

# 项目计划书

## 蜘蛛爬网

1 引言.....	2
1.1 编写目的.....	2
1.2 背景.....	2
1.3 定义.....	2
1.4 参考资料.....	2
2 项目概述.....	3
2.1 工作内容.....	3
2.2 主要参加人员.....	3
3 项目实施.....	4
3.1 流程图.....	4
a 总体项目流程图.....	4
b 网址分析流程图.....	4
c 数据库操作流程图.....	5
3.2 程序.....	5
a 主函数：.....	5
b 各函数模块功能：.....	6
4 实施计划.....	8
4.1 工作任务的分解与人员分工.....	8
4.2 进度.....	8
4.3 预算.....	8
4.4 关键问题.....	8
5 支持条件.....	9
5.1 计算机系统支持.....	9
5.2 老师提供的支持.....	9

# 1 引言

## 1.1 编写目的

通过此项目的学习，我们对一个项目的整体流程有所掌握，编程能力有新的提高，团队的合作更加默契。

## 1.2 背景

项目名称：蜘蛛爬网

项目提议者：何老师（亚嵌金牌讲师）

项目起因：随着互联网的飞速发展，寄存在网络中的网址越来越多，该项目的目的就是把互联网中的网址都找出来。

## 1.3 定义

给你一个起始网址 e.g. [www.haol23.com](http://www.haol23.com) 查找出其下的所有网址，再在其中的每个网址为基点查找出其下的所有网址，该过程像试蜘蛛织网，所以起名为蜘蛛爬网。理论上给你一个起始网址，能将整个互联网的所有网址找出来。

## 1.4 参考资料

- [1] 《Linux 开发工具 - 项目开发的最有效途径》清华大学出版社（美）John Fusco 著
- [2] 《亚嵌项目实训营 - 理论教材》
- [2] 《C 语言教程》
- [3] 《嵌入式 Linux 应用开发完全手册》韦东山 编著 人民邮电出版社

## 2 项目概述

### 2.1 工作内容

第一阶段:文献调研,理论分析。查阅各种资料,对蜘蛛爬网的前景,实施方案,以及难点做一个估计。同时进行必要的需求分析。

第二阶段:熟悉 linux 下 C 编程知识,准备项目所需的软环境。从软件角度,包括操作系统与应用软件:linux 操作系统, vim 文本编辑器, gcc 编译工具, gdb 调试工具等。

