

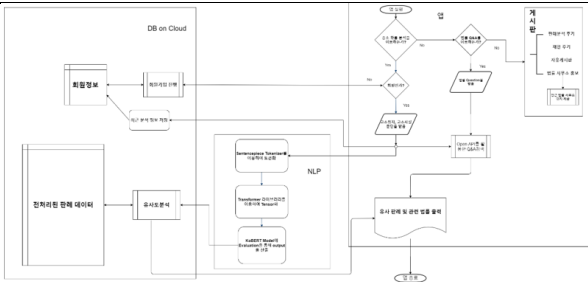
## 2020 한이음 공모전 개 발 보 고 서

2020. 9.

프로젝트명	국문	자연어처리 및 비전을 이용한 AI 기반 맞춤형 법률 서비스 앱 개발
	영문	Development of artificial intelligence-based legal service app using natural language processing
작 품 명	LAWBOT	

 정보통신기획평가원

요 약 본

작품 정보			# 필히 1장으로 작성
프로젝트명	국문	자연어처리 및 비전을 이용한 AI 기반 맞춤형 법률 서비스 앱 개발	
	영문	Development of artificial intelligence-based legal service app using natural language processing	
작품명	LAWBOT		
작품 소개	<p>저희는 고소장 전문을 텍스트로 직접 입력 받거나 사진으로 입력 받아 OCR/자연어 처리/TFIDF 기술을 사용하여 문자 데이터로 변환해 특징 데이터들을 추출하고 분석하여 AI를 통해 법률 조항을 제시하고, 더 나아가 사용자가 처한 법적 상황과 유사한 상황의 판례를 제공해주는 앱을 만들고자 합니다. 또한 키워드를 이용해 검색하는 간편한 법률 지식 QnA 서비스와 커뮤니티 게시판도 만들어 법에 대한 접근성을 고취시키고자 합니다. 저희가 만든 앱이 대중화된다면 법률 지식이 없어 부당한 일을 당할 수 밖에 없었던 사회 취약계층에게 큰 도움이 될 것으로 예상하고 있습니다.</p>		
작품 구성도			
작품의 개발배경 및 필요성	<p>1) 일반인들에게 법률은 아직 어려운 영역임</p> <p>2) 법률 플랫폼이 활성화 된다면 사회 취약계층에게 큰 도움이 될 것임</p> <p>3) 친숙한 매개체를 통해 접근성이 확대될 수 있음</p>		
작품의 특징점	<p>1) 법률 커뮤니티 서비스</p> <p>2) 유사 판례 제공 서비스</p> <p>3) 플랫폼을 이용한 기능 제공</p>		
작품 기능	<p>1) 앱 - 친숙한 매개체로 기능을 제공함</p> <p>2) 법률 QnA - 법률 전문가로부터 직접 답변을 받을 수 있도록 함</p> <p>3) 유사 판례 제공 - 사용자의 상황과 유사한 판례를 제시함</p> <p>4) 법률 문서 해석 - 전문적인 법률 용어를 앱 안에서 바로 검색할 수 있도록 함</p> <p>5) 커뮤니티 서비스 - 사용자간의 정보 및 의견 교환을 위한 커뮤니티를 조성함</p> <p>6) DB(클라우드) - 중요한 데이터를 클라우드 기술을 통해 안전하게 보관함</p>		
작품의 기대효과 및 활용분야	<p>1) 시장성 및 높은 자율성</p> <p>2) 법률문서 가독성 향상</p> <p>3) 변호사 홍보의 장 마련</p> <p>4) 높은 신뢰도 및 접근성 확대</p> <p>6) 사회취약계층 지원, 개인정보보호</p>		

## I. 작품 개요

※ 평가항목 : 기획력 (필요성, 차별성)

