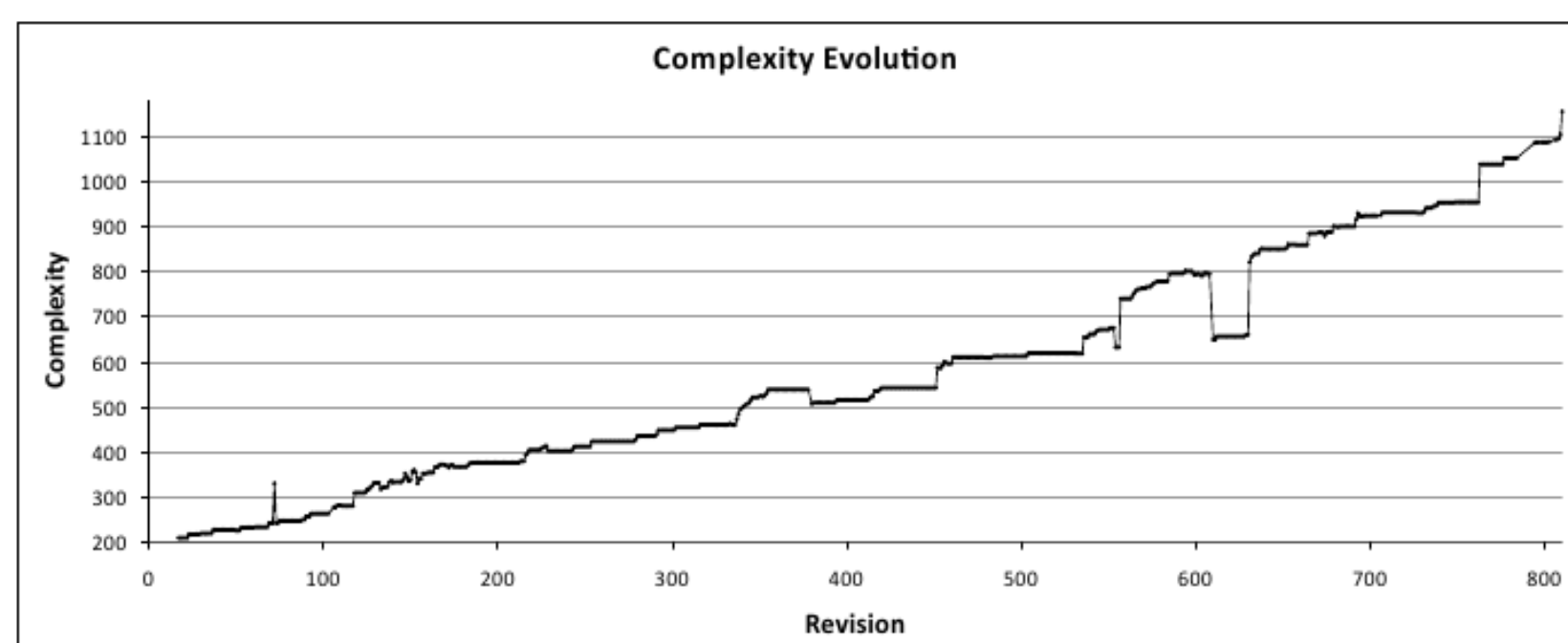
Os dados estão disponíveis para consulta em <http://metricminer.org.br>.

