KARMA: 후속 연구

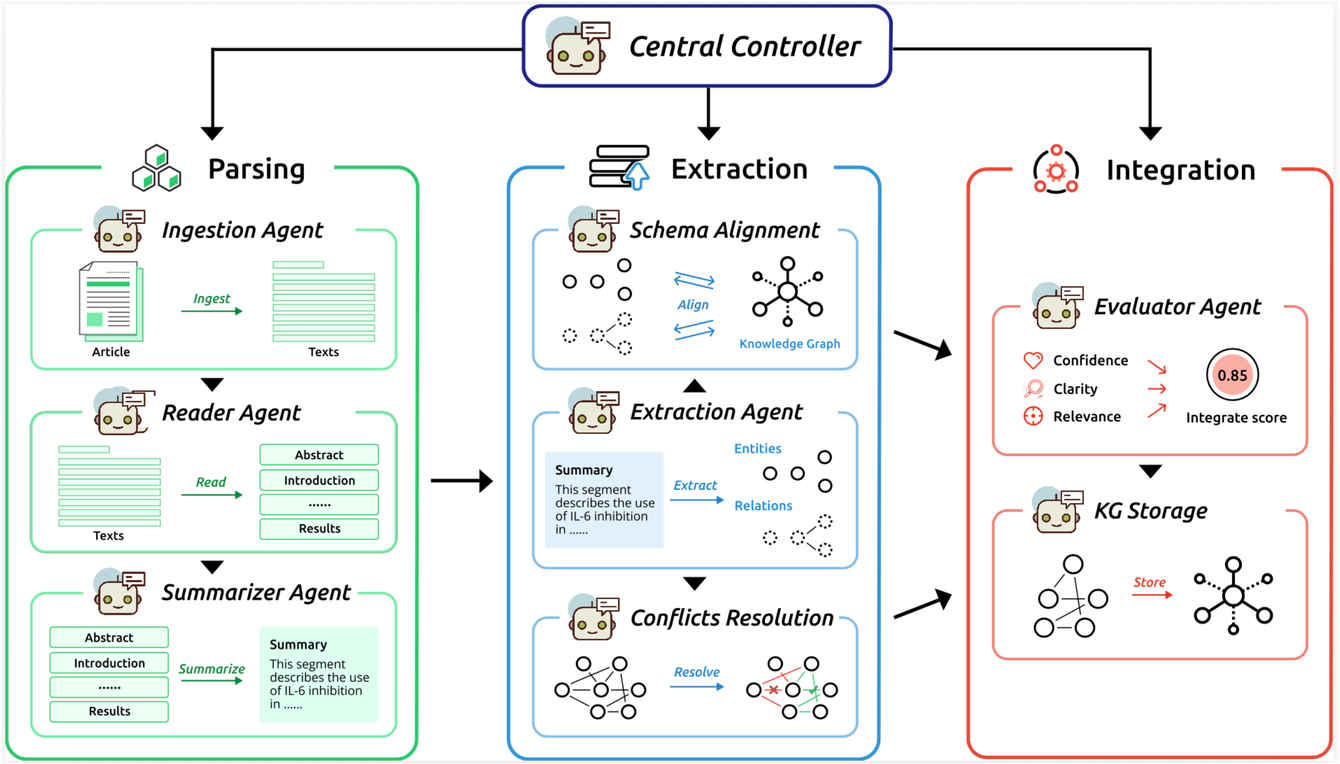
Yuxing Lu(PKU, Georgia Tech) | Wei Wu(PKU) | Xukai Zhao(Tsinghua) | Rui Peng(PKU) | Jinzhuo Wang(PKU, †)
NeurIPS 2025

