# 一、数据库说明：

报刊新闻量化舆情数据平台包含从 1998 年起由报纸媒体发布的与 A 股（及 B股）全部上市公司相关新闻，以香港中文大学、美国斯坦福大学和南加州大学等高校教授的相关学术研究为指导[2]， 采用由香港中文大学授权使用的分析技术和相关数据，并结合业界领先的机器学习和自然语言处理技术对收集到的上千万篇原始新闻进行了严谨的清理分析及核验。

# 二、数据总体结构：

本数据平台主要包括了五张数据表（如下表所示） ，用户可以通过自定义不同的条件关联检索出更复杂细致的信息以支持不同的研究应用。

表 数据平台总体结构表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 字段数 | 区间 | 内容说明 | 字段内容 |
| NEWS\_MENT\_STKS | 6 | 1998-04-02~ | 新闻提及的上市公司及相关信息 | 新闻编号、股票代码、提及次数、提及句子数、首次提及位置、标题是否提及 |
| NEWS\_SENTIMENT | 16 | 1998-04-02~ | 新闻的情感分析结果 | 新闻编号、句子总数、负面句子数、中性句子数、正面句子数、首尾段句子总数、首尾段负面句子数、首尾段中性句子数、首尾段正面句子数、段首尾句子总数、段首尾负面句子数、段首尾中性句子数、段首尾正面句子数、标题情感评分、正文情感评分、整体情感评分 |
| NEWS\_SIMI | 4 | 1998-04-02~ | 新闻与历史新闻的内容相似度分析结果 | 新闻编号、历史新闻编号、间隔时长、相似度 |
| PRESS\_INFO | 5 | 1998-04-02~ | 报社基本信息 | 报社编号、发行报刊名称、是否海外媒体、报社所在地、创刊年份 |
| NEWS\_BASIC\_INFO | 15 | 1998-04-02~ | 新闻基本信息表，包括提及最多的公司，情感倾向评分，发布日期和报社等 | 新闻编号、发布日期、报社编号、最相关股票代码、标题是否提及、提及句子数、提及次数、提及到不同上市公司数、提及不同上市公司的句子数、提及不同上市公司的次数、整体情感评分、是否原创、新闻热度、是否有外部链接、外部链接 |