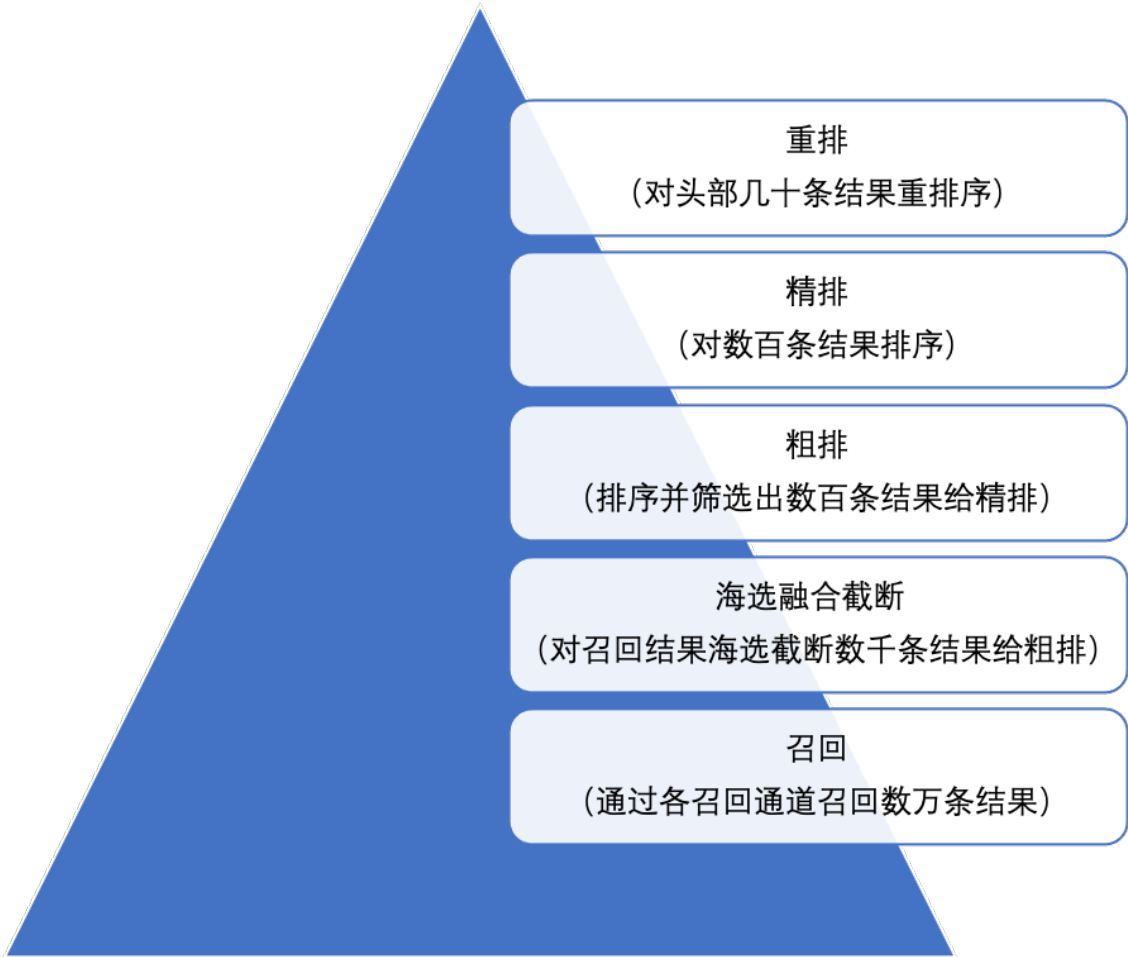


搜索排序

排序 是搜索链路中的最后一个阶段，排序系统普遍采用级联架构，以此平衡效率和效果，并为业务需求提供灵活的调整空间。



排序系统的整体定位和要求主要是基于搜索系统的核心目标，在提升用户体验、确保业务目标、以及系统可扩展性和计算效率的基础上，进行精细的排序优化。

排序的核心定位和具体要求应当如下：

- 用户体验
 - 排序的最终目的是提供用户最相关且高质量的内容，确保用户在搜索或推荐过程中能快速找到所需信息或产品
 - **相关性与个性化**：根据用户的查询意图、历史行为和偏好来提供精准匹配的结果。通过理解查询的语义和用户的需求，返回最相关的文档或项。个性化排序则应该考虑用户的历史点击行为、兴趣标签、社交关系等因素
 - **点击率与转化率优化**：对于电商或内容平台，排序应当能够优化用户的点击行为和后续转化（如购买、分享、收藏、评论等）。排序系统需要动态地根据不同场景和目标调整排序策略
 - **多样性与新鲜度**：排序应能平衡内容的相关性和多样性，避免返回过于相似的内容，并提供新鲜、时效性强的内容，确保系统的响应内容不单调

- **速度与低延迟**：排序的响应时间应尽可能低，尤其是对于实时搜索和高并发场景，保证用户体验流畅

- **业务目标**

- 排序系统不仅需要关注用户体验，还需要在符合用户需求的基础上达成业务目标。不同业务场景可能会有不同的排序需求：
 - **商业化目标**：对于电商或广告类排序，重点在于优化广告主的曝光和用户的转化，特别是在广告和自然搜索结果的平衡上，避免广告内容影响搜索体验，同时又能有效推广商业产品
 - **内容曝光与流量控制**：对于信息流或新闻类搜索，排序系统需要保证内容的多样性和公平性，避免过度曝光某些来源的内容或重复的文章。同时，需要控制某些内容的流量（例如冷启动内容、特定作者、优质新内容等）
 - **防作弊与内容质量控制**：排序系统应包含防作弊机制，确保低质量或虚假内容不会通过排序进入前列，防止垃圾信息和恶意内容干扰用户体验

- **数据反馈机制**

- **实时监控与调优**：系统应能够实时监控排序结果，及时发现异常或偏差，调整排序算法和策略，保证排序目标的长期一致性
- **在线反馈学习**：根据用户的点击、停留、转化等行为数据，实时更新排序模型的权重和参数，持续优化排序效果

排序体系通常分层为粗排、精排、重排、混排，其中：

- **粗排**

- 目标是从大量召回候选文档中快速筛选出一个较小的候选集。粗排通常会选择计算复杂度低、处理速度快的算法，确保能在较短时间内从大量数据中快速筛选出相关性较强的文档

- **精排**

- 目标是从用户体验和业务需求的角度，采用相对复杂的模型对粗排给到的候选文档进行精细排序

- **重排**

- 对精排结果的进一步调整，通常结合业务规则、个性化需求、流量控制等进行干预，确保排序结果符合特定的业务目标和用户需求

- **混排**

- 将不同类型的内容（如广告和自然搜索结果）或不同业务场景下的内容，按照一定的策略和算法进行统一排序和展示

特别的，对于搜索系统而言，Query-Doc 的相关性是决定用户体验的关键，并在粗排、精排中建模。本章节将相关性模块单独用一个篇幅介绍，粗排、精排、重排、混排分别用四个篇幅介绍。

[相关性](#)

[粗排](#)

[精排](#)

[重排](#)

[混排](#)