



LAWMATE

사용자 중심의 법령 검색 챗봇





법률 검색



변호사 검색



**전자증거
개시**



**법률자문 및
전략 수립**



Legal Tech
Law + Technology

세계 리걸 테크 스타트업 투자는
2011년 9천140만 달러에서
지난해 2억9천200만 달러로
3배 가량 증가하였다.
법률서비스 소프트웨어 시장은
지난해 38억 2천 800만 달러
규모이며 앞으로 2019년까지
57억 6천 300만 달러 규모로
성장할 것이란 전망이다.



DATA

데이터 출처: <http://kin.naver.com/index.nhn>
<http://www.easylaw.go.kr>



16440

지식 in
변호사 답변



1243

생활법령
백문백답



생활법령

쉽게 풀어 쓴
법령 내용

BACKGROUND

1

**TF-IDF
가중치**

벡터 기반 모델

2

**TS-SS
유사도**

벡터 기반 모델

3

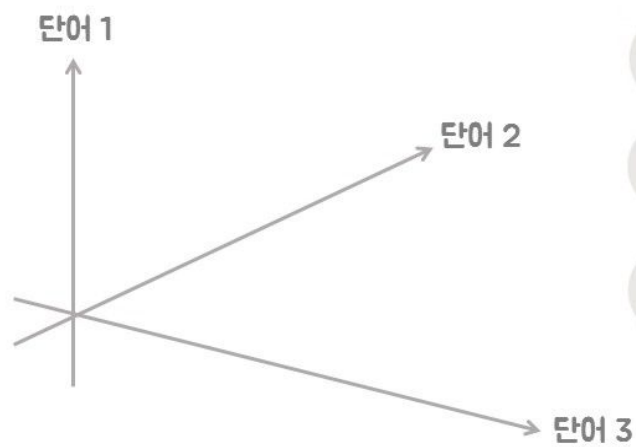
**Text
Rank**

그래프 기반 모델

TF-IDF

벡터 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델

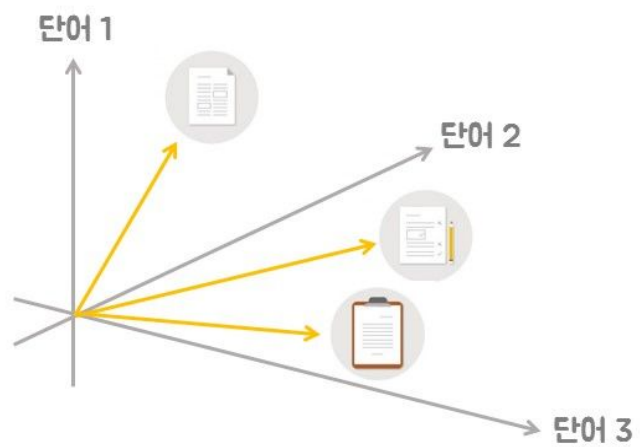


- 01 n 개의 개별 단어가 출현하는 문서를 n -차원의 벡터 공간 안의 벡터로 표현(단어는 축을 의미)
- 02 쿼리를 같은 공간 안의 벡터로 표현
- 03 벡터 공간 내에서 쿼리와 문서 간의 거리에 따라 순위화

TF-IDF

벡터 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델

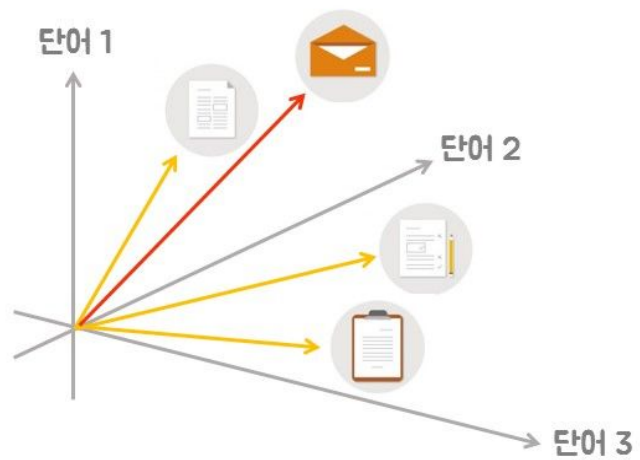


- 01 n 개의 개별 단어가 출현하는 문서를 n -차원의 벡터 공간 안의 벡터로 표현(단어는 축을 의미)
- 02 쿼리를 같은 공간 안의 벡터로 표현
- 03 벡터 공간 내에서 쿼리와 문서 간의 거리에 따라 순위화

