Evaluating Document Representations for Content-based Legal Literature Recommendations

Malte Ostendorff[1,2], Elliott Ash[3], Terry Ruas[4], Bela Gipp[4], Julian Moreno-Schneider[2], Georg Rehm[2]

[1]Open Legal Data, Germany (mo@openlegaldata.io)
[2]German Research Center for Artificial Intelligence, Germany (firstname.lastname@dfki.de)
[3]ETH Zurich, Switzerland (ashe@ethz.ch)
[4]University of Wuppertal, Germany (lastname@uni-wuppertal.de)

ABSTRACT

Recommender systems assist legal professionals in finding relevant literature for supporting their case. Despite its importance for the profession, legal applications do not reflect the latest advances in recommender systems and representation learning research. Simultaneously, legal recommender systems are typically evaluated in small-scale user study without any public available benchmark datasets. Thus, these studies have limited reproducibility. To address the gap between research and practice, we explore a set of state-of-the-art document representation methods for the task of retrieving semantically related US case law. We evaluate text-based (e.g., fast-Text, Transformers), citation-based (e.g., DeepWalk, Poincaré), and hybrid methods. We compare in total 27 methods using two silver standards with annotations for 2,964 documents. The silver standards are newly created from Open Case Book and Wikisource and can be reused under an open license facilitating reproducibility. Our experiments show that document representations from averaged fast-Text word vectors (trained on legal corpora) yield the best results, closely followed by Poincaré citation embeddings. Combining fast-Text and Poincaré in a hybrid manner further improves the overall result. Besides the overall performance, we analyze the methods depending on document length, citation count, and the coverage of their recommendations. We make our source code, models, and datasets publicly available.

CCS CONCEPTS

• **Information systems** → *Recommender systems*; *Similarity measures*; *Clustering and classification*; • **Applied computing** → *Law*.

KEYWORDS

Legal literature, document embeddings, document similarity, recommender systems, Transformers, WikiSource, Open Case Book

1 INTRODUCTION

Legal professionals, e.g., lawyers and judges, frequently invest considerable time to find relevant literature [24]. More so than most other domains, in law there are high stakes for finding the most relevant information (documents) as that can drastically affect the outcome of a dispute. A case can be won or lost depending on whether or not a supporting decision can be found. Recommender systems assist in the search for relevant information. However, research and development of recommender systems for legal corpora poses several challenges. Recommender system research is known to be domain-specific, i.e., minor changes may lead to unpredictable variations in the recommendation effectiveness [4]. Likewise, legal English

is a peculiarly obscure and convoluted variety of English with a widespread use of common words with uncommon meanings [31]. Recent language models like BERT [15] may not be equipped to handle legal English since they are pretrained on generic corpora like Wikipedia or cannot process lengthy legal documents due to their limited input length. This raises the question of whether the recent advances in recommender system research and underlying techniques are also applicable to law.

In this paper, we empirically evaluate 27 document representation methods and analyze the results with respect to the aforementioned possible issues. In particular, we evaluate for each method the quality of the document representations in a literature recommender use case. The methods are distinguished in three categories: (1) word vector-based, (2) Transformer-based, and (3) citation-based methods. Moreover, we test additional hybrid variations of the aforementioned methods. Our primary evaluation metric comes from two silver standards on US case law that we extract from Open Case Book and Wikisource. The relevance annotations from the silver standards are provided for 2,964 documents.

In summary, our contributions are: (1) We propose and make available two silver standards as benchmarks for legal recommender system research that currently do not exist. (2) We evaluate 27 methods of which the majority have never been investigated in the legal context with a quantitative study and validate our results qualitatively. (3) We show that the hybrid combination of text-based and citation-based methods can further improve the experimental results.

2 RELATED WORK

Recommender systems are a well-established research field [3] but relatively few publications focus on law as the application domain. Winkels et al. [55] are among the first to present a content-based approach to recommend legislation and case law. Their system uses the citation graph of Dutch Immigration Law and is evaluated with a user study conducted with three participants. Boer and Winkels [9] propose and evaluate Latent Dirichlet Allocation (LDA) [7] as a solution to the cold start problem in collaborative filtering recommender system. In an experiment with 28 users, they find the user-based approach outperforms LDA. Wiggers and Verberne [52] study citations for legal information retrieval and suggest citations should be combined with other techniques to improve the performance.

Kumar et al. [22] compare four different methods to measure the similarity of Indian Supreme Court decision: TF-IDF [46] on all document terms, TF-IDF on only specific terms from a legal dictionary, Co-Citation, and Bibliographic Coupling. They evaluate the similarity measure on 50 document pairs with five legal domain experts. In their experiment, Bibliographic Coupling and TF-IDF on

legal terms yield the best results. Mandal et al. [29] extend this work by evaluating LDA and document embeddings (Paragraph Vectors [26]) on the same dataset, whereby Paragraph Vectors was found to correlate the most with the expert annotations. Indian Supreme Court decisions are also used as evaluation by Wagh and Anand [50], where they use document similarity based on concepts instead of full-text. They extract concepts (groups of words) from the decisions and compute the similarity between documents based on these concepts. Their vector representation, an average of word embeddings and TF-IDF, shows IDF for weighting word2vec embeddings improve results. Also, Bhattacharya et al. [6] compare citation similarity methods, i.e., Bibliographic Coupling, Co-citation, Dispersion [33] and Node2Vec [17]), and text similarity methods like Paragraph Vectors. They evaluate the algorithms and their combinations using a gold standard of 47 document pairs. A combination of Bibliographic Coupling and Paragraph Vectors achieves the best results.

