**修订记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本号 | 修订日期 | 修改内容 | 备注 | 修订人 |
| USS\_video \_v1.0 | 2021-8-10 | 1：初稿 |  | 孙玲 |

# 超编编辑部需求

超级编辑部系统搜索功能接入统一搜索。支持5家子公司内容的搜索及筛选。8月版本支持视频数据；9月版本支持音乐数据。

## 统一搜索增加音乐热词文件生成

【需求方案】

运营管理台在视频的定时任务中增加逻辑： 解析

conversionRate\_music\_keyword\*\*\*\*\*\*\*\*.csv文件，**去掉第一行，取第二、第三列**（排序不变），生成新的csv文件，文件名为music\_hot\_words\_\*\*\*\*\*\*\*\*.csv，编码格式还是gbk，如music\_hot\_words\_20210901.csv 并上传到ftp目录下。

## 统一搜索提供全内容的筛选能力

【需求背景】

统一搜索支持音乐内容筛选；

【具体方案】

音乐接口根据**系统标签**及**人工标签**字段进行筛选，筛选命中歌曲按照为score > 搜索飙升热度> hot > ID；

音乐筛选接口返回的音乐数据要求为**ID标识**、**音乐类型**（MV（包含liveshow）、歌曲、专辑）、**音乐标识**。ID标识需要拼接，具体样例：

单曲：11位版权id+{1位单曲类型}+内容id 如：6005663GKHF2600919000004719727 6005663GKMD2600919000004720015

专辑：素材id: 1138947547 1138947533

mv: 内容id : 600906000000374103 600906000000371844

**数据处理：**

1、数据处理对音乐标签数据进行补齐，超编音乐数据**增加通用标签字段**，存放**一二三级标签**内容、人工标签字段、存放人工运营标签；参考超编标签体系，将音乐侧当前音乐数据上标签按照超编的标签级别逐级补齐：

1. **一级标签仅包含风格、情感、语言、场景、主题、节日六类**
2. 如果音乐数据中标签为超编侧通用的五级标签，也需设置到标签字段中；
3. 人工运营标签存入超编音乐的人工运营标签字段内。

2、音乐数据导入歌曲、专辑、mv（包括liveshow）数据到**超编音乐表**，取最优展示的那条数据入表；同时新增ID字段满足如下要求：

单曲：11位版权id+{1位单曲类型}+内容id 如：6005663GKHF2600919000004719727 6005663GKMD2600919000004720015

专辑：素材id: 1138947547 1138947533

mv: 内容id : 600906000000374103 600906000000371844

3、音乐新增**搜索飙升热度**（累计近7天搜索点击次数），转换为**0~1区间内值**；1000次点击数据入库，同时更新时，上一个周期的数据失效；

## 统一搜索提供全内容的搜索能力

【需求背景】

【具体方案】

**音乐搜索场景：**

全匹配搜索： 名称（歌曲、专辑、MV名）搜索、按歌手搜索、影视剧名称字段；

组合搜：歌曲+歌手、专辑+歌手、mv+歌手搜索

纠错搜索：去掉专辑名搜索逻辑；

模糊搜索：mv名，歌曲名、专辑名前中后缀、分词搜（当无前中后缀匹配结果时，走分词搜）；

**组间：全匹配》组合》纠错》模糊搜**；

**组内：score > newhot > hot > ID;**

**人物搜索场景：**

人物下歌曲、专辑、mv优先展示，按照**score > 搜索飙升热度 > hot > ID**

再展示**标题命中人物**的数据；

**标签搜索场景：**

当搜索词为**音乐标签**、或者**音乐标签 + 子内容类型**、**标签+人名**搜索时，进入标签搜索逻辑：

标签搜索**命中标签字段**；子内容类型、人名字段、歌曲专辑mv名称字段也命中的加分；排序更靠前；标签搜索排序也为score > 搜索飙升热度> hot > ID；

**语义搜索场景：**

专辑、mv搜索支持语义搜索；语义支持当前音乐搜索类型外，同时语义支持**解析标签搜索词、标签+人名搜索**；语义搜索排序同上；

**视频、音乐间排序：**

1. 当有明确倾向为视频搜索或者音乐搜索时，同级别搜索结果中，优先排识别的结果，比如**精确匹配组**（包括系列剧）、**模糊匹配组**（前中后缀匹配）、**分词匹配**（包括关键字搜索）组；如果搜索词为标签、标签+一级分类，则标签搜到的**媒资类型**的数据优先；
2. 对于**精确组**：如果搜索词为人物搜索，判断当前人物为视频人物还是音乐人物，如果为音乐人物，通过歌手字段精确匹配的歌曲、专辑、mv优先排；当人物为视频人物时，该人物长视频优先排序；如果该人物全匹配的长视频也有、音乐也有，优先排该人物isnew为1的音乐数据，再排最近1年内主演的长视频数据，再按热度排人物的其他数据；
3. 当无明确倾向时，精确匹配组视频优先；其他组按照热度排序
4. 音乐的热度优先用newhot比，newhot为0的，取热度字段进行比较；

**音乐搜索返回筛选条件**：单曲、专辑、MV。

**返回总数：**返回总数为视频总数+ 音乐总数

语义能力： 1）视频、音乐标签能识别出具体倾向；

2）视频、音乐一级分类能识别具体倾向；

3）歌曲、视频实体能识别出倾向；

4）语义支持大数据量语义

## 统一搜索提供联想提示能力

**数据处理：**

取导入到超编库中的歌曲名、专辑名、歌曲+歌手、专辑+歌手、歌手+歌曲、歌手+专辑数据入联想库，同时有字段标识为音乐联想词；入库时，热度数据取歌曲、专辑热度数据；同时歌曲、专辑名均为清洗过的数据；上下线和内容一致；

**联想词接口：**

检索联想词表中音乐数据，**全匹配 + 前缀匹配**；

排序：按如下字段降序排： score > hot> id

当搜索词<2个时，hot>20W;当搜索词>=2个时，无此限制

联想词做去重处理；

视频、音乐联想词排序： 视频5个；音乐5个；当视频或者音乐不足5个，另外一个类型又多余5个时，另外一种类型数据补齐；

联想词返回搜索总数；

## 统一搜索提供听歌识曲查询接口

超编增加听歌识曲查询接口，具体逻辑为输入搜索词后，调用搜索接口获得匹配的歌曲ID列表，再调用媒资接口，查询内容详情返回。