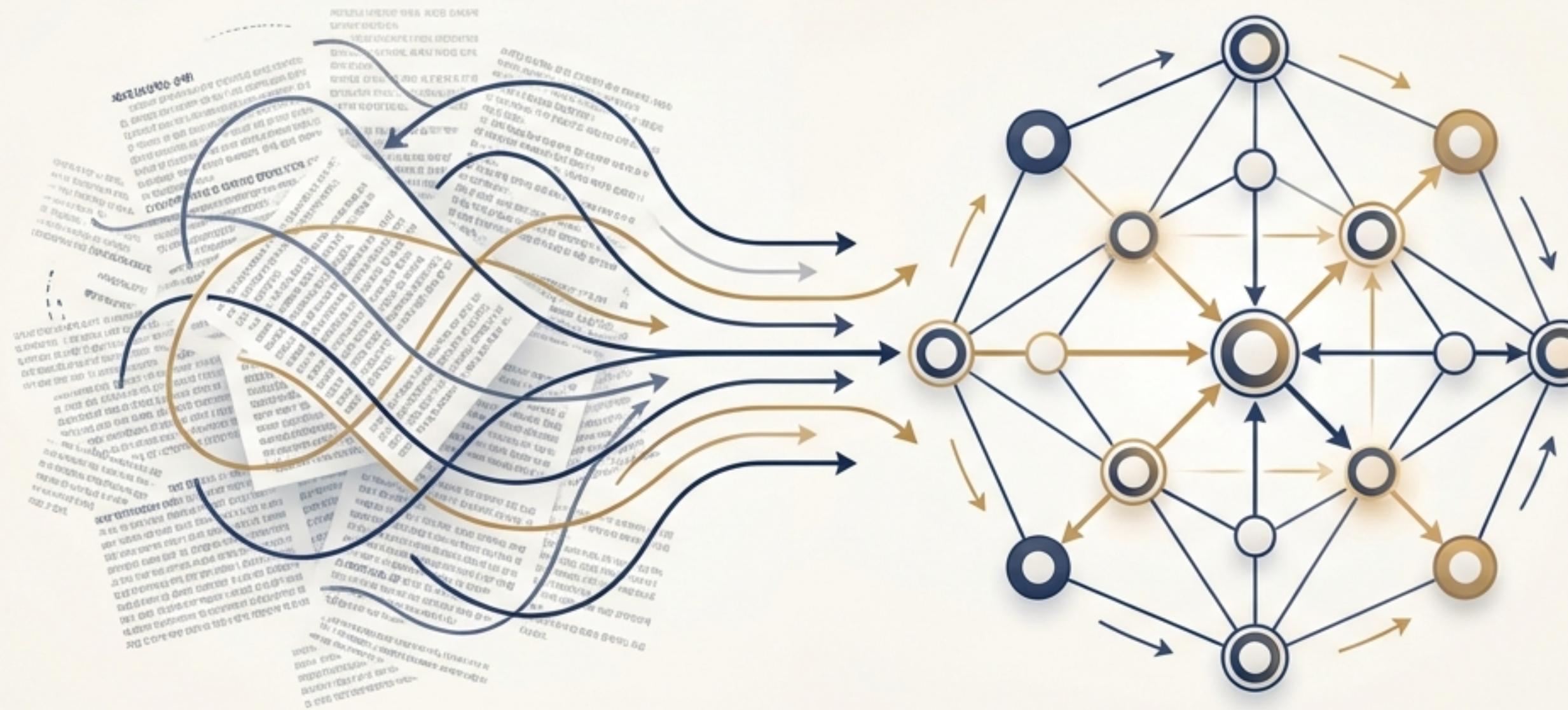


보험 약관, 텍스트의 미로에서 지식의 네트워크로

그래프 기반 RAG를 활용한 차세대 보험 약관 분석 시스템



복잡하게 얹힌 조항들 속에서 명확한 해답을 찾는 방법

우리는 텍스트의 미로에 갇혀 있습니다

보험 약관은 단순한 텍스트 덩어리가 아닙니다.
수많은 조항, 항, 호가 서로를 참조하며 복잡한
논리적 구조를 이룹니다.



