vivo综搜SBS评测方案V2.0

**修订表**

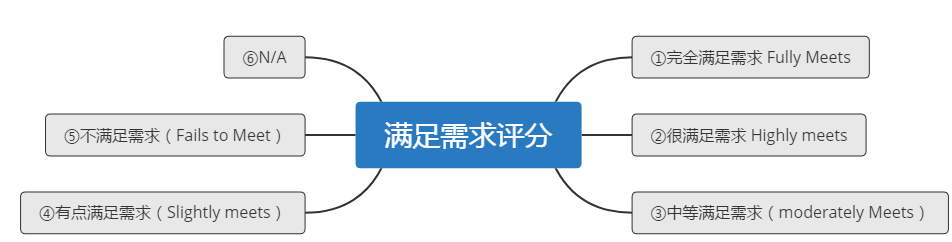
|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **状态** | **修改描述** | **作者** |
| 2019-4-2 | NEW | 新起草 | 马积珍 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 评测方法

1. 采用两两对比（side by side）评测方式对vivo综搜引擎与竞品（baidu/Sogou）进行比较评测。
2. 评分前需要充分理解query搜索意图，以用户体验为出发点，按照搜索结果对移动用户的需求满足程度评分衡量。

# 满足需求打分总则

评测者以移动用户需求为标准，判断搜索结果对用户有多大帮助，在多大程度上满了用户需求。评分分为以下6个等级：

详细定义如下：

|  |  |
| --- | --- |
| **评分等级** | **描述** |
| 1. 完全满足需求Fully Meets | 用户需求被完全、完美满足，不用再看其它结果了。通常，查询词有明确答案的才会有完全满足需求的评分，比如搜索“亚马逊”，亚马逊官网就是FullyM。 |
| 1. 很满足需求   Highly Meets | 对大部分用户来说非常有帮助。有小部分用户可能还想看看其它结果。通常是质量高、有权威度、最新的内容。一个查询可以有很多“很满足需求”的结果。比如用户搜索“海底捞”，标出附近海底捞地点的地图结果是个HM，但不是FullyM，因为可能有小部分用户想看的是海底捞官网，所以海底捞官网是另一个HM。 |
| 1. 中等满足需求   Moderately-Meets | 对很多用户有帮助，或者对一些用户很有帮助。有些或很多用户可能还想看看其它结果。能回答查询，但没那么全面、及时、权威，同时不是低质量、过时、不准确的内容通常属于MM。 |
| 1. 有点满足需求Slightly Meets | 对部分用户有帮助。很多或大部分用户还得看其它结果。可能是质量较低、过时、不大准确、太宽泛或太狭隘的内容。标题有误导性或太夸张的也属于SM。 |
| 1. 不满足需求   Fails to Meet | 完全不满足移动用户需求。几乎所有用户都得看其它结果。经常是与查询无关、事实错误、很低质量、很过时的内容，或者在手机上完全没办法用的功能。在页面质量评分中被评为“最低质量”的页面也应该被归为“不满足需求”。 |
| 1. N/A | 0结果 |

