



腾讯云知文

# 问答引擎在工业级实战中的演化

钟黎

腾讯云 AI语义研发负责人

# 腾讯云知文产品矩阵



TABLE OF

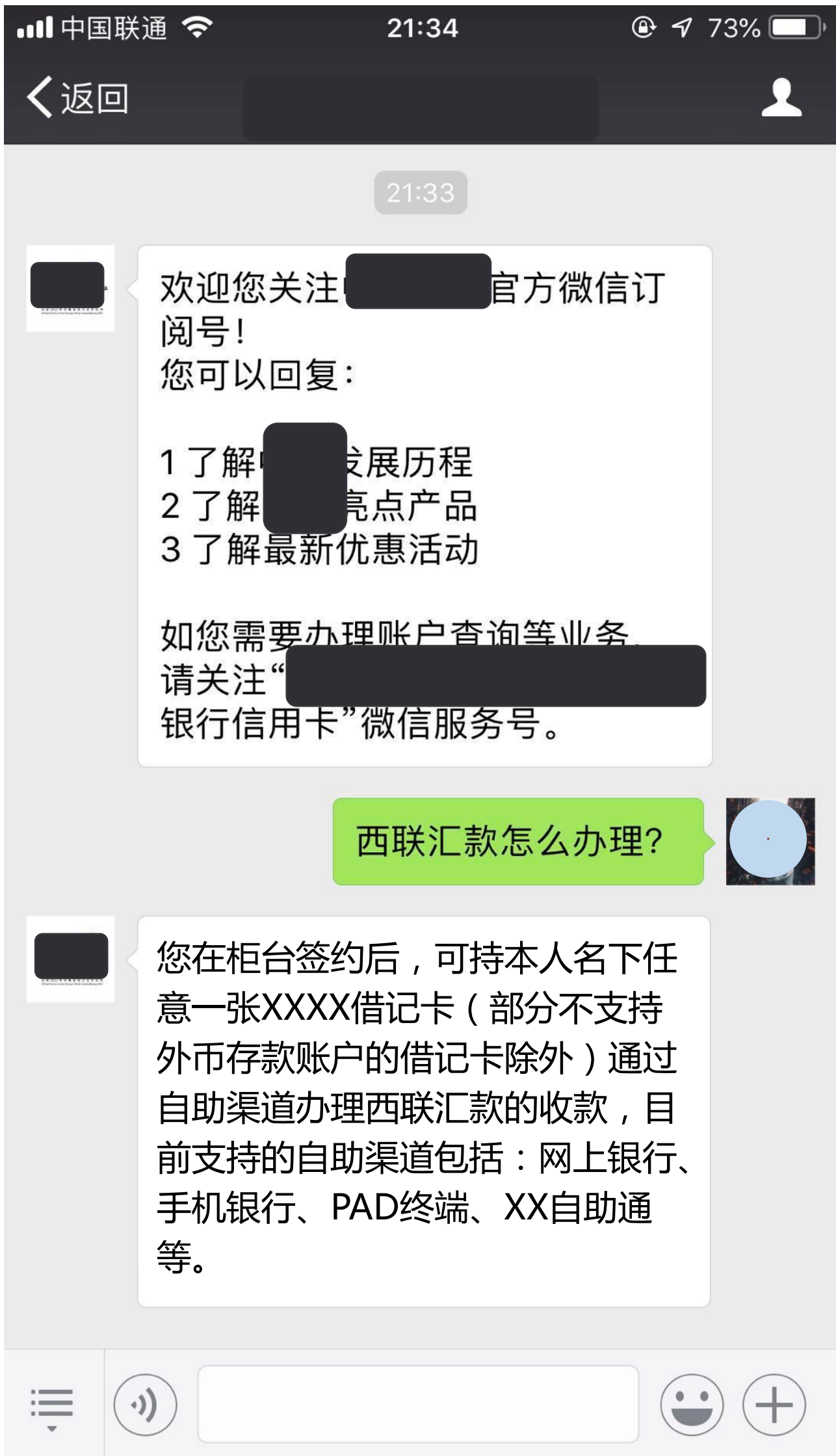
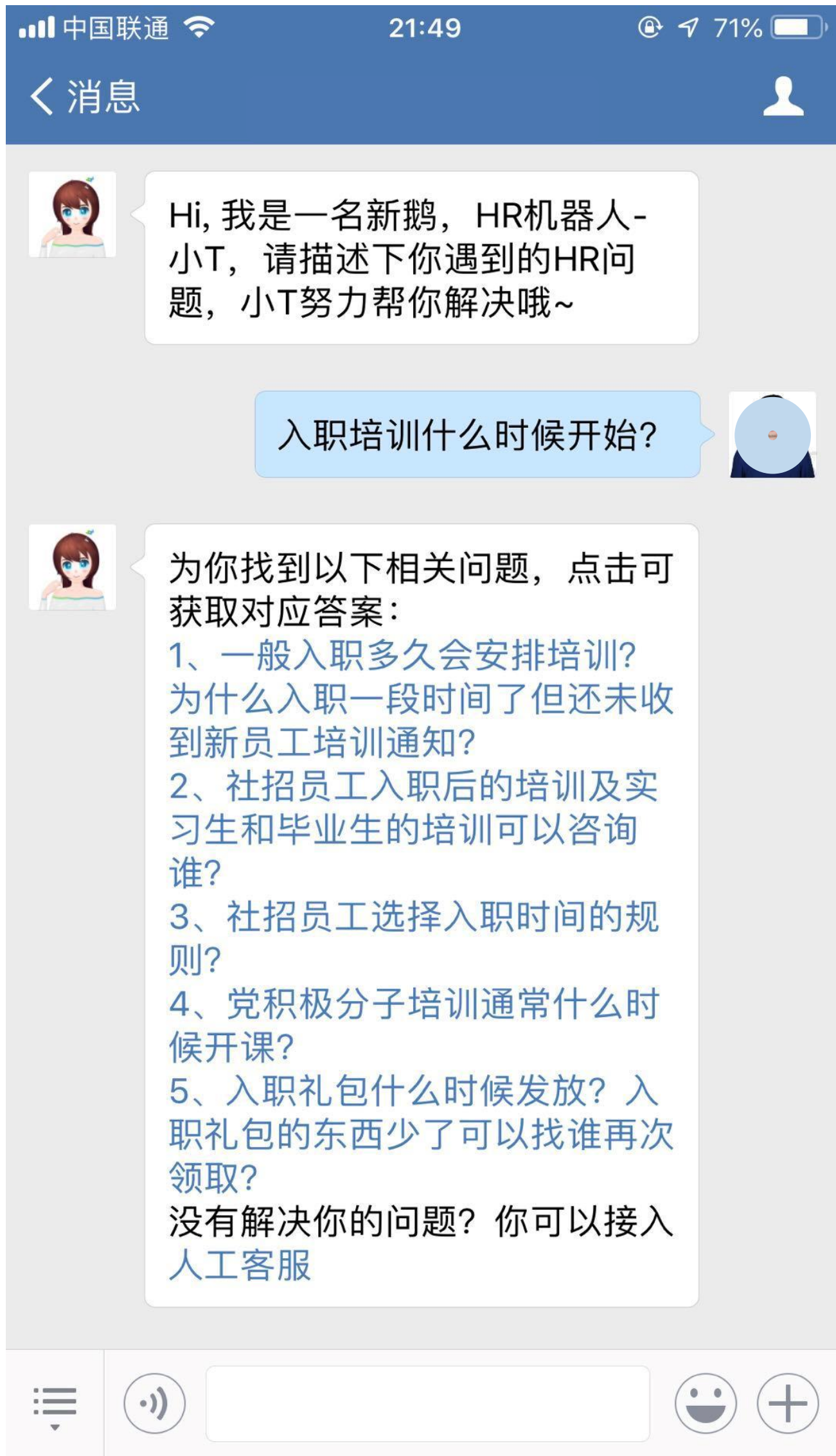
# CONTENTS 大纲

---

- 知文问答：从FAQ机器人说起
- 用活存量文档：文档型机器人
- 走向结构化数据：知识图谱机器人
- 知文技能树：任务型机器人
- Query处理、多轮对话与知识管理
- 总结



# 知文问答v0.1：FAQ问答的形式与本质



本质上是Q-Q匹配

## 找到的相似问题集：

自助渠道首次签约西联汇款解付的规定

自助渠道解付速汇金

自助渠道解付西联汇款的办理流程

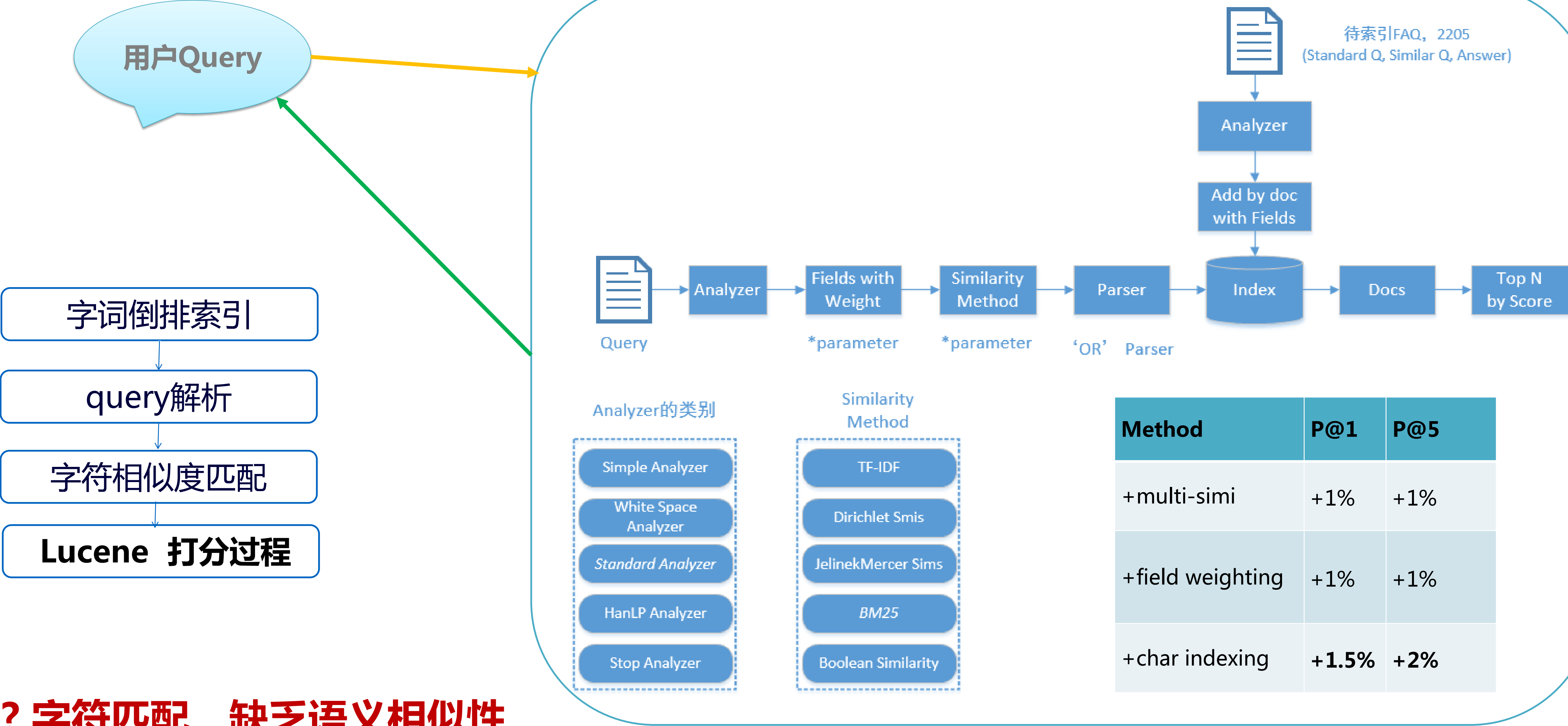
个人网银西联汇款解付操作步骤

XXXX卡充值办理渠道

西联汇款支持币种

.....

# 知文问答v0.1 : Lucene内核revisited



? 字符匹配，缺乏语义相似性

# 多通道召回：从字符召回到向量召回

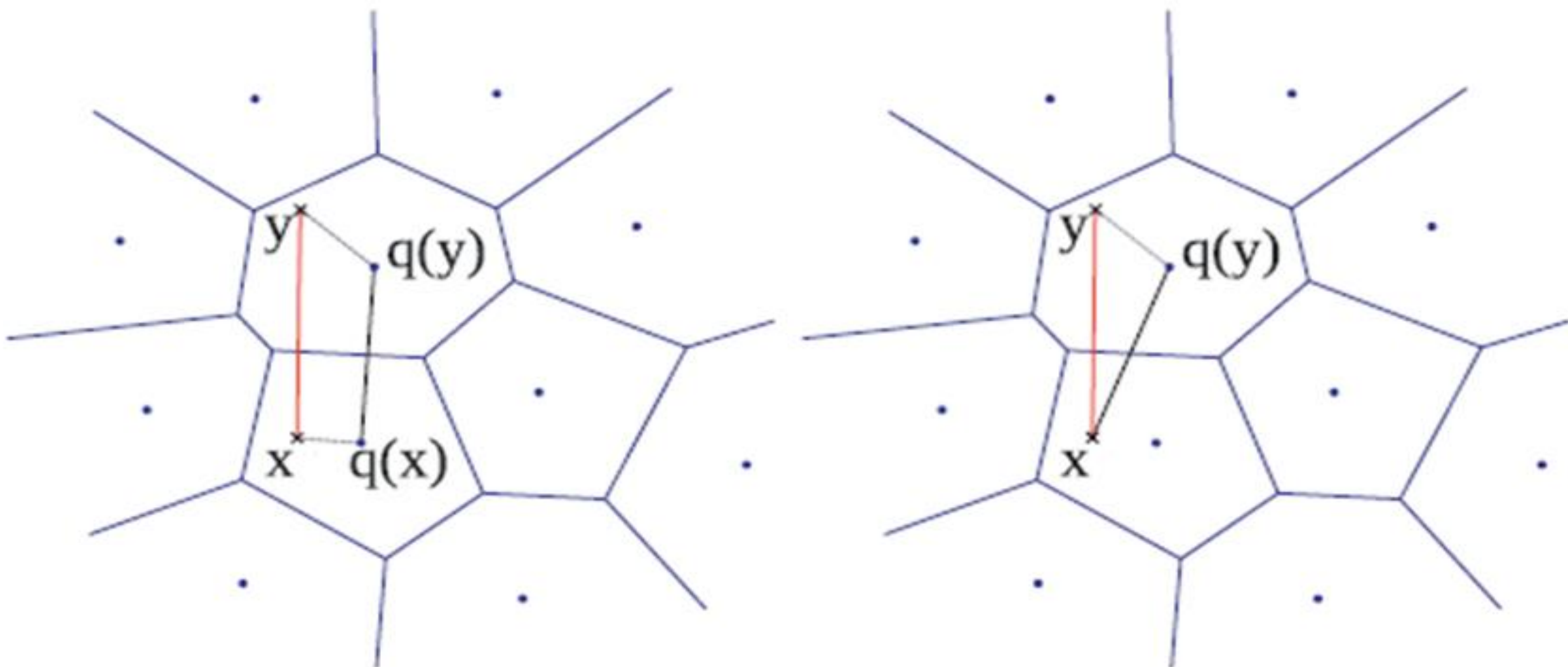
## 高维向量的召回

- K-d tree and its variants
- Local sensitive hashing
- Vector quantization

属性	KD-Tree	LSH	Vector Quantization
离线构造			
加速方式	空间分割	分桶哈希	聚类
降维特点	最大方差维度分割	保距映射	编码量化
维度要求	低维（<20）	高维	高维
存储	大	大	小
在线查找			
精度	高	对参数敏感	高
效率	低	高	低（可加速）

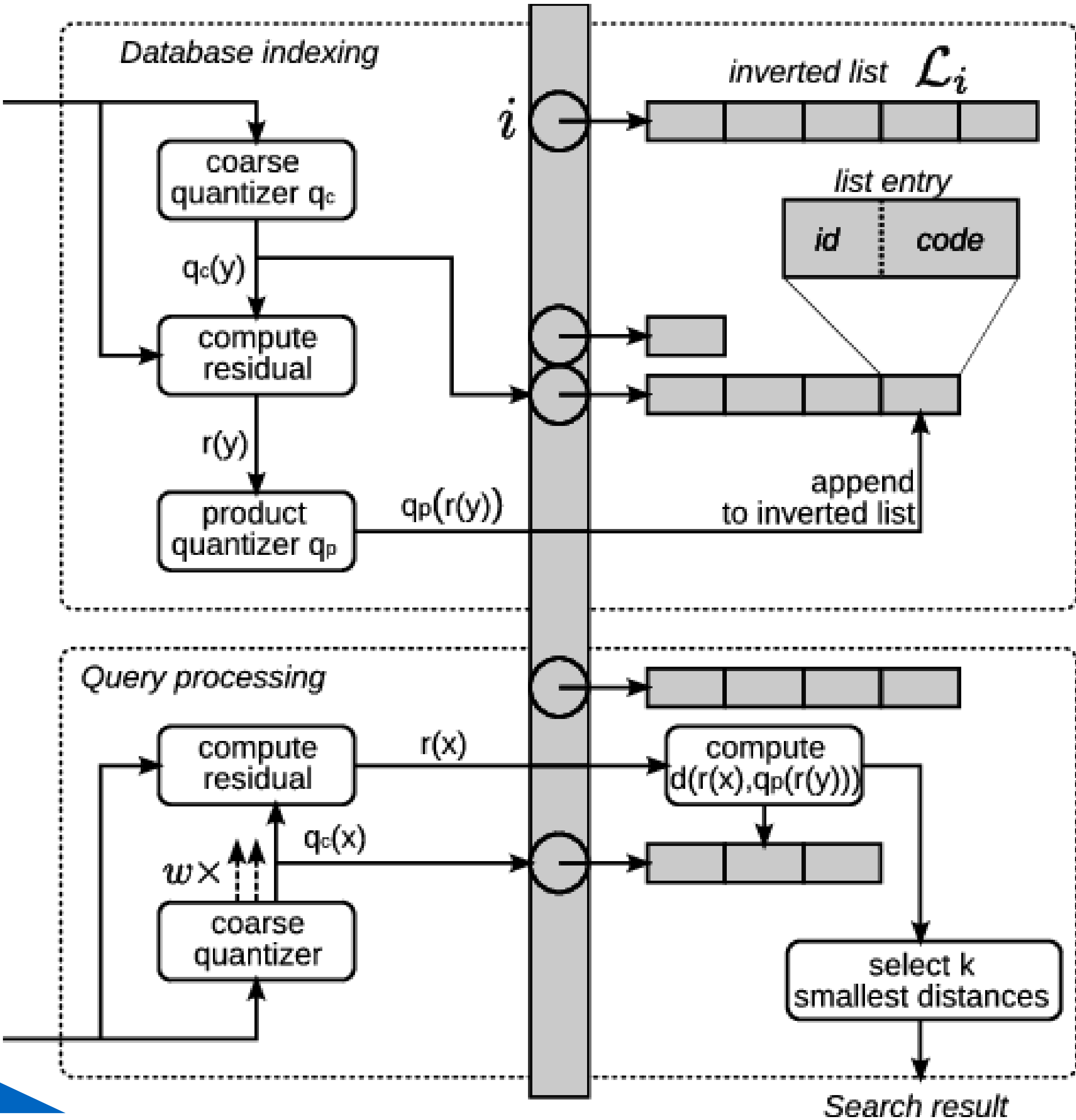
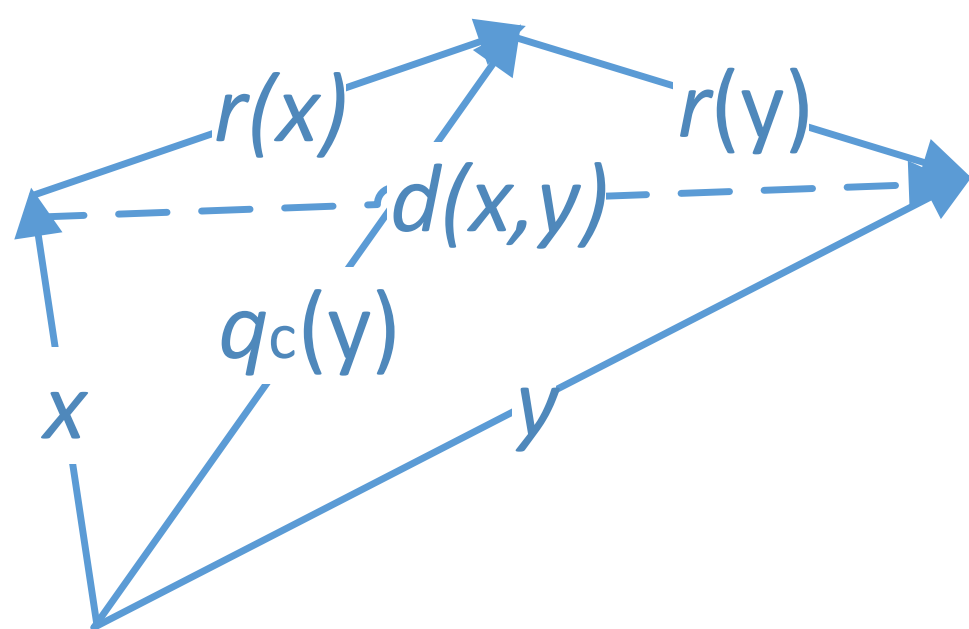


