

**课 程 设 计 报 告**

**题目： 迷你中文搜索引擎**

**课程名称：**

**专业班级：**

**学 号：**

**姓 名：**

**指导教师：**

**报告日期：**

**计算机科学与技术学院**

**任□务□书** （黑体小2号加粗居中）

* **设计内容**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

选取至少50个中文web页面，由此构成文档集；首先对文档进行解析与分词，以哈希表、BST或AVL等查找树建立文档索引，采用倒排索引技术；然后处理用户搜索与查询，基于用户输入的查询关键字与条件，在上述索引中进行查找；显示满足查找条件的WEB文档。

* **设计要求**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

（1）以文件形式保存索引结构，文档集可扩展；

（2）查询条件支持简单的布尔操作：AND与OR；

（3）分词处理自行实现，不可调用分词库或软件，html文档解析可调用它方支持库；

（4）实现查询结果文档分值量化功能或者界面友好可适当加分。

* **参考文献**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

[1] 严蔚敏, 吴伟民. 数据结构（C语言版）. 北京: 清华大学出版社,1997

[2] 严蔚敏, 吴伟民, 米宁. 数据结构题集（C语言版）. 北京: 清华大学出版社,1999

[3] Christopher D. Manning等著，王斌译. 信息检索导论，人民邮电出版社，2010

[4]搜索引擎技术揭密：中文分词技术, <http://www.williamlong.info/archives/333.html>

[5] **□**任丽芸. 搜索引擎中文分词技术研究. 重庆理工大学硕士学位论文，2011

**目□□录**（黑体小2号加粗居中）

**任务书** I

**1□□□引言** 1

1.1□□×××××× 1

1.2□□×××××× 3

1.2.1□□×××××× 7

1.3□□×××××× 10

**……**

**……**

**……**

3**□□□××××××** 20

3.1□□×××××× 20

3.2□□×××××× 23

3.2.1□□×××××× 25

3.3□□×××××× 30

**……**

**……**

**4□□□** 40

**参考文献** 44

**附录□××××××** 45

（章为宋体小4号加粗，其余宋体小4号，字母、阿拉伯数字为Time New Roman小4号）

**1□引言**（黑体小2加粗居中）

**1.1□课题背景与意义**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

××××××××××××××××××××××××××××××××××××××××××××××××××××××××（宋体小4号，行间距固定1.5倍行距，字符间距为标准）

**1.1.1□××××××**（黑体小4号加粗, 字母、阿拉伯数字为Time New Roman小4号加粗）

××××××××××××××××××××××××

**1.1.2□××××××**

××××××××××××××××××××××××

**1.2□国内外研究现状**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman小4号加粗）

**1.3□课程设计的主要研究工作**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman小4号加粗）

××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

--------章与章之间插入分页符----------

**2□系统需求分析与总体设计**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**2.1□系统需求分析**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

这部分应该写的是用户需求，明确你做的系统要实现的目标，能处理一些什么样的事务、事务处理流程等。

**2.2□系统总体设计**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

这部分可根据用户需求，设计和规划一个系统，说明清楚系统应该有哪些功能模块，每个模块做什么。最后给出完整的系统模块结构图。

以下是文中有图标的形式，要求对图标引用说明。

××××××××××××××××××××××××××××××××××××××××××××××××，其×××××可表示如下：

 (3-1）

 (3-2）

×××××××××××××××××××××××××××× (如表3-1所示)

表3-1□××××××××××

|  |  |  |  |
| --- | --- | --- | --- |
| ××××× | ××× | ××× | ××× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |

（表标题：位于表格上方，黑体小4号，字母、阿拉伯数字为Time New Roman小4号，表内容：宋体5号，字母、阿拉伯数字为Time New Roman 5号）

××××××××××××××××××××××××××× (如图3-1所示)



图3-1□××××××××××

（图标题：位于图下方，黑体小4号，字母、阿拉伯数字为Time New Roman小4号）

.......

--------章与章之间插入分页符----------

**3□系统详细设计**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**3.1□有关数据结构的定义**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

这部分要写的：（1）首先描述系统中要处理那些数据，每种类型的数据包括哪些数据项，每个数据项的数据类型，最后可用一个表格表示出来；（2) 描述这多种数据在系统中如何关联，可通过图直观的说明这多种数据间的关联。

**3.2□主要算法设计**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

这部分主要描述系统中的模块实现的流程，可采用文字配合流程图的方式表示各模块的算法思想及流程。

**4□系统实现与测试**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**4.1□系统实现**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

这部分应该写的是用户需求，明确你做的系统要实现的目标，能处理一些什么样的事务、事务处理流程等。

这部分可首先叙述一下你的系统实现的软硬件环境；

根据3.1的设计，用C语言定义各种数据类型；

程序代码部分在这里不需要给出来，只需要叙述清楚在系统中包括哪些函数，各函数的说明，如何利用这些函数实现系统各模块的功能，以及函数间的调用关系（可用图表示出来）。程序详见附录。

**4.2□系统测试**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

首先叙述一下常用的软件测试方法，在选择几个主要的功能模块（自行掌握数量，关键要体现你的水平的一些模块）描述测试过程，（1）先明确模块的功能、设计目标等。（2）分析、叙述如何选取测试数据，要求有完整的测试大纲。（3）运行结果（这时可用截图）。（4）分析运行结果、确认程序满足该模块的设计目标。

**5□总结与展望**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**5.1□全文总结**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

对自己的工作做个总结，主要工作如下：

（1）对。

（2）。

（3）

。

**5.1□工作展望**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

在今后的研究中，围绕着如下几个方面开展工作。。。。。。。。

（1）。

（2）

# 6□体 会(黑体小2号加粗居中)

这部分就自由发挥了。

**参考文献** (黑体小2号加粗居中)

[1]□王静康,张凤宝,夏淑倩等.论化工本科专业国际认证与国内认证的“实质性”.高等工程教育研究,2014,5:1-4

[2]□Stone J A, Howard L P. A simple technique for observing periodic nonlinearities in Michelson interferometers. Precision Engineering,1998,22(4):220-232

[3]□朱印红,袁衍明.Dreamweaver完美网页设计——技术入门篇.(第一版).北京:中国电力出版社,2006:19～20

[4]□Lewis S L. Physics and chemistry of the solar system.北京:北京大学出版社,2014.1～2

[5]□陈剑.上博简《民之父母》“而得既塞於四海矣”句解释[EB/OL］.简帛研究网站，http://www.bamboosilk.org/Wssf/2003/chenjian03.htm．2003-01-18

( 宋体小4号)

……

……

……

**附录** (黑体小2号加粗居中)

×××××××××××××××××××××××××××

( 宋体小4号)

……

……

……