

OOP上机实验(4)

(Version: 0.79 Date: 2025/4/9 Author: wangxp@fudan.edu.cn)

1 实验目标

通过本实验，拟达到如下目标

- 熟悉标准库的容器使用及容器的性能；
- 熟悉标准库的算法；
- 熟悉C++的接口与实现相分离的做法与实践。

2 实验内容

2.0 背景

在互联网中，计算机是通过IP地址进行标识的。一个IP地址是一个32位的无符号整数，不过通常我们会把32位的无符号整数分成4个八位的数值，例如复旦的邮件服务器地址为

202.120.224.10

转换成整数就是 3396919306。而我们对数字不是很敏感，更容易记住有意义的名字，例如

mail.fudan.edu.cn

这个就是学校的邮件服务器的名称。

因此，互联网上的任何一台机器，通常有一个主机名(网络术语叫做域名, Domain Name)，还有对应的IP地址。整个互联网要正常运行的话，必须要把域名转换成IP地址--实际上，每一次我们通过浏览器访问网站时都可能发生好多次这样的转换；如果有转换不成功的话，我们就不能访问 互联网。

本次实验设计一个在本地提供服务的域名查找程序。

2.1 目标

实验目标：设计一个高效率的域名查找程序，使域名的查找、删除、创建效率都非常高效。

本次实验，提供给大家的文件包括：

数据文件: ns20k.txt, ns40k.txt, ns80k.txt, ns224k.txt
源文件: nstest.cc, ns_common.h, nsvector.cc, nsmap.cc, nsumap.cc
库文件: stopwatch.cc, stopwatch_ut.cc; split.h, split.cc
cmake文件: CMakeLists.txt
批处理文件: run.bat, run.sh

其中stopwatch和split为库文件, 不需要修改也不需要提交, 直接使用即可。

2.1.1 文本文件

数据文件的内容包含多行, 每一行中包含域名和对应的IP地址, 中间使用空格分隔, 如下所示

```
coloradovtc.com 427663021 # 在第 20000 行
```

文本文件中包含20k/40k/80k/224k, 分别表示2万个/4万个/8万个/22.4万个域名与IP的对---实际的域名已经超过几千万个, 因此编写的程序效率至关重要。

数据来自[1], 把2020/4/14和2020/4/15日的数据放在一起, 然后通过脚本语言(未提供)生成了现在的文本文件。

2.1.2 接口文件

ns_common.h是一个接口文件, 考虑到会使用不同的数据结构(std::vector、std::map或std::unordered_map)分别实现域名查找服务程序; 同时规定: 不管使用哪一种数据结构实现, 都必须完整实现以下接口函数, 具体如下

```
using HostName = std::string;
using IPAddress = unsigned int;
const IPAddress NON_EXISTING_ADDRESS = 0;

void create(std::istream& in);
bool contain(const HostName&);
void insert(const HostName&, const IPAddress&);
bool remove(const HostName&);

IPAddress lookup(const HostName&);
```

此外, 把所有的代码放在名字空间EXP4中, 以保证使用的名字和其他的名字不冲突。

2.1.3 补充代码

nsvector.cc, nsmap.cc和nsumap.cc中已经有部分代码, 请根据下面的要求补充完整。

- 使用vector实现
nsvector.cc使用std::vector实现名称服务程序。具体地, 把名字和IP地址作为一个pair放在

std::vector中。在实现contain/remove/lookup时，不能使用循环，需使用STL库函数find_if，而且第三个参数使用lambda函数的形式。关于lambda函数的更多信息请参考[2]。

- 使用map实现

nsmap.cc使用std::map实现名称服务程序，其中key是名称，value是IP地址；需要实现ns_common.h中的所有函数。

- 使用unordered_map实现

nsumap.cc使用std::unordered_map实现名称服务程序，其中关于unordered_map的详细信息参考[3]；需要实现ns_common.h中的所有函数。

2.1.4 其他文件

nstest.cc包含测试程序，主要流程是：读入数据；查找在数据库中的主机名；查找不在数据库中的主机名。且在每一个步骤都输出相应的信息。

run.bat是批处理程序(只能在windows下面运行, run.sh可以在WSL或macOS下运行，请注意文件的权限)，执行该程序，会自动进行运行程序，最终的效果如下图：

- Windows下的截图

```
D:\course\2025S\OOP\doc\experiments\lab4\key>run
test create:                using time 25478.9742 ms
test contains:              using time 239.0031 ms
test not contains:         using time 314.2583 ms
test insert/lookup/remove: using time 646.4714 ms
test create:                using time 669.7981 ms
test contains:              using time 0.4938 ms
test not contains:         using time 0.4686 ms
test insert/lookup/remove: using time 1.0352 ms
test create:                using time 665.5264 ms
test contains:              using time 0.5054 ms
test not contains:         using time 0.1960 ms
test insert/lookup/remove: using time 0.2971 ms
```

- WSL下的截图

```
hduser@T40Z:/mnt/d/course/2025S/OOP/doc/experiments/lab4/key$ ./run.sh
test create:                using time 27919.8953 ms
test contains:              using time 236.3060 ms
test not contains:         using time 314.6753 ms
test insert/lookup/remove: using time 632.3943 ms
test create:                using time 857.9741 ms
test contains:              using time 0.2011 ms
test not contains:         using time 0.1816 ms
test insert/lookup/remove: using time 0.6058 ms
test create:                using time 962.7450 ms
test contains:              using time 0.0283 ms
test not contains:         using time 0.0235 ms
test insert/lookup/remove: using time 0.1186 ms
```

2.3.2 提交要求

提交内容：[只要提交nsvector.cc](#)、nsmap.cc和nsumap.cc三个文件；在提交前，确保每个程序都可以正常运行。

3 参考资料

[1] 新注册的域名。参考 <https://whoisds.com/newly-registered-domains>

[2] Lambda表达式。参考 <https://zh.cppreference.com/w/cpp/language/lambda>

[3] unordered_map。参考 https://zh.cppreference.com/w/cpp/container/unordered_map

4 实验要求

4.1 实验提交要求

- 如果有源代码提交的要求，请不要改变源代码原来的名字；
- 如果有截屏的要求，把截屏的内容放在PDF文件中(文件名为：学号_实验编号.pdf)，并在前面写明实验内容的编号和源代码名称；
- 把所有要提交的源代码和PDF文档放在同一个文件夹内，把该文件夹压缩；
- 压缩文件支持zip或rar两种形式。

4.2 压缩文件命名

压缩文件名为：学号_姓名_实验编号.rar 或学号_姓名_实验编号.zip。例如第1次实验，学号为19307130022,姓名为张其的学生，提交的压缩文件名为 22307130022_张其_1.rar或22307130022_张其_1.zip)。请大家务必按照命名规则来正确命名文件的名称，否则可能导致评分无法进行，影响实验评分！

请注意：提交的文件不包括*.obj和*.exe，只需要包含*.cc 和*.h 文件即可。如果有说明，请把说明包含在文本文件中，例如代码demo.cc的说明文件名为demo.cc.txt。

4.3 截至时间

截至时间通常为下周五晚上12点前，请大家安排时间准时完成作业。