# 多通道召回：从字符召回到向量召回

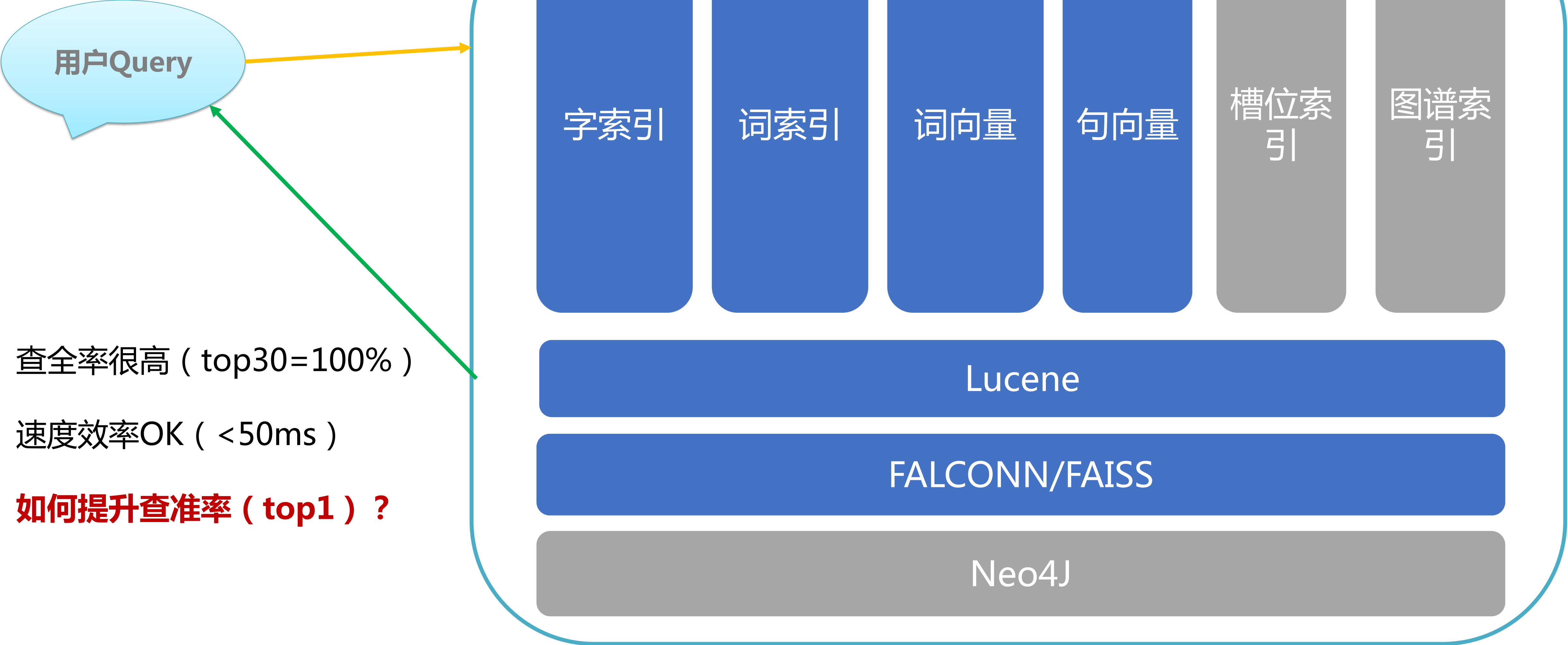


symmetric case

asymmetric case



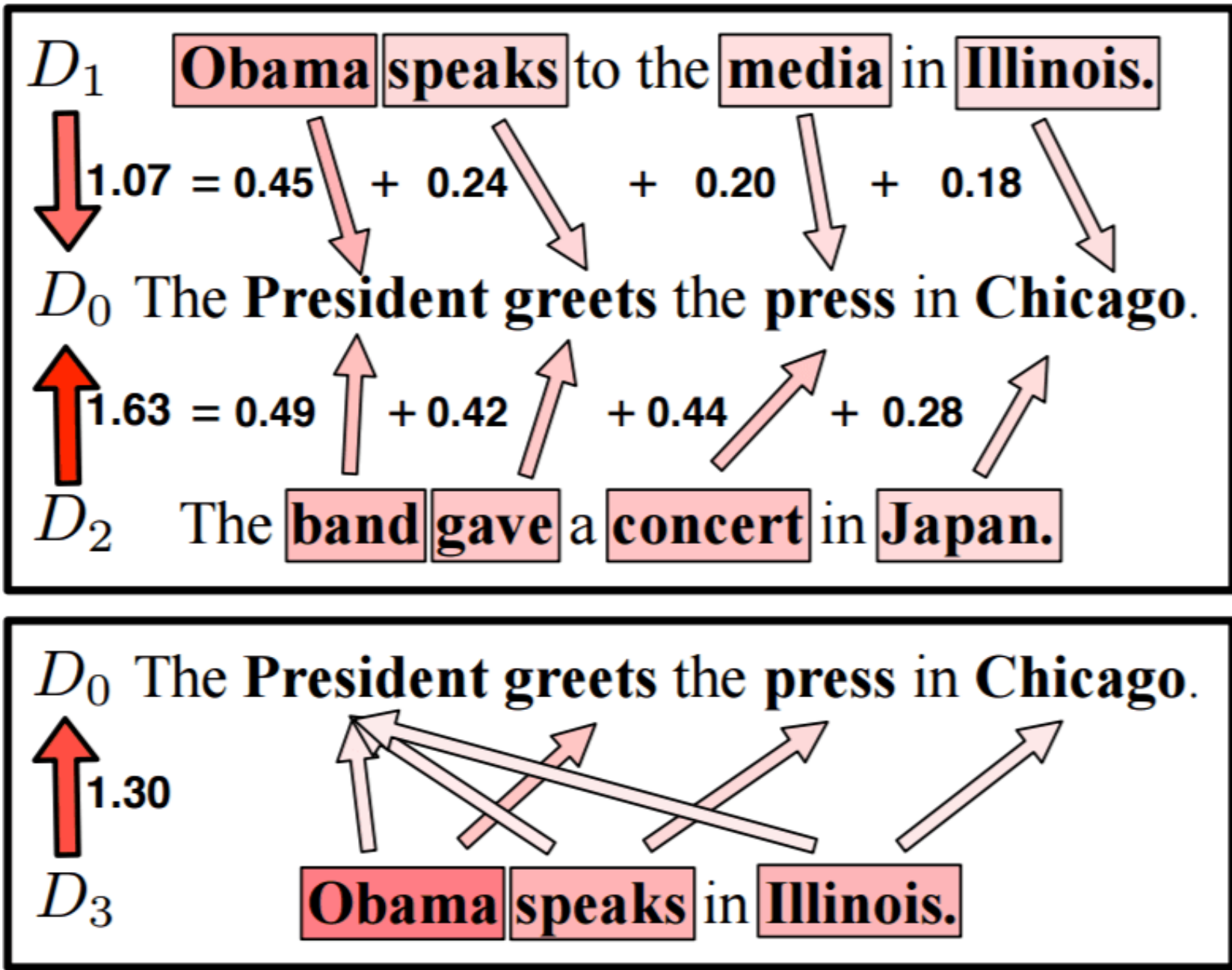
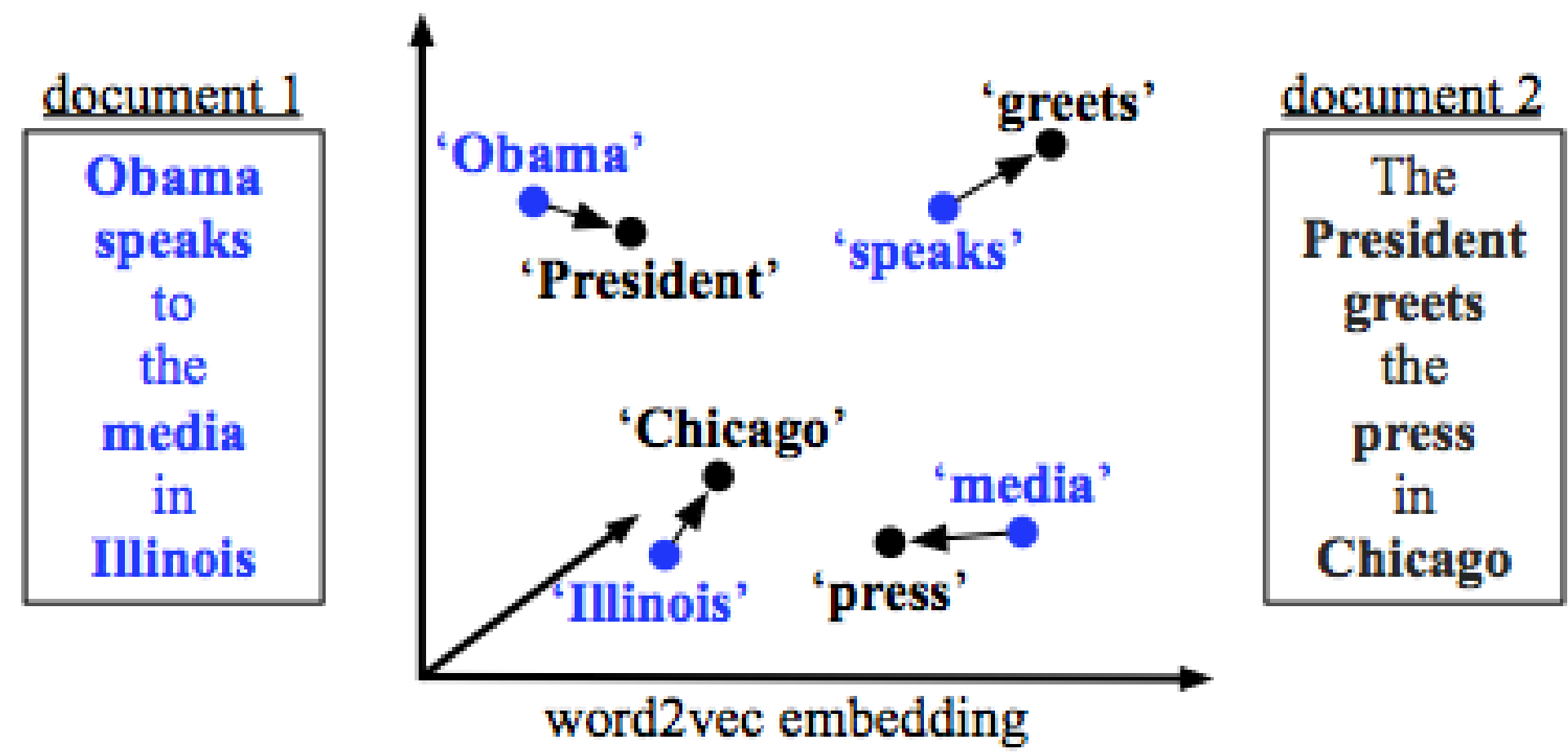
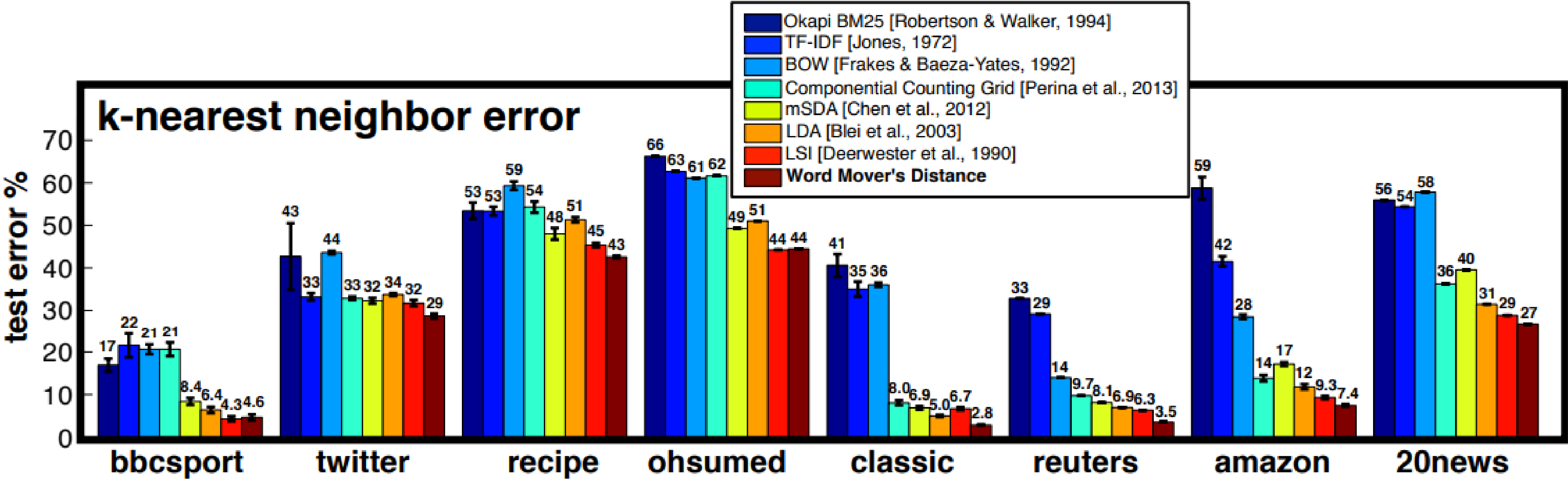
# 知文问答v0.2：多通道召回



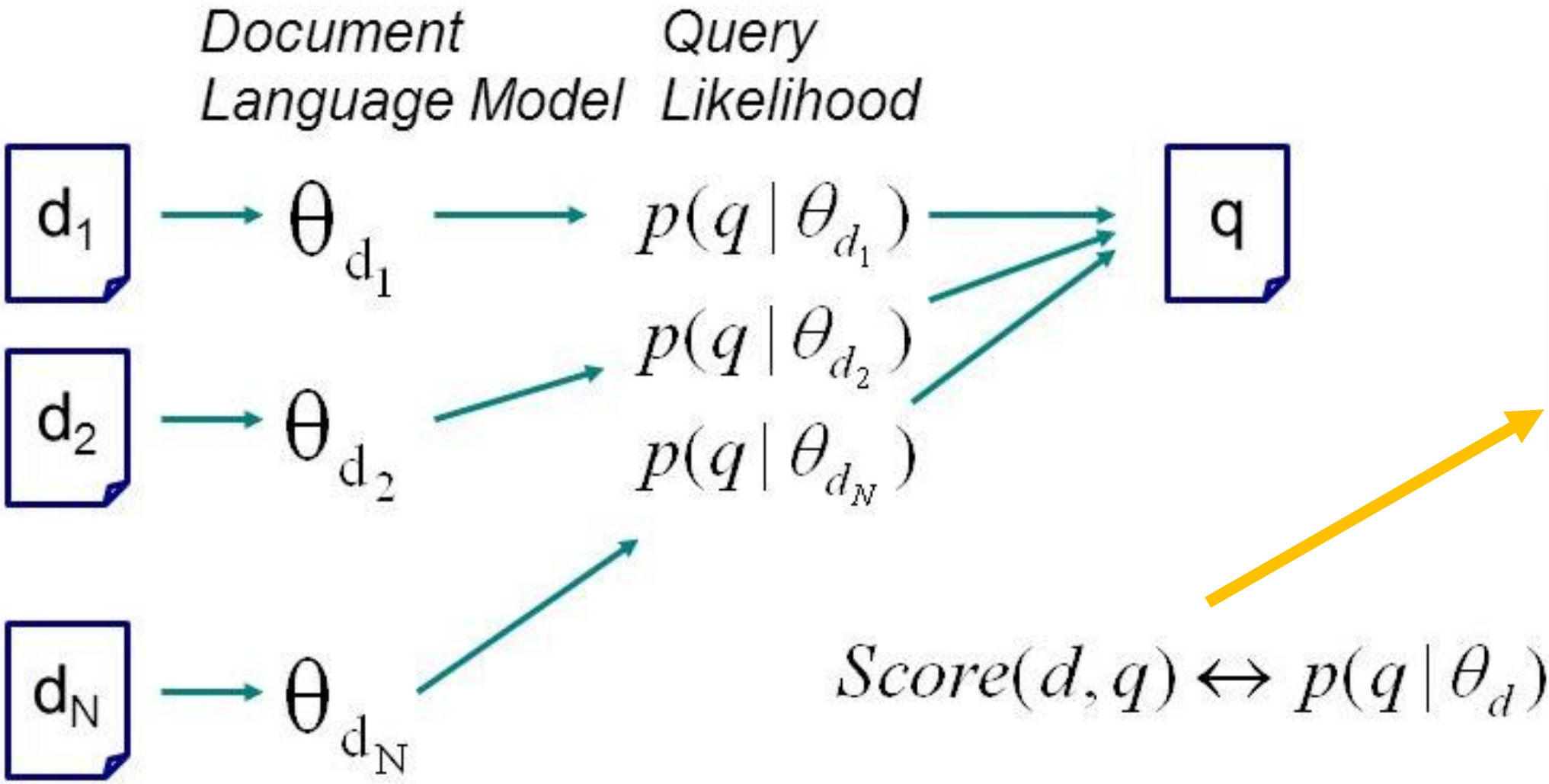
- 查全率很高 ( top30=100% )
- 速度效率OK ( <50ms )
- 如何提升查准率 ( top1 ) ?



# 从召回到（无监督）排序：WMD



从召回到（无监督）排序：语言模型



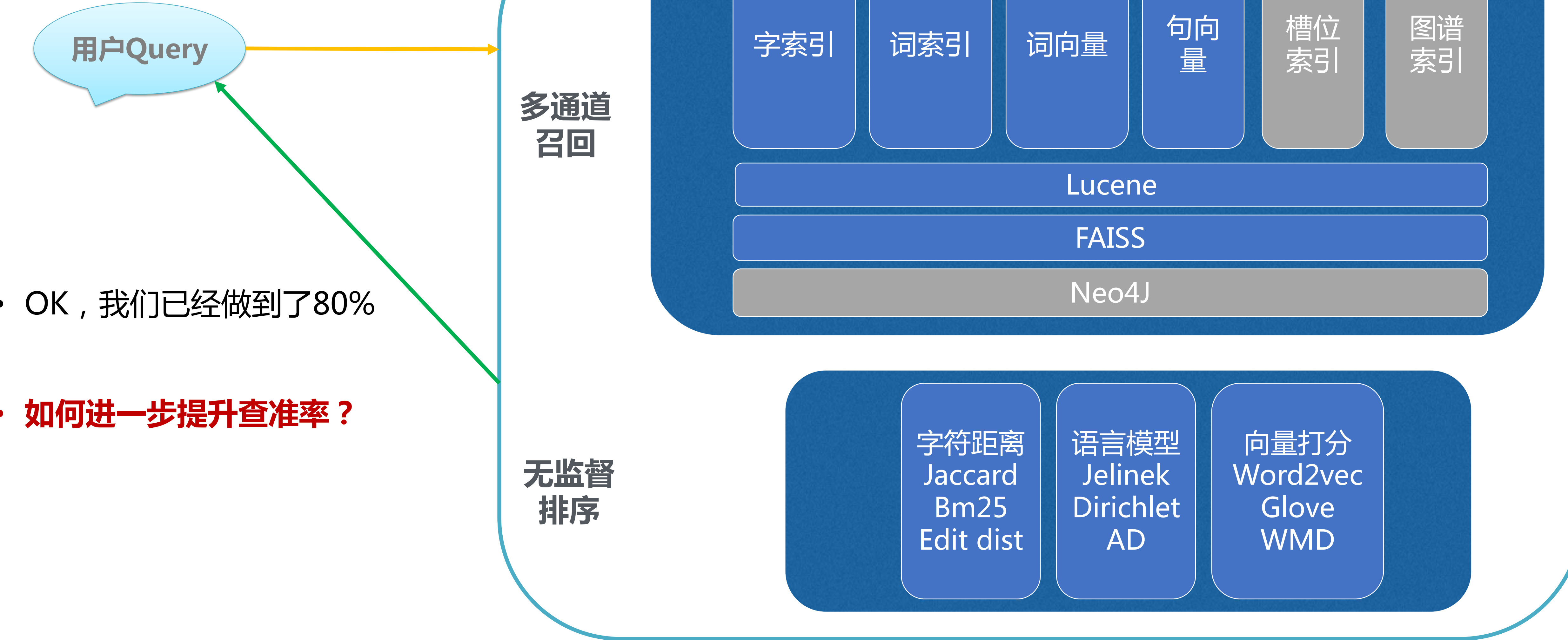
$$P_{MLE}(w_i | w_{i-1}) = \frac{c(w_{i-1}w_i)}{c(w_{i-1})}$$

$$p(w_n | w_{n-1}) = \lambda_1 p(w_n | w_{n-1}) + \lambda_2 p(w_n)$$

Method	$p_s(w   d)$	$\alpha_d$	Parameter
Jelinek-Mercer	$(1 - \lambda) p_{ml}(w   d) + \lambda p(w   \mathcal{C})$	$\lambda$	$\lambda$
Dirichlet	$\frac{c(w; d) + \mu p(w   \mathcal{C})}{\sum_w c(w; d) + \mu}$	$\frac{\mu}{\sum_w c(w; d) + \mu}$	$\mu$
Absolute discount	$\frac{\max(c(w; d) - \delta, 0)}{\sum_w c(w; d)} + \frac{\delta  d _u}{ d } p(w   \mathcal{C})$	$\frac{\delta  d _u}{ d }$	$\delta$

$$P(t | D) = \alpha \underbrace{P(t | \theta_D)}_{\text{the probability given to the term by the document language model}} + (1 - \alpha) \underbrace{P(t | \theta_C)}_{\text{the probability given to the term by the collection language model}}$$

# 知文问答v0.3：召回+无监督粗排





# 有监督精排：复杂特征+简单模型 => 端到端模型

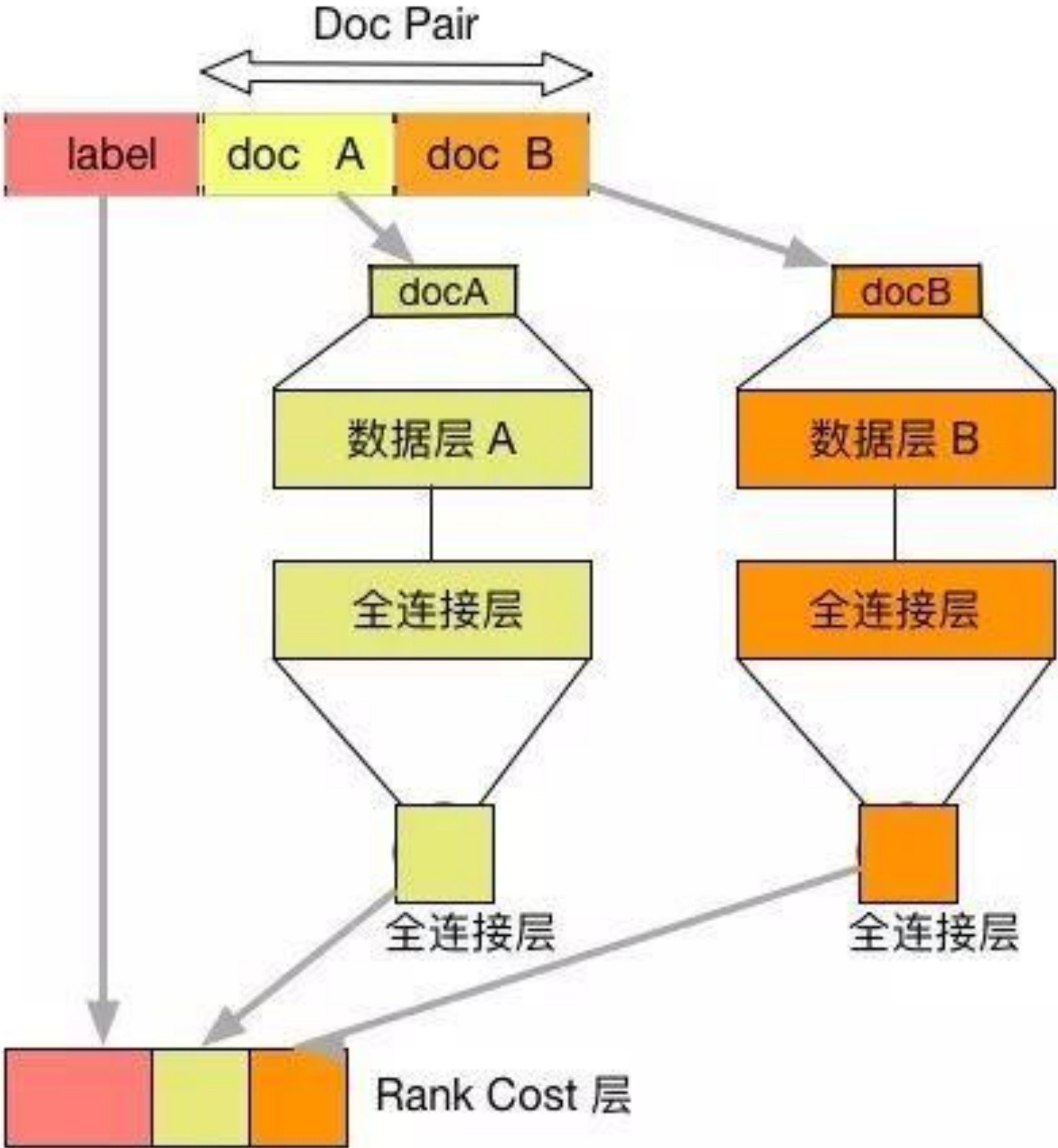
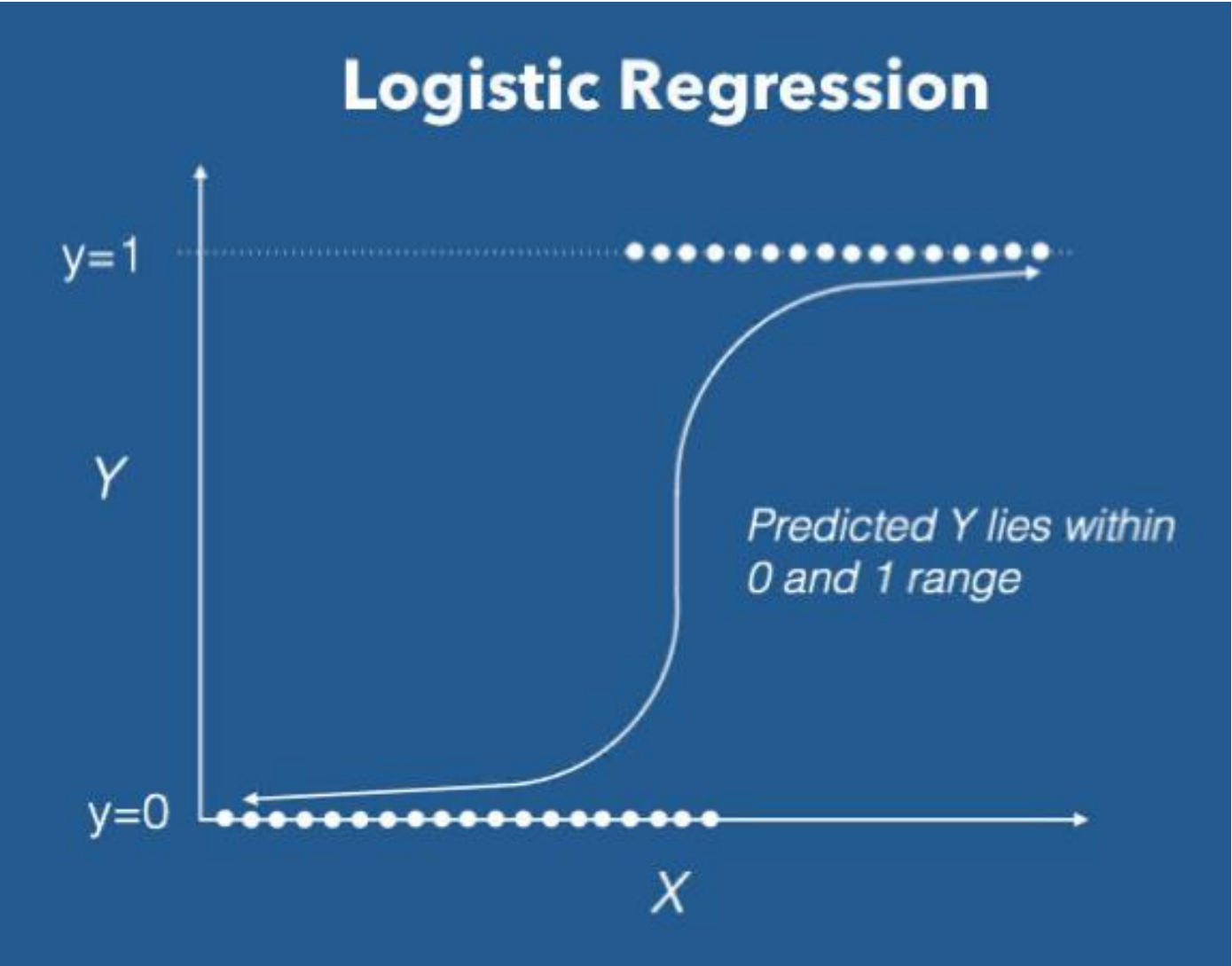
	Candidate 1	Candidate 2	Candidate 3	Candidate 4	Candidate 5	...	Candidate N	
Ranker 1	S <sub>11</sub>	S <sub>12</sub>	S <sub>13</sub>	S <sub>14</sub>	S <sub>15</sub>		S <sub>1N</sub>	Norma
Ranker 2	S <sub>21</sub>	S <sub>22</sub>	S <sub>23</sub>	S <sub>24</sub>	S <sub>25</sub>		S <sub>2N</sub>	
Ranker 3	S <sub>31</sub>	S <sub>32</sub>	S <sub>33</sub>	S <sub>34</sub>	S <sub>35</sub>		S <sub>3N</sub>	
...								
Ranker M	S <sub>M1</sub>	S <sub>M2</sub>	S <sub>M3</sub>	S <sub>M4</sub>	S <sub>M5</sub>		S <sub>MN</sub>	

