

법률 취약계층을 위한 가까운 판례 데이터

# 가판데

📅 2024.09~2024.11 (2개월)

👤 빅데이터분석가 양성과정 수업의 일환

🏠 법률취약계층을 위한 판례데이터 서비스제공

## 개요

법제처 OPEN API활용하여 판례 제목과 본문상에  
'근로,사기,상속,상해'키워드가 포함된 데이터 수집하여  
사용자가 선택한 소주제와 유사도가 높은 판례데이터를 제시

## 역할

'상속' 키워드 데이터수집, 전처리, 일반분석 및 키워드,주제분석  
Kobert 모델활용, 모델유사도 개선, 분석결과보고서 작성,  
PPT 전체 레이아웃 및 내용 작성, 서비스 시뮬레이션 구현

## 활용도구



GitHub



slack



anaconda



VSCode



python



MySQL



Streamlit

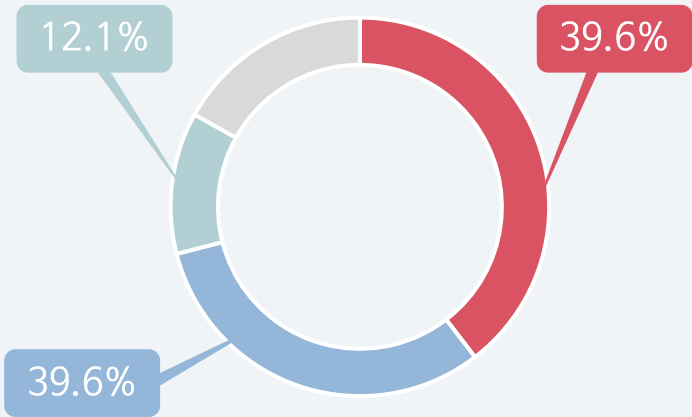


자본력에 따라 법의 보호 편차가 발생하며,  
모두가 동등한 법률서비스를  
받지 못하는 상황 발생



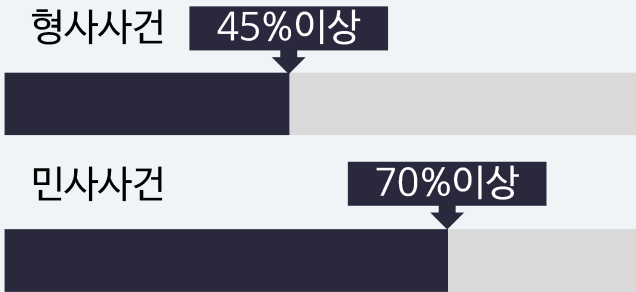
출처 :성주원, “2024년판 ‘유전무죄 무전유죄’ 비극[생생확대경]” 이데일리, 2024.02.06

법률서비스의 미충족 원인은  
어떻게 해야 하는지 잘 몰라서 39.6%,  
상담료와 수임료 등 경제적인 이유가 31.4%,  
법적 절차가 무섭고 스트레스 12.1%로 나타남



출처 :KDI 「디지털 전환에 따른 전문서비스업 현황과 과제: 법률 서비스업을 중심으로」, (2024)

