|  |  |  |
| --- | --- | --- |
| 文件状态：  [ ] 草稿  [√] 正式发布  []正在修改 | 文件标识： | SYZH1104-TYPT-02 |
| 当前版本： | V1.0 |
| 作者： | 邓虹玉 |
| 完成日期： | 2018/1/22 |

基于语义网的站内搜索引擎模块改造项目需求分析说明书

**北京慧博科技有限公司**

**2018年1月**

修订历史版本

|  |  |  |
| --- | --- | --- |
| **日期** | **修订内容** | **修订人** |
| 2018-01-22 | 创建需求规格说明书。 | 邓虹玉 |
| 2018-02-1 | 根据各部门交流的情况，修正部分内容。 | 邓虹玉 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

目录

第一章 引言 1

1.1 文档目的 1

1.2 预期读者 1

1.3 术语解释 1

1.4 参考资料 1

第二章 需求概述 2

2.1 项目背景 2

2.2 项目范围 2

2.3 项目目标 2

2.4 计划以后完成的需求 3

2.5 不提供的需求 3

2.6 运行环境说明 3

2.6.1 服务器运行环境 3

2.6.1.1 Web服务器 3

2.6.1.2 数据库服务器 3

2.6.2 集客CRM程序运行环境 4

2.6.2.1 处理程序 4

2.6.3 客户端运行环境 4

第三章 整体说明 5

3.1 业务模型 5

3.1.1 集客CRM 5

3.1.2 积分系统 6

3.2 用例模型 7

3.2.1 集客CRM 7

3.2.2 积分系统 7

第四章 软件功能需求 8

4.1 集客CRM 8

4.1.1 订单关怀 8

4.1.1.1 下单关怀 8

4.1.1.2 催付提醒 9

4.1.1.3 物流提醒 9

4.1.1.4 付款关怀 10

4.1.1.5 回款提醒 10

4.1.1.6 退款关怀 10

4.1.1.7 评价关怀 10

4.1.2 客户管理 11

4.1.2.1 会员列表 11

4.1.2.2 高级分组 11

4.1.2.3 会员互动 12

4.1.2.4 黑名单管理 12

4.1.3 精准营销 13

4.1.3.1 短信营销 13

4.1.3.2 邮件营销 14

4.1.3.3 优惠券发送 15

4.1.3.4 支付宝红包发送 16

4.1.3.5 流量营销 16

4.1.3.6 促销活动 17

4.1.3.7 自动化营销 17

4.1.4 数据洞察 18

4.1.4.1 会员分析 18

4.1.4.2 订单分析 19

4.1.4.3 商品分析 19

4.1.4.4 物流分析 20

4.2 积分系统 20

4.2.1 会员等级积分设置 20

4.2.2 赚积分 21

4.2.3 花积分 22

4.2.4 会员积分管理 23

4.2.5 积分分析 23

第五章 软件非功能需求 25

5.1 性能 25

5.2 用户界面需求 25

5.2.1 主页面需求 25

1. 引言
   1. 文档目的

本业务需求规格说明文档的编写目的，是为用户及软件开发方双方充分理解业务需求而编写的。它阐述了系统的使用范围及背景，明确了所要达到的目标以及达到目标所需的条件，从而使双方对软件的业务相关规定有一个共同的理解，使之成为整个开发工作的基础，为该系统的需求定义、分析、设计、开发与测试的提供指导。

* 1. 预期读者
* 研发组、测试组、设计组等相关项目干系人；
* 基于语义网站内搜索引擎项目相关成员。
  1. 术语解释

|  |  |  |
| --- | --- | --- |
| **序号** | **名词** | **相关解释** |
| 1 | 知识图谱 | 知识图谱（Knowledge Graph/Vault）又称为科学知识图谱，在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。 |
| 2 | 语义网 | 语义网是对未来网络的一个设想，现在与Web 3.0这一概念结合在一起，作为3.0网络时代的特征之一。简单地说，语义网是一种智能网络，它不但能够理解词语和概念，而且还能够理解它们之间的逻辑关系，可以使交流变得更有效率和价值。 |
| 3 | solr | Solr是一个独立的企业级搜索应用服务器，它对外提供类似于Web-service的API接口。用户可以通过http请求，向搜索引擎服务器提交一定格式的XML文件，生成索引；也可以通过Http Get操作提出查找请求，并得到XML格式的返回结果。 |

* 1. 参考资料
* 基于语义网的站内搜索项目系统建设方案；

1. 需求概述
   1. 项目背景

站内搜索引擎功能模块是电商类网站用户通往商品页面的重要渠道，模块功能是否满足用户需求，直接决定电商商品的销量，是电商平台可持续发展的关键因素。

站内搜索引擎功能模块旨在帮助用户对店铺、商品等信息进行精确检索和信息挖掘，提供全面、准确、快速的信息检索服务，为用户第一时间找到目标店铺和商品。同时，用户使用站内搜索引擎模块功能，产生大量的搜索日志和浏览历史等数据，数据真实的反馈用户的浏览习惯、行为以及偏好，是网站进行网站架构优化、提升用户体验、大数据应用建设等其他应用的重要底层数据。