### 1. 작품 소개

- 1) 법률에 대한 접근성을 높이기 위한 프로젝트
  - 법률 분야에서는 업계 종사자를 제외한 사람들이 관련 지식을 얻기 쉽지 않음
  - 법이 사회 질서 유지 및 개인의 도덕관념 형성에 중요한 요소임에도 일반 사회 구성원들이 이에 쉽게 접근할 수 없다는 점은 개선되어야 한다고 느낌
  - SW기술을 통해 '법률'에 대한 접근성을 높이기 위한 프로젝트를 고안하게 됨
- 2) 사용된 SW 기술
  - 사용자의 상황을 판단하기 위해 OCR/NLP(TFIDF) 기술을 사용
  - API를 활용하여 판례 정보를 얻었고, 이를 TFIDF 기술을 활용하여 전처리
  - 회원가입, 게시판 등을 위한 DB를 생성하였고 이를 클라우드(AWS)에 연동

### 2. 작품의 개발 배경 및 필요성

- 아직 법률 분야에서는 업계 종사자를 제외한 사람들이 관련 지식을 얻기 쉽지 않음
- 법률 지식이 없어 부당한 일을 당할 수밖에 없었던 사회 취약계층에게 큰 도움이 될 것임
- 친숙한 매개체인 앱을 통해 법에 대한 접근성이 확대될 필요가 있음

### 3. 작품의 특징 및 장점



- 1) 법률 커뮤니티 서비스
  - 기존 벤치마킹 제품에는 없던 커뮤니티 서비스를 제공하여 법률 정보 및 의견을 자유롭게 공유할 수 있는 포럼을 형성하여 차별화
- 2) 유사 판례 제공 서비스
  - 기존 벤치마킹 제품에는 없는 유사 판례 제공 서비스를 통해 사용자가 직접 자신의 상황과 유사한 판례를 열람할 수 있도록 하여 법률 전문가에 대한 의존을 낮출 수 있음
- 3) 플랫폼을 이용한 기능 제공
  - 판례 정보 제시, 커뮤니티, 변호사 추천 기능을 모두 하나의 플랫폼에서 제공하여 어플리케이션 하나로 여러 가지 기능을 사용할 수 있도록 함

## II. 작품 내용



구분	기능	설명	현재진척도(%)
S/W	앱	기능을 제공하기 위한 어플리케이션	20%
	법률 QnA	사용자의 질문을 줄글로 받아 법률전문가로부터 적절한 답변을 받을 수 있게함	70%
	유사 판례 제공	사용자가 처한 상황을 분석하여 유사한 판례를 제시함	70%
	법률 문서 해석	어려운 법률 전문 용어를 앱에서 바로 검색할 수 있게 함	70%
	커뮤니티 서비스	앱 사용자 간의 커뮤니티 서비스를 제공하여 정보 공유 및 포럼 형성	70%
	DB (클라우드)	AWS를 이용한 Web Based DB 구축	80%

## 2) S/W 주요 기능

기능	설명	작품실물사진
법률 QnA	QnA 게시판을 통해 법률 전문가로부터 질문에 대한 전문적인 답변을 받을 수 있다.	
유사 판례 제공	사용자가 본인의 상황 양식을 서식에 맞게 채우거나 이미지 파일을 통해 소장 등을 제시하면 OCR과 자연어처리 기술을 통해 유사한 판례를 찾는데 필요한 키워드만을 뽑아 판례 DB에서 이와 유사한 판례를 찾아 제시해주는 기능이다.	

		<p>· 안녕하세요 최화정이에요. 부바바방ㅇ 움에에에아?</p>
커뮤니티 서비스	자유롭게 법률 정보 및 의견을 나눌 수 있는 커뮤니티를 조성하여 법에 대한 접근성을 제고하고 앱 활성화를 도모한다.	
법률 문서 해석	판례, 고소장 등에서 잘 모르는 법률 용어가 나타나면, 앱 안에서 바로 검색하여 용어에 대한 뜻풀이를 얻을 수 있다.	

