

GitHub Parser

Студент: Антон Мордберг

Научный руководитель: Дмитрий Калашников

СПбАУ, весна 2016



Введение

- Цель: требуется определять популярность используемых библиотек в проектах с открытым исходным кодом на GitHub
- Проблемы
 - Ограничения на количество запросов
 - Отсутствие необходимых запросов среди предоставляемых GitHub API
 - Необходимость интегрировать сервис в общую систему обработки

Задачи

- Построение дерева запросов к GitHub API
- Пред- и постобработка запросов на основании файла конфигурации
- Получение файлов с исходным кодом
- Разработка схемы хранения результатов
- Интеграция сервиса в общую систему

Конфигурация запросов

```
"update": {  
    "credentials": [...],  
    "query" : {  
        "id": "main search",  
        "type": "search_repo",  
        "mongo_collection": "java_repos2",  
        "language": "Java",  
        "iterate": {  
            "start": "2010-11-01T11:23:28Z",  
            "end": "2010-11-02T11:23:28Z",  
            "step_size": 1,  
            "granularity": "day"  
        },  
        "nested" : {  
            "id": "search files",  
            "type": "search_files",  
            "filenames": [  
                "pom.xml",  
                "build.gradle"  
            ],  
            "mongo_collection": "build_files2",  
            "nested": {  
                "id": "links",  
                "type": "get_link",  
                "mongo_collection": "build_files_links2",  
                "nested": {  
                    "id": "download",  
                    "type": "download_file"  
                }  
            }  
        }  
    }  
},  
}
```

Конфигурация постобработки

```
"dependencies": {  
  "language": "Java",  
  "period": {  
    "start": "2008-11-01T11:23:28Z",  
    "end": "2016-12-01T11:23:28Z"  
  },  
  "listOfDependencies": {  
    "slf4j": "org.slf4j",  
    "json": "org.json",  
    "mongodb": "org.mongodb",  
    "appcompat": "com.android.support",  
    "android.gms": "com.google.android.gms"  
  },  
  "delim": ";",  
  "output": "dep-result.csv"  
}
```