민사 사건의 70% 이상,  
형사 사건의 45% 이상이  
‘나 홀로 소송’을 진행하며  
나홀로 소송의 주된 이유로는 수임료 부담



출처 :신예림, “[재판 지연 해소] “수임료 부담 돼” 나 홀로 소송... 재판 지연의 원인?” 법률방송뉴스, 2024.01.26

법률 취약계층 지원을 위한 법률 지원 서비스를 개발하고, 리걸테크 분야의 신시장 개발 및 전략 기술로 성장하는 것을 목표

데이터수집 및 저장



환경설정

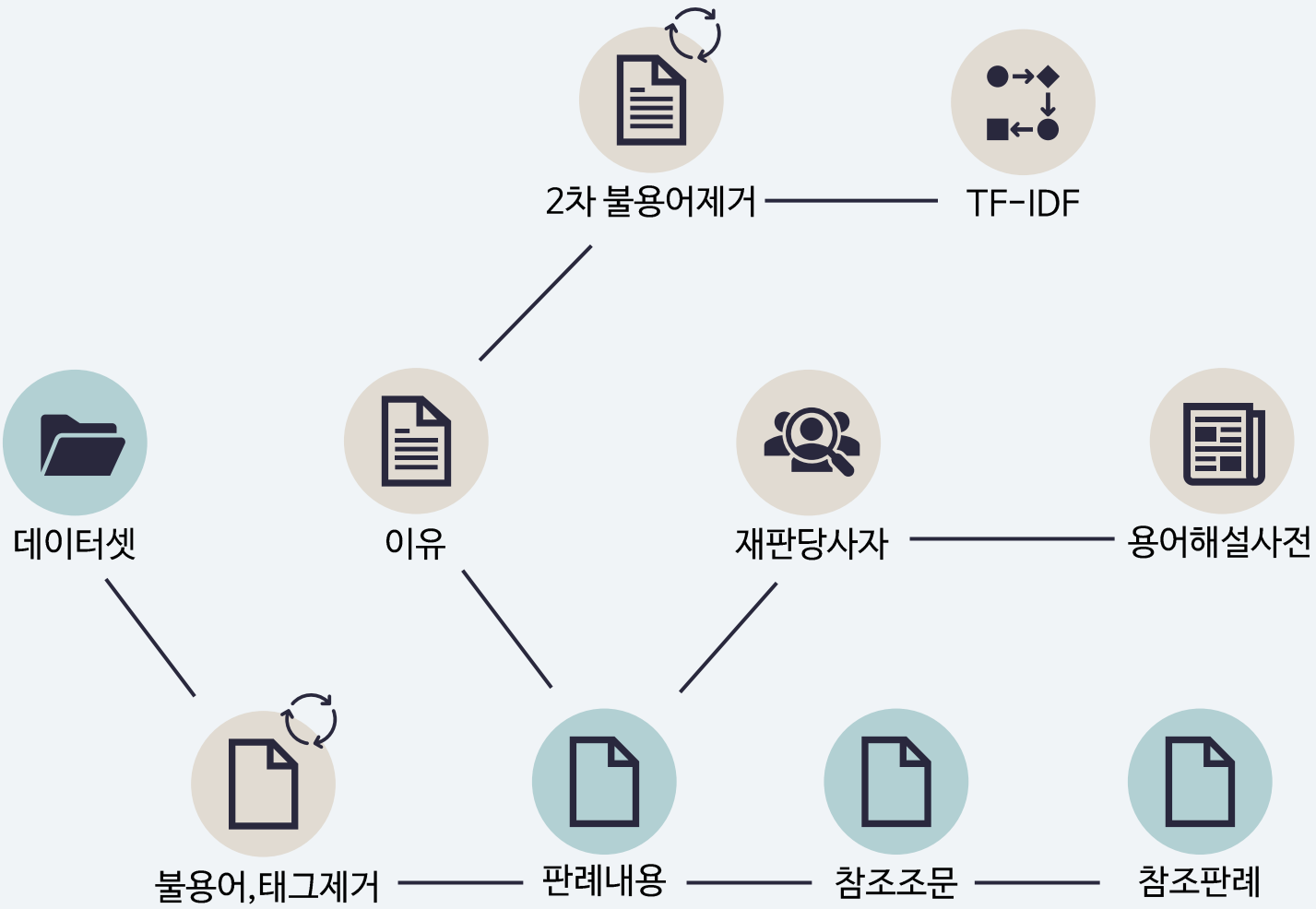


데이터 수집



데이터 저장

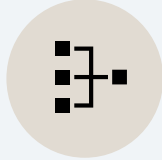
데이터전처리



분석 및 서비스



데이터분석



라벨 태깅



서비스환경 구축



서비스 시뮬레이션

## 프로젝트 문제점



'판례내용'이 재판당사자 / 주문 / 이유 등 범위가 넓은 정보를 포함하고 있어 정규표현식을 활용하여 사건과 연관이 적은 카테고리들을 제외하고, 판결에 대한 주요 정보에 해당하는 '이유' 부분을 분리하는 과정에서 결측치가 발생하였음



