



中國計算機學會
CHINA COMPUTER FEDERATION



CCF计算经济学专业组



北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University



北大
创新评论
PKU
Innovation
Review



博时基金
BOSERA FUNDS

阿里云



金证股份

未名数创
Undiscovered Innovation

2023 CCF 计算经济学比赛 博金挑战赛

蔡晖

梦想还是要有的



中国计算机学会
CHINA COMPUTER FEDERATION



CCF计算经济学专业组



北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University



北大
创新评论
PKU
Innovation
Review



博时基金
BOSERA FUNDS



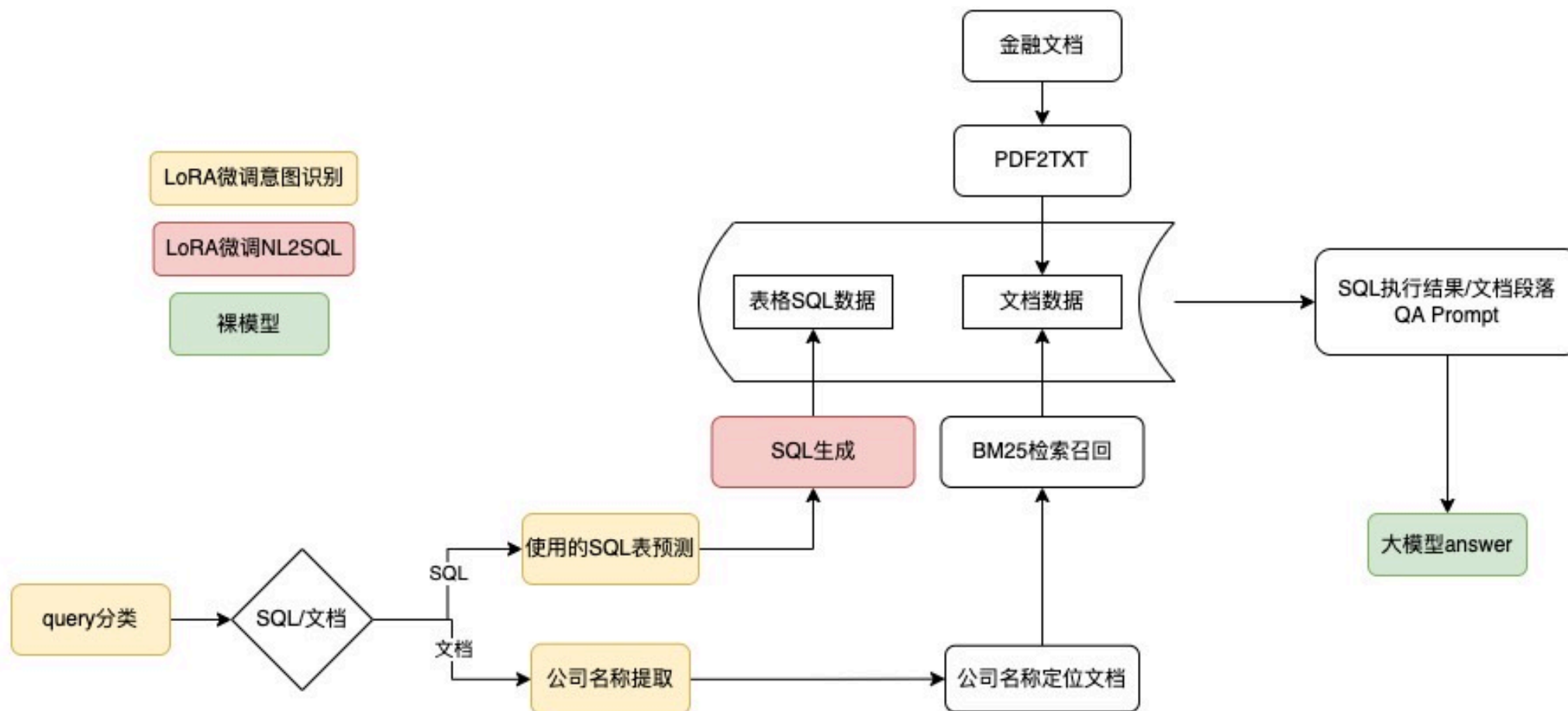
阿里云



金证股份

未名数创
Undiscovered Innovation

整体链路





中国计算机学会
CHINA COMPUTER FEDERATION



CCF计算经济学专业组



北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University



北大
创新评论
PKU
Innovation
Review

博时基金
BOSERA FUNDS

阿里云



金证股份

未名数创
Undiscovered Innovation

Query分类/意图识别 & 信息提取

LoRA微调Tongyi-14B，测试集上准确率**100%**

Instruction

现在，你个问题分类器。用户会输入一些金融类问题，我们需要对问题做分类，来确认最适合的回答方式。
一共有两类，一类是通过查询结构化的SQL数据表来回答，还有一类是查询非结构化的文档（招股说明书）进行回答。
你需要根据输入，判断该问题是更适合使用SQL查询数据库计算得到，还是从文档中寻找答案。
请用Json格式输出问题分类，如果是SQL查询，还需要输出所有需要使用的表名，如果是文档召回，请输出公司名称。