### 3. 주요 적용 기술

#### 1) NLP(TFIDF

- NLP의 한 방식인 TF-IDF란 특정 문서의 집합속에서 단어들의 등장 빈도수를 바탕으로 각각 개별문서의 키워드를 알아내기에 적합한 알고리즘
- TF-IDF란 이름은 한 문서에서 특정 단어가 등장하는 횟수를 나타내는 TF라는 용어와 문서 집합속에서 해당 단어를 가지고 있는 문서의 개수를 나타내는 DF의 역수인 IDF를 곱한 값이라는 뜻에서

명명

- 현대의 빅데이터를 다루기 위해서는 수천 수만 수십만개의 나이가 수백 수천만개의 문서집합을 다뤄야 할 필요성이 제기되었고 이에 따른 IDF값의 적절한 추세를 확보하기 위해 보통은 DF값의 역수에 Log를 취한 형태를 사용
- 본 프로젝트에서는 데이터의 처리를 위해 python을 이용하고 있기 때문에 사용의 편의성을 위해 Scikit-learn python package에서 제공하는 tf-idf모델을 이용하기로 함
- 대법원 판례 1851개(8월 기준) 데이터의 판시사항(summary), 판결요지(judgement), 판례내용(content)를 각각 나누어 판시사항 1850개를 포함하는 문서집합, 판결요지 1161개를 포함하는 문서집합, 판례내용 1851개를 포함하는 문서집합을 각각 만들어 각각의 문서집합에 대하여 tf-idf모델을 학습시키고 파라미터를 pickle python package를 이용하여 저장하여 이용

- 유사도 분석의 경우 python의 Scikit-Learn package를 이용하여 만든 tf-idf모델을 이용하여 1851개의 문서집합 각각에 대하여 tf-idf값을 key\_weight컬럼으로 DB에 저장한 후 새롭게 들어올 소장에 대하여 동일한 모델로 tf-idf값을 산출해내 (단어개수)\*1의 행렬로 만든 후 1851개의 문서의 tfidf값인 1851\*(단어개수)행렬과 연산을 통하여 각각의 문서와의 유사도를 구함 이때 행렬 연산에 사용되는 알고리즘은 Cosine similarity, Manhattan similarity, Euclidean similarity, Jaccard similarity 등을 이용하여 유사도를 구하게 됨

## 2) API

- 유사 판례 제공 기능을 제공하기 위해 판례 정보를 얻어야 하여 국가 법령정보 공동활용 사이트로부터 API를 신청하여 key값을 얻었고, HTTP 요청을 보내는 파이썬의 모듈인 request를 사용하여 판례 정보에 접근할 수 있었음
- 법률 문서 해석 기능을 제공하기 위해 단어의 검색 결과를 얻어야 하여 네이버 검색 API를 신청하여 key값을 얻었고, 이 역시 request를 이용하여 검색 결과에 접근할 수 있었다. 원하는 사전(두산백과, 법률용어사전)에 접근하는 추가적인 코드를 통해 보다 정확한 용어 검색 결과를 얻을 수 있었음

## 3) DB(MySQL)

- 유사판례분석 : 소장의 내용과 비교분석할 판례 데이터를 DB에 저장하기 위해 Precedent, Summary, Judgement, Content 테이블을 만듦
- 판례api에서 받아온 각 판례의 판례일련번호와 판례 사건명을 넣을 수 있는 Precedent 테이블이 있고, 각 판례의 판시사항(Summary), 판결요지(Judgement), 판례내용(Content)을 nlp모델로 돌려 결과 값들을 각 테이블에 넣었다. 각 테이블은 Precedent의 ID를 외래키로 가지며, 한 판례가 여러 단어를 가지고 있기 때문에 Precedent\_ID와 keyIndex를 기본키로 설정
- 법률 QA : 법률 QA의 질문과 답변을 DB에 저장하기 위해 Question, Answer 테이블을 만듦  
Question 테이블의 경우, 모든 사용자가 질문을 작성할 수 있으므로 User 테이블의 ID를 외래키로 지정
- 답변의 경우, 변호사만 작성 가능하고 일반 사용자는 작성 불가하므로 User\_ID 대신 Lawyer 테이블의 ID를 외래키로 설정하였으며, 하나의 질문에 여러 답변이 달릴 수 있으므로 Question 테이블과 Answer 테이블을 1대N 관계로 설정