Average

Resort

- Step1:归一化 (Min-max normalization)
- Step2:计算平均分
- Step3:重排序

$$s' = (s - min)/(max - min)$$
$$score_i = (s'_{1i} + \dots + s'_{Mi})/M$$
$$sort(score_{1\dots n})$$

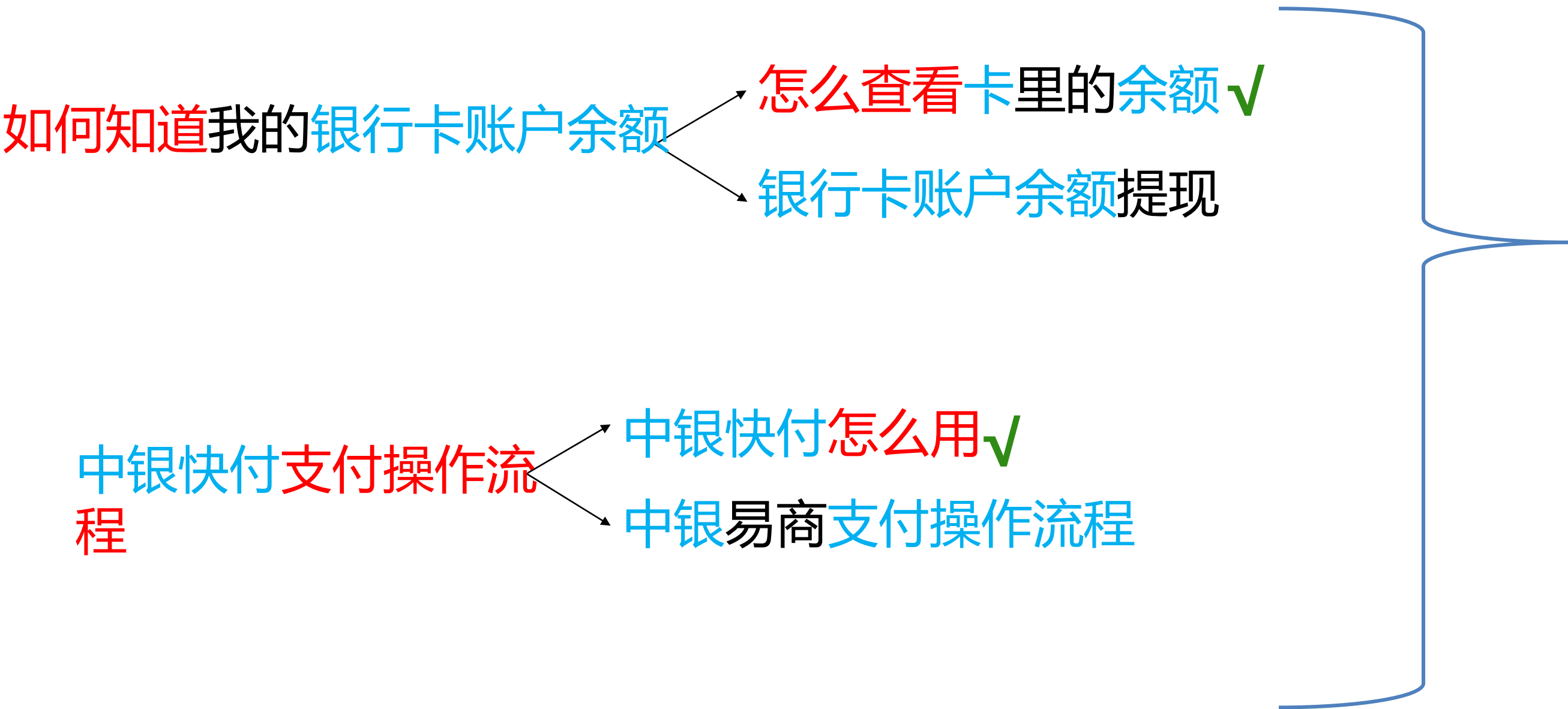




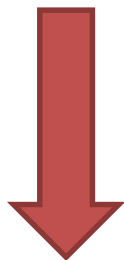
# 有监督精排：端到端模型 => 深度模型

字面上相似程度低，却在很多场景下表示相似含义。

字面上相似程度高，语义却不同。



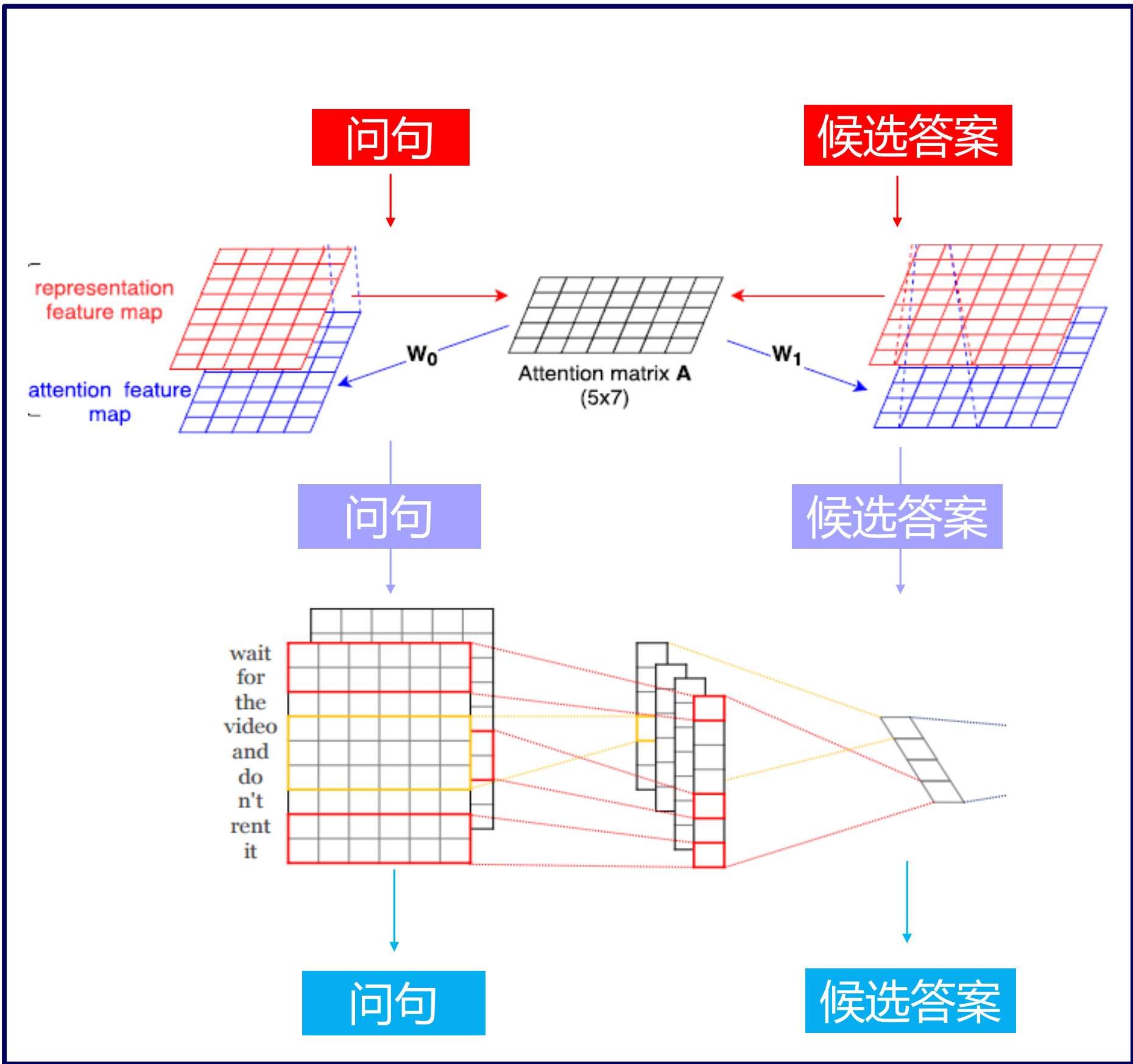
浅层模型解决不太好的问题



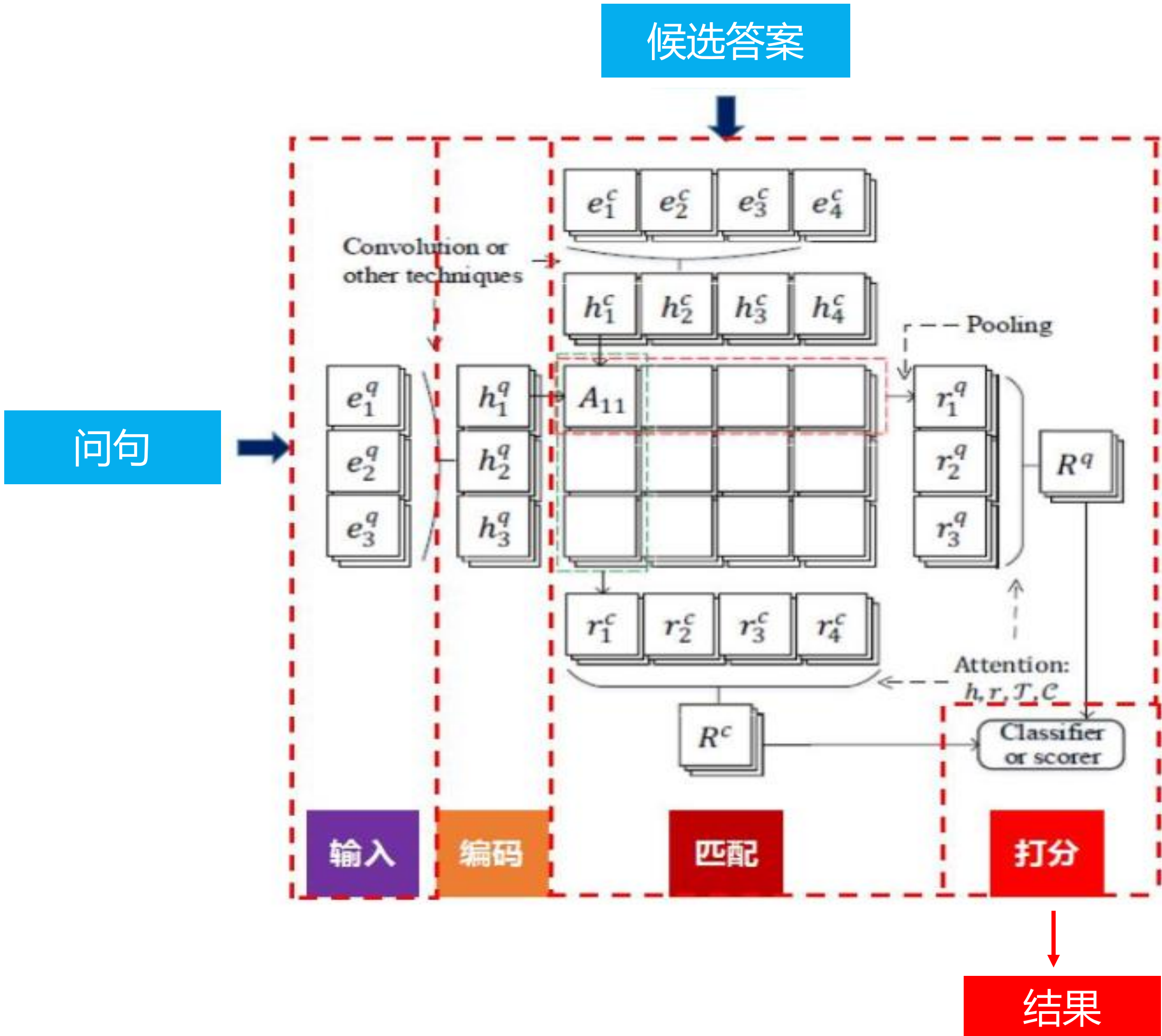
DeepRank，挖掘深层语义匹配关系

## 有监督精排：DeepRank

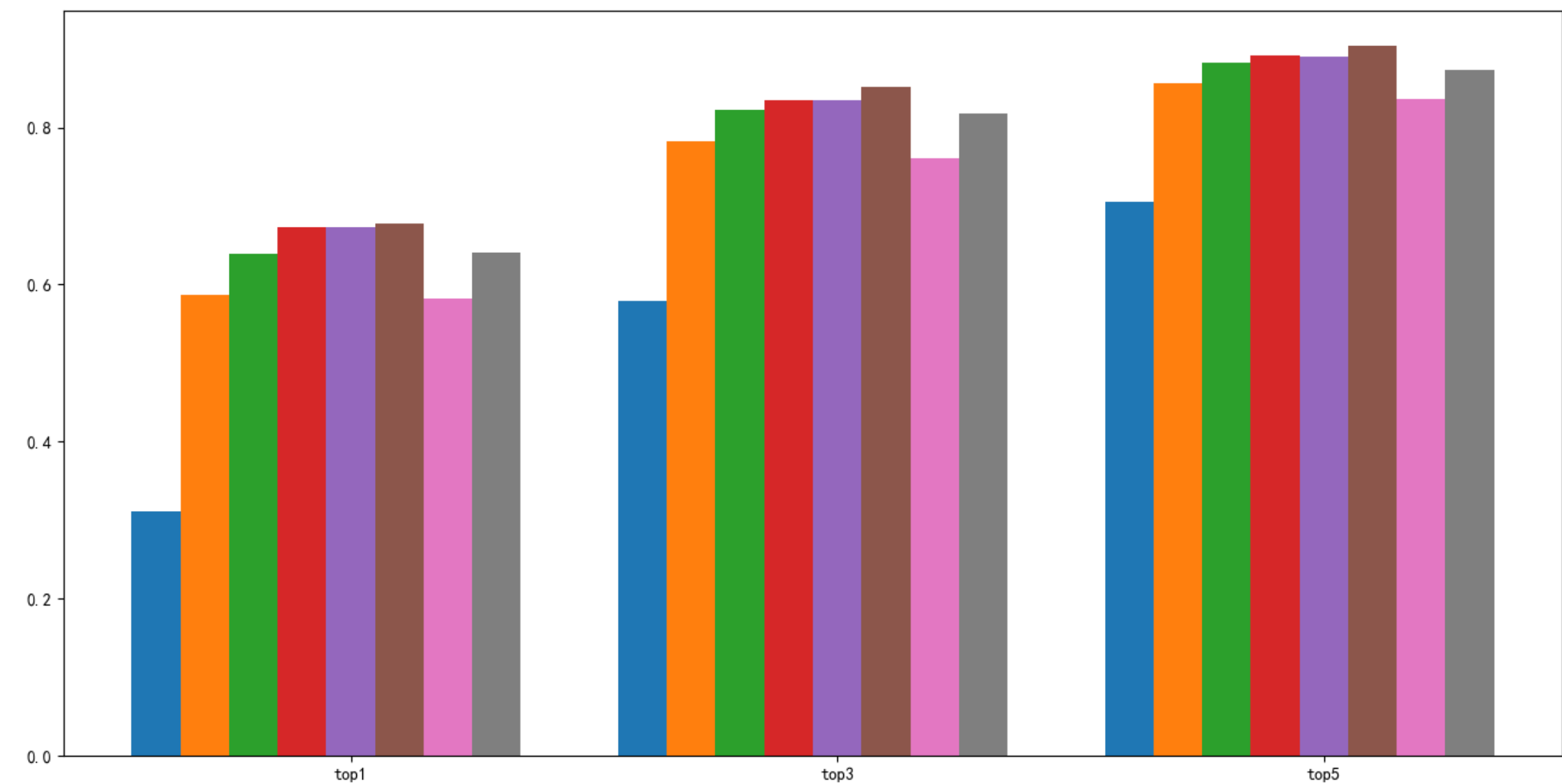
# 1 表征学习



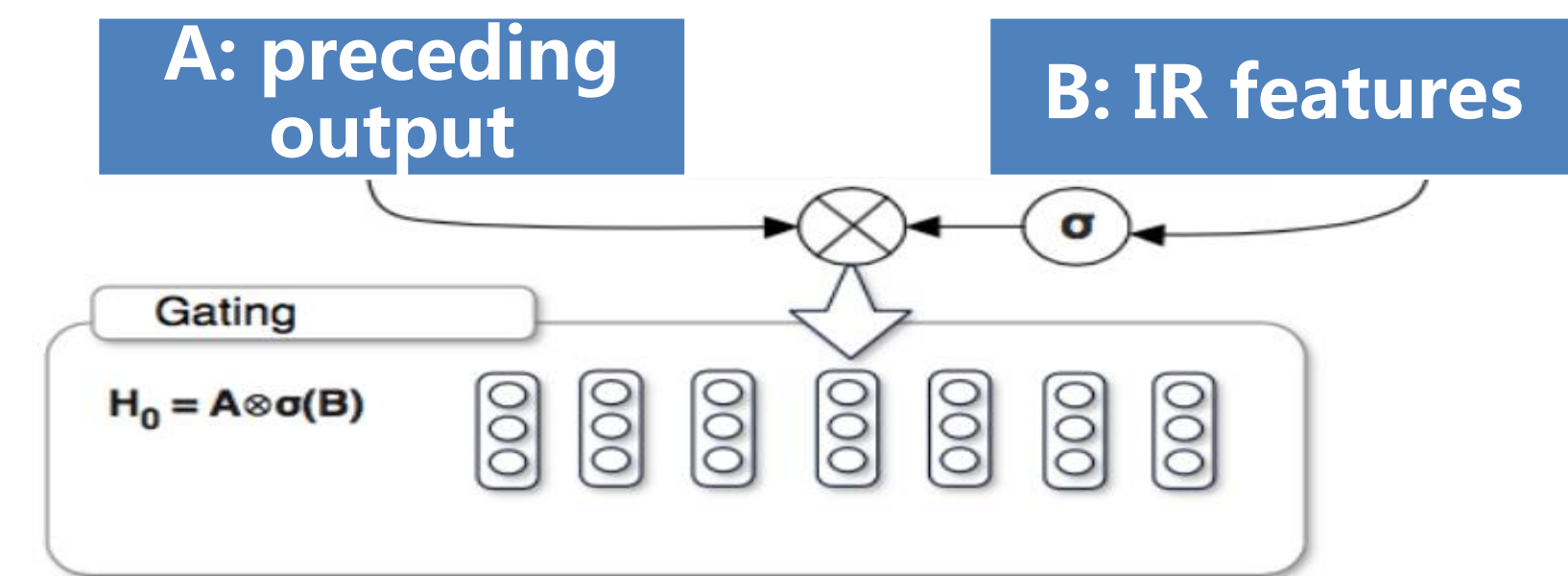
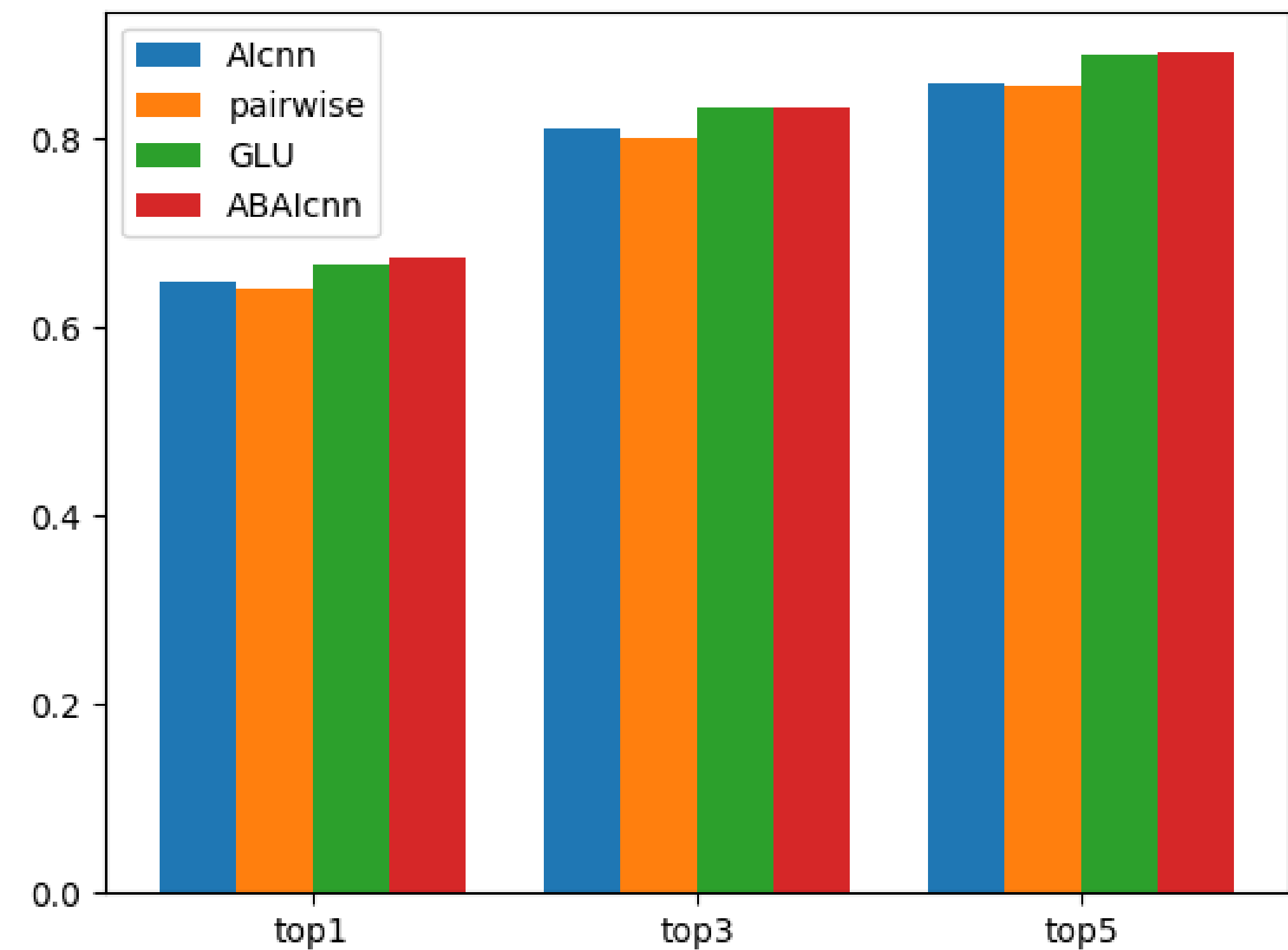
## 2 匹配打分



