

## 관련 연구 자동 생성을 위한 LLM의 활용 및 정제 기법 제안

최승민\*, 정유철\*\*

### Proposal for the Utilization and Refinement Techniques of LLMs for Automated Research Generation

Seung-min Choi\*, Yu-chul Jung\*\*

**요약** 과거부터 꾸준히 지식 그래프(Knowledge Graph)와 언어 모델(LM, Language Model)의 통합에 대한 많은 연구가 이뤄지고 있다. 그 중, 지식 그래프의 구조화된 지식을 이용해 자동 텍스트 생성을 다루는 연구는 그리 활성화되지 않았다. 본 연구에서는 기존 논문들과 비슷한 수준의 특정 도메인 관련 연구 항목(Related Work)을 자동 생성하기 위한 방법론, 즉, '1) 최적의 Prompt 선정, 2) 4단계 정제기법을 통해 Triple 추출, 3) 지식 그래프 구축, 4) 관련 연구 자동 생성'을 제안한다. 제안된 방법론은 대규모 언어 모델(LLM) 중, GPT-4를 활용하고, 4단계 정제 기법을 적용하여 관련 연구를 자동으로 생성하도록 설계했다. 그렇게 설계된 모델은 Triple 추출에서 #Supp, #Cont, Fluency에서 17.3, 14.1, 4.2의 성능과 GPT-4 자동 평가 기준, 100점 기준 정제 전, 88.5점에서 정제 후, 96.5점으로 기존 논문과 비슷한 수준의 유의미한 관련 연구 자동 생성 능력을 보였다.

**Abstract** Research on the integration of Knowledge Graphs (KGs) and Language Models (LMs) has been consistently explored over the years. However, studies focusing on the automatic generation of text using the structured knowledge from KGs have not been as widely developed. In this study, we propose a methodology for automatically generating specific domain-related research items (Related Work) at a level comparable to existing papers. This methodology involves: 1) selecting optimal prompts, 2) extracting triples through a four-step refinement process, 3) constructing a knowledge graph, and 4) automatically generating related research. The proposed approach utilizes GPT-4, one of the large language models (LLMs), and is designed to automatically generate related research by applying the four-step refinement process. The model demonstrated performance metrics of 17.3, 14.1, and 4.2 in Triple extraction across #Supp, #Cont, and Fluency, respectively. According to the GPT-4 automatic evaluation criteria, the model's performance improved from 88.5 points before refinement to 96.5 points after refinement out of 100, indicating a significant capability to automatically generate related research at a level similar to that of existing papers.

**Key Words** : GPT-4, Knowledge Graph, Prompt, Refining Techniques, Triple Sentence Extraction

#### 1. 서론

지식 그래프(Knowledge Graph)와 언어 모델(LM)

은 상호 보완적 성질을 지닌다[1]. 지식 그래프의 구조화된 지식을 활용하여 언어 모델은 구조화된 지식 기반 텍스트를 생성할 수 있는데, ChatGPT와 같은 대규모

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(ITP-2024-RS-2022-00156394) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)\*

\*Department of Computer Engineering, Kumoh National Institute of Technology University

\*\*Corresponding Author : Department of Artificial Intelligence Engineering, Kumoh National Institute of Technology University (jyc@kumoh.ac.kr)

Received July 20, 2024

Revised August 19, 2024

Accepted August 26, 2024

모 언어 모델(Large Language Model, 이하 LLM)의 소개 이후 더욱 발전하고 있다. LLM을 활용한 정보 추출 및 지식 그래프 구축에 관한 연구는 여러 주제 도메인에서 광범위하게 이뤄졌다. 특히 해당 논문의 연구 지식을 지식 그래프화하는 [2], [3], [4] 연구들이 대부분인데, 본 논문에서는 관련 연구 (Related Work) 부분을 자동으로 지식 그래프화 하는 문제에 집중한다.

통상적으로 논문 내 관련 연구 항목에서는 '누가(제목, 저자 등), 무엇을, 어떻게'의 형식으로 타 논문의 기법, 기술이 서술된다. 이렇게 서술된 관련 연구 항목의 문장들은 논문 내의 제안 방법론, 이용된 기술, 연구 결과 등의 정보를 간략하지만, 정확한 내용을 포함하는 Triple(삼중문항)이 있다. 하지만, LLM에 추출용 Prompt를 적용하여 Triple 문장을 직접 추출한 결과는 해당 분야의 전문가가 수동으로 추출한 결과와 차이가 있다. 다양한 상용/오픈 LLM 있지만, 가장 성능이 우수하다고 알려진 GPT-4를 활용하는 것이 일반적이다.

특정 도메인에 대한 GPT-4의 정보 추출 성능은 [5], [6]과 같은 최근 연구에서 증명되었는데, 추출되는 지식 구조는 엔티티 사이의 관계를 구조화된 형태를 가지기에 특정 도메인의 분야에서 높은 성능을 보인다. 이러한 구조화된 지식은 본 연구와 같이 전문적 지식을 요구하는 부분에서 유의미한 결과를 나타냈다.