지식 그래프의 지식 확장을 위한 Multi-Agents 기반 프레임워크

DeepShark Lab 학부연구생 김승겸

KARMA 개념 회상

지식 추출 및 통합을 위한 9개의 특화된 LLM 기반 에이전트 협업 시스템

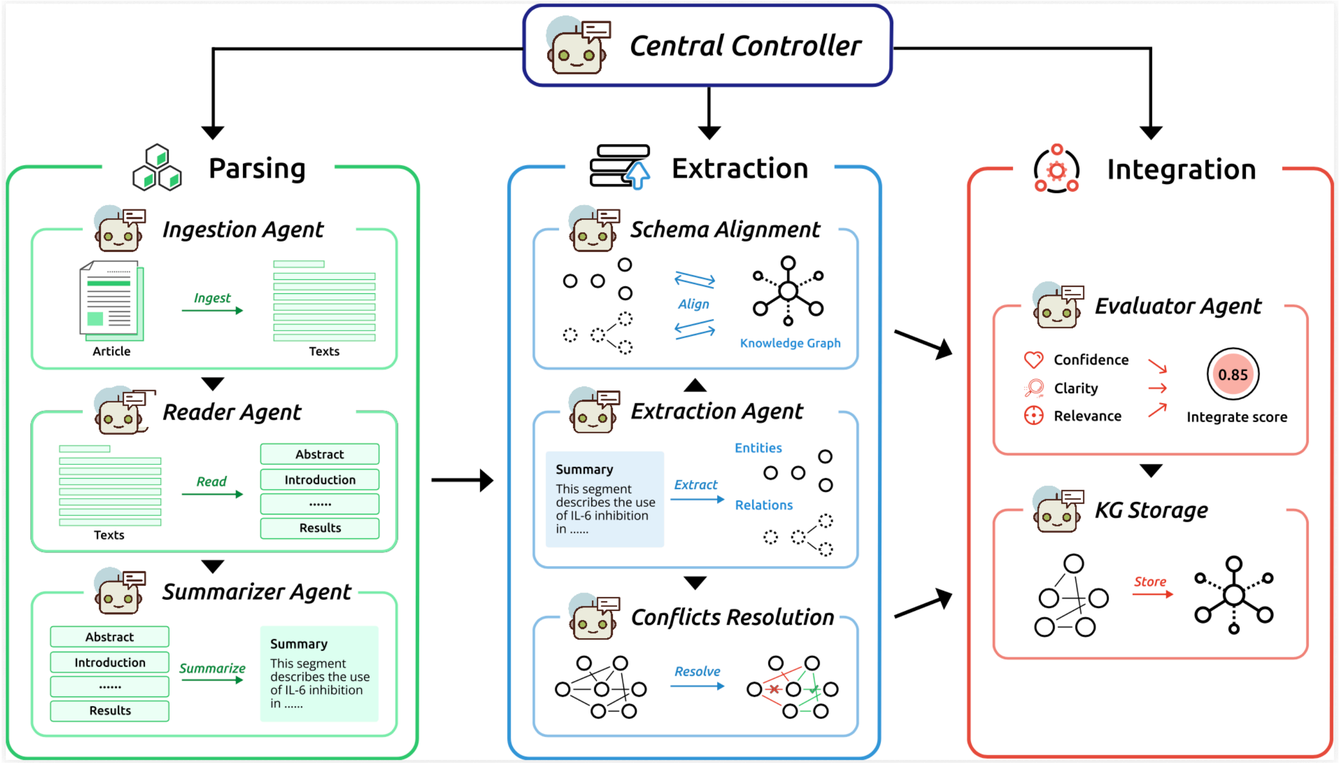


Knowledge-Aware Reasoning with Multi-Agents

목차

KARMA 프레임워크 구성

핵심 기능 분석



⚙️Base Agent

모든 에이전트의 부모 클래스

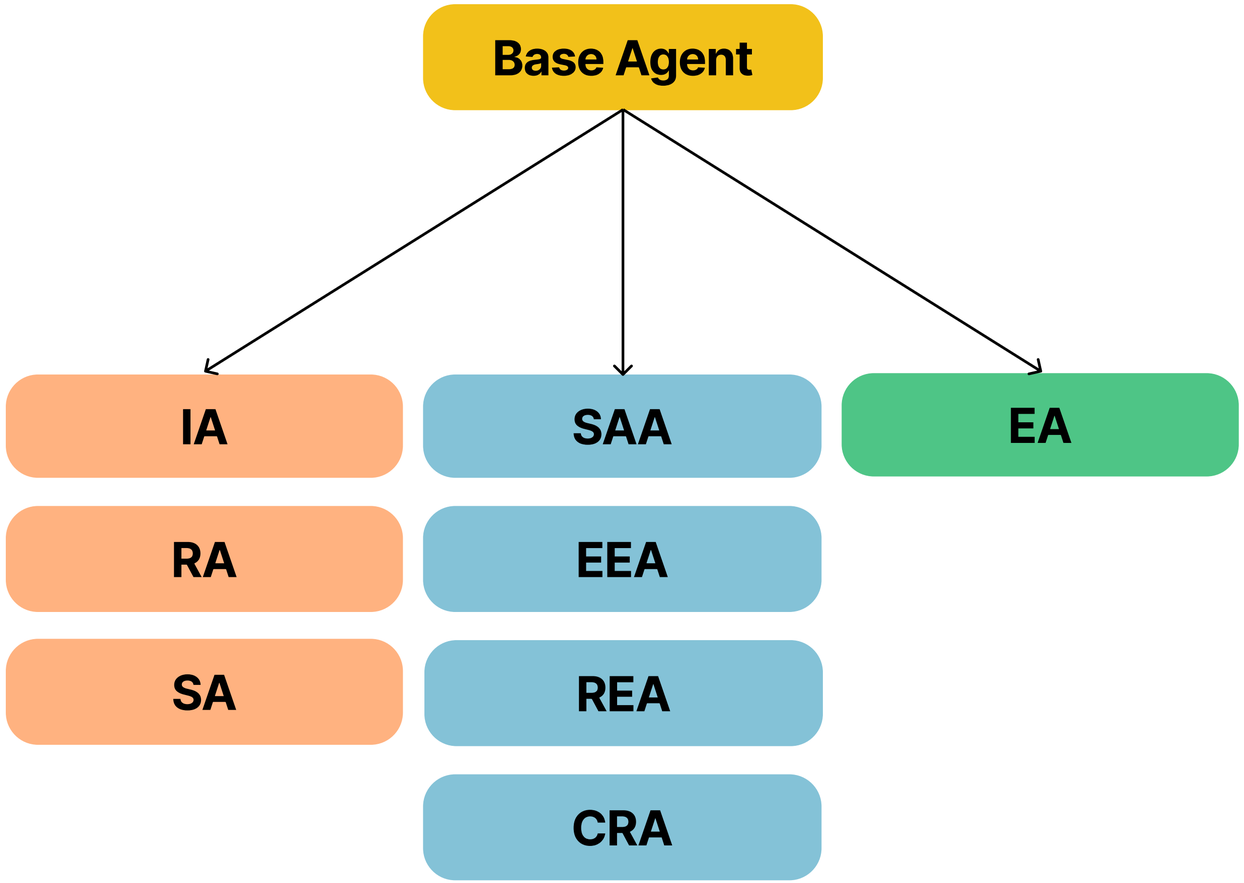
- KARMA 프레임워크 내 9개의 에이전트가 상속받는 추방 클래스
- LLM 호출, 성능 모니터링, 데이터 파싱