# 有监督精排：DeepRank



- 随机构造
- 随机构造+Lucene召回
- Lucene召回
- 交换句子, 打乱单词
- 相似词增强
- 交换句子, 打乱单词, 相似词增强
- 相似词增强
- 相似词标准化



# 知文问答v0.4：FAQ机器人的基础框架

用户Query

- 精度速度兼顾
- 无成本冷启动，无监督粗排兜底
- 一定的标注成本下，有监督精排能够进一步提升

如果没有FAQ呢？

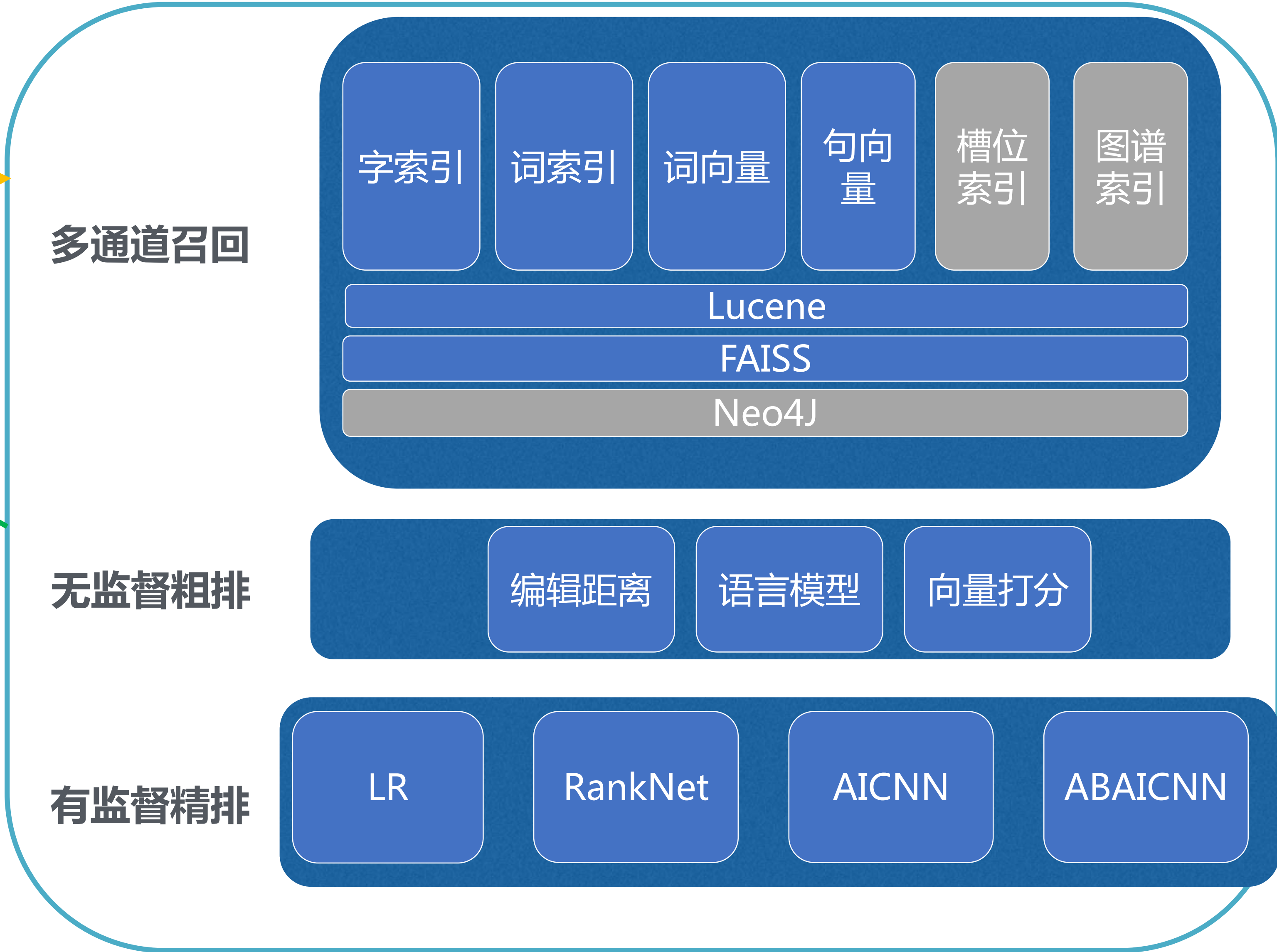




TABLE OF

# CONTENTS 大纲

---

- 知文问答：从FAQ机器人说起
- 用活存量文档：文档型机器人
- 走向结构化数据：知识图谱机器人
- 知文技能树：任务型机器人
- Query处理、多轮对话与知识管理
- 总结

# 知文问答v0.5：文档阅读理解

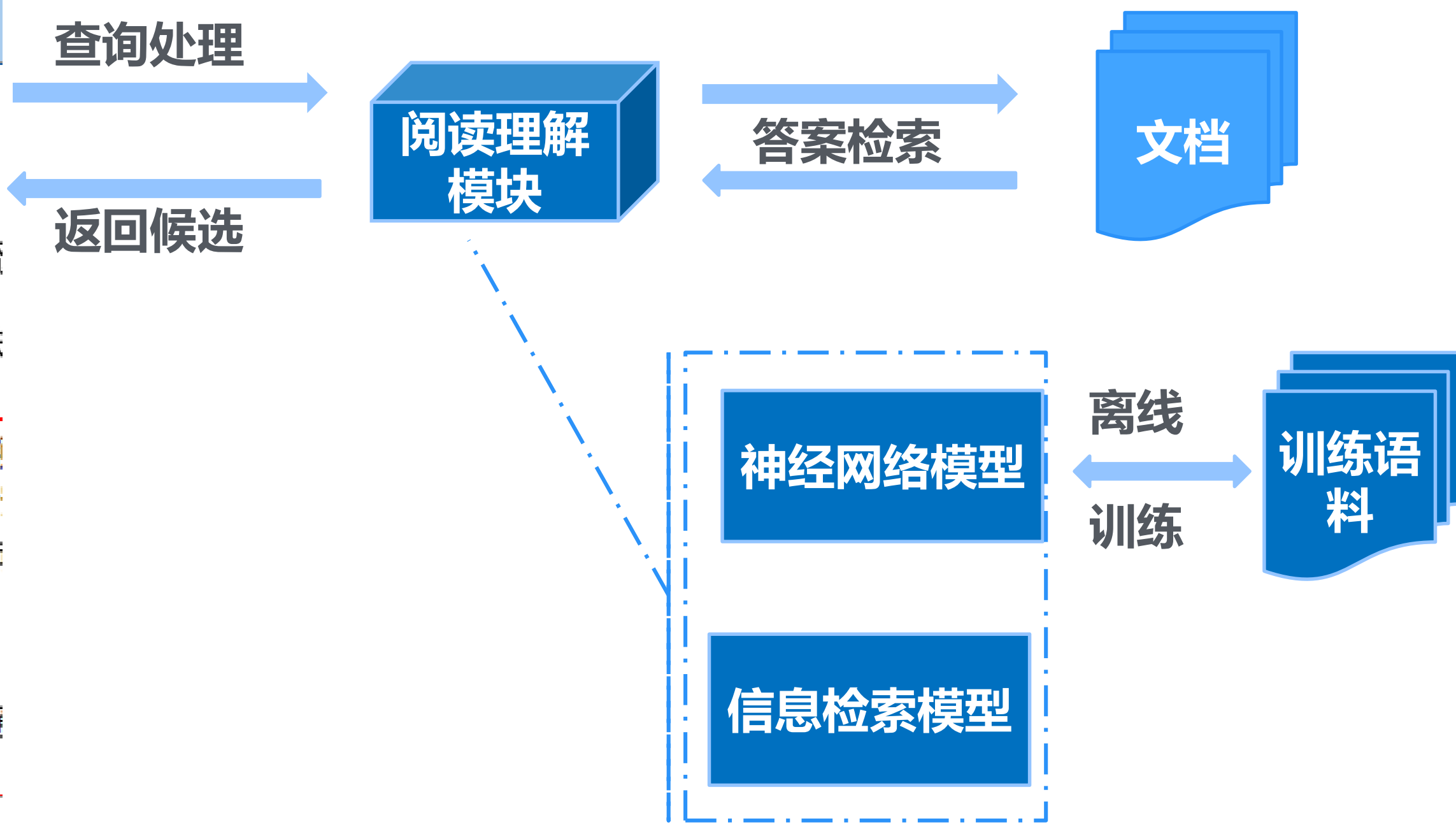
Question: 为什么养老宝买入后资产没有增加？

## 对相关文档做阅读理解

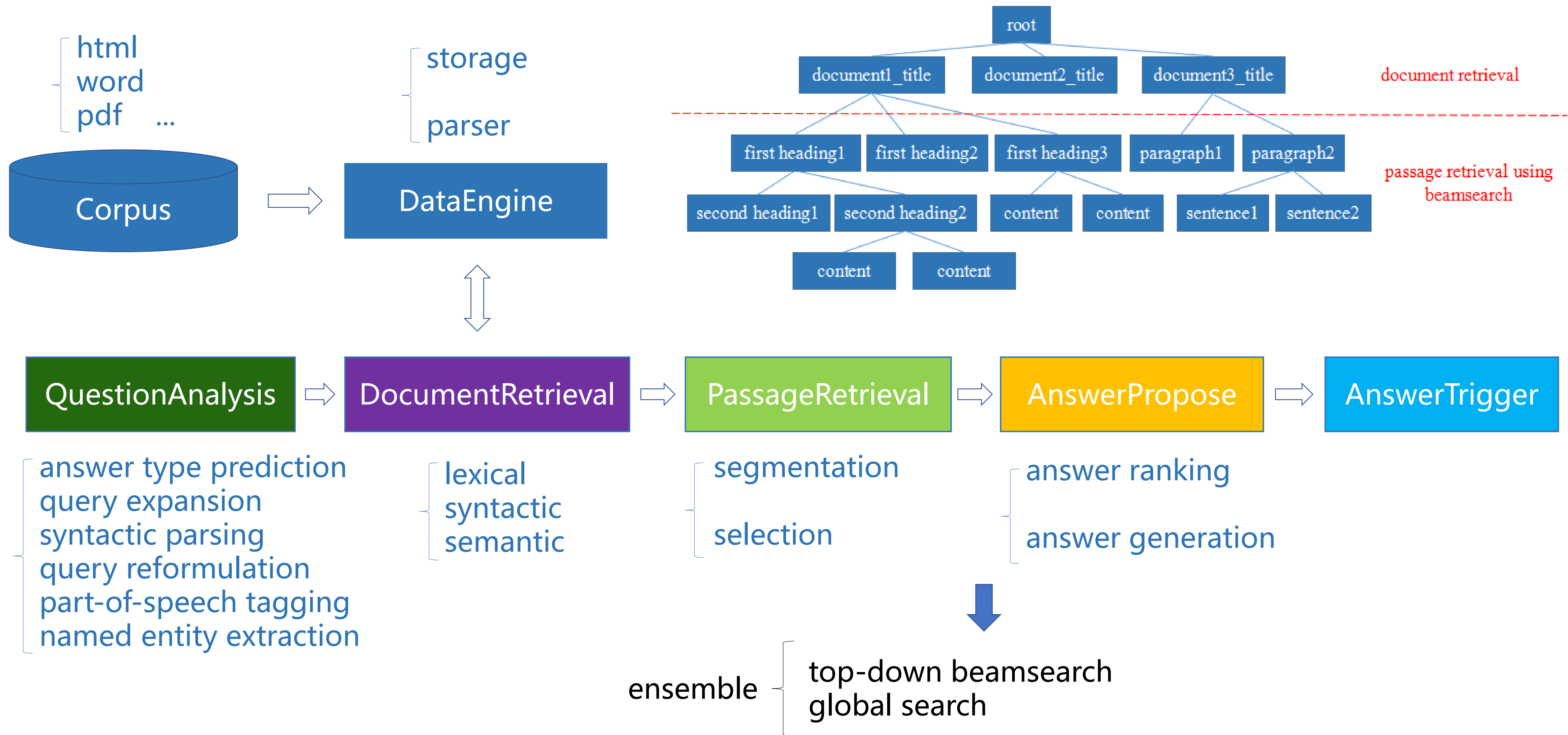
- (一) APP登录后资产无法正常显示怎么办？
- 1、网络原因：刷新页面，或安全退出后重新登录（登录→我的→应用设置→安全退出）。
  - 2、APP原因：尝试登录养老宝官网（e.boc.cn/yib）进行操作。
  - 3、未签约：可尝试用原交易的银行卡买入1元，资产即可正常显示。如果不是原银行卡，请勿买入，按换卡流宝变更客户信息》。
  - 4、如仍无法显示资产，请通过记录工单形式进行反馈。请提示客户我们将尽快排查问题，但客户资产不会
- (二) 买入后资产未增加怎么办？
- 1、网络问题：如银行卡扣款成功，但养老宝资产份额未增加，请点击：我的→查询，看看是否存在买入\*\*金额可能由于网络通讯超时所致，请您放心，实际已买入成功，您可安全退出后重新登录查看（我的→应用设置→
  - 2、开户失败：如果是首次购买，银行卡扣款成功，但养老宝资产份额未增加，也查不到买入成功的记录，有可作日退款。
  - 3、自动买入：自动买入扣款后，不是实时增加资产，资产一般在中午前显示（特殊情况下有延时）。
  - 4、其他原因：少数情况下，由于基金公司划款排队原因，可能出现延时，具体以基金公司为准。其他情况请通我们将尽快排查问题。

Answer: 买入后资产未增加怎么办？

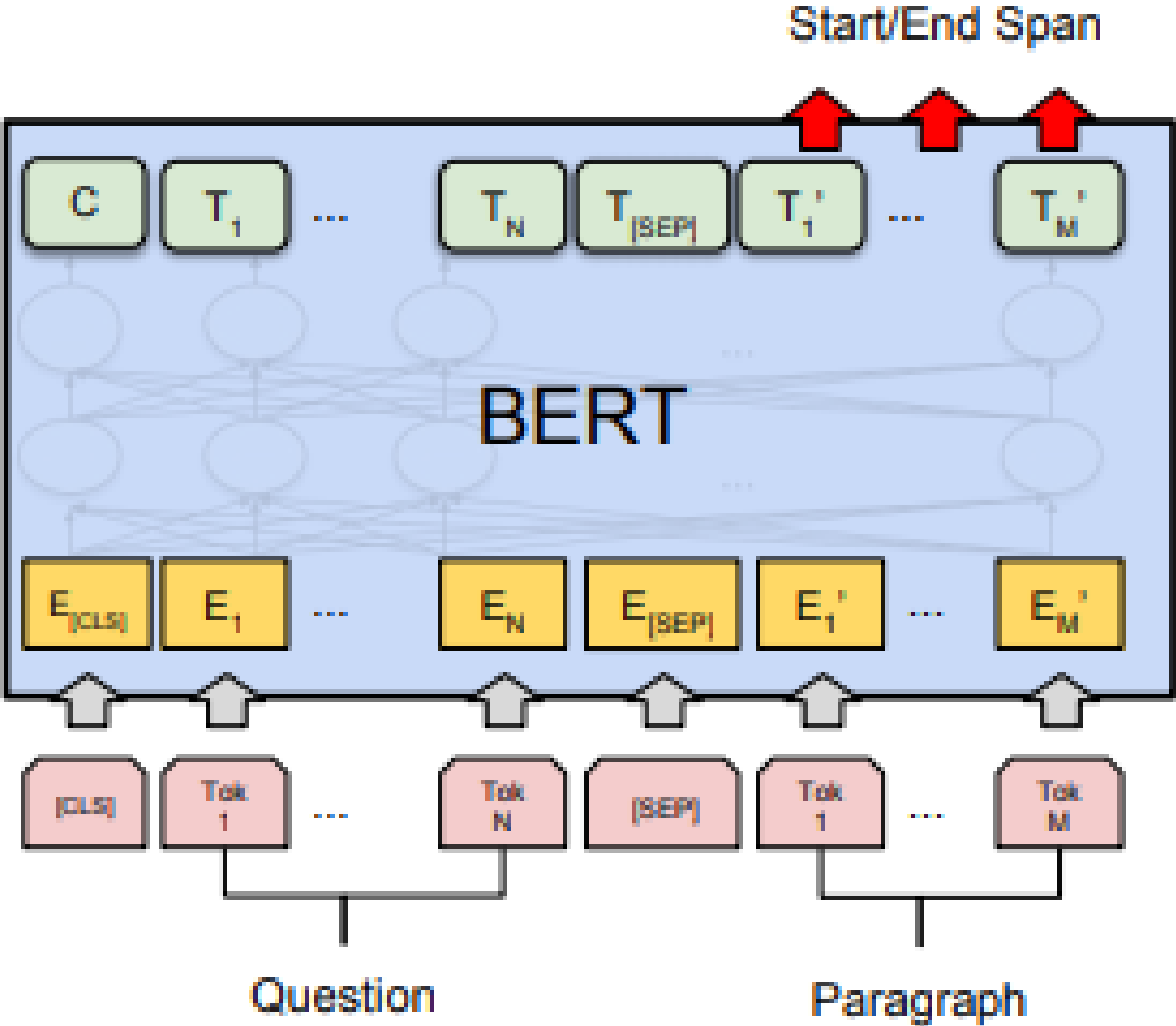
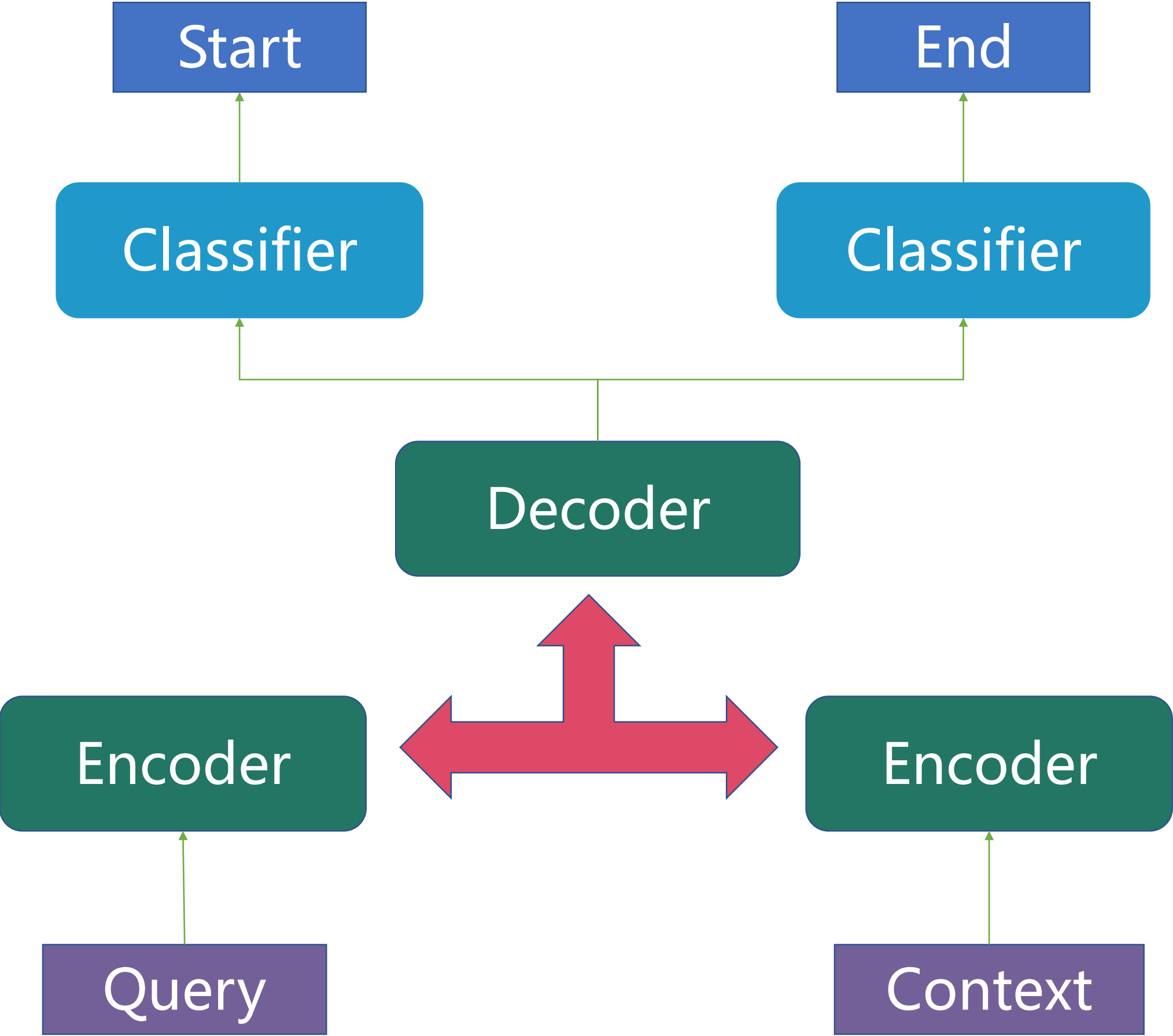
- 1、网络问题：如银行卡扣款成功，但是养老宝资产份额未增加，请点击
- 2、开户失败：如果是首次购买，银行卡扣款成功，但...
- 3、自动买入：自动买入扣款后，不是实时增加资产...
- 4、其他原因：少数情况下，由于基金公司划款排队原因，...