- 법률게시판 : 게시글과 게시글에 달린 댓글을 DB에 저장할 수 있도록 Post 테이블과 Reply 테이블을 만들 한 게시글에 여러 댓글이 달릴 수 있으므로 두 테이블은 1대N 관계로 설정
- 변호사정보 : 변호사 정보를 저장할 수 있는 테이블로 Career, Qualification, Education, Activity 테이블을 만들 한 변호사가 같은 내용의 정보를 여러 번 입력할 수 없도록 하기 위하여 외래키인 Lawer\_ID와 상세내용이 저장된 detail 컬럼을 묶어 기본키로 설정

- 회원관리 : 회원을 관리하기 위해 User 테이블과 Lawyer 테이블을 만들 변호사의 경우, 일반회원인 동시에 변호사이므로 정보가 User 테이블과 Lawyer 테이블에 모두 저장 Lawyer 테이블은 User 테이블의 ID 값을 외래키로 받아왔으며 동시에 기본키로 설정

#### 4. 작품 개발 환경

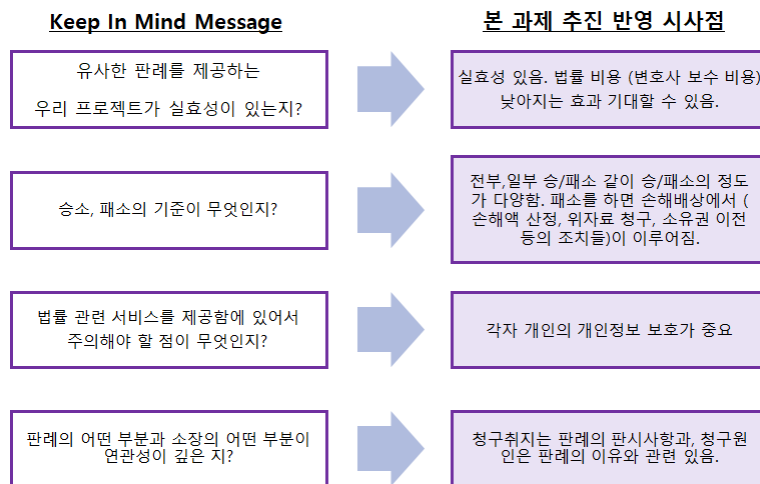
구분		상세내용
S/W 개발환경	OS	Window/MacOS
	개발환경(IDE)	Jupyter Notebook, colab, visual studio code
	개발도구	git, UML
	개발언어	Python, javascript
	기타사항	해당 없음
H/W 구성장비	디바이스	해당 없음
	센서	해당 없음
	통신	해당 없음
	언어	해당 없음
	기타사항	해당 없음
프로젝트 관리환경	형상관리	gitlab
	의사소통관리	google drive, trello
	기타사항	해당 없음



