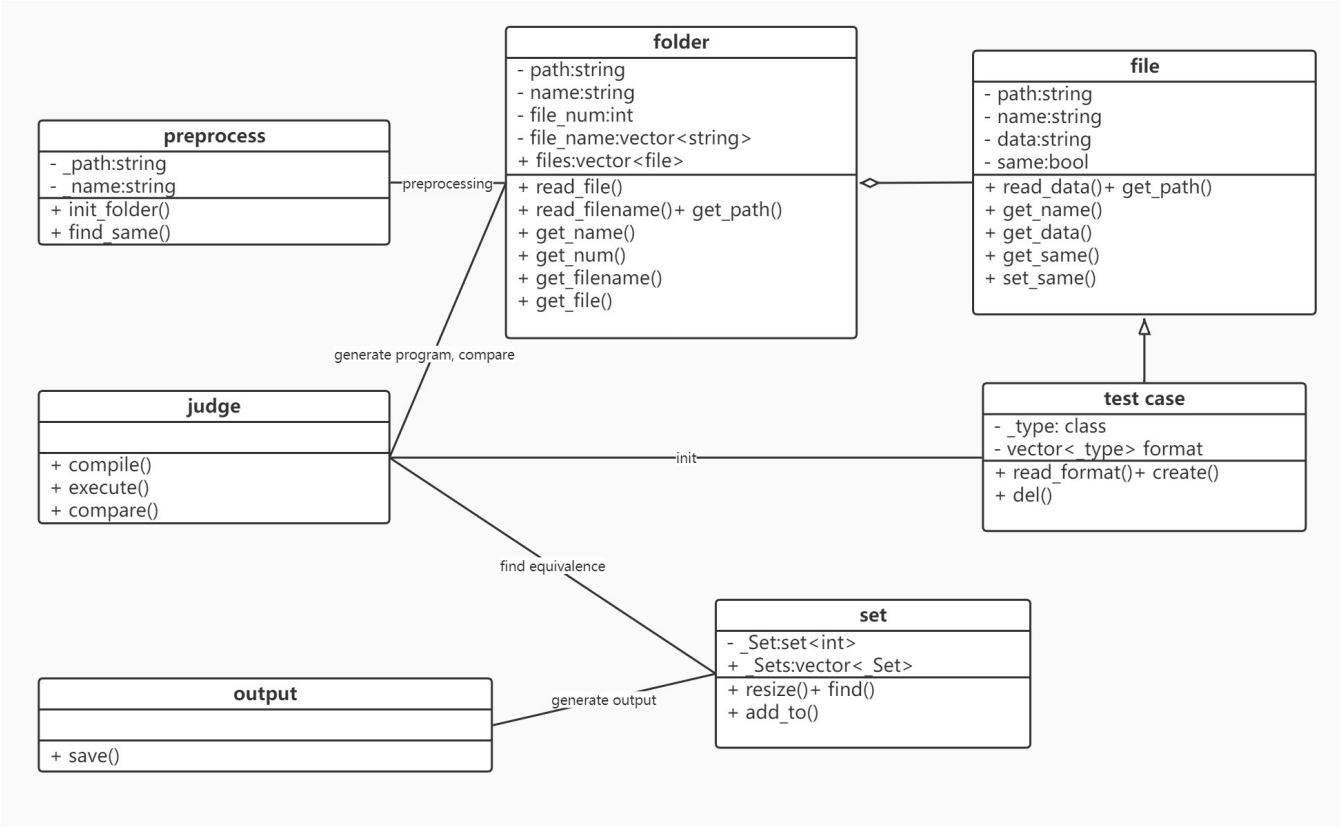
lab4 实验报告

一、模块介绍

设计的等价判断工具可以划分为三个模块：预处理preprocess模块，等价判断judge模块，输出output模块。



程序默认需要与输入input文件夹放置在同一目录下。预处理模块初始化时，以此目录和input文件夹名为参数。之后对input下的指定的文件夹，检查并标记完全相同的文件。若两个文件相同，则后续可以忽略序号较大的文件，只考虑序号较小者与其他文件比较，最后直接得到与这个文件等价的相应结果。

等价判断模块是主体模块，包含三个方法：`compile()`方法调用g++编译不相同的文件，并对每个不相同文件生成同名文件夹，用于后续存放输出结果；`execute()`方法对每个可执行文件，运行，使用生成的测试用例并得到若干相应输出，以`i.txt`形式存放于先前生成的文件夹中；`compare()`方法对两两文件夹中的相对应的输出进行比较，假设一个子文件夹dir中包含n个互不相同的OJ程序，两两比较，需要比较C_n^2次，若所有输出都对应相同，则说明等价，在set中记录。最后为减少冗余，将删除之前生成的文件夹。

