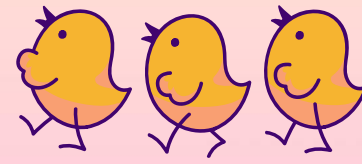


LangCon 2020



법률 텍스트에 대한 비 학습 접근

국회예산정책처,
서울대학교 행정대학원 김재윤

1. 배경: 법률 텍스트의 규칙성과 정보 추출(IE)
2. 사례: 30년간 형벌 수준 변화 추적
3. 사례: 특정 유형 조문의 추출
4. 사례: 법률 404 error 디버깅
5. 결론

1. 배경: 법률 텍스트의 규칙성과 정보 추출

- 언어와 법률 (Tiersma)
 - 법률은 자연어이나, 상당히 전형적(stereotypical)
 - 텍스트 생산이 규칙적
- 한국 법률 텍스트의 구조
 - 제목
 - 버전정보
 - 시행일, 공포번호, 공포일
 - 본문
 - 분류단위: 편-장-절
 - 조문: 조-항-호-목
 - 조에는 조제목이 있음
 - 부칙

국어기본법

[시행 2020.5.27.] [법률 제16589호, 2019.11.26., 일부개정]

문화체육관광부(국어정책과), 044-203-2533

제1장 총칙 <개정 2011. 4. 14.>

제1조(목적) 이 법은 국어 사용을 촉진하고 국어의 발전과 보전의 기반을 마련하여 국민의 창조적 사고력의 증진을 도모함으로써 국민의 문화적 삶의 질을 향상하고 민족문화의 발전에 이바지함을 목적으로 한다.

[전문개정 2011. 4. 14.]

제2조(기본 이념) 국가와 국민은 국어가 민족 제일의 문화유산이며 문화 창조의 원동력임을 깊이 인식하여 국어 발전에 적극적으로 힘씀으로써 민족문화의 정체성을 확립하고 국어를 잘 보전하여 후손에게 계승할 수 있도록 하여야 한다.

[전문개정 2011. 4. 14.]

제3조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

1. "국어"란 대한민국의 공용어로서 한국어를 말한다.
2. "한글"이란 국어를 표기하는 우리의 고유문자를 말한다.
3. "어문규범"이란 제13조에 따른 국어심의회의 심의를 거쳐 제정한 한글 맞춤법, 표준어 규정, 표준 발음법, 외래어 표기법, 국어의 로마자 표기법 등 국어 사용에 필요한 규범을 말한다.
4. "국어능력"이란 국어를 통하여 생각이나 느낌 등을 정확하게 표현하고 이해하는 데에 필요한 듣기·말하기·읽기·쓰기 등의 능력을 말한다.

[전문개정 2011. 4. 14.]

제27조(권한의 위임·위탁) ① 이 법에 따른 문화체육관광부장관의 권한은 대통령령으로 정하는 바에 따라 그 일부를 특별시장·광역시장·도지사 또는 특별자치도지사에게 위임할 수 있다.

② 문화체육관광부장관은 이 법에 따른 업무의 일부를 대통령령으로 정하는 바에 따라 관련 기관·단체 등에 위탁할 수 있다.

[전문개정 2011. 4. 14.]

부칙 <법률 제7368호, 2005. 1. 27.>

제1조 (시행일) 이 법은 공포 후 6월이 경과한 날부터 시행한다.

제2조 (다른 법률의 폐지) 한글전용에관한법률은 폐지한다.

제3조 (공문서의 작성에 관한 적용례) 제14조의 규정은 이 법 시행 후 최초로 작성하는 공문서부터 적용한다.

제4조 (어문규범에 관한 경과조치) 이 법 시행 당시 종전의 문화예술진흥법 제7조의 규정에 의한 어문규범은 제11조의 규정에 의한 어문규범으로 본다.

제5조 (국어심의회에 관한 경과조치) 이 법 시행 당시 종전의 문화예술진흥법 제6조의 규정에 따라 설치된 국어심의회는 제13조의 규정에 따라 설치된 국어심의회로 본다.

제6조 (다른 법률의 개정) 문화예술진흥법 중 다음과 같이 개정한다.

제2장(제5조 내지 제8조)을 삭제한다.

1. 배경: 법률 텍스트의 규칙성과 정보 추출

- 닭 잡는데 어찌 소 잡는 칼을 쓰시려 하십니까?
- 닭 = 법률
- 소 = 비정형 텍스트
- 닭 잡는 칼 = 정보 추출 = 화웅
- 소 잡는 칼 = NLP = 여포



- 규칙 기반(Rule-based) NLP
 - 1980년대 '통계혁명' 이후 NLP는 통계, 기계학습 등을 이용하게 되었다.
 - 2010년대 이후에는 뉴럴넷 발전, NLP 도구의 대중화 등으로 새로운 국면
 - 그러나, 옛날 옛적 언어처리 시스템은 hand-coded 규칙으로 작동했다.
- 정보 추출(Information Extraction)
 - NLP의 끝자락 또는 정보검색(Information Retrieval)과 NLP 사이의 분야
 - 정형화되지 않은 문서에서 자동으로 정형화된 데이터를 추출하는 기술
 - <https://www.slideserve.com/ohio/scalable-information-extraction-and-integration>

- 정보 추출의 익숙한 예
: html parser
 - BeautifulSoup
 - 정규표현식
- 법률이 html과 닮은 점: 규칙성
 - 법률 데이터 자체가 html로 제공
 - 다만, 법률의 규칙은 html보다는 더 낮은 수준이라는 차이점

```
27 <h1 class="bon_tit">국어기본법</h1>
28 <p class="bon_revision">
29 [시행 2017.9.22.]
30
31
32
33 [법률 제14625호, 2017.3.21., 일부개정]
34
35 </p>
36 <p class="department">&nbsp;</p>
37 </div>
38 <style>
39 table {border-collapse: collapse; border: 1px;}
40 .INL_TSTYLE { display: inline-table; }
41 .INL_ASTYLE { vertical-align: bottom; }
42 </style>
43 <div class="PRE_CONT"></div><div id="제1장"></div><h2
44 class="jang" id="제1장" class="">제1장 총칙<span class="revision"> &lt;개정
class="jang" id="제1장" class="">제1장 총칙<span class="revision"> &lt;개정
2011.4.14.&gt;</span>
45 </h2><div class="jo_area" id="제1장제1조"><div class="jo_txt">
<strong class="jo" class="">제1조(목적<span class="revision"></span>)
</strong> 이 법은 국어 사용을 촉진하고 국어의 발전과 보전의 기반을 마련하여
국민의 창조적 사고력의 증진을 도모함으로써 국민의 문화적 삶의 질을 향상하고
민족문화의 발전에 이바지함을 목적으로 한다.<p></p>
46 <span class="revision">[전문개정 <a
href="javascript:joHistCall({contId: '20050127000000003', contSid: '0008',
tocAno: '제1조', promDt: '20110414', basicDt: '20200116', viewGb: 'EXE',
hanTranceYn: 'Y'});">2011.4.14.</a>]</span></div><dl class="jo_btn">
<dt class="hidden">조문버튼</dt>
```

1. 배경: 법률 텍스트의 규칙성과 정보 추출

Dataframe, csv, xls

xlm, json, html

txt, doc, pdf, hwp

정형 데이터

반 정형 데이터

비정형 데이터

범죄 통계

법안처리 통계

법률

법안

판례

회의록

국민청원

	C	D	E
1	죄종별(3)	2018	
2		발생건수 (건)	검거건수 (건)
3	소계	1,580,751	1,328,609
4	소계	26,787	25,811
5	소계	309	298
6	살인	226	217
7	영아살해	7	7
8	존속살해	44	43
9	촉탁·승낙살인	1	1
10	자살교사·방조	30	29
11	위계·위력·촉탁·강요	-	-
12	특가법(보복살인)	1	1
13	소계	488	484

```

ass="floatL w40P">
v class="pr30">
<div class="subti01 subti01_" >
<h4>정무이송</h4>
</div>
<div class="contln">
<h5 class="subti02">▶ 정 무이송정보</h5>
<div class="tableCol01">
<table summary="정 무이송정보">
<caption class="hidden">정 무이송정보</caption>
<colgroup>
<col style="width:100px" /><col>
</colgroup>
<thead>
<tr>
<th scope="col">정 무이송일</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-01-23</td>
</tr>
</tbody>
</table>
</div>
</div>
liv>
    
```

법률 제 호

정보통신망 이용촉진 및 정보보호 등에 관한 법률

일부개정법률안

정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부를 다음과 같이 개정한다.

제1조 중 “자의 개인정보”를 “자”로 한다.

제2조제1항제6호를 삭제한다.

제3조제1항 중 “이용자의 개인정보”를 “이용자”로 하고, 같은 조 제3항 중 “개인정보보호”를 “정보보호”로 한다.

