

정보통신관련법 및 판례 관련 지원 서비스

컴퓨터공학과 15학번 최현호

지도교수: 이영구 교수님

요약

법조계 종사자 및 정보통신분야 종사자 간의 간극 해소를 위하여 법, 판례 및 정보통신용어 연계 서비스를 제안한다. 링크 예측 및 키워드 추출 기법을 활용하여 법 및 판례와 정보통신용어 간의 연결관계를 밝히고 이를 기반으로 직관성 높은 서비스를 제공한다.

1. 서론

1.1. 연구배경

2020년 12월 정식 서비스를 시작하였던 스캐터랩(SCATTER LAB)의 열린 주제 대화형 인공지능 챗봇(Open-Domain Conversational AI Chatbot)인 이루다는 여러 사회적 파장을 일으켰지만, 그 중 직접적으로 법적인 제재를 받았던 이유는 개인정보보호법 제28조의2제2항을 위시한 8개 조항을 위반하였기 때문이다.¹⁾ 비단 이루다의 경우뿐만 아니라, 현재 제4차 산업혁명의 최전선을 담당하는 인공지능을 연구하는 데에 있어 연구 데이터는 정보통신정책과 법에서 자유로울 수 없다. 위반했을 경우의 처벌 수위 역시 마찬가지인데, 현행 징벌적 손해보상의 범위인 3배 배상제도의 수위를 사건의 경중에 따라 강화해야 한다는 목소리도 있으며²⁾, 한편으론 강력한 제제는 산업 전반의 발전저해를 초래한다는 목소리가 동시에 있어 양측의 의견을 적절히 절충할 필요성이 있다.

법은 필연적으로 기술 발전 속도에 발빠르게 대처하지 못한다. 산업 현장에서 특정 기술이

빠르게 등장하고 사라지는 탓도 있겠지만, 그 말은 그만큼 기술이 발빠르게 발전한다는 뜻이다. 기술 발전 속도에만 맞춰서 법을 제정하지 않니 산업 현장의 목소리를 제대로 반영할 수 없고, 그렇다고 모든 이해관계를 고려한 법을 도입하자니 역으로 산업 현장과 그 기술로 인하여 피해를 본 사람들을 구제할 수 없는 문제가 발생한다. 하지만, 법을 이해하고 전공하는 사람과 산업을 이해하는 사람은 그 이해의 깊이가 다를 수밖에 없으나, 가장 기초적인 단계에서부터 정보 및 의견의 교류가 가능한 체계가 존재한다면 이 문제는 이전보다는 발빠르게 대처할 가능성이 높아진다. 따라서 산업 현장, 특히 정보통신 분야 종사자 및 법조계의 이해 간극을 좁히기 위하여 별도로 존재하던 법 및 정보통신 관련 정보를 하나로 모을 수 있는 플랫폼을 제안한다.

1.2. 연구목표

현재 대한민국 법원의 종합법률정보 포털을 제외하고 양질의 판례를 제공하는 대중성 있는

플랫폼은 약 4개가 있다³⁾. 저작권법 제7조에 의거하여, 법률 및 판례는 저작권법의 보호를 받지 못한다. 하지만 프로젝트 진행에 있어서의 문제를 최대한 피하기 위해, 판례 수집은 대한민국 법원 종합법률정보에서 조회 가능한 정보통신 분야의 판례에 한한다. 해당되는 판례 및 관련법 그리고 정보통신기술 및 그 용어를 수집, 데이터베이스를 구축하여 각 노드 간의 관계를 가장 잘 나타낼 수 있는 그래프 데이터베이스 및 링크 예측 기술을 활용하여 현재 존재하는 정보 간의 관계 및 추후 등장할 기술의 예상 논점을 쉽게 파악할 수 있도록 한다.

가능한 키워드는 수집된 판례에서 최대한 추출한다. 추출된 결과 중 정보통신에 관련된 키워드는 정보통신용어 DB의 유관 키워드 및 인접 노드 간의 유사도 비교를 통하여 최대한 비슷한 경우를 추천할 수 있도록 한다. 추가로 판례에 등장하지 않거나, 혹은 잠재적 법적 분쟁 가능성이 있는 기술에 대한 저촉 가능성을 예측하며, 더 나아가 특정 기술을 기준으로 관련법 및 판례를 제시할 수 있도록 구현한다.

