**爬虫目标网站评估**

**1. Wall Street Journal（华尔街日报）**

· **类型**: 新闻资讯

· **链接**: https://www.wsj.com/news/latest-headlines?mod=nav\_top\_section

· **大体描述**: 该网站提供覆盖多个行业、领域和地域的新闻资讯，其中仅部分栏目具有爬取价值。

· **栏目分析**:

1. 全球新闻对美股市场的影响有限，难以提供准确的市场反应。

2. 政治类新闻（如选举）对美股的直接影响较弱，通常是通过间接途径产生影响，但不够直接。

3. 经济新闻在一定程度上能反映宏观经济环境，但难以直接影响特定股票价格；重大经济决策如美联储利率调整可能对整个美股市场产生重要影响。

4. 前沿技术板块的人工智能和生物技术新闻对某些特定行业具有较大的直接影响。

5. 市场与金融板块中的商品和期货类新闻与大宗商品市场密切相关，股票类新闻则直接影响美股市场。

6. 观点、艺术、生活方式、房地产、个人理财、健康等栏目不具备爬取价值。

· **落库形式**: 资讯标题、内容、发布日期、链接及所属类型。

· **潜在应用场景**: 可用于大宗商品、美股及特定行业美股的关联性资讯快速检索，并结合语义分析判断新闻对市场的利好或利空影响。

**2. Bloomberg（彭博社）**

· **类型**: 新闻资讯

· **链接**: https://www.bloomberg.com/

· **大体描述**: 提供针对不同国家和地区的财经新闻、市场数据、市场分析、研报和经济指标等内容。

· **栏目分析**:

1. 国家/地区：仅爬取美国市场（US）相关内容，符合美股市场研究需求。

2. Deals：包含大量企业相关简讯，信息丰富但可能难以区分企业的国别。

3. 固定收入：涉及期货（如小麦）、银行、债券、美联储货币政策及企业动态（如并购信息），但获取成本较高。

4. ETF：与大型金融企业如贝莱德相关的资讯具有一定价值，此外部分数字货币资讯可能对区块链行业产生影响，但总体影响较小。

5. 外汇：主要与各国央行及货币政策相关，缺乏企业直接关联性，获取价值较低。

6. 因子投资/要素投资：与行业存在一定关联，偶尔能获取与企业直接相关的资讯，但信息更新频率较低，性价比需进一步评估。

7. 日历：与金十数据相比，性价比偏低。

8. 股票：不具备爬取价值。

9. 期货：虽然与大宗商品有相关性，但历史数据不多，需评估性价比。

10. 债券利率、货币：不具备爬取价值。

11. 商品：与大宗商品相关，但历史数据较少，需进一步评估。

12. 行业表现：不具备爬取价值。

13. 能源：与大宗商品相关，历史数据少，性价比需考量。

14. 经济学：不具备爬取价值。

15. 行业/财经：与美国企业关联性不强，不具备爬取价值。

16. 科技/人工智能 & 科技/大型科技公司 & 科技/更多资讯：由于人工智能和科技领域的头部企业主要集中在美国，具有较强的针对性，因此具备较高爬取价值。

17. 商业周刊、政治、观点：不具备爬取价值。

18. 媒体、探索：不具备爬取价值。

**3. Financial Times（金融时报）**

· **类型**: 新闻资讯

· **链接**: https://www.ft.com/

· **大体描述**: 报道国际事件（如战争）、全球政策及商业动态。

· **栏目分析**:

1. home：新闻与美国企业、大宗商品无明显关联，不具备爬取价值。

2. 全球：虽然按洲分类，但信息与美国企业无直接关联，不具备爬取价值。

3. 我们/投资美国：与美国企业无直接关联，不具备爬取价值。

4. 我们/美国主页：与美国企业关联性弱，不具备爬取价值。

5. 我们/美国经济：与美国企业无直接关联，不具备爬取价值。

6. 我们/美国公司：与美国和加拿大企业有较强关联，具备较高爬取价值。

7. 公司/科技领域：与科技类企业相关，尤其是美国科技企业，有较高爬取价值。

8. 技术：与企业信息密切相关，具备爬取价值。

9. 技术/人工智能 & 技术/半导体：与企业信息高度相关，具备较高爬取价值。

10. 市场：信息宏观性较强，缺乏针对性，不具备爬取价值。

11. 气候、观点：不具备爬取价值。

**4. financialdatadb（财务报表）**

· **类型**: 付费制，财务报表获取（季度），支持csv和api格式。

· **链接**: https://financialdatadb.com/cos/us/featured

· **示例文件1**:

o 字段

§ **公司名称**（例如：Zscaler）

§ **股票代码**（例如：ZS）

§ **开始日期**（例如：2017-Jul-31）

§ **结束日期**（例如：2024-Jul-31）

· 示例文件2：

· 示例文件3：

**5. discountingcashflows（财务报表）**

· **类型**: 付费制，财务报表获取（年度），支持csv和api格式。

· **链接**: https://discountingcashflows.com/company/AAPL/balance-sheet-statement/

· **示例文件**:

**6. macrotrends（财务报表）**

· **类型**: 免费，支持获取年度及季度财务报表。

· **链接**: https://www.macrotrends.net/stocks/charts/KO/cocacola/financial-statements

· 具备获取价值。

**7. annualreports（年度报表）**

· **类型**: 免费，提供企业年度报告

· **链接**: https://www.annualreports.com/Company/apple-inc

· 具备获取价值。

**8. FXstreet（交易平台）**

· **类型**: 聚合交易平台

· **链接**: https://www.plus500.com/zh-sg/

· 不具备获取意义。

**9. AHDB（期货商品相关）**

· **类型**: 期货商品

· **链接**: https://ahdb.org.uk/

· 对于获取美国企业的信息而言，不具备太大的爬取价值；对于商品期货而言，具备极大的爬取价值。

**10. USDA（美国农业部）**

· **类型**: 政府官方网站

· **链接**: https://www.usda.gov/equity

· 对于商品期货，具备极大的爬取价值；但对于美国企业而言，不具备太高的爬取价值。

**11. 其他资源**

· **Yahoo Finance（雅虎财经）**

· **Reuters（路透社）**: https://www.reuters.com/

· **Agri Census**

· **CNBC（媒体）**: https://www.cnbc.com/world/?region=world