

问答知识库建设说明

一. 普通一问一答（无上下文和场景）

问答的编写应注意三个总原则：

1. **越多越好**：知识库内容越多，可供调用的信息就越多，相当于大脑越发达。
几万条的库和几十条的库，用起来差别会很大。
2. **前期人工维护**：知识库初步使用时还需要人工监测，如调整标准问答的设置，根据“无效问答”一栏中的内容做增添等。后期这样的工作会逐渐减少甚至避免。
3. **尽量设置标准问题**：便于算法衍生更多问法，减少人工维护量。

具体设置：

第一步：Excel 样表：

question	answer
法国的首都是哪个城市	是巴黎啦
《白鹿原》是谁的作品	陈忠实
什么季节最热	夏天，北半球的 6~8 月
.....

格式和操作都很简单，只需在 excel 里建立如上的两栏表格，左栏为“question”，右栏为“answer”（表头这两个单词不能省略，且必须为英文），然后在下面依次列出问答即可。

第二步：

问答库内的问题应设置为“标准问题”，以便计算。有以下几点需要注意：

- ① 使用标准普通话，尽量避免明显的方言或个人习惯用语，如：

你们公司给哪儿替晃呀？

依知不知地址啊

伦家想鸡道怎么去了啦

最佳设置应为：

你们公司地址在哪里？

你们公司具体位置在哪里？

贵公司在什么位置？

等

② 标准问题中不要出现错别字，如：

你们公司低（地）点在哪儿？

几天之内可以退火（货）？

③ 尽量设置语法成分较为完整的句子，不要只写短语，如：

官方网站（×） 你们官方网站是什么（√）

微信公众号（×） 你们有没有微信公众号？（√）

④ 库内问题最好用简明扼要的单句，不要设置内容过于复杂的长句，如：

我想问一下你们正在搞活动的性价比很高的至尊套餐，如果我是会员的话多少钱？（×）

至尊套餐会员价多少钱？（√）

嗨你好你好你好，你叫什么名字呀？（×）

你叫什么名字？（√）

我上星期在你们网站上买了一个吸尘器，觉得不好用，想问下是不是7天内可以退？（×）

7天内是否可以退货？（√）

7 天内能不能退货？（√）

7 天内可以退货吗？（√）

第三步：检查

Excel 表格中的问答上传完毕后，可根据自动检索功能提示，修改无法上传的问题，如减少题目或答案字数等。

注：库内已有的，完全重复的问题无法上传，可以稍作几个字的改动

补充说明：

1. **以上标准是为了与算法相结合，减少人工维护成本，不是硬性要求。**如第③条，其实也可以将知识库中的问题设置为“官方网站”，那么当问到“你们的官方网站是什么”时可能无法作答；但如果设置为后者，那么当问到“你们的官方网站是”，“你们的官方网站”，“你们官方网站”，“官方网站”均可作答。
2. 我们的算法会自动处理用户的真实问题语句，进行“模糊”、“去口语化”、“同意替换”和“容错”等处理，使之趋近于库内标准问题并调出精准答案。也就是说，库内如果多设置不同问法，回答准确率就能成倍增长。建议前期多根据“无效问答”维护知识库，增加问题，一劳永逸。

二. 上下文场景限制（如有需要）

如果需要编辑多轮对话，希望同一个问题在不同语境下呈现不同的答案，这时就需要用到“场景”功能。详见下图所示：

A14 fx				
	A	B	C	D
1	question	answer	scene	
2	我好想吃甜点啊	别吃啦，再吃都成小猪啦！	情侣	
3	讨厌	哈哈，逗你的。就喜欢你圆乎乎的样子	情侣	
4	看地上都是她吐的瓜子皮	你也讨厌这种行为吧	讨论	
5	讨厌	必须的！真没素质	讨论	
6				
7				
8				
9				
10				
11				

“场景”的设置也很简单，只要再加一栏“scene”即可。scene 栏内容最多不能超过 20 个字符，汉字和英文均可。

在不强制场景的情况下，输入“讨厌”，答案会随机出现；如果自己设置了场景，就会按照设置出现符合上下文的答案。

建议：如果不是对多轮对话以及上下文有特别的需求，不要设置 scene 栏，由系统自动判断场景即可。