项目题目

# 项目简介

使用爬虫抓取各类新闻，在网站上展示

# 二．项目市场分析

满足人们闲暇时间看新闻的需求

# 三．项目组成员

组长：施赛华

组员：施赛华 张亚明 周伦浩 王伟良

时间： 2015/10/10 ~ 2015/10/19

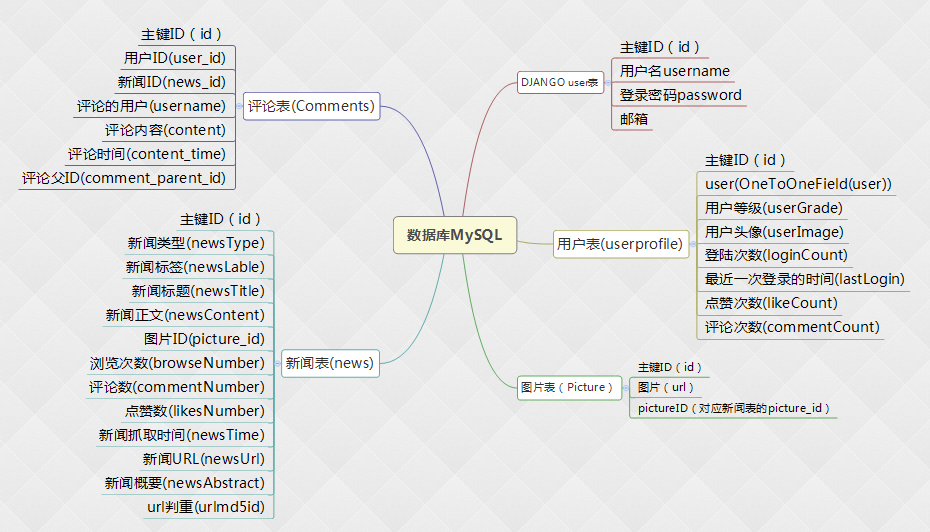
项目版本: v1.0

# 四．概要设计



# 五．数据库设计

# 六．模块详细设计



1, app目录为app应用模块

2，spider\_web目录为项目管理目录

3，static 存放静态资源文件

4，template 模板文件

5，见代码注释

# 七．总体功能设计

1. 数据库模型的设计

2. 爬虫方面抓取网络上的新闻，存入数据库

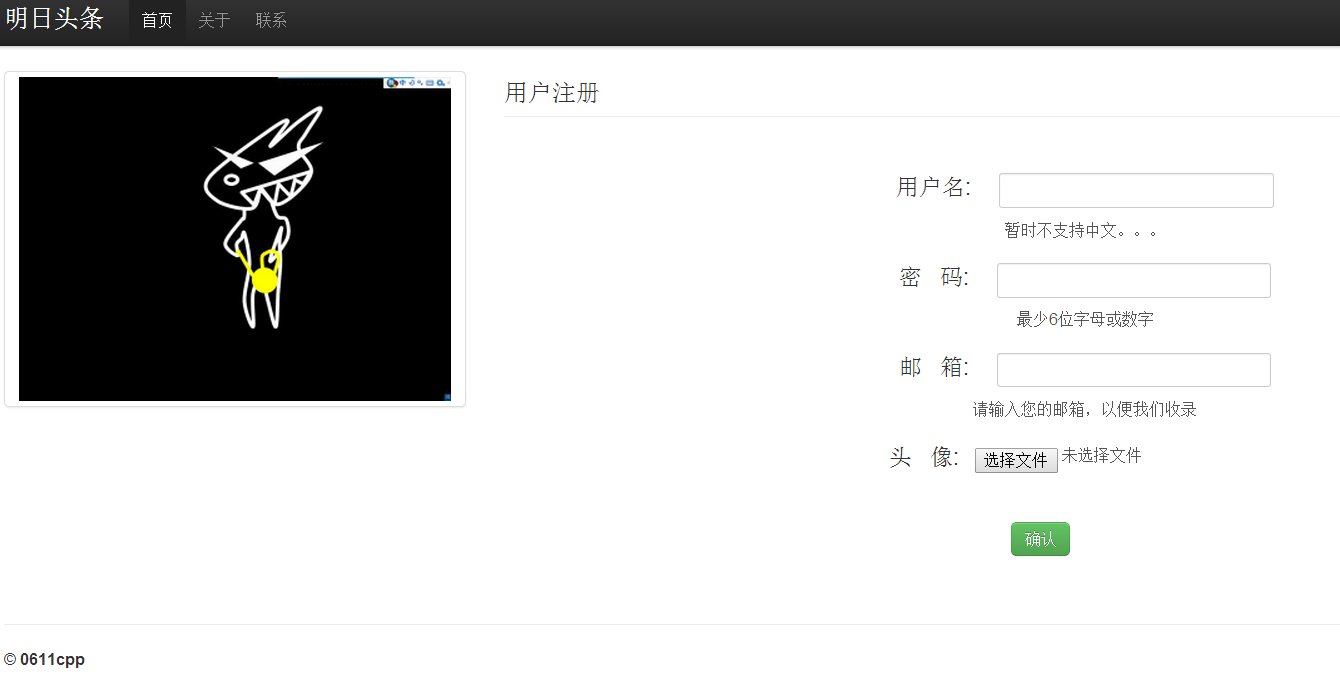
3. 实现新闻的显示，浏览，评论

4. 用户注册，登录，修改密码

# 八．已完成功能

1. 大部分功能都已实现

# 九．效果图



# 十．待完成功能

1. 爬虫抓取其他网站的新闻，存入数据库

2. 网页排版优化

3. 点赞功能

4. 实现新闻分类显示

# 十一.代码github地址

https://github.com/yoyomaster/spider\_web

# 十二.项目使用说明

1, 爬虫(使用scrapy 框架)(环境建议装在python虚拟环境)

环境搭建

安装openssl

pip install lxml

pip install Scrapy

pip install scrapyd

pip install scrapy-crawlera

sudo apt-get install python-lxml

2，django环境搭建

3，uwsgi web服务与nginx服务环境搭建

4， 数据库使用mysql

创建数据库 create database spider\_web character set utf8

Django: python manager.py migrate 生成数据库表

5，在爬虫项目目录下（spider\_163目录下 ）

使用命令 scrapy crawlall 命令 启动爬虫