With Eunomos, Boella et al. [8] present a legal document and knowledge management system that allows searching legal documents. The document similarity problem is handled using TF-IDF and cosine similarity. Other experiments using embeddings for document similarity include Landthaler et al. [23], Nanda et al. [34], and Ash and Chen [2].

Even though different methods have been evaluated in the legal domain, most results are not coherent and rely on small-scale user studies. This finding emphasizes the need for a standard benchmark to enable reproducibility and comparability [4]. Moreover, the recent Transformer models [49] or novel citation embeddings have not been evaluated in legal recommendation research.

3 METHODOLOGY

In this section, we describe our quantitative evaluation of 27 methods for legal document recommendations. We define the recommendation scenario as follows: The user, a legal professional, needs to research a particular decision, e.g., to prepare a litigation strategy. Based on the decision at hand, the system recommends other decisions to its users such that the research task is easy to accomplish. The recommendation is relevant when it covers the same topic or provides essential background information, e.g., it overruled the seed decision [48].

3.1 Case Corpus and Silver Standard

Most of the previous works (Section 2) evaluate recommendation relevance by asking domain experts to provide subjective annotations [9, 22, 29, 55]. Especially in the legal domain, these expert annotations are costly to collect and, therefore, their quantity is limited. For the same reason, expert annotations are rarely published. Consequently, the research is difficult to reproduce [4]. In the case of the US court decisions, such expert annotations between documents are also not publicly available. We construct two ground truth datasets from publicly available resources allowing the evaluation of more recommendations to mitigate the mentioned problems of cost, quantity, and reproducibility.

3.1.1 Open Case Book. With Open Case Book, the Harvard Law School Library offers a platform for making and sharing open-licensed casebooks [1]. The corpus consists of 222 casebooks containing 3,023 cases from 87 authors. Each casebook contains a manually curated set of topically related court decisions, which we use as relevance annotations. The casebooks cover a range from broad topics (e.g., *Constitutional law*) to specific ones (e.g., *Intermediary Liability and Platforms' Regulation*). The decisions are mapped to full-texts and citations retrieved from the Caselaw Access Project (CAP)[2]. After duplicate removal and the mapping procedure, relevance annotations for 1,601 decisions remain.

3.1.2 Wikisource. We use a collection of 2,939 US Supreme Court decisions from Wikisource as ground truth [53]. The collection is categorized in 67 topics like *antitrust*, *civil rights*, and *amendments*. We map the decisions listed in Wikisource to the corpus from CourtListener[3]. The discrepancy between the two corpora decreases the number of relevance annotations to 1,363 court decisions.

Table 1: Distribution of relevant annotations for Open Case Book and Wikisource.

	Relevant annotations per document						
	Mean	Std.	Min.	25%	50%	75%	Max.
Open Case Book	86.42	65.18	2.0	48.0	83.0	111.0	1590.0
Wikisource	130.01	82.46	1.0	88.0	113.0	194.0	616.0

We derive a binary relevance classification from Open Case Book and Wikisource. When decisions A and B are in the same casebook or category, A is relevant for B and vice versa. Table 1 presents the distribution of relevance annotations. This relevance classification is limited since a recommendation might still be relevant despite not being assigned to the same topic as the seed decision. Thus, we consider the Open Case Book and Wikisource annotations as a silver standard rather than a gold one.

3.2 Evaluated Methods

We evaluate 27 methods, each representing legal document d as a numerical vector $\vec{d} \in \mathbb{R}^s$, with s denoting the vector size. To retrieve the recommendations, we first obtain the vector representations (or document embeddings). Next, we compute the cosine similarities of the vectors. Finally, we select the top $k = 5$ documents with the highest similarity through nearest neighbor search[4]. Mean Average Precision (MAP) is the primary and Mean Reciprocal Rank (MRR) is the second evaluation metric [30]. We compute MAP and MRR over a set of queries Q, whereby Q is equivalent to the seed decisions with $|Q_{WS}| = 1363$ available in Wikisource and $|Q_{OCB}| = 1601$ for Open Case Book. In addition to the accuracy-oriented metrics, we evaluate the coverage and Jaccard index of the recommendations. The coverage for the method a is defined as in Equation 1 where D denotes the set of all available documents in the corpus and D_a denotes the recommended documents by a [16].

---

[1]https://opencasebook.org
[2]https://case.law
[3]https://courtlistener.com
[4]We set $k = 5$ due to the UI [36] into which the recommendations will be integrated.

$$Cov(a) = \frac{|D_a|}{|D|} \tag{1}$$

We define the Jaccard index [19] for the similarity and diversity of two recommendation sets R_a and R_b from methods a and b for the seed d_s in Equation 2:

$$J(a,b) = \frac{|R_a \cap R_b|}{|R_a \cup R_b|} \tag{2}$$

We divide the evaluated methods into three categories: Word vector-, Transformer-, and citation-based methods.

3.2.1 TF-IDF Baseline.
As a baseline method, we use the sparse document vectors from TF-IDF [46], which are commonly used in related works [22, 34][5].