표준화된 LLM 통신 로직

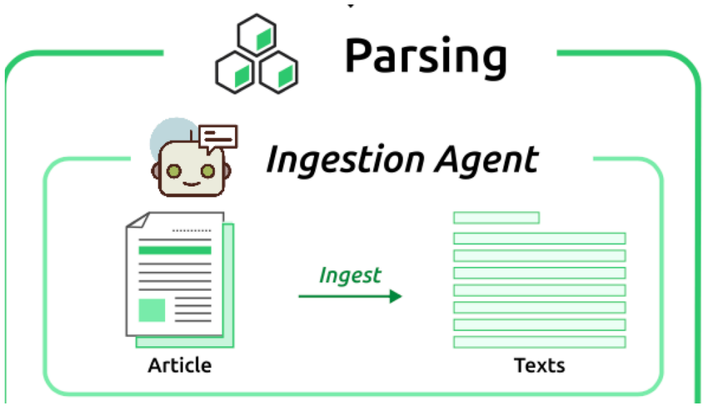
- OpenAI API 호출 과정을 캡슐화

추상 메서드 process() 제공

- 다형성을 보장하여 각 에이전트 자신만의 process() 로직 실행



📋Ingestion Agent



BaseAgent 상속

- API 호출, 토큰 계산, 에러 핸들링은 부모 클래스를 통해 수행
- IngestionAgent는 비정형 문서(논문)를 정형화된 데이터 구조로 변환

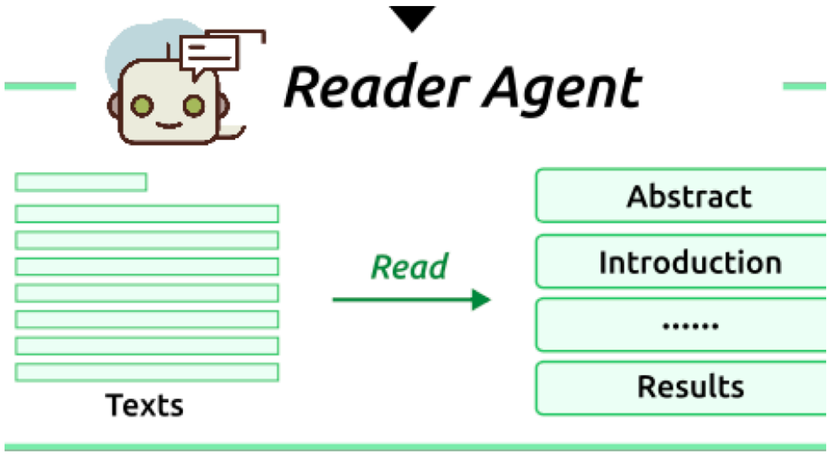
추출 및 정규화

- 비정형 문서를 가져와 표준 포맷으로 변환
- 제목/저자/DOI 등 핵심 메타데이터 추출
- OCR 과정에서 발생하는 글자 깨짐 및 특수 기호 보정

환각 억제

- 정보가 없을 때 LLM이 임의로 데이터를 지어내는 것을 예방하기 위해 “Unknown”, “N/A”를 사용하는 방어 규칙 삽입

Reader Agent



문서 Segmentation

- 정규화된 텍스트에서 불필요한 정보 제거 후 유의미한 내용 식별
- 정규표현식 기반으로 각 텍스트 패턴을 분석하여 섹션 정보 식별
- 제공된 KARMA는 Biomedical 문헌 분석에 특화

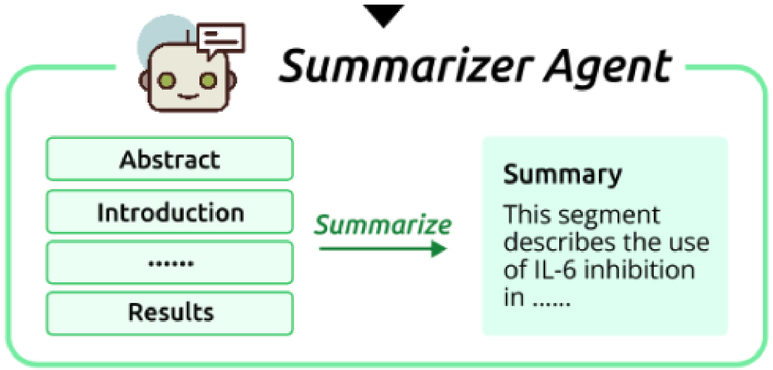
관련성 점수 산정

- Biomedical 지식(개체 간 관계 등) 추출 가능성 평가(0.0~1.0)
- 높은 관련성(0.8~1.0)
- 중간 관련성(0.4~0.7)
- 낮은 관련성(0.0~0.3)

