

# 招股书分析报告

## 目标

提前设置有关新上市公司招股书（<https://data.eastmoney.com/xg/xg/default.html>）需要分析的问题列表，上传招股书后，系统自动对预置的问题逐个执行问答，最终生成招股书分析报告。

我们选取三个具有代表性的问题：

1. 该公司的审计机构是哪家公司？该审计机构去年的上市项目数量与排名分别是多少？

需要从招股书中查询信息并进一步结合外部信息（审计机构行业排名情况）进行回答

2. 该公司处在哪个行业？搜索该行业最近几年中国的市场规模是多少？

需要从招股书中查询信息并进一步结合外部信息（市场规模）进行回答

3. 该公司流动比率有何变化？

数值信息计算

除了以上三个问题，专业人员一般会从[商业模式画布](#)中的八个分析维度以及一些重要财务指标的分析来完成对一家新上市公司总体情况的了解，我们在毕昇演示环境（[dataelem.com/bisheng](https://dataelem.com/bisheng)）中提供了总计42个问题的全量招股书分析技能，欢迎试用。

## 效果示例

文件

达梦数据库招股说明书-0830.pdf

文件解析中

文件解析完成，分析开始

该公司聘请的审计机构是哪家公司？该审计机构去年的上市项目数量与排名分别是多少？

Thought: 我需从文档中查询该公司聘请的审计机构是哪家公司以及该审计机构去年的上市项目数量和排名。

Action: Basic\_information

Action Input: 该公司的审计机构是哪家公司？

Tool input: 该公司的审计机构是哪家公司？

langchain 不支持异步，进入同步模式

Thought: 我需从文档中查询该公司聘请的审计机构是哪家公司以及该审计机构去年的上市项目数量和排名。

Action: Basic\_information

Action Input: 该公司的审计机构是哪家公司？

Tool input: 该公司的审计机构是哪家公司？

Observation: 该公司的审计机构是中天运会计师事务所（特殊普通合伙）。

Thought: 我发现文章中提到该公司的审计机构是中天运会计师事务所（特殊普通合伙）。

Action: Auditors\_info

Action Input: 中天运会计师事务所（特殊普通合伙）

Tool input: 中天运会计师事务所（特殊普通合伙）

Tool input: 中天运会计师事务所（特殊普通合伙）

Observation: 中天运会计师事务所排名第16，项目量为4。

我发现中天运会计师事务所（特殊普通合伙）在去年的审计项目数量为4，排名第16。

Final Answer: 该公司聘请的审计机构是中天运会计师事务所（特殊普通合伙），该审计机构去年的上市项目数量是4，排名第16。

该公司聘请的审计机构是中天运会计师事务所（特殊普通合伙），该审计机构去年的上市项目数量是4，排名第16。

该公司处在哪个行业？搜索该行业最近几年中国的市场规模是多少？

Thought: 我需要从文档中查询该公司所处的行业，然后再搜索该行业在中国的市场规模

Action: Basic\_information

Action Input: 该公司处在哪个行业？

Tool input: 该公司处在哪个行业？

langchain 不支持异步，进入同步模式

Thought: 我需要从文档中查询该公司所处的行业，然后再搜索该行业在中国的市场规模

Action: Basic\_information

Action Input: 该公司处在哪个行业？

Tool input: 该公司处在哪个行业？

Observation: 该公司处在基础软件行业。

Thought: 我发现达梦股份有限公司所处的行业是基础软件行业。我需要搜索该行业在中国的市场规模。

Action: bing\_search

Action Input: 基础软件行业中国市场规模

Tool input: 基础软件行业中国市场规模

Thought: 我找到了关于基础软件行业在中国市场规模的信息。根据搜索结果显示，2022年全国软件和信息技术服务业规模以上企业超过3.5万家，累计完成软件业务收入108126亿元，同比增长11.2%。此外，根据2020年的数据，基础软件行业在中国整体IT开支中的占比非常低，相比于美国基础软件投入占比可以达到三分之一甚至一半，中国的基础软件投入可能只有个位数。这说明中国整个基础软件产业竞争力相对较弱，基础软件的价值在产业中也没有得到广泛认同。所以国产操作系统、数据库企业的市场份额相对较低。

