爱唱整体优化方案

# 统一搜索优化措施

## 音乐栏目搜索入口页-热词

【需求背景】

当前咪咕爱唱默认界面比较空，并且当前的热搜曲目无法做到动态更新，可以接入统一搜索音乐搜索热词能力，实现搜索内容的完善及动态更新。

【需求设计】

爱唱接入当前热词接口，爱唱需要过滤跳转链接类热词，同时建议热搜词不要超过15个。搜索侧需要支持对接联调。

## 音乐栏目搜索入口页-搜索发现

接入现在的搜索发现接口。需要支持对接联调。

## 音乐栏目搜索结果页

【优化目标与背景】

当前TV侧搜索结果召回逻辑和排序完全复用APP端，导致简拼和全拼的命中效果不理想。优化TV侧简拼、全拼搜索结果，提升搜索结果命中率。

【优化设计】

歌曲：

当前爱唱歌曲搜索存在如下问题：

1. 简拼前缀无法召回部分歌曲，例如输入“JHY”，无法召回“交换余生”，



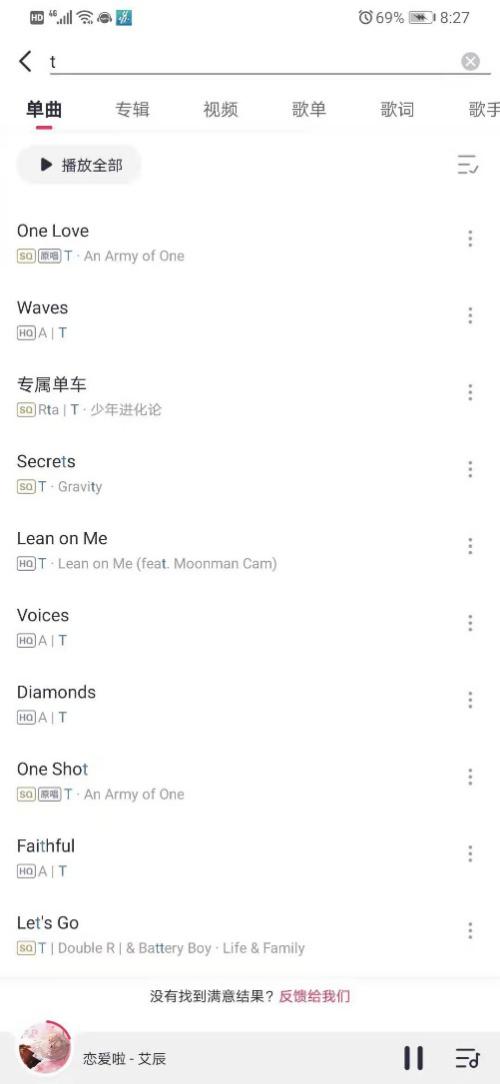
1. 简拼全匹配无法召回部分内容。比如：输入“ＮＧ”　无法召回“逆光”。竞品可以召回。



1. 英文歌曲不支持首拼搜索，仅支持全拼搜索。例如输入“ＹＢＷＭ”，无法召回“You belong with me”



1. 简拼全匹配召回的内容由于有专辑召回的，ＴＶ端不展示专辑信息，在ＴＶ端看不出逻辑。



基于此，当搜索请求为爱唱搜索请求并且搜索关键字为非汉字时，简化当前的搜索逻辑：

**不需要纠错、分词、词曲作者、中后缀、语义搜索**。提高简拼、全拼的搜索结果。支持

歌曲名、歌曲名+歌手名简拼全拼全匹配搜索及前缀搜索；支持歌手名简拼全匹配和全拼 。包括英文歌曲、歌手的首字母匹配。文档中统一归为简拼一类中。原词匹配归为全拼一类中。

歌曲名、歌曲名+歌手名命中时，是清洗后的歌曲名进行命中，比如“孤梦（网剧《山河令》插曲）”，歌手张哲瀚；则输入“GMZ”就可以从Solr召回。

歌曲名检索时，不支持歌曲别名检索

根据歌手名命中时，不需要匹配歌手别名。同时输入字符少于2个字符时，不做歌手名简拼、全拼命中、不做歌曲名+歌手名简拼、全拼命中。

特殊的一点：如GEM邓紫棋，输入DZQ可以命中；输入“GEMDZQ”也可以命中。

目前歌曲分冷库、热库；对于冷库内容：只做歌曲、歌手全拼、简拼全匹配；

排序： 人工置顶>hotlevel>newhot>hot>programId

数据处理：

由于当前的hot字段都是爬取等方式获得，不是很准，并且同一个名称的歌曲热度相同，在TV侧无法使用，系统中连续累计的点击字段对于歌曲上线时间越长的越有利，需要新的更真实的热度字段，才能达到越热门的内容越优先召回。当前搜索系统中歌曲的每日点击数是最真实的热度数据，系统中新增前日点击量这个字段，**数据处理根据点击意愿表中数据计算每个内容id的前一周的总点击量，保留点击量大于100的作为newhot的值**。