본 연구에서는 특정 주제 분야의 논문을 대상으로 주요 관련 연구들을 추출하여 지식 그래프 화하고 이를 이용하여 관련 연구 텍스트를 자동으로 생성하는 LLM의 활용 및 정제방안을 제안한다. 특히, LLM을 활용한 지식 그래프 구축에서 적용할 수 있는 Prompt 방법과 네 가지 정제 방법 (① 관계(Relation) 통제, ② 엔티티(Entity) 일반화, ③ 구조적 중복 제거, ④ 의미적 중복 최소화)을 제시한다. 일련의 정제 과정을 통해 추출된 Triple 문장을 기반으로 한 지식 그래프는 논문의 관련 연구 항목을 자동으로 생성하는 데 활용된다.

제안 기법의 적절성을 검증하기 위해 지식 그래프와 LLM 논문 각 30건을 분석하여 총 1,700여 개의 Triple 문장을 추출 후 정제 전/후에 대한 비교 실험 및 시각화를 진행하였다. 그 결과, 정제 후 Triple 문

장들 대상의 검증에서, 중요 기술 언급 점수는 평균 17.3 (Supp), 누락된 정보는 평균 14.1 (Cont)로 측정되었으며, 이로 구성된 지식 그래프는 노드 간, 연결성과 강도에 대해 더 유의미한 결과를 보였다. 이는 지식 그래프와 대규모 언어 모델 관련 논문들의 지식이 유의미하게 통합되어 관련 연구 항목을 자동 생성할 수 있음을 확인하였고, 생성된 관련 연구 항목 또한 정제 전, 후의 차이를 극명하게 보였다.

## 2. 관련 연구

지식 그래프에 자동 구축에 관련된 연구는 꾸준히 진행되었는데, [7] 연구에서 최근 언어 모델(LM)과 결합하여 추출 및 생성 작업하는 연구가 활발히 이어지고 있다. 특히, 거대 언어 모델을 활용하여 지식 그래프 구축을 위한 Triple 문장을 추출하는 [8]과 같은 연구들이 제안된 바 있다. 이 연구는 복잡한 텍스트 데이터로부터 관계와 개체를 효율적으로 식별하고 추출하는 능력을 개선하는 데 중점을 두고 있다.

개체명 인식(Named Entity Recognition, 이하 NER) 기술을 활용하여 문서 내 중요 개체를 인식하고 이를 기반으로 한 연구들도 활발히 진행되고 있다. [9], [10]의 연구에서는 포괄적인 도메인에서도 GPT-3.5는 평균 정확도 79점으로 유의미한 성능을 보여줬다. 이러한 접근 방식은 문서로부터 구조화된 정보를 추출하여 지식 그래프에 통합함으로써 [11] 연구에서는 다른 목적의 문서에서 추출된 지식의 표현력을 강화하고, 연구 목적에 맞는 문서를 생성하는 작업을 통해, GPT-NEO가 NER에서 좋은 성능을 보였다는 것을 간접적으로 증명한다.

### 2.1 LLM을 이용한 Triple 추출

여러 대규모 언어 모델(LLM)을 활용하여 지식 그래프(KG) 구축 및 추론 작업의 성능을 평가한 연구 중, [12] 연구에서는 특정 도메인에서의 지식 추출 능력이나 구조화 측면에서 GPT-3가 높은 성능을 보인다는 것을 연구에서 보였다. [13],

[14] 연구에서는 BioGPT 등 Fine-tuning 된 언어 모델을 사용하여 추출된 Triple 문장이 다양한 특정 도메인에서 Prompt의 기법 (Zero-shot, Few-shot 등)에 따라 Triple 추출 작업에서 점수(본 연구 기준 F1-macro)가 달라질 수 있음을 보였다.

지식 그래프의 불완전성 문제나 특정 도메인의 지식 추출을 해결하기 위해, [15], [16] 연구에서는 LLM을 활용한 지식 그래프 완성 방안을 제시했다. 지식 그래프의 노드와 엣지 사이의 관계를 학습하고 노드 사이의 잠재적 관계를 예측할 수 있는데, 이러한 잠재적 관계는 누락된 정보를 예측하므로 세부적인 정보를 지닐 수 있다. 이를 위하여 KG-LLM과 같은 모델을 개발하여 지식 그래프 특화의 LLM을 개발했다.

[11] 연구에서는 구조화된 정보를 추출하기 위해 미세 조정된 대규모 언어 모델을 이용해 구조화된 정보를 추출하였는데, Inverted Plot Pyramid(IPP) 구조와 5W로 뉴스 기사의 정보를 세분화하여 분류 및 해당 연구자가 설정한 IPP 구조로 진행된다. 이러한 방식은 문서의 자동 생성 및 Triple 추출에 체계적으로 원하는 정보를 표현할 수 있음을 알 수 있다.

한편, 지식 그래프를 자연어 문장으로 변환하고 이를 언어 모델의 사전 훈련 코퍼스에 통합하는 연구도 존재한다. [17] 연구는 지식 그래프의 구조적 차이를 극복하고, 지식 그래프의 통합을 통해 언어 모델의 정확성을 개선할 방법을 모색했다. 이러한 연구 [18]에서는 LLM의 정보 추출에 대한 일반화 능력을 탐구하고, 이를 위해 과학적 기사에서 사전 훈련된 언어 모델로부터 구조화된 정보를 직접 추출하는 작업인 Virtual Knowledge Extraction 작업을 제안하고 관련 VINE 데이터셋을 개발했다.