필터링

- 임계값 미만의 세그먼트 제외
- 저자는 0.2로 임계값 설정

📋Summarizer Agent



핵심 요약문으로 변환

- 높은 관련성의 텍스트 세그먼트를 핵심 요약문으로 변환
- 100 단어 이내의 정보 밀도가 높은 요약 생성
- 문장 간 인과 관계 및 논리적 흐름 보존

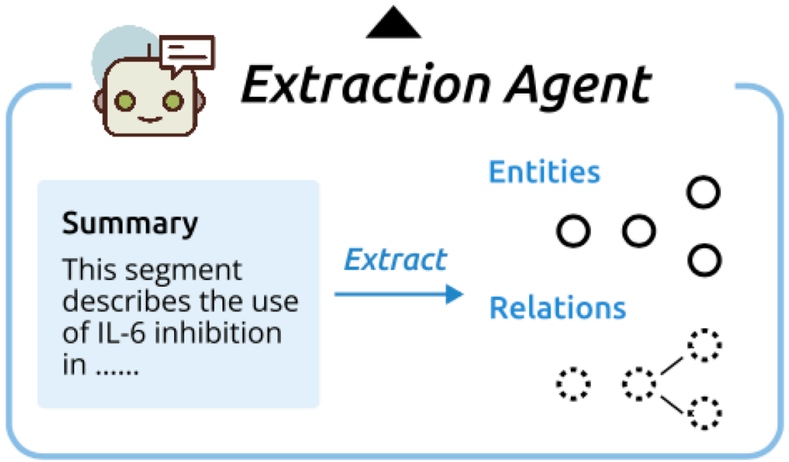
데이터 무결성 보존(feat. Biomedical)

- 생물학적 개체 (유전자, 약물, 질병, 화학 물질 등)
- 정량적 데이터 (농도, 퍼센트, 복용량 등)
- 통계적 지표 (p-value, 95%CI, Fold-change 등)
- 관계 지표 (활성화, 억제, 치료 등의 동사)

필터링

- relevance_threshold 0.2 미만 세그먼트 제외
- 정밀도 중심으로 Temperature 0.2 설정

📋Entity Extraction Agent



핵심 개체 식별 및 분류

- Biomedical 논문에서 지식 그래프의 노드가 될 엔터티 식별
- DRUG, DISEASE, GENE/PROTEIN 등
- _deduplicate_entities: 문서 전체에서 언급된 동일 개체를 통합
- KGEntity 리스트 출력

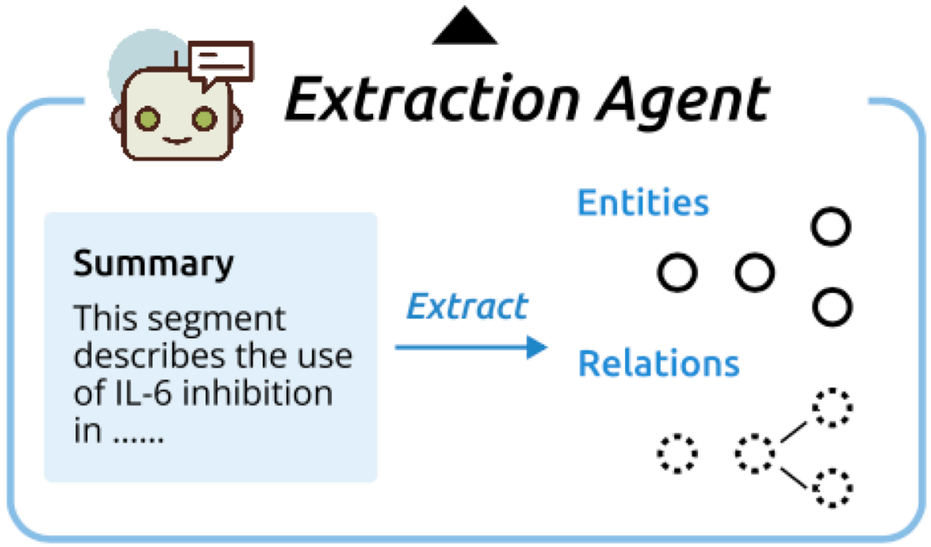
온톨로지 매핑

- 고유 식별자 연결
- 프롬프트 수준에서 사실 기반 정밀한 매핑하여 JSON 출력
- {“mention”: “Aspirin”, “type”: “Drug”, “normalized_id”: “MESH:D001241”}

이중 추출

- LLM 호출 실패 시 정규표현식 기반 Fallback 로직 실행
- 바이오 용어 특유의 명명 규칙(패턴)을 활용하여 최소한의 핵심 정보 보존

📋Relationship Extraction Agent



추출된 개체 간의 의미론적 관계 식별 및 분류

- 두 개체 사이의 관계를 Triple(머리-관계-꼬리) 구조로 변환
- _extract_relationships_from_text를 통한 LLM 추론
- 추출된 개체가 실제 목록에 존재하는지 확인(_entity_exists)
- 중복 관계 제거 및 신뢰도 기반 병합