TF-IDF

벡터 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델



- 01 n 개의 개별 단어가 출현하는 문서를 n -차원의 벡터 공간 안의 벡터로 표현(단어는 축을 의미)
- 02 쿼리를 같은 공간 안의 벡터로 표현
- 03 벡터 공간 내에서 쿼리와 문서 간의 거리에 따라 순위화

TF-IDF

벡터 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델

문서



벡터

= (단어1 가중치 x 단어1 빈도,
단어2 가중치 x 단어2 빈도, ...)



$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

- 01 Tf(Term Frequency) = 문서 내 단어의 빈도
단어가 특정 문서 내에서 얼마나 자주 등장하는
지
- 02 IDF(Inversed Document Frequency)
= $\log(\text{전체 문서의 개수} / \text{단어가 포함된 문서의 수})$
- 03 단어가 문서 집합 내에서 얼마나 자주 등장하는
지

특정 문서 내에 자주 등장하고, 다른 문서에는 자

TS-SS 유사도

벡터 기반 모델

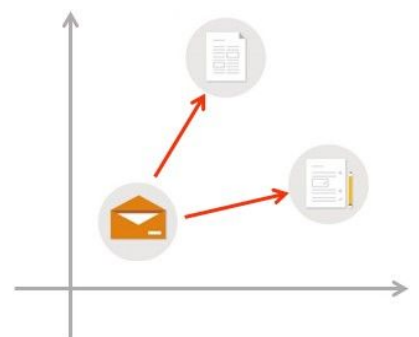
문서 간 유사도 판별에 있어서
기존의 코사인 유사도와
유클리드 유사도의 단점을
개선하기 위한 유사도



01 코사인 유사도

단어가 등장하는 비율을
유사도에 반영

방향은 같지만 등장하는
단어의 규모가 다른 문서
들을 서로 일치하는 것으
로 판별



02 유클리드 유사도

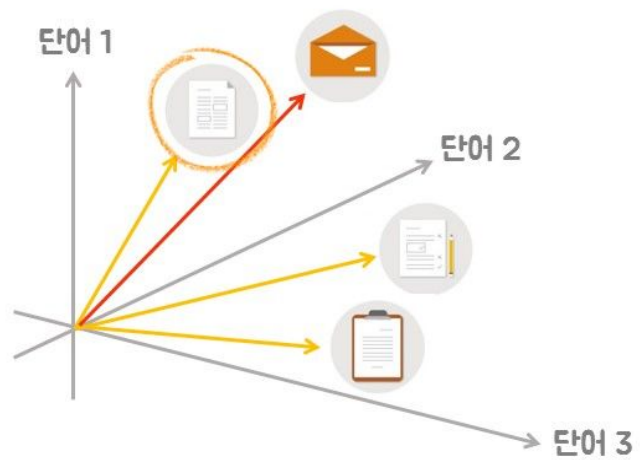
좌표 평면 상의 거리를
유사도에 반영한다.

단어의 규모가 다른 경
우를 식별해낼 수 있으
나, 단어의 등장 비율은
반영 불가

TF-IDF

벡터 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델



- 01 n 개의 개별 단어가 출현하는 문서를 n -차원의 벡터 공간 안의 벡터로 표현(단어는 축을 의미)
- 02 쿼리를 같은 공간 안의 벡터로 표현
- 03 벡터 공간 내에서 쿼리와 문서 간의 거리에 따라 순위화

TS-SS 유사도

벡터 기반 모델

문서 간 유사도 판별에 있어서
기존의 코사인 유사도와
유클리드 유사도의 단점을
개선하기 위한 유사도



03 삼각형 면적 유사도

두 문서와 원점이 이루는
삼각형의 넓이를 유사도
에 반영

$$\text{TS-SS 유사도} \\ = (\text{삼각형 면적 유사도} \\ \times \text{부채꼴 면적 유사도})$$

A Heidarian, 2016, A Hybrid Geometric
Approach for Measuring Similarity Level Among
Documents and Document Clustering