输出模块对于set中的记录，在input文件夹同级目录下生成output文件夹，文件夹中共两个文件，`equal.csv`保存等价的OJ程序的相对路径对，`inequal.csv`保存不等价的OJ程序的相对路径对。

二、关键代码介绍

程序的数据结构组成主要是三个类：文件夹folder类，文件file类，和存放等价结果的sets类，而与测试用例有关的testcase类继承自file类。

1. file 类

file以相对路径和文件名为参数初始化，same变量表示是否与其他文件相同。

`read_data()`

获取该文件内容，使用`ifstream`下`istreambuf_iterator<char>`直接得到。

2. testcase 类

每个子文件夹dir中有一个特殊文件`stdin_format.txt`，保存OJ标准输入格式。继承自`file`的`testcase`类以`read_data()`方法可以获得文本内容。在`./input/dir`目录下新建`/testcase`目录，其中存放生成的测试用例。

考虑到OJ程序的标准输入格式中，每行由多个type组成，其中type只包括int、char和string，以及(a,b)表示相应上下界，使用了一个类__type保存这些信息，并用一个数组`vector<_type>`存储所有type。

`read_format()`

要从读取的文本字符串中获取每个__type的信息，使用正则表达式解析，int、char和string对应的正则表达式如下：

```
regex r0("(int)\\((-?[0-9]+),(-?[0-9]+)\\)");
regex r1("(char)");
regex r2("(string)\\(([0-9]+),([0-9]+)\\)");
```

使用smatch容器与分组直接得到类型type和a、b，存入vector中。

`create()`

根据vector中保存的解析结果构造相应的随机输入测试用例。随机字符串string可以考虑先随机生成一个要求字符串长度范围内的int，再拼接int次随机字符char，而随机字符char可以考虑字母ASCII码范围内生成随机int来转换，因此只需要考虑生成随机int。随机int使用中的`uniform_int_distribution`得到。

3. folder 类

假设我们要对input文件夹下的某文件夹dir中的所有文件得出等价结果，则文件夹folder类以`./input`路径和`dir`文件夹名为参数初始化。

`read_file()`

可以得到dir文件夹下的所有源文件。使用了C++17中的文件系统`filesystem`，用`directory_iterator`找到所有的源文件，将文件名放入`file_name`vector中，根据路径和文件名初始化一个file类，放入`files`vector中。`read_filename()`方法的区别是只读取文件名而不初始化文件存入vector。

4. set 类

思路是使用一个vector，每个单元存放一个int型set，每个文件根据在文件夹中读取的序号对应一个集合。若序号i对应的文件与序号j对应的文件等价，则`set[i]`中包含j，且`set[j]`中包含i。

`add_to()`

两个序号为参数，将序号j放入`set[i]`中所有的序号对应的集合，且将`set[i]`中所有序号放入`set[j]`，最后将j放入`set[i]`中，另一序号也进行对称操作。

5. output 类

`save()`方法根据`sets`中的结果，对每个集合查询所有比该集合大的序号是否在集合中，若存在则说明序号对应的源文件等价，在equal.csv中写入，否则说明不等价，写入到inequal.csv中。

三、运行流程

新建output文件夹，初始化equal.csv和inequal.csv；

初始化folder类，表示input文件夹，`read_filename()`获取所有的子文件夹dir名。

对子文件夹数量循环，对每个子文件夹dir：

初始化folder类表示dir文件夹，`read_file()`获取所有源文件；

初始化sets类存放结果；

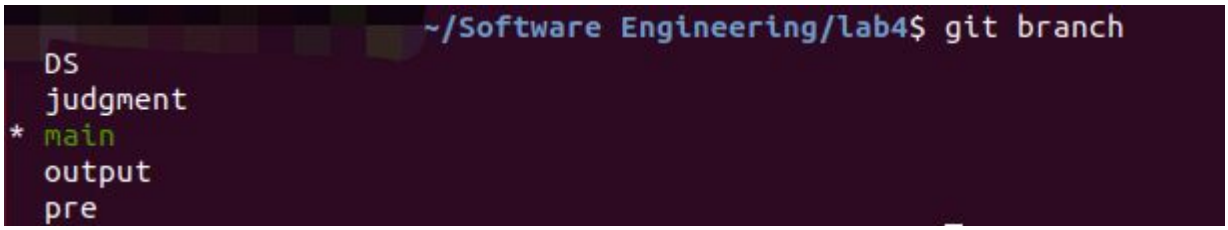
初始化preprocess类，`find_same()`找出相同的文件，标记并保存结果；

