**咪咕文化**

**产品需求文档(PRD)**

**（V3.7.0）**

C:\Users\ADMINI~1\AppData\Local\Temp\WeChat Files\590637531216426107.png

**咪咕文化科技有限公司**

**2020年12月**

**修订记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本号 | 修订日期 | 修改人 | 修改内容 | 修改原因 |
| USS\_music\_v3.7.0 | 2020-12-30 |  | 1、视频彩铃搜索优化  2、热度计算增加校验  3、支持多维度自动置顶  4、歌手tab支持语义搜索&增加头像过滤  5、非热门内容不支持不完整纠错  6、特定场景下歌曲名模糊搜索支持按编辑距离匹配  7、使用正则优化影视类搜索词的搜索效果  8、歌手下歌曲、视频数量统计规则调整  9、若干现网体验问题优化 | 新增需求 |

目录

[目录 3](#_Toc15272)

[1. 前言 4](#_Toc2469)

[1.1. 文档说明 4](#_Toc20045)

[1.2. 术语及缩略语 4](#_Toc5735)

[2. 产品背景 4](#_Toc22040)

[2.1. 产品概念及目标 4](#_Toc122)

[2.2. 产品形式 5](#_Toc28831)

[2.3. 业务服务对象及业务范围 5](#_Toc3906)

[3. 产品业务需求 6](#_Toc28504)

[4. 产品功能需求 7](#_Toc16267)

[4.1. 视频彩铃搜索优化 7](#_Toc18278)

[4.2. 热度计算增加校验 7](#_Toc23834)

[4.3. 支持多维度自动置顶 8](#_Toc10899)

[4.4. 歌手tab支持语义搜索&增加头像过滤 8](#_Toc3483)

[4.5. 非热门内容不支持不完整纠错 9](#_Toc23703)

[4.6. 特定场景下歌曲名模糊搜索支持按编辑距离匹配 10](#_Toc14359)

[4.7. 使用正则优化影视类搜索词的搜索效果 10](#_Toc31279)

[4.8. 歌手下歌曲、视频数量统计规则调整 11](#_Toc27727)

[4.9. 若干现网体验问题优化 11](#_Toc2599)

[5. 其他要求 18](#_Toc15276)

[5.1. 兼容性需求 18](#_Toc3812)

[5.2. 可测试性需求 18](#_Toc790)

# 前言

## 文档说明

本文档主要针对咪咕统一搜索项目咪咕音乐子公司的需求提出全面要求，详细描述了需求的业务交互流程, 用于指导后续的系统改造开发和联调测试，是后续统一认证相关技术方案和产品实现的依据之一。

本文档由tsg归口。

本文档主要起草人：段伟伟

## 术语及缩略语

| 词语 | 解释 |
| --- | --- |
| 产品 | 歌曲的MV、彩铃、振铃、全曲、随声听、无损、等属性。 |

# 产品背景

## 产品概念及目标

针对咪咕音乐的搜索功能主要指通过提供多种搜索入口和方式，让用户能够快速且便捷地进行搜索，同时通过一系列的算法和规则等，向用户输出其想要的或推荐的内容。

* **提供搜索入口**
  + 主要功能目标1：拓宽并优化用户搜索的渠道和方法。
* 引导搜索输入
  + 主要功能目标1：帮助用户快速输入。
  + 主要功能目标2：进行搜索词匹配。
  + 主要功能目标3：帮助用户快速定位搜索词。
* 优化搜索结果
  + 主要功能目标1：准确且快速露出歌曲、专辑、视频、歌单、歌词、活动等用户搜索目标内容。
  + 主要功能目标2：优化搜索结果内容排序。
  + 主要功能目标3：优化搜索结果界面呈现方式。
  + 主要功能目标3：推荐相关内容，增加优质资源露出机会。