이로 인해 발생하는 핵심적인 문제점들:



분산된 정보 (Scattered Information): 하나의 질문에 대한
답이 문서 전체에 흩어져 있습니다.



숨겨진 참조 관계 (Hidden Referential Relationships):
'제 0조 참고'와 같은 구문 뒤에 숨은 맥락을 파악하기 어렵습니다.



맥락적 이해의 부재 (Lack of Contextual Understanding):
단순 키워드 검색은 조항 간의 논리적 연결을 놓칩니다.

선형적 텍스트에서 지식 네트워크로의 전환

기존 방식



선형적, 단절된 정보

(Linear, Disconnected Information)

새로운 접근

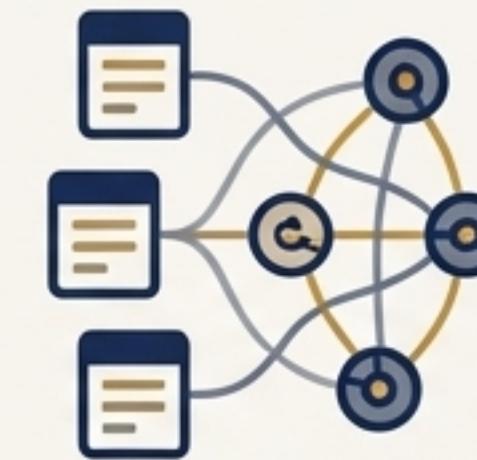


관계 중심, 맥락적 이해

(Relationship-centric, Contextual Understanding)

약관을 단순한 문서가 아닌, 관계와 의미를 담은
‘지식 그래프(Knowledge Graph)’로 재구성합니다.

시스템 작동 방식: 3단계 프로세스



1. 구조 분해 (Deconstruction)

PDF 약관을 의미 있는 최소 단위(조, 항, 호)로 파싱하고 구조화합니다.

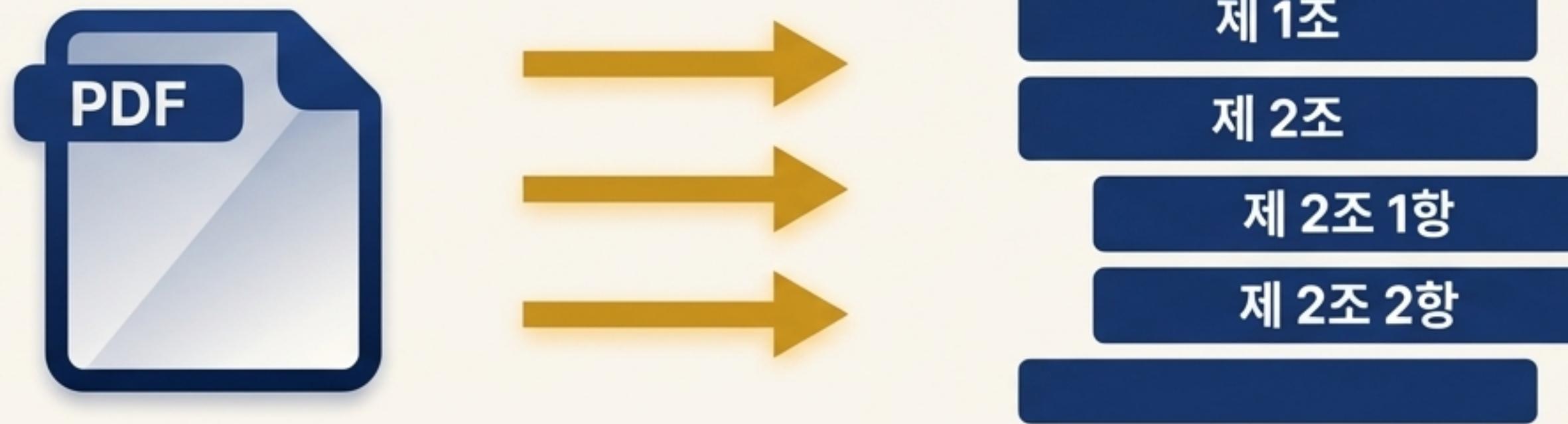
2. 관계 연결 (Connection)

파싱된 데이터 간의 계층 구조와 상호 참조 관계를 추출하여 그래프를 생성합니다.

