

系统开发

G04 网络热词聚合搜索引擎

组长：徐霄雯
组员：杨樾人、李冻秋、李珏宇

联系方式
Sylviehsu@outlook.com

日期
2018/12/21

目录

- 1 文档概述..... 2
 - 1.1 编写目的
 - 1.3 范围
 - 1.4 定义
- 2 项目概述..... 4
 - 2.1 工作内容
 - 2.2 产品
 - 2.4 计划批准者和批准日期
 - 2.5 项目完成期限
- 3 系统编码..... 6
 - 3.1 编码语言
 - 3.2 版本控制
 - 3.3 开发工具
 - 3.4 编码规范
- 4 实现计划..... 8
 - 4.1 开发过程
 - 4.2 接口人员
 - 4.3 进度
- 5 支持条件..... 9
 - 5.1 计算机系统支持
 - 5.2 需要用户承担的工作
 - 5.3 需由外单位提供的服务
- 6 输出要求.....16
- 7 编码人员及负责人18
 - 7.1 编码工作分工
 - 7.2 工作进度安排
 - 7.3 计划制定人姓名及制定日期

1 文档概述

1.1 编写目的

为了保证网络热词聚合搜索引擎系统开发项目组 G04 按时保质地完成项目目标，为了便于项目团队成员更好地了解项目开发的情况，同时，为了便于项目小组的成员能够更好地了解项目系统的编码组成和实现部分的具体情况，使得系统编码工作和实现工作的开展过程能够合理有序，所以需要文档来对整个编码过程进行描述，直观的将系统编码的实现计划展示出来。

本计划将系统实现生命周期内的工作任务范围，各项编码工作的任务分解、人员分工、开发进度，以及系统实现过程中的各项事务编成文档，作为项目组成员以及项目干系人之间的共识和约定，和系统实现过程中各项工作的基础和约定。

1.2 背景

1.2.1 项目名称

项目名称: 网络热词聚合搜索引擎

1.2.2 项目委托单位

浙江大学计算机学院软件工程实践本科教学组

1.2.3 项目用户

用户范围: 广大喜欢吃瓜的网友

1.2.4 任务提出者

软件工程项目实训教师——邢卫，邵建

1.2.5 项目承担者

浙江大学软件工程专业 16 级

G06 项目组长: 李冻秋

开发人员: 杨樾人、徐霄雯、李珏宇

1.2 背景

随着互联网的不断发展，网络流行语已经成了年轻人表达情感的主要语言方式之一。它被认同，反映某种现实，类似大碗宽面、这些网络用语，生动形象地体现了网民们当时的心情，也被用来表达情感。

基于以上的现状，网络流行语的聚合搜索引擎以小鸡词典、b站、微博等网络平台的新词热梗为数据来源，为渴望与时俱进使用新梗的广大网友提供搜索平台，在遇到新词时进行方便的搜索。

1.3 范围

本编码与实现计划文档是在基本已经全面了解了用户的需求之后设立，它将为教学辅助网站提供开发过程中的编码模块指导，它包括了本项目的项目概述、系统编码、系统实现计划等多方面的内容，为项目编码过程中各个模块的资源调配以及编码过程中的规范提供文档依据，在项目的整个生命周期中都有相当重要的地位。

1.4 定义

MySQL:系统服务器所使用的数据库关系系统(DBMS)

Mongodb: 非关系型数据库

Elasticsearch：一个分布式的搜索引擎

Java:静态面向对象编程语言，用以开发平台的后端。

Spring:一个轻量级的Java开发框架。XML:可扩展标记语言，标准通用标记语言的子集，是一种用于标记电子文件使其具有结构性的标记语言，用来做前端页面设计。

UML(Unified Modeling Language):统一建模语言、是一套用来设计软件蓝图的标准建模语言，是一种从软件分析、设计到编写程序规范的标准化建模语言

1.5 参考资料

[1]软件需求(第三版), Karl E. Wiegers,清华大学出版社, 2004.