## 5. 기타 사항 [본문에서 표현되지 못한 작품의 가치(Value)] 및 제작 노력

### 1) 법률전문가 인터뷰를 통한 방향성 검토

- 현 변호사님과의 인터뷰를 통해 기획한 서비스의 방향성을 검토하고 제약사항을 확인



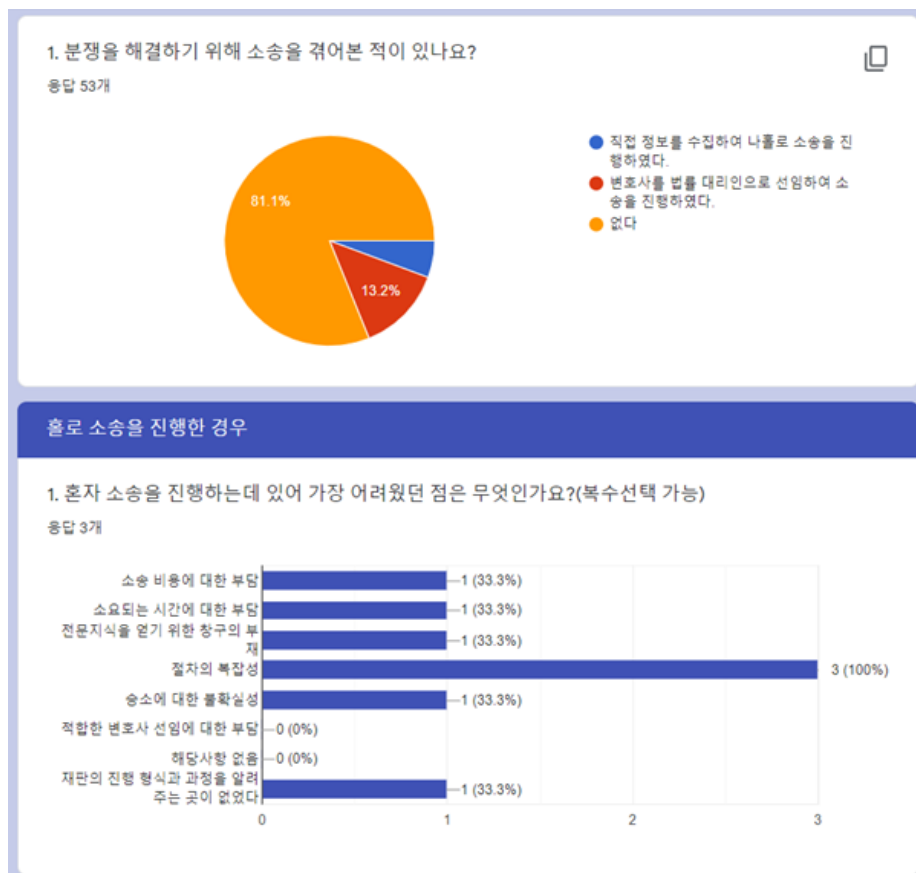
<인터뷰 결과>

### 2) 법률서비스 필요성에 대한 설문 수행

- 설문기간 및 방법

항목	내용
설문 명	법률 서비스 플랫폼에 대한 설문
설문 기간	2020년 7월 28일 ~ 2020년 8월 6일
설문응답	53명
표본 집단	대학생, 학부모
표본 집단 선정이유	특정 연령대에서만 설문을 조사하면 결과가 편향되어 나타날 수 있기 때문에 다양한 연령대에서 조사하고자 위와 같이 표본 집단을 선정하였다.
설문도구	구글 설문지, 학교 커뮤니티 및 카카오톡

- 설문결과: 대부분의 응답자들은 소송을 경험하지 않았지만 소송이나 분쟁경험이 있는 응답자의 대부분은 변호사 등의 법률서비스를 이용하는 것으로 집계됨
- 설문시사점: 소송을 혼자 하는 경우 가장 어려운 것은 절차의 복잡성으로 나타나 본 플랫폼에서 절차에 대한 안내를 포함하 필요해 보임



### 3) 법률서비스 시장동향 분석 및 차별화 분석

구분	로엔비	법률메카	로톡	유렉스	LAWBOT
형식	웹사이트, 앱	웹사이트	웹사이트, 앱	웹사이트, 앱	웹사이트, 앱
법률 Q&A	○	○	X	○	○
법령/판례검색	○	○	X	○	○
유사판례 제공	X	X	X	X	○
커뮤니티	X	X	X	X	○
변호사 추천	X	○	○	X	○

