

Qcot相关标准

0、背景&目的

2、Qcot具体维度要求

2.1、需求信息补齐

2.2、扩展与泛化

2.3、问题拆解

2.4、表达规范化

3、Qcot 评测 Rubric (0-3 分)

3.0整体打分规则

3.1 需求信息补齐 (Completeness)

3.2 扩展与泛化 (Expansion)

3.3 问题拆解 (Decomposition)

3.4 表达规范化 (Normalization)

0、背景&目的

这里定义的仅针对query的改写，一个“好”的改写并非有一个绝对统一的标准（毕竟每个人的思考方式都不一致），但它必须服务于下游任务的目标（即召回材料的相关比例），但一般需要从以下场景来进行相关改写

一句话总结：不改变原意图（最高原则），通过补全、扩展、拆解、规范化等手段策略提升召回RAG质量

- **信息查漏补缺：**用户需求未明确但有一定指向的具体主体的条件信息、LBS等默认信息等进行信息补全，形成更完整的查询query
- **扩展与泛化：**在不改变核心意图的前提下，提升信息的覆盖广度与深度
- **问题的拆解：**将复杂、多跳的问题拆解为多个子问题，逐一解答后整合结果
- **表达规范化：**将口语化、模糊、不规范的表达，转化为适合检索的、书面化的规范表达

2、Qcot具体维度要求

2.1、需求信息补齐

用户需求未明确但有一定指向的具体主体的条件信息、LBS等默认信息等进行信息补全，形成更完整的查询query

分类	具体说明	改写示例
需求主体细化	用户需求未明确具体主体的条件信息，整体补全细化	专升本读几年书—> 统招2025年最新学制、自考2025年最新学制、成考专升本最新学制、国开专升本最新学制 虚拟键盘怎么调出—>“ 虚拟键盘调出方法”“iOS虚拟键盘调出步骤”“安卓虚拟键盘设置”“Windows虚拟键盘开启”“Mac虚拟键盘显示”
默认信息补齐	1、时间、地点等默认信息补齐 2、查询有相关官方信息时，带上官方、官网、公告等限定条件	"国庆放假多少天"—>"2025年国庆放假天数" "周末去哪里玩比较好"—>"2025年10月北京周末游玩推荐" "京东白条密码怎么设置"—>“ 京东白条支付密码格式要求2025”“京东白条密码设置路径 官方指引” "国土局招聘"—> 国土局招聘官网、国土局2025招聘公告、国土局考试科目地区差异
消除歧义	用户输入信息有误or不规范，但有明确指向	PS机刷信用卡费率高违法吗—> POS机刷信用卡费率高违法吗 水果14价格—> “苹果14价格”
补全上下文	补充 <u>隐含上下文</u> ，仅多轮对话场景	

2.2、扩展与泛化

在不改变核心意图的前提下，提升信息的覆盖广度与深度

分类	细化	改写示例
主体分点扩展	主体条件明确，但主体下的分点进行分析，可做一定扩展	百合花写作背景—> 茹志鹃百合花创作时间、茹志鹃人生经历 大疆action5怎么样—>“ 大疆action5用户评价”“大疆Action 5发布时间”“大疆 Action 5防抖技术”“大疆Action 5电池续航”“大疆Action 5 拍摄参数”“大疆Action 5防水性能”“大疆Action 5 评测”

主体横向对比	主体高度相似含义的横向其他实体对比，作为信息补充	我国第一部国别体史书——>中国第一部国别体史书、国语成书时间、 春秋 左传 战国策的体裁 大疆action5怎么样—> " 大疆 Action 5 vs GoPro 14 "
--------	--------------------------	--

2.3、问题拆解

将复杂、多跳的问题拆解为多个子问题，逐一解答后整合结果

分类	细化	示例query
问题拆解	原query有多个并列关系的查询意图或者查询实体	小红书WiFi数据流量推流区别—> “小红书数据流量 推送限制” “小红书WiFi推送频率” 360安全浏览器和360极速浏览器哪个好—>"360安全浏览器 极速浏览器 对比"， " 360安全浏览器特点 "， " 360极速浏览器特点 "

