### 研究内容

从Github上获取开源项目信息，提取该项目存在与变异算子对应的bug修复模式。

1. 如何获取项目信息？获取到的信息包含哪些？信息之间的关联是什么？
2. 如何定义变异算子？
3. 如何识别相应的模式？

### 总体架构



图 1：总体架构图（简单）

### Github提供的API

### MongoDB使用

### 爬取关键字

### 词向量

### 数据集定义

项目名称：owncloud/android

## 程序文档：

### GithubRepo

GithubRepo：github数据仓库。一个GithubRepo对象实例对应一个代爬取的Android项目。

#### 字段self

user：仓库的用户

repo：仓库的名称

根据这两个即可确定Github获取信息的地址：

如 user = owncloud，repo=android，则：

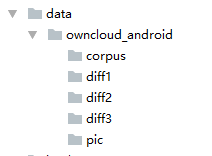
项目访问地址（浏览器） <https://github.com/owncloud/android>

##### name:

owncloud\_android

##### data\_dir:

F:\LM\A\_Android\_bugģʽƥ���ھ�\AndroidBugMutator\data\owncloud\_android



##### url:

* <https://api.github.com/repos/owncloud/android/issues>
* <https://api.github.com/repos/owncloud/android/commits>
* {'Authorization': 'token ef373502bf40d0705e70a73d4935390bde7aebf8', 'Accept': 'application/vnd.github.v3+json', 'user-agent': 'Mozilla/5.0 (X11; Linux x86\_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36'}

##### mongodb

* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'comments')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'issues')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'commits')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'commitsinfo')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'commentsinfo')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'issuesinfo')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'issuesinfo2')
* Collection(Database(MongoClient(host=['localhost:27017'],document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'issuesinfo3')
* Collection(Database(MongoClient(host=['localhost:27017'], document\_class=dict, tz\_aware=False, connect=True), u'owncloud\_android'), u'diffinfo')

#### \_\_init(self,user,repo)\_\_方法:

初始化以上方法

##### self.init\_url()：初始化爬取url

##### self.init\_filepath()：初始化数据存储路径

##### self.init\_mongodb\_coll()：初始化数据库集合