初始化testcase类，`create()`生成测试用例；

初始化judge类，`compile()`将所有不同的源文件编译为可执行文件，`execute()`执行测试用例并以`.txt`方式将输出结果保存在同名文件夹中，每个输出结果以序号命名，`compare()`是双层循环，对每一个文件，将其输出结果与序号比他大的文件的输出结果进行比对，若均相同，则等价，保存在sets类中。

四、**git**操作

分为5个分支，代表上述不同模块。



add & git status

```
~/Software Engineering/lab4$ git add .
~/Software Engineering/lab4$ git status
位于分支 DS
您的分支与上游分支 'origin/DS' 一致。

要提交的变更:
  （使用 "git reset HEAD <文件>..." 以取消暂存）

        删除:      input/4A/101036360.cpp
        删除:      input/4A/117364748.cpp
        删除:      input/4A/127473352.cpp
        删除:      input/4A/134841308.cpp
        删除:      input/4A/173077807.cpp
        删除:      input/4A/48762087.cpp
```

commit

```
~/Software Engineering/lab4$ git add .
~/Software Engineering/lab4$ git commit -m "delete irrele
vant"
[judgment 14d0c90] delete irrelevant
 66 files changed, 452 deletions(-)
```

reset

```
:~/Software Engineering/lab4$ git reset
重置后取消暂存的变更:
D       lab4/input/4A/101036360.cpp
D       lab4/input/4A/117364748.cpp
D       lab4/input/4A/127473352.cpp
D       lab4/input/4A/134841308.cpp
D       lab4/input/4A/173077807.cpp
D       lab4/input/4A/48762087.cpp
D       lab4/input/4A/84822638.cpp
D       lab4/input/4A/84822639.cpp
```

merge

```
...~/Software Engineering/lab4$ git merge pre
Merge made by the 'recursive' strategy.
 lab4/pre/preprocess.cpp | 35 ++++++++++++++++++++++++++++++++++
 lab4/pre/preprocess.h   | 16 ++++++++++++++++
 2 files changed, 51 insertions(+)
 create mode 100644 lab4/pre/preprocess.cpp
 create mode 100644 lab4/pre/preprocess.h
```