- 본 프로젝트에서 개발하고자 하는 LAWBOT은 타 서비스들에서 종합적으로 제공하지 못하는 서비스를 통합하여 제공하고자 함

#### 4) 변호사 추천 알고리즘 자체 개발

- 사용자 맞춤형 변호사를 추천하기 위해 자체적으로 추천 알고리즘을 개발함 일반 회원과 변호사 회원 모두 회원가입을 할 때 관심 분야를 선택하도록 했으며, 일반 회원이 변호사를 추천받을 경우, 해당 회원의 관심 분야를 고려하여 변호사를 추천하기 위해 등록된 변호사들의 관심 분야와 해당 회원의 관심 분야가 일치하는 정도를 수치화 하여 가장 높은 점수를 얻은 변호사 5명을 추천하게 구성함

III. 프로젝트 수행 내용

※ 평가항목 : 수행능력 (문제해결능력, 수행충실성)

1. 프로젝트 수행일정

프로젝트 기간 (한이음 사이트 기준)		2020.00.00. ~ 2020.00.00.											
구분	추진내용	프로젝트 기간											
		1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
계획	아이디어 회의												
분석	벤치마킹 조사 및 분석												
	활용 데이터 조사 및 분석												
설계	아키텍처 설계												
	모바일 앱 기능 및 화면 설계												
	OCR 및 NLP 모델 설계												
	판례정보 수집 및 분석모델 설계												
개발	모바일 앱 기능 개발												
	자연어 인식 및 처리 기능 개발												
	판례 정보 수집 및 분석 모델 개발												
테스트	통합 테스트 및 보안												
종료	Lessons Learned												

2. 프로젝트 추진 과정에서의 문제점 및 해결방안

### 1) 프로젝트 관리 측면

- 코로나 사태로 인하여 직접 만나지 못하여 소통의 어려움을 느꼈다. 줌과 마이크로소프트 팀즈를 이용하여 최대한 자주 일주일에 한 번씩은 회의를 하며 이 문제를 해결
- 팀 내의 각자 맡은 파트에 대한 진행상황을 알지 못하는 어려움을 느낌 Trello를 통해 구체적으로 To Do, Doing, Done, Delayed으로 파트를 나눠 각자의 진행상황을 알기 쉽게 표현하도록 함으로써 해결

### 2) 작품 개발 측면

- 프로젝트가 법률관련 프로젝트이지만, 팀원 중 법률관련 전문가가 없음에 따른 디테일 부족 문제가 있었다. 변호사와의 면담을 직접 의뢰하고 하여 부족했던 부분을 보충

- 형법이 적용되는 사례는 수많은 변수가 존재하므로 서비스 범위를 민법으로 축소하고 그 중 손해배상에 집중
- 자연어처리 기술 적용에 있어, 형태소 분석 및 의미역 분석을 어떤 식으로 할 것인지에 대한 많은 논의 있었음 먼저 API를 단순적용하는 법을 사용하기 위해 각종 자연어처리 API를 사용해보고 테스트해 보았음 SKT 에서 만든 KOBERT를 사용하기로 하였으나, 법률 분야의 특성 상 시중에 제공되고 있는 API에서는 사용되지 않을 전문용어들이 많을 것 같아 SentencePiece를 이용하여 직접 언어 학습을 시키고 형태소 분석기를 만들기로 결정
- 대법원 제공 민사소송 중 손해배상 목록에 해당하는 문서들을 활용하여 SentencePiece를 학습시켜 프로젝트에 알맞은 형태소 분석기와 Vocab 리스트를 만들어냄
- 인공지능 모델 학습 및 평가면에서는, Sentimental Analysis처럼 학습데이터를 승소/패소로 이원화하여 의뢰인이 입력한 데이터가 각각 어떤 카테고리에 더 알맞는지 체크하여 예상 승소/패소 확률을 제공하기로 결정 하지만 여러가지 이유로 인해 새로운 방법을 도입하기로 했음
  - 첫째, 판례가 정확하게 승소/패소로 나누어 떨어지지 않았음
  - 둘째, 고소장에 드러나지 않는 정보들 중 판례에 영향을 크게 미치는 요소가 많음
  - 셋째, 실제 고소의 승소/패소 여부에 법정대리인의 능력이 크게 영향을 미침
- 위와 같은 이유로, 애초 계획했던 ‘승소확률분석’ 서비스에서 ‘유사판례분석’ 서비스로 방향을 변경하게 됨
- 따라서 TFIDF 라는 새로운 방법을 도입하여 형태소를 분석하고, 다양한 유사도 분석 기법(유클리드, 코사인 등)을 활용하여 고소장의 vocab 데이터와 판례의 vocab 데이터의 유사도를 분석하여 가장 유사한 판례순으로 보여주는 식으로 서비스를 변경

