

# 基于 GAKG 知识图谱的 Acemap 搜索增强系统

作者：张亦弛 (523030910235)

## 1. 任务与数据

本项目构建一个智能 **Search Agent**，旨在解决 Acemap 现有系统在长难句理解、拼写纠错及跨语言检索方面的不足，实现从自然语言到精准结构化查询的转化。

- 数据源**：Acemap API (提供基础检索服务) 与 **GAKG 知识图谱** (含 ~165 万地学实体，作为权威术语库进行 Grounding)。

## 2. 方法架构 (LLM + KG 协同增强)

系统采用流水线架构，包含三个关键步骤：

- 意图识别 (Intent Recognition)**：利用 LLM (Qwen-Plus/GPT-4o) 提取关键词、年份等要素。**创新点**：在 Prompt 中植入翻译指令，强制将中文口语转化为英文学术术语 (如“板块构造” → “Plate tectonics”)。
- 知识校准 (Knowledge Grounding)**：利用 `fuzzywuzzy` 算法将提取词与 GAKG 词表匹配。当相似度 > 85% 时触发机制，自动将拼写错误 (如 *Grnite*) 修正为标准术语 (*Granite*)。
- 查询生成与后处理**：生成标准化 API 参数，并在客户端执行逻辑过滤 (如 `institute=MIT`, `Year>=2023`) 以确保结果精度。

## 3. 实验评估

基于测试集，系统在多类场景下均显著优于原始搜索，典型案例对比如下：

评估场景	测试 Query	原始搜索表现	Agent 增强后效果
3.1 拼写纠错	recent papers about <b>Grnite</b>	结果含大量噪音，Precision 低	成功映射为 <i>Granite</i> ，召回量 <b>6543</b> → <b>10000+</b> ，Top-5 均为核心文献
3.2 跨语言检索	帮我找关于 <b>板块构造</b> 的论文	仅返回 0 篇或少量中文论文	自动译为 <i>Plate tectonics</i> ，召回 <b>3031+</b> 篇高质英文文献，打通中英壁垒
3.3 复杂逻辑	Find papers on Basalt <b>from 2023</b>	缺乏时间约束过滤	准确识别时间约束，通过客户端过滤将时间准确率提升至 <b>100%</b>