# 文档阅读理解：pipeline

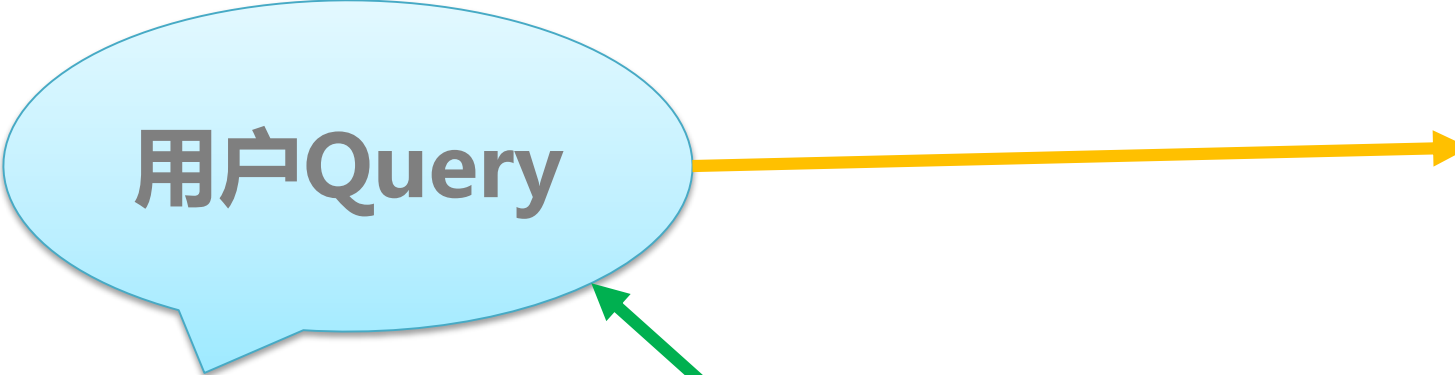


# 文档阅读理解：end-to-end





# 知文问答v0.5 : FAQBot + DocBot



- 支持半结构化FAQ数据
- 支持非结构化存量文档

如果是更为结构化的数据呢？

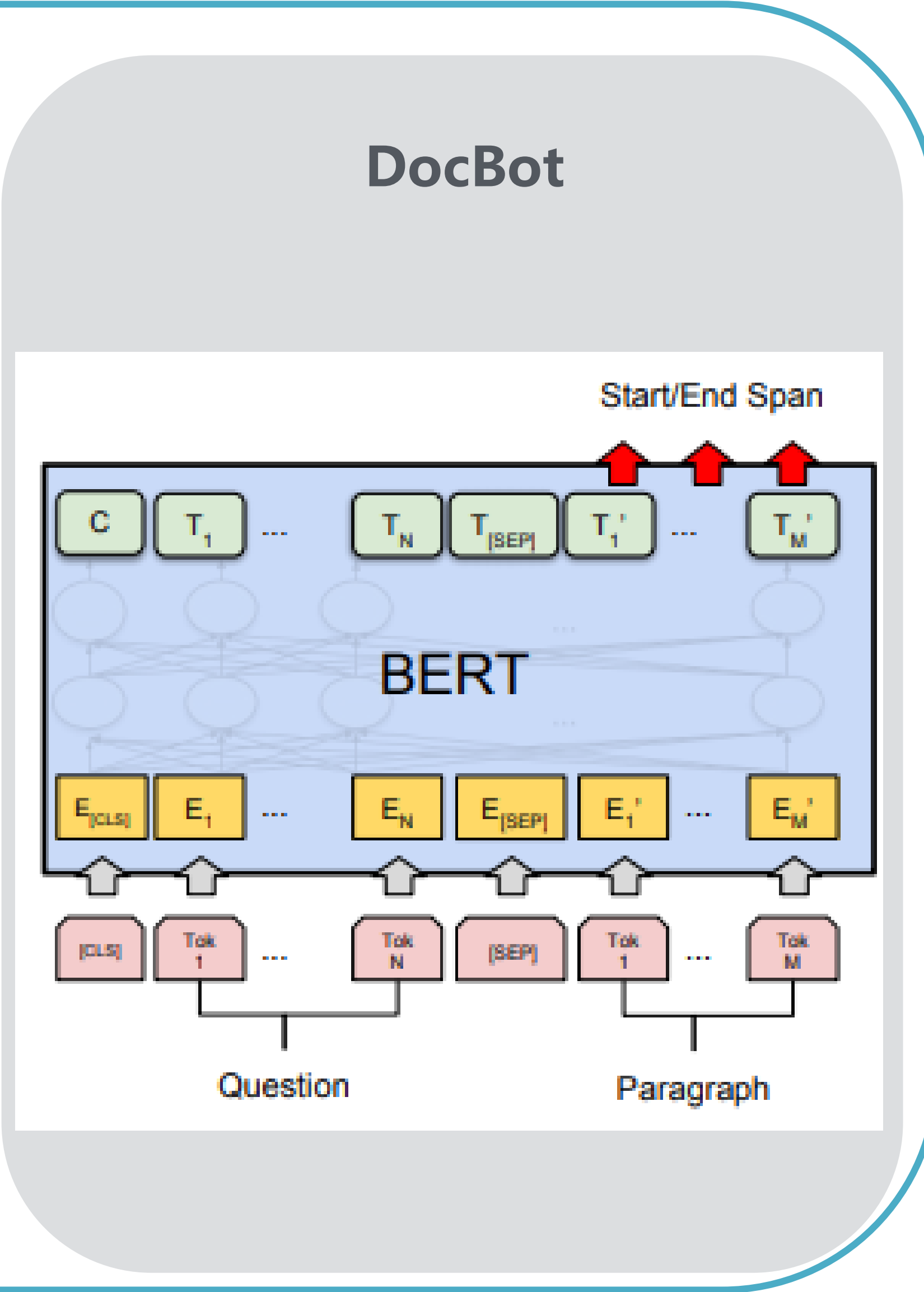
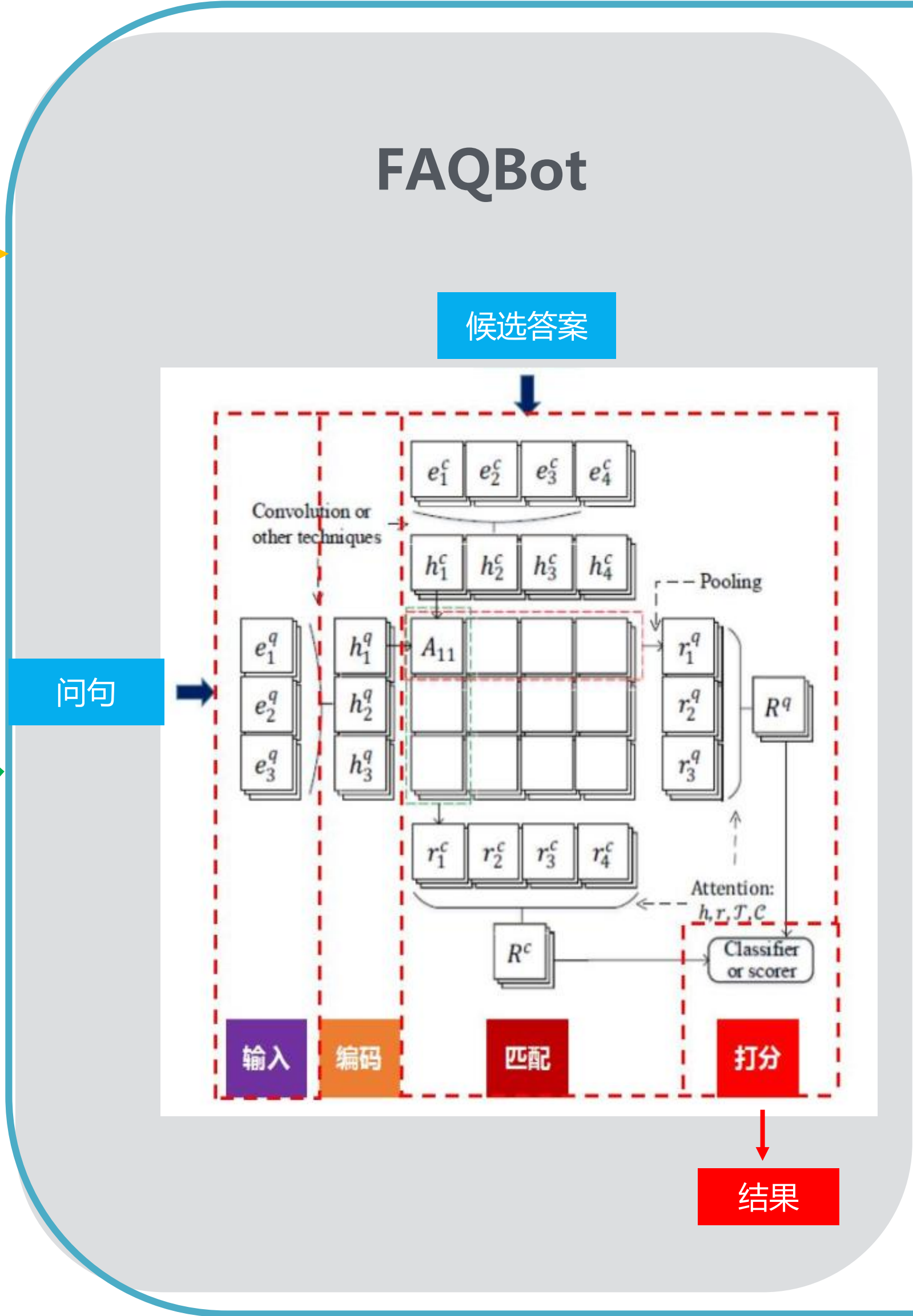


TABLE OF

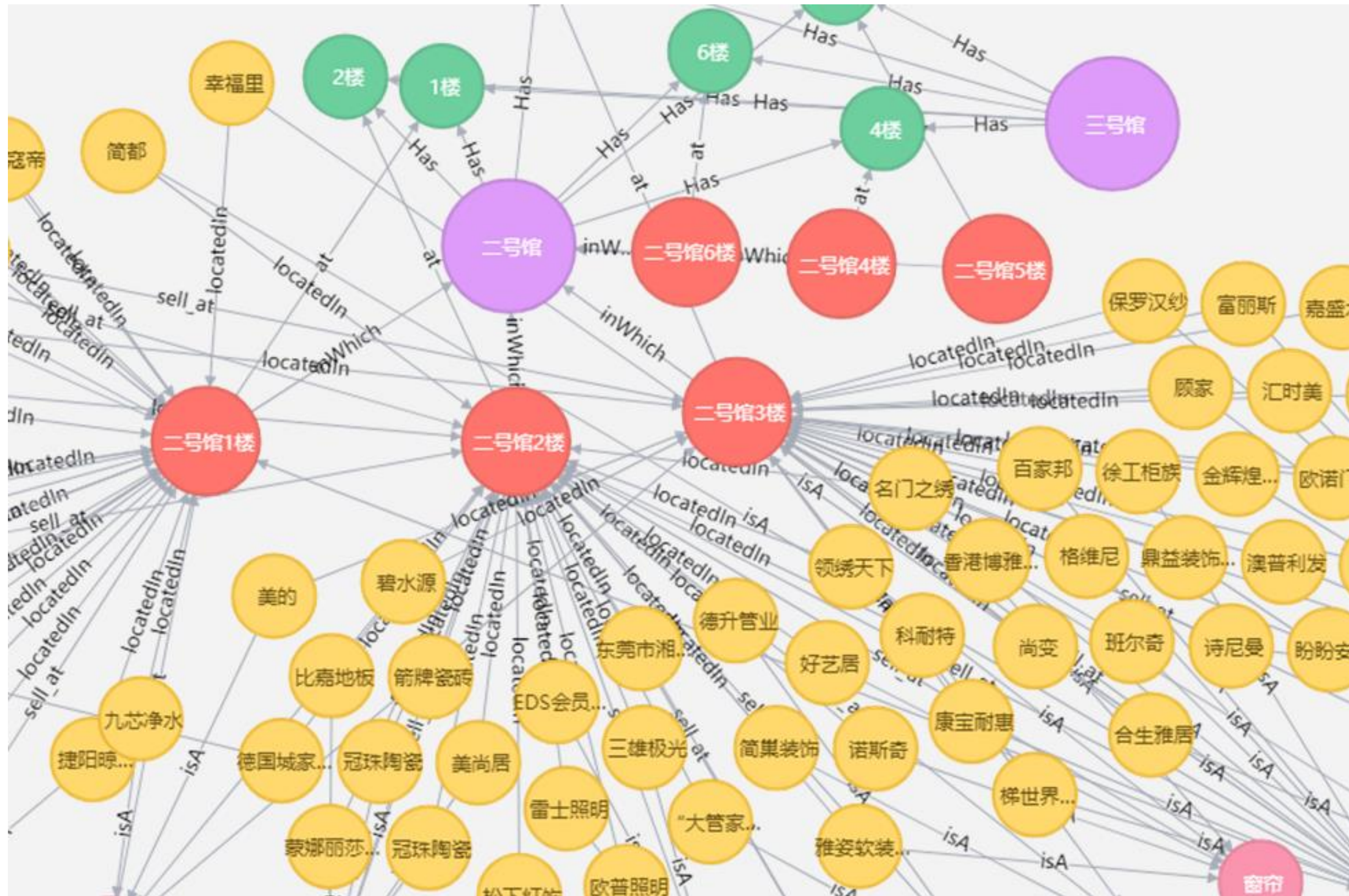
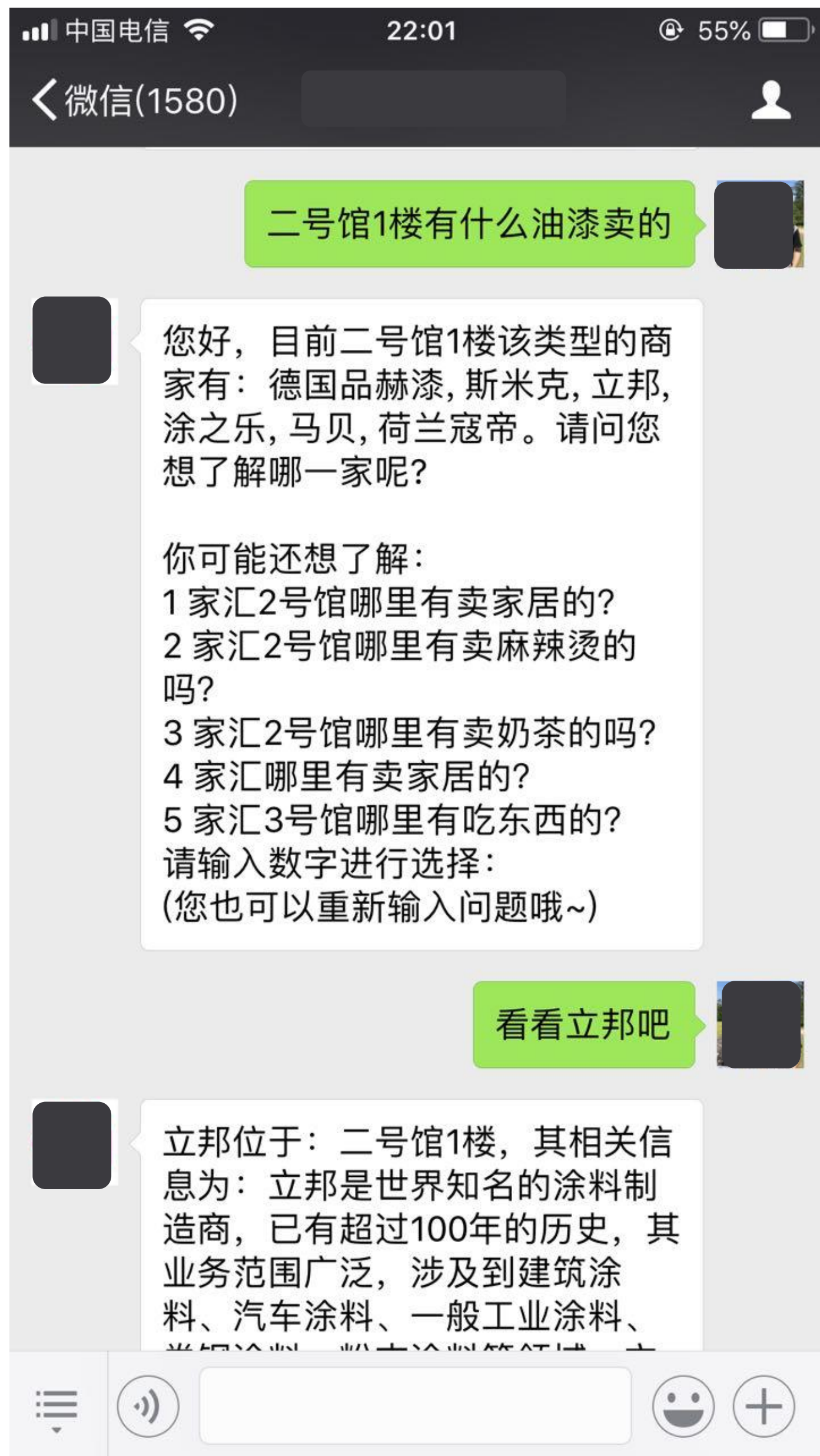
# CONTENTS 大纲

---

- 知文问答：从FAQ机器人说起
- 用活存量文档：文档型机器人
- 走向结构化数据：知识图谱机器人
- 知文技能树：任务型机器人
- Query处理、多轮对话与知识管理
- 总结

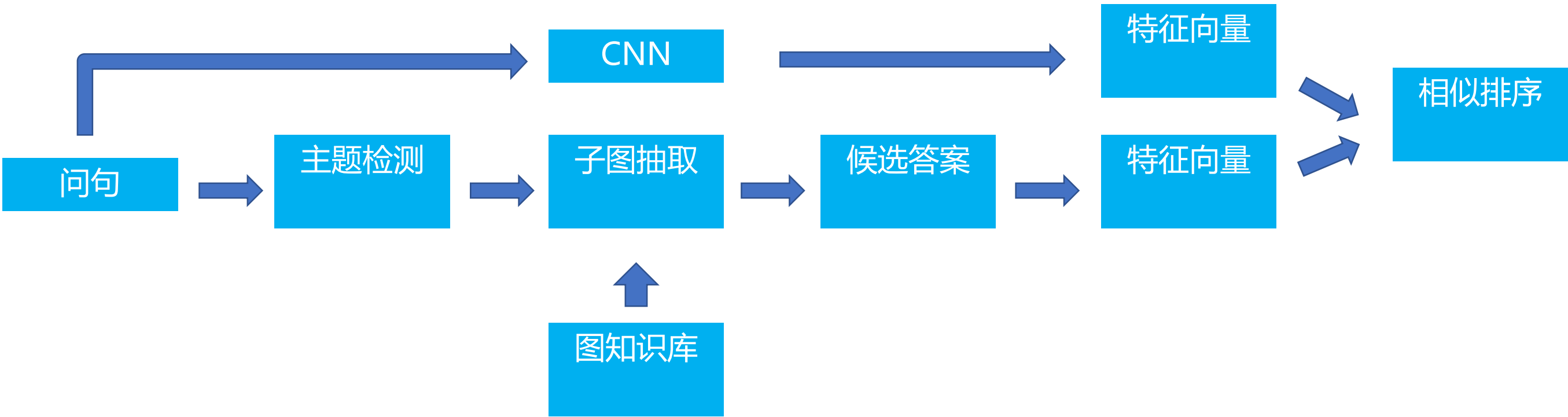
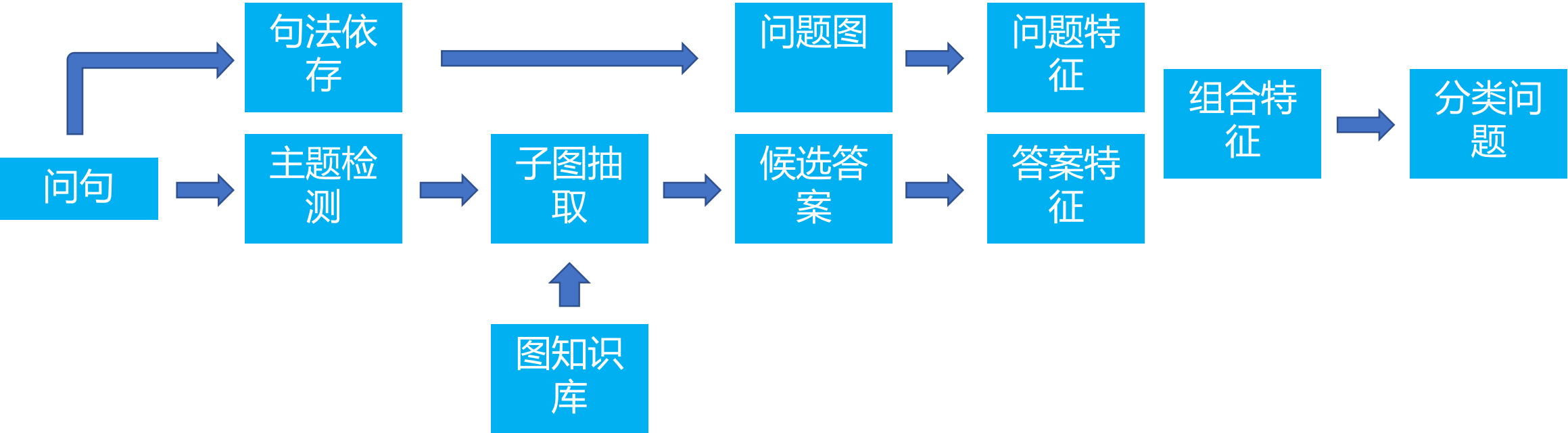
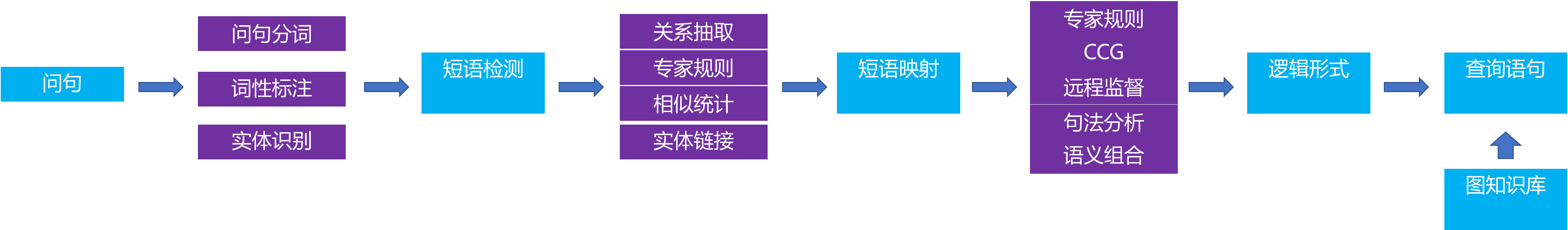