### 3. 프로젝트를 통해 배우거나 느낀 점

- 다함께 프로젝트를 진행함에 있어서 팀워크의 힘과 중요성을 느낌
- 개별적으로 작은 프로그램을 진행하는 것과 달리 큰 프로젝트를 진행할 때 시스템의 구조를 그려보고, 필요한 서류를 준비하면서 큰 그림을 파악하는 것의 중요성을 느낌

- 법률 분야의 접근성을 높이기 위해 시작한 프로젝트인데, 직접 부딪혀보니 왜 그 진입장벽이 높은지 알게 되었음. 전문용어의 남발과 너무나 많은 법률조항, 판례, 그리고 복잡한 법률체계 등으로 인해 서비스를 제공하는 우리조차 이해하기 힘들었다. 그래서 더더욱 이 서비스를 통해 사람들이 많은 도움을 받을 수 있었으면 좋겠다고 생각하게 되었음

- 개발 포지션으로만 팀을 구성하다보니, 팀 구성에 있어서 포지션의 중요성을 깨닫게 됨
- 처음에는 개발 파트를 나누지 않고 함께 진행하려고 했으나, 프로젝트가 진행됨에 따라 효율적인 협업을 위해서는 각자 파트를 나누어 진행하는 것이 더 좋다는 것을 깨닫게 됨
- 어플 개발이 생각보다 어렵다는 것을 느꼈음

#### IV. 작품의 기대효과 및 활용분야

##### ※ 평가항목 : 기획력 (활용가능성)

##### 1. 작품의 기대효과

- 시장성 및 높은 자율성 : 유사판례분석 서비스와 법률게시판에서의 자유로운 소통, 그리고 법률QA 서비스로 얻을 수 있는 실제 변호사로부터의 답변을 통해 소송을 진행하고자 하는 개인이 변호사 상담 없이 소송을 준비할 수 있게 함으로써 변호사상담에 드는 비용을 줄이고 개인의 자율성을 높일 수 있을 것을 기대
- 법률문서 가독성 향상 : 일반인이 이해하기 어려운 법률문서 상의 법률용어의 의미를 앱상에서 바로 띄어줌으로써 일반인의 법률문서 가독성을 높이는 것을 기대
- 변호사 홍보의 장 마련 : 신규 변호사가 앱에서의 활발한 법률QA 답변을 통해 자신을 홍보할 수 있는 장을 마련할 것을 기대하며 더불어 일반 사용자는 이를 통해 자신에게 맞는 변호사를 찾을 수 있을 것을 기대
- 높은 신뢰도 : 앱 가입시 변호사로 가입할 경우, 변호사 신분증을 대한변호사협회 사이트를 통해 인증하고 실제 변호사인 경우에만 법률QA에 답변을 할 수 있도록 함으로써 네이O 지식IN의 신뢰도 문제를 해결할 수 있음을 기대

##### 2. 작품의 활용분야

- 접근성 확대 : 전문 지식 분야인 법률 서비스의 진입장벽을 낮추고 그에 따른 법률 사각 지대의 축소
- 사회취약계층 지원 : 거동이 불편하거나 사회취약계층에 속한 사람들이 부당한 대우를 받거나 정당한 권리를 행사하지 못하는 것을 방지
- 개인정보보호 : 상담 및 정보수집 과정에서 발생하는 불필요한 접촉을 방지하고 익명성을 보장

