

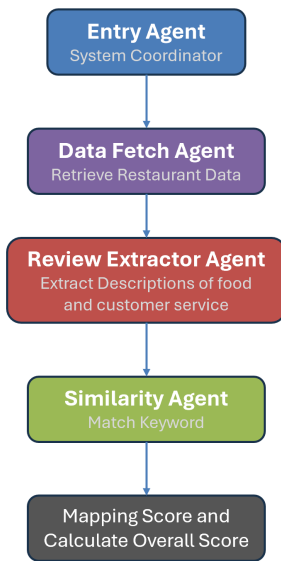
# Report

HW3 (Multi-Agent Restaurant Rater)

R13946016 陳采妍

## Structure

### Pipeline overview



### Agent Structure

1. **Entry Agent** - 管理整個流程的中心協調者
2. **Data Fetch Agent** - 從文字檔中抓取餐廳評論
3. **Review Extractor Agent** - 分析評論以提取食物和服務描述
4. **Similarity Agent** - 將提取的描述與預定義的評分關鍵字匹配

### Key Functions

- `fetch_restaurant_data()` - 通過餐廳名稱檢索餐廳評論
- `calculate_overall_score()` - 使用幾何平均值計算最終評分
- `get_score_from_adjectives()` - 將描述映射到數字分數 (1-5)
- `string_similarity()` - 使用SequenceMatcher檢查文本相似性

### Workflow Process

1. 使用者查詢被發送到Entry Agent

2. Entry Agent委託Data Fetch Agent檢索評論
3. Review Extractor Agent處理每個評論以找到食物/服務描述
4. Similarity Agent將描述映射到預定義的評分關鍵字
5. 系統計算數字分數和最終評分

## Prompt Design Analysis

### Data Fetch Agent

使用最小的提示，僅專注於生成正確的函數調用 JSON 格式：

```
{
  "call": "fetch_restaurant_data",
  "args": {
    "restaurant_name": "<name>"
  }
}
```

### Similarity Agent

使用詳細的提示，包含：

- 清晰的職責定義 ("professional word similarity analyzer")
- 步驟列表，逐步處理
- 明確的JSON輸出格式要求
- 處理不清楚的匹配時的回退指令

### Review Extractor Agent

有最複雜的提示，包含：

- 特定的目標描述示例（形容詞、副詞、短語）
- 強大的反幻覺保護 ("DO NOT generate or infer...")
- 對否定詞的特殊關注
- 詳細的輸出格式

## Discussion of Success and Failure Cases

- Success cases:
  - 當評論中的描述詞直接對應到評分關鍵字時，系統能夠準確評分
  - 當描述詞與關鍵字有高度相似性時，系統能透過相似度分析找到對應評分

- 系統能夠處理多條評論並整合評分
- Failure cases:
  - 某些描述詞需要上下文才能理解其真實含義，可能導致誤判
  - 當評論使用非標準或模糊的描述詞時，可能無法準確匹配

## Optimizations and Improvements

- Agent 分工更細
  - Baseline: 直接從評論提取分數
  - 優化版: 提取評論描述詞 → 取得相似關鍵字 → 對照出成績
- Prompt 細節
  - Ex. 僅擷取 review 中有的內容、處理否定詞等
- 流程更細
  - 將每條評論分次處理、並作錯誤驗證
  - 若評論描述詞已在分數對應表中，就不需再取得相似字
- 錯誤/驗證機制
  - 驗證
    - 驗證 Agent 回傳的格式是否符合期待
    - output 與 input 內容，進行相似性比對
  - 3次重試機會
  - (logger 日誌彈性處理錯誤訊息)

## Screenshots of all passed cases

```
[V] Test 0 Passed. Pyautogen 0.9.0
[V] Test 1 Passed. Expected: 3.000 Predicted: 2.982 Query: How good is the restaurant taco bell overall?
[V] Test 2 Passed. Expected: 9.350 Predicted: 9.376 Query: How good is the restaurant Chick-fil-A overall?
[V] Test 3 Passed. Expected: 8.060 Predicted: 8.060 Query: What is the overall score for Starbucks?
[V] Test 4 Passed. Expected: 9.540 Predicted: 9.423 Query: What is the overall score for In-n-Out
[V] Test 5 Passed. Expected: 3.650 Predicted: 3.671 Query: What is the overall score for McDonald's?
----- 5/5 Tests Passed. MAE: 0.0364 -----
```