제4조제1항 중 “이용자의 개인정보보호”를 “이용자 보호”로 하고, 같은 조 제2항제6호 및 제6호의2를 삭제한다.

제5조 단서 중 “제4장의 개인정보의 보호에 관하여 이 법과 「개인정보 보호법」의 적용이 경합하거나 제7장을 “제7장”으로 한다.

제4장의 제목 “개인정보의 보호”를 “정보통신서비스의 안전한 이용환경 조성”으로 한다.

제22조 앞의 “제1절 개인정보의 수집·이용 및 제공 등”을 삭제한다.

제22조를 삭제한다.

제23조를 삭제한다.

제23조의2제1항제2호를 삭제하고, 같은 항 제3호를 다음과 같이 한다.

저작권침해금지규칙

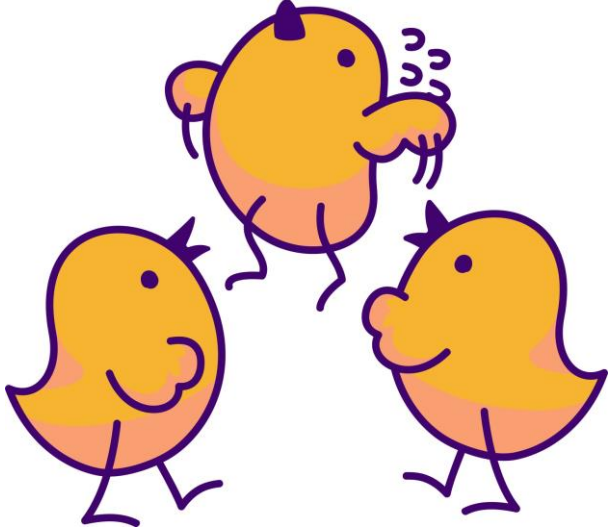
[대법원 2001.6.29, 선고, 99다23246, 판결]

【판시사항】

- [1] 서체파일이 구 컴퓨터프로그램보호법 제2조 제1호에서 정한 컴퓨터프로그램에 해당한다고 판단한 사례
- [2] 윤곽선의 수정 내지 제작작업을 한 서체파일에 창작성이 인정되는지 여부(적극)
- [3] 신청인과 피신청인의 서체파일의 소스코드가 동일하다면 피신청인의 서체파일은 신청인의 서체파일에 의존하여 작성된 것으로 추정되므로 프로그램저작권 침해를 인정함이 상당하다고 한 사례

【판결요지】

- [1] 구 컴퓨터프로그램보호법(1998. 12. 30. 법률 제5605호로 개정되기 전의 것)상의 컴퓨터프로그램은 특정한 결과를 얻기 위하여 컴퓨터 내에서 직접 또는 간접으로 사용되는 일련의 지시·명령으로 표현된 것으로 정의되는바, 이 사건 서체파일의 소스코드는 ① 그것이 비록 다른 응용프로그램의 도움 없이는 바로 실행되지 아니한다고 하여도 컴퓨터 내에서 특정한 모양의 서체의 윤곽선을 크기, 장평, 굵기, 기울기 등을 조절하여 반복적이고 편리하게 출력하도록 특정한 결과를 얻기 위하여 프로그래밍 언어의 일종인 포스트스크립트(PostScript) 언어로 제작된 표현물이고, ② 서체파일 제작 프로그램에서 마우스의 조작으로 서체의 모양을 가감하거나 수정하여 좌표값을 지정하고 이를 이동하거나 연결하여 저장함으로써, 제작자가 특정한 결과를 얻기 위하여 스스로의 알고리즘(algorithm)에 따라 프로그래밍 언어로 직접 코드를 작성하는 보통의 프로그램 제작과정과는 다르다 하여도, 포스트스크립트 언어로 작성되어 사람에게 이해될 수 있고 그 내용도 좌표값과 좌표값을 연결하는 일련의 지



- 김재윤(2019) 논문에서 가져왔습니다.
- 일부 내용은 PyCon 2019 “법률을 디버깅하다” 발표에서 가져왔습니다.

2. 사례: 30년간 형벌 수준 변화 추적

2. 사례: 30년간 형벌 수준 변화 추적

- **징역 7년**의 위기를 겪은 박나래씨
 - 수제향초를 제작해서 선물했다는 혐의
 - 환경부의 행정지도에 따라 수제향초를 모두 수거하여 처벌을 모면
- 신설된 규제와 형벌규정 때문
 - 가습기 살균제 사건을 계기로 사전에 화학안전성 검사를 받도록 규제 신설
- 법률상 **형벌 규정의 수와 수준**은 어떻게 변해 왔는가?
 - 참고: 이 작업을 과거에 여러 명의 직원을 동원하여 한 시점에 대해서만 수작업으로 했다



2. 사례: 30년간 형벌 수준 변화 추적

1. 수집

- 시점별 현행법에 해당하는 법률 웹페이지 주소를 생성
- 11~19대국회 말을 기준시점으로 9번 웹크롤링(1985, 1988, 1992 ... 2016.5)

2. 가공

- 텍스트를 문장별로 분리하고, 형벌 조항만 발라낸다.

3. 정보 추출

- 징역이나 벌금에 해당하는 숫자만 가져온다.

2. 사례: 30년간 형벌 수준 변화 추적

1. 수집 결과: 모든 법률 텍스트를 1개 조문당 한칸씩으로 정리

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
140	국회법													
141	국회법	제1조(목적) 이 법은 국회의 조직·의사 기타 필요한 사항을 규정함으로써 국민의 대의기관인 국회의 민주적이고 효율적인 운영에 기여함을												
142	국회법	제2조(당선통지 및 등록) ① 중앙선거관리위원회위원장은 국회의원당선인이 결정된 때에는 그 명단을 즉시 국회에 통지하여야 한다.;② 국												
143	국회법	제3조(의석배정) 국회의원(이하 "의원"이라 한다)의 의석은 의장이 각 교섭단체대표의원과 협의하여 이를 정한다. 다만, 협의가 이루어지지												
144	국회법	제4조(정기회) 정기회는 매년 9월 1일에 집회한다. 그러나 그 날이 공휴일인 때에는 그 다음날에 집회한다. 개정 2000.2.16.												
145	국회법	제5조(임시회) ① 임시회의 집회요구가 있을 때에는 의장은 집회기일 3일전에 공고한다. 이 경우 2이상의 집회요구가 있을 때에는 집회일												
146	국회법	제5조의2(연간 국회운영기본일정등) ① 의장은 국회의 연중 상시운영을 위하여 각 교섭단체대표의원과의 협의를 거쳐 매년 12월 31일까지												
147	국회법	제5조의3(법률안제출계획의 통지) 정부는 부득이한 경우를 제외하고는 매년 1월 31일까지 당해연도에 제출할 법률안에 관한 계획을 국회												
148	국회법	제6조(개회식) 국회는 집회일에 개회식을 행한다. 다만, 임시회의 경우에는 개회식을 생략할 수 있다. 개정 2000.2.16.												
149	국회법	제7조(회기) ① 국회의 회기는 의결로 이를 정하되, 의결로 연장할 수 있다.;② 국회의 회기는 집회후 즉시 이를 정하여야 한다.												

2. 가공 결과: 텍스트를 문장별로 분리하고, 형벌 조항만 발라낸다

	A	B
1	title	text
2	국회법	제166조(국회의회의방해죄)①제165조를위반하여국회의회의를방해할목적으로회의장또는그부근에서폭행,체포감금,협박,주거침입퇴거를
3	국회법	②제165조를위반하여국회의회의를방해할목적으로회의장또는그부근에서사람을상해하거나,폭행으로상해에이르게하거나,단체또는다

2. 사례: 30년간 형벌 수준 변화 추적

```
text = '제166조(국회회의방해죄) ① 제165조를  
위반하여...방해한자는5년이하의징역또는1천  
만원이하의벌금에처한다.'
```

```
if keywords[0] in text:  
    loc = text.find(keywords[0])  
    result_text = re.findall(  
        "\d+", text[loc - 2 : loc])  
    year_0 = float(result_text[0])
```

- 키워드 위치 찾기
 - '년이하의징역', '천만원이하의벌금'
- 정규표현식(re)으로 키워드 위치 바로 앞에서 숫자만 추출
 - "Wd+"
 - 결과: 5, 1
- 추출된 형량을 저장

2. 사례: 30년간 형벌 수준 변화 추적

keywords =

['년이하의징역', '년이하의금고', '년이상징역',
'년이상금고', '년상유기징역',
'년상유기금고', '월이하징역',
'월이하금고', '사형', '무기징역', '무기금고',
'무기또는', '억원이하벌금',
'천만원이하벌금', '백만원이하벌금',
'만원이하벌금', '가중처벌', '가중한다.']