원본데이터를 확인한 결과 괄호 표기양식이 달라 결측치가 발생한 것을 확인하였고 정규표현식에 패턴을 추가하여 결측치를 처리하였음



의미상 중요한 단어를 찾기위해 문서를 분석하는 과정에서 원고, 피고, 피상고인 등 재판당사자가 많은 빈도로 등장함



TF-IDF점수를 기반으로 워드클라우드를 생성하여 도출된 재판당사자를 불용어 사전에 추가하고 2차 불용어 제거작업을 진행, 추출된 재판당사자 등 법률용어는 사전으로 구성하여 사용자에게 제공함



사용자가 선택한 주제와 코사인유사도가 높은 문서를 사용자에게 제공하기 위하여 판시사항, 판결요지, 판례내용과 클러스터별 중심단어를 Kober 모델에 활용해 벡터화하여 코사인유사도를 확인하였을때 판례내용 유사도 제일 낮았음



판례내용은 유사도 결과가 가장 낮았지만 법원의 결정을 서술한 부분으로 법적 주요 쟁점부분인 판시사항과 판결문을 간략히 요약한 판결요지보다 문서의 중요도가 높다고 판단하여 유사도 개선방안을 검토함

판례내용 분할

- 표기양식([ ], [ ]) 상이하여 정규표현식 패턴을 추가함

【원고, 상고인】 원고 (소송대리인 법무법인 참진 담당변호사 이영동) 【피고, 피상고인】 주식회사 경남은행 외 3인 (소송대리인 변호사 최준화) 【원심판결】 창원지법 2023. 1. 19. 선고 2021나59370 판결【주 문】 원심판결의 원고 패소 부분 중 MMF 계좌 관련 청구 부분을 파기하고, 이 부분 사건을 창원지방법원에 환송한다.【이 유】 상고이유를 판단한다. 1. 사안의 개요…대법관 천대엽(재판장) 민유숙(주심) 이동원 권영준

【피 고 인】 피고인 【상 고 인】 검사 【원심판결】 대전지법 2018. 11. 9. 선고 2018노1367 판결【주 문】 원심판결을 파기하고, 사건을 대전지방법원에 환송한다. 【이 유】 상고이유를 판단한다. 1. 원심은 그 판시와 같은 이유로, 피고인이 인터넷을 통하여 출력한 등 기사항전부증명서 하단의 열람 일시 부분을 수정 테이프로 지우고 복사한 행위는 공공적 신용을 해할 위험이 있는 정도의 …대법관 민유숙(재판장) 김재형 이동원 노태약(주심)

- 판결내용상단과 판결에 대한 주요 정보에 해당하는 ‘이유’를 나누어 저장

판례내용_상단	판례내용_이유
【원고, 상고인】 원고 (소송대리인 법무법인 참진 담당변호사 이영동) 【피고, 피상고인】 주식회사 경남은행 외 3인 (소송대리인 변호사 최준화) 【원심판결】 창원지법 2023. 1. 19. 선고 2021나59370 판결【주 문】 원심판결의 원고 패소 부분 중 MMF 계좌 관련 청구 부분을 파기하고, 이 부분 사건을 창원지방법원에 환송한다.	'이유': '상고이유를 판단한다. 1. 사안의 개요가. 사실관계 원심판결 이유와 기록에 따르면 다음과 같은 사실을 알 수 있다. 1) 망소외 1(이하 '망인'이라 한다)은 원심판결 별지 2 기재와 같이 피고들 등의 은행에 대한 예금채권과 해당 은행들이…원심법원에 환송하기로 하여, 관여 대법관의 일치된 의견으로 주문과 같이 판결한다. 대법관 천대엽(재판장) 민유숙(주심) 이동원 권영준'

불용어 사전

분류	불용어 목록
대명사	'그', '이', '저', '이것', '저것', '그것', '그녀', '그들', '그분', '누구', '너', '나', '우리', '너희', '당신', '당신들', '자신', '각자', '어느', …
수사(숫자)	'하나', '둘', '셋', '넷', '다섯', '여섯', '일곱', '여덟', '아홉', '열', '여든', '아흔', '백', '천', '만', '십만', '백만', '첫째', '둘째', '셋째', …
법률 용어	'피고(심판청구인)', '원고(부대상고인)', '피상고인겸상고인', '신청인및사건본인', '피항소인', '신청인(준재심피고)', '공동관리인', …
분야별 불용어	'제', '호로', '등', '것', '항', '의', '금', '검사', '판결', '거나', '유1', '상고이유', '기초', '사실', '기초사실', '사고', '피고', '원고', '피고인'…