# 知文问答v0.6：知识图谱机器人



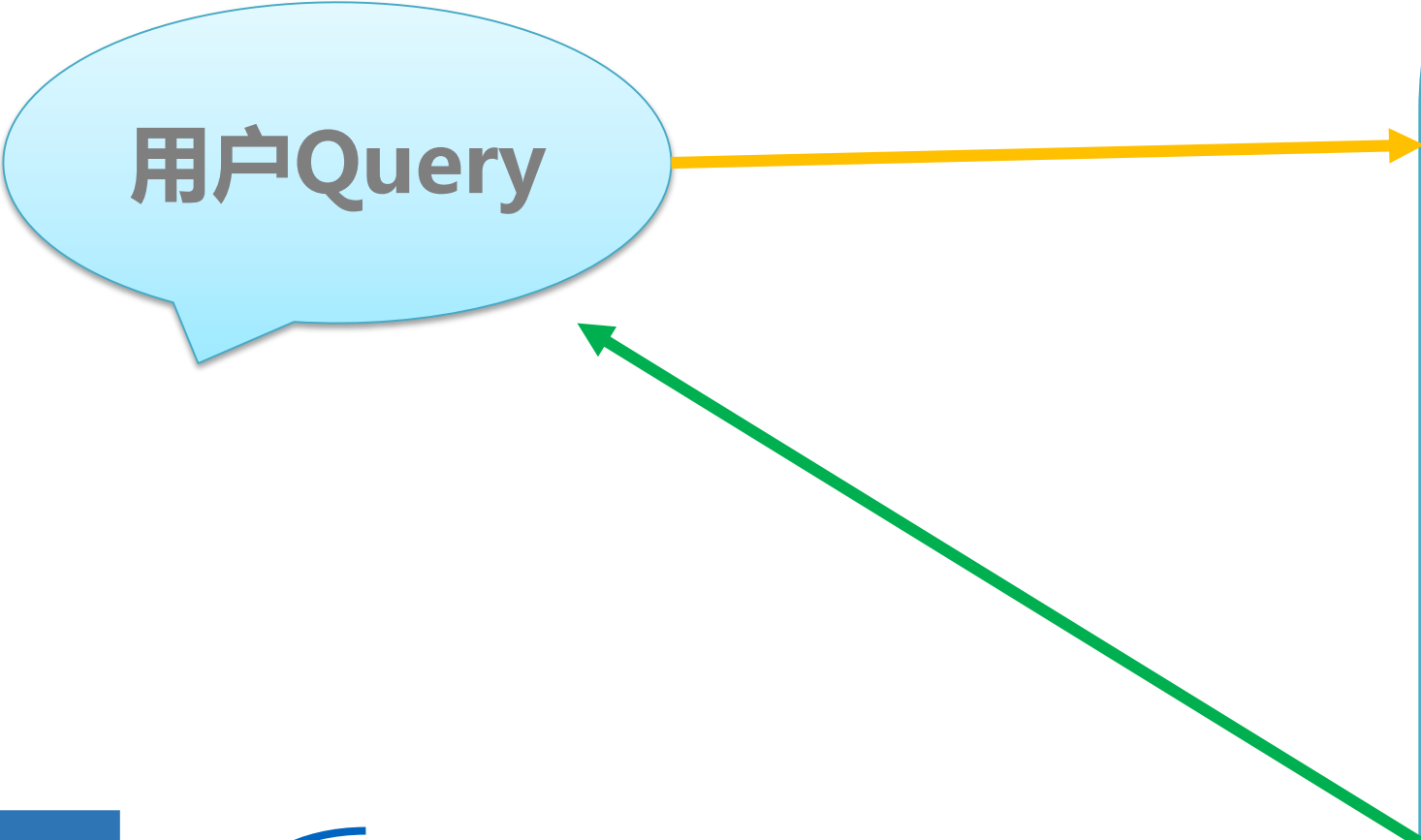


# 知识图谱机器人





# 知文问答v0.6 : FAQBot + DocBot + KGBot



信息获取

- 支持半结构化FAQ数据
- 支持非结构化存量文档
- 支持结构化知识图谱

如果执行指令或者操作？

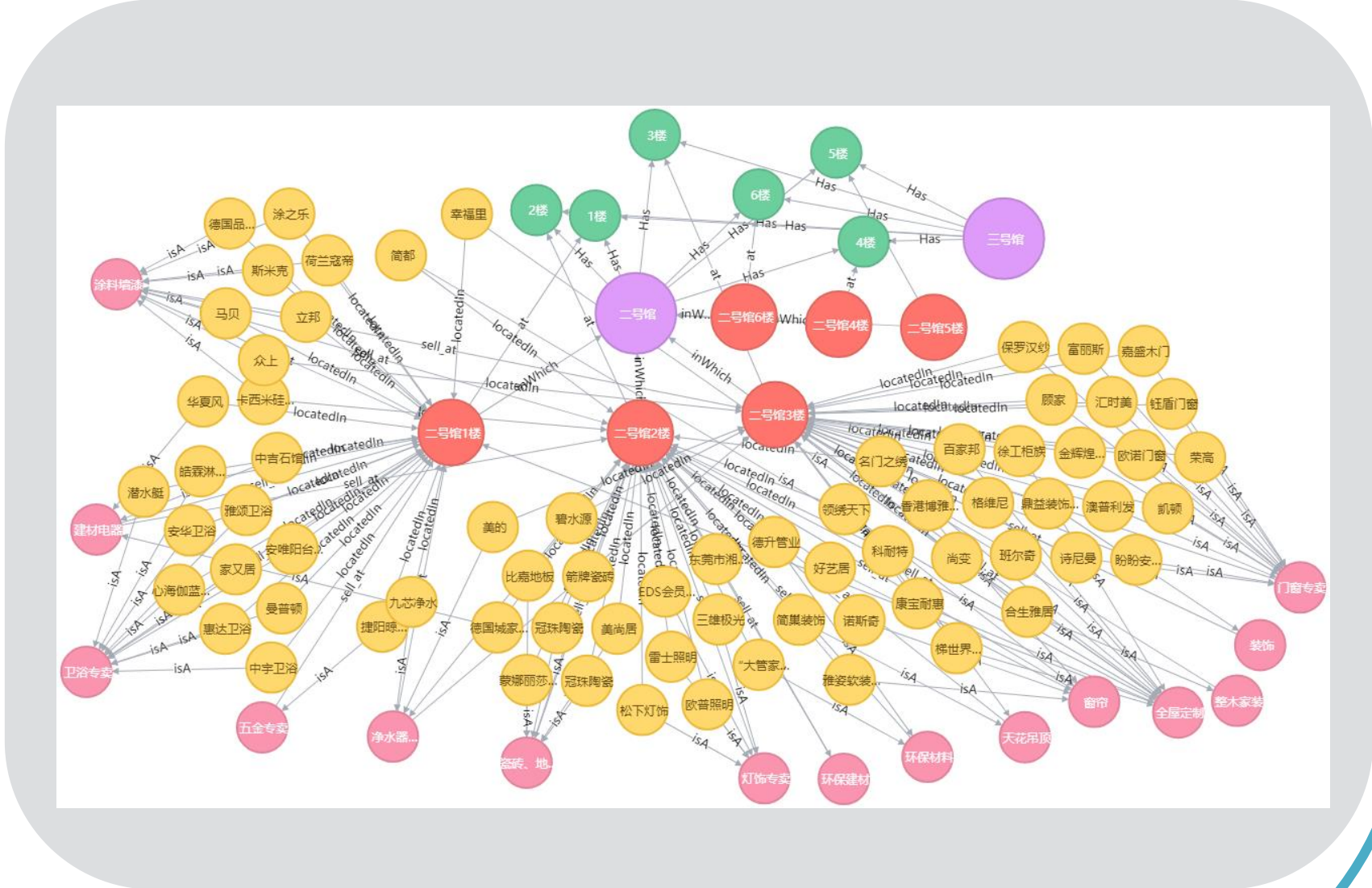
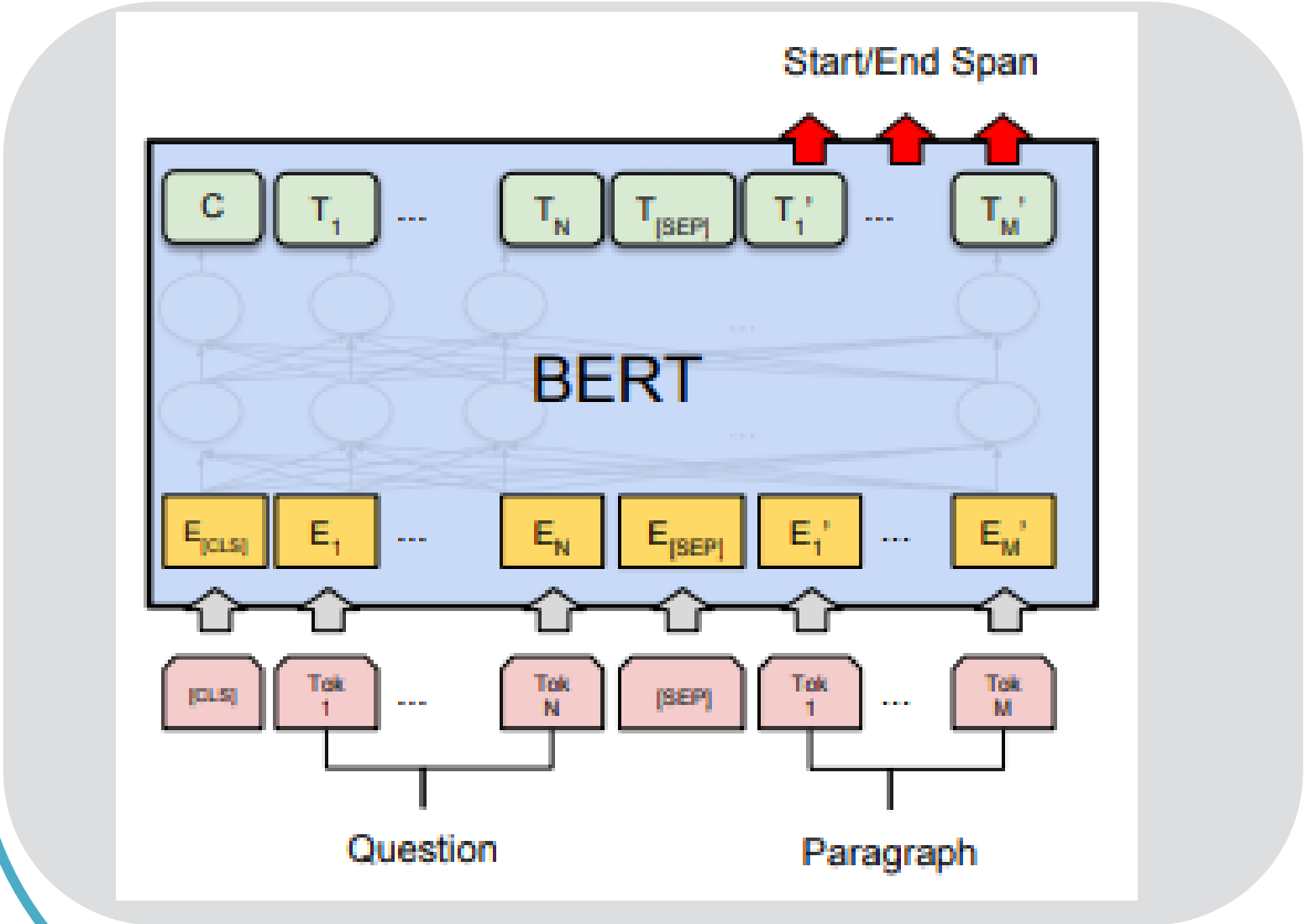
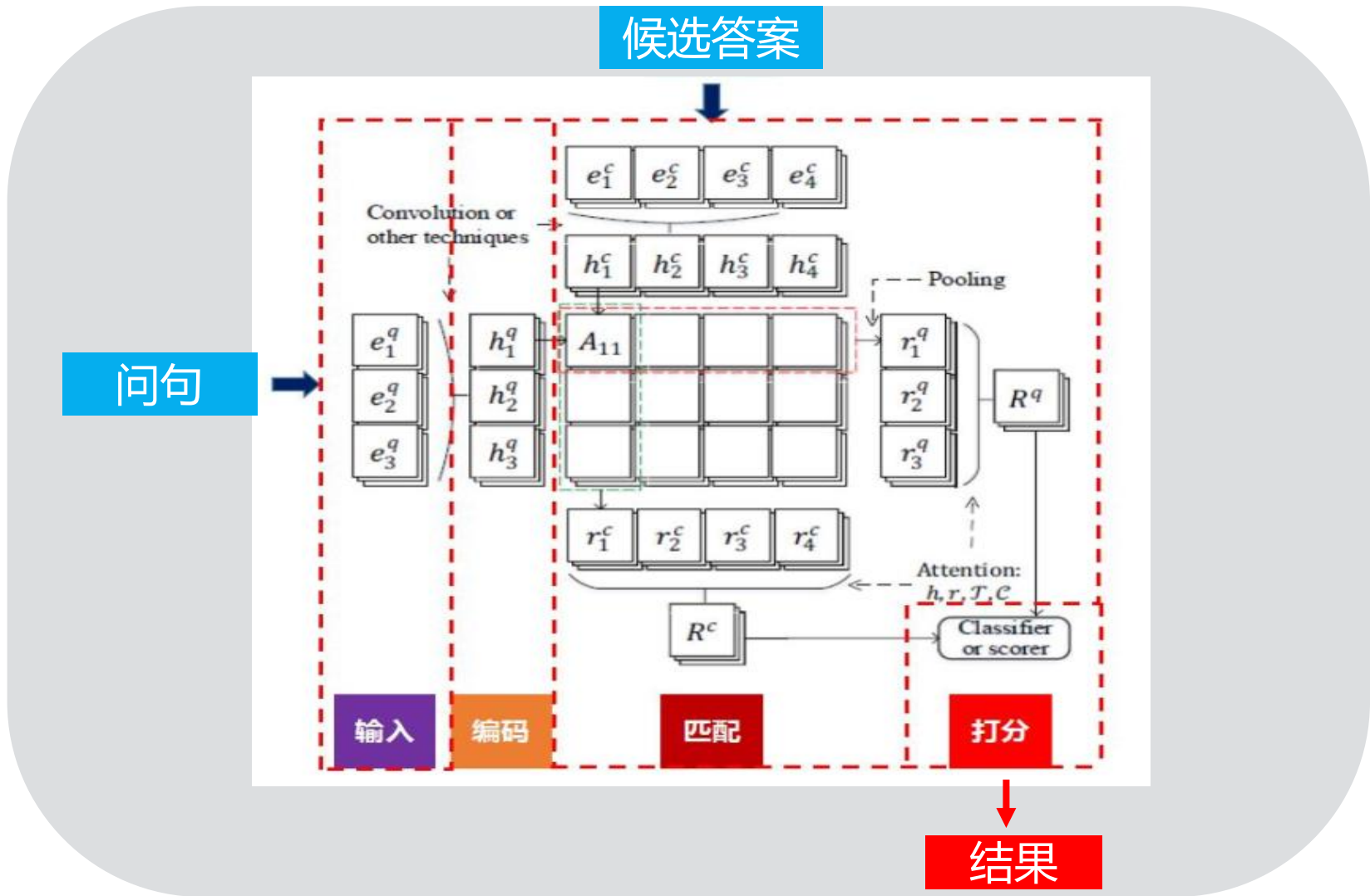


TABLE OF

# CONTENTS 大纲

---

- 知文问答：从FAQ机器人说起
- 用活存量文档：文档型机器人
- 走向结构化数据：知识图谱机器人
- 知文技能树：任务型机器人
- Query处理、多轮对话与知识管理
- 总结



# 知文问答v0.7：任务型机器人（pipeline）



云南有什么好玩的城市？

小云给您推荐了热门城市：



丽江市  
代表景点： 丽江东...



普洱市  
代表景点： 普洱国...



西双版纳州  
代表景点： 曼听御...

丽江推荐几个酒店吧

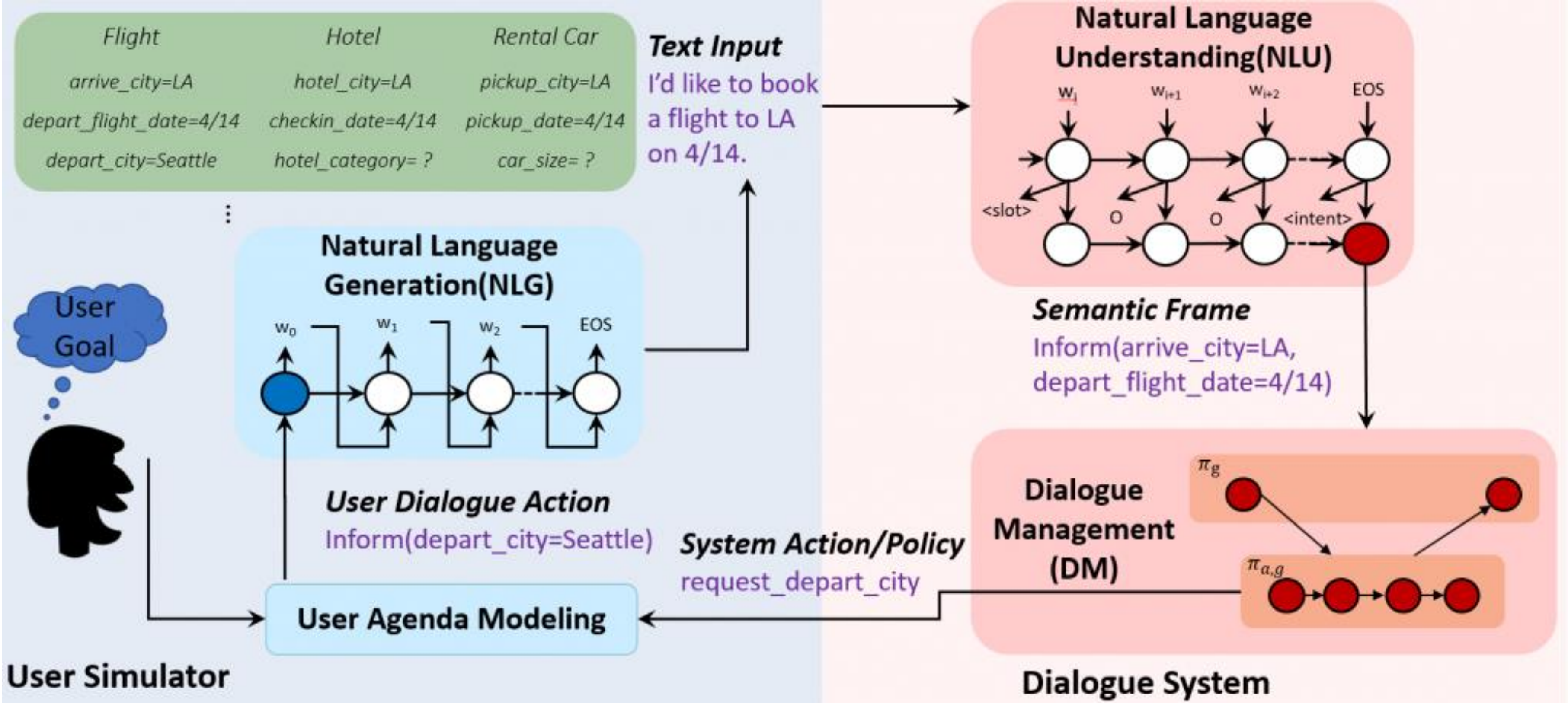
热门城市

热门景点

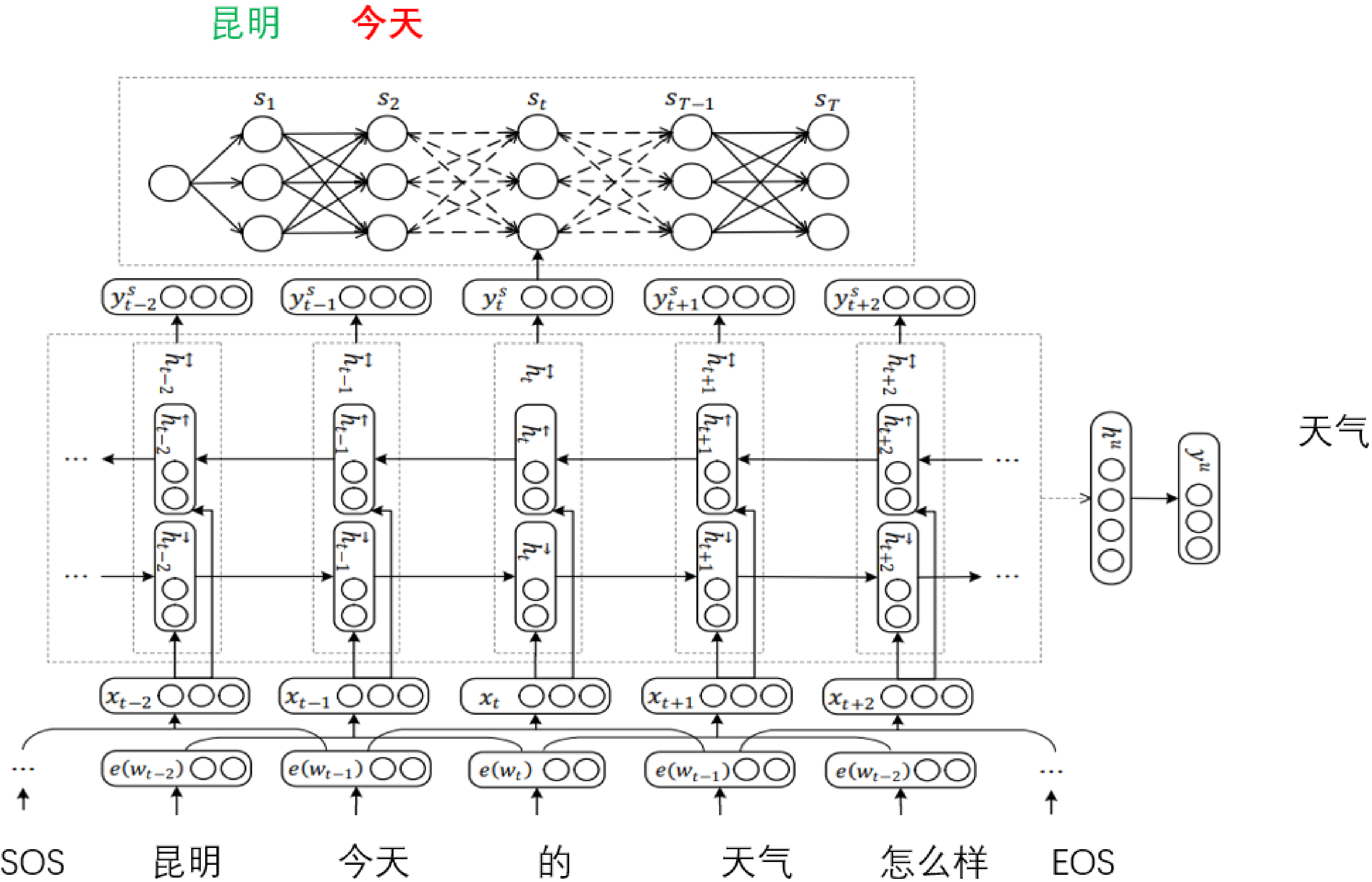
住酒店

吃东西

输入任何你想问的问题...

