

浅谈 FAQ

分享人：杨夕

一、引言

- 场景：假设 你有一个 标准的问题库，此时 有一个 新 query 进来，你会做什么操作？
- 灵魂三连问：
 - 如何根据 这个 query，你怎么返回一个标准答案呢？
 - 如何从 问题库 里面找 答案？
 - 如何 判断 你 找到 答案 是 对的？

二、问答系统是什么？

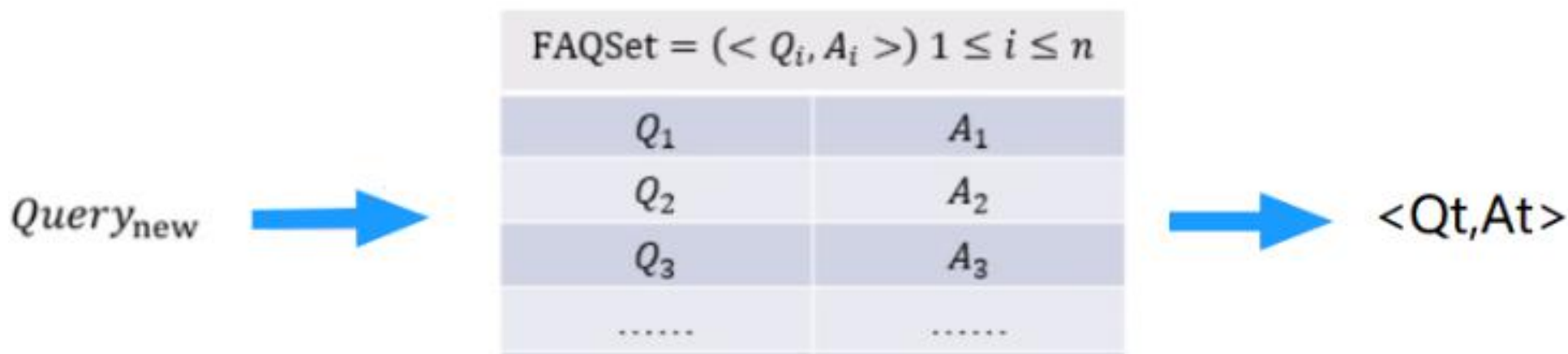
- 介绍：问答系统是信息检索的一种高级形式，能够更加准确地理解用户用自然语言提出的问题，并通过检索语料库、知识图谱或问答知识库返回简洁、准确的匹配答案。相较于搜索引擎，问答系统能更好地理解用户提问的真实意图，进一步能更有效地满足用户的信息需求。
- 目标对象：用户输入的 query 和 答案；

三、问答系统有哪些？

- 领域类型分类：
 - 面向限定域的问答系统；
 - 面向开放域的问答系统；
 - 面向常用问题集的问答系统；
- 根据答案来源分类：
 - 基于结构化数据的问答系统（eg: KBQA）；
 - 基于文本的问答系统（eg: 阅读理解）；
 - 基于问答对的问答系统（eg: FAQ）；
- 答案的反馈机制分类：
 - 基于检索式的问答系统；
 - 基于生成式的问答系统；

四、FAQ 检索式问答系统 是什么？

- 名称： (Frequently asked Questions) ， 是检索式问答系统
- 目标： 给定 标准问题库， 匹配与用户输入 query 最相近的问题并返回答案；
- 流程：
 - 分析 用户输入的 query;
 - 计算 与 query 最相近 的标准问题;
 - 做 Rank, 并取 Top1 的标准问题;
 - 返回 Top1 的标准问题的答案;



五、 query 匹配标准 QA 的核心是什么？

- 核心组成
 - 用户输入的 query 和 标准问题的 的相似度；
 - 常用方法：
 - 关键字匹配；
 - TF-idf 相似度计算；
 - 用户输入的 query 和 标准答案 的 匹配度；
 - 常用方法：
 - DL 模型 相似度计算；
 - Bert 句向量 语义相似度计算；

六、FAQ 检索式问答系统 方案篇

• 6.1 常用 方案有哪些？

- 方案一：QA 匹配：输入Query与候选的所有Answer 计算相似度，去找哪一个最合适
- 方案二：QQ 匹配：输入Query与历史语料库里的Query去找哪一个最相似，然后把找到的query对应的Answer取出即可；【常用】

六、FAQ 检索式问答系统 方案篇

- 6.2 为什么 QQ 匹配比较常用？

- | 3.2.1 QQ 匹配的优点有哪些？

- 原因 1：语义空间
 - 原因 2：语料的稳定性
 - 原因 3：业务回答与算法模型的解耦
 - 原因 4：新问题发现与去重
 - 原因 5：新问题发现与去重

七、FAQ 标准问题库构建篇

- 7.1 如何发现 FAQ 中标准问题？

- s1: 对 问题库 中的问题进行预处理;
- s2: 将 问题库 中的问题向量化;
- s3: 对 问题库 进行聚类;
- s4: 对比已有的FAQ, 发现并补足未覆盖的知识点。将FAQ与知识点——对应

七、FAQ 标准问题库构建篇

• 7.2 FAQ 如何做合并？

- FAQ 的最终目标：用户的每一个意图都能对应到唯一的FAQ，这样用户每次提问的时候，可能会出现以下两种情况：
 - 一一对应：后台能够根据意图对应的FAQ直接给出一个直接的回答；
 - 模糊回复：当用户的提问，所对应的多个FAQ意思过于相近，那么当用户问到相关问题时，就不会出现一个直接的回答，而是多个意图相关的推荐问题，此时，需要用户选择比较符合的回答，后台需要对触达率较高的FAQ进行自动分析，分析哪一些问句总是被推荐相同的答案，将问句对应的意图合并。

七、FAQ 标准问题库构建篇

- 7.3 FAQ 标准库如何实时更新？
 - 淘汰机制：后台需要定期分析历史日志，采用淘汰机制淘汰废弃知识点，如已下线业务知识点等。

八、FAQ 标准问题库答案优化篇

- 8.1 FAQ 标准问题库答案如何优化？
 - 挖掘对话，进行答案优化：如果机器人已经正确识别意图但最后仍然转人工，说明知识库的答案不对，需要进一步修正这一类知识点相对应的答案。
 - 分析头部场景，回答应用文本、图片、自动化解决方案等多元化方式：比如在电商场景中，经常会有查询发货到货时间、订单状态等的场景。利用图示指引、具体订单处理等方式让用户操作更便捷。

参考

- 1. 【FAQ 检索式问答系统】 那些你不知道的事
(https://github.com/km1994/nlp_paper_study/blob/master/QA_study/FAQ.md)
- 2. 基于句子相似度的FAQ问答系统_aogogo的博客-程序员宅基地_faq问答系统(<https://www.cxyzjd.com/article/zhouboke/80591208>)
- 3. 基于深度学习的FAQ问答系统
(<https://cloud.tencent.com/developer/article/1196826>)
- 4. 你问我答之「YiBot知识体系运营知多少」
(<https://zhuanlan.zhihu.com/p/27145792>)

感谢倾听！！！！



个人微信



关于NLP那些你不知道的事