Tool input: Invalid Format: Missing 'Action:' after 'Thought:'

I now know the final answer

Final Answer: 达梦股份有限公司所处的行业是基础软件行业。根据搜索结果，2022年全国软件和信息技术服务业规模以上企业超过3.5万家，累计完成软件业务收入108126亿元，同比增长11.2%。基础软件行业在中国整体IT开支中的占比相对较低，国产操作系统、数据库企业的市场份额也相对较低。

达梦股份有限公司所处的行业是基础软件行业。根据搜索结果，2022年全国软件和信息技术服务业规模以上企业超过3.5万家，累计完成软件业务收入108126亿元，同比增长11.2%。基础软件行业在中国整体IT开支中的占比相对较低，国产操作系统、数据库企业的市场份额也相对较低。

公司流动比率有何变化？

Thought: 我需要查询公司流动比率的变化

Action: Basic\_information

Action Input: 公司流动比率的变化

Tool input: 公司流动比率的变化

langchain 不支持异步，进入同步模式

Thought: 我需要查询公司流动比率的变化

Action: Basic\_information

Action Input: 公司流动比率的变化

Tool input: 公司流动比率的变化

Observation: 根据提供的信息，公司的流动比率在报告期内有所变化。具体数据如下：

- 2021年度流动比率为3.74倍
- 2020年度流动比率为2.82倍
- 2019年度流动比率为2.05倍

可以看出，公司的流动比率在这三个年度中呈逐年增加的趋势，即公司的流动性逐渐提高。流动比率是衡量公司偿付短期债务能力的指标，数值越高表示公司具有更强的短期偿付能力。因此，根据给出的数据，可以得出结论公司的短期偿债能力逐年提升。

我现在知道了公司流动比率的变化

Final Answer: 公司的流动比率在报告期内呈逐年增加的趋势，分别为3.74倍、2.82倍和2.05倍。这表明公司的流动性逐渐提高，短期偿付能力逐年提升。

公司的流动比率在报告期内呈逐年增加的趋势，分别为3.74倍、2.82倍和2.05倍。这表明公司的流动性逐渐提高，短期偿付能力逐年提升。

该公司聘请的审计机构是哪家公司？该审计机构去年的上市项目数量与排名分别是多少？

该公司聘请的审计机构是中天运会计师事务所（特殊普通合伙），该审计机构去年的上市项目数量是4，排名第16。

该公司处在哪个行业？搜索该行业最近几年中国的市场规模是多少？

达梦股份有限公司所处的行业是基础软件行业。根据搜索结果，2022年全国软件和信息技术服务业规模以上企业超过3.5万家，累计完成软件业务收入108126亿元，同比增长11.2%。基础软件行业在中国整体IT开支中的占比相对较低，国产操作系统、数据库企业的市场份额也相对较低。