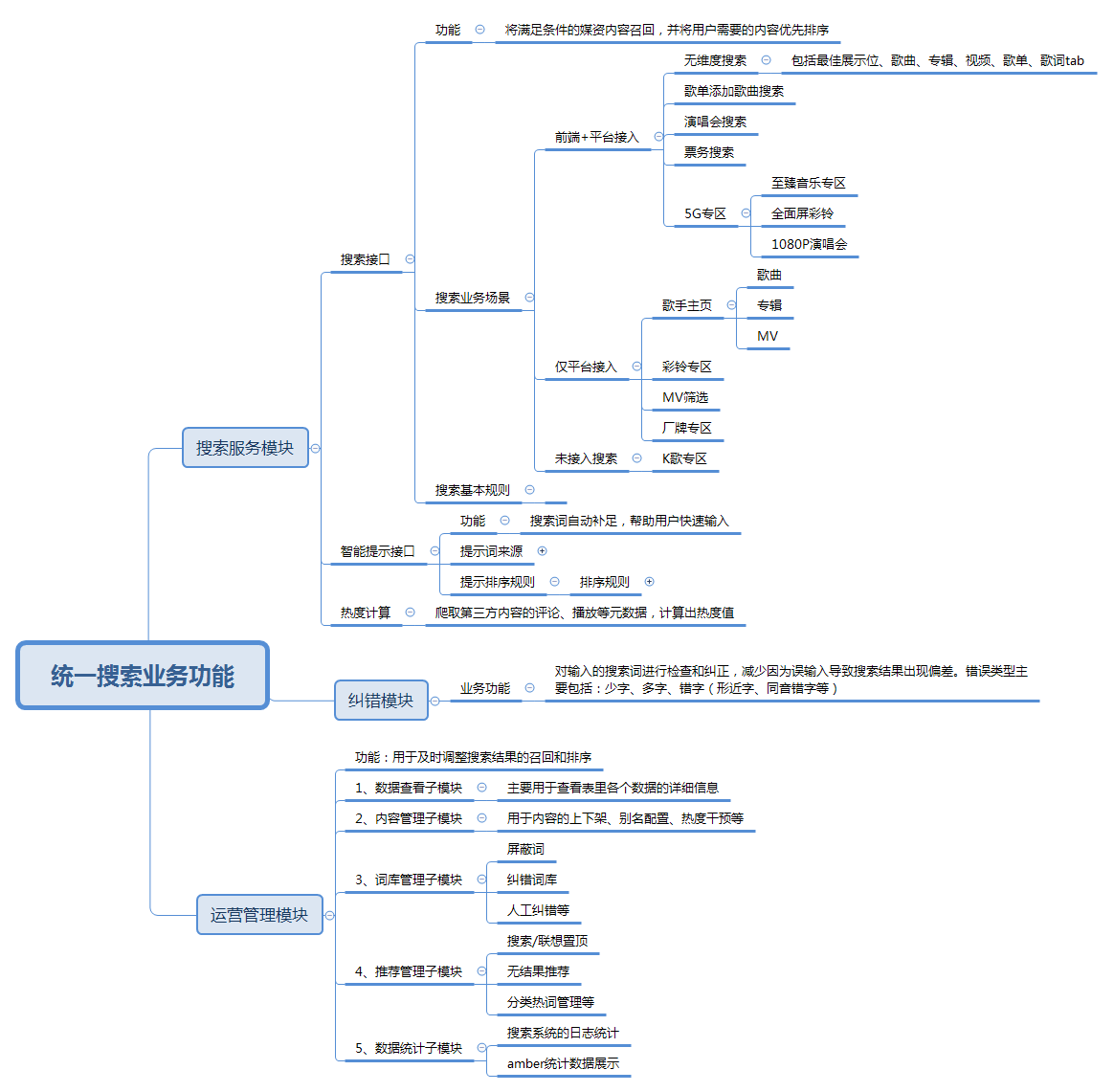
## 产品形式

咪咕统一搜索主要包括统一搜索系统和运营管理系统两个部分。前者为各子公司提供搜索能力API接口，根据不同的搜索需求，返回特定的搜索结果。后者通过WEB管理系统为运营人员和管理员提供运营管理、搜索规则、分析统计等功能。

## 业务服务对象及业务范围

|  |  |
| --- | --- |
| 用户类型 | 服务内容 |
| 互联网用户 | 咪咕音乐APP/WEB产品的互联网使用者。 |
| 搜索运营人员 | 咪咕音乐的搜索相关业务的运营人员。 |
| 管理员 | 统一搜索平台的管理员。 |

# 产品业务需求

**

用户在APP前端通过不同的搜索入口（搜索框、搜索历史、搜索发现、热门搜索、联想提示、多重筛选搜索、分类标签搜索等），利用文字或语音等搜索方式，在APP前端输入想要获取的信息，发起搜索请求，搜索后端接收到搜索请求后，进行意图识别，在数据库中进行相关数据的检索，然后利用各种算法等规则对相关内容进行排序，同时人工可通过在运营平台上进行查看及操作来对结果进行干预，最后向用户输出最终搜索结果。