[19], [20]의 연구에서는 샘플데이터가 필요 없는 Zero-shot 학습 기법을 이용하여 클래스 예측을 위한 명시적인 예시 없이도 성능을 발휘할 수 있음을 보여줬다. 이 접근 방식은 특히 데이터가 부족한 상황이나 새로운 도메인에 적용될

때 유용하며, 지식 그래프와의 통합을 통해 언어 모델의 이해도와 추론 능력을 더욱 향상할 수 있다. Zero-shot 학습 기법을 통해 기존 지식의 정보를 기반으로 예측을 진행하여, 특정 도메인의 지식과 같이 한정된 지식으로도 다른 대규모 데이터 셋과 비슷한 이해도와 추론 능력을 지닐 수 있다.

## 2.2 객체명 인식 (NER)

지금까지의 객체명 인식, NER 기법(Named Entity Recognition) 연구는 다양한 도메인에서의 지식 그래프 구축과 밀접하게 연관되어 있다. 효율적이고 복잡한 특정한 구조에서 중요도가 높은 객체를 중심으로 추출하여 기용할 수 있는 NER기법은 지식을 인식하고 추출하는 것에 있어 매우 효과적인 것을 알 수 있다.

특히, [21], [22], [23] 연구에서는 NLP 분야의 지식 그래프를 중단 간 구축하는 새로운 접근 방법을 제시했다. 각각 신경망 아키텍처, NER을 span 예측으로 재구성하는 접근 방식, 외부 컨텍스트 검색과 협력적 학습을 통해 NER 성능을 향상한다. 이 연구들은 복잡한 데이터에서 효율적인 NER을 통해 지식 그래프를 구축하는 방법을 연구했다.

[18] 연구는 NER에 대해서 과거와 비교해 신경망 모델로 더 향상된 성능을 서술한다. 이러한 NER은 향상된 성능을 보이며 과거의 시스템을 통합한다면 더 좋은 성능을 낼 수 있음을 보여준다.

또한, [24] 연구는 법률 텍스트에서의 NER 작업을 보여준다. 수동으로 주석이 달린 데이터셋과 특정 컨텍스트(맥락) 내에서 필요한 정보를 바탕으로 학습하는 방식을 데이터 셋에 실험을 적용하여 제시한다. GPT-4로 추출한 이러한 지식 정보들은 완벽하지는 않지만, NER 작업에 다른 연구에서 소개된 모델과 비교했을 때, CL-KL+SEMI 모델이 외부 문맥(W/CONTEXT)으로 평가했을 때, 84.10으로 좋은 성능을 보였다.

### 3. 자동 지식 그래프 구축

논문의 관련 연구 항목을 자동 생성하기 위해 특정 도메인(KG, LLM) 분야의 관련 연구 항목이 존재하는 기존 논문들을 분석하여 지식 그래프를 구축하는 과정을 구성한다.

GPT-4를 이용한 Triple 문장 추출에 최적인 Prompt를 선정하고, 해당 Prompt와 논문들을 입력으로 GPT-4를 이용한 Triple 문장 추출, 그리고 일련의 정제 과정을 거쳐 최종 Triple 문장들을 획득하게 된다. 자동 지식 그래프 구축의 실험에서는 LLM과 KG 분야의 최근 5년 이내의 논문 중에 연구 분야 항목이 존재하는 논문 각 30건, 총 60건을 수집했다.

#### 3.1 Triple 추출 및 정제

논문 내의 관련 연구 항목을 자동 생성을 위한 지식 그래프 구축을 위해, 특정 도메인의 논문들을 대상으로 관련 연구 항목에서 Triple sentence sample을 추출한다. Triple 문장은 (Entity1, Relation, Entity2)와 같은 구조를 지니고 있다.

위 구조에서 Entity1은 논문에서 나타나는 기법이나 기술을, Entity2에는 Entity1의 서술 정보를 담고 있는데 이러한 구조는 논문의 지식에서 Entity를 설명하기 위한 표현이다.

##### 3.1.1 추출 Prompt 선정

수집한 논문의 관련 연구 항목에 서술된 내용에서 GPT-4를 사용해 각 논문의 Triple 문장을 추출했다. 지식 그래프를 구성하기 위한 Triple 문장을 추출하기에 목적과 기준에 맞게 추출되어야 한다.

추출을 위해 단순한 추출 요청 prompt를 사용하는 경우, GPT-4는 다음과 같은 Triple 문장들이 추출된다.

(Gustav Freytag, outlined, Plot Pyramid Model)  
 (Plot Pyramid Model, includes, Introduction)  
 (Fichtean Curve, begins with, Rising Action)

그림 1. 일반적인 추출 Prompt 적용 시  
 Fig. 1. Application of Generic Extraction Prompts

하지만, 추출된 Triple 문장들은 상세함이 다소 부족하며, 단어로 끊어져 있어서 지식 그래프의 목적보다는 요약문의 형태와 비슷하다. 연구의 목적과 기준에 맞는 Triple 문장을 추출하기 위해서는 적절한 Prompt 기법이 요구된다.

본 실험에서 Chain of Thought(CoT)[25], Zero-Shot[19] 등의 Prompt 기법을 시도해 본 결과, 원문에서 지식을 추출하기에는 Instruction-based Prompting이 적합하다고 판단하였으며, 이러한 실험은 [표1]에서 확인할 수 있다. 아래는 Triple을 추출할 때의 예시이다.

- Extracting entities that contain information about techniques or challenges mentioned in the target paper, not just the authors or publication details.

- Providing a detailed explanation, not just a brief description, of predicates (entity2) related to the subject (entity1) in triples.