2 项目方案陈述

2.1 工作内容

2.1.1 项目用户方

使用网页端的广大消费者。

2.1.2 项目开发方

姓名职位	
徐霄雯	开发人员
李珏宇	开发人员
杨樾人	开发人员
李冻秋	测试人员

2.1.3 方案目标：

我们的系统将实现以下的功能：

1. 查看网络流行语的含义
2. 搜索网络流行语
3. 查看图片和视频
4. 按标签查看网络流行语
5. 创建词条
6. 点赞点踩词条
7. 可视化分析
8. 智能搜索和推荐

2.2 产品

2.2.1 程序

网络热词聚合搜索引擎

2.2.2 服务

开发、测试及部署，后期技术维修支持。

2.3 验收

2.3.1 验收方式

验收由项目开发小组和授课教师(邢卫，邵建)共同验收。项目的验收采用一次性验收的方式，在达成项目阶段验收的目标后，按照计划提交阶段的工作成果，包括前期的软件需求规格说明书，系统概要设计等，以及提交代码。验收的标准如下表，在项目小组提交成果后，由验收方进行评估，如果验收方对于小组提交的成果存在意见，应及时反馈，开发小组对验收方的意见进行评估，并进行改进。经过调整后直至双方达成共识。

2.3.2 验收标准

优秀	合格	不合格
1. 完成全部需求	1. 完成全部需求	未完成全部需求
2. 用户体验良好	2. 用户体验感较差	
3. 界面美观，网站架构清晰易用	3. 界面美观度一般或者架构较不清晰	
4. 系统运行稳定流畅	4. 系统运行较不稳定	
5. 系统比较安全	5. 系统有严重安全问题	

2.4 计划批准者和批准日期

计划批准者	批准日期
邢卫，邵建	2019 年 7 月 5 日

2.5 项目完成期限

2019 学年暑假小学期

3 系统编码

在系统设计完成后，要进入编码阶段。编码是把系统设计转化为程序代码的过程。为了能够让团队在编码过程中能够更好地配合，需要给整个团队指定统一的编码标准。

计划批准者	批准日期
邢卫，邵建	2019 年 7 月 6 日

3.1 编码语言

本次开发后端使用的开发编程语言为 Java，前端的开发使用 html 和 javascript、css 等。

Java 是一门面向对象编程语言，Java 语言具有功能强大和简单易用两个特征。Java 语言作为静态面向对象编程语言的代表，极好地实现了面向对象理论，允许程序员以优雅的思维方式来进行复杂的编程。Java 具有简单性、面向对象、分布式、健壮性、安全性、平台独立与可移植性、多线程、动态性等特点。Java 可以编写桌面应用程序、Web 应用程序、分布式系统和嵌入式系统应用程序等。

超文本标记语言（英语：HyperText Markup Language，简称：HTML）是一种用于创建网页的标准标记语言。HTML 是一种基础技术，常与 CSS、JavaScript 一起被众多网站用于设计令人赏心悦目的网页、网页应用程序以及移动应用程序的用户界面[3]。网页浏览器可以读取 HTML 文件，并将其渲染成可视化网页。HTML 描述了一个网站的结构语义随着线索的呈现，使之成为一种标记语言而非编程语言。

ES 分布式搜索：ElasticSearch 是一个基于 Lucene 的搜索服务器。它提供了一个分布式多用户能力的全文搜索引擎，基于 RESTful web 接口。Elasticsearch 是用 Java 开发的，并作为 Apache 许可条款下的开放源码发布，是当前流行的企业级搜索引擎。设计用于云计算中，能够达到实时搜索，稳定，可靠，快速，安装使用方便。

MongoDB 是由 C++ 语言编写的，是一个基于分布式文件存储的开源数据库系统。在高负载的情况下，添加更多的节点，可以保证服务器性能。可以为 WEB 应用提供可扩展的高性能数据存储解决方案。