## 추천시스템 구현을 위한 모델사용

## KoBERT 코사인유사도 계산결과

## - 판결내용

사건번호	사건명	유사도
2010누30330	상속세 부과처분 취소	0.3974
87누898	증여세등부과처분취소	0.3690
2000오2	상속재산관리인선임	0.3683
83누221	상속세부과처분취소	0.3666
2012누8801	상속세부과처분취소	0.3662
2009브23	상속 한정승인 결정에 대한 즉시항고	0.3646
80다2360	소유권이전등기말소	0.3645
79다1009	소유권보존등기말소	0.3628
84누649	상속세부과처분취소	0.3625
2006다2179	상속분양수	0.3621

## - 판결요지

사건번호	사건명	유사도
4294민상168	덕대계약이행완료확인	0.5380
81다카550	소유권이전등기말소	0.5114
86다카1312	소유권이전등기말소등	0.4960
4294민상1255	가옥명도	0.4863
62다343	부동산소유권이전등기말소	0.4846
87다카663	소유권이전등기말소등	0.4839
90다카17214	소유권이전등기말소	0.4827
80구796	양도소득세부과처분취소	0.4824
80구844	취득세부과처분취소	0.4824
90다카3789	소유권이전등기말소등	0.4777

## - 판시사항

사건번호	사건명	유사도
95누9327	상속세부과처분취소	0.8622
2007다9719	소유권이전등기	0.8622
85누804	증여세부가처분취소	0.8622
83다카893	부당이득금	0.8622
83다카745	대여금	0.8622
82누138	양도소득세부과처분취소	0.8622
81나459	공유물분할청구사건	0.8622
80다2360	소유권이전등기말소	0.8622
80나4240	손해배상청구사건	0.8622
79다434	소유권보존등기말소	0.8622

판결내용은 법원의 최종 결정을 가장 구체적으로 서술한 부분이지만 코사인유사도 결과는 가장 낮았음,  
판결요지와 판시사항은 유사도는 높았지만 일부 문서에서 표기누락되어 각 1,200건, 900건의 문서에서 결측치가 발생함

추천시스템 구현을 위한 모델사용

KoBERT 코사인유사도 개선



“단일 키워드를 벡터화하는 방식이 아닌 ‘클러스터 중심단어’ 10개를 10차원의 벡터화하여 유사도를 개선”

기존코드

```
def encode_sentence(sentence):
    inputs = tokenizer(sentence, padding=True,
                        truncation=True, return_tensors="pt")
    with torch.no_grad():
        outputs = model(**inputs)
    return outputs.last_hidden_state.mean(dim=1).numpy()
```

사건번호	사건명	유사도
2010누30330	상속세 부과처분 취소	0.3974
87누898	증여세등부과처분취소	0.3690
2000으2	상속재산관리인선임	0.3683
83누221	상속세부과처분취소	0.3666
2012누8801	상속세부과처분취소	0.3662
2009브23	상속 한정승인 결정에 대한 즉시항고	0.3646

변경코드

```
def get_keyword_embedding(user_keyword, df):
    matching_rows = df[df['키워드'].apply(lambda x: user_keyword in x
                                         if isinstance(x, list) else False)]
    if not matching_rows.empty:
        return matching_rows['키워드임베딩'].iloc[0]
    else:
        return None
```

사건번호	사건명	유사도
2015나6250	건물명도등	0.8054
2015구합66579	상속세부과처분취소청구의소	0.7779
2003가합1087	부당이득금	0.7697
2002구합4914	상속세부과처분취소	0.7560
2009나16058	유류분 반환·손해배상(기)	0.7557
2002누1657	상속세등부과처분취소	0.7540