04 부채꼴 면적 유사도

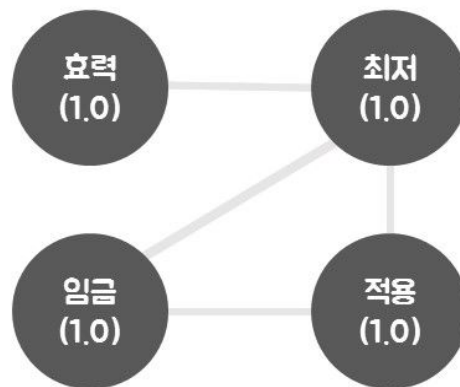
두 문서 간의 유클리드
거리에 문서 벡터의 크
기 차이를 합한 값을 반
지름으로 하는 부채꼴을
그려서 그 넓이를 유사
도로 사용

Text Rank

그래프 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델

효력: 사용자는 최저임금의 적용을 받는 근로자에게
최저임금 이상의 임금을 지급해야 한다.



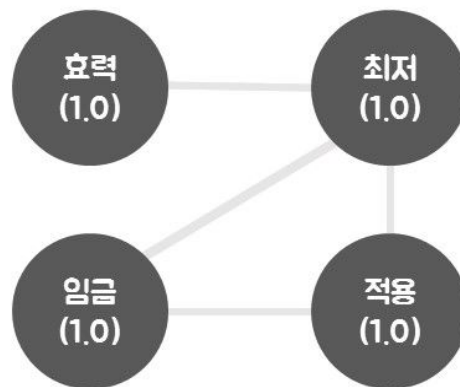
- 01 한 문장 또는 특정 윈도우 내에 같이 등장할 경우, 단어 간의 선을 연결시킨다.
- 02 모든 노드에 1의 가중치를 준다.

Text Rank

그래프 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델

효력: 사용자는 최저임금의 적용을 받는 근로자에게
최저임금 이상의 임금을 지급해야 한다.



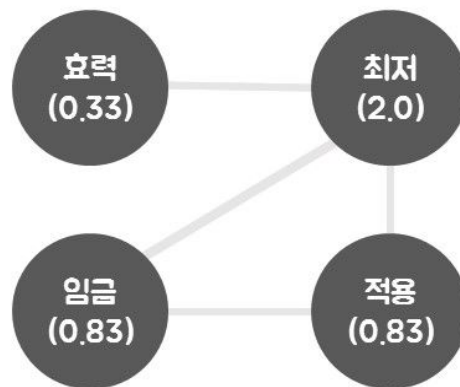
- 03 각 노드가 가지고 있는 가중치를 이웃에게 나누어 준다. 노드의 새로운 가중치 합은 이웃으로부터 넘겨받은 가중치의 합으로 업데이트된다.
- 04 이 과정을 반복하면 가중치들이 일정한 값으로 수렴한다.

Text Rank

그래프 기반 모델

문서를 벡터로 표현하여
쿼리와 문서들 간의 유사도를
판별하고, 이를 바탕으로
순위를 매기는 모델

효력: 사용자는 최저임금의 적용을 받는 근로자에게
최저임금 이상의 임금을 지급해야 한다.



- 03 각 노드가 가지고 있는 가중치를 이웃에게 나누어 준다. 노드의 새로운 가중치 합은 이웃으로부터 넘겨받은 가중치의 합으로 업데이트된다.
- 04 이 과정을 반복하면 가중치들이 일정한 값으로 수렴한다.

MODEL

유사한
질문
찾기

사용자의
새로운 질문에
법령 매칭

답변과
법령
매칭

유사 질문 찾기

유사한 질문을 올리는 사람들은
서로 유사한 상황에 처해있을 것이다

새로운 질문이 들어오면
카테고리를 분류하여 그 안에
서 유사한 질문을 찾는 모델



새로운 질문



데이터 전처리

data preprocessing

꼬꼬마 세종 말뭉치를 이용하여
질문의 형태소 분석, 불용어 제거

관측치 임베딩

document-embedding

각 질문의 벡터는 질문에 등장한 단어별
임베딩 벡터에 대한 tfidf 가중합 벡터



Word2vec 단어 임베딩

word-embedding

빈도수가 5 이상인 모든 단어에
대해 각 단어마다 100차원으로
임베딩한 벡터 생성



나이프베이즈 모델 학습

ensemble

카테고리 분류를 위한 두 번째 모델인
Tfidf 가중치를 이용한
나이프베이즈 모델 학습

SVM 모델 학습

training classifier model

질문이 속하는 법 카테고리를 분류하기
word2vec 임베딩을 기반으로 한
SVM을 학습

SVM 모델과 나이프베이즈 모델을
결합하여 voting을 통해 카테고리를 분류




유사도 비교

document-similarity

카테고리 내의 질문들과 새로운 질문 간
tfidf 가중치를 이용한 코사인 유사도와
word2vec embedding 벡터를
기반으로 한 TS_SS 유사도 계산