각각의 Triple 문장은 (기법이나 모델, 관계, Entity1에 대한 서술)을 가지고 있으며 문장을 이룰 수 있는 자세한 정보를 가져야 한다. 위 서술된 조건을 만족하는 Triple 문장은 각 특정 도메인 당, 약 800개의 Triple 문장을 얻을 수 있었다.

표 1. Prompt 별 추출 Triple 비교 및 특징  
Table 1. Comparison and Features of Extracted Triples by Prompt

Selected Prompt	Extracted Triples	Remarks
<p>Extract sentences that meet the following conditions: 1) Each sentence should be formatted as (Entity1, Relation, Entity2). 2) The grammar of the Relations should maintain consistent word use and temporal expression. 3) Entity1 should include the techniques and methods used, and Entity2 should contain the object (information or function) of Entity1. 4) ...(Omitted)</p>	<p>1. (GPT-3, contributes to, development of LLMs capable of diverse NLP tasks) 2. (ChatGPT, utilizes, RLHF for enhancing alignment with human preferences) 3. (GPT-4, processes, both image and text inputs) ...(Omitted)</p>	<p>Useful for structuring relationships or information, with Triple sentences specifically expressed.</p>
<p>These triples are extracted to highlight the techniques, strategies, and features described in the text for the development of large language models, knowledge graph construction, and application to other NLP tasks ...(Omitted)</p>	<p>1. (GPT-3, led to the development of, highly capable LLMs) 2. (ChatGPT, utilizes, RLHF during fine-tuning) ...(Omitted)</p>	<p>There are some limitations in indicating direct influence or dependency between studies or technologies.</p>
<p>Review the related work sections and focus on extracting triples that describe the significant contributions of technological advances or previous research. Organize information using a format(research or technology, contribution/activation/enhancement, specific area or ability).</p>	<p>1. (GPT-3, led to the development of, highly capable LLMs) 2. (ChatGPT, utilizes, RLHF during fine-tuning) 3. (GPT-4, exhibits, generalization and problem-solving capabilities across diverse domains) ...(Omitted)</p>	<p>Focused on emphasizing significant contributions to the advancement of technology or research. However, structural limitations exist..</p>
<p>The relationships and dependencies between different studies, models, or technologies mentioned in the related work are investigated. Use the format(which affects or influences previous work/model, subsequent work/model) to extract triples that clearly show this connection.</p>	<p>1. (GPT-3, influences, ChatGPT and GPT-4) 2. (ChatGPT, is influenced by, RLHF during fine-tuning) 3. (GPT-3 and ChatGPT, influence, GPT-4's development and capabilities)</p>	<p>Concentrated on identifying and expressing the influence relationships between studies or technologies in detail, but structural limitations are present.</p>

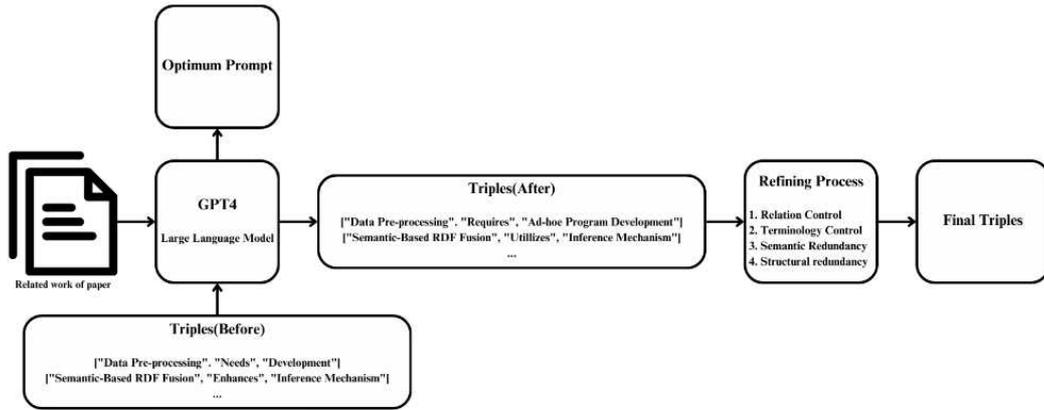


그림 3. Triple 추출 및 정제 과정  
 Fig. 3. Triple extraction and refining process

3.1.2 정제 모듈

위에 언급된 방법을 사용하여 Triple 문장들을 획득하였어도, 제대로 된 용어통제(약어, 일반화 등)의 미비하거나, 의미가 중복되는 문장들의 존재하는 등의 문제가 발생한다. 이를 해결하기 위해 생성된 Triple 문장을 정제하는 총 Relation 통제, 용어통제, 의미론적/구조적 유사도 및 정제 작업 4단계의 절차를 적용하였다.

1단계인 Relation의 통제는 유사하거나 동일한 의미를 지닌 단어와 문법이 일정하지 않은 동사의 특징을 가진 Relation을 일반화한다.

- Focuses on → Targets**
- Represents → Demonstrates**
- Problem, Face → Challenges**
- Improves, Learns, Facilitates → Enhances**

그림 2. Relation 함축  
 Fig. 2. Relation Implication

2단계로는 용어통제이다. 한가지 예로 대규모 언어 모델인 LLM의 경우 Triple 문장 내의

Entity는 LLM, Large Language Model, LM 등 약어로 서술되어 있거나 서술어 형식으로 추출되었기에 통제가 필요하다.