- 형벌규정의 규칙성을 이용
 - 실제 법률과 추출된 정보를 대조하여 빠진 키워드가 없는지 확인 과정 필요
- 겹치는 키워드는 논리식을 이용하여 경우의 수를 조정
 - '천만원이하벌금'과 '만원이하벌금'의 경우

2. 사례: 30년간 형벌 수준 변화 추적

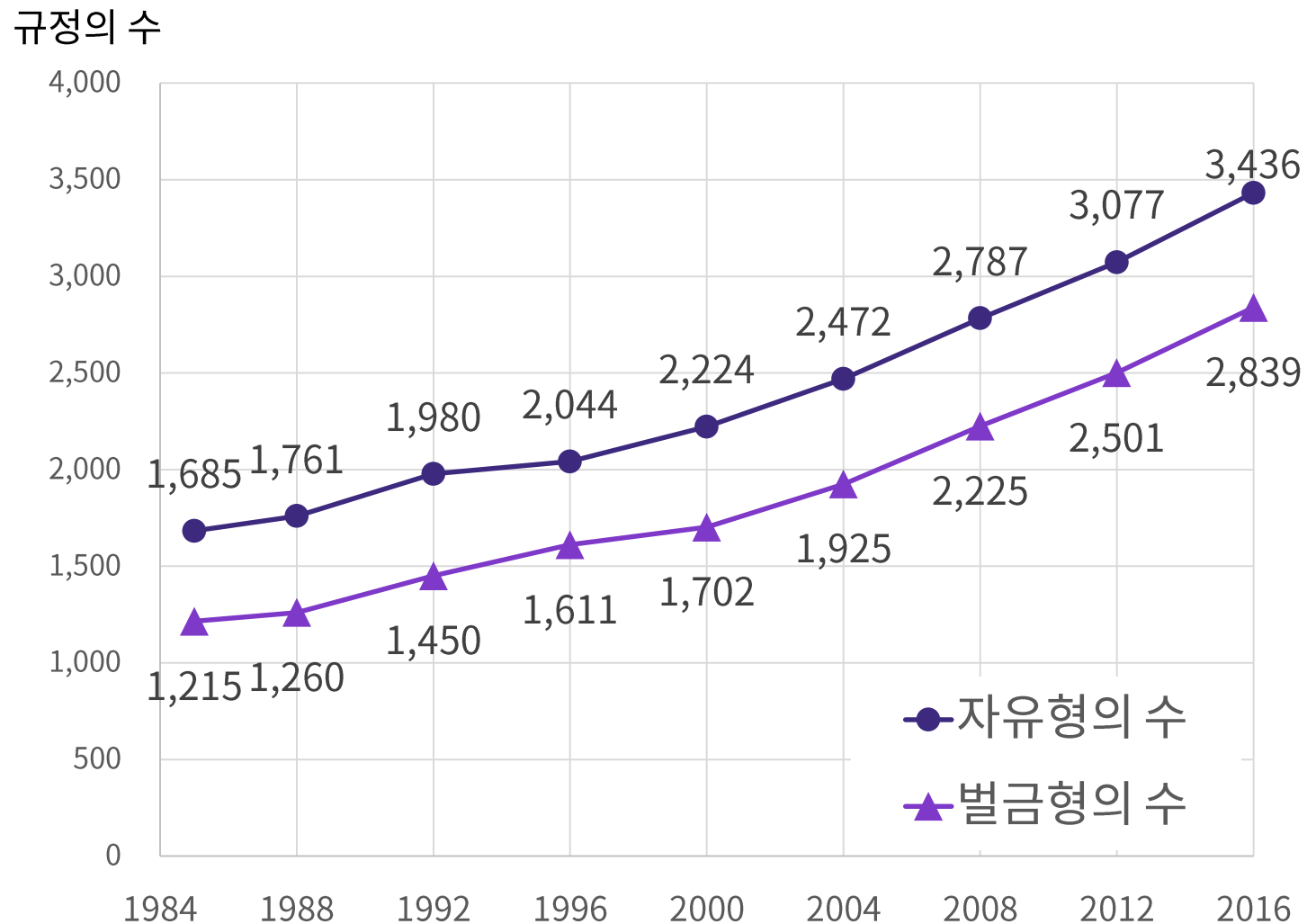
- 정보 추출 과정에서 만난 장애물
 - 띄어쓰기 문제
 - 띄어쓰기가 틀린 조문이 있어서, 띄어쓰기를 모두 제거한 후 키워드 매칭 수행
 - '~년 이상의 징역'과 '~년 이상의 유기징역'의 차이
 - 전자는 무기징역을 포함하고, 후자는 형법상 유기징역 상한까지(제도 변화 있었음)
 - '벌금'과 '벌금', '김고'와 '금고'
 - 크롤링한 웹페이지에서 애초에 한자 '罰金'을 '벌금'으로 잘못 변환한 경우
 - 체계적인 측정오차의 우려!: 이 오류는 분석결과를 의미있게 다르게 만들

- 추출결과: 앞서 예시한 제166조 ①의 '5'와 '1'을 성공적으로 추출

[illegible]

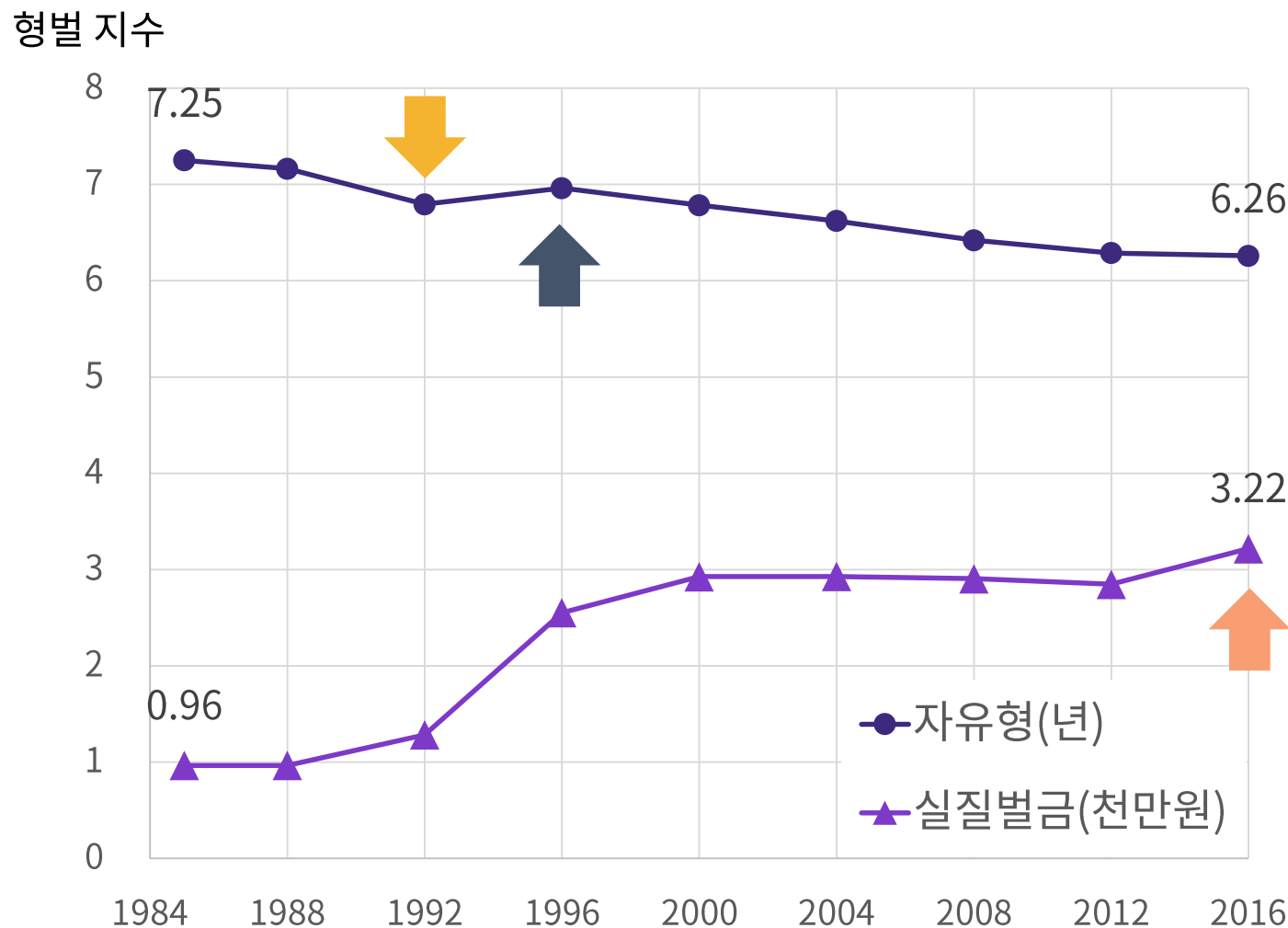
2. 사례: 30년간 형벌 수준 변화 추적

- 자유형 연2.3% 증가
 - 자유형 = 사형, 무기징역, 유기징역
 - 31년간 2.0배 증가
- 벌금형 연2.8% 증가
 - 31년간 2.3배 증가
- 참고로 전체 법률 조문의 성장률은 2.7% (Kim, 2019)