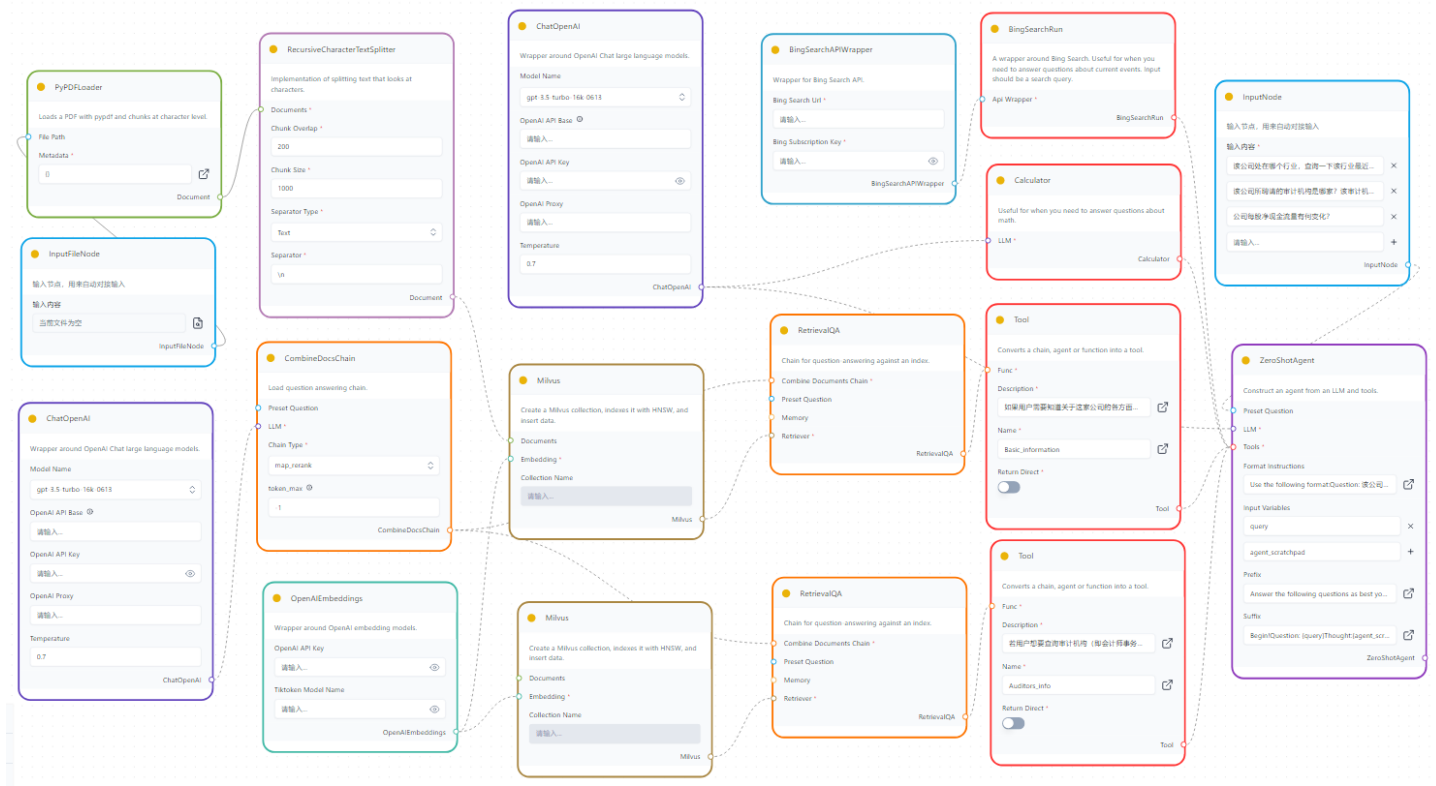
公司流动比率有何变化？

公司的流动比率在报告期内呈逐年增加的趋势，分别为3.74倍、2.82倍和2.05倍。这表明公司的流动性逐渐提高，短期偿付能力逐年提升。

## 构建思路

针对这三个问题分析技能所需的组件：

1. 需要根据文本内容进行问答，因此选择语言模型 [ChatOpenAI](#) 组件（或其他语言模型组件）、[RetrievalQA](#) 组件、[CombineDocsChain](#) 组件、[Milvus](#) 组件、[OpenAIEmbeddings](#) 组件，组成对文本进行问答的链；
  2. 除了根据招股书之外，还要根据包含审计机构项目数量和排名的文本内容回答，所以我们需要构建两条对文本进行问答的链，其中一条链需要让用户上传招股书报告，因此选择 [PyPDFLoader](#) 组件和 [InputFileNode](#) 组件；另一条链我们提前把包含审计机构项目数量和排名的文本导入知识库，在Milvus组件的collection name参数中填写对应知识库的信息；
  3. 回答第二个问题需要联网进行搜索，选择 [BingSearchRun](#) 组件、[BingSearchAPIWrapper](#) 组件；
  4. 回答第三个问题涉及到对比计算，选择 [Calculator](#) 组件提供计算能力（大语言模型具备一定的计算能力，也可以选择不添加Calculator组件）；
  5. 选择 [ZeroShotAgent](#) 组件把所有部分连接起来，BingSearchRun组件和Calculator组件是[工具/Tools]类型下的组件，因此可以直接和[代理/Agents]连接，对文本进行问答的链想要与[代理/Agents]连接，需要引入 [Tool](#) 组件，把整个对文本进行问答的链作为工具与[代理/Agents]连接；
  6. 需要预置问题，因此选择 [Input Node](#) 组件。
- 连接示例如图：