第三阶段:编程部分的实现,先出个个模块的实现到整体的完成。

第四阶段:进行项目的调试,测试其性能指标。

### 2.2 主要参加人员

项目经理:黄炳

项目技术总监:陈志强

项目开发成员:尧彦帮、陈练俊、饶周

文档编写:陈志强

## 3 项目实施

### 3.1 流程图

#### a 总体项目流程图

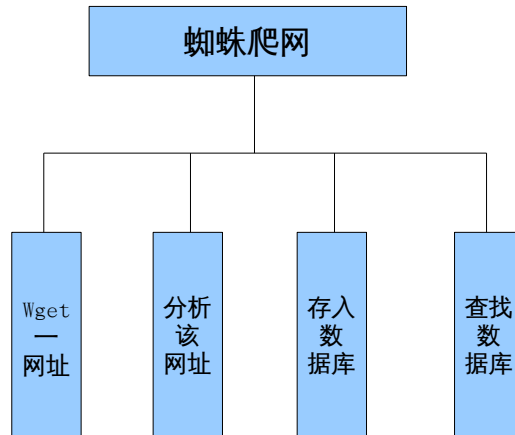


图 1 总体项目流程图

#### b 网址分析流程图

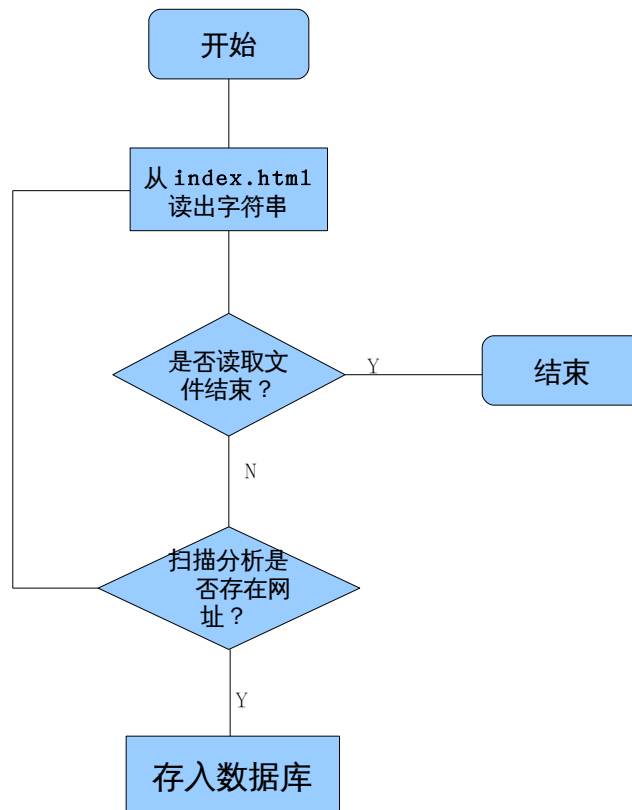


图 2 分析网址流程图

### c 数据库操作流程

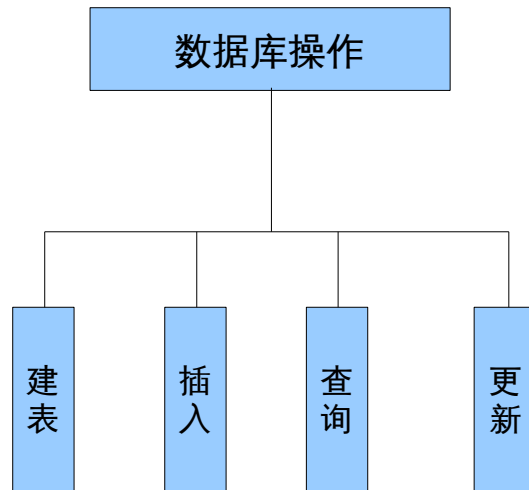


图 3 数据库操作框图

## 3.2 程序

### a 主函数：

```
int main (void)
{
    char *temp ;
    int i = 5 ;

    get_web = malloc(sizeof(struct db_web));

    init_db () ;

    while (1)
    {
        temp = malloc(512);
```

```
select_db ();

wget_op (get_web->domain , get_web->page) ;
free(get_web->page) ;
get_web->page = NULL;

get_link_op () ;           //获取出网址

update_db (get_web->id , 2);

free(temp);
temp = NULL;

}

free(get_web);
get_web = NULL;

sqlite3_close (db);

return 0 ;
}
```

## **b 各函数模块功能：**

void create\_db (void) ;

创建数据库。

int search\_callback(void \*p, int argc, char \*\*argv, char \*\*argvv) ;

查找回调函数，通过返回值判断要插入数据库的数据是否已经存在。

void insert\_db (char \*domain , char \*page , int status ) ;

对数据库进行插入操作函数。

void is\_insert\_db (char \*domain , char \*page , int status) ;

判断该网址是否要插入数据库函数，如已存在则不插入，不存在则插入。

```
int call_back (void *p , int argc , char **argv , char **argvv );
```

查询数据库数据用到的回调函数，用来传出查询的结果，用来下一个 wget 操作。

```
int select_db ();
```

查询数据库函数。

```
void update_db (int id , int status);
```

数据库更新函数，用改变已存的状态属性，达到对一数据只 wget 一次。

```
void wget_op (char *domain , char *page);
```

完成 wget 操作函数。

```
void init_db (void);
```

初始化数据库函数，将第一个网址存入数据库。

```
char get_char (char *data);
```

获取一个数据函数。

```
char * is_string (char *data,char * dst);
```

字符串判断函数，用来判断 data 中有没有 dat 字符串。

```
char * get_address (char * data);
```

提出一网址函数，当判断完网址前面的几个关键字就调用此函数提出该网址。

```
char * get_one_result (char *data);
```

获取一网址函数，包括完成网址关键字的判断工作。

```
void separate_insert_web (char *temp);
```

将一个绝对网址的域名和网页分开后并插入数据库函数。

```
void web_insert_into_db (char *domain , char *temp);
```

判断一个 download 分析下来的连接是绝对还是相对网址，并存入数据库。

```
void get_link_op ( void );
```

对每一次 wget 得到的文件进行分析处理的相关操作函数。

## 4 实施计划

### 4.1 工作任务的分解与人员分工

项目经理：黄炳，负责整个项目的规划我任务安排。

项目技术总监：陈志强，负责技术方面问题的功课，和对组成员关于技术方面问题进行交流，设计出合理的方案。

项目开发成员：尧彦帮、陈练俊、饶周，负责项目各大板块的实现。

### 4.2 进度

第一天：在组成员的谈论下，对该项目有明确的思路，在组长的协调分工下，已完成项目的计划书。

第二天：在组长的协调下，对每个成员进行了任务的分配，各自都完成自己的任务，也就是项目实施的前期准备工作，查看相关文献等等。

第三天：组成员都拿出各自的问题进行了讨论，在大家的共同努力下，各自的问题得到了充分的解决，各成员对自己的任务有了大致的框架。

第四天：各自在完成自己的任务，同时也在不断的交流。

第五天：各成员的任务都已大部分完成。

第六天：项目技术总监陈志强将各成员的成果加于整合，并进行调试。

第七天：项目经理总结了该项目完成过程，并宣布该项目圆满完成。

### 4.3 预算

该项目是在亚嵌项目实训营中完成的，其中涉及到的软件、硬件设备都已具有，所以耗资为0元。

### 4.4 关键问题

项目过程中遇到的一些苦难问题：

- a.项目计划书不知从何下手
- b.项目的分工问题比较棘手
- c.数据库操作不熟练
- d.项目完成过程中的一些细节问题导致调试时十分的困难



## 5 支持条件

### 5.1 计算机系统支持

操作系统：(Linux) Ubuntu-8.04

### 5.2 老师提供的支持

何老师（亚嵌金牌讲师）为我们讲述了整个项目的大概流程，并在技术上提供了大量的帮助支持。