3.2.2 Word vector-based Methods.
The following methods are derived from word vectors, i.e., context-free word representations. **Paragraph Vectors** [26] extend the idea of word2vec [32] to learning embeddings for word sequences of arbitrary length. Paragraph Vectors using distributed bag-of-words (dbow) performed well in text similarity tasks applied on legal documents [2, 29] and other domains [25]. We train Paragraph Vectors' dbow model to generate document vectors for each court decision. Like word2vec, **GloVe** [39] and **fastText** [10, 20] produce dense word vectors but they do not provide document vectors. To embed a court decision as a vector, we compute the weighted average over its word vectors, $\vec{w}_i$, whereby the number of occurrences of the word i in d defines the weight c_i. Averaging of word vectors is computationally effective and yields good results for representing even longer documents [1]. For our experiments, we use word vectors made available by the corresponding authors and custom word vectors. While GloVe vectors are pretrained on Wikipedia and Gigaword [39], fastText is pretrained on Wikipedia, UMBC webbase corpus and statmt.org news dataset [10]. Additionally, we use custom word vectors[6] for both methods (namely **fastText**$_{\text{Legal}}$ and **GloVe**$_{\text{Legal}}$) pretrained on the joint court decision corpus extracted from Open Case Book and Wikisource (see Section 3.1). Using word vectors pretrained on different corpora, allows the evaluation of the method's cross-domain applicability.

3.2.3 Transformer-based Methods.
As the second method category, we employ language models for deep contextual text representations based on the Transformer architecture [49], namely **BERT** [15], **RoBERTa** [28], Sentence Transformers (**Sentence-BERT** and **Sentence-RoBERTa**) [44], **LongFormer** [5] and variations of them. In contrast to Paragraph Vectors and average word vectors, which neglect the word order, the Transformers incorporate word positions making the text representations context-dependent. BERT significantly improved the state-of-the-art for many NLP tasks. In general, BERT models are pretrained on large text corpora in an unsupervised fashion to then be fine-tuned for specific tasks like document classification [37]. We use four variations of BERT. The original BERT [15] as base and large version (pretrained on Wikipedia and BookCorpus) and two BERT-base models pretrained on legal corpora. **Legal-JHU-BERT-base** from Holzenberger et al.

[18] which is a BERT base model but fine-tuned on the CAP corpus. Similarly, **Legal-AUEB-BERT-base** from Chalkidis et al. [14] is as well fine-tuned on the CAP corpus but also on other corpora (court cases and legislation from the US and EU, and US contracts). RoBERTa improves BERT with longer training, larger batches, and removal of the next sentence prediction task for pretraining. Sentence Transformers are fine-tuned BERT and RoBERTa models in a Siamese setting [12] to derive semantically meaningful sentence embeddings that can be compared using cosine similarity (Sentence-BERT and Sentence-RoBERTa). The provided Sentence Transformers variations are *nli-* or *stsb*-version that are either fine-tuned on the SNLI and MNLI dataset [11, 54] or fine-tuned on the STS benchmark [13]. As the self-attention mechanism scales quadratically with the sequence length, the Transfomer-based methods (BERT, RoBERTa and Sentence Transformers) bound their representation to 512 tokens. Longformer includes an attention mechanism that scales linearly with sequence length, which allows to process longer documents. We use pretrained Longformer models as provided by Beltagy et al. [5] and limited to 4096 tokens. All Transformer models apply mean-pooling to derive document vectors. We experimented with other pooling strategies but they yield significantly lower results. These findings agree with Reimers and Gurevych [44]. We investigate each Transformer in two variations depending on their availability and w.r.t. model size and document vector size (base with $s = 768$ and large with $s = 1024$).

3.2.4 Citation-based Methods.
We explore citation-based graph methods in which documents are nodes and edges correspond to citations to generate document vectors. Like text-based representations, citation graph embeddings have the vector size $\vec{d} \in \mathbb{R}^{300}$. With **Deep-Walk**, Perozzi et al. [40] were the first to borrow word2vec's idea and applied it to graph network embeddings. DeepWalk performs truncated random walks on a graph and the node embeddings are learned through the node context information encoded in these short random walks similar to the context sliding window in word2vec. **Walklets** [41] explicitly encodes multi-scale node relationships to capture community structures with the graph embedding. Walklets generates these multi-scale relationships by sub-sampling short random walks on the graph nodes. **BoostNE** [27] is a matrix factorization-based embedding technique combined with gradient boosting. In [27], BoostNE is applied on a citation graph from scientific papers and outperforms other graph embeddings such as DeepWalk. Hence, we expect comparable results for the legal citation graph. Nickel and Kiela [35] introduced **Poincaré** embeddings as a method to learn embedding in the hyperbolic space of the Poincaré ball model rather than the Euclidean space used in the aforementioned methods. Embeddings produced in hyperbolic space are naturally equipped to model hierarchical structures [21]. Such structures can also be found in the legal citation graph in the form of different topics or jurisdictions. For DeepWalk, Walklets, BoostNe, we use the Karate Club implementation [45].

3.2.5 Variations & Hybrid Methods.
Given the conceptional differences in the evaluated methods, each method has its strength and weakness. For further insights on these differences, we evaluate all methods with **limited text**, **vector concatenation**, and **score summation**: Unlike the Transformers, the word vector-based methods

---

[5]We use the TF-IDF implementation from the scikit-learn framework [38].
[6]The legal word vectors can be downloaded from our GitHub repository.

have no maximum of input tokens. Whether an artificial limitation of the document length improves or decreases the results is unclear. Longer documents might add additional noise to the representation and could lead to worse results [47]. To make these two method categories comparable, we include additional variations of the word vector-based methods that are limited to the first 512 or 4096 tokens of the document. For instance, the method *fastText$_{Legal}$ (512)* has only access to the first 512 tokens.