약어로 되어있는 기법, 기술 Entity1 을 언어 모델이 하나의 기법, 기술로 인식하게끔 통일시킨다. 이 과정을 통해서 통일된 Entity로 동일하게 인식 시킴으로써 지식 그래프를 형성할 때, Entity가 여러 서술어 Entity2를 가질 수 있도록 한다.

3단계로는 추출 및 사용자 정제가 제대로 이뤄졌음에도 불구하고, 사용자가 의도한 목적의 Triple 문장이 적절하게 이뤄지지 않을 가능성이 존재한다. 동일 기술이나 기법을 포함하는 Entity1이 Relation이나 Entity2의 부분적인 차이로 인해 다양한 논문에서도 비슷한 주제를 가진 경우, 중복이라고 볼 수 있을 정도의 Triple 문장이 구성된다. 그렇기에 본 연구에서 중복도를 의미적 중복도와 구조적 중복도를 중점으로 중복도 체크를 진행했다.

Triple 문장의 의미적 중복을 해결하기 위해 과학 문서를 대상으로 학습되어 여러 도메인의 과학적 문서에 특화되어있는 Allenai의 Sci-BERT모델[27]을 기용해 각 Triple 문장들의 의미 유사도를 검사하였다. 본 연구에서 추출된 Triple 문장의 의미적 유사도는 Sci-BERT 모델



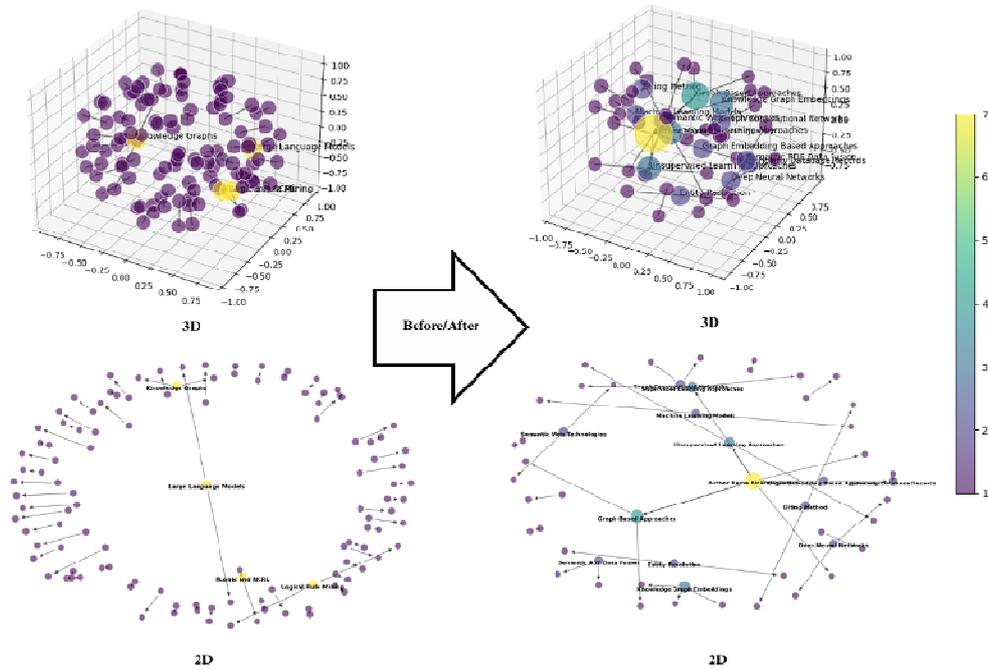


그림 5. Triple 정제 전/후 지식 그래프 비교도  
 Fig. 5. Comparison of Knowledge Graphs Before and After

러한 점으로 정제 전과 후를 비교했을 때, 정제 전은 각 Triple 문장끼리 연결되어 독립적인 경향을 보인다. 하지만 정제 후의 Triple 문장은 노드 연결의 강도와 빈도에 강하게 나타나며, Triple 문장에서 연결 강도 또한 강하게 나타내는 것에 대한 정제 전, 후에 큰 차이를 보였다. 이러한 점을 미뤄보았을 때, 정제 후의 Triple 문장이 확연히 지식 그래프의 지식 구조를 확연하게 볼 수 있다.

#### 4. 실험 및 평가

본 연구에서는 제안된 지식 그래프의 유효성을 검증하기 위한 두 가지 주요 실험을 수행하였다. 첫 번째 실험은 기존 논문의 관련 연구 부분에서 추출 및 정제된 Triple 문장에 대한 평가이며, 두 번째 실험은 자동 생성된 논문의 관련 연구 항목에 대한 평가이다. 이 두 실험에서 공통적인

문제점은 현존하는 평가 데이터셋이나 평가 방법론이 존재하지 않는다는 점이다. 따라서, 본 연구에서는 사용자의 정성적 평가를 기반으로 실험 결과를 평가하였다.

특히, 첫 번째 실험의 평가 과정은 GPT-4의 개체명 인식(NER) 및 정보 추출(IE) 성능을 기준으로 진행되었다. 기존 연구에서 GPT 모델은 NER 및 IE 작업에서 높은 성능을 보인 것으로 [26], [16] 연구에 평가됐으며, 본 연구 역시 입력된 원문으로부터 중요 기술 및 기법을 인식하고 정보를 추출하는 과정에서 GPT-4가 높은 성능을 나타냈다.