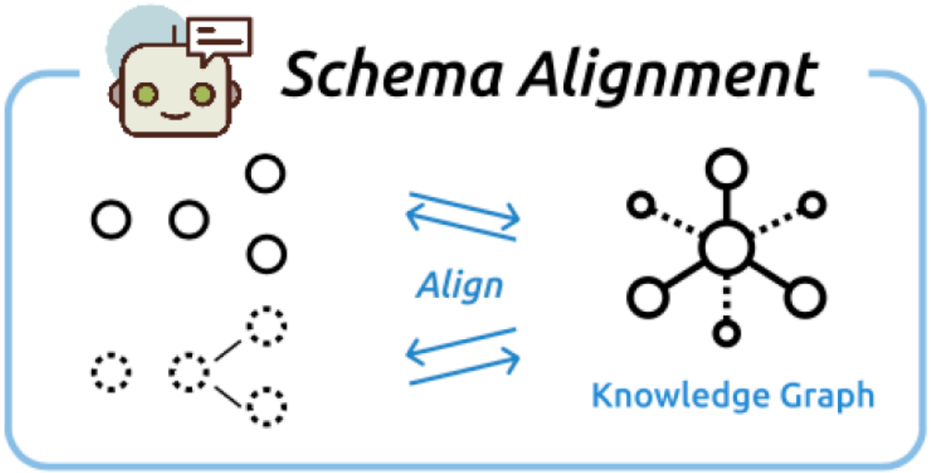
관계 타입 분류

- 9가지 관계 타입 정의
- 치료 및 예방 (TREATS), 기능 조절(INHIBITS, ACTIVATES, REGULATES), 인과 및 상관(CAUSES, ASSOCIATED_WITH) 등

품질 관리

- 자신이 추출한 결과에 대해 스스로 신뢰도 점수 산정(추측인지 명확한 인과적 표현인지)
- 자가 평가 지표: Clarity(관계 타입의 명확성), Relevance(주제와 추출한 관계의 관련성)

📋Schema Alignment Agent



지식 그래프 내의 개체 타입과 관계 레벨 표준화

- 서로 다른 표현들을 통일된 온톨로지와 명명 규칙에 맞게 매핑
- “inhibits”, “inhibiting”, “inhibited” 등 다양한 시제의 관계 표현 통일
- 타입이 지정되지 않은 (Unknown) 개체에 대한 분류

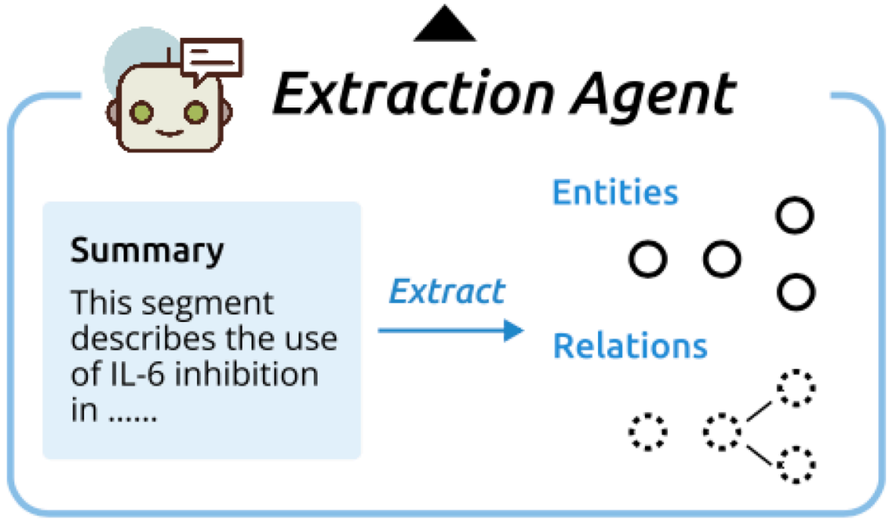
개체 분류

- 이 단어의 정체가 무엇인가?
- 개체의 타입을 표준 카테코리로 재분류
- Drug, Disease, Gene, Protein, Chemical, Pathway, Anatomy
- 단어 끝에 ‘-stain’ → Drug, ‘-ase’ → Protein

관계 정규화

- 둘 사이의 관계를 어떻게 부를 것인가
- 관계 술어를 정형화된 형태로 변환
- ex) “inhibits”, “inhibiting”, “inhibited” → “inhibit”