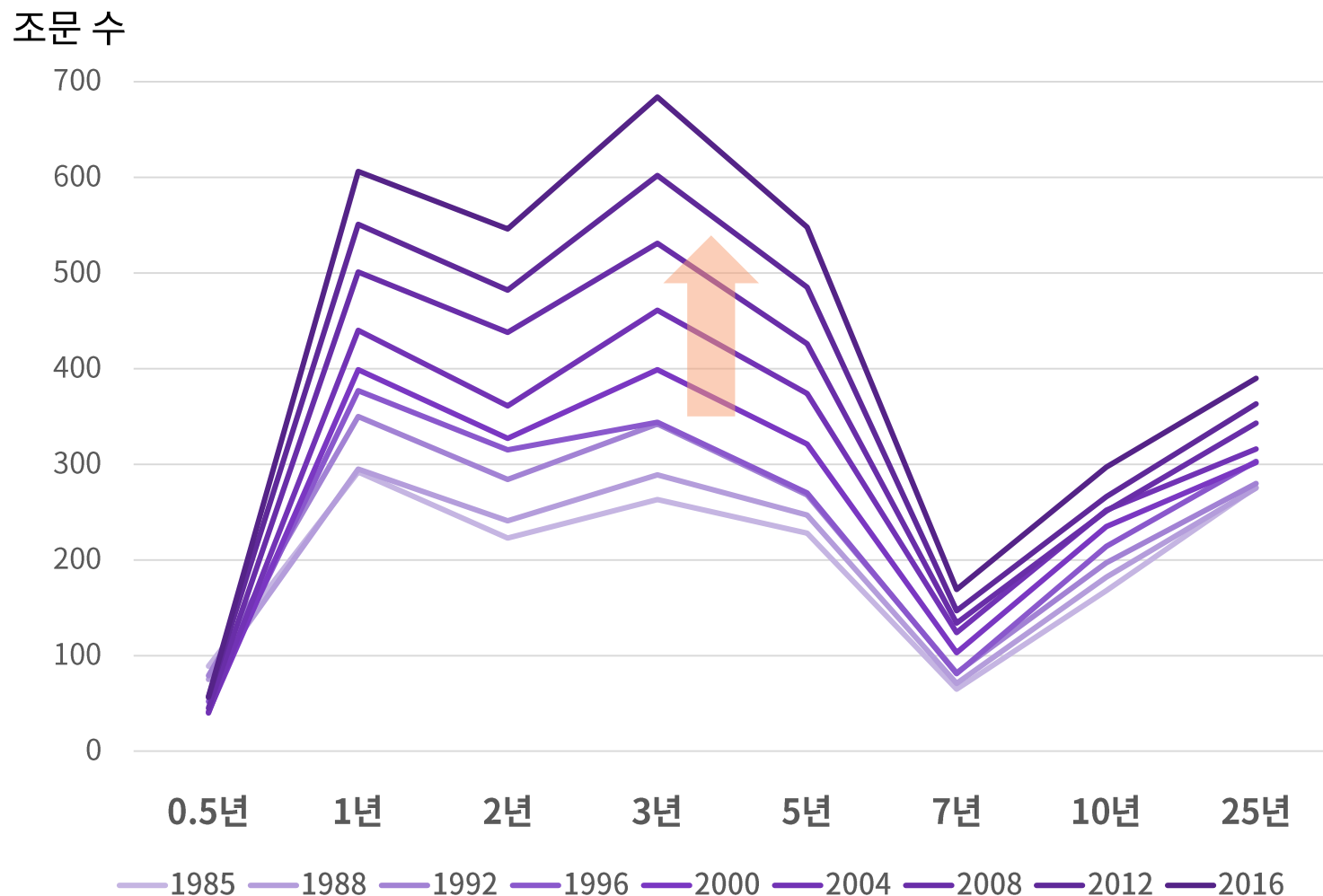
2. 사례: 30년간 형벌 수준 변화 추적

- 자유형 지수 하락
 - '85년 7.25년 → '16년 6.26년
- 시기별 트렌드 분석
 - 5공화국 법령 정비
 - 건설안전사고 대응 입법
 - 법정형 정비사업
- 의문: 입법자들이 정말 형량을 낮춘 것인가?



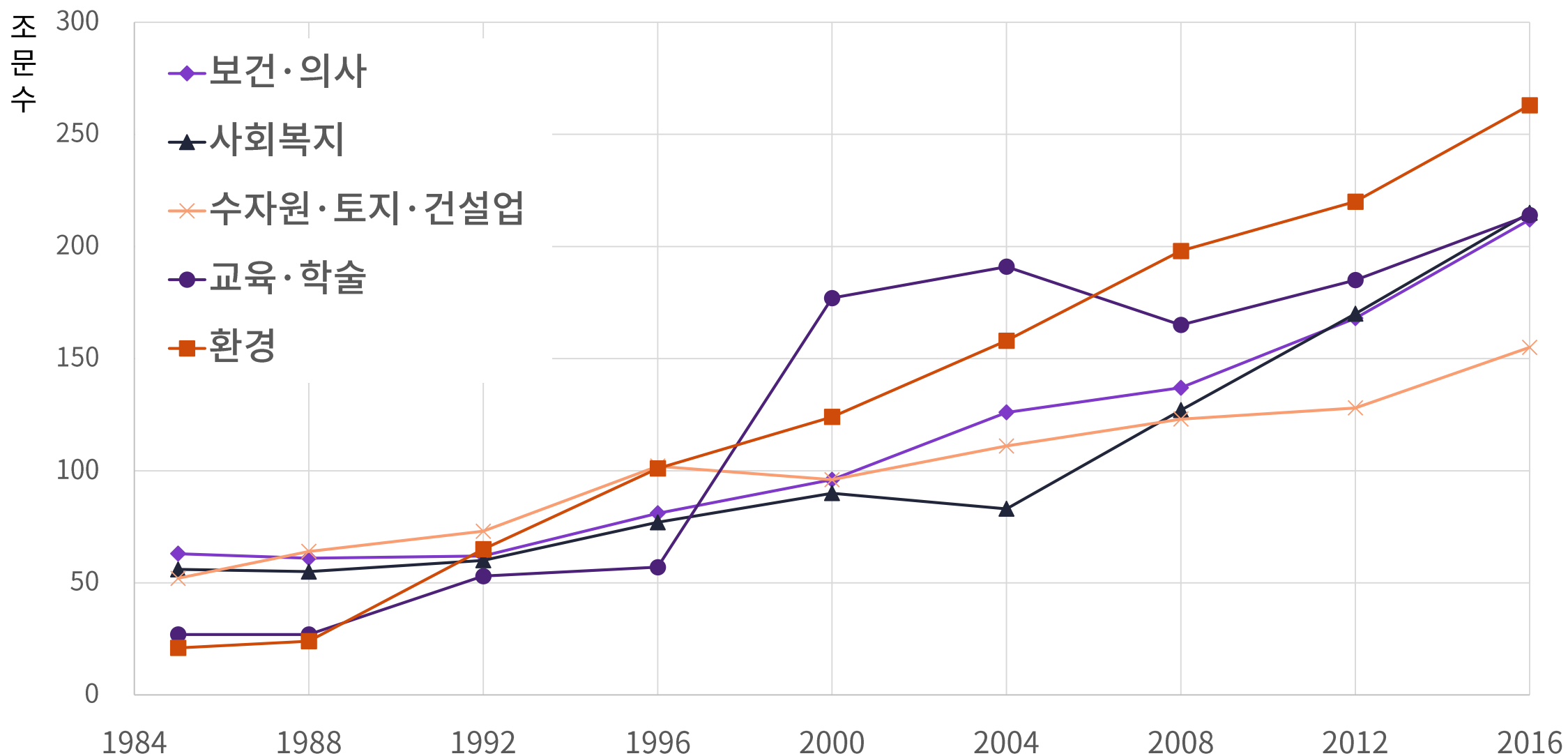
2. 사례: 30년간 형벌 수준 변화 추적

- 답: **착시효과**이다.
 - 가벼운 형벌이 더 빨리 늘어나면서 평균이 낮아졌을 뿐
- 모든 수준에서 징역형, 사형의 수 증가
 - 입법자가 형벌을 감경했다고 보기 어려움