Additionally, we explore hybrid methods that utilize text and citation information. Each of the single methods above yields a vector representation $\vec{d}$ for a given document d. We combine methods by concatenating their vectors. For example, the vectors from fastText $\vec{d}_{\text{fastText}}$ and Poincaré $\vec{d}_{\text{Poincaré}}$ can be concatenated as in Equation 3:

$$\vec{d} = \vec{d}_{\text{fastText}} || \vec{d}_{\text{Poincaré}} \qquad (3)$$

The resulting vector size is the sum of the concatenated vector sizes, e.g., $s = 300 + 300 = 600$. Recommendations based on the concatenated methods are retrieved in the same fashion as the other methods, with cosine similarity. Moreover, we combine methods by adding up their cosine similarities [51]. The combined score of two methods is the sum of the individual scores, e.g., for method X and method Y the similarity of two documents d_a and d_b is computed as in Equation 4. Methods with score summation are denoted with $X + Y$, e.g., Poincaré + fastText$_{Legal}$.

$$sim(\vec{d}_a, \vec{d}_b) = sim(\vec{d}_{X_a}, \vec{d}_{X_b}) + sim(\vec{d}_{Y_a}, \vec{d}_{Y_b}) \qquad (4)$$

Lastly, we integrate citation information into Sentence Transformers analog to the fine-tuning procedure proposed by Reimers and Gurevych [44]. Based on the citation graph, we construct a dataset of positive and negative document pairs. Two documents d_a, d_b are considered as positive samples when they are connected through a citation. Negative pairs are randomly sampled and do not share any citation. **Sentence-Legal-AUEB-BERT-base** is the Sentence Tranformer model with Legal-AUEB-BERT-base as base model and trained with these citation information.

4 RESULTS

For our evaluation, we obtain a list of recommendations for each input document and method and then compute the performance measures accordingly. We compute the average number of relevant recommendations, precision, recall, MRR, MAP, and coverage.

4.1 Quantitative Evaluation

4.1.1 Overall Results. Table 2 presents the overall evaluation metrics for 27 methods and the two datasets. From the non-hybrid methods, fastText$_{Legal}$ yields with 0.05 the highest MAP score on Open Case Book, whereas on Wikisource, fastText$_{Legal}$, Poincaré, and Walklets all achieve the highest MAP score of 0.031. The hybrid method of Poincaré || fastText$_{Legal}$ outperforms the non-hybrids for Wikisource with 0.035 MAP. For Open Case Book, the MAP of Poincaré + fastText$_{Legal}$ and fastText$_{Legal}$ are equally high.

Due to space constraints, we remove 14 methods from Table 2 (excluded methods are in the supplementary materials[9]). From the word vector-based methods, we discard the 512 and 4096 tokens variations of Paragraph Vectors, GloVe and GloVe$_{Legal}$, as they show a similar performance deterioration as fastText$_{Legal}$. The base versions

of some Transformers are also excluded in favour of the better performing large versions. Similarly, the *nli* version always outperform the *stsb* version of Sentence Transformers (sBERT and sRoBERTa). For the hybrid variations, we show only the best methods. We also tested Node2Vec [17] and but exclude it given its low MAP scores.

Regarding the word vector-based methods, we see that the methods which are trained on the legal corpus (Paragraph Vectors, fastText$_{Legal}$, GloVe$_{Legal}$) perform similarly well with a minor advantage by fastText$_{Legal}$. Moreover, there is a margin between the generic and legal word vectors even though the legal word vectors are trained on a small corpus compared to ones from the generic vectors. The advantage of Paragraph Vectors over TF-IDF is consistent with the results from Mandal et al. [29]. Limiting the document length to 512 or 4096 decreases the effectiveness of fastText$_{Legal}$. A limit of 512 tokens decreases the MAP score to 59% compared to all tokens on Open Case Book. With 4096 tokens, the performance decline is only minor (90% compared to all tokens). The token limitation effect is also larger on Open Case Book than Wikisource. The 4096 tokens version of fastText$_{Legal}$ even outperforms all Transformer methods.

Longformer-large is the best Transformer for Open Case Book with 0.031 MAP. For Wikisource, Legal-AUEB-BERT achieves the highest MAP of 0.022, closely followed by Legal-JHU-BERT. The Longformer's theoretical advantage of processing 4096 instead of 512 tokens does not lead to better results for Wikisource, for which even BERT scores the same MAP of 0.018. We generally observe that large models outperform their base counterparts[7]. Likewise, RoBERTa has higher scores than BERT as Liu et al. [28] suggested. From the Transformers category, Sentence Transformers yield the worst results. We assume that fine-tuning on the similarity datasets like NLI or STSB does not increase the performance since the models do not generalize well to other domains. However, the language model fine-tuning from Legal-JHU-BERT and Legal-AUEB-BERT does improve the performance, whereby Legal-AUEB-BERT generally outperforms Legal-JHU-BERT. For Open Case Book, Legal-AUEB-BERT is the best model in the Transformer category in terms of MAP even though it is only used as base version.

Poincaré and Walklets are by far the best methods in the citation category. For Wikisource, the two citation-based methods, score the same MAP of 0.031 as fastText$_{Legal}$. Compared to the word vector-based methods, the citation methods do better on Wikisource than on Open Case Book.