📋Conflict Resolution Agent



서로 모순되는 지식 Triple 탐지 및 해결

- 서로 다른 논문이나 소스에서 상반된 연구 결과가 보고될 경우 필요
- 증거 기반의 의사결정을 통해 지식 그래프의 데이터 무결성 확보

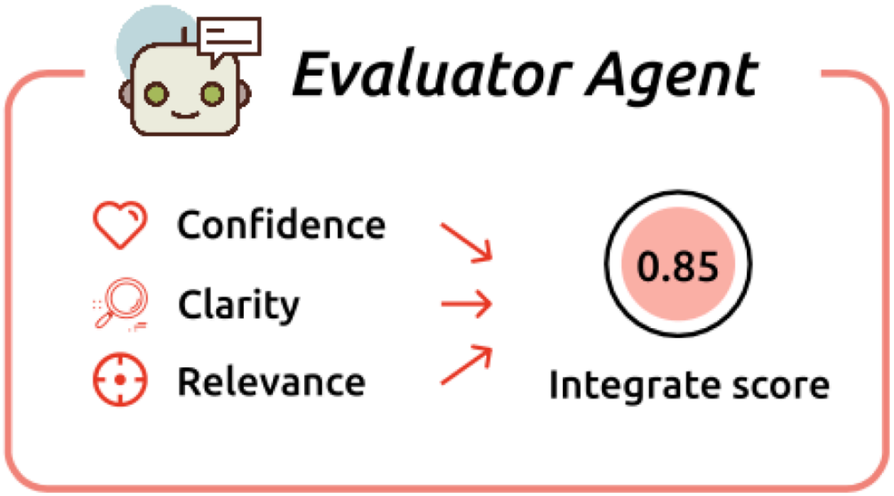
충돌 유형

- **직접적 모순**: Drug A **TREATS** Disease B vs **CAUSES** Disease B
- **의미론적 충돌**
  - 발현 방향의 충돌(Increases vs Decreases)
  - 조절 효과의 충돌(Upregulates vs Downregulates)

해결 전략 및 우선순위

1. 관계의 신뢰도 점수
2. 증거의 품질(얼마나 구체적인지)
3. 관계 기술의 구체성
4. 정보의 최신성
5. 기존 지식 체계와의 일관성

Evaluator Agent



추출된 지식 트리플의 최종 품질 평가 및 통합 여부 결정

- 가중치 기반의 통합 점수 산출
- 설정 가능한 임계값을 통한 데이터 품질 제어
- 투명한 의사결정 프로세스 및 품질 기준 준수

3대 품질 평가 지표

1. CONFIDENCE (가중치 50%)
2. CLARITY (가중치 25%)
3. RELEVANCE (가중치 25%)

통합 스코어링 및 결정 로직

- IntegrationScore = (0.5 x Confidence) + (0.25 x Clarity) + (0.25 x Relevance)
- 임계값 = 0.6
- EXCELLENT (>= 0.8): 고품질, 강력한 근거 확보
- GOOD (0.6 ~ 0.79): 신뢰 가능, 마이너한 한계 존재

KARMA 예제 논문 실험

J Appl Physiol 123: 1610–1616, 2017.
First published July 13, 2017; doi:10.1152/japplphysiol.01119.2016.

RESEARCH ARTICLE

Aspirin as a COX inhibitor and anti-inflammatory drug in human skeletal muscle

Stephen M. Ratchford, Kaleen M. Lavin, Ryan K. Perkins, Bozena Jemiolo, Scott W. Trappe, and Todd A. Trappe

Human Performance Laboratory, Ball State University, Muncie, Indiana

Submitted 22 December 2016; accepted in final form 8 July 2017

예제 논문 실험 결과


```
"entities": [
  "sarcopenia",
  "skeletal muscle",
  "Aspirin",
  "PGE2",
  "aspirin",
  "PGE2/COX pathway",
  "muscle inflammation",
  "COX"
 ]
```

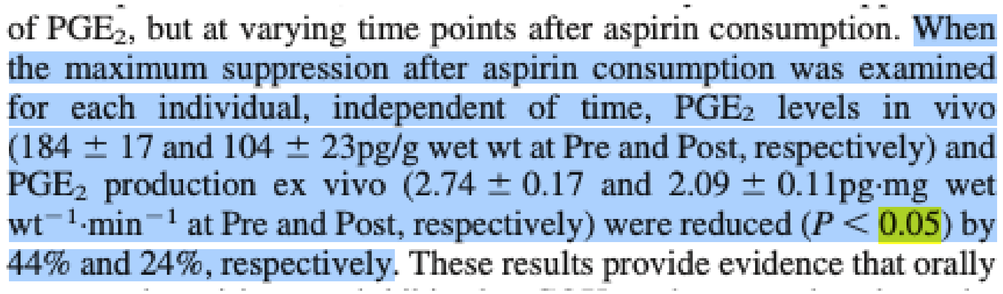
```
  {
    "head": "aspirin",
    "relation": "inhibits",
    "tail": "COX",
    "confidence": 0.85,
    "source":
 "relationship_extraction",
    "relevance": 0.7,
    "clarity":
 0.8999999999999999
  }
```

```
{
    "head": "aspirin",
    "relation": "decreases",
    "tail": "PGE2",
    "confidence": 0.9,
    "source":
 "relationship_extraction",
    "relevance": 0.5,
    "clarity": 0.7
  }
```

```
{
    "head": "aspirin",
    "relation": "associated_with",
    "tail": "muscle inflammation",
    "confidence": 0.75,
    "source":
"relationship_extraction",
    "relevance": 0.5,
    "clarity": 0.6
  }
```

```
{
    "head": "aspirin",
    "relation": "associated_with",
    "tail": "sarcopenia",
    "confidence": 0.7,
    "source":
"relationship_extraction",
    "relevance": 0.5,
    "clarity": 0.6
  }
```