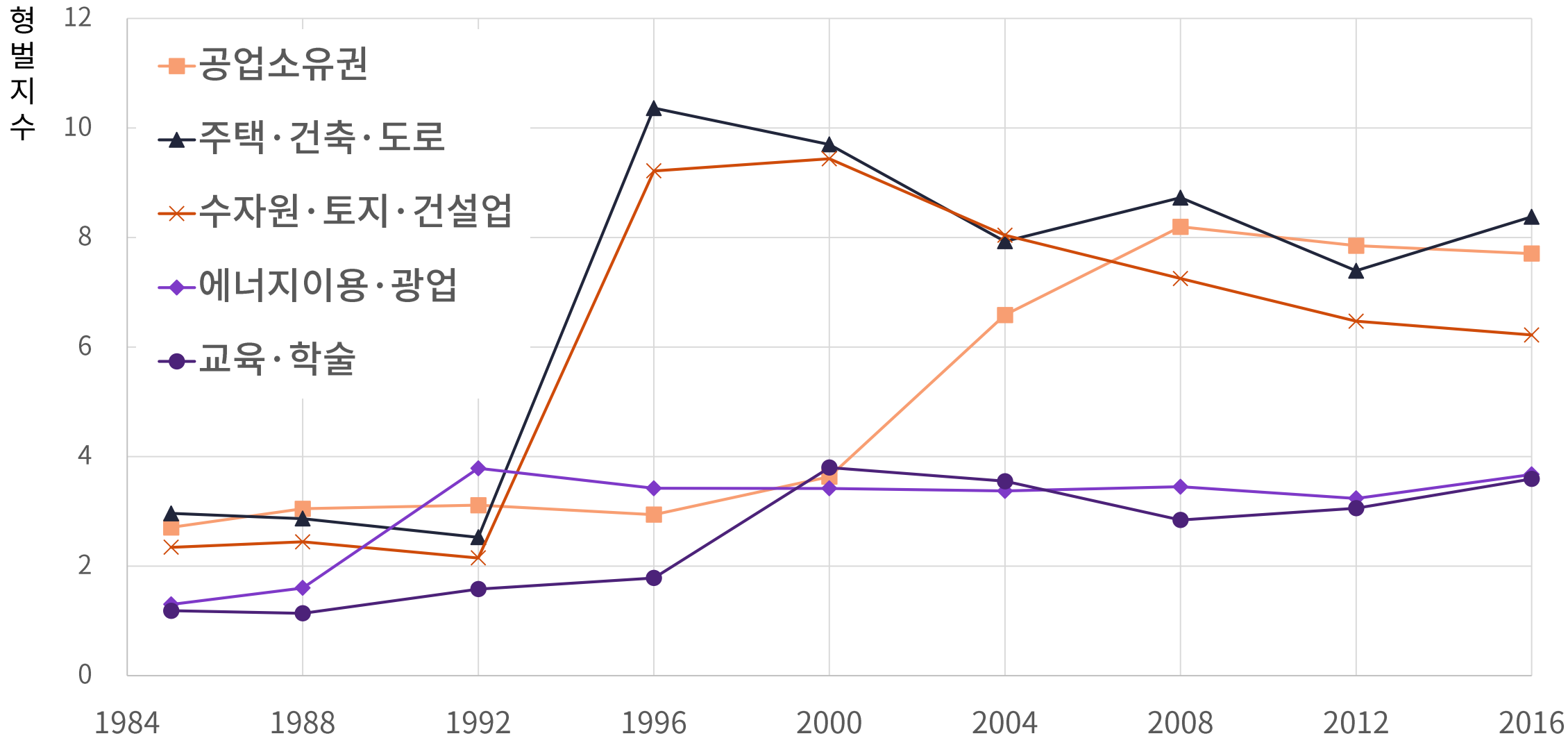
2. 사례: 30년간 형벌 수준 변화 추적

- 법률 분야별 분석 결과: 조문 수가 빠르게 증가한 상위 5개 편



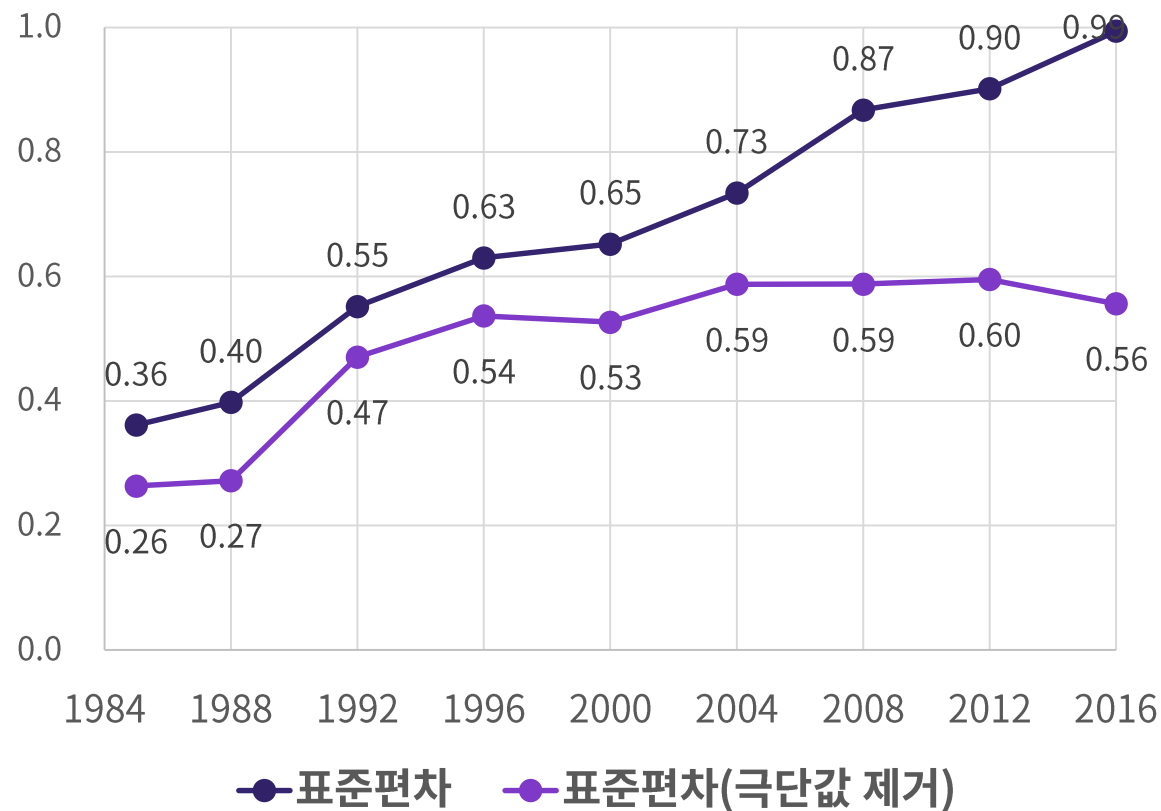
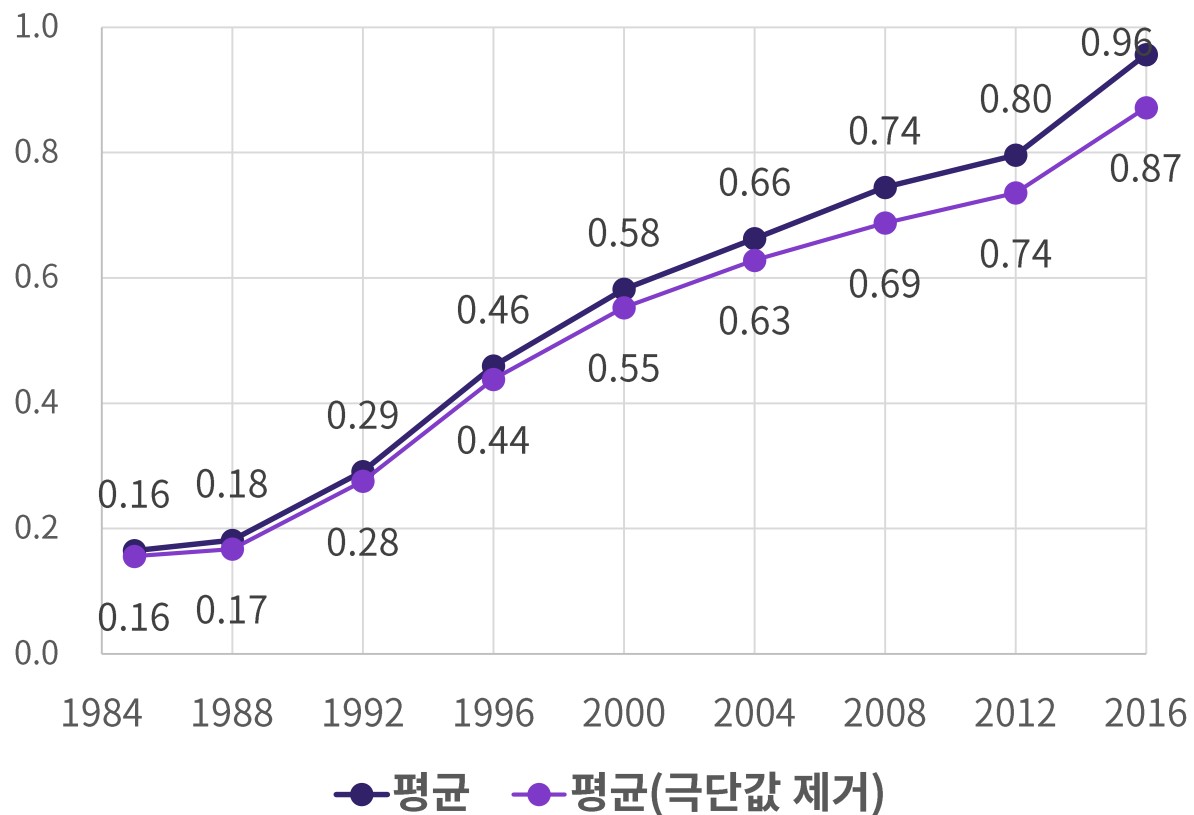
2. 사례: 30년간 형벌 수준 변화 추적