* 1. 项目范围

本项目建设的内容涉及到站内搜索，具体包含：

* 智能提示；
* 智能搜索；
  1. 项目目标
* 在电商知识图谱关键技术支撑研究方面，形成总体研究方案，并基于语义搜索、智能推荐、图谱构建关键技术等研究，形成专利及高水平研究论文。
* 形成搜索推荐的智能提示原型系统，根据用户输入完成汉语补全、拼音补全、语义补全、纠错、首字母不全。
* 形成搜索推荐的语义搜索原型系统，智能搜索根据用户输入，通过商品知识图谱，完成意图识别。根据运营知识图谱的规则，进行二次加权排序。根据生成商品图谱商品属性，生成过滤筛选项目。对于存在歧义的检索，生成用户搜索建议。
* 形成搜索推荐的智能推荐原型系统，智能推荐根据用户兴趣模型，混合基于商品内容、分类体系、用户行为协调过滤的推荐模型进行推荐。
* 完成商品知识图谱的构建系统原型，支持自动化及手工的数据更新及相关可视化。
  1. 计划以后完成的需求

无。

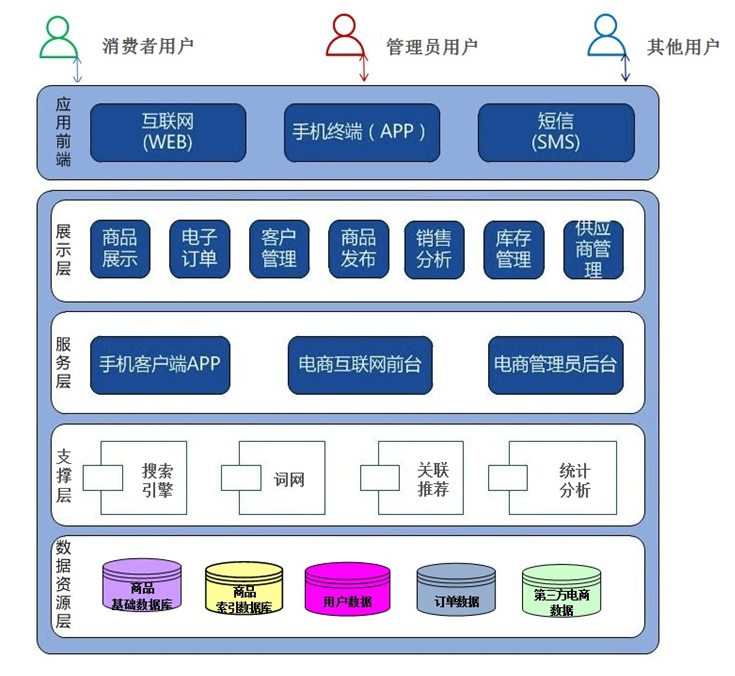
* 1. 不提供的需求

无。

* 1. 运行环境说明
     1. 服务器运行环境

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ip** | **机器配置** | **操作系统** | **网络带宽** | **安装应用** |
| 10.1.33.80 | 知识图谱服务器 | redhat 6.8 | 百兆 | toknumx\jdk\python\数据同步脚本 |
| 10.1.33.82 | 检索索引服务器 | redhat 6.8 | 百兆 | elasticsearch\jdk |
| 10.1.33.84 | WEB服务器 | redhat 6.8 | 百兆 | tomcat\jdk\mysql\search.war\sse.war |
| 10.1.33.86 | 日志服务器 | redhat 6.8 | 百兆 | nginx |

1. 整体说明
   1. 业务架构

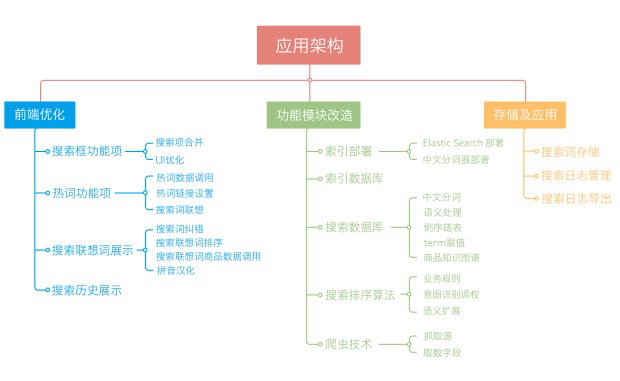


业务架构示意图

基于知识图谱技术的搜索引擎功能模块需要以多来源多格式商品数据汇聚分析为支持，实现商品数据多维度多属性的商品知识图谱构建及更新，配合电商平台各产品线实现商品的搜索展现。整个总体架构分为：数据资源层、支撑层、服务层、展示层和应用前端组成。