예제 논문 실험 결과: KARMA의 정교성



of PGE_2, but at varying time points after aspirin consumption. When the maximum suppression after aspirin consumption was examined for each individual, independent of time, PGE_2 levels in vivo (184 ± 17 and 104 ± 23pg/g wet wt at Pre and Post, respectively) and PGE_2 production ex vivo (2.74 ± 0.17 and 2.09 ± 0.11pg·mg wet wt^{-1}·min^{-1} at Pre and Post, respectively) were reduced ($P < 0.05$) by 44% and 24%, respectively. These results provide evidence that orally

- **P**: 실험 결과가 우연히 나올 확률

→ 아스피린이 PGE2 수치를 감소시킨 것은 우연이 아니다

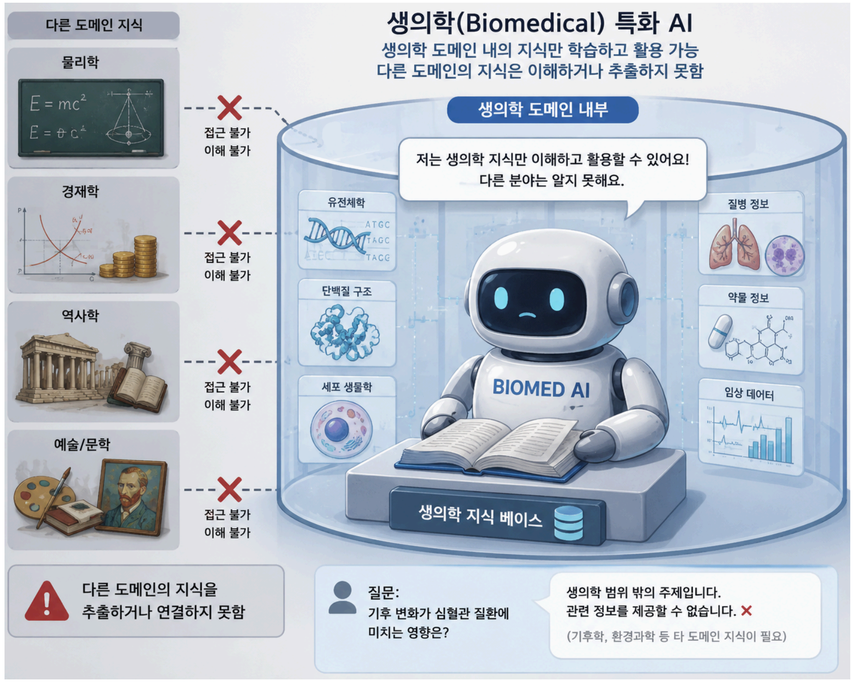

```
{
    "head": "aspirin",
    "relation": "decreases",
    "tail": "PGE2",
    "confidence": 0.9,
    "source":
"relationship_extraction",
    "relevance": 0.5,
    "clarity": 0.7
  }
```