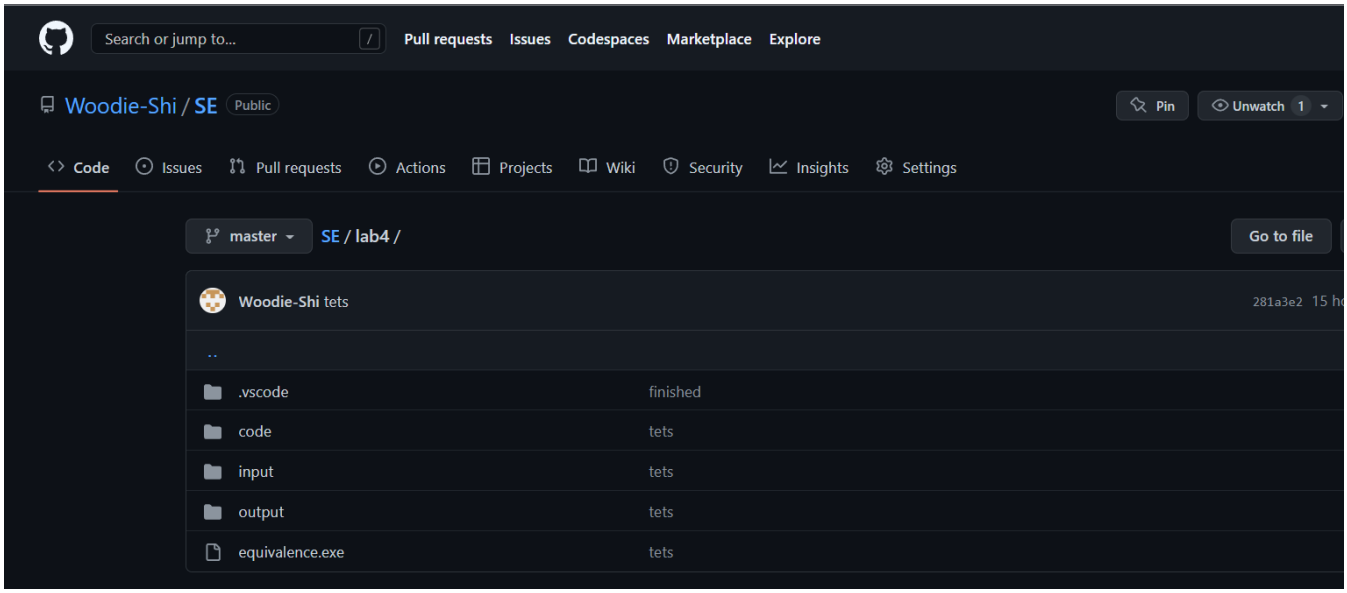
分支合并图

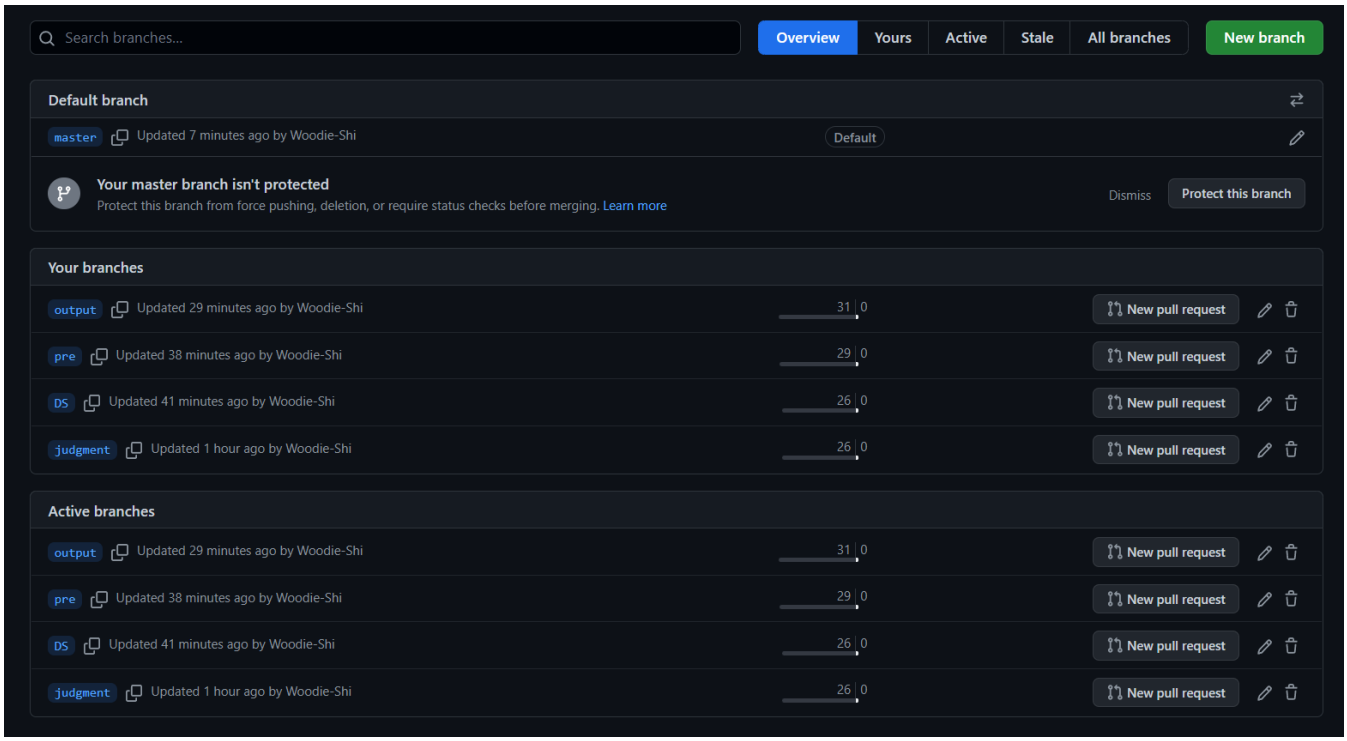
```
* commit 281a3e261dc9aabcedba2b8cf9624ee77b4d67a4 (HEAD -> master, origin/master
)
| Author: Woodie-Shi <woodie.shichenglu@foxmail.com>
| Date:   Wed Nov 16 23:54:26 2022 +0800
|
|     tets
|
*   commit 180d030d326a620aa8eed4d4a00b9175a26da37e
|\  Merge: dcfe40f 6671419
| | Author: Woodie-Shi <woodie.shichenglu@foxmail.com>
| | Date:   Wed Nov 16 23:43:12 2022 +0800
| |
| |     Merge branch 'output' into main
| |
| * commit 66714193d8bac15b7a9f687031e63147199343ff (origin/output, output)
| | Author: Woodie-Shi <woodie.shichenglu@foxmail.com>
| | Date:   Wed Nov 16 23:20:29 2022 +0800
| |
| |     delete input
| |
| * commit 8fb90f81f7c667443bec2e8c78b60fee1eaf04f7
| | Author: Woodie-Shi <woodie.shichenglu@foxmail.com>
| | Date:   Wed Nov 16 22:42:13 2022 +0800
:
```