두 번째 실험에서는 [11]의 방법을 참고하여 정성적 평가 기준을 적용하였다. 이는 Gold 표준과 일치하는 plot point의 수(#Supp), 동일 기준에서 누락된 plot point의 수(#Cont), 그리고 문법적 정확성 및 유창성(Fluency)을 포함한

다. 본 연구에서는 이러한 기준을 Triple과 자동 생성된 논문의 관련 연구 항목에 적용하여 각각에 대한 정성적 평가를 수행하였다.

이러한 접근 방식은 제안된 지식 그래프와 자동 생성 기법의 유효성을 검증하는 것에 중요한 역할을 하며, 기존 방법론이나 데이터셋의 부재 상황에서도 의미 있는 평가가 가능하게 된다. 이를 통해, 본 연구는 지식 그래프를 활용한 연구 및 자동문서 생성 분야에 기여될 수 있을 것으로 기대된다.

#### 4.1 Triple 평가

위에서 언급한 타 논문을 참고한 평가 방식은 본 연구의 Triple 문장과 자동 생성된 관련 연구 항목에서 다음과 같이 적용된다. 또한, GPT-4 기준 객체 인식과 정보 추출 부분에 대해서, 특정 도메인에 별도의 Fine-tuning 없이 좋은 성능을 보인다는 것은 타 연구 [5], [6]에서 증명되었다.

그렇기에 본 연구에서의 Triple 문장은 기존 논문을 기준으로 언급된 Entity는 Support(이하 #Supp), 누락된 Entity는 Count(이하 #Cont), Entities의 문법과 문장의 매끄러움은 Fluency로 점수를 매긴다. 위의 평가 방식에서 임의로 선정된 60개의 Triple 문장은 표 2와 같이 평가되었다.

표 2. Triple 정량적 평가  
Table 2. Triple Quantitative Evaluation

Qualitative Results	
#Supp	17.3
#Cont	14.1
Fluency	4.2

위의 평가는 Triple 문장과 기존 논문의 관련 연구 항목을 비교하여 언급된 Triple 문장의 점수, 누락된 Triple 문장의 개수를 토대로 도출된 평가 점수는 각각 17.3, 14.1의 점수가 도출된다. 이러한 점수는 Triple 문장이 정제 전/후 비교했을 때, 지식 형성에 매우 유의미한 결과를 가질 수 있다는 점을 알 수 있다. Fluency 와

같은 경우, Triple 문장이 원문과 비교했을 때, 적절하게 Triple 문장이 구성되어 있는지에 따라 평가된다.

본 평가는 각 도메인에서 추출된 Triple 300 건씩, 총 600개를 무작위로 추출하여 평가를 진행했다.

Triple 문장의 점수를 측정하기 위해, GPT-4를 이용하여 4가지 기준(정확도, 사실성, 기여도, 유창성)을 중심으로 자동 평가를 진행했다.

표 3. GPT-4 자동 평가 (Triple)  
Table 3. GPT-4 Automatic Evaluation (Triple)

Criteria	Score	Full Score
Accuracy	25	25
Factuality	20	25
Contribution	20	25
Fluency	25	25
Total Score	90	100

#### 4.2 관련 연구 자동 생성 평가

선정된 추출 Prompt와 4단계의 정제기법이 적용된 Triple 문장을 활용하여 생성된 관련 연구 항목은 Knowledge Graph (KG)와 Large Language Model (LLM), 이렇게 두 가지 분야에서 적용됐다.

평가에 대해서는 Triple과 관련 연구 자동 생성 자체에 대한 비교 정답군이 존재하지 않고, 기존 논문의 관련 연구에 존재하는 도메인의 핵심 기술이 언급되는 정도로 평가하기 위해, 정성적, 적량적 평가를 진행하였다.

이렇게 자동으로 생성된 관련 연구 항목을 [그림 6]과 같이 생성하여 자동 평가를 진행했다. 이러한 평가에 GPT-4를 활용하였으며, 4가지 기준(정확도, 내용 깊이, 조직성, 참조의 적절성)을 기준으로 채점을 진행했다. 기준의 선정은 순서대로 제공된 지식 그래프를 적용했는지, 관련 연구 항목의 특징을 적용했는지, 관련 연구 항목의 구조와 흐름이 올바른지, 인용된 텍스트의 참조가 적절히 이루어졌는지 평가한다.

관련 연구 자동 생성 실험 결과에 대한

GPT-4의 자동 평가는 정제 전/후 4가지 기준(정확도, 내용 깊이, 조직성, 참조의 적절성)을 순서대로 각각 평균, 정제 전(21.5, 22, 24, 21), 정제 후(24.5, 23.5, 25, 23.5)로 나타났으며 Triple 정제 이후, 정제 전과 비교했을 때, 전체적으로 유의미한 결과를 보였다.

평가 점수를 미루어 보았을 때, 선정된 prompt와 4단계 정제 과정을 거친 Triple 문장을 활용하여 자동 생성을 진행했을 때의 결과다. 이러한 결과는 추출된 Triple 문장으로 형성한 지식 그래프가 논문의 관련 연구 자동 생성 시스템에 효과적으로 적용될 수 있음을 보여준다.

## 5. 결론

본 연구는 특정 도메인(KG, LLM)에서 선별된 논문을 기반으로 GPT-4를 이용해, 선정된 prompt와 4단계 정제 기법(약어/술어 통제, Relation 통제, 의미, 구조적 정제)을 통해 지식 그래프를 구성했다. 이렇게 구성된 지식 그래프를 통해서 논문의 관련 연구 항목을 자동 생성하는 시스템을 보였다. 새로운 정제 단계를 적용해서 더 좋은 지식 구조를 가진 지식 그래프를 도출해냈다.

