
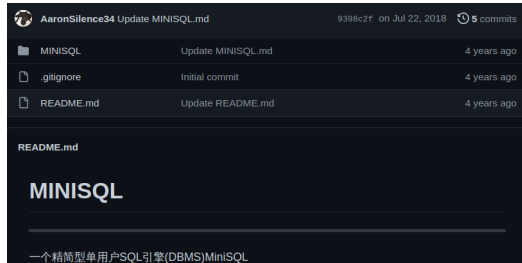

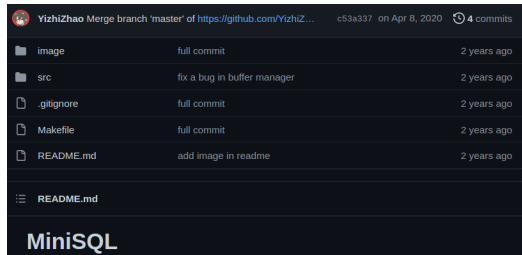


爬取了github上80个（共264个）C++ minisql项目，以默认的80%相似度为阈值检测克隆，得到结果（摘取很小一部分结果分析）：

```
1 38,428,1,6
2 38,431,1,9
3 40,459,1,2
4 40,463,1,6
5 40,467,1,10
6 40,463,38,428
7 60,674,40,467
8 60,674,1,10
```

不妨挨个分析

项目id和项目的对应关系在 [bookkeeping_projs/](#) 中

| project_id | project | 缩略图 |
|------------|------------------------|--|
| 1 | callMeName/MiniSql |  |
| 38 | AaronSilence34/MINISQL |  |
| 40 | trainsn/MiniSQL |  |
| 60 | YizhiZhao/MiniSQL |  |

| | |
|---------------------|---------------------------|
| project_id, file_id | file |
| 1,2 | code/API.h |
| 1,6 | code/BufferManager.cpp |
| 1,9 | code/CatalogManager.h |
| 1,10 | code/Condition.cpp |
| 38,428 | src/BufferManager.cpp |
| 38,431 | src/CatalogManager.h |
| 40,459 | minisql/API.h |
| 40,463 | minisql/BufferManager.cpp |
| 40,467 | minisql/Condition.cpp |
| 60,674 | lib/Condition.cpp |

于是分析结果就是：

- callMeName/MiniSql/code/API.h 与 trainsn/MiniSQL/minisql/API.h 相似：

```

1 #ifndef API_H
2 #define API_H
3
4 #include "Attribute.h"
5 #include "Condition.h"
6 #include "Minisql.h"
7 #include "IndexInfo.h"
8 #include <string>
9 #include <cstring>
10 #include <vector>
11 #include <stdio.h>
12
13 class CatalogManager;
14 class RecordManager;
15 class IndexManager;
16 class API{
17 public:
18     RecordManager *rm;
19     CatalogManager *cm;
20     IndexManager *im;
21     API(){}
22     ~API(){}
23
24     void tableDrop(string tableName);
25     void tableCreate(string tableName, vector<Attribute>* attributeVector,
26 string primaryKeyName,int primaryKeyLocation);
27
28     void indexDrop(string indexName);
29     void indexCreate(string indexName, string tableName, string
30 attributeName);
31
32     void recordShow(string tableName, vector<string>* attributeNameVector =
33 NULL);
34     void recordShow(string tableName, vector<string>*
35 attributeNameVector, vector<Condition>* conditionVector);
36
37     void recordInsert(string tableName,vector<string>* recordContent);
  
```

《不演了》

- callMeName/MiniSql/code/BufferManager.cpp 与 AaronSilence34/MINISQL/src/BufferManager.cpp 与 trainsn/MiniSQL/minisql/BufferManager.cpp 相似：

导入数据库之后的克隆结果:

```
MariaDB [minisqls]> select * from projectClones;
+-----+-----+-----+-----+-----+-----+-----+-----+
| id | cloneId | cloneClonedFiles | cloneTotalFiles | cloneCloningPercent | hostId | hostAffectedFiles | hostTotalFiles | hostAffectedPercent |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | 2 | 1 | 23 | 4.350 | 35 | 1 | 38 | 2.630 |
| 2 | 2 | 5 | 23 | 21.740 | 36 | 5 | 20 | 25.000 |
| 3 | 14 | 1 | 34 | 2.940 | 15 | 1 | 21 | 4.760 |
| 4 | 16 | 1 | 17 | 5.880 | 42 | 1 | 18 | 5.560 |
| 5 | 26 | 1 | 25 | 4.000 | 27 | 1 | 13 | 7.690 |
| 6 | 28 | 14 | 30 | 46.670 | 28 | 14 | 30 | 46.670 |
| 7 | 1 | 6 | 20 | 30.000 | 40 | 6 | 20 | 30.000 |
| 8 | 1 | 2 | 20 | 10.000 | 25 | 2 | 18 | 11.110 |
| 9 | 5 | 4 | 4 | 100.000 | 49 | 4 | 4 | 100.000 |
| 10 | 23 | 1 | 25 | 4.000 | 42 | 1 | 18 | 5.560 |
| 11 | 25 | 1 | 18 | 5.560 | 40 | 1 | 20 | 5.000 |
| 12 | 35 | 1 | 38 | 2.630 | 50 | 1 | 29 | 3.450 |
| 13 | 74 | 8 | 178 | 4.490 | 74 | 8 | 178 | 4.490 |
| 14 | 76 | 4 | 206 | 1.940 | 76 | 4 | 206 | 1.940 |
+-----+-----+-----+-----+-----+-----+-----+-----+
14 rows in set (0.001 sec)

MariaDB [minisqls]> select * from projects where projectId = 5;
+-----+-----+-----+
| projectId | projectPath | projectUrl |
+-----+-----+-----+
| 5 | /home/qiu/Downloads/minisqls/elfff1_minisQL.zip | NULL |
+-----+-----+-----+
1 row in set (0.000 sec)

MariaDB [minisqls]> select * from projects where projectId = 49;
+-----+-----+-----+
| projectId | projectPath | projectUrl |
+-----+-----+-----+
| 49 | /home/qiu/Downloads/minisqls/elfff1_Interpreter.zip | NULL |
+-----+-----+-----+
```