处理冲突：rm

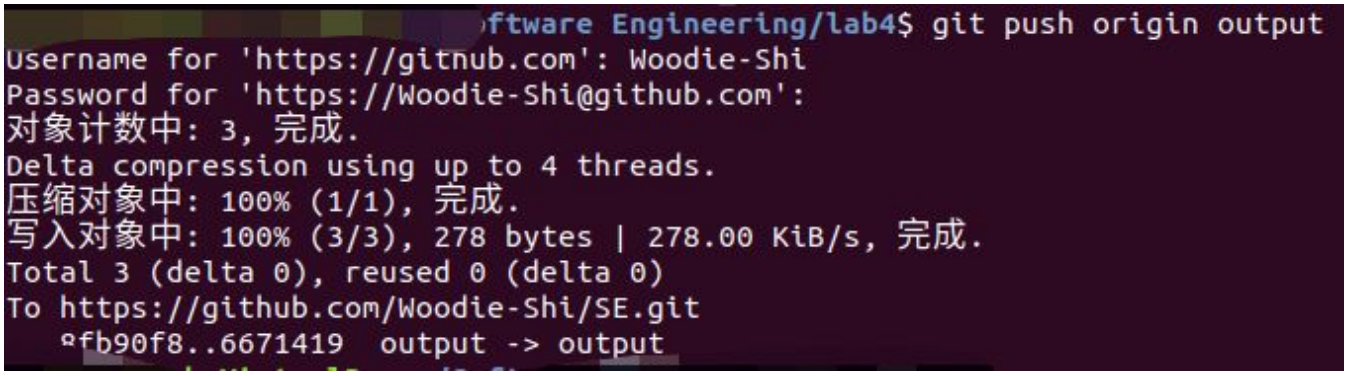
```
:~/Software Engineering/lab4$ git merge judgment
删除 lab4/data structure/main.cpp
冲突（修改/删除）：lab4/data structure/data_structure.h 在 judgment 中被删除，在
 HEAD 中被 修改。lab4/data structure/data_structure.h 的 HEAD 版本被保留。
冲突（修改/删除）：lab4/data structure/data_structure.cpp 在 judgment 中被删除，
在 HEAD 中被 修改。lab4/data structure/data_structure.cpp 的 HEAD 版本被保留。
删除 lab4/data structure/.vscode/settings.json
自动合并失败，修正冲突然后提交修正的结果。
```

github


Search or jump to...
Pull requests
Issues
Codespaces
Marketplace
Explore
Woodie-Shi / SE
Public
Pin
Unwatch 1
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
master
SE / lab4 /
Go to file
Woodie-Shi tets
281a3e2 15 h
..
.vscode
finished
code
tets
input
tets
output
tets
equivalence.exe
tets


Search branches...
Overview
Yours
Active
Stale
All branches
New branch
Default branch
master
Updated 7 minutes ago by Woodie-Shi
Default
Your master branch isn't protected
Protect this branch from force pushing, deletion, or require status checks before merging. Learn more
Dismiss
Protect this branch
Your branches
output
Updated 29 minutes ago by Woodie-Shi
31 0
New pull request
pre
Updated 38 minutes ago by Woodie-Shi
29 0
New pull request
DS
Updated 41 minutes ago by Woodie-Shi
26 0
New pull request
judgment
Updated 1 hour ago by Woodie-Shi
26 0
New pull request
Active branches
output
Updated 29 minutes ago by Woodie-Shi
31 0
New pull request
pre
Updated 38 minutes ago by Woodie-Shi
29 0
New pull request
DS
Updated 41 minutes ago by Woodie-Shi
26 0
New pull request
judgment
Updated 1 hour ago by Woodie-Shi
26 0
New pull request

push


ftware Engineering/lab4$ git push origin output
Username for 'https://github.com': Woodie-Shi
Password for 'https://Woodie-Shi@github.com':
对象计数中: 3, 完成.
Delta compression using up to 4 threads.
压缩对象中: 100% (1/1), 完成.
写入对象中: 100% (3/3), 278 bytes | 278.00 KiB/s, 完成.
Total 3 (delta 0), reused 0 (delta 0)
To https://github.com/Woodie-Shi/SE.git
fb90f8..6671419 output -> output

五、运行说明

可在`lab4-code`文件夹中执行`make`，即得到程序`equivalence.exe`，在`input`同级目录下运行，得到output。