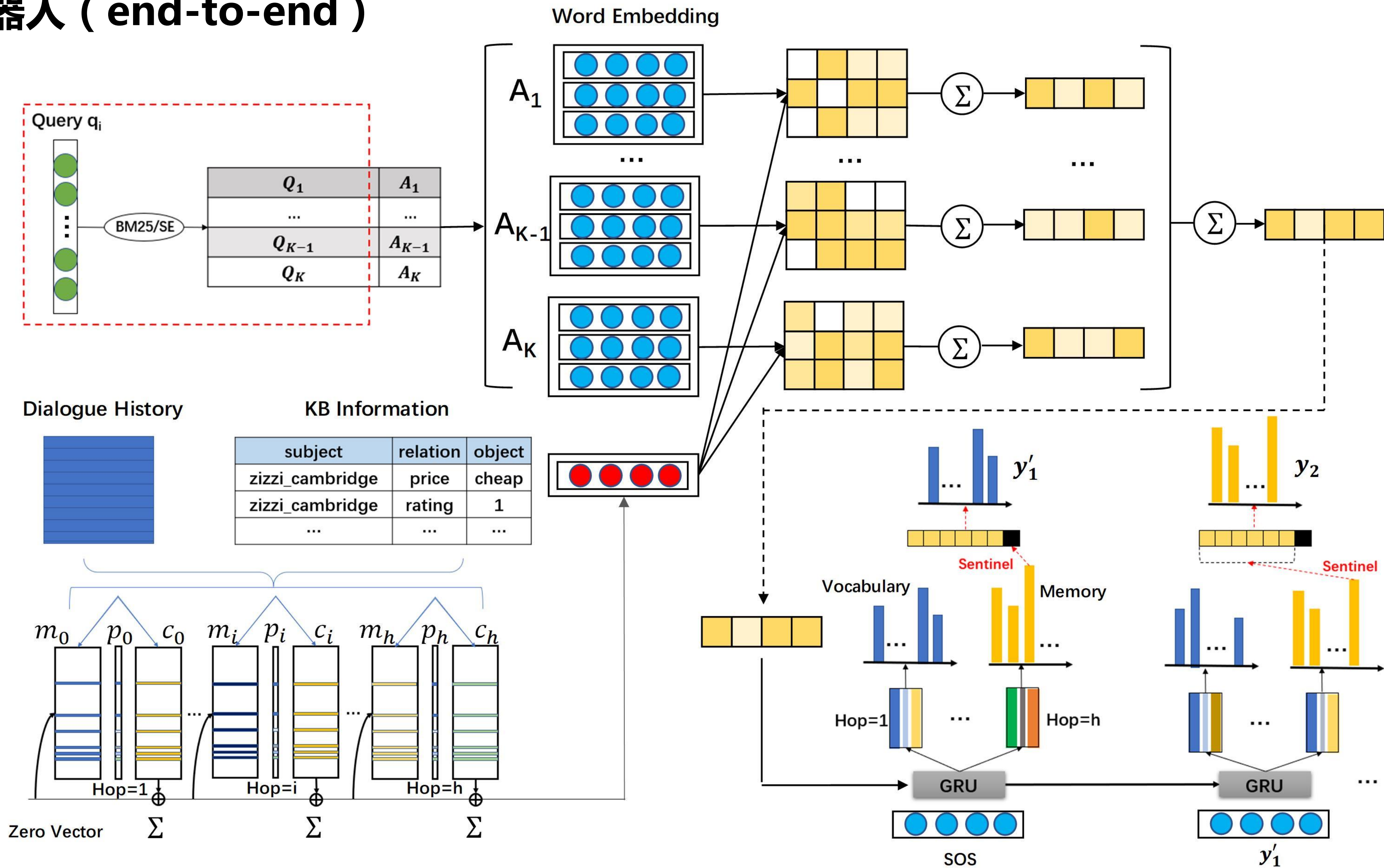


# 任务型机器人 ( pipeline )





# 任务型机器人 ( end-to-end )





# 知文问答v0.7 : FAQBot + DocBot + KBBot+TaskBot

信息获取

任务执行

用户Query

- 半结构化FAQ数据
- 非结构化存量文档
- 结构化知识图谱
- 自定义指令和技能

多bot如何耦合？

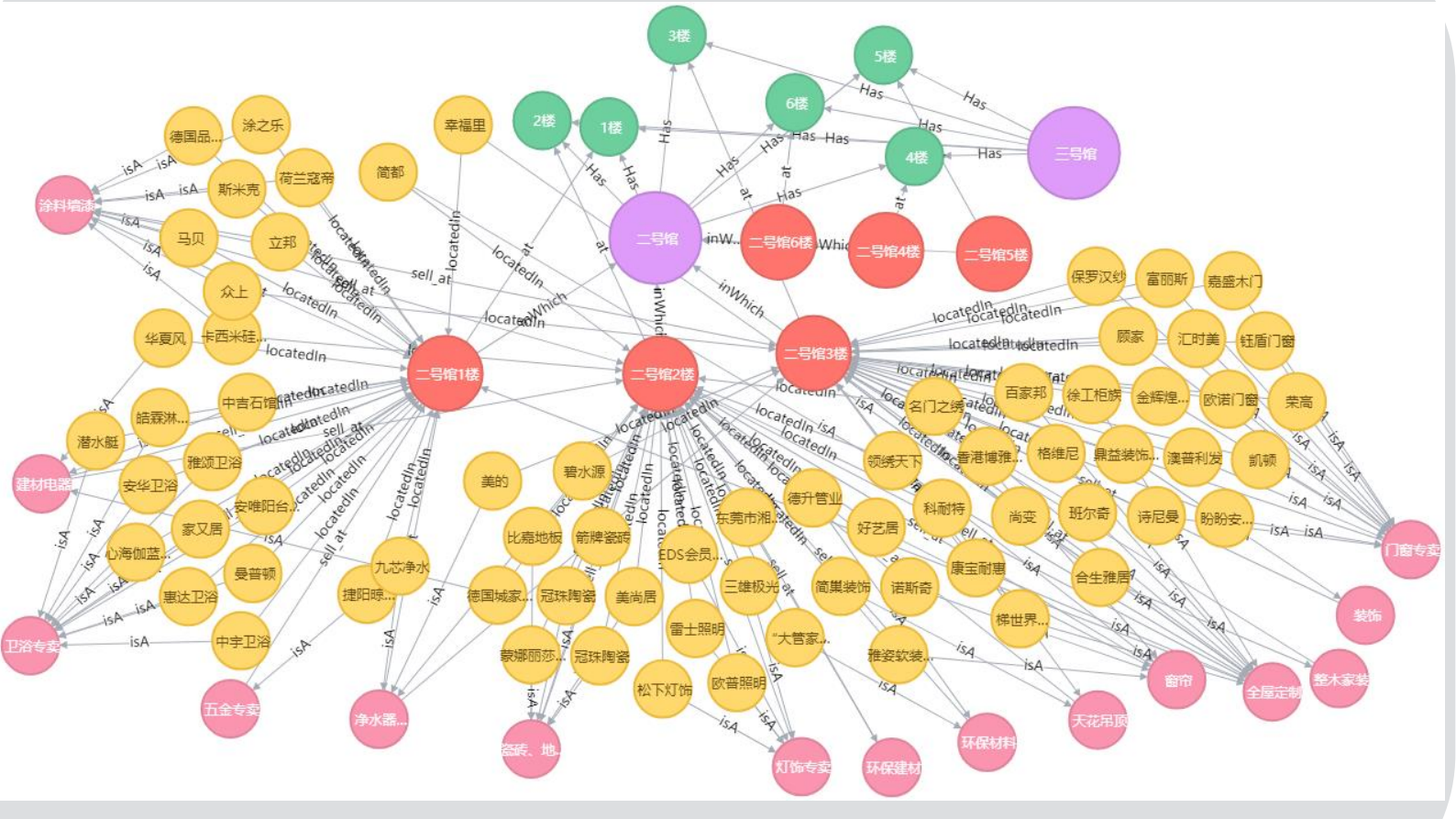
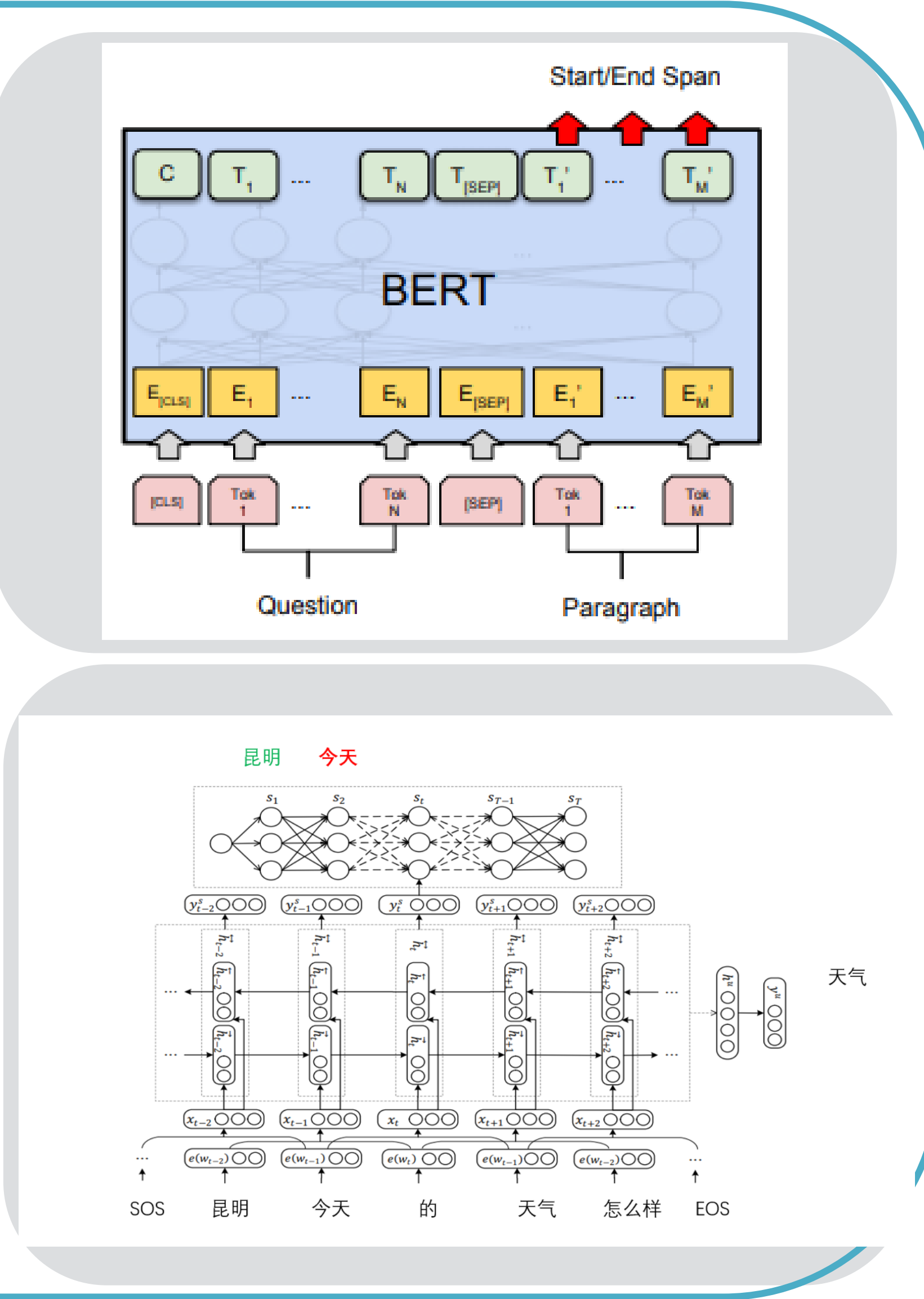
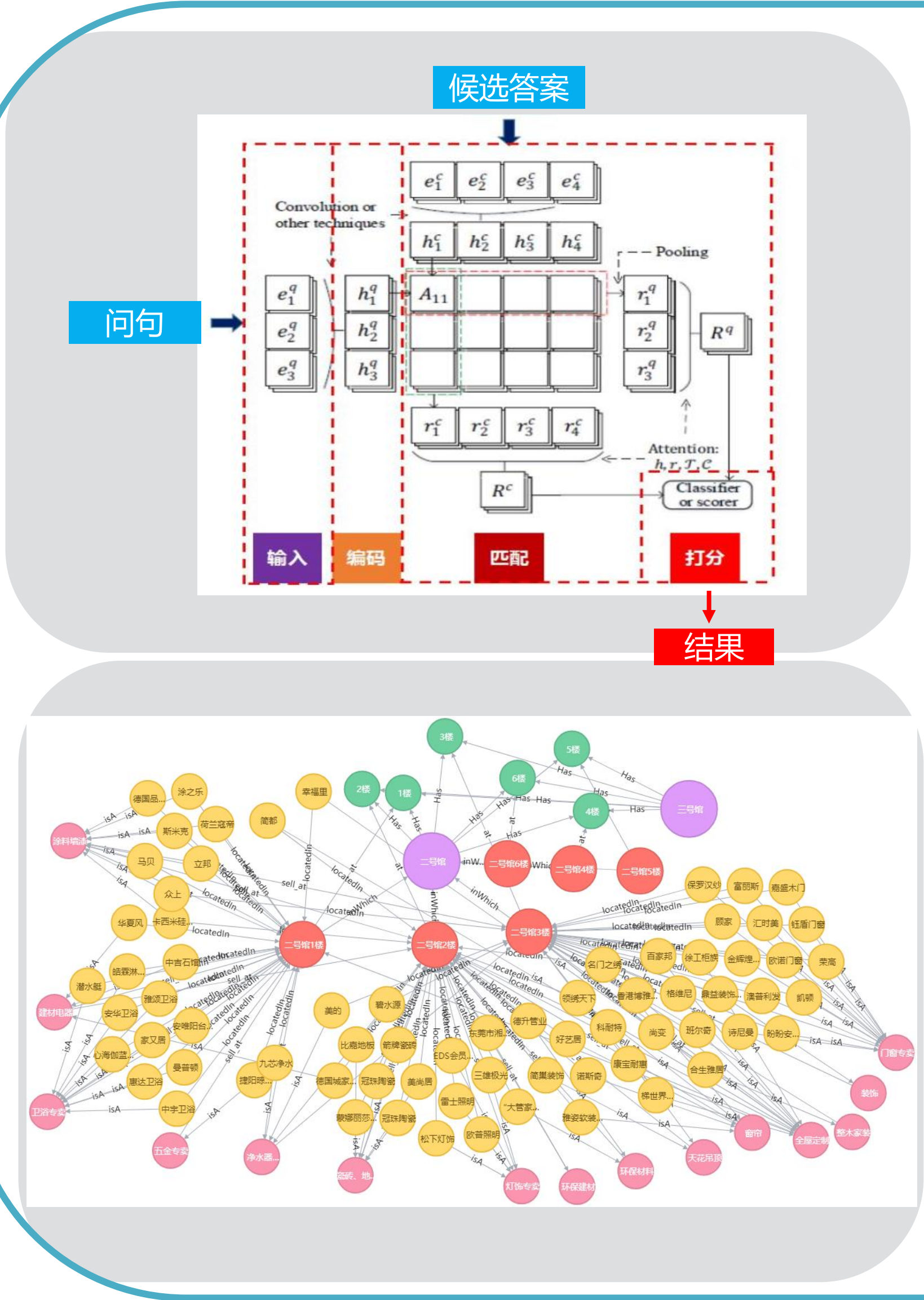


TABLE OF

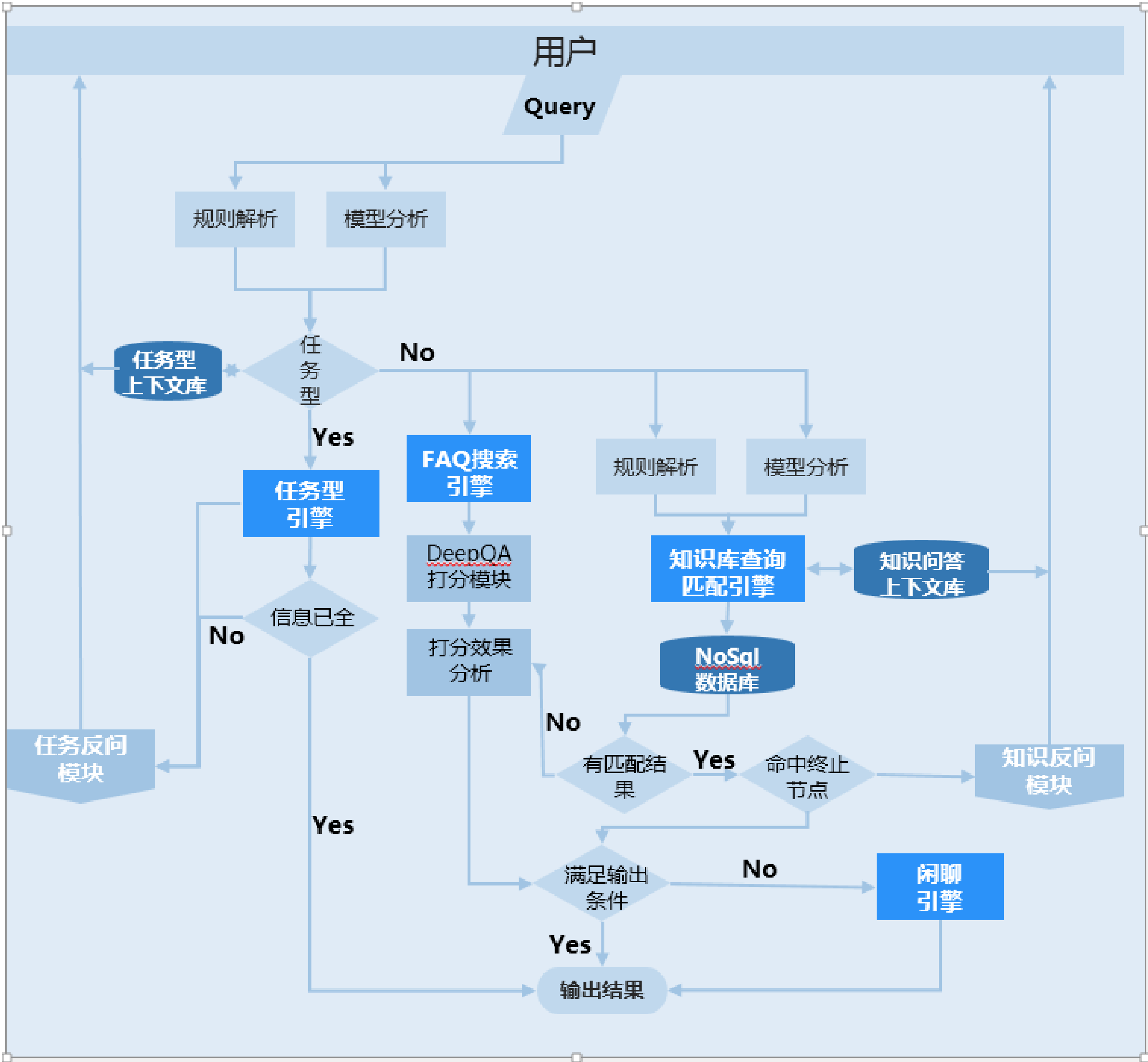
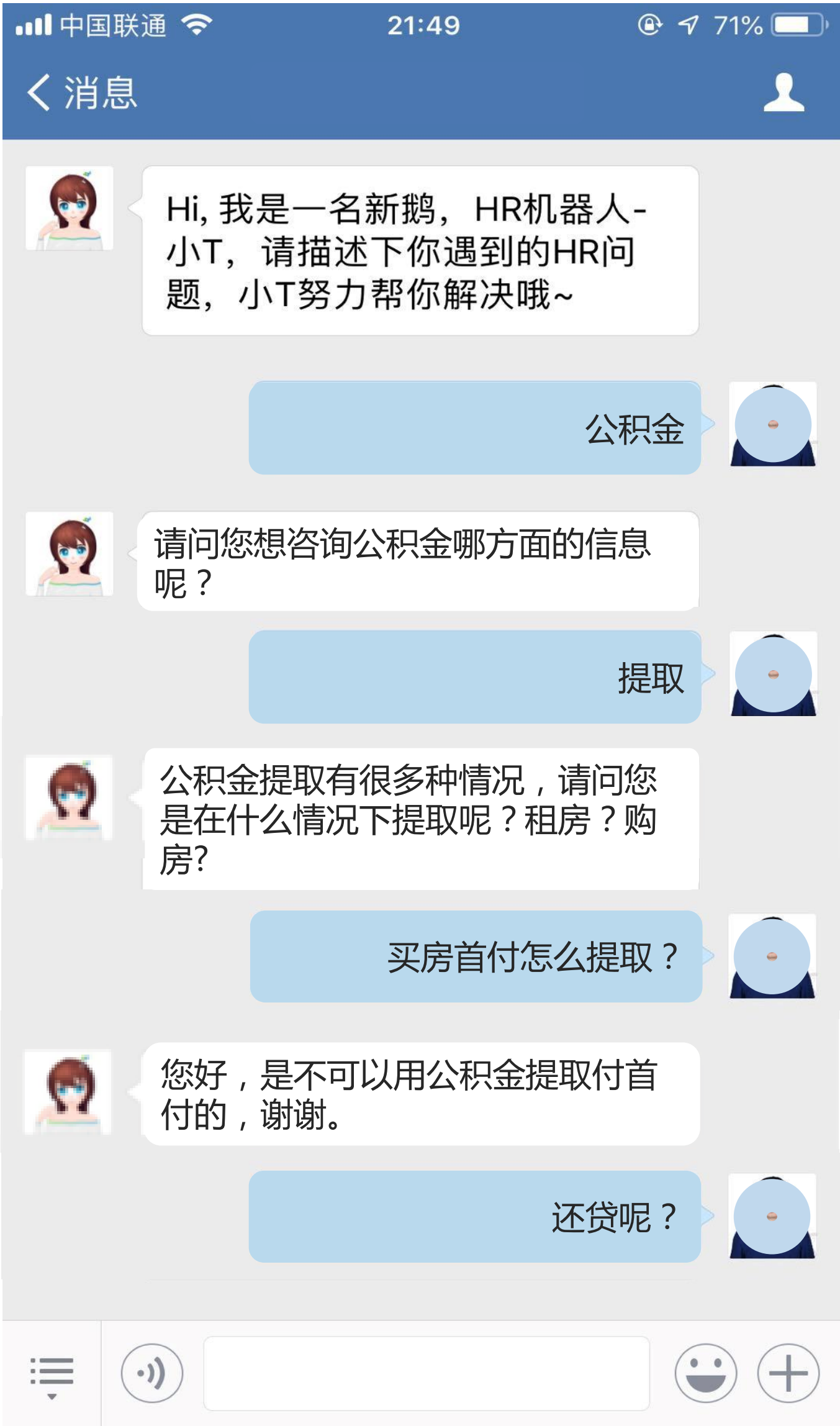
# CONTENTS 大纲

---

- 知文问答：从FAQ机器人说起
- 用活存量文档：文档型机器人
- 走向结构化数据：知识图谱机器人
- 知文技能树：任务型机器人
- 对话管理、Query处理与知识库管理
- 总结

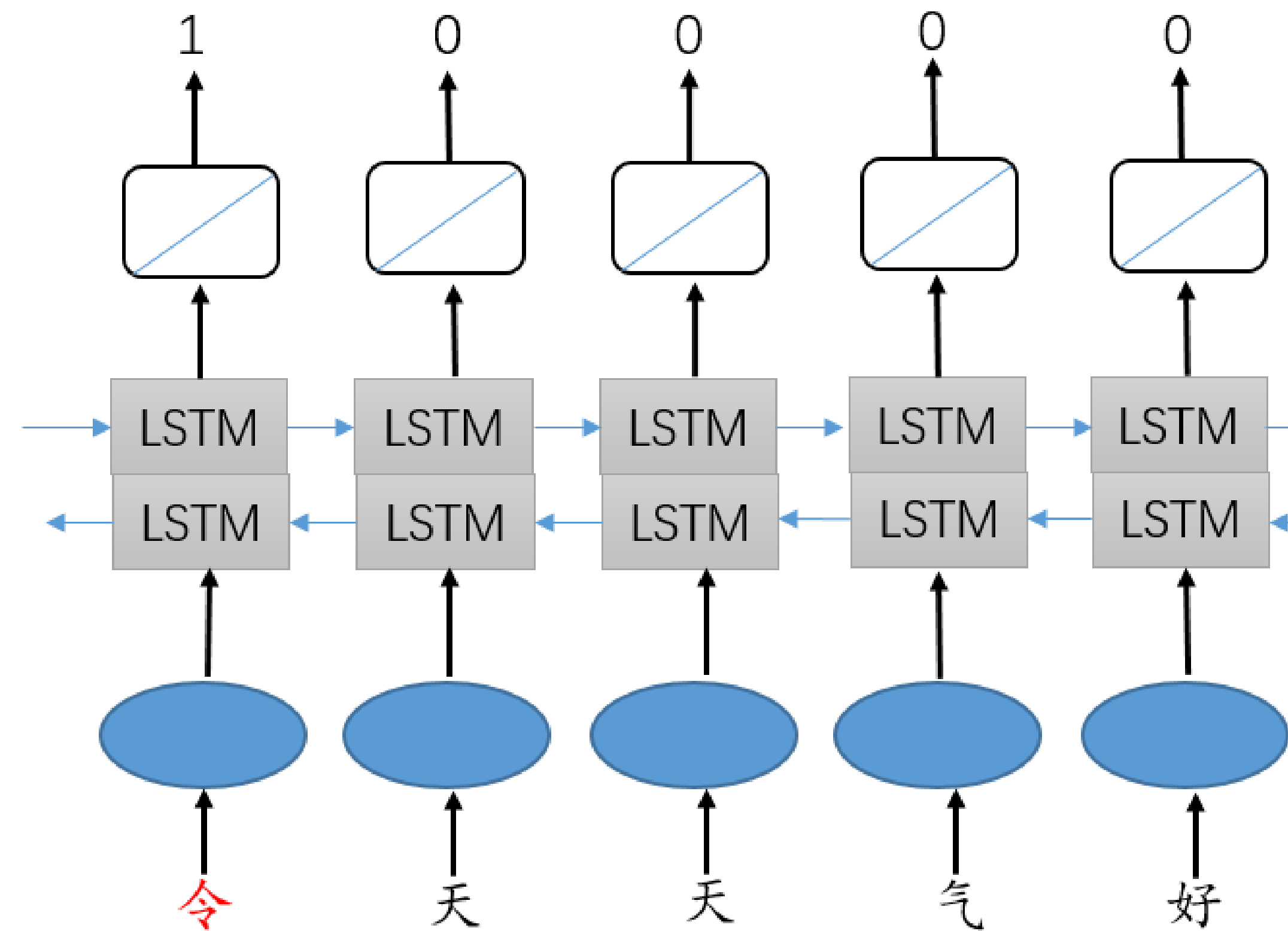
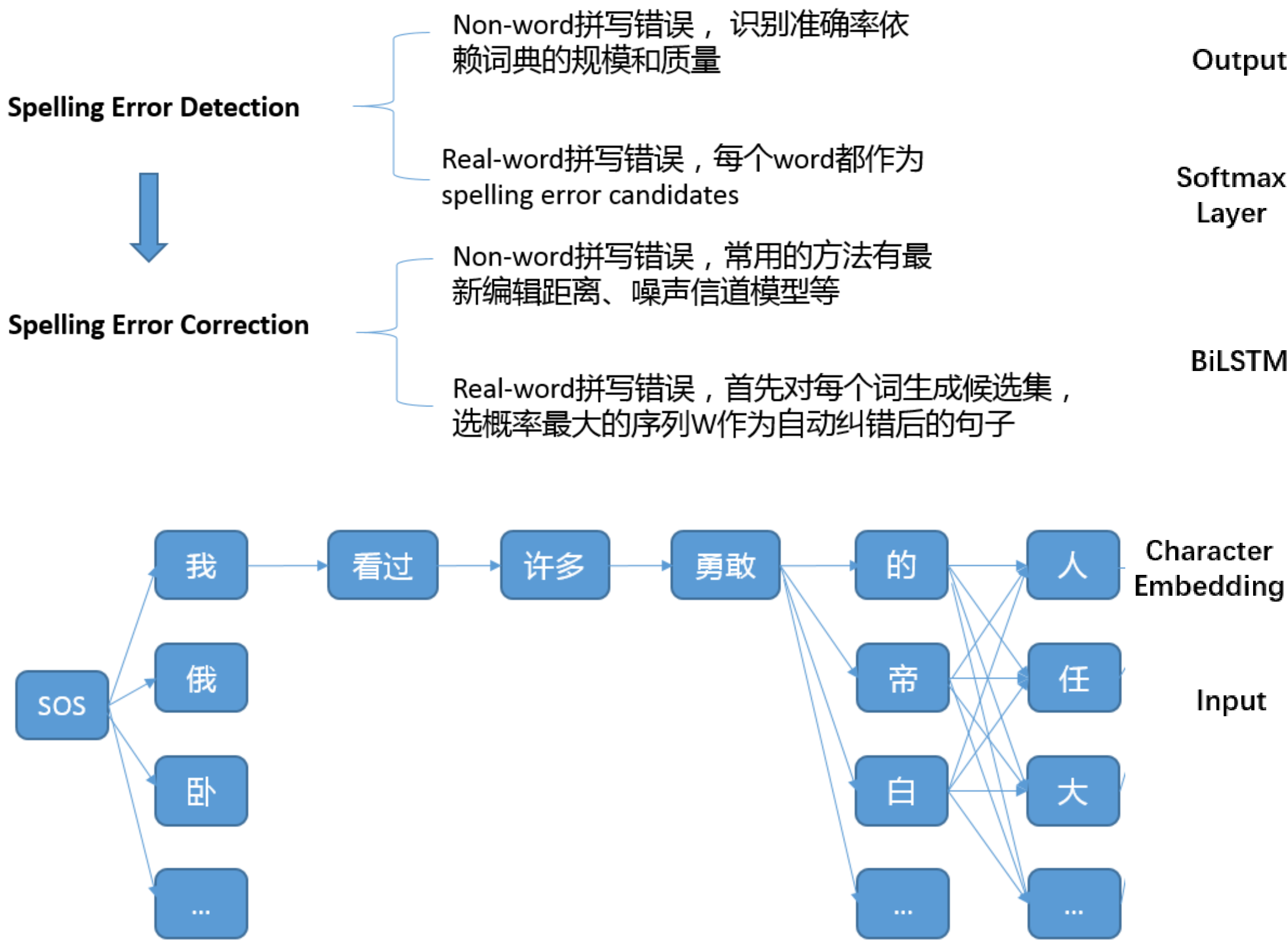