Диаграмма классов

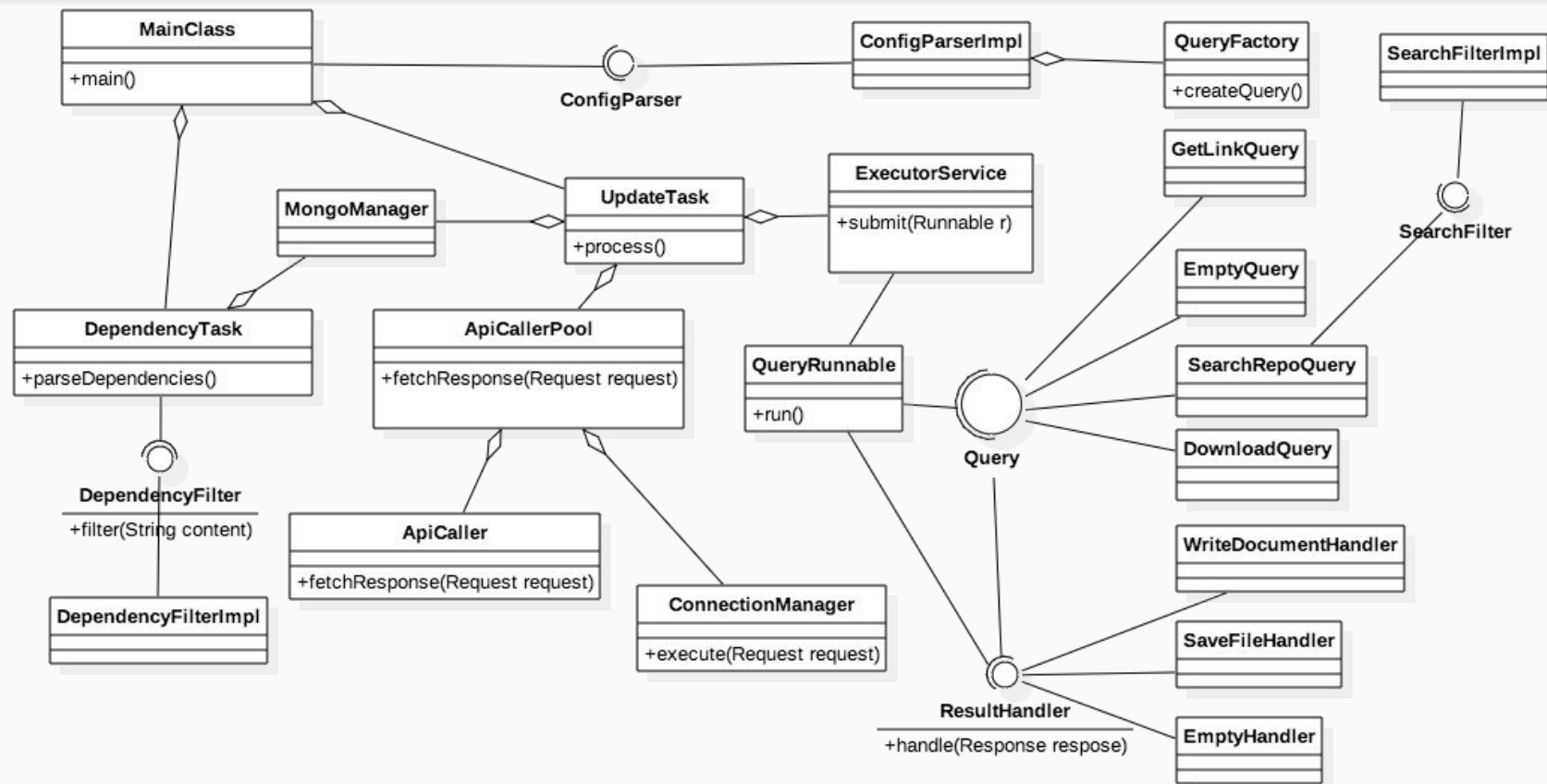


Диаграмма последовательности: запросы

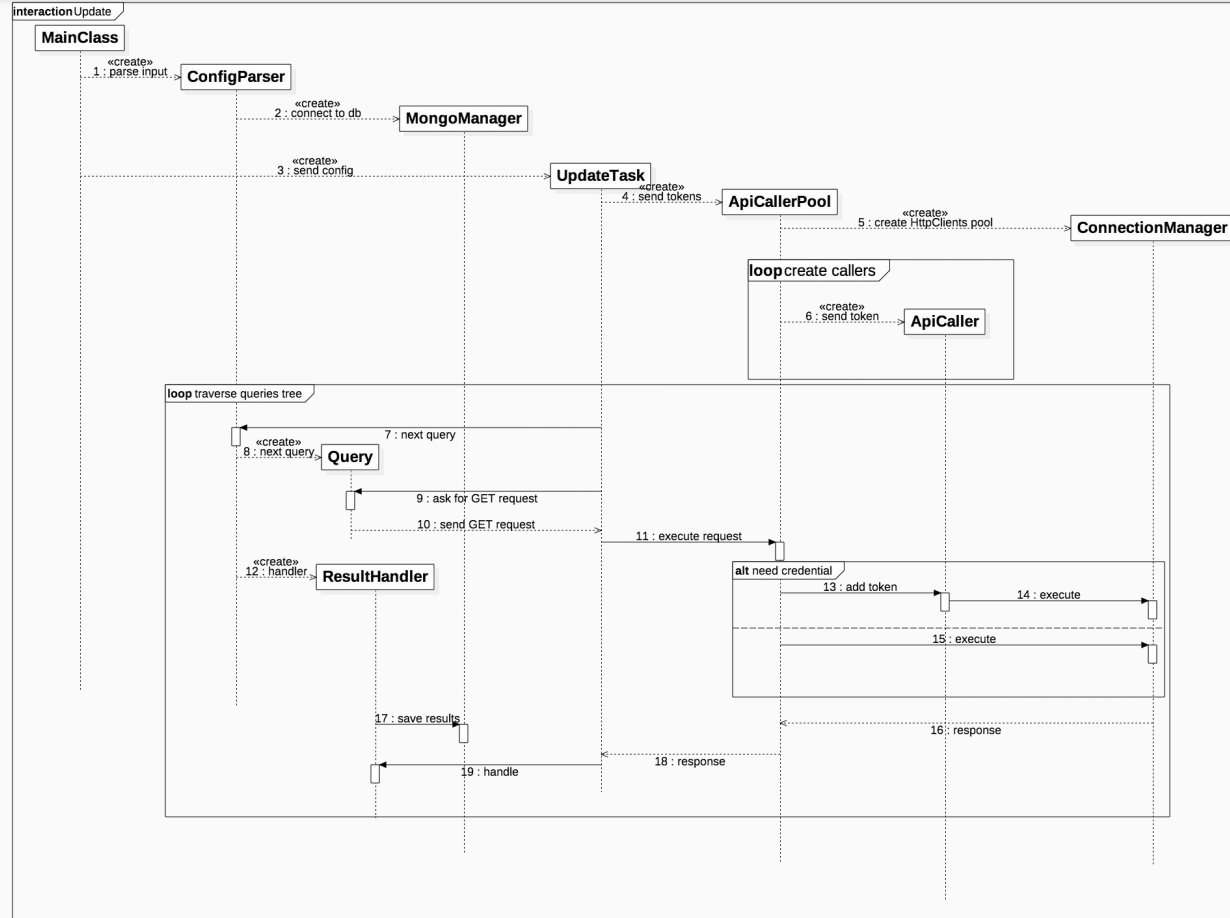


Диаграмма последовательности: поиск зависимостей