- REA 에이전트: 통계 수치를 지식의 핵심 근거로 파악
- EA가 신뢰도를 0.9로 매우 높게 판정

KARMA 한계

OpenAI

- OpenAI API 의존성 및 비용 문제



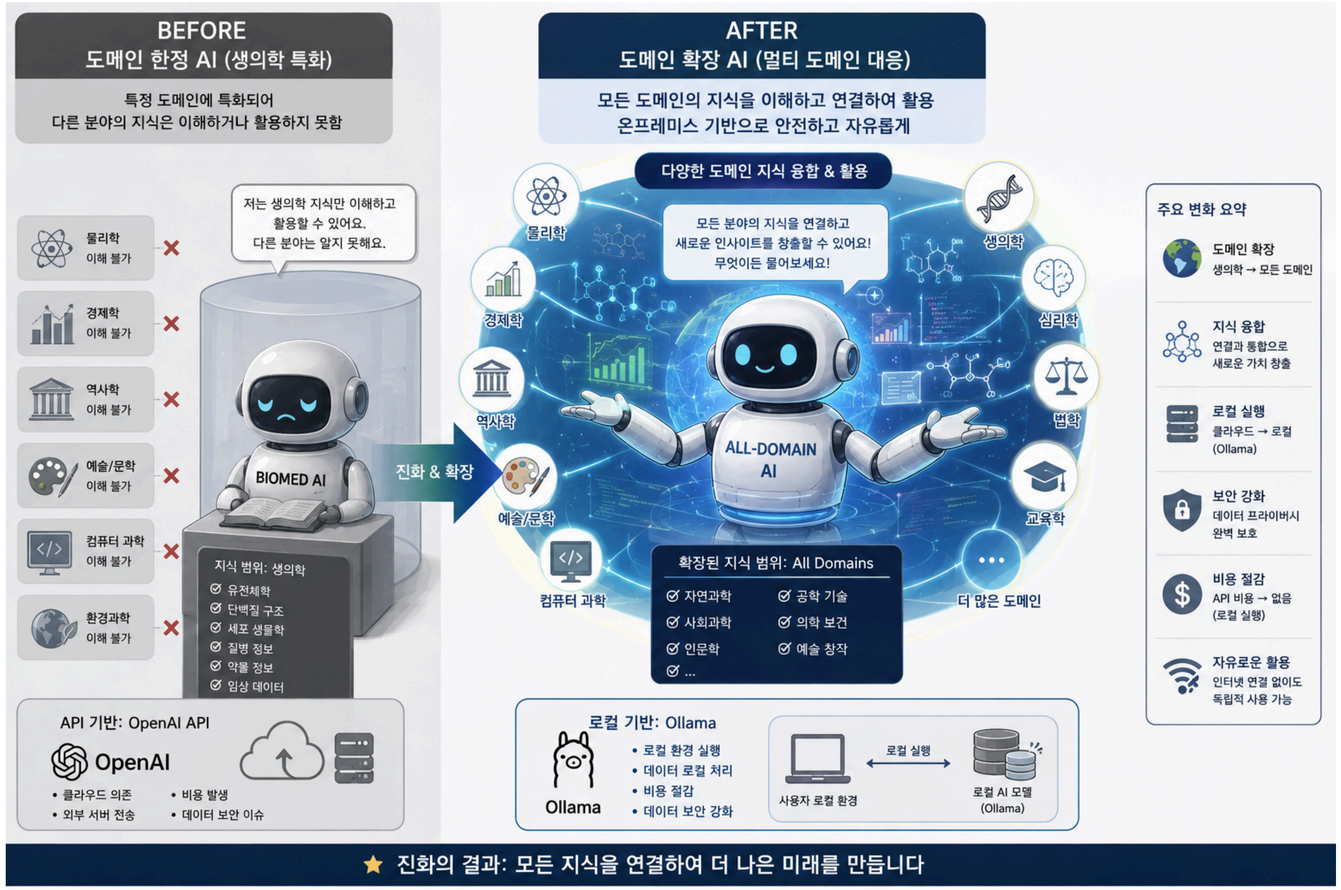
- Biomedical 도메인 한정

Open AI → Local LLM

- OpenAI 호환 Ollama3.1 사용
- 대량의 문헌 처리 시 발생하는 API 비용 문제 해결

Biomedical → {Domain}

- domain 파라미터 도입하여 도메인 확장성 확보



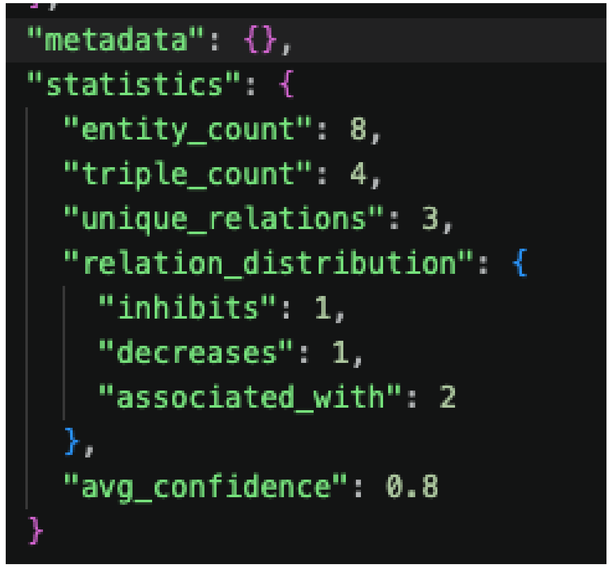
Ollama 기반 KARMA 결과 (Local LLM 구축)

광범위한 추출

- **비용 문제 해결**
- **재현율 확보**

트리플 증가 및 Overfitting 발생

- **관계 파편화**: 도메인 적응형 프롬프트 주입으로 인해 관계 생성 단계에서 파편화 발생
- **노이즈 발생**: 폭증한 엔터티로 인해 불필요한 트리플 발생



<기존 KARMA>

<Ollama KARMA>

Ollama 기반 KARMA 결과 (도메인 확장)

Advancements in Natural Language Processing:
Exploring Transformer-Based Architectures for Text Understanding

Tianhao Wu[1], Yu Wang[2], Ngoc Quach[2]*

엔터티 중복

- transformer-based architectures와 transformer 구별 불가
- 단순 문자열 비교의 한계

관계 표준화 부족

- 관계 파편화: 도메인 적응형 프롬프트 주입으로 인해 관계 생성 단계에서 파편화 발생


```
"metadata": {},
"statistics": {
  "entity_count": 119,
  "triple_count": 131,
  "unique_relations": 58,
  "relation_distribution": {
    "outperforms": 3,
    "uses": 10,
    "involves": 6,
    "achieves": 1,
    "has_impact_on": 2,
    "improved": 1,
    "builds_upon": 1,
    "influenced_by": 6,
    "extends": 4,
    "improves": 6,
    "relies_on": 4,
    "concerned_with": 1,
```


<도메인을 computer_science로 입력했을 때>

후속 연구 (최종 목표)

1. Ollama 기반 KARMA 고도화

- 도메인 적응형 추출 정교화
- 다중 문서 지식 통합
- 성능 지표 확립

2. 환각 제어 및 방지 연구

- GraphRAG 시스템 구현
- 추론 경로 검증
- 상충하는 지식 필터링

3. 판사 AI 구축

감사합니다

DeepShark Lab 학부연구생 김승겸