## 판례데이터 라벨링

- 판례별 클러스터와 키워드로 라벨링한 데이터를 DB에 저장

Limit to 1000 rows

1 • **SELECT** 판례일련번호, 사건번호, 클러스터, 키워드 **FROM** nof.inheritance\_keyword;

2

<

**Result Grid** | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

	판례일련번호	사건번호	클러스터	키워드
▶	238913	2023다221144	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주
	240673	2021두52143	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주
	237913	2023두44061	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	237589	2022다302237	5	망인, 손해, 자랑, 위자료, 지급, 손해배상, 사망, 유족, 운전, 운행
	236923	2022누32308	7	상속세, 세액, 부과, 납부, 과세, 상속, 재산, 가액, 신고, 납세
	238065	2022누67403	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주
	234693	2020다292626	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	234385	2020그42	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	238109	2021구합72130	7	상속세, 세액, 부과, 납부, 과세, 상속, 재산, 가액, 신고, 납세
	237653	2022스625	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주
	227703	2019다29853	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	235383	2022스597, 598	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	227279	2021다294674	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주
	234461	2017스98, 99, ...	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	226751	2015다59801	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	226685	2017두62716	1	가액, 시가, 평가, 상속, 세법, 주식, 재산, 과세, 증여, 상속세
	212581	2014두46485	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	227875	2019다232918	9	상속, 재산, 분할, 유류분, 상속인, 부동산, 민법, 한정승인, 피상, 포기
	234835	2019나2050091	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주
	214533	2019구합83052	3	신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주

## 클러스터 중심단어

클러스터	주제	키워드
1	상속 및 재산 평가	'가액', '시가', '평가', '상속', '세법', '주식', '재산', '과세', '증여', '상속세'
2	부동산 소유권 및 취득	'점유', '토지', '취득시효', '등기', '소유권', '완성', '자주', '부동산', '취득', '의사'
3	부동산 상속 및 사망 관련	'신청', '상속', '부동산', '주택', '사망', '건물', '계약', '회사', '재산', '호주'
4	부동산 등기 및 소유권 이전	'등기', '임야', '소유권', '부동산', '이전', '명의', '원인', '상속', '명의신탁', '토지'
5	손해배상 및 사망 보상	'망인', '손해', '차량', '위자료', '지급', '손해배상', '사망', '유족', '운전', '운행'
6	농지 및 경작	'농지', '분배', '농지개혁법', '토지', '상환', '등기', '본건', '소유권', '경작', '완료'
7	상속세 및 납부	'상속세', '세액', '부과', '납부', '과세', '상속', '재산', '가액', '신고', '납세'
8	부동산 등기 및 거래	'토지', '등기', '소유권', '이전', '명의', '분할', '지번', '주소', '환지', '매매'
9	상속 분할 및 유류분	'상속', '재산', '분할', '유류분', '상속인', '부동산', '민법', '한정승인', '피상', '포기'
10	보험 및 보상	'보험', '보험금', '계약', '지급', '망인', '보험료', '자동차', '연금', '수익', '상해'





FUNCTION 1

전체 4개 분야 상세페이지 버튼 제공

FUNCTION 2

판례 최다 빈출 법원명, 재판장, 빈출 횟수,  
전체 데이터 중 선택분야 비율연도별 관련사건수 그래프 등 제공

FUNCTION 3

키워드 네트워크로 전체 키워드를 계층구조로 시각한 결과와  
키워드 클릭 시 관련있는 노드가 하이라이트되도록 제공

FUNCTION 4

자세한 판례검색 페이지상에 선택한 분야의 상세주제와  
상세주제 선택시 관련된 중심 키워드 10개를 제공

FUNCTION 5

유사도 상위 10개의 판례와 가장 많이 참조된 조문 상위20개  
제공, 하이퍼링크로 연결하여 원본페이지로 이동가능

FUNCTION 6

법률용어를 버튼형식으로 제공, 버튼 클릭시 해당 용어의  
해설을 출력

04 서비스 구축

# 서비스 시뮬레이션

▶ 시연영상 <https://youtu.be/JthNdAvafgA>

### 메인화면

Welcome to NoF

가까운 판례 데이터 가판대입니다. 원하는 카테고리를 선택하세요.