3.2 版本控制

本次项目小组采用 **Github** 作为版本控制系统。**Github** 是一个分布式的版本控制系统，作为开源代码库以及版本控制系统，**Github** 目前拥有 140 多万开发者用户。随着越来越多的应用程序转移到了云上，**Github** 已经成为了管理软件开发以及发现已有代码的首选方法。

3.3 开发工具

IntelliJ IDEA 是一种商业化销售的 **Java** 集成开发环境（**Integrated Development Environment, IDE**）工具软件，由 **JetBrains** 软件公司(前称为 **IntelliJ**)发展，提供 **Apache 2.0** 开放式授权的社区版本以及专有软件的商业版本，开发者可选择其所需来下载使用。

- ✓基于 **Gradle** 的构建支持。
- ✓**Android** 专属的重构和快速修复。
- ✓提示工具以捕获性能、可用性、版本兼容性问题。
- ✓支持 **ProGuard** 和应用签名。
- ✓功能强大的布局编辑器，可以让你拖拉 **UI** 控件并进行效果预览。

SublimeText 是一个代码编辑器(**SublimeText2** 是收费软件，但可以无限期试用)，也是先进的文本编辑器。**SublimeText** 具有漂亮的用户界面和强大的功能，例如代码缩略图，**Python** 的插件，代码段等。还可自定义键绑定，菜单和工具栏。**SublimeText** 的主要功能包括:拼写检查，书签，完整的 **PythonAPI**，**Goto** 功能，即时项目切换，多选择，多窗口等等。**SublimeText** 是一个跨平台的编辑器，同时支持 **Windows**、**Linux**、**MacOSX** 等操作系统。**SublimeText2.0** 进行了如下改善:支持 **Retina** 视网膜屏、快速跳到下一个、文本拖放、改善构建系统、**CSS** 自动完成和高亮设置等。

3.4 编码规范

3.4.1 前端开发

前端和代码统一使用 **IntelliJIDEA** 进行开发；

测试在本地测试；

项目完成后迁移到服务器上线。

3.3.2 后端开发

Java 实现后台的操作，包括类的定义以及对数据库的查询与修改。使用 Spring 框架，Java 编码参照《Google-Java 编码规范》(<https://google.github.io/styleguide/javaguide.html>)。

4 实现计划

4.1 开发过程

4.1.1 需求分析

需求分析是整个设计中重要的一环，当可行性分析完成，项目立项，确定开发角色后，从 12 月 13 开始至 12 月 14 日，有关的设计开发人员与相关业务人员共同对业务流程、管理方式进行分析，并进行资料的收集、整理。通过前期对有关数据信息的收集、归纳和分析整理后，确定了用户需求，对软件必须完成的功能进行了定义，在此基础上完成了数据定义，建立了数据字典。

4.1.2 系统设计

从 12 月 15 日开始至 12 月 17 日，完成对整个系统的分析设计，对概念模型、存储模式、完整性控制、存取权限等进行了定义，对系统功能各模块进行了详细设计、定义了数据库总体结构、编码命名规范。

4.1.3 编码及测试阶段

从 12 月 18 日开始至 12 月 22 日，完成程序设计和系统测试，完成了数据库建立及程序的编制调试。为了避免错误积累，采用边开发边测试的基本模式，对每个模块都安排开发人员进行单独测试，系统联调及系统测试，对系统处理逻辑、例外处理能力、容错能力等进行大规模的测试，对发现的问题进行彻底纠正。