두 가지 유사도를 모두 고려
(tfidf 유사도가 높을수록
TS_SS 거리가 가까울수록 유사)



새로운 질문과
가장 유사한
상위 5개 질문 채택

상위 5개 질문의
답변에 대한 분석을 토대로
상황에 맞는 법령을 제공하자

각 유사질문에 대한 답변



답변-법령 매칭

변호사의 답변에는 법령의 내용이 포함되어 있을 것이다

답변과 유사도가 높은 법령을 찾아 연결하는 모델

데이터 전처리

data preprocessing

꼬꼬마 세종 말뭉치를 이용하여
법령의 형태소 분석,
법 분야별로 불용어 제거



가중치 모델 탐색

Keyword-extraction

tfidf, rake, textrank 등의 키워드 추출 알고리즘을 이용하여 각 법령의 주요 단어를 추출

tfidf 가중치 선택

Tfidf similarity

가장 키워드가 잘 추출된 tfidf를 선택

이와 같은 과정을 3~4단계를
반복하여 각 답변마다 법령의 가장
작은 소분류까지 매칭



답변과 법령 매칭

Matching answer and law

각 중분류에 속하는 법령들을 묶어
하나의 문서로 보고, 각 중분류와
답변 간의 tfidf 유사도 계산
가장 유사도가 높은 중분류 선택

유사 질문 찾기



사용자의 질문과
가장 유사한 상위 5개
질문 채택

법령

답변과 법령 매칭



유사한 질문들의
답변에 매칭된 법령을 토대로
투표를 통해 법령 제공

전 여친에게 돈을 돌려받고 싶어요.
전 여친에게 카드와 현금을 빌려줘서
450만원 가량 사용을 했는데 갚기로
해놓고 잠수를 타버렸습니다
현재 카톡은 차단 당해있고 번호도
바꾼 상태인데 돈 돌려 받을 수 있을
까요?

제가 1월초 아르바이트에서 만난 동료와
함께 여행을 다녀왔습니다. 그런데 제 카
드를 빌려 쓰고 자신이 대금을 대신 납
부하겠다고 하여 제 카드를 빌려주었습니
다. 그렇게 2주간 약 1500만원 가량의
카드 지출이 있었고 서울로 돌아와서 대
금을 납부해달라 말하자 계속 이런 저런
핑계를 대며 대금 납부기한을 미루는 것
입니다. 고소할 수 있는 지 알고 싶어요.

채무불이행에 대한 대응 개관
대여금(빌려준 돈)을 반환받지 못한
채권자는 민사소송을 통해 대여금
을

청구할 수 있습니다.

민사소송절차에 따라 집행권을
얻으면 이를 근거로 채무자의 재산
을

현금화하여 돈을 받을 수 있습니다.
(이하 생략)



사용자의 질문



유사 질문 추출



법령 선택

무면허음주걸렸습니다.
알콜농도 0.07인데
벌금얼마나나올가요
초범입니다.

음주운전 문의 어제 단속에 걸려서
혈중알콜 농도가 0.066% 나왔습니
다.현재 저는 23살이고 전역 한지
얼마 안되서 현재 놓고있는 사람입
니다. (대학생도아니구요) 여기서 공
금한거는 처음 단속이 되는 것이고,
벌금이 얼마나 나오는지 분납이 가
능한지 궁금합니다.

운전자는 술에 취한 상태 및 과로, 질병
또는 약물의 영향과 그 밖의 사유로 인해
정상적으로 운전할 수 없는 우려가 있는
상태에서 운전을 하면 30만원 이하의 벌
금이나 구류에 처해지고, 약물의 영향으
로 정상적으로 운전하지 못할 우려가 있
는 상태에서의 운전을 하면 3년 이하의
징역이나 1천만원 이하의 벌금에 처해집
니다



사용자의 질문



유사 질문 추출



법령 선택