|  |  |  |
| --- | --- | --- |
| 文件状态：  [ ] 草稿  [√] 正式发布  []正在修改 | 文件标识： | SYZH1104-TYPT-14 |
| 当前版本： | V1.0 |
| 作者： | 梁志敏 |
| 完成日期： | 2018/6/18 |

基于语义网的站内搜索引擎模块改造项目

系统用户意见

**北京慧博科技有限公司**

**2018年6月**

修订历史版本

|  |  |  |
| --- | --- | --- |
| **日期** | **修订内容** | **修订人** |
| 2018-06-17 | 创建文档 | 梁志敏 |
| 2018-06-18 | 审核文档 | 张亚情 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# 项目实施介绍

本项目建设过程如下：

|  |  |  |
| --- | --- | --- |
| **工作项** | **工作内容** | **时间** |
| 可研编制 | 根据汇报内容，编制可行性研究报告 | 7.15-7.23 |
| 软硬件环境准备 | 准备在知识图谱平台软硬件环境及域名 | 7.24-8.2 |
| 差异分析 | 分析现有功能与知识图谱平台的差异 | 8.3-8.8 |
| 图谱平台构建 | 数据清洗及学习 知识图谱构建及校验 智能提示、搜索组件定制及优化 各组件集成及测试 | 8.9-8.27 |
| 测试环境部署 | 修改后的产品部署至测试环境，并联调 | 8.28-8.31 |
| 安全扫描 | 将套装软件送至汇通金财安全实验室进行安全漏洞扫描 | 9.1-9.18 |
| 正式环境部署 | 修改后的产品部署至正式环境，并联调 | 9.22-9.27 |
| CA部署 | 对接第三方CA，具备平台使用条件 | 9.22-9.28 |
| 与商城融合方案 | 开展套装软件与商城融合技术研究，主要从用户、展现形式、数据融合和CA融合4方面开展（本次计划主要完成用户和数据融合的技术方案制定和评审） | 9.10-9.31 |
| 安全测试 | 将套装软件送至中国电科院开展安全测试 | 10.8-10.30 |
| 培训 | 测试环境部署后进行培训及演练，生产环境部署后进行演练 | 9.22-9.31 |

本项目实现功能清单如下：

| **序号** | **一级功能** | **状态** | **二级功能** | **状态** | **三级功能** | **状态** | **功能点说明** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 搜索框功能项 | 完善 | 搜索项合并 | 完善 | 一站式检索 | 完善 | 整合商品、物料、编号三个选项为“商品”，增加搜索提示。 |
| 关联推荐 | 新增 | 关联推荐是通过词网智能运算产生的。词网是基于国网电商基础商品数据库，通过商品/语义关联分析，计算搜索词与搜索词之间、商品与商品之间、搜索词与商品之间的关联关系，从而为用户推荐相关的检索词、检索问题和检索答案 |
| 分类导航 | 新增 | 多种分类查询 |
| 页面完善 | 完善 | WEB总体展示 | 完善 | 主界面展示 |
| 布局 | 完善 | 布局优化 |
| 搜索词 | 完善 | 搜索联想词展示 | 完善 | 增加拼音汉化功能，相同前缀词联想功能 |
| 搜索联想词商品摘要展示 | 新增 | 根据搜索联想词关联出的商品，提供商品摘要展示功能 |
| 联想词数据调用 | 新增 | 方便用户快速查找所需商品 |
| 商品摘要数据调用 | 新增 | 搜索词及搜索联想词无相关的商品摘要提示 |
| 搜索历史 | 新增 | 在用户点击搜索框时展示搜索历史数据 |
| 拼音关联 | 新增 | 针对以拼音方式输入商品搜索词，实现相关汉子化的相关搜索词推荐 |
| 2 | 热词功能项 | 完善 | 热词管理 | 完善 | 热词数据调用 | 完善 | 提取用户搜索词存储库72小时内搜索次数最多且在索引库有匹配商品的关键词，更换频率为24小时。 |
| 热词链接设置 | 完善 | 通过热词关联相关商品 |
| 词频统计 | 新增 | 对用户键入搜索词进行词频统计 |
| 3 | solr部署 | 完善 | solr部署到招标专区 | 完善 | 搜索合并 | 完善 | 合并相关搜索项 |
| 中文分词器接入 | 新增 | 分词模型 | 新增 | 构建基准分词器 |
| 关键词提取 | 新增 | 通过方法将用户输入的一段文字或多个搜索组合词中用于表达商品主要内容的词汇提取出来，使用标签云揭示关键词的主要程度 |
| 分词标注 | 新增 | 根据商品内容进行分词并且获得词语的词性，然后根据不同的词性进行标注展示 |
| 内容分类 | 新增 | 针对输入单个商品搜索词以及多个组合词进行归类，然后将所属商品类别进行标记。 |
| 语义联想 | 新增 | 根据用户输入的商品搜索内容进行关键词提取，进行语义联想。推荐出相关联的多个搜索词共用户选择 |
| 4 | 索引库建设 | 新增 | 数据导入 | 新增 | 商品主数据导入 | 新增 | 提供索引基础数据 |
| 索引建立 | 新增 | 索引字段处理 | 新增 | 索引商品页面html原始数据，提取字段 |
| 字段权重赋值处理 | 新增 | 建设词元为基点的索引库，给词元赋值 |
| 5 | 搜索联想词 | 新增 | 汉字映射 | 新增 | 拼音汉化 | 新增 | 增加拼音汉化功能，相同前缀词联想功能 |
| 联想词管理 | 新增 | 联想词数据计算统计 | 新增 | 对用户的站内搜索数据进行统计 |
| 联想词数据存储 | 新增 | 联想词存储 |
| 联想词数据导出 | 新增 | 联想词导出 |
| 相同前缀词语算法设计 | 新增 | 筛选排序算法设计 | 新增 | 对筛选条件排序重新进行算法设计 |
| 6 | 排序算法设计 | 新增 | 排序算法设计 | 新增 | 动态相关度排序 | 新增 | 根据检索词的词频、位置、相对关系等因素计算出该结果与检索词间的相关度，并利用该相关度进行结果排序。 |
| 检索相关度排序 | 新增 | 增加用户偏好的权重，形成个性化的动态相关度 |
| 7 | 搜索词存储 | 新增 | 用户搜索词日志管理 | 新增 | 用户搜索词日志存储 | 新增 | 用户的站内搜索数据进行存储 |
| 用户搜索词日志导出 | 新增 | 用户的站内搜索数据进行导出 |
| 8 | 搜索词应用 | 新增 | 搜索词应用 | 新增 | 搜索词应用 | 新增 | 搜索词智能应用 |