# 知文问答v0.7：对话管理

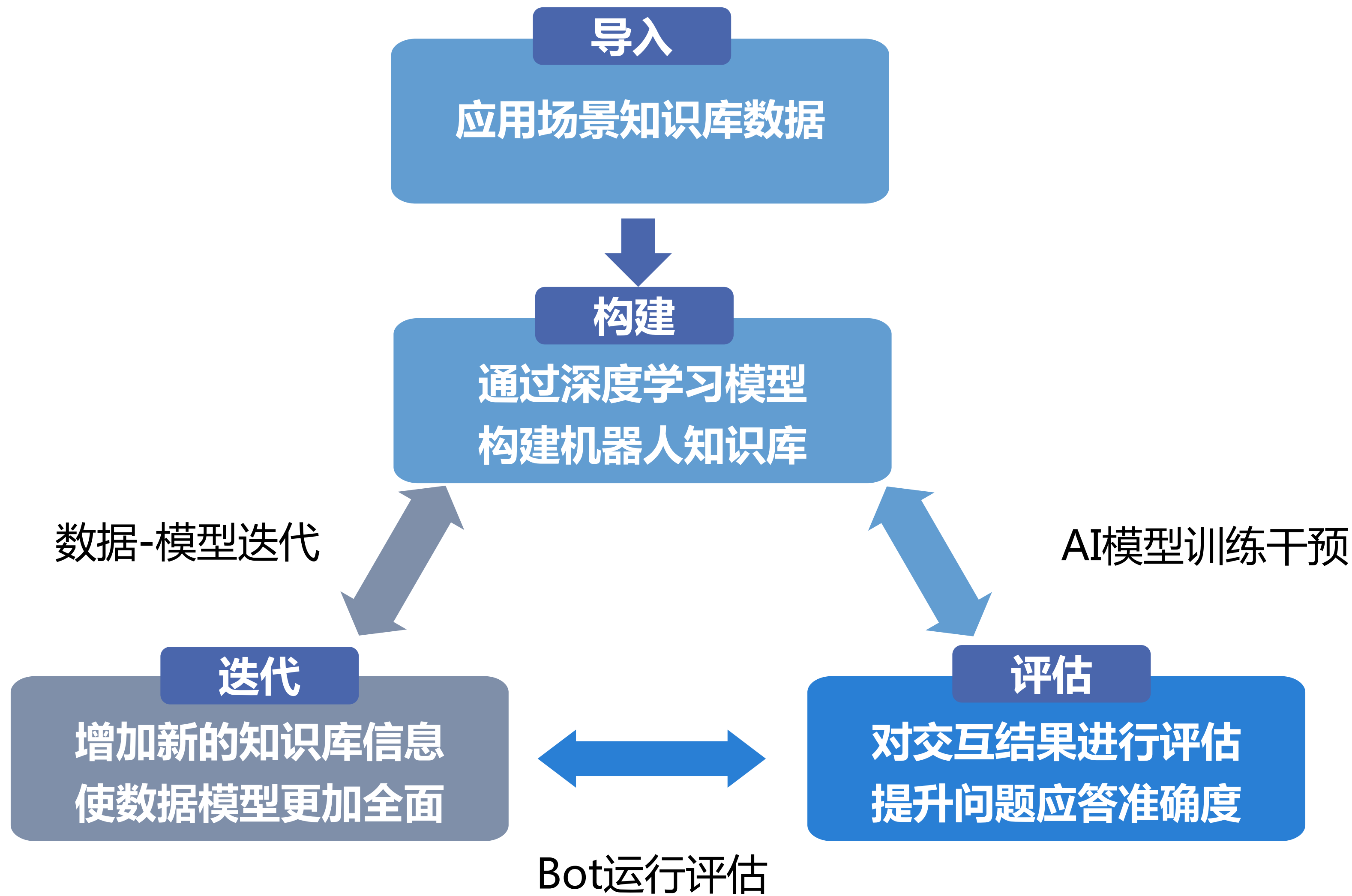




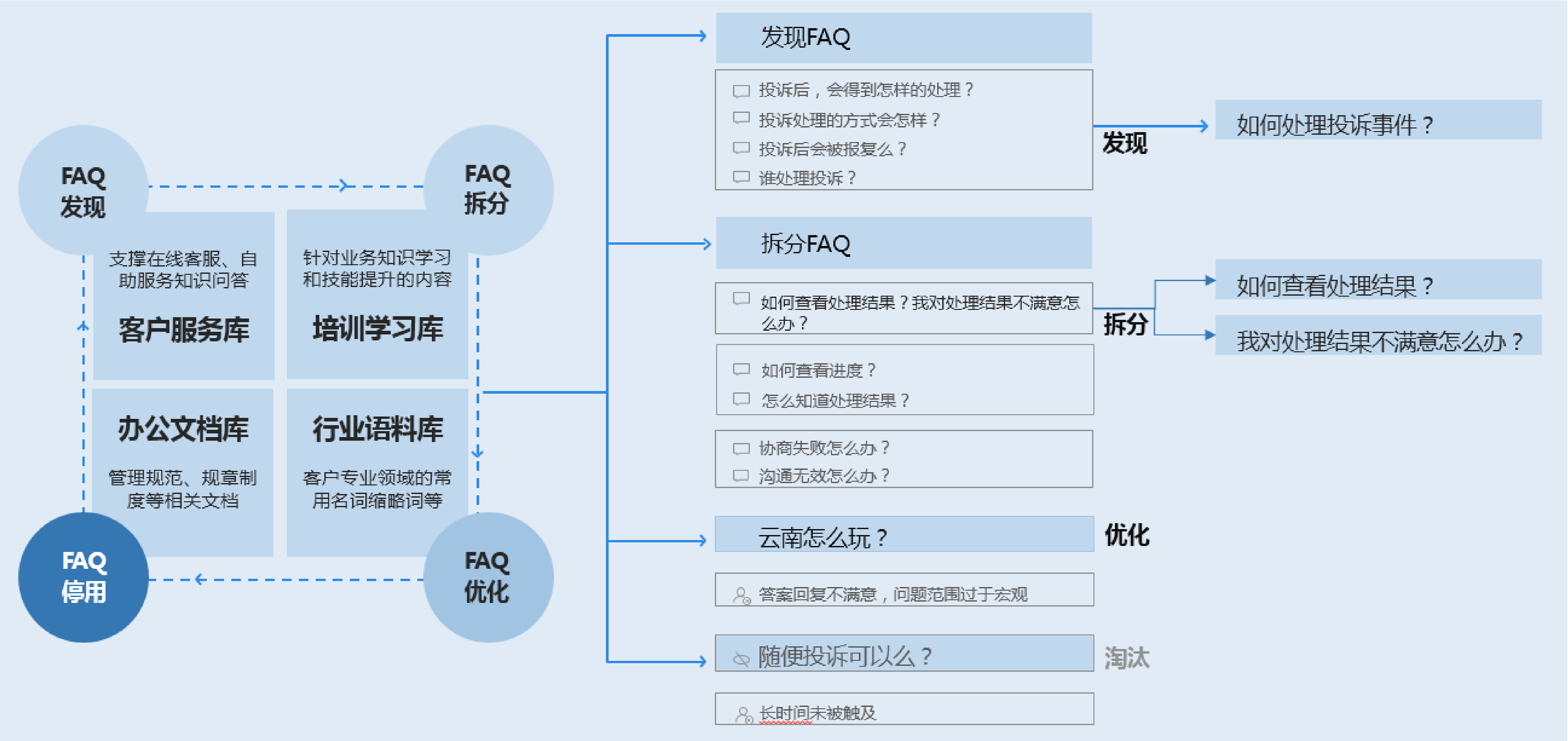
# 知文问答v0.8 : Query处理



# 知文问答v0.9：知识库管理



# 知文问答v0.9：知识库管理





# 知文问答v1.0：智能问答分析

## 客服Bot使用统计

- 客服Bot调用趋势
- 客服Bot反馈速率
- 客服Bot答复准确率
- 不同Bot使用情况

## 客户咨询词云分析

- 客户阶段内反馈问题统计
- 客户反馈问题主题词云
- 关联关键词分布
- 对应Query示例

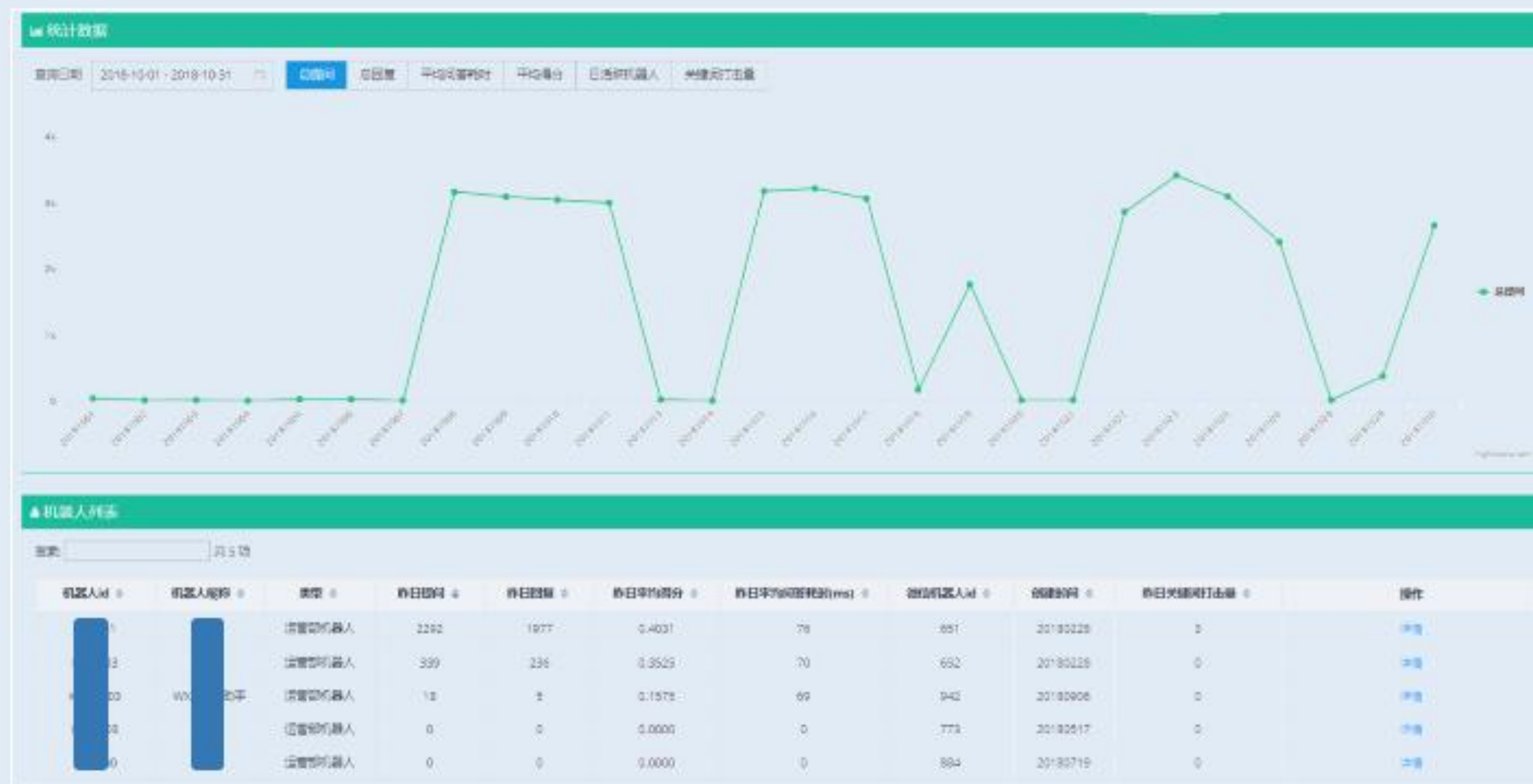


TABLE OF

# CONTENTS 大纲

---

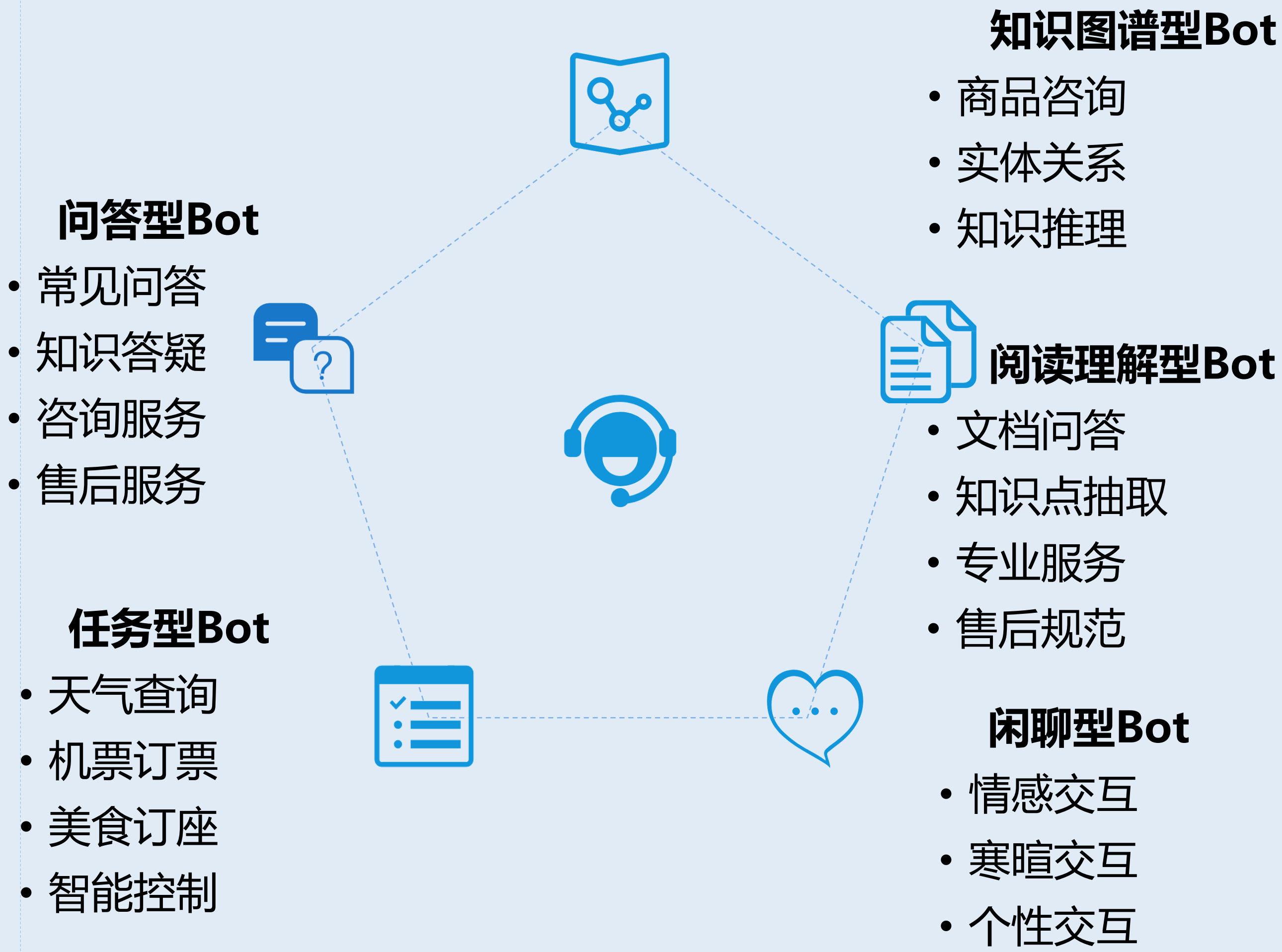
- 知文问答：从FAQ机器人说起
- 用活存量文档：文档型机器人
- 走向结构化数据：知识图谱机器人
- 知文技能树：任务型机器人
- 对话管理、Query处理与知识库管理
- 总结



# 知文问答回顾

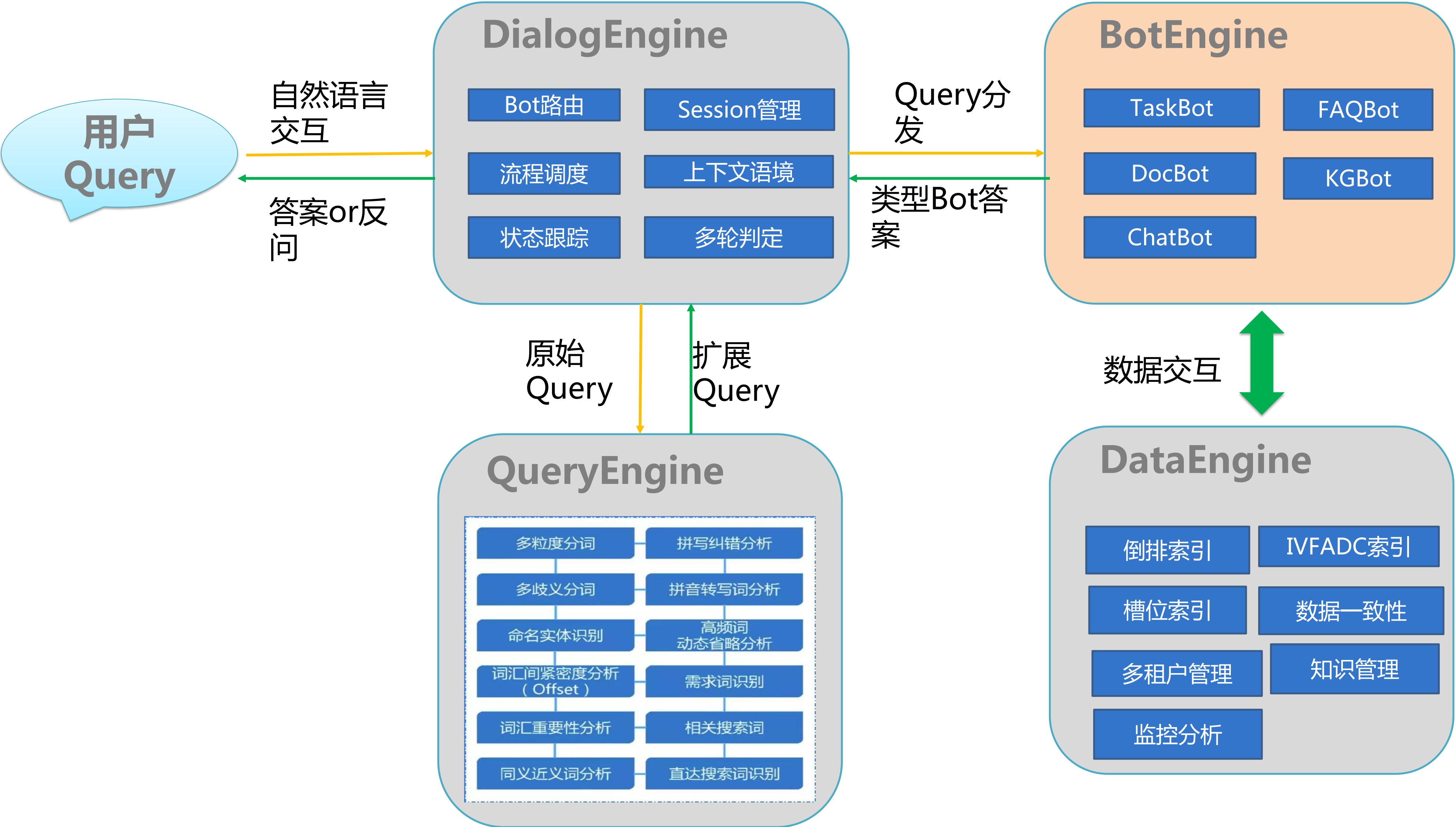
知文问答系统提供了多种通用机器人形态进行组合，满足用户多层次多场景的智能交互需求。

- **FAQ-Bot**：问答型机器人，常见业务问题问答机器人，解决90%业务问题
- **Task-Bot**：任务型机器人，明确具体任务的人机交互，例如：查天气、定机票
- **Chat-Bot**：闲聊型机器人，具有温度的智能互动闲聊
- **Doc-Bot**：阅读理解型机器人，直接从文档匹配答案
- **KG-Bot**：知识图谱型机器人，输出实体或关系型答案

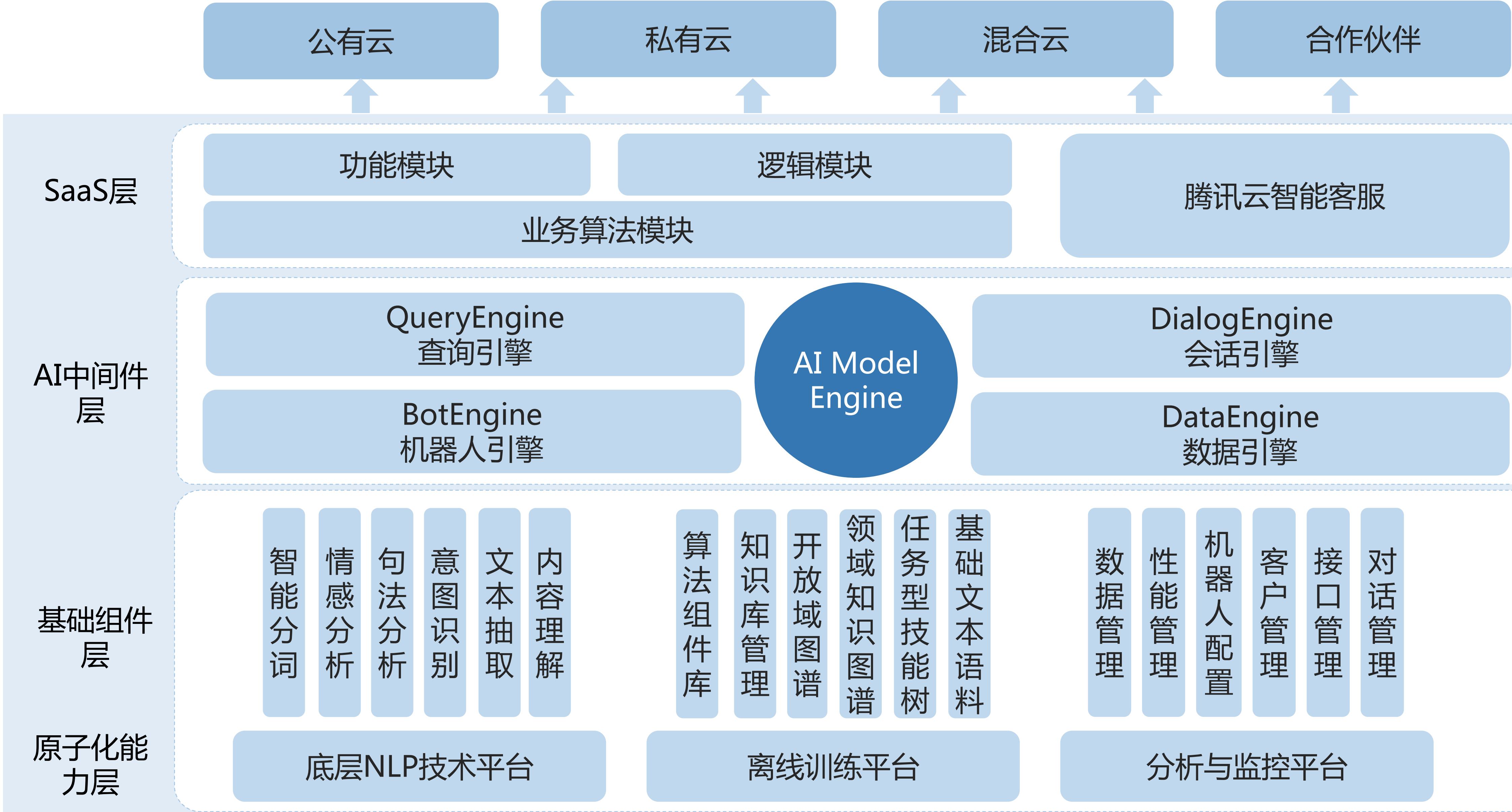




# 知文问答v1.0：算法架构



# 腾讯云知文开放平台



# THANKS

**AiCon**  
全球人工智能与机器学习技术大会