Few Shot

例子：

输入：湖南长远锂科股份有限公司变更设立时作为发起人的法人有哪些？
输出：{"问题分类": "文档召回", "公司名称": "湖南长远锂科股份有限公司"}

输入：嘉实基金管理有限公司2019年成立了多少基金？
输出：{"问题分类": "SQL查询", "表名": ["基金基本信息"]}

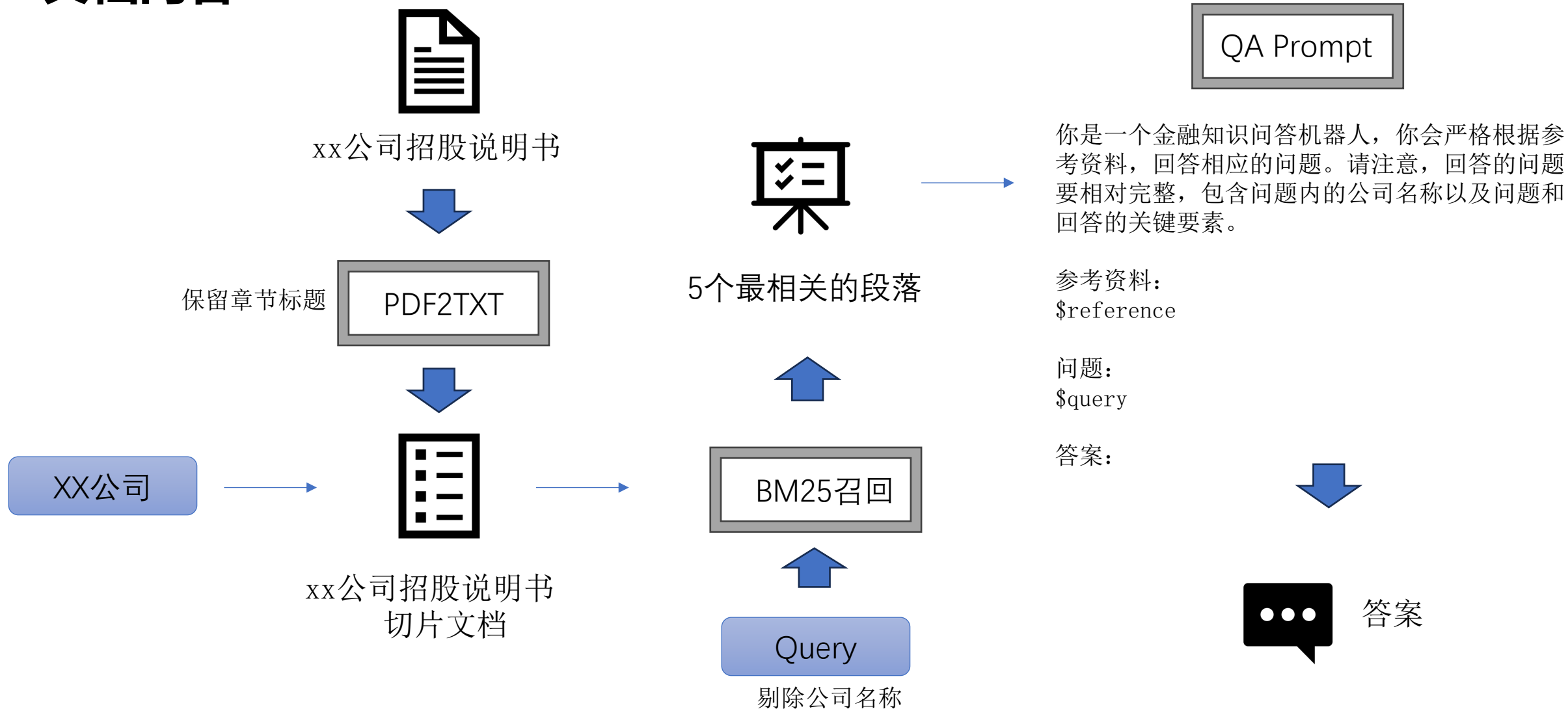
Information

我们的数据库和字段有：
{ '基金基本信息': ['基金代码', '基金全称', '基金简称', '管理人', '托管人', '基金类型', '成立日期', '到期日期', '管理费率', '托管费率'],
.....