## ● 工作原理

用户上传文档并解析完成后，ZeroShotAgent组件会自动逐个运行预置的问题，并且每个问题自行选择调用哪个工具（一个或多个）来回答，逐个问题得到答案后最终将所有问题的答案拼接在一起生成分析报告。

## ● 组件参数设置

### ○ ChatOpenAI

- Model Name：推荐选择gpt-3.5-turbo-16k-0613
- OpenAI API Key：API接口密钥
- OpenAI API Base/OpenAI Proxy：服务的地址，填写其中之一即可（如果是自己代理的国内的地址则填写OpenAI Proxy参数）

### ○ Milvus

- Collection Name

只能有一个Milvus组件连接[装载机/Loaders]中的组件，供用户上传招股书报告，此时用户上传报告后会自动创建一个库，无需填写；

其他Milvus组件需填写对应毕昇平台[知识]功能模块中创建知识库的Collection Name；

### ○ OpenAIEmbeddings

- Tiktoken Model Name：模型名称，推荐使用text-embedding-ada-002
- OpenAI API Key：API接口密钥
- OpenAI API Base/OpenAI Proxy：服务的地址，填写其中之一即可（如果是自己代理的国内的地址则填写OpenAI Proxy参数）

- CombineDocsChain
  - Chain Type: 选择“stuff”，原因详见 [📖 CombineDocsChain](#)
  - token\_max: “-1”，即对传给大模型的文本长度不做限制，详见 [📖 CombineDocsChain](#)
- BingSearchAPIWrapper
  - Bing Search Url
  - Bing Subscription Key
- Tool
  - Description: 对工具的描述，ZeroShotAgent组件根据描述判断是否调用这个工具。连接导入招股书文档的Tool组件可以填写“如果用户需要知道关于这家公司的各方面信息，可以使用该工具。”，连接审计机构信息文档的Tool组件可以填写“若用户想要查询审计机构（即会计师事务所）的背景信息，可以使用该工具。”
  - Name: 工具的名称，连接导入招股书文档的Tool组件可以设置为“Basic\_information”，连接审计机构信息文档的Tool组件可以设置为“Auditors\_info”
- InputNode: 预置的问题，这里填入开头列出的三个问题
- ZeroShotAgent
  - Format Instructions: 指导Agent运行的方式
  - input\_variables: 最终的提示词所需要的输入变量列表，对应Suffix中的变量，即{query}和{agent\_scratchpad}
  - prefix: 放在工具列表之前的字符串，作为提示词模板的开头
  - suffix: 放在工具列表后面的字符串，作为提示词模板的结尾

## 优化

如果问题的答案不完全正确，或无法给出答案，除 [📖 合同审核](#) 中提到的问题优化、ZeroShotAgent组件参数优化外，我们还可以对其他组件的参数进行优化。

### • Tool组件参数优化

Tool组件的参数是作为提示词的一部分传递给大语言模型的，详见 [📖 ZeroShotAgent](#)，所以，当我们有多条文本问答的链的时候，可以在Tool组件的“Description”参数中，更详细地描述这个Tool组件连接的知识库所包含的文本信息、能回答的问题种类等。

如可以在连接招股书文档的Tool组件的“Description”参数中添加如下问题，让语言模型学习到如果需要回答这类问题，可以使用这个工具：

- 1 如果用户需要知道关于这家公司的各方面信息，可以使用该工具。如：
- 2 - 某公司的盈利模式是什么

- 3 - 某公司的保荐机构是哪家公司？
- 4 - 某公司的审计机构是哪家公司？
- 5 - 某公司的主要产品是什么？

可以在连接审计机构信息文档的Tool组件的“Description”参数中添加“如在去年一年中审计项目的项目量和排名”，更加明确这个工具的知识库所能提供的信息：

- 1 若用户想要查询审计机构（即会计师事务所）的背景信息，如在去年一年中审计项目的项目量和排名，可以使用该工具。

- CombineDocsChain参数优化

如果语言模型总是因为超token数无法输出答案，可以通过“token\_max”参数设置进行优化。原理详见 [📖 CombineDocsChain](#)