技术动态跟踪试验验证报告

项目名称：技术动态跟踪

测试阶段：验收测试

测试负责人：[张蕊保]

报告日期：2025年12月26日

一、测试概述

1.1 项目背景

随着化学化工领域文献数量的爆炸式增长，从海量非结构化 PDF 文献中高效、准确地提取结构化实验数据（如催化剂组分、反应条件及活性数据）成为化学信息学研究的关键。本项目旨在开发一套自动化的技术动态跟踪系统，实现对化学文献中复杂表格和文本数据的深度抽取。

1.2 测试目的

本测试旨在评估系统提取的结构化数据与人工标注数据（Ground Truth）的一致性，验证技术动态跟踪系统从非结构化数据中抽取关键化学信息的准确性。

1.3 测试范围

测试模块	包含功能点	是否测试
技术动态跟踪	领域覆盖指标完整性（合成氨催化剂：反应压力、反应活性、温度）、信息抽取准确性	是

二、测试环境与资源

2.1 硬件环境

环境类型	CPU型号	GPU型号/数量	内存	存储	其他硬件	备注
主测试环境	Intel(R) Xeon(R) Gold	NVIDIA H100（1块）	64GB	1TB SSD	GPU加速卡	无

	6430					

2.2 软件环境

环境类型	操作系统	核心依赖库/软件版本	备注
主测试环境	Windows 11	Python 3.13; information-extractor	无

2.3 测试数据

本次测试采用深度人工标注的基准数据集（Ground Truth），数据特征如下：

- 数据规模：覆盖 **18 篇** 催化领域核心论文，总计 **181 条** 实验记录。
- 核心字段清单：
  - 催化剂信息：催化剂名称
  - 反应条件：反应压力（P）、反应温度（T）。
  - 性能指标：催化活性（单位统一归一化为 $\mu mol \cdot g_{cat}^{-1} \cdot h^{-1}$ 或 $mmol \cdot g_{cat}^{-1} \cdot h^{-1}$）。

三、测试策略和方法

3.1 测试策略

测试策略：通过构建“原始PDF-助手输出-人工金标准（Ground truth）”对比的链路，验证系统对合成氨催化剂领域关键指标参数的提取准确率。

信息抽取指标计算：

- 准确率（Precision，P）：衡量提取结果的准确性。

$$P = \frac{TP}{TP + FP} \times 100\%$$

- 召回率（Recall，R）：衡量对原文信息的覆盖能力。

$$R = \frac{TP}{TP + FN} \times 100\%$$

参数定义如下：

1. 真阳性（True Positive，TP）：模型抽取的属性值于Ground Truth完全一致。
2. 假阳性（False Positive，FP）：模型抽取了错误的信息。包括：
   a. 幻觉提取：原文中不存在该数值，但模型输出了结果。
   b. 错误关联：提取的数值正确，但将其关联到错误的实体。

c. 格式错误：数值正确但量纲/单位转换错误。

3. 假阴性（False Negative, FN）：原文中存在关键技术指标，但模型未能抽取（漏抽）。

接口是否开发可以被外部调用：curl -X POST http://192.168.120.2:8080/task -F "files=@{文件绝对路径}"

示例：curl -X POST http://192.168.120.2:8080/task -F "files=@C:\Users\Administrator\Downloads\3\1\Fang 等 - 2022 - Combining molybdenum carbide with ceria overlayers.pdf"

3.2 测试方法及工具

测试类型	测试方法	使用工具	领域专属测试指标
技术动态跟踪展示准确性测试	标准数据对比法	curl，pandas，excel,apifox	抽取技术动态数值正确率>70%

四、测试执行情况

4.1 测试用例执行统计

测试模块	计划用例数	实际执行用例数	通过数	失败数	阻塞数	通过率
技术动态跟踪	33	33	33	0	0	100%

4.2 测试进度说明

- 计划测试周期：2025/12/03~2025/12/25
- 实际测试周期：2025/12/03~2025/12/25

五、测试结果

5.1 主要指标

测试模块	准确率	召回率
技术动态跟踪	**72%**	**73%**

接口已提供并可以正常调用获取响应结果

5.2 性能测试结果

模块	单用户访问平均响应时间（s）	10 并发用户访问平均响应时间（s）	CPU 使用率峰值（%）	内存使用率峰值（%）	响应成功率	性能要求（是否达标）
技术动态跟踪	35	100	15.5	20	100%	[达标]

六、测试结论与建议

6.1 测试结论

经过本次对“物质科学助手”技术动态跟踪模块的验收测试，得出以下结论：

1. 功能达标情况：技术动态跟踪模块在合成氨催化剂领域的关键指标提取（催化剂名称、压力、温度、活性指标）表现稳定。实测准确率（72%）与召回率（73%）均超过了预设的 70% 指标阈值，满足项目验收的基本要求。
2. 系统性能表现：在 NVIDIA H100 硬件加速下，系统具备较强的并发处理能力。虽然单用户响应时间为 35s，但在 10 并发压力下系统仍能保持 100% 的响应成功率，且资源占用（CPU 15.5%, 内存 20%）处于健康区间，具备后续扩展示验规模的硬件冗余。
3. 接口可用性：系统已成功开放 RESTful API 接口，经 `curl` 命令验证，外部调用链路通畅，能够支持下游业务系统的集成需求。

综合评价：该项目模块已通过验收测试准则，系统运行稳定，数据抽取质量符合预期。

6.2 改进建议

1. 强化单位转换与后处理逻辑：对抽取到的活性数据进行单位自动化校验与标准化转换。
2. 提升算法性能：准确率 (72%) 领先阈值较小，需进一步提高算法的性能。