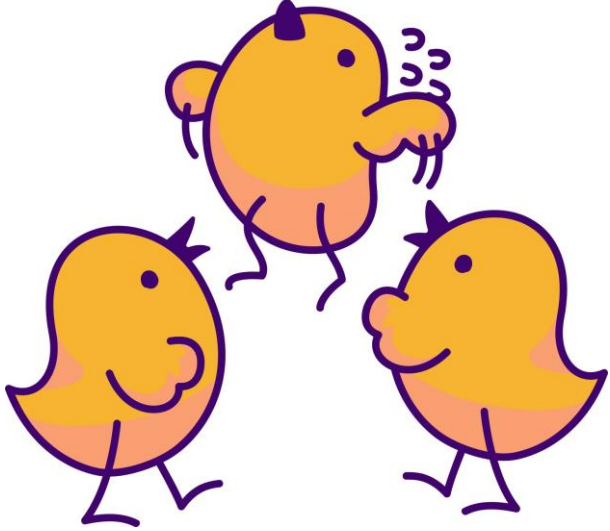
- 법률 분야별 분석 결과: 형벌지수가 빠르게 높아진 상위 5개 편



2. 사례: 30년간 형벌 수준 변화 추적

- 징역 1년 대비 벌금액 1천만원의 비율





- 이 부분은 기획 중인 연구프로젝트에서 가져온 것입니다. 활용이나 인용에 유의하여 주시기 바랍니다.

3. 사례: 특정 유형 조문의 추출

3. 사례: 특정 유형 조문의 추출

- '가치조항'이란? 좋은 말

- 법률은 일반적으로 무엇인가 강제하기 위해서 존재한다는 통념과는 다른 법조문들
- 강제력이 없으며, 사회적 가치를 담고 있다

- 예시1: 인문학 및 인문정신문화의 진흥에 관한 법률

- 제2조(**기본이념**) 이 법은 인문학 및 인문정신문화의 진흥이 인간의 존엄을 바탕으로 사회적·문화적 가치와 조화를 이루고 경제·사회 발전의 원동력이 되도록 하며, 국민의 자율성과 창의성이 존중받도록 하고, 인문학이 자연과학 및 사회과학과 균형있게 발전하도록 함을 기본이념으로 한다.

- 예시2: 문학진흥법

- 제3조(**국가와 지방자치단체의 책무**) ① 국가와 지방자치단체는 문학 진흥에 관한 시책을 강구하고, 문학 창작 및 향유와 관련한 국민의 활동을 권장·보호·육성하도록 노력하여야 한다.
② 제1항에 따른 문학 진흥 시책은 국민생활의 질적 향상을 위한 생활 문화의 개발·보급에 관한 사항을 포함하여야 한다.
③ 국가와 지방자치단체는 제1항에 따른 책무를 다하기 위하여 이에 수반되는 예산상의 조치를 취하도록 노력하여야 한다.

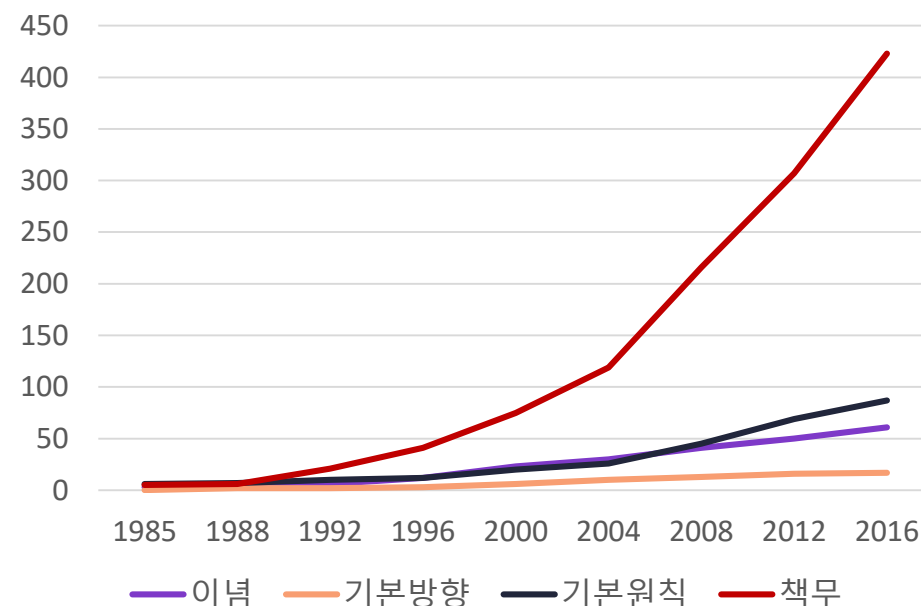
- 가치조항의 분류 = 조 제목으로 파악 가능
 - 목적규정
 - 목적 규정은 그 법령의 입법 목적을 간결하고 명확하게 요약한 문장
 - 기본이념에 관한 규정
 - 법령의 기본이념 규정은 법령의 제정 이념이나 정신을 표현한 것
 - 그 법령의 조문을 통해서 구현되어야 하는 이념에 관한 선언적 규정
 - 국가 등의 책무·책임 등에 관한 규정
 - 법령의 목적 달성을 위해 국가나 지방자치단체 등이 수행해야 할 책무 등을 정한 것
- 가치조항의 의의
 - 국민 또는 기업 등에 대해서 입법자들이(유권자들이) 바라는 바를 직접 당부하는 취지의 법률 규정
 - 강제적이고 재정적인 수단보다는, 설득적이고 권고적인 수단을 사용하는 경향

- 정보 추출 전략: 키워드로 식별
 - 조 제목에 관련된 키워드가 있는 경우를 찾기
- 데이터: 김재윤(2019, 형벌조항 전수조사) 이용
- 분석 과정
 - 법률 데이터 획득
 - 법률 데이터 전처리
 - 목표한 조문 찾기
 - 유형별 조문의 기술통계와 텍스트 데이터 얻기

3. 사례: 특정 유형 조문의 추출

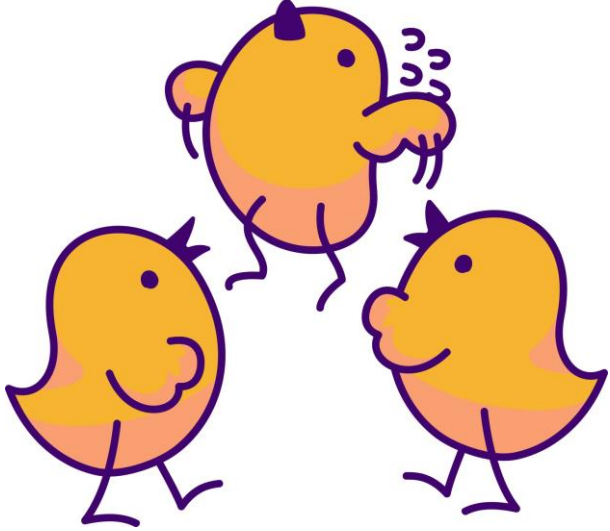
	year	총 법률	기본법	총 조항	가치조항 합계	목적	이념	방향	원칙	책무
0	1985	751	9	31173	663	649	3	0	6	5
1	1988	788	9	32864	692	673	4	2	7	6
2	1992	875	12	36457	802	763	6	2	10	21
3	1996	928	18	39837	891	823	12	3	12	41
4	2000	986	24	43845	1016	892	23	6	20	75
5	2004	1104	36	48968	1198	1013	30	10	26	119
6	2008	1263	49	57446	1497	1182	41	13	45	216
7	2012	1362	57	63245	1719	1277	50	16	69	307
8	2016	1500	65	72820	2051	1463	61	17	87	423