4.1.4 文档、产品部署

从 12 月 23 日开始至 12 月 25 日，完成系统的测试和调试并撰写测试报告，完成用户培训工作，编写用户手册，系统投入运行阶段。

4.1.5 项目总结

项目结束后用剩余的两天时间，对项目研发、部署等开发过程中的问题、经验教训总结备案，以利于项目经验的积累，同时撰写课程报告与论文。

4.2 接口人员

负责本项目同用户的接口人员为李珏宇，由软件开发方派专人，按客户要求运行并给客户演示，在后期中负责维护和更新。

4.3 进度

- 方法：采用结构化开发
- 设置模块开发优先级：通过对“教学辅助网站”的特点和信息流程等的分析，确定各模块开发的先后次序。

5 支持条件

5.1 计算机系统支持

5.1.1 硬件

- 服务器：阿里云服务器，ECS 实例 5 台

- 客户端：任意联网的 PC 机，并安装谷歌等浏览器

5.1.2 软件

- 服务器操作系统为 ubuntu16.04 ；
- 使用 IDE 为 IntelliJIDEA, vim ；
- 数据库采用 MySQL ；
- 服务器采用 Nginx。

运行时需要的支持条件:

- 服务器的要求
 - 服务器建议选用 IntelCPU，选择 windowsserver2012。
 - 服务器内存必须使用服务器专用 ECC 内存
 - 开发平台选 apache 平台
 - 为了保证数据存储的绝对可靠，硬盘应使用磁盘冗余阵列(RAID01)
 - 为了防止服务器不可预测的故障，或者服务器的定期维护对系统运行的影响，所以建议使用两台服务器。两台服务器应构成双机热备份系统。中间使用 Watchdog 电路。这样的结构可以保证整个系统的长时间不间断工作，即使在服务器定期维护的时候也可以使用另一台后备服务器工作。
 - 服务器应该必须有固定 IP 地址。
 - 经济条件允许的情况下，应该尽量使用高速稳定的配件。
 - 服务器上应该配备的软件
 - 操作系统:ubuntu16.04
 - 数据库:MySql
 - 服务器必须使用专业的防火墙和反病毒软件。
 - 除了为了运行必须配备的程序以外，服务器上建议尽量不要安装其他无关程序，以减少程序的混乱或者程序的意外冲突。
 - 各系列的操作系统尽量统一。这样可以避免管理软件因为操作系统版本不一致造成的过多的开销。

各系列的机器必须也安装反病毒软件和防火墙。以防止网络上的蠕虫病毒在整个网络范围内的蔓延。

5.2 需要用户承担的工作

1. 提出“掌柜”App 的功能需求不完善的方面;
2. 参与系统的评审与验收;
3. 与项目开发小组保持联系, 并配合开发小组完成项目开发工作。

5.3 需由外单位提供的服务

本系统为独立开发, 不需要外单位提供条件

6 白盒测试和黑盒测试

白盒测试：

后端开发的同学在开发阶段通过分析控制构造的环路复杂性, 导出基本可执行路径集合, 对测试用例进行了测试。设计出的测试用例保证了在测试中程序的每个可执行语句至少执行一次。

在此次的开发过程中, 白盒测试阶段绘制了程序控制流图, 通过分析控制构造的环路复杂性, 导出了基本可执行路径集合。主要包括以下 4 个步骤和一个工具方法：

1. 程序的控制流图：描述程序控制流的一种图示方法。
2. 程序圈复杂度：McCabe 复杂性度量。从程序的环路复杂性可导出程序基本路径集合中的独立路径条数, 这是确定程序中每个可执行语句至少执行一次所必须的测试用例数目的上界。
3. 导出测试用例：根据圈复杂度和程序结构设计用例数据输入和预期结果。
4. 准备测试用例：确保基本路径集中的每一条路径的执行。

以下是对 api 接口的测试

6.1 查询词条

REST client interface showing a POST request to `http://127.0.0.1:8080/api/v3/entry`. The request body is `x-www-form-urlencoded`. The response is a JSON object with status `200 OK`, time `615 ms`, and size `4.05 KB`.