系统当前运行情况良好。

# 2.系统主要功能

描述系统涉及平台内容、各平台主要功能点及功能点内容描述。

## 2.1 基于语义网的站内搜索平台

基于语义网的站内搜索平台在内/外网实施或部署智能提示模块、智能搜索模块。具体功能如下：

|  |  |
| --- | --- |
| **功能点名称** | **功能点内容描述** |
| 模块1 智能提示 | 模块1包含搜索词匹配补全、搜索词语义补全、搜索词语义消歧、搜索词全拼提示、搜索词首字母提示、搜搜词混合拼音提示、统计提示等7个功能，主要提供搜索智能提示服务。 |
| 模块2 智能搜索 | 模块2包含搜索意图识别、搜索语义拓展、商品过滤筛选、搜索拼音及错词纠正等4个功能，主要实现提供智能搜索服务。 |

# 3.系统使用情况及结果

## 3.1 系统运行的稳定性

自系统上线试运行以来，系统运行情况良好、指标监控情况属正常范围，未出现影响用户正常使用的重大事故。

## 3.2系统的性能情况

自系统上线运行以来，

1、当系统进行多用户并发操作时，应满足如下要求：主搜索框进行搜索时，搜索提示响应不超过2秒，提交搜索后，搜索结果页的访问平均响应时间不超过5秒；系统登录平均响应时间不超过5秒；在打开带有主搜索框的网页时，搜索热词加载速度不低于2秒；筛选条件在执行二次搜索时，平均响应时间不超过5秒。

2、日常平均CPU占用率小于40%，忙时小于75%，内存占用率小于50%，最大并发时小于75%。

3、系统稳定试运行三个月以上后，运行安全、稳定，达到7×24h的可靠运行能力，年可用率＞99.97%。

# 4.需求和设计的符合性

基于知识图谱技术的站内搜索引擎的优化改造，一是对前端体验优化，改善前端搜索功能、提供搜索增项数据、筛选条件排序优化；二是搜索引擎模块功能的改造，在solr的基础上对搜索引擎进行二次改造，研究利用爬虫技术，抓取采购专区商品页面目标数据，建设商品知识图谱，优化中文分词和搜索结果排序算法；三是开发数据存储功能，存储底层的用户数据，以供数据分析和其他项目应用拓展。设计符合国网相关要求，与需求相符。

## 5.用户反馈及总结

用户在系统试运行期间，对系统功能以及性能方面、文档完整性等方面都持有较好的反馈意见，基于语义网的站内搜索平台满足用户基本需求的基础上，操作更加便捷、人性化，搜索更加智能，因此同意验收。

确认签字：

日 期： 年 月 日