3. 지능 부여 (Enrichment)

각 노드에 의미를 이해할 수 있는 임베딩 벡터를 생성하여 지능형 검색을 준비합니다.

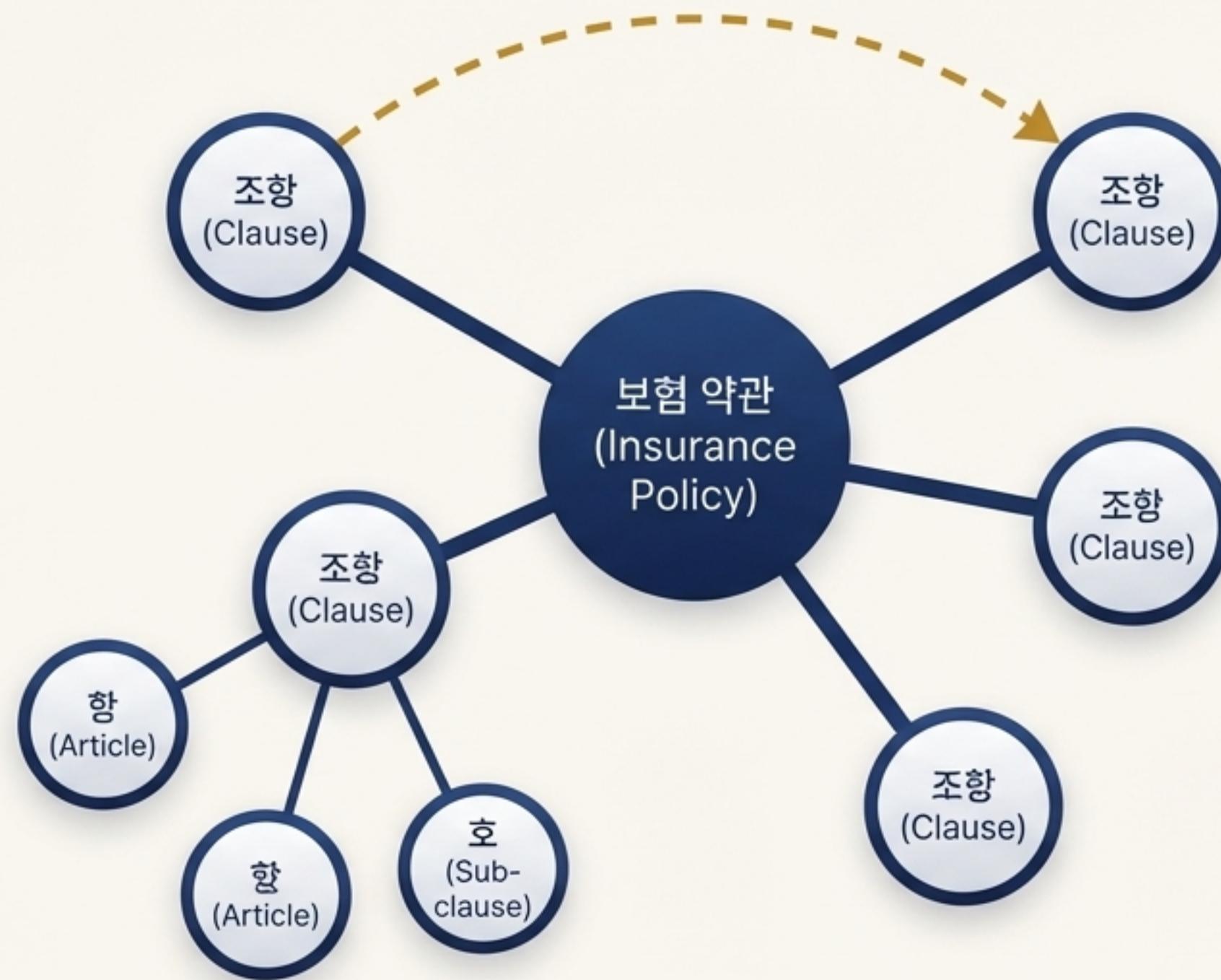
Step 1: 원본 PDF로부터 의미 있는 구조 분해하기



- 1. 업로드 (Upload):** 분석할 PDF 약관 파일을 시스템에 업로드합니다.
- 2. 기본 정보 입력 (Input Basic Information):** 상품 코드, 상품 설명 등 메타데이터를 입력합니다.