그렇게 설계된 모델은 Triple 추출에서 #Supp, #Cont, Fluency에서 17.3, 14.1, 4.2의 성능과 GPT-4 자동 평가 기준, 100점 기준 96.5점으로 준수한 관련 연구 자동 생성 능력을 보였다.

4단계의 정제 단계를 적용해서 더 좋은 지식 구조를 가진 지식 그래프를 도출하였으며, 도출된 지식 그래프의 지식 구조를 통해서 생성된 관련 연구 항목은 자동 평가 결과, 더 우수함을 보였기에, 이는 본 연구가 제시하는 방법론의 가능성을 긍정적으로 평가할 수 있다.

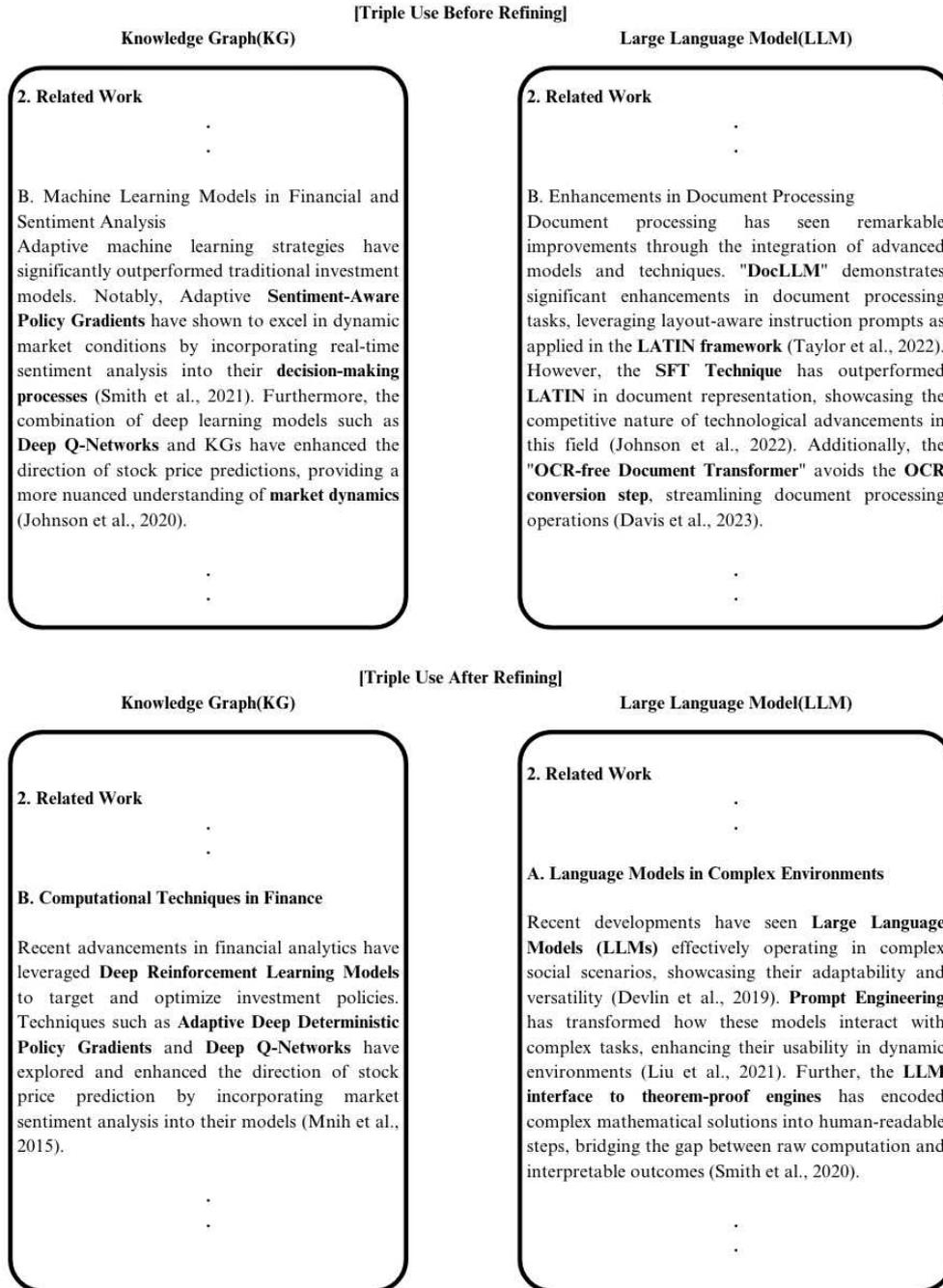


그림 6. 관련 연구 자동 생성 비교 (Triple 정제 전/후)

Fig. 6. Automatic Generation and Comparison of Related Work (Before/After Triple Refinement)