Input

现在我们开始：
输入：\$query
输出：

文档问答





中国计算机学会
CHINA COMPUTER FEDERATION



CCF计算经济学专业组



北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University



北大
创新评论
PKU Innovation Review



博时基金
BOSERA FUNDS



阿里云



金证股份

未名数创
Undiscovered Innovation

NL2SQL 流程

问题

请帮我查询在2019年,海富通基金管理有限公司成立哪种类型的基金个数最多?

关键要素1:
需要的表名称

关键要素2:
表里的字段

关键要素2:
每个字段的示例数据

关键要素2:
所有枚举值



千问14B
Chat - SFT

SQL结果

```
SELECT COUNT(*) AS  
基金数量, [基金类型]  
FROM [基金基本信息]  
WHERE [管理人] = '海  
富通基金管理有限公司'  
AND [成立日期] >=  
'20190101' AND [成立  
日期] <= '20191231'  
GROUP BY [基金类型]  
ORDER BY 基金数量  
DESC LIMIT 1
```

查询结果

```
[[{'基金数量': 9, '  
基金类型': '债  
券型'}]]
```



千问14B
Chat +
Few-shot

最终答案

在2019年,
海富通基金
管理有限公司成立的债
券型基金个
数最多, 共
有9只。

Database



中國計算機學會
CHINA COMPUTER FEDERATION



CCF计算经济学专业组



北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University



北大
创新评论
PKU
Innovation
Review

博时基金
BOSERA FUNDS

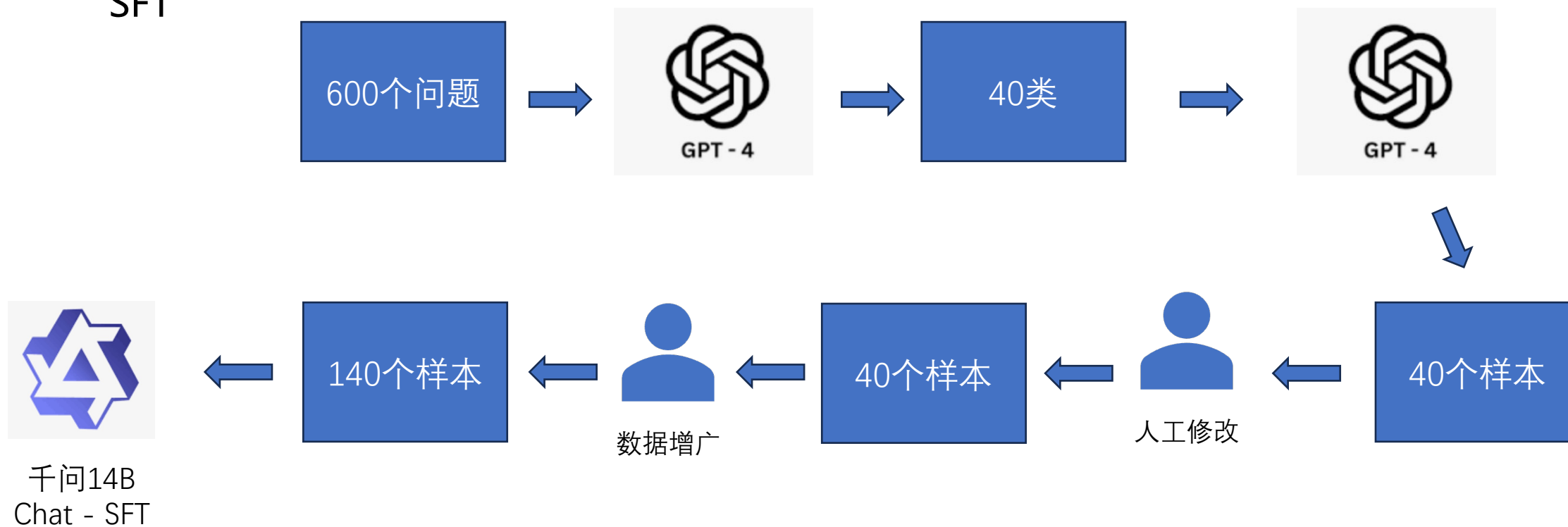
阿里云



金证股份

未名数创
Undiscovered Innovation

NL2SQL
SFT





中國計算機學會
CHINA COMPUTER FEDERATION



CCF计算经济学专业组



北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University



北大
创新评论
PKU
Innovation
Review



未名数创
Undiscovered Innovation

Future Work

文档问答

召回:

1. 利用目录信息, 进一步缩小召回范围
2. 多路召回 (向量数据库)

回答:

1. 降噪/信息 (指令) 遵循

SQL问答

召回:

1. 表/字段精细化召回

NL2SQL:

1. COT, 将问题拆分, 分别计算并汇总

回答:

1. 信息 (指令) 遵循