

第一次实验报告

软71 沈冠霖 2017013569

November 4, 2018

1 实验目标

本次实验有四个目标：首先，自己实现字符串，栈，链表，哈希表四个基本类，并且尽量做成模板类来复用。其次，能够读取几十万长度的词库，并且进行保存，散列变换。之后，能够对符合要求的网页进行解析，提取其标题，时间，来源，正文。最后，对正文进行中文分词。这一切都是为了构建之后的新闻检索与推荐系统。

2 实验环境

Win10, VS2017下进行开发。

3 抽象数据定义

字符串 定义了一个字符串类NCharString，可以进行空字符串构造（空构造函数），通过char来构造（重载构造函数），赋值($m_{setValue}$) $fi\bar{f}6(m_{Duplicate}), * \Phi m_{cutString} \Psi fi(f_{con})$

链表 定义了查找（查找第i个位置的 $m_{searchAtPlace}, \sim, \Gamma * pnIm_{searchEqual}), (Mn\bar{f}ie(m_{insertAtPlace})$

栈 定义了一个栈NStack，基于链表，里面存储标签和标签内的内容，便于进行解析。可以实现入栈 ($mm_{pushBack} \Psi, \Phi(m_{popBack}), \$/\& :z(m_{judgeEmpty})$).

哈希表 定义了一个29万长度的哈希表，可以把一个我自定义类型的字符串链表转换成哈希，也可以读入一个我自定义的字符串来搜索其位置。此处有引用代码[1]。

具体模块 定义了一个常数类Constant和一个全局常数类变量g_Constant，负责读入开始的几个数字，并存储4个数字，24结果等信息。定义了一个类NumberBase来存储每种情况的结果，及其使用的数字，还有其父母情况。定义了一个类NumberList来生成，存储所有情况，并生成算式。

4 网页解析