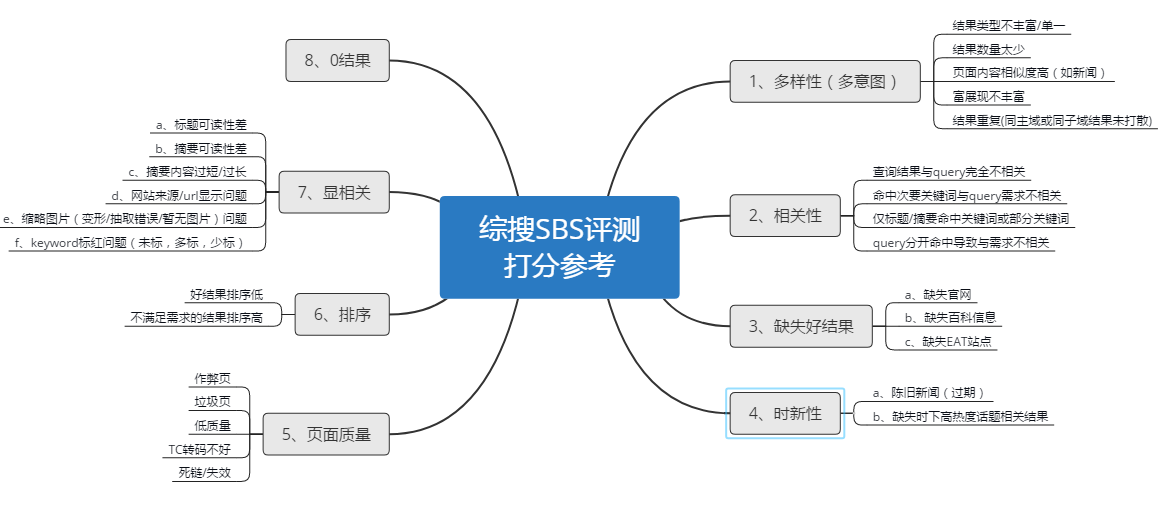
# 打分细则

## 打分依据

总依据：即既要页面质量高，又要内容相关，才被认为满足了需求。评测者需要仔细思考query与用户意图。

## 打分参考维度

打分时，原则上不要求评测者逐条点击页面查看，在拿捏不准，疑似网页质量不够满意时才打开页面查看。打分判断时参考以下8个维度考虑：



**说明：**

①多样性，一指多意图query希望结果类型更多更丰富；二指展现形态更多样化；

②相关性，主要指结果与query的相关（满足）程度；

③时新性，针对时效性较大的query，如新闻类query；

④显相关，如果不点击搜索结果页面，如普通的搜索结果，那么（标题、摘要、url、图片、时间来源）就是判断依据；

⑤页面质量：在与query相关的前提下，能从页面中获得的信息满足用户需求的程度。

⑥ E-A-T：指的专业、权威、可信任度高的页面或站点结果。

## 不同query类别打分参考

|  |  |
| --- | --- |
| **Query类别** | **打分参考** |
| 导航寻址类 | 精准的导航寻址，  NO.1是目标结果——【FullyM，5分】  No.2～3有目标结果——【Highly Meets 4分】  其他酌情递减得分。 |
| App/功能类 | NO.1是目标结果且能下载——【FullyM，5分】  其他根据结果能否下载以及排位酌情评分 |
| 电视剧/电影/动画片 | NO.1直达结果，可直接消费播放剧集，查看详情等——【FullyM，5分】  需要点击页面进入找到播放源播放——【Highly Meets 4分】  其他以排序位置，以满足用户需求程度酌情降级评分。 |
| 音频类 | NO.1直达结果，可直接消费点击播放——【FullyM，5分】  其他根据排序和满足程度酌情递减得分。 |
| 错写类 | 根据纠错情况，评分 |
| …… | 略 |

## YMYL页面判断要求

YMYL即：(Your Money or Your Life )

|  |  |
| --- | --- |
| **页面概述** | **要求** |
| 购物，以及可以线上转账、付账等交易的页面 | 真实，有效，不存在虚假诱骗行为 |
| * 提供有关投资、税务、退休计划、买房、大学学费、买保险等方面建议或信息的页面 | 真实，有效，不存在虚假诱骗行为 |
| 提供健康、药物、疾病、精神健康、营养等方面建议或信息的页面 | 真实，内容安全，不误导用户的行为 |
| * 提供重要公众信息的官方或新闻页面，如涉及到当地/国家政府政策、流程、法律、灾难应急服务等信息的页面，涉及重要国际事件、商业、政治、科技等的新闻页面。这部分需要评估员运用自己的知识和判断力，不是所有新闻页面都属于YMYL页面。 | 真实，时效性高 |

在YMYL页面的判别上，要求谨慎，不然可能影响用户的健康、安全、财务等。

# Side by side打分要求

要求评测者在对Side By Side评测的两个引擎打分尺度上，保持统一的水平，以使得评测结果的公平和公正性。

# 评测指标及统计公式

## 输出指标

满意率、伤害率、综合平均分

## 指标计算公式

1. 满意率=\*100%
2. 伤害率=\*100%
3. 综合平均分

用于直观统计两引擎结果对用户的需求满足程度，得分越高说明越能满足用户需求。按照等级赋予分值：

完全满足需求（5分）、很满足需求（4分）、中度满足需求（3分）、有点满足需求（2分）、不满足需求（1分）、N/A（0分）。

通过用户评测可获得每个Query在各个引擎的满意度得分，则各引擎的得分可通过计算N个评测Query的均值获得。

（说明：其中 N：评测Query总数；：第i个评测Query在A引擎的整体满意度得分。）

# 附录

参考文献：Google July 20,2018《search quality evaluator guidelines.pdf》。