가치 조항의 증가 추세(목적 제외)



law

10519	자원순환기본법	제3조(기본원칙) 국가 및 지방자치단체와 사업자, 국민 등 사회의 모든 구성원은 자...	2016	39	원칙
10520	자원순환기본법	제5조(국가 및 지방자치단체의 책무) ① 국가는 자원순환사회로의 전환을 종합적·체계...	2016	39	책무
10521	자원순환기본법	제6조(사업자의 책무) ① 사업자는 자원의 투입과 에너지 사용을 효율적으로 하고 공...	2016	39	책무
10522	자원순환기본법	제7조(국민의 책무) ① 모든 국민은 자연환경과 생활환경을 청결히 유지하고, 1회용...	2016	39	책무
10523	수중레저활동의안전및활성화등에관한법률	제1조(목적) 이 법은 수중레저활동의 안전과 질서를 확보하고 수중레저활동의 활성화 ...	2016	42	목적



- 김재윤·최선영(2019) 논문에서 가져왔습니다.
- 일부 내용은 PyCon 2019 “법률을 디버깅하다” 발표에서 가져왔습니다.

4. 사례: 법률 404 error 디버깅

- '특례제한법'이라는 입법 장치
- 취지
 - 국회 스스로 특례입법을 규제하자
 - 특례란, 원칙의 예외가 되는 규정으로서, 많을 수록 복잡성 증대
- 규칙
 - 특례를 규정하려면, 특례제한법 별표(우측 그림)에 먼저 목록을 규정
 - 이 목록에 있어야만, 개별법에서 특례를 정할 수 있다.
 - 일종의 목록 관리 규제

[별표] <개정 2017. 1. 17.>

이 법에 따라 설치하는 부담금(제3조 관련)

1. 「개발이익 환수에 관한 법률」 제3조에 따른 개발부담금
2. 「개발제한구역의 지정 및 관리에 관한 특별조치법」 제21조에 따른 개발제한구역 보전부담금
3. 「고압가스 안전관리법」 제34조의2에 따른 안전관리부담금
4. 「공적자금상환기금법」 제5조제4항 및 제5항에 따른 공적자금상환기금 출연금
5. 「공항소음 방지 및 소음대책지역 지원에 관한 법률」 제17조에 따른 소음부담금

4. 사례: 법률 404 error 디버깅

- 그러나, 아래와 같이 **목록과 개별 규정이 불일치하는 경우가 있다.**
- 「집단에너지사업법」 제18조(건설비용 **부담금**)
사업자는 공급시설 건설비용을 사용자에게 부담시킬 수 있다.

404 **Law** not found!



- 「부담금관리 기본법」 **부담금**은 **별표**에 규정된 법률에 따르지 아니하고는 설치할 수 없다.
- 별표에 그런 법 없는데? → 소송 (대법원 2013다25927, 25934)
 - 별표에 없는 특례를 규정했는데, 이 규정에 따라 부담금을 부과할 수 있는가?

분석 순서

1. 특례 목록에 있는 법조문을 찾아가는 URL 생성
2. 해당 URL에서 조문 텍스트 추출
3. 키워드 검사 (고정 키워드, 케이스마다 변하는 키워드)
4. 이상이 있는 경우를 보고

찾아 갈 URL 생성

- 국가법령정보
한글주소 서비스

- 예:
<http://www.law.go.kr/법령/도시철도법/제10조>

법령

신뢰할 수 있는 법제처
국가법령정보센터

화면내검색 漢 漢 漢

도시철도법

[시행 2019. 7. 1.] [법률 제16146호, 2018. 12. 31., 타법개정]

국토교통부(광역도시철도과), 044-201-3959

판 규

제10조(토지 등의 수용 또는 사용) ① 도시철도건설자는 도시철도건설사업을 위하여 필요하면 「공익사업을 위한 토지 등의 취득 및 보상에 관한 법률」 제3조에 따른 토지·물건 및 권리(이하 "토지 등"이라 한다)를 수용 또는 사용할 수 있다.

② 제7조제1항에 따른 사업계획의 승인과 같은 조 제6항에 따른 고시는 「공익사업을 위한 토지 등의 취득 및 보상에 관한 법률」 제20조제1항 및 제22조에 따른 사업인정 및 사업인정고시로 보며, 재결신청(裁決申請)의 기한은 같은 법 제23조제1항 및 제28조제1항에도 불구하고 제7조제1항에 따라 승인을 받은 사업계획에서 정한 도시철도사업기간의 종료일로 한다.

③ 토지 등의 수용 또는 사용에 관하여는 이 법에 규정이 있는 경우를 제외하고는 「공익사업을 위한 토지 등의 취득 및 보상에 관한 법률」을 준용한다.

4. 사례: 법률 404 error 디버깅

텍스트 가져오기

'http://www.law.go.kr/법령/도시철도법/제10조' html 소스

```
1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
2 <html xmlns="http://www.w3.org/1999/xhtml" lang="ko" xml:lang="ko" style="height: 100%;">
3 <head>
4 <style type="text/css">
5 body {
6     overflow: hidden;
7     width: 100%;
8     height: 100%;
9 }
10 </style>
11 </head>
12 <body topmargin="0" leftmargin="0" rightmargin="0" bottommargin="0" marginheight="0" marginwidth="0">
13 <table cellpadding="1" cellspacing="0" border="0" height="100%" width="100%" style="border: 1px solid #cacaca;">
14 <tbody>
15 <tr>
16 <td> <iframe id="lawService" name="lawService" title="lawService" scrolling="auto" frameborder="0" style="position: absolute; left: 0px; right: 0px; top: 0px; bottom: 0px; width: 100%; height: 100%; marginheight="0"
17 marginwidth="0" tabindex="-1" src="/LSW//lsSideInfo.do?lsiSeq=206364&icNo=0010&icBrNo=00&docCls=io&urlMode=lsScJorItInfoR" >/iframe> </td>
18 </tr>
19 </tbody>
20 </table>
21 </body>
22 </html>
```

1차 크롤링: 이 네모 안의 주소를 얻어서 2차 크롤링에 사용

```
670 <div class="lawcon">
671 <p class="pty1_de2_1"> ① <input type="checkbox" id="Y001000" name="joNoList" value="10:0" onclick="javascript:getJoSelectLayer(this, '');" style="vertical-align: middle;" /> <span class="bl"><label for="Y001000">제10조(토지 등
의 수용 또는 사용)</label></span> ① 도시철도건설자는 도시철도건설사업을 위하여 필요하면 <a title="팝업으로 이동" class="link sfon1" onclick="javascript:fncLsLawPop('1002172690', 'ALLJO');" href="javascript:;">「공익사업을 위한
토지 등의 취득 및 보상에 관한 법률」</a> <a title="팝업으로 이동" class="link sfon2" onclick="javascript:fncLsLawPop('1002172691', 'JO');" href="javascript:;">제3조</a>에 따른 토지·물건 및 권리(이하 "토지등"이라 한다)를 수용 또
는 사용할 수 있다.</p>
672 <p class="pty1_de2_1"> ② <a title="팝업으로 이동" class="link sfon2" onclick="javascript:fncLsLawPop('1002172692', 'JO');" href="javascript:;">제7조</a> <a title="팝업으로 이동" class="link sfon3"
onclick="javascript:fncLsLawPop('1002172693', 'JO');" href="javascript:;">제1항</a>에 따른 사업계획의 승인과 같은 조 제6항에 따른 고시는 <a title="팝업으로 이동" class="link sfon1"
onclick="javascript:fncLsLawPop('1002172683', 'ALLJO');" href="javascript:;">「공익사업을 위한 토지 등의 취득 및 보상에 관한 법률」</a> <a title="팝업으로 이동" class="link sfon2"
onclick="javascript:fncLsLawPop('1002172685', 'JO');" href="javascript:;">제20조</a> <a title="팝업으로 이동" class="link sfon3" onclick="javascript:fncLsLawPop('1002172686', 'JO');" href="javascript:;">제1항</a> 및 <a title="팝업
으로 이동" class="link sfon2" onclick="javascript:fncLsLawPop('1002172684', 'JO');" href="javascript:;">제22조</a>에 따른 사업인정 및 사업인정고시로 보며, 재결신청(裁決申請)의 기한은 <a title="팝업으로 이동" class="link sfon1"
onclick="javascript:fncLsLawPop('1002172687', 'ALLJO');" href="javascript:;">같은 법</a> <a title="팝업으로 이동" class="link sfon2" onclick="javascript:fncLsLawPop('1002172688', 'JO');" href="javascript:;">제23조</a> <a title="팝
업으로 이동" class="link sfon3" onclick="javascript:fncLsLawPop('1002172689', 'JO');" href="javascript:;">제1항</a> 및 <a title="팝업으로 이동" class="link sfon2" onclick="javascript:fncLsLawPop('1002172681', 'JO');" href="
javascript:;">제28조</a> <a title="팝업으로 이동" class="link sfon3" onclick="javascript:fncLsLawPop('1002172694', 'JO');" href="javascript:;">제1항</a>에도 불구하고 <a title="팝업으로 이동" class="link sfon2"
onclick="javascript:fncLsLawPop('1002172692', 'JO');" href="javascript:;">제7조</a> <a title="팝업으로 이동" class="link sfon3" onclick="javascript:fncLsLawPop('1002172693', 'JO');" href="javascript:;">제1항</a>에 따라 승인을 받은
사업계획에서 정한 도시철도사업기간의 종료일로 한다.</p>
673 <p class="pty1_de2_1"> ③ 토지등의 수용 또는 사용에 관하여는 이 법에 규정이 있는 경우를 제외하고는 <a title="팝업으로 이동" class="link sfon1" onclick="javascript:fncLsLawPop('1002172682', 'ALLJO');" href="
javascript:;">「공익사업을 위한 토지 등의 취득 및 보상에 관한 법률」</a>을 준용한다.</p>
674 </div>
```