- 3. 조항 단위 파싱 (Clause-level Parsing):** 텍스트를 정밀하게 분석하여 개별 조항과 그 하위 구조(항, 호)를 식별하고 추출합니다.

“먼저 어 PDF로 된 이존의 약관을 업로드를 합니다... pdf를 먼저 파싱하고 관련된 조항호를 추출합니다.”

Step 2: 조항 간의 숨겨진 관계를 그래프로 연결하기

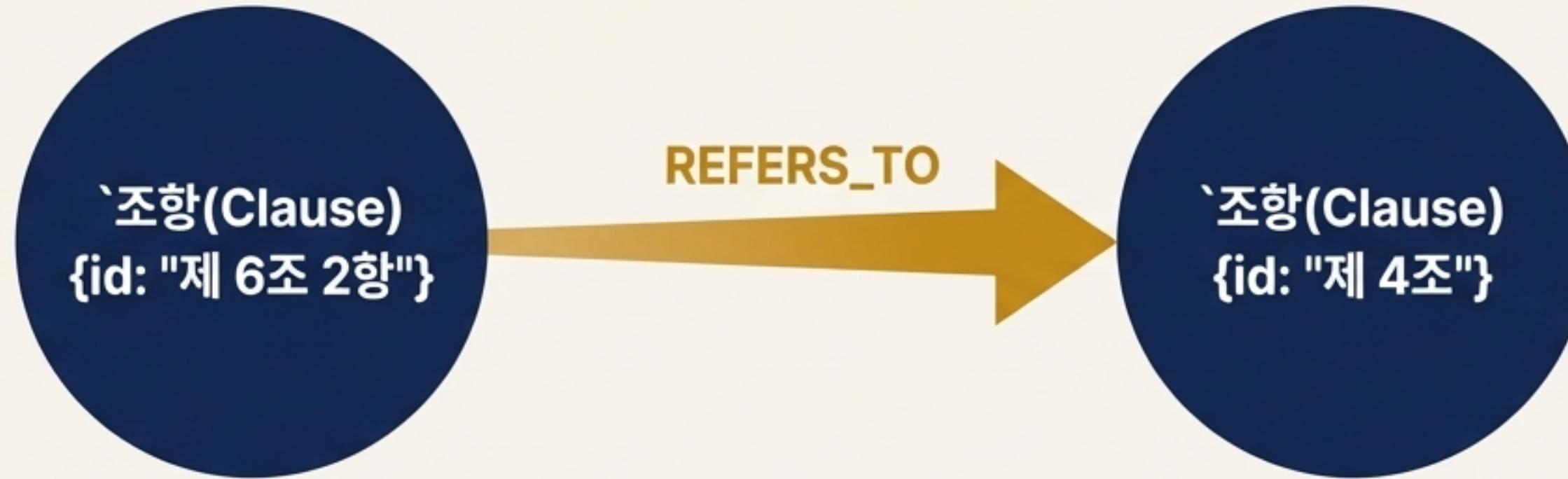


HAS_SUBCLAUSE: 조항-항-호 간의 명확한 계층 구조를 정의합니다.

REFERS_TO: 한 조항이 다른 조항을 참조하는 핵심적인 상호 참조 관계를 포착합니다. 이것이 바로 맥락적 이해의 열쇠입니다.