KEY	VALUE	DESCRIPTION
<input checked="" type="checkbox"/> id	367	
<input checked="" type="checkbox"/> video_limit	5	
<input checked="" type="checkbox"/> img_limit	5	
<input checked="" type="checkbox"/> weibo_limit	5	

The response body is displayed in JSON format:

```
1 {
2   "code": "0",
3   "msg": "请求成功",
4   "data": [
5     {
6       "id": 367,
7       "name": "童年",
8       "time": "2017-07-12",
9       "image": "https://api.jikipedia.com/upload/6fd8e804b77f8dbd39bafdd8fa82e116.jpg",
10      "view": 3203,
11      "like": 19,
12      "dislike": 0
13    }
14  ]
15 }
```

6.2 搜索

POST http://127.0.0.1:8080/api/v3/search

Send

Save

Params

Authorization

Headers (9)

Body

Pre-request Script

Tests

Cookies

Code

Comments (0)

none

form-data

x-www-form-urlencoded

raw

binary

GraphQL BETA

KEY	VALUE	DESCRIPTION	...	Bulk Edit
<input checked="" type="checkbox"/> key	-小鸡 快跑			
<input checked="" type="checkbox"/> tag				
<input checked="" type="checkbox"/> page_number	0			
<input checked="" type="checkbox"/> page_size	5			
<input checked="" type="checkbox"/> order_by	[]			
<input checked="" type="checkbox"/> time_begin				
<input checked="" type="checkbox"/> time_end				
Key	Value	Description		

Body

Cookies

Headers (3)

Test Results

Status: 200 OK

Time: 162 ms

Size: 5.67 KB

Save

Download

Pretty

Raw

Preview

JSON

```

3      "msg": "请求成功",
4      "data": [
5        {
6          "id": 1048,
7          "name": "细软跑",
8          "time": "2017-10-31",
9          "image": "",
10         "view": 13800,
11         "like": 18,

```

Bootcamp

Build

Browse

6.3 推荐

► http://127.0.0.1:8080/api/v3/recommend Examples (0) ▼

POST http://127.0.0.1:8080/api/v3/recommend Send Save

Params Authorization Headers (9) Body ● Pre-request Script Tests Cookies Code Comments (0)

Query Params

KEY	VALUE	DESCRIPTION	...	Bulk Edit
Key	Value	Description		

Body Cookies Headers (3) Test Results Status: 200 OK Time: 125 ms Size: 4.15 KB Save Download

Pretty Raw Preview JSON ▼

```
1 {
2   "code": "0",
3   "msg": "请求成功",
4   "data": [
5     {
6       "id": 6972,
7       "name": "小鸡",
8       "time": "2018-08-02",
9       "image": "",
10      "view": 104,
11      "like": 3,
12      "dislike": 0,
13      "tagList": [],
14      "content": "一种动物在幼年时代两个腿一个鼻子一个嘴。还是一种不要脸的东西。 ",
15      "videoList": [
16        {
17          "video_id": "53885190",
18          "title": "刚孵化的小鸡太萌太可爱了！鸡太美是有迹可循的",
19          "watch": "7.3万",
20          "subtitle": "354".
```

6.4 点赞

POST http://127.0.0.1:8080/api/v3/vote Send Save

Params Authorization Headers (9) **Body** Pre-request Script Tests Cookies Code Comments (0)

☐ none ☐ form-data ☒ x-www-form-urlencoded ☐ raw ☐ binary ☐ GraphQL BETA

	KEY	VALUE	DESCRIPTION	...	Bulk Edit
<input checked="" type="checkbox"/>	id	1568			
<input checked="" type="checkbox"/>	like	true			
	Key	Value	Description		

Body Cookies Headers (3) Test Results Status: 200 OK Time: 482 ms Size: 495 B Save Download

Pretty Raw Preview JSON ↕

```

1 {
2   "code": "0",
3   "msg": "请求成功",
4   "data": [
5     {
6       "id": 1568,
7       "name": "89.5分",
8       "time": "2017-11-29",
9       "image": "",
10      "view": 3400,
11      "like": 17,
12      "dislike": 1,
13      "tagList": [
14        "学校",
15        "学生",
16        "考试"
17      ],
18      "content": "一个把小学生逼得私自篡改答案的分数。其实诚信比分数更重要 ",