In the category of hybrid methods, the combination of text and citations improves the performance. For Open Case Book, the score summation Poincaré + fastText$_{Legal}$ has the same MAP of 0.05 as fastText$_{Legal}$ but a higher MRR of 0.746. The MRR of Poincaré + fastText$_{Legal}$ is even higher than the MRR of its sub-methods Poincaré (0.629) and fastText$_{Legal}$ (0.739) individually. The concatenation of Poincaré || fastText$_{Legal}$ is with 0.035 MAP the best method on Wikisource. Using citation as training signal as in Sentence-Legal-AUEB-BERT also improves the performance but not as much as concatenation or summation. When comparing the three hybrid variations, score summation achieves overall the best results. In the case of Wikisource, the concatenation's scores are below its sub-methods, while summation has at least the best sub-methods score. Moreover,

---

[7]Legal-JHU-BERT and Legal-AUEB-BERT are only available as base version.

Table 2: Overall scores for top $k = 5$ recommendations from Open Case Book and Wikisource as the number of relevant documents, precision, recall, MRR, MAP and coverage for the 27 methods and the vector sizes. The methods are divided into: baseline, word vector-based, Transformer-based, citation-based, and hybrid. High scores according to the exact numbers are <u>underlined</u> (or bold for category-wise). * values were rounded up.

Datasets →		Open Case Book						Wikisource					
Methods ↓	Size	Rel.	Prec.	Recall	MRR	MAP	Cov.	Rel.	Prec.	Recall	MRR	MAP	Cov.
TF-IDF	500000	1.60	0.320	0.032	0.363	0.020	0.487	1.59	0.318	0.026	0.389	0.015	0.446
Paragraph Vectors	300	2.78	0.555	0.056	0.729	0.049	**0.892**	2.39	0.477	0.036	0.629	0.030	**0.841**
fastText	300	2.66	0.532	0.053	0.713	0.045	0.811	2.11	0.422	0.031	0.581	0.025	0.772
fastText$_{Legal}$	300	<u>**2.87**</u>	<u>**0.574**</u>	<u>**0.059**</u>	**0.739**	<u>**0.050**</u>	0.851	**2.39**	**0.478**	**0.037**	**0.631**	**0.031**	0.815
fastText$_{Legal}$ (512)	300	1.97	0.394	0.037	0.591	0.028	0.835	2.16	0.433	0.034	0.587	0.027	0.809
fastText$_{Legal}$ (4096)	300	2.76	0.552	0.054	0.727	0.045	0.867	2.33	0.466	0.035	0.620	0.029	0.817
GloVe	300	2.68	0.536	0.054	0.702	0.046	0.814	2.06	0.412	0.033	0.577	0.026	0.789
GloVe$_{Legal}$	300	2.82	0.564	0.057	0.724	0.048	0.834	2.31	0.461	0.037	0.621	0.030	0.804
BERT-base	768	1.26	0.253	0.021	0.428	0.015	0.815	1.62	0.323	0.021	0.485	0.015	0.784
BERT-large	1024	1.35	0.270	0.022	0.443	0.016	0.841	1.82	0.364	0.023	0.530	0.018	0.794
Legal-JHU-BERT-base	768	1.47	0.295	0.025	0.482	0.018	0.848	1.85	0.371	0.027	0.537	0.020	0.796
Legal-AUEB-BERT-base	768	1.66	0.331	0.028	0.506	0.021	0.884	**2.01**	**0.401**	**0.027**	**0.573**	**0.022**	0.813
Longformer-base	768	1.91	0.382	0.033	0.572	0.026	**0.892**	1.65	0.329	0.020	0.514	0.016	**0.841**
Longformer-large	1024	**2.09**	**0.419**	**0.039**	**0.614**	**0.031**	0.885	1.80	0.360	0.023	0.535	0.018	0.826
RoBERTa-large	1024	1.52	0.305	0.026	0.481	0.019	0.843	1.93	0.387	0.026	0.553	0.020	0.782
Sentence-BERT-large-nli	1024	1.03	0.206	0.018	0.352	0.013	0.872	1.37	0.273	0.017	0.443	0.012	0.782
Sentence-BERT-large-nli-stsb	1024	0.98	0.196	0.018	0.338	0.013	0.848	1.36	0.272	0.015	0.434	0.011	0.777
Sentence-RoBERTa-large-nli	1024	0.92	0.183	0.016	0.321	0.011	0.884	1.18	0.236	0.013	0.409	0.009	0.795
BoostNE	300	1.29	0.258	0.022	0.442	0.016	0.800	1.24	0.248	0.016	0.398	0.013	0.832
DeepWalk	300	1.34	0.267	0.028	0.473	0.021	0.818	1.82	0.364	0.030	0.533	0.025	<u>**0.856**</u>
Poincaré	300	**2.24**	**0.447**	**0.044**	0.629	**0.036**	<u>**0.930**</u>	2.33	0.465	0.038	0.598	**0.031**	0.837
Walklets	300	**2.24**	**0.448**	0.043	**0.636**	0.035	0.816	**2.35**	**0.470**	**0.038**	**0.611**	0.031	0.826
Poincaré ‖ fastText$_{Legal}$	600	2.36	0.473	0.048	0.656	0.041	0.737	<u>**2.52**</u>	<u>**0.505**</u>	<u>**0.041**</u>	0.638	<u>**0.035**</u>	0.818
Longformer-large ‖ fastText$_{Legal}$	1324	2.26	0.451	0.043	0.642	0.035	0.876	1.91	0.383	0.025	0.547	0.020	0.829
Poincaré + fastText$_{Legal}$	300 / 300	**2.85**	**0.571**	**0.058**	<u>**0.746**</u>	**0.050**	0.860	2.48	0.497	0.040	<u>**0.646**</u>	0.034	0.835
Poincaré + Longformer-large	300 / 1024	2.09	0.419	0.039	0.630	0.033	0.885	1.80	0.360	0.023	0.548	0.019	0.826
Sentence-Legal-AUEB-BERT-base	768	2.19	0.438	0.039	0.603	0.031	**0.917**	2.36	0.471	0.038	0.602	0.032	**0.849**

combining two text-based methods such as Longformer-large and fastText$_{Legal}$ never improves its sub-methods.