2차 크롤링: <div class="lawcon">과 </div> 사이에서 법조문 내용을 얻음

키워드 검사

- 특례조항이 되기 위하여 꼭 필요한 단어가 있는지 검사
- [부담금관리 기본법] 별표의 경우 찾아가야 할 주소와 검사해야 할 키워드가 같이 서술되어 있음
- ‘에 따른’ 으로 주소와 키워드 분리
 - ‘항’ 이하 부분과 ‘및’ 이하 부분은 find와 string slicing을 이용하여 제거
- 주소의 텍스트를 불러와서 키워드가 있는지 확인: if 키워드 in 텍스트:
 - 키워드가 있으면 ‘이상 없음’ 출력,
키워드가 있으면 ‘이상 있음’ 출력

3. 「고압가스 안전관리법」 제34조의2에 따른 안전관리부담금

4. 「공적자금상환기금법」 제5조제4항 및 제5항에 따른 공적자금상환기금 출연금

제34조의2(안전관리부담금) ① 산업통상자원부장관은 가스의 안전관리와 유통구조의 개선을 위하여 다음 각 호의 어느 하나에 해당하는 자로부터 안전관리부담금(이하 "부담금"이라 한다)을 부과·징수할 수 있다.

제5조(기금에의 출연) ④ 우체국예금특별회계는 매년 「우체국예금·보험에 관한 법률」에 따른 우체국예금의 전년도 평균 잔액에 1천분의 3을 초과하지 아니하는 범위에서 대통령령으로 정하는 비율을 곱한 금액(그 금액이 10만원보다 적은 경우에는 10만원)을 기금에 출연하여야 한다.

⑤ 우체국보험특별회계는 매년 전년도에 적립한 책임준비금 등을 고려하여 대통령령으로 정하는 금액에 1천분의 3을 초과하지 아니하는 범위에서 대통령령으로 정하는 비율을 곱한 금액(그 금액이 10만원보다 적은 경우에는 10만원)을 기금에 출연하여야 한다.

이상이 있는 경우를 보고

- 3번의 경우, 키워드가 있으니 이상이 없다고 나옴
- 4번의 경우 키워드가 없으니, 이상이 있다고 나옴
- 그러나, 4번은 사실 오류가 아니며, 정상케이스임
: **False-positive**
- 4번은 정보 추출 방법이 분류에 실패한 케이스

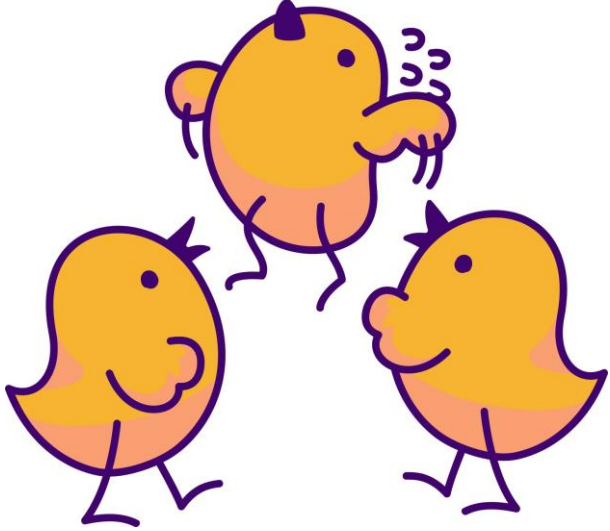
3. 「고압가스 안전관리법」 제34조의2에 따른 안전관리부담금
4. 「공적자금상환기금법」 제5조제4항 및 제5항에 따른 공적자금상환기금 출연금

제34조의2(안전관리부담금) ① 산업통상자원부장관은 가스의 안전관리와 유통구조의 개선을 위하여 다음 각 호의 어느 하나에 해당하는 자로부터 **안전관리부담금**(이하 "부담금"이라 한다)을 부과·징수할 수 있다.

제5조(기금에의 출연) ④ 우체국예금특별회계는 매년 「우체국예금·보험에 관한 법률」에 따른 우체국예금의 전년도 평균 잔액에 1천분의 3을 초과하지 아니하는 범위에서 대통령령으로 정하는 비율을 곱한 금액(그 금액이 10만원보다 적은 경우에는 10만원)을 기금에 출연하여야 한다.
⑤ 우체국보험특별회계는 매년 전년도에 적립한 책임준비금 등을 고려하여 대통령령으로 정하는 금액에 1천분의 3을 초과하지 아니하는 범위에서 대통령령으로 정하는 비율을 곱한 금액(그 금액이 10만원보다 적은 경우에는 10만원)을 기금에 출연하여야 한다.

분류가 가끔 실패하는데, 무슨 쓸모인가?

- 다행히 이 분류문제는 false-positive의 확률은 있으나, false-negative의 확률은 거의 없다.
- '안전관리부담금'이라는 정확한 키워드가 법률에서 우연히 매칭될 확률은 없기 때문(도메인 지식으로 봤을 때)
- 즉, '이상 없음'으로 나온 케이스는 검토할 필요가 없고, '이상 있음'으로 나온 케이스만 수작업으로 검토하면 됨
→ 300번의 수작업을 10번으로 줄여주는 쓸모가 있음
- 또한, 더 규칙적인 고정 키워드를 사용하는 국유재산특례 등의 경우에는 분류 실패가 없음



5. 결론

- 법률 데이터의 특성상 현대적 NLP가 아니어도 할 수 있는 일이 있다.
- 만약, 법률 데이터를 좀 더 정형화시킨다면, 정보추출이 좀 더 쓸모가 있어질 것임

가능성: 법률별, 조문별로 태그 달기

- 조문 특성을 메타정보로 저장: 조문의 종류, 조문의 규정 방식(기속 vs 재량) ...
- 실무적으로도 필요: 입법할 때 비슷한 유형의 조문들을 참고할 필요
- 머신러닝 NLP에도 일종의 label로서 도움이 될 것
- 일종의 annotated corpora?

싸움터가 법률 도메인이었다면,
화웅이 관우를 이겼을지도 모른다.

<http://blog.naver.com/qkdwoa5/100054457590>



법률은 많고,
분석할 것도 많다.
도구를 가리지 말고
쓸모 있으면 뭐든
써보자.



- 김재윤. 2019. "형벌 규정의 전체 현황, 법정형의 수준 및 균형성 분석 - 30년간의 전체 형벌 규정에 대한 텍스트 분석을 이용하여 -." 『법경제학연구』 제16권 제2호, 한국법경제학회.
- _____. 2019. "법률을 디버깅하다(Debugging law)." PyCon Korea 2019 발표문.
- 김재윤·최선영. 2019. "특례제한규정의 효력 · 법체계성 및 입법적 개선방안에 대한 연구: 자동화된 텍스트 분석 방법 적용." 입법과 정책 제11권 제2호, 국회입법조사처.
- Kim, Jaeyoon. 2019. "Evolution of Korean legal code after democratization in 1987." Proceedings of Conference on Data Science and Law, ETH Zurich.
- Tiersma, Peter. The Creation, Structure, and Interpretation of the Legal Text, <http://www.languageandlaw.org/LEGALTEXT.HTM>
- Source code: <https://sites.google.com/view/jaeyoon-kim>

감사합니다.

문의: suitbread@gmail.com