```

6.5 创建词条

▶ http://127.0.0.1:8080/api/v3/create Examples (0)

POST http://127.0.0.1:8080/api/v3/create Send Save

Params Authorization Headers (9) **Body** Pre-request Script Tests Cookies Code Comments (0)

☐ none ☐ form-data ☒ x-www-form-urlencoded ☐ raw ☐ binary ☐ GraphQL BETA

	KEY	VALUE	DESCRIPTION	...	Bulk Edit
<input checked="" type="checkbox"/>	name	小鸡			
<input checked="" type="checkbox"/>	img_url	imgurl			
<input checked="" type="checkbox"/>	tag_list	["tag1","tag2"]			
<input checked="" type="checkbox"/>	content	自做集代扣代缴			
	Key	Value	Description		

Body Cookies Headers (3) Test Results Status: 200 OK Time: 57 ms Size: 175 B Save Download

Pretty Raw Preview JSON ↕

```

1 {
2   "code": "0",
3   "msg": "创建成功",
4   "data": null
5 }

```


7 设计变更

数据库变更

在工作的初始阶段，我们是使用 MySQL 进行的数据库开发，但是在开发的第二天，发现用 MySQL 进行建表，关系过于复杂，需要建立太多的表格，所以我们使用了 MongoDB。

ES 变更：

我们一开始只使用了一个 ES 的节点，后来发现用 ES 的集群速度更快。

API 的变更：

我们的 api 设计经过了八次迭代，具体的跌倒版本可以查看 ftp 里面的附件。

8 工作进度

我们的项目进度尽可能与设计说明书里面的进度相一致，但是仍有略微的出入，故在此详细列出。

文档名称	项目阶段	文件格式	完成日期
《项目可行性报告》	项目准备	Word 文档	2019.7.6
《项目章程》	项目准备	Word 文档	2019.7.6
《项目总体计划》	项目准备	Word 文档	2019.7.6
《质量保证计划》	项目准备	Word 文档	2019.7.6

《前景与范围》	项目准备	Word 文档	2019.7.6
《需求工程计划》	项目准备	Word 文档	2019.7.6
小组例会纪要(中期)	项目准备	Word 文档	2019.7.7
《软件需求规格说明书》	项目准备	Word 文档	2019.7.7
《系统设计计划》	蓝图设计	Word 文档	2019.7.7
《需求变更控制会规程》	蓝图设计	Word 文档	2019.7.8
《系统编码与实现计划》	系统实现	Word 文档	2019.7.8
《测试计划》	系统实现	Word 文档	2019.7.13
《用户手册》	系统实现	Word 文档	2019.7.15
《软件概要设计说明书》	系统实现	Word 文档	2019.7.13

9 输出要求

编码阶段最终的输出是一个涵盖客户端，数据库文件，以及本系统所需各种图片等非文本文件及其它与本系统相关文件的文件夹。其中每一个客户端、后端代码文件分别只能实现一个完整的业务功能，并以该功能来命名文件。

10 编码人员及负责人

10.1 编码工作分工

➤前端:徐霄雯、李珏宇

➤负责人: 徐霄雯

➤后端:陈清源、杨樾人

➤测试:李冻秋

➤文档:全体成员

10.2 工作进度安排

进度	开始时间-截止时间
第一次会议(项目计划)	2019/7/6-2019/7/7
软件需求分析	2019/7/6-2019/7/7
系统设计	2019/7/7-2019/7/16
编码实现与部署	2019/7/16-2019/7/18
系统测试	2019/7/16-2019/7/17
项目总结	2019/7/17-2019/7/19

10.3 计划制定人姓名及制定日期

计划制定人: 李冻秋

制定日期:2019/7/7