

# Construção de um Dataset para Análise da Evolução dos bugs de Sistemas

---

Pedro Araujo Pires

# O que é?

---

- O que é?
  - Conjunto de informações sobre alguns sistemas, geradas para várias versões (180, uma a cada 2 semanas).
- Quais informações?
  - Modelo FAMIX do código fonte.
  - Métricas de código fonte (CK + 11).
  - Tempo de vida de bugs no sistema.

# Quais sistemas?

---

- Eclipse JDT Core
- Eclipse PDE UI
- Lucene
- Equinox Framework
- Geronimo
- Hadoop MapReduce
- TV-Browser
- Hibernate

# O que já foi feito?

---

- Repositório SCM:
  - Download das 180 versões de cada sistema.
  - Download do *log* dos *commits* dos sistemas, e extração das seguintes informações:
    - Data de criação do *commit*.
    - Mensagem do desenvolvedor.

# O que já foi feito?

---

- Código fonte:
  - Remoção das classes de teste.
  - Geração do modelo FAMIX (arquivo MSE) para cada versão.

# O que já foi feito?

---

- Issue Tracker
  - Download das *issues* dos repositórios.
    - *Bugs*
    - Fixed
  - Extração das informações relevantes:
    - Data de abertura.
    - ID do *bug*.

# O que já foi feito?

---

- Geração dos dados:
  - Mapeamento de *bugs* para classes.
    - Identificação do ID do *bug* na mensagem de *commit*.
    - Presença do ID na mensagem indica quais arquivos (classes) foram modificados para a resolução do *bug*.
  - Cálculo do tempo de vida dos *bugs* em cada sistema.
    - Início: data de criação da *issue* no *issue tracker*.
    - Fim: data do *commit* onde o bug foi corrigido.
  - Geração das métricas de código fonte.

# O que falta fazer?

---

- Análise dos dados:
- Validação das informações do dataset, correlacionando anomalias nos dados com eventos ocorridos no período sendo analisado.