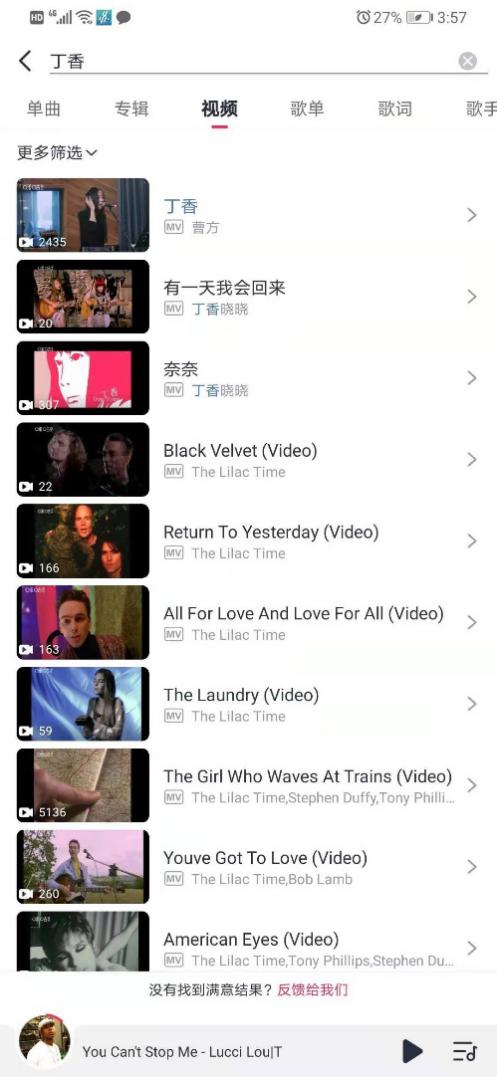
MV（SearchMVController）：

当前MV检索时存在如下问题：

1. 部分内容简拼前缀匹配无法召回，例如输入“chj”无法召回“彩虹金刚”，需要输入“chjg”才可以。

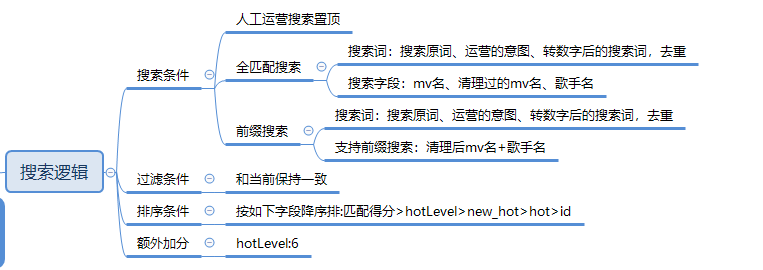
 

1. 部分内容通过简拼全匹配无法召回。比如输入“DX”，无法召回“丁香”

1. 基于此，当搜索请求为爱唱渠道的搜索请求，如果输入搜索词均为非汉字时，

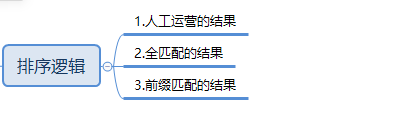
全匹配搜索、前缀搜索都支持全拼、简拼匹配。



输入字符少于2个字符时，全匹配搜索时不匹配歌手名字段

对于歌手名称需要做预处理：比如G.E.M.邓紫棋，需要支持对这个名称的搜索，也需要支持去掉英文字符的全汉字搜索。 有MV“透明”，那么比如搜索“toumingGE”、“TMDZQ” 都能召回透明这个mv。

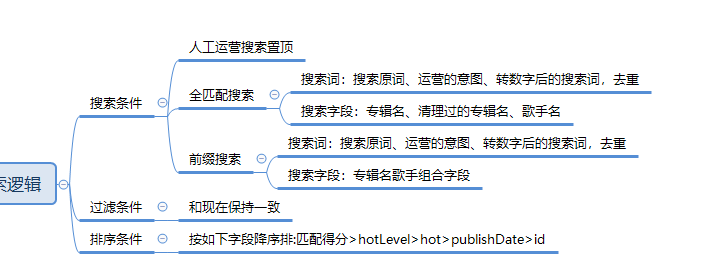
排序：



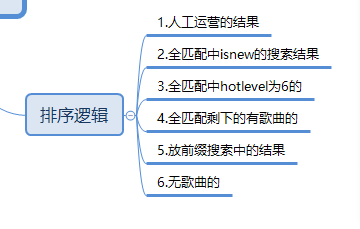
专辑（SearchAlbumController）

当搜索请求为爱唱搜索渠道请求并且仅拼音字母搜索时，搜索逻辑做简化。

1. 去掉当前根据专辑下歌曲、模糊搜、纠错搜，搜索出这个专辑的逻辑。去掉匹配影视名字段和歌手别名、专辑别名字段。保留如下搜索逻辑：



1. 在全匹配搜索中支持简拼全匹配。不搜索专辑别名、歌手别名。
2. 当检索关键字小于2个时，全匹配搜索时不搜索歌手名匹配；不做专辑名+歌手名前缀匹配。
3. 排序：



歌手（SearchSingerController）

当前歌手检索时存在如下问题：

1. 简拼前缀匹配无法快速召回热门歌手。比如输入“CY”无法召回“陈奕迅”“蔡依林”等热门歌手



竞品效果：



1. 简拼全匹配无法召回热门歌手。比如输入“TS” 无法优先召回“Taylor Swift”热门歌手。



基于以上问题，搜索请求为爱唱搜索渠道请求时，如果搜索词为拼音+数字搜索，则支持置顶、精确匹配（全拼精确匹配、简拼精确匹配）、前缀匹配（全拼前缀匹配、简拼前缀匹配）；原来的纠错搜、语义搜去掉。同时不命中运营别名字段。搜索结果排序：

--优先人工置顶/标签自动填充置顶；

--精确匹配和前缀匹配按照hotlevel\*10万+hot > 粉丝数 > 歌手id倒排；

数据处理：

MV、专辑、歌手的热度需要全量重新跑一遍。