근로 페이지로 이동

사기 페이지로 이동

상속 페이지로 이동

상해 페이지로 이동

법률용어사전 이동

사용자 카테고리 선택

### 상속페이지(상단)

전체 데이터 수  
17549

최다 빈출 법원명  
대법원 2,253회

최다 빈출 재판장  
박만호

상속데이터비율

연도별 관련사건수

23.7%

76.3%

Count

Year

기타 상속

상속 및 재산 평가  
부동산 상속 및 사망 관련  
손해배상 및 사망 보상  
상속세 및 납부  
상속 분할 및 유류분

부동산 소유권 및 취득  
부동산 등기 및 소유권 이전  
농지 및 경작  
부동산 등기 및 거래  
보험 및 보상

### 상속페이지(하단)

키워드 계층구조 시각화

자세한 판례검색 페이지 이동

### 자세한 판례검색 페이지

#### 상속 페이지

여기는 상속 페이지입니다.

주제를 선택하세요:

부동산 상속 및 사망 관련

부동산 상속 및 사망 관련과 관련된 키워드:

신청, 상속, 부동산, 주택, 사망, 건물, 계약, 회사, 재산, 호주

주제 선택 완료

홈으로 돌아가기

사용자 주제 선택

### 판례검색결과페이지

판례검색결과 참조조문

#### 부동산 상속 및 사망 관련과 관련된 사건명

사건번호	사건명
2018나2006493	유해인도
2009브20	상속한정승인결정에대한즉시항고
2018두32927	취득세등경정청구거부처분취소
2012나9392	소유권보존등기
2010다27793	특별채용계약무효확인
2021나31998	공유물분할

원본페이지 연결

### 법률용어사전 제공페이지

#### 법률용어사전

가압류

가압류피신청인

감호청구인

검사

검찰권을 행사할 권한을 갖는 국가기관

검찰관

법률용어해설 제공

## 제약사항 및 추가 연구방안

### 제약사항

판례데이터를 분석하는 과정에서 정형화 되지 않은 판례 작성 형식으로 데이터의 활용과 처리에 어려움이 발생함

구분	사건번호		
	92마783	2013다30882	2018도19043
구분번호	가. 나.	[1], [2]	[1], [2]
괄호기호	【원고, 피상고인】	【원고, 상고인】	[피 고 인]
재판장명	표기누락	대법관고영한 양창수(주심) 박병대김창석	대법관민유숙(재판장) 김재형이동원 노태약(주심)



미국  
24시간 이내



유럽  
1주~1개월이내



중국  
7영업일이내



한국  
1건당 1,000원

- 미국, 유럽, 중국은 선고 후 모든 판결문을 공개하고 있으나 우리나라는 판결문 1건당 1,000원에 제공하고 있음
- 공공데이터로 제공되는 데이터는 고등법원 이상의 상급심 판례로 구성되어있어 하급심 판례데이터가 부족함

### 추가 연구방안

#### ‘판례 표준양식에 따른 자동 데이터화’

서비스 제작 과정에서 수행한 전처리 과정을 학습하여, 사용자가 판례를 입력하면 동일한 전처리 과정을 자동으로 적용하는 기능을 추가적으로 개발하면 판례의 가독성 향상뿐만 아니라 타 서비스 개발에도 도움이 되어 국내 리걸테크 산업에 기여할 것임

#### ‘판결문 공개확대를 위한 비실명화 작업 자동화’

법률서비스의 정확도와 질을 높이기 위해 판결문 공개 확대가 필요한 상황이지만 비실명화 작업을 위한 인력이 부족해 판결문 공개 확대가 어려움  
더 많은 판결문을 공개하기 위해 기계학습 모델이나 딥러닝 기술을 이용한 시스템 구축과 활용 방안 고려가 필요함

#### ‘법률 취약계층을 위한 서비스개발 고도화’

주제를 선택하는 방식 외에도 키워드를 직접 입력하면 판례를 제공하거나 판례의 키워드 외에도 재판장명 등 다양한 조건과 필터를 활용한 검색기능을 개발해 법률 취약계층을 위한 폭넓은 검색 옵션을 제공할 필요가 있음