

# 需求分析

项目名称： 针对资讯的用户建模和个性推荐系统设计与开发

项目类别：  
☐ 电子商务  
☐ 移动终端应用  
☒ 大数据分析  
☐ 物联网应用  
☐ 人机交互应用  
☐ 其他( )

命题企业： 浙江网新恒天软件有限公司

咨询邮箱： xyjian@hengtiansoft.com

2017 年 12 月 1 日

# 项目需求分析

## 一、项目背景

近几年，全球数据量出现爆炸式增长，数据成了当今社会增长最快的资源之一。如此增长迅速、庞大繁杂的数据资源，给传统的数据分析、处理技术带来了巨大的挑战。大数据时代的出现简单的讲是海量数据同完美计算能力结合的结果，大数据计算技术完美地解决了海量数据的收集、存储、计算、分析的问题。大数据时代开启人类社会利用数据价值的另一个时代。

## 二、项目概述

大数据时代，数据量够大，涵盖范围够广，新数据源在不断产生并持续产生中。每个人都有很多信息接收渠道，却无法很好地从这些信息中找到真正自己感兴趣的内容。为此，有很多基于个性推荐的应用脱颖而出，这里面最出名的，就是今日头条。通过对用户的兴趣建模，这些应用能有效地发现用户关心的内容，并推送给用户。

## 三、项目需求

### （一）总体要求

设计并开发一个资讯网站平台，通过对用户兴趣的追踪和建模，为用户推荐符合其阅读兴趣的内容。命题单位会提供相关资讯数据。要求：

- （1）要识别优质的和低质的资讯内容，优先推送优质内容。

- (2) 只有用户仔细阅读的资讯才能作为兴趣来源；
- (3) 不能持续推送过于相似的资讯，以免影响阅读体验；
- (4) 能通过用户兴趣进行隐性的用户分组，同组用户的内容可以组内传播；
- (5) 能识别用户强烈不感兴趣的内容，并主动避推；
- (6) 通过训练的模型对资讯进行实时分类；
- (7) 资讯有美观的展示，阅读体验佳。

## **（二）功能需求**

- (1) 用户功能：支持用户注册，登录，修改用户信息及密码，找回密码
- (2) 资讯浏览功能：浏览资讯列表，可以动态加载，列表项可以点赞或点踩；浏览资讯详情，可以点赞或点踩；显示资讯的分类信息；可以查看本人兴趣模型。

## **（三）技术要求**

项目前端为 HTML5 展示，通过调用 RESTful 接口与后台通讯。后台开发语言不限，平台及操作系统不限，数据存储形式不限。

训练语料由参赛团队自行搜集。

开发期间与评选期间命题单位提供的数据内容可能不一致。开发用数据为 JSON 格式的静态资讯数据；评选期间，利用命题单位提供相同 JSON 格式的实时数据接口进行资讯兴趣匹配、推送与展示，该接口接收当前 offset（数字，初始是 0，按已接收资讯量递增）作为

参数。数据接口地址可作为网站配置参数。

#### （四）运行环境需求

##### 1. 软件环境

服务器操作系统及版本：Linux (CentOS 6.3 以上)；

Web 应用服务器：Tomcat 8 版本及以上；

前端页面符合 HTML5 标准，兼容主流现代浏览器，如 Chrome 等；

数据库服务器：MySQL 5.6 及以上；

##### 2. 硬件环境

服务器部署（建议）

序号	服务器	配置	数量	类型
1	数据分析服务器	8 核或以上、16G 内存、500G 硬盘	1 台	1、云平台/PC 服务器。 2、充分考虑大量数据处理和分析的情况。
2	资讯网站服务器	4 核或以上、8G 内存、500G 硬盘	1 台	1、云平台/PC 服务器，建议应用支持水平扩展。 2、数据库可另选服务器进行搭建。

##### 3. 网络环境

资讯网站平台能够接受外网访问。