* 1. 应用架构

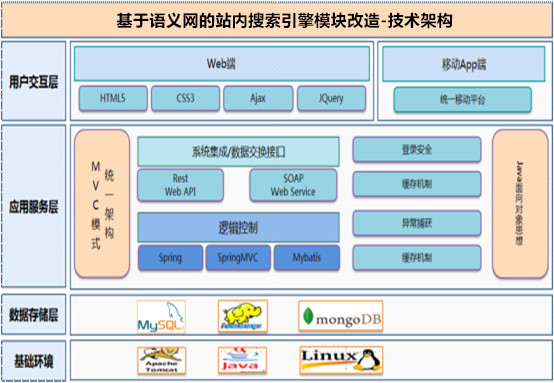
基于知识图谱技术采购专区站内搜索引擎改造三块功能，前端展示功能、搜索模块功能和数据存取功能，详细如下图：



应用架构示意图

* 1. 技术架构

基于语义网的站内搜索引擎模块改造技术架构确定了目标系统的核心技术路线，明确了应用系统内部采用分层模型，明确了各个层次在应用系统中的作用和承载关系，以及各个层次与规范体系的对应。以此为基础，明确了本系统在国网公司的IT规划中与国网公司其他业务应用系统、基础架构的关系。本系统的技术架构设计要满足总部一级部署，健壮的灾备机制，易维护性，性能的可扩展性，渠道的可扩展性，高可用性，健全的安全防护体系，合理保护投资等。



搜索引擎模块技术架构图

基础环境主要采用Linux操作系统，中间件环境采用Java运行环境，采用Tomcat应用服务器。

数据存储层采用MySQL关系型数据库；支持非关系型数据库，如MongoDB等。

应用服务层基于MVC的多层架构，实现模型、数据、业务及展现的分离，以及基于面向服务的总体架构；针对海量资源的浏览和下载，提供数据缓存机制。

用户交互层采用当前主流的HTML5、CSS3、Ajax及JQuery等技术，实现良好的用户体验和用户交互效果；提供移动APP服务，实现多平台访问。

1. 软件功能需求
   1. 电商搜索业务流程及效果优化

业务总体情况：用户进入电商平台后，通过站内搜索引擎模块功能查找目标商品，搜索引擎模块返回搜索数据的过程。全过程包括：用户在搜索框键入搜索词，搜索引擎模块读取搜索词，在后台对搜索词进行计算匹配，把匹配结果传递给用户完成搜索过程。

本次站内搜索引擎模块功能改善的业务需求主要为支撑电商平台用户使用的功能改善需求，包括：

（1）发起搜索键入搜索词；

（2）系统响应增加搜索联想、纠错等处理；

（3）确认关键词后提交搜索数据；

（4）接收数据，进行数据处理，返回搜索结果；

（5）搜索结果筛选，确认目标商品。

业务流程表

|  |  |  |
| --- | --- | --- |
| **流程节点** | **业务步骤** | **业务信息** |
| 键入搜索词 | 选择搜索方式：选择搜索范围； 点击搜索框：点击搜索框 键入关键词：输入关键词 修改：修改键入词 | 键入的搜索词 |
| 返回联想词 | 数据读取：读取输入的数据； 数据计算处理：对数据进行匹配处理； 联想词展示：展示联想词、商品数量、拼音汉化等信息； | 搜索联想 |
| 选择确认搜索词 | 选择：选择符合意向的联想词； 点击：点击搜索按钮，提交搜索数据； | 确认的搜索词 |
| 数据分析处理返回 | 分词：对搜索词进行分词处理； 处理：对搜索词进行语义处理； 检索：在索引库检索词元； 算法排序：对匹配的商品进行排序处理； 数据返回：返回搜索结果； | 搜索结果商品清单 |
| 搜索结果选择 | 筛选：筛选商品，点击目标商品。 | 目标商品页 |

* 1. 基础功能完善及底层数据库搭建

基于语义网的智能索引库建设，涉及了机器学习的范畴，系统在上线后可以应用爬虫技术不断抓取数据，solr的各个组件模块对抓取的基础数据进行智能处理，通过机器学习，对数据进行语义分析、分词、算法匹配等处理，逐步完善、扩充基础底层数据，在电商平台进行人工智能技术的初步应用。

本项目最终会完成了三个底层数据库的建设，一是索引库，二是搜索库，三是搜索词库，三个数据库都是网站应用的底层数据，可以为用户画像、商品推荐、大数据应用等其他产品供底层数据支持。

* 1. 人工智能应用

按照国网公司电子文件管理要求，应用语义网技术研发电商资源专业服务平台，将把电商品数据及功能纳入统一管理和服务。

基于语义网的智能索引库建设，涉及了机器学习的范畴，系统在上线后可以应用爬虫技术不断抓取数据，solr的各个组件模块对抓取的基础数据进行智能处理，通过机器学习，对数据进行语义分析、分词、算法匹配等处理，逐步完善、扩充基础底层数据，在电商平台进行人工智能技术的初步应用。