4.1.2 Document Length. The effect of the document length on the performance in terms of MAP is displayed in Figure 1. We group the seed documents into eight equal-sized buckets (each bucket represents the equal number of documents) depending on the word count in the document text to make the two datasets comparable.

Both datasets, Open Case Book and Wikisource, present a similar outcome. The MAP increases as the word count increases. Table 2 presents the average overall documents and, therefore, the overall best method is not equal to the best method in some subsets. For instance, Paragraph Vectors achieve the best results for several buckets, e.g., 4772-6172 words in Open Case Book or 6083-8659 words in Wikisource. The text limitation of fastText$_{Legal}$ (4096 tokens) in comparison to fastText is also clearly visible. The performance difference between the two methods increases as the document length increases. For the first buckets with less than 4096 words, e.g., 187-2327 words in Open Case Book, one could expect no difference since the limitation does not affect the seed documents in these buckets. However, we observe a difference since target documents are not grouped into the same buckets. Remarkable is that the performance difference for very long documents is less substantial. When comparing Longformer-large and Legal-AUEB-BERT, we also see an opposing performance shift with changing word count. While Legal-AUEB-BERT's scores are relatively stable throughout all buckets, Longformer depends more on the document length. On
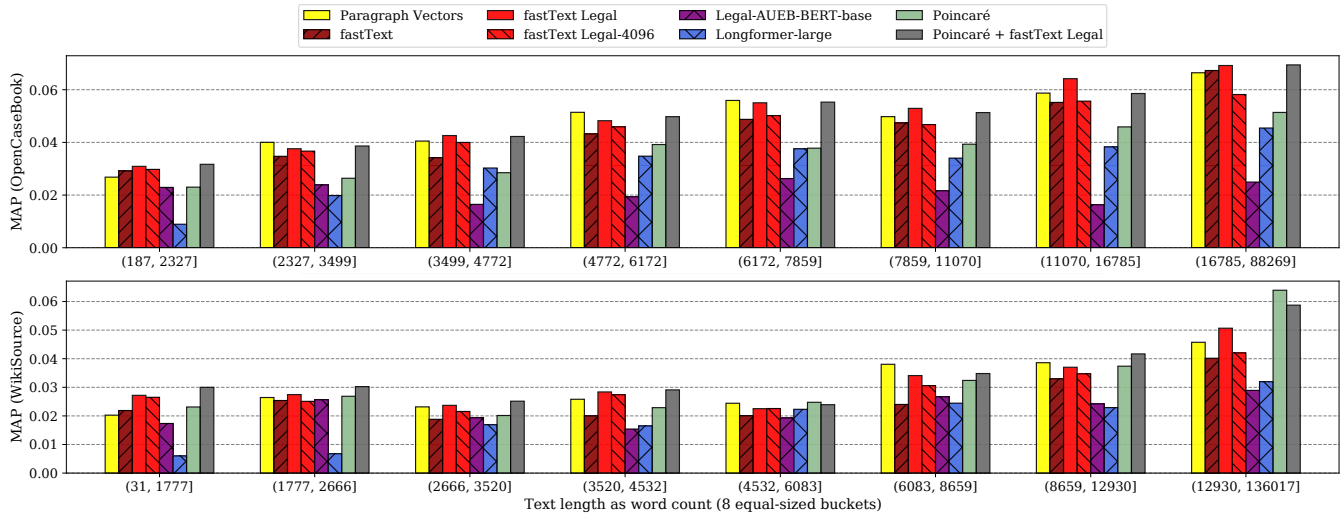
Figure 1: MAP wrt. words in the seed document of Open Case Book (top) and Wikisource (bottom). The more words, the better the results, no peak at medium length. fastText$_{Legal}$ outperforms Legal-BERT and Longformer for short documents.

the one hand, Longformer performs worse than Legal-AUEB-BERT for short documents, i.e., 187-2327 words in Open Case Book, and 31-1777 words in Wikisource. On the other hand, for documents with more words, Longformer mostly outperforms Legal-AUEB-BERT by a large margin. The citation-based method Poincaré is as well affected by the document length. However, this effect is due to a positive correlation between word count and citation count.

4.1.3 Citation Count. Figure 2 shows the effect of the number of in- and out-citations (i.e., edges in the citation graph) on the MAP score. The citation analysis for Wikisource confirms the word count analysis. More data leads to better results. Instead, for Open Case Book, the performance of the citation-based methods peak for 31-51 citations and even decrease at 67-89 citations. When comparing Poincaré and Walklets there is no superior method and no dependency pattern is visible. The performance effect on DeepWalk is more substantial. The number of citations must be above a certain threshold to allow DeepWalk to achieve competitive results. For Open Case Book, the threshold is at 51-67 citations, and for Wikisource, it is at 30-50 citations. Figure 2 also shows the on average higher MAP of Poincaré + fastText$_{Legal}$ in comparison to the other approaches. Citation-based methods require citations to work, whereas text methods do not have this limitation (see 0-14 citations for Open Case Book). When no citations are available, citation-based methods cannot recommend any documents, whereas the text methods still work (see 0-14 citations for Open Case Book).