이 모든 데이터는 Neo4j 그래프 데이터베이스에 저장되어 관계를 영구적으로 보존합니다.

핵심 관계 포착: 'REFERS_TO'의 작동 방식



원본 텍스트 예시 (Original Text Example)

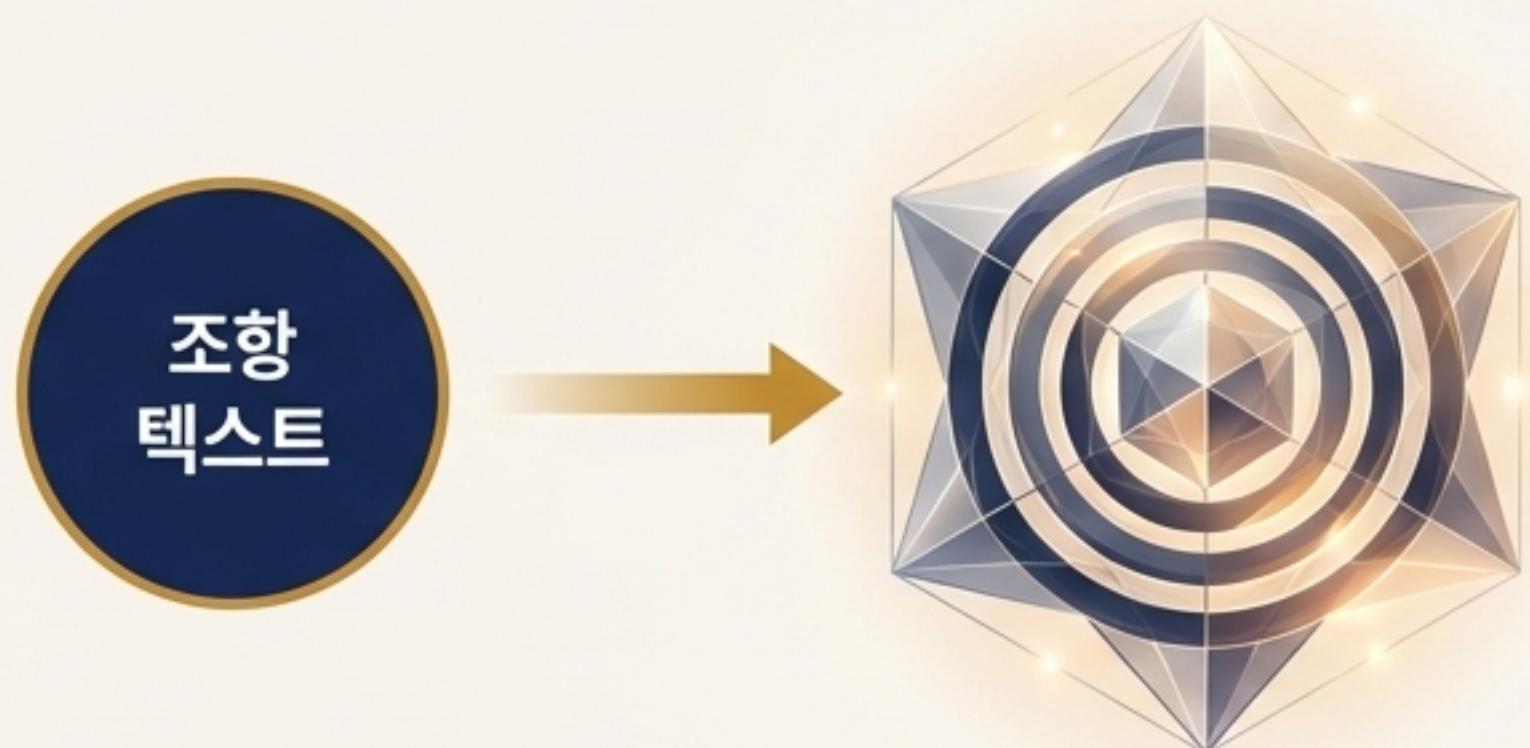
...[제 6조 2항 내용]... 이는 제 4조의 규정에 따른다."