## 4. Reprodução de estudo da literatura

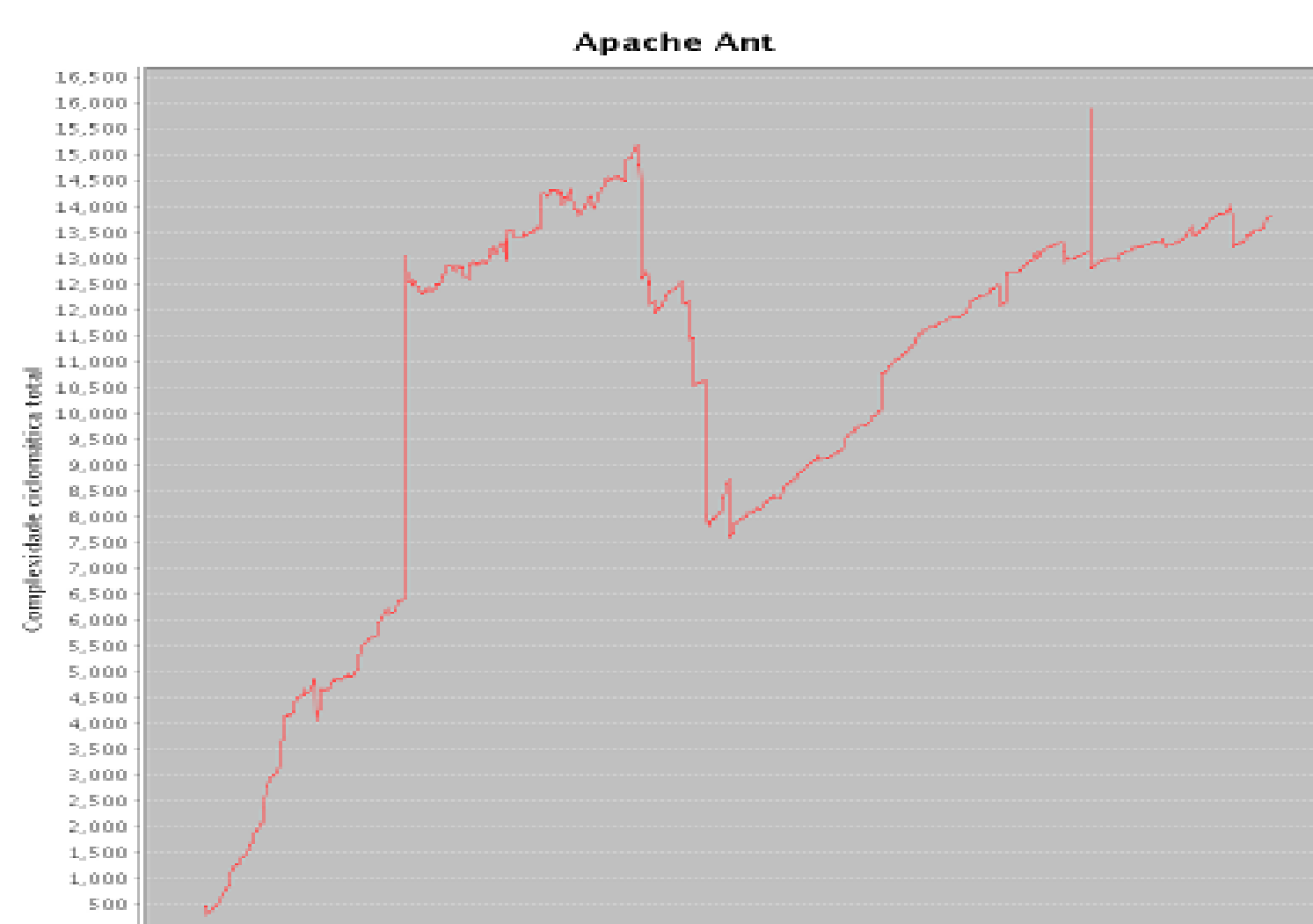
Para avaliar o valor da ferramenta no contexto de pesquisa em Mineração de Repositórios, o estudo publicado em [2] foi reproduzido e estendido utilizando os dados minerados no MetricMiner.

Neste trabalho, os autores estudam o efeito de refatorações sobre a complexidade do sistema. Os autores consideraram a métrica de complexidade ciclomática para medir a complexidade de um software.

O gráfico na Figura 2 exibe a evolução da complexidade ciclomática do software PMD, publicado no artigo original e a Figura 3 exibe a evolução do Ant, com os dados dos minerados no MetricMiner.



**Figura 2:** Complexidade ciclomática do projeto PMD ao longo do tempo publicado em [2]



**Figura 3:** Complexidade ciclomática do projeto Ant ao longo do tempo

Os autores classificam os commits em duas categorias, os que contêm “refatorações documentadas” e os que não contêm. Uma “refatoração documentada” é um commit que possui em sua mensagem palavras como “refactoring”, “refactored”, “refactor”.

Com essa classificação, os autores analisam o efeito sobre a métrica de complexidade ciclomática (CC) dos commits de cada categoria. A tabela 1 exibe os resultados da classificação do conjunto de commits analisado.

	Decrementam CC	Não alteram CC	Incrementam CC
Refatorações documentadas	14	17	12
Sem refatoração documentada	27	580	136

**Tabela 1:** Resultados obtidos em [3]

A mesma análise foi reproduzida sobre um volume de dados muito maior, minerados de 256 projetos da Apache com o MetricMiner. A tabela 2 exibe os resultados das refatorações sobre a complexidade desses projetos. Assim como no estudo original, observa-se que as refatorações de fato não contribuem com a diminuição da complexidade de um software.

	Decrementam CC	Não alteram CC	Incrementam CC
Refatorações documentadas	1504	1603	3230
Sem refatoração documentada	30145	99580	121239

**Tabela 2:** Resultados obtidos com o MetricMiner

## 5. Conclusão

O MetricMiner é uma ferramenta de valor na área de Mineração de Repositórios de Software, permitindo que estudos nessa área sejam realizados sem a necessidade de configuração de ferramentas locais. Além disso, o sistema já conta com uma base de dados grande disponível para consultas de pesquisadores na área.

## 6. Referências bibliográficas

- [1] Kagdi, H., Collard, M. L., and Maletic, J. I. (2007). *A survey and taxonomy of approaches for mining software repositories in the context of software evolution*.
- [2] Soetens, Q. D. and Demeyer, S. (2010). *Studying the effect of refactorings: a complexity metrics perspective*.