Our citation-based methods use only a fraction of original citation data, 70,865 citations in Open Case Book, and 331,498 citations in Wikisource, because of limitation to the documents available in the silver standards. For comparison, the most-cited decision from CourtListener (the underlying corpus of Wikisource) has 88,940 citations, whereas in experimental data of Wikisource the maximum number of in- and out-citations is 386. As a result, we expect the citation-based methods, especially DeepWalk, to work even better when applied on the full corpus.

4.1.4 Coverage and Similarity of Recommendations. In addition to the accuracy-oriented metrics, Table 2 reports also the coverage of the recommendation methods. A recommender systems for an expert audience should not focus on small set of most-popular items but rather provide a high coverage of the whole item collection. However, coverage alone does not account for relevancy and, therefore, it must be contextualized with other metrics, e.g., MAP.

Overall, two citation-based methods yield the highest coverage for both datasets, i.e., Poincaré for Open Case Book and DeepWalk for Wikisource. In particular, Poincaré has not only a high coverage but also high MAP scores. Yet, the numbers do not indicate that citation-based methods have generally a higher coverage since the text-based Paragraph Vectors or Longformer-base also achieve a considerably high coverage. The lowest coverage has by far the TF-IDF baseline. Notable, the hybrid methods with concatenation and summation have a different effect on the coverage as on the accuracy metrics. While the hybrid methods generally yield a higher MAP, their coverage is lower compared to their sub-methods. Only, the Sentence-Legal-AUEB-BERT-base yields a higher coverage compared to Legal-AUEB-BERT-base.

Besides the coverage, we also analyze the similarity or diversity of the recommendations between two methods. Figure 3 shows the similarity measured as Jaccard index for selected methods. Method pairs with $J(a, b) = 1$ have identical recommendations, whereas $J(a, b) = 0$ means no common recommendations. Generally speaking, the similarity of all method pairs is considerably low ($J < 0.8$). The highest similarity can be found between a hybrid method and one of its sub-methods, e.g., Poincaré + fastText$_{Legal}$ and fastText$_{Legal}$ with $J = 0.76$. Apart from that, substantial similarity can be only found between pairs from the same category. For example, the pair of the two text-based methods of GloVe$_{Legal}$ and fastText$_{Legal}$ yields $J = 0.67$. Citation-based methods tend to have a lower similarity compared to the text-based methods, whereby the highest Jaccard index between two citation-based methods is achieved for Walklets and
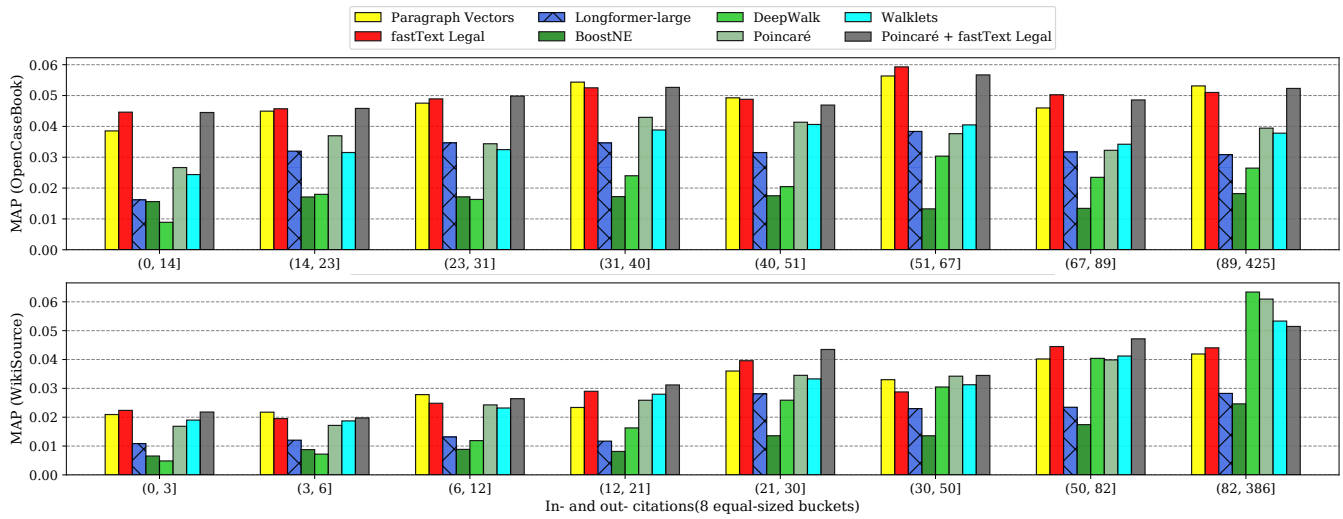
Figure 2: MAP scores wrt. citation count for Open Case Book (top) and Wikisource (bottom). Among citation-based methods, Poincaré and Walklets perform on average the best, while DeepWalk outperforms them only for Wikisource and when more than 82 citations are available (rightmost bucket).

Figure 3: Jaccard index for similarity or diversity of two recommendation sets (average over all seeds from the two datasets).

Poincaré with $J = 0.32$. Like the coverage metric, the Jaccard index should be considered in relation to the accuracy results. GloVe$_{Legal}$ and fastText$_{Legal}$ yield equally high MAP scores, while having also a high recommendation's similarity. In contrast, the MAP for Wikisource from fastText$_{Legal}$ and Poincaré is equally high, too. However, their recommendation's similarity is low $J = 0.11$. Consequently, fastText$_{Legal}$ and Poincaré provide relevant recommendations that are diverse from each other. This explains the good performance of their hybrid combination.