그래프 표현 (Graph Representation)

"6조 2항에서 제 4조를 참고하고 있으면, 그 4조에 대한 내용이 이렇게 그래프로 '레퍼스트(REFERS_TO)'라고 하는 관계로 연결이 된 것을 알 수가 있습니다."

이 연결을 통해 시스템은 두 조항이 동떨어져 있어도 서로 관련이 있음을 인지하게 됩니다.

Step 3: 각 노드에 의미적 지능 부여하기



임베딩 벡터
(Embedding Vector)

그래프의 모든 노드(각 조항, 항, 호)는 단순한 텍스트 저장소가 아닙니다.

각 노드의 텍스트는 자연어 처리 모델을 통해 고차원의 **임베딩 벡터**로 변환됩니다.

이 벡터는 텍스트의 '의미'를 수학적으로 표현한 것으로, 유사한 의미를 가진 텍스트는 벡터 공간에서 가까운 위치에 존재하게 됩니다.

Function: 이는 키워드 일치를 넘어, 질문의 '의도'를 파악하고 가장 관련성 높은 조항을 찾는 것을 가능하게 합니다.

증명의 시간: 기존 방식으로는 답할 수 없는 질문

66

“제 6조에 명시된 보험금 지급 사유가 발생했을 때,
제 4조에 따른 계약 무효 사유와 관련하여 보험금
지급이 거절될 수 있는 경우는 무엇입니까?”

99

두 개의 분리된 맥락 (Two Separate Contexts): 이 질문은 '제 6조(지급 사유)'와 '제 4조(계약 무효)'라는, 문서상 멀리 떨어진 두 정보를 동시에 이해해야 합니다.

관계 추론 필요 (Requires Relational Inference): 단순히 두 조항을 찾는 것을 넘어, 둘 사이의 논리적 관계(참조 관계)를 바탕으로 답을 종합해야 합니다.

지식 그래프를 통한 답변 생성 과정

복잡한 질문 (Complex Query):
"제 6조에 명시된 보험금 지급
사유가 발생했을 때, 제 4조에
따른 계약 무효 사유와 관련하여
보험금 지급이 거절될 수 있는
경우는 무엇입니까?"



초기 노드 검색 및 관계 추론
(Initial Node Search & Relational Inference)

“질문의 의도를 분석해서... 유사한 조항들을 찾고 그 조항에 맞춰서
이렇게 6조와 4조가 연결된 답변을 만든 것을 알 수가 있습니다.”

분산된 정보로부터 종합된 하나의 완전한 답변

제 6조에 따른 보험금 지급 사유가 발생하더라도,
<Text from Clause 4 regarding contract invalidation>에 해당하는 경우
보험 계약이 무효 처리되어 보험금 지급이 거절될 수 있습니다.
특히, **<Text from Clause 6 that references Clause 4>** 조항에 따라,
계약 무효 사유는 보험금 지급 여부에 직접적인 영향을 미칩니다.

답은 단순히 검색된 것이 아니라,
조항 간의 ‘관계’를 이해함으로써 구성된 것입니다.

단순 검색을 넘어, 진정한 문서 이해로



맥락의 무결성 (Contextual Integrity)

문서의 원본 구조와 조항 간의 논리적 관계를 그대로 보존하여 정보의 왜곡이나 손실을 방지합니다.



심층적 정보 검색 (Deep Retrieval)

키워드 매칭을 넘어, 조항 간의 숨겨진 연결을 따라가며 필요한 정보를 심층적으로 찾아냅니다.



복합 추론 능력 (Complex Reasoning)

문서 전체에 흩어진 정보를 종합해야 하는 '멀티-홉(multi-hop)' 질문에 대한 추론 및 답변 생성이 가능해집니다.



이 기술은 보험 약관을 넘어, 법률문서, 기술 매뉴얼, 학술 논문 등
모든 복잡하고 상호 참조가 많은 텍스트에 적용될 수 있습니다.

우리는 정보 검색의 패러다임을 '찾는 것'에서 '이해하는 것'으로 바꾸고 있습니다.