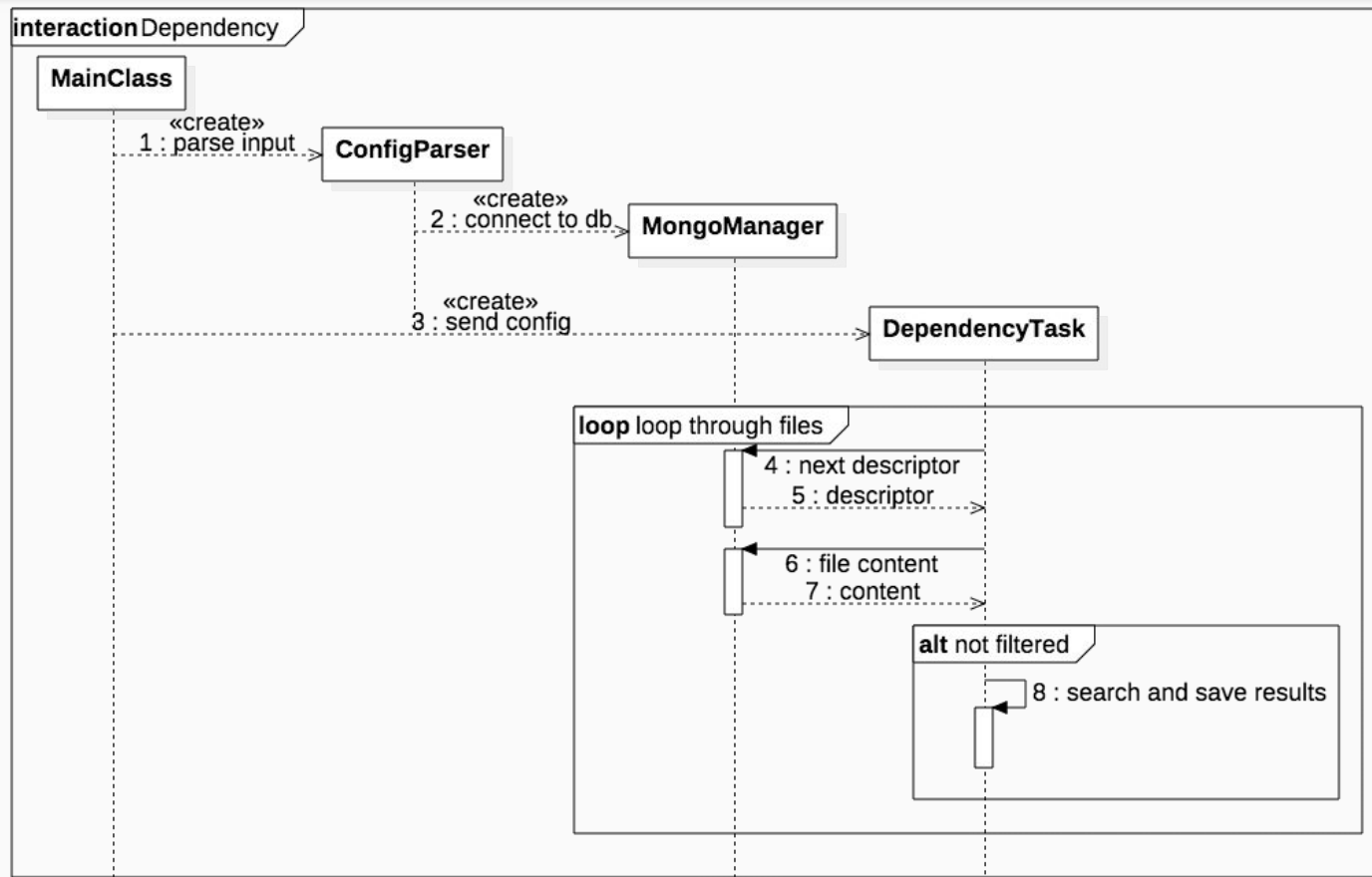


Диаграмма использования

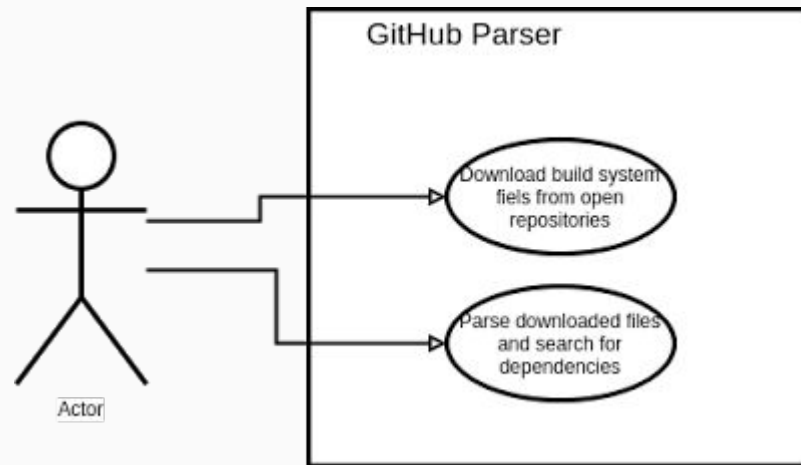
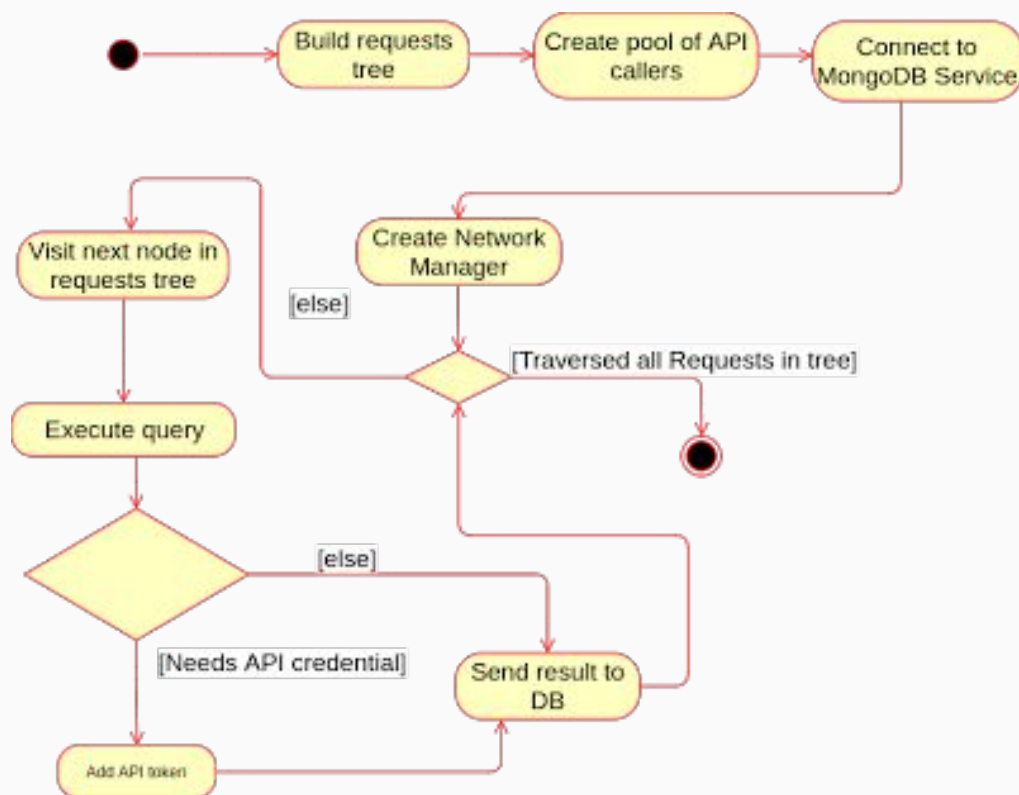


Диаграмма активности



Следующие этапы

- Увеличить покрытие кода тестами
- Создание REST сервиса на основе Spring MVC
- Интеграция сервиса в общую систему обработки

Спасибо за внимание