4.2 Qualitative Evaluation

Due to lack of openly available gold standards, we conduct our quantitative analysis using silver standards. Thus, we additionally conduct a qualitative evaluation with domain experts to estimate the quality of our silver standards.

Table 3 lists one of the randomly chosen seed decisions (Mugler vs. Kansas[8]), and five recommended similar decisions, each from fastText$_{Legal}$ and Poincaré. In Mugler vs. Kansas (1887), the court held that Kansas could constitutionally outlaw liquor sales with constitutional issues raised on substantive due process (Fourteenth Amendment) and takings (Fifth Amendment). We provide a detail description of the cases and their relevance annotations in Appendix A.

The sample verification indicates the overall usefulness of both text-based and citation-based methods and does not contradict our quantitative findings. Each of the identified cases have a legal important connection to the seed case (either the Fourteenth Amendment or Fifth Amendment), although it is difficult to say whether the higher-ranked cases are more similar along an important topical dimension. The rankings do not appear to be driven by facts presented in the case as most of them have not to do with alcohol bans. Only Kidd vs. Pearson (1888) is about liquor sales as the seed decision. The samples also do not reveal considerable differences between text- and citation-based similarity. Moreover, we cannot confirm the findings from Schwarzer et al. [47], which suggests that text-based methods are focused on specific terms and citation yield mostly broadly related recommendations. With regards to the silver standards, the domain expert annotations agree in 14 of 20 cases (70%). In only two cases the domain expert classify a recommendation as irrelevant despite being classified as relevant in the silver standard.

---

[8]https://www.courtlistener.com/opinion/92076/mugler-v-kansas/

Table 3: Examples from fastText$_{\text{Legal}}$ and Poincaré (other methods are in the supplementary material) for *Mugler v. Kansas* with relevance annotations by the silver standards (S) and domain expert (D).

	#	Open Case Book Recommendations	Year	S	D	Wikisource Recommendations	Year	S	D
fastText$_{\text{Legal}}$	1	Yick Wo v. Hopkins	1886	N	N	Kidd v. Pearson	1888	N	Y
	2	Munn v. Illinois	1876	Y	Y	Lawton v. Steele	1894	N	Y
	3	LS. Dealers' & Butchers' v. Crescent City LS.	1870	N	Y	Yick Wo v. Hopkins	1886	N	N
	4	Butchers' Benevolent v. Crescent City LS.	1872	Y	Y	Geer v. Connecticut	1896	N	Y
	5	Lochner v. New York	1905	Y	Y	Groves v. Slaughter	1841	Y	N
Poincaré	1	Yick Wo v. Hopkins	1886	N	N	Rast v. Van Deman & Lewis Co.	1916	Y	N
	2	Allgeyer v. Louisiana	1897	Y	Y	County of Mobile v. Kimball	1881	N	N
	3	Calder v. Wife	1798	N	N	Brass v. North Dakota Ex Rel. Stoeser	1894	Y	Y
	4	Davidson v. New Orleans	1877	Y	Y	Erie R. Co. v. Williams	1914	Y	Y
	5	Muller v. Oregon	1908	Y	Y	Hall v. Geiger-Jones Co.	1917	Y	Y

5 DISCUSSION

Our experiments explore the applicability of the latest advances in research to the use case of legal literature recommendations. Existing studies on legal recommendations typically rely on small-scale user studies and are therefore limited in the number of approaches that they can evaluate (Section 2). For this study, we utilize relevance annotations from two publicly available sources, i.e., Open Case Book and Wikisource. These annotations does not only enable us to evaluate the recommendations of 2,964 documents but also the comparison of in total 41 methods and their variations of which 27 methods are presented in this paper.

Our extensive evaluation shows a large variance in the recommendation performance. Such a variance is known from other studies [4]. There is no single method that yields the highest scores across all metrics and all datasets. Despite that, fastText$_{\text{Legal}}$ is on average the best of all 41 methods. fastText$_{\text{Legal}}$ yields the highest MAP for Open Case Book, while for Wikisource only hybrid methods outperform fastText$_{\text{Legal}}$. Also, the coverage of fastText$_{\text{Legal}}$ is considerably high for both datasets. Simultaneously, fastText$_{\text{Legal}}$ is robust to corner cases since neither very short nor very long documents reduce fastText$_{\text{Legal}}$'s performance substantially. These results confirm the findings from Arora et al. [1] that average word vectors are "simple but tough-to-beat baseline". Regarding baselines, our TF-IDF baseline yields one of the worst results. In terms of accuracy metrics, only some Transformers are worse than TF-IDF, but especially TF-IDF's coverage is the lowest by a large margin. With a coverage below 50%, TF-IDF fails to provide diverse recommendations that are desirable for legal literature research.

The transfer of research advances to the legal domain is one aspect of our experiments. Thus, the performance of Transformers and citation embeddings is of particular interest. Despite the success of Transformers for many NLP tasks, Transformers yield on average the worst results for representing lengthy documents written in legal English. The other two method categories, word vector-based, and citation-based methods, surpass Transformers.

The word vector-based methods achieve overall the best results among the non-hybrid methods. All word vector-based methods with in-domain training, i.e., Paragraph Vectors, fastText$_{\text{Legal}}$, and GloVe$_