# 产品功能需求



## 视频彩铃搜索优化

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750337774 | 1、数据处理  从数据集市同步视频彩铃的播放量、点赞数、评论数、彩铃订阅量。数据集市每日将数据同步搜索的ftp上；  2、搜索排序  根据上述数据计算自有热度 = 点赞数\*100 + 采用歌曲热度计算公式计算热度（评论+播放量） --- 公式暂定如此，并将自有热度加入视频彩铃搜索排序中。  IMG_256  IMG_256 |
| 需求规格：  无 | |

## 热度计算增加校验

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750337779 | 【需求背景】  当前爬取第三方数据的时候存在较大的失败率，因此会因为前一次爬取失败导致本次计算的增量很大，影响热度计算的准确性。  【需求描述】  每次爬取的时候增加一个标志位用于表示本次爬取是否成功，建议当标志位为成功的时候才计算增量，计算热度。 |
| 需求规格：  无 | |

## 支持多维度自动置顶

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750338799 | 为了进一步实现运营自动化，增加支持多维度自动置顶功能。支持在运营管理平台配置置顶搜索词（以及搜索词的别名）、配置置顶维度（复用当前人工置顶的维度，包括歌手、歌曲等）、配置内容入库的条件、配置置顶的排序规则。如：  1、设置“独立音乐人”为【歌手】置顶，置顶内容选择条件是歌手的singerNormal为1，且歌手头像不为空，歌手的歌曲数或者专辑数不为0，排序规则是 粉丝数 > hotlevel > hot ，最大个数限制设置为100；  2、设置“全面屏彩铃”【视频彩铃】置顶，置顶内容选择条件是彩铃纵横比是16:9，排序规则是 hotlevel > publishtime；  3、设置咪咕汇的置顶为【视频彩铃】，置顶内容选择条件是彩铃标题里包含“2020咪咕汇”，排序规则是publishtime > hotlevel。  注：  一个搜索词可能会对应多个维度置顶，运营管理平台开发的时候可复用咪咕视频上自动置顶的页面样式。 |
| 需求规格：  无 | |

## 歌手tab支持语义搜索&增加头像过滤

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750338990 | 1、歌手tab支持歌手语义搜索（仅用于歌手tab）。语义维度包括 性别（男、女、组合）、标签、别名，如：  -- 别名+歌手名，如：奶茶刘若英；  -- 性别+标签，如：香港流行男歌手；  -- 标签组合，如：香港知名唱作人（地域标签：香港、职业：唱作人）；  注：  单个标签和性别的，可以通过自动置顶填充来实现。  2、支持歌手+歌手组合。如林俊杰、周杰伦，陈奕迅薛之谦；  3、歌手tab搜索时过滤无头像的歌手；  4、排序 ：  -- 优先人工置顶/自动填充置顶；  -- 优先按照匹配度排序。各组内按照hotlevel\*10万+hot > 粉丝数 > 歌手id倒排； |
| 需求规格：  无 | |

## 非热门内容不支持不完整纠错

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750338991 | 当前：只要满足条件的内容，都支持不完整纠错（即：纠错结果只是名称的一部分）；  要求：改为只有当歌手满足热度要求（hot > 0或者hotlevel > 0 ）或者歌曲/专辑/视频/歌词/影视名等满足热度要求（hot > 10万或者后台level > 0）才支持不完整纠错，否则入纠错词典时只入原完整名称。 （现有设定一方面是为了减少不必要的索引，提高纠错速度，一方面是为了避免这部分数据影响热门内容的纠错效果。）  badcase：  搜索“起风啦”，预期是可以纠错为"起风了"，目前返回“起风啦”，就是因为有个非热门歌曲“又起风啦”。 |
| 需求规格：  无 | |

## 特定场景下歌曲名模糊搜索支持按编辑距离匹配

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750338992 | 【需求背景】  在上个版本“非精确有版权热门歌曲排序前置”的需求里，当模糊搜索时，经常因为分词的原因导致部分结果无法召回，更无法优先排序，因此增加如下需求，以优化模糊搜索时的效果，增加热门内容模糊召回的概率。  【需求描述】  当纠错结果返回的纠错词rel=0，且类型为歌曲/专辑/视频/歌词/影视名时热度小于10万，模糊搜索支持按照编辑距离匹配，具体要求如下：  1、汉字大于等于2个字，数字字母大于等于5个字时才支持按照编辑距离召回；  2、仅召回编辑距离为1的结果，最多取两个，且要求召回内容热度大于20万（hotlevel\*10万+hot）；  3、仅原词搜索，不包括简拼、全拼；  4、注意按照编辑距离召回的和普通分词召回的，需要去重处理。 |
| 需求规格：  无 | |