해당 플랫폼을 웹을 통해 구현한다면, 조회수를 별도로 계산하여 해당 시점에서 가장 많은 조회수를 기록한 용어나 법, 판례를 제시하여 가장 관심도가 높은 부분에 대해 파악할 수 있도록 구현한다. 가능하다면 댓글 기능을 추가 구현하여 해당 기술 및 법에 대한 자유로운 의견 교류를 장려할 예정이나, 프로젝트를 진행함에 있어 불필요한 관리 소요가 발생할 가능성이 커 우선도를 낮게 배정한다.

2. 관련연구

2.1. 링크 예측

링크 예측(Link Prediction)이란 현재시점의 네트워크가 주어졌을 때, 미래시점의 네트워크에 추가될 엣지를 예측하는 문제를 의미한다. 또한 현재 관계가 나타나지 않은 노드 간의 연결성을 예측하는 문제 역시 포함한다⁴⁾. 노드를 단어의 개념으로 인식한다면 자연어 처리 분야의 Word2Vec 기술과 유사하다고 볼 수 있으나, 선형적인 형태의 자연어와는 별개로 그래프는 매우 다양한 방향성을 띌 수 있다.

예측 방법에는 Translation-based model 및 Neural-based model이 존재한다. Translation-based model의 경우 Entity와 Relation을 모두 벡터로 표현하며, Relation은 Entity의 위치를 다른 위치로 옮겨주는 연산자 역할로 간주한다. 따라서 주체 Entity(h)와 객체 Entity(t) 그리고 이들 사이의 관계를 나타내는 Relation(r)으로 이루어진 트리플 <h, r, t>로 두었을 때 $h + r = t$ 와 같이 귀결되는 경우로써 단순하지만 Neural-based보다 우수한 성능을 보인다. 하지만 1대1 대응이 아니면 표현할 수 없다.

Neural-based model의 경우 하나의 트리플 <l_{hs}, rel, r_{hs}>에 대해 2개의 임베딩 표현을 형성한다. 그리고 그 두 표현 사이의 유사도를 이용하여 스코어 함수를 정의한다. 낮은 계산 복잡도를 지녀 대규모 지식 그래프에 적용 가능하다.

2.2. 판례 제공 서비스

현재 판례 제공 플랫폼의 서비스 양상은 전체 공개되지 않은 판결문에 대한 서비스가 주가 된다. 독보적인 판례를 확보하여 타 플랫폼

과의 차별점을 강조하는 서비스가 대부분이며, 빅케이스의 경우 단순 검색이 아닌 쟁점별 판례 검색이나 핵심요약, 유사도 높은 판례를 제공하는 등 NLP를 도입한 검색 서비스가 특징이다. 다만, 아직 특정 산업군에 특화된 판례 서비스는 전무하다. 따라서, 특정 산업군을 겨냥한 서비스는 충분한 가치가 있을 것이다.

3. 프로젝트 내용

3.1. 시나리오

사용자 측면에서의 UI 및 사용 기능은 우선적으로는 검색 서비스와 유사하게 구현한다. 검색 결과로 법, 판례, 정보통신용어를 각각 정리하여 나열하며, 각 항목을 클릭하였을 때 연관 법, 연관 판례, 연관 정보통신용어를 같이 보여주는 방식으로 구현한다. 인접 거리 제한을 설정하여 표시할 수 있도록 구현한다.

키워드 추출을 위해 법 및 판례 DB를 기반으로 Korean KeyBERT 라이브러리나 TextRank 등을 활용한다. 해당 키워드를 기반으로 정보통신용어와의 유사도를 검증하여 링크 예측을 수행한다. 해당 예측 정보를 토대로 별개의 DB를 구축하여 검색 결과와 연동할 수 있도록 작업한다.

3.2. 요구사항

3.2.1 NLP 모델에 대한 요구사항

NLP를 위하여 구글에서 2018년부터 제공하는 BERT 및 고전적인 TextRank를 통하여 키워드를 추출한다. 다만, 추출에 사용되는 데이터

가 한국어인 만큼, 전처리기로써 한국어 형태소 분류기인 mecab-ko를 활용하며, 법률용어 및 정보통신용어를 최대한 전처리기에 탑재한 후 추출을 진행하도록 한다. 위 과정이 생략되면 일상어가 아닌 전문용어는 전처리기에 의해 분해되어 정상적인 결과값을 얻지 못할 확률이 크다. 따라서, 가능하다면 추가적으로 전처리기에 용어를 탑재하기 전과 후의 추출 결과를 비교하여 보고서에 게재할 수 있도록 한다.