## REFERENCES

- [1] Abu-Rasheed, Hasan, Christian Weber, and Madjid Fathi. "Knowledge Graphs as Context Sources for LLM-Based Explanations of Learning Recommendations." arXiv preprint arXiv:2403.03008 (2024).
- [2] Turki, Houcemeddine, et al. "Enhancing knowledge graph extraction and validation from scholarly publications using bibliographic
- [3] Trajanoska, Milena, Riste Stojanov, and Dimitar Trajanov. "Enhancing knowledge graph construction using large language models." arXiv preprint arXiv:2305.04676 (2023).
- [4] Dong, Qian, et al. "Incorporating explicit knowledge in pre-trained language models for passage re-ranking." Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022.
- [5] Tang, Yiyi, et al. "Large Language Model in Medical Information Extraction from Titles and Abstracts with Prompt Engineering Strategies: A Comparative Study of GPT-3.5 and GPT-4." medRxiv (2024): 2024-03.
- [6] de Faria, Joana Ribeiro, Huiyuan Xie, and Felix Steffek. "Automatic Information Extraction From Employment Tribunal Judgements Using Large Language Models." arXiv preprint arXiv:2403.12936 (2024).
- [7] Trajanoska, Milena, Riste Stojanov, and Dimitar Trajanov. "Enhancing knowledge graph construction using large language models." *arXiv preprint arXiv:2305.04676*(2023).
- [8] Wadhwa, Somn, Silvio Amir, and Byron C. Wallace. "Revisiting relation extraction in the era of large language models." Proceedings of the conference. Association for Computational Linguistics. Meeting. Vol. 2023. NIH Public Access, 2023
- [9] Xu, Derong, et al. "Large language models for generative information extraction: A survey." arXiv preprint arXiv:2312.17617 (2023).
- [10] Shamsabadi, Mahsa, Jennifer D'Souza, and Sören Auer. "Large Language Models for Scientific Information Extraction: An Empirical Study for Virology." arXiv preprint arXiv:2401.10040 (2024).
- [11] Ranade, Priyanka, and Anupam Joshi. "FAB ULA: Intelligence Report Generation Using Retrieval-Augmented Narrative Construction." arXiv preprint arXiv:2310.13848 (2023).
- [12] Dagdelen, John, et al. "Structured information extraction from scientific text with large language models." Nature Communications 15.1 (2024): 1418.
- [13] Hellberg, Ebba. "Exploring GPT models as biomedical knowledge bases: By evaluating prompt methods for extracting information from language models pre-trained on scientific articles." (2023).
- [14] Zhang, Zhi, et al. "BTDM: A Bi-Directional Translating Decoding Model-Based Relational Triple Extraction." Applied Sciences 13.7 (2023): 4447.
- [15] Yao, Liang, et al. "Exploring large language models for knowledge graph completion." arXiv preprint arXiv:2308.13916 (2023).
- [16] Shu, Dong, et al. "Knowledge Graph Large Language Model (KG-LLM) for Link Prediction." arXiv preprint arXiv:2403.07311 (2024).
- [17] Dunn, Alexander, et al. "Structured information extraction from complex scientific text with fine-tuned large language models." arXiv preprint arXiv:2212.05238 (2022).
- [18] Zhu, Yuqi, et al. "Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities." arXiv preprint arXiv:2305.13168 (2023).
- [19] Halike, Ayiguli, Aishan Wumaier, and Tuergen Yibulayin. "Zero-Shot Relation Triple Extraction with Prompts for Low-Resource Languages." Applied Sciences

13.7 (2023): 4636.

[20] Agrawal, Monica, et al. "Large language models are few-shot clinical information extractors." arXiv preprint arXiv:2205.12689 (2022).

[21] Mondal, Ishani, Yufang Hou, and Charles J ochim. "End-to-end NLP knowledge graph construction." arXiv preprint arXiv:2106.01167 (2021).

[22] Lample, Guillaume, et al. "Neural architectures for named entity recognition." arXiv preprint arXiv:1603.01360 (2016).

[23] Fu, Jinlan, Xuanjing Huang, and Pengfei Liu. "SpanNER: Named entity re-/recognition as span prediction." arXiv preprint arXiv:2106.00641 (2021).

[24] Wang, Xinyu, et al. "Improving named entity recognition by external context retrieving and cooperative learning." arXiv preprint arXiv:2105.03654 (2021).

[25] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in neural information processing systems* 35 (2022): 24824-24837.

[26] Bose, Priyankar, et al. "A survey on recent named entity recognition and relationship extraction techniques on clinical texts." *Applied Sciences* 11.18 (2021): 8319.

[27] Beltagy, Iz, Kyle Lo, and Arman Cohan. "SciBERT: A pretrained language model for scientific text." arXiv preprint arXiv:1903.10676 (2019).

---

저자약력

---

최 승 민 (Seung-Min Choi)

[학부생회원]



- 금오공과대학교 입학(2020)
- 금오공과대학교 인공지능 연구실 학부연구생(2023)

〈관심분야〉 자연어 처리, 지식 그래프, 인공지능 기술 기반 응용

정 유 철 (Yu-Chul Jung)

[정회원]



- 1996년~2003년 아주대학교 정보 및 컴퓨터공학과 졸업학사
- 2003년~2005년 한국과학기술원 정보통신공학로봇 전공졸업석사
- 2005년~2011년 한국과학기술원 전산학 졸업박사
- 2009년~2013년 한국전자통신연구원선임연구원
- 2013년~2017년 한국과학기술정보연구원선임연구원
- 2017년~현재국립금오공과대학교 컴퓨터공학과조교수

〈관심분야〉 정보 검색 및 자연어 처리 분야, 소프트웨어기반 로봇, 서비스 플랫폼, 공공 건강, 기계학습 관련, 인공지능 기술 기반 응용