2.4、表达规范化

- 能将口语化、模糊、不规范的表达，转化为适合检索的、书面化的规范表达。

分类	细化	示例query
口语化	常用的提法、去口语化、书面化表达	手机老卡还烫 -> 手机频繁卡顿和发热 原因和解决方法 马化腾是谁 -> 马化腾简介 美债是什么 -> 美债介绍
语句	疑问句变成陈述句	"去哪玩周末" -> "周末去哪玩" "怎么样iphone15" -> "iPhone15怎么样"
缩写	补全缩写	浙大 -> 浙江大学
特殊字符	修正乱码或特殊字符、拼音改汉字（只修正文字错误，而不改变原意）	"天猫精灵【使用说明】" -> "天猫精灵使用说明"

3、Qcot 评测 Rubric (0–3 分)

3.0 整体打分规则

总分	等级说明
3 分 (优秀)	意图完全保真；改写精准、有必要的补齐；扩展与拆解合理；表达规范；显著提升召回率与覆盖面。
2 分 (可用)	意图可靠；补齐与扩展基本到位但有少量欠缺；整体召回价值较高。
1 分 (有限价值)	基本保持意图，改写过于接近原句，缺少应补齐的信息；
0 分 (不可用)	意图偏离、拆解错误、补齐虚构、表达混乱；不具备检索价值。

3.1 需求信息补齐 (Completeness)

用于判断改写是否补齐 主体 / 时间 / 地点 / 默认前提 / 官方信息 / 范围限制 等用户未明确但必要的信息。

分数	标准	case (Q=「大疆 Action5 怎么样」)
3 分	补齐关键缺失信息，重要条件完整明确	大疆 Action5 官方参数 (含发布时间/配置)
2 分	补齐部分关键信息但仍有遗漏	Action5 参数介绍
1 分	补齐不足，未显著提升检索价值	大疆 Action5 视频拍摄模式
0 分	补齐错误、虚构或无关信息	大疆 Action5 在 2024 年销量 (错误前提)

3.2 扩展与泛化 (Expansion)

评估模型是否围绕主体进行 合理的分点扩展 / 横向对比 / 关联维度补充，以提升召回覆盖面。

分数	标准	case (Q=「大疆 Action5 怎么样」)
3 分	扩展维度充足、合理、有明显召回收益；不过度、不跑偏	大疆 Action5 实拍样张评测

2 分	有扩展但不够全面，或略显保守	大疆 Action5 对比 GoPro 13
1 分	扩展稀少或缺失应有维度	Action5 参数（非扩展，只是重述）
0 分	跑题或无扩展（需扩展的场景下）	GoPro 13 功能介绍（主体偏移）

3.3 问题拆解 (Decomposition)

适用于 **复杂、多跳、多实体、多维度 Query**，需要拆解为可检索的若干子问题，并在必要时提供“总括 Query”。

示例 Query：

分数	标准	case (Q= “对比下大疆 Action5 和 GoPro 13 的参数和稳定性”)
3 分	准确识别多层结构，完整拆解为多个可检索子问题；如需整合，补充总括 query	<ul style="list-style-type: none"> - Action5 参数 - GoPro 13 参数 - Action5 vs GoPro 13 稳定性对比
2 分	部分拆解正确，但略有遗漏或粒度不均衡	<ul style="list-style-type: none"> - Action5、GoPro 13 参数对比（缺少稳定/场景）
1 分	仅做表层拆解；关键步骤未拆开	<ul style="list-style-type: none"> - Action5 与 GoPro13 对比（笼统）
0 分	未识别复杂结构、未拆解或拆解错误	<ul style="list-style-type: none"> - 哪个更好（未拆解复杂任务）

3.4 表达规范化 (Normalization)

评估语言是否从口语化、模糊、非标准表达 → 转化为 **正式、清晰、可搜索、可结构化理解** 的表达。

分数	标准	示例
3 分	完全规范：陈述句化、补全缩写、纠错、消歧义	「大疆 Action5 用户评价汇总」
2 分	大部分规范化完成但仍略口语	「Action5 评测好不好」

1 分	规范化不足；多为用户原话	「action5 好不好」
0 分	未正规化或表达混乱	「action5 啥水平啊」

松汝 421361

松汝 421361