3.2.2. 데이터베이스에 대한 요구사항

법 DB의 경우, 가능한 한 법 조항의 조, 항, 호, 목을 구별하여 분류된 노드에 대한 관계를 확실히 표시할 수 있도록 작업한다. 이를 기반으로 판례 및 정보통신용어와의 직접적인 관련성을 더욱 잘 나타낼 수 있도록 작업한다. DBMS는 가능하면 노드간의 관계를 잘 표현할 수 있는 Neo4j를 우선적으로 고려하며, 프로젝트 규모에 더 적합한 DBMS가 있다면 해당 DBMS를 채용한다. 데이터베이스에는 인접 노드와의 관계, 용어 및 정의와 같은 필수적인 정보만을 담아 경량화할 수 있도록 한다.

3.2.3. 색인 웹 페이지에 대한 요구사항

메인 페이지에는 조회수가 높은 법, 판례, 정보통신용어를 일부 표시하며, 검색 시 특정 유사도 이상 및 키워드 포함 결과를 출력할 수 있도록 구현한다. 그래프 시각화를 통해 사용자가 정보를 한 눈에 파악할 수 있도록 작업한다. 가능하다면, 해당 그래프에서 바로 해당 항목에 해당하는 사이트에 접속할 수 있도록 구

현한다.

4. 향후 일정

법률정보 API 및 정보통신용어사전 크롤링을 11월 2일까지 완료한다. 정보통신용어는 우선적으로 용어만을 수집하며, 해당 용어를 기반으로 정의부 스크래핑을 진행한다. 해당 데이터를 사용자가 접근할 수 있도록 하는 플랫폼은 웹을 이용하며, 그래프 시각화를 포함하여 11월 16일까지 모든 작업을 마무리한다. 법 및 판례에서 키워드 추출 작업을 진행하며, 정보통신용어는 유사도 검증 처리를 통하여 인접 단어 추출 작업을 동시에 진행하며, 해당 작업은 데이터 수집이 완료되는 시점에서 11월 9일까지 마무리한다. 추가로, 키워드 추출 작업과 링크 예측 작업이 병행이 가능하다면 해당 작업을 11월 16일까지 마무리한다. 11월 23일까지는 사이트 정상 동작 여부를 파악하고 미비점을 보완하며, 해당 데이터를 기반으로 본 프로젝트의 성패를 최종보고서에 28일까지 반영한다. 최종보고서를 기반으로 11월 30일까지 최종 발표 PPT를 마무리한다. 일정에 대한 간트 차트는 [그림 1]과 같다.

Task Name	10월			11월					
	13 - 19	20 - 26	27 - 02	03 - 09	10 - 16	17 - 23	24 - 30		
데이터 수집									
웹 페이지 설계 및 DB 결합									
키워드 추출 및 그래프DB 적용									
링크 예측 수행 및 결과 DB 결합									
최종 검증 및 미비점 보완									
최종보고서 작성 및 발표준비									

※ 향후 일정은 상황에 따라 변경될 수 있음.

[그림 1]

5. 결론 및 기대효과

법조계에서는 정보통신분야에 직결되는 정보를 편하게 접할 수 있는 기회를 얻게 될 것이

며, 정보통신분야 종사자는 정책 수행 혹은 연구 시 저촉 가능한 법이나 관련 판례에 대해 더 나은 접근성을 지닐 수 있게 될 것이다. 결론적으로, 양 집단의 이해 간극을 좁혀 법조계의 입장에서 산업 현장과 동떨어진 법령이나 정책을 낼 가능성이 낮아질 것이며, 산업 현장에서는 법적인 문제를 손쉽게 확인할 수 있어 별도 법무담당이 존재하지 않는 소규모의 집단에서도 충분히 법적인 문제점을 예상하고 대응할 수 있을 것이라 전망한다.

6. 참고문헌

- 1) 「개인정보위, '이루다' 개발사 (주)스캐터랩에 과징금·과태료 등 제재 처분」, 『개인정보보호위원회』, 2021년 4월 28일.
- 2) 「개인정보 유출로 인한 손해배상책임의 최근 동향」, 『보안뉴스』, 2016년 7월 7일.
- 3) 「온라인 기업 '판례 검색 사이트' 본격 경쟁 돌입」, 『법률신문』, 2022년 5월 12일.
- 4) David Liben-Nowell·Jon Kleinberg, 「The Link Prediction Problem for Social Networks」, 『CIKM』 제12회, Association for Computing Machinery, 2003.