

RAG实体统计

RAG 模型抽取结果统计报告

生成时间: 2025-10-17 15:22:24

1. 模型对比汇总

模型	成功论文数	失败论文数	总实体数	总关系数	平均实体数	平均关系数	平均实体类型数	平均关系类型数	总实体类型数	总关系类型数
DeepSeek	10	0	289	234	28.9	23.4	10.4	13.1	20	74
Gemini	10	0	304	225	30.4	22.5	10.0	14.3	23	94
Kimi	10	0	163	143	16.3	14.3	7.6	9.0	20	52

2. 关键发现

- 最多平均实体数: Gemini (30.4 个)
- 最多平均关系数: DeepSeek (23.4 个)
- 最多平均实体类型数: DeepSeek (10.4 种)
- 最多平均关系类型数: Gemini (14.3 种)
- 最丰富实体类型: Gemini (共 23 种不同类型)
- 最丰富关系类型: Gemini (共 94 种不同类型)

3. 各模型详细数据

DeepSeek

- 成功论文数: 10
- 总实体数: 289
- 总关系数: 234
- 平均实体数: 28.9
- 平均关系数: 23.4
- 平均实体类型数: 10.4
- 平均关系类型数: 13.1
- 总实体类型数 (去重) : 20
- 总关系类型数 (去重) : 74
- 详细数据: `deepseek/paper_statistics.csv`

Gemini

- 成功论文数: 10
- 总实体数: 304
- 总关系数: 225
- 平均实体数: 30.4
- 平均关系数: 22.5
- 平均实体类型数: 10.0
- 平均关系类型数: 14.3
- 总实体类型数（去重）: 23
- 总关系类型数（去重）: 94
- 详细数据: `gemini/paper_statistics.csv`

Kimi

- 成功论文数: 10
- 总实体数: 163
- 总关系数: 143
- 平均实体数: 16.3
- 平均关系数: 14.3
- 平均实体类型数: 7.6
- 平均关系类型数: 9.0
- 总实体类型数（去重）: 20
- 总关系类型数（去重）: 52
- 详细数据: `kimi/paper_statistics.csv`

4. 文件说明

```
statistics/
├── extraction_summary.md          # 本文件（汇总报告）
├── deepseek/
│   └── paper_statistics.csv      # DeepSeek 每篇论文的统计数据
├── gemini/
│   └── paper_statistics.csv      # Gemini 每篇论文的统计数据
└── kimi/
```

5. CSV 文件列说明

- **paper:** 论文名称
- **entities:** 实体数量
- **relations:** 关系数量
- **entity_types:** 实体类型数量（去重）
- **relation_types:** 关系类型数量（去重）