## 使用正则优化影视类搜索词的搜索效果

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750339050 | 【需求背景】  原有正则表达式是放在纠错模块，后来因为影响其他搜索词的效果，该逻辑被取消，但是取消后会影响影视类搜索词的搜索效果。  【需求描述】  影视类搜索词的正则表达处理逻辑放在搜索接口。  正则表达式包括：  XXX + 主题曲/片尾曲/背景音乐/片头曲/原声；  若搜索词满足上述正则表达式，则搜索接口支持用过滤了“主题曲/片尾曲/背景音乐/片头曲/原声”之后的词去匹配影视名（支持去符号、去大小写、繁简转换），优先展示原词精确匹配的，再展示清洗之后匹配影视名的结果（清洗之后匹配影视名的放在前中缀匹配分组里）。  【测试重点】：  1、关注影视歌曲的搜索效果；  2、用例需要多找一些非影视歌曲名但是满足正则表达的词进行测试，避免此需求影响其他搜索词效果。 |
| 需求规格：  无 | |

## 歌手下歌曲、视频数量统计规则调整

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750338681 | 1、独立音乐人歌曲数量计算  原有：对于独立音乐人，会将作词作曲的歌曲统计到歌手的歌曲数量下，但是在SAM中不展示作词作曲的歌曲，会导致数量和展示结果对不上；  改为：歌手下歌曲数量计算时，不考虑作词作曲数量。  2、视频数量计算  原有：统计了所有有版权视频数量；  改为：统计所有有版权“在线状态”视频数量。 |
| 需求规格：  无 | |

## 若干现网体验问题优化

|  |  |
| --- | --- |
| 需求标示 | 需求描述 |
| 00750339414 | 1. 歌曲搜索：纠错前缀补充的时候，如果有召回的结果需要设置纠错提示。   Badcase：搜“nengmianchangjiang”，纠错前缀补充结果时未设置纠错提示。     1. 组合场景纠错优化   原有：组合场景时，增加了前缀匹配，默认调用对与专辑同名的歌曲提前的处理方法，该方法将导致组合场景为歌曲+歌手时，把其他歌曲提前。  改为：查询时，boost里对songName\_qp或songNeatName\_qp:"全匹配"的增大得分；同时限制当组合场景有专辑纠错结果时再调用对与专辑同名的歌曲提前的处理方法。  Badcase：搜“m滨崎步”时，其他歌曲排序在歌曲m前面     1. 数据处理对歌名、专辑名、歌手名、mv名中的不可截断空格替换为普通空格。   Badcase：搜“ice paper”，概率出现最佳展示位没有歌手结果  4、人工运营的容错、意图数据不作为高亮词返回  Badcase：泰密容错为泰蜜，结果显示高亮  66ee4b7217b3f55b9f479113970c983   1. 去除计算得数的转换   Badcase：搜“二十四+七”，最佳展示位是313专辑   1. 别名原名拼接问题修复  * 别名包含在原名中的不需要拼接原名。 * 对原名和别名进行统一的格式处理，即去特殊符括号之类的字符之后比较。 * 对原名和别名去括号、中括号、大括号再比较是否一致。   注：歌曲搜索的场景同样处理  Badcase：搜索fx，最佳展示位歌手f(x)拼接了别名(fx)     1. 修复词曲作者召回结果和其他非热门组结果拼接按热度倒排的排序不佳问题。   Badcase：搜“周杰伦”，周杰伦为词曲作者的歌曲排序在周杰伦演唱的歌曲上面。     1. 修复继续搜索和简拼纠错热门内容的代码逻辑冲突问题   Badcase：dx    9、解决因置顶内容仍参与抽屉排序，评论数超过500的单独提前造成的hotlevel和点击意愿高的歌曲排序靠后的问题。  Badecase：夜空中最亮的星，hotlevel和点击意愿更高，但排序在第二位    10、无维度接口当是默认搜索请求(搜歌曲时)才查询返回跳转标识，其他tab下不再返回跳转的数据和flag。  11、歌曲搜索下的语义分支中的标签搜索增加标签同义词的判断  12、修复带符号的歌曲搜索结果排序不佳或搜不到的问题  Badcase：搜索“同根生 《泰雅梦游》：〈相聚歌〉、〈战舞〉”，精确匹配结果排序靠后  bff079a16a351e663a41e9c16f4d8fc |
| 需求规格：  无 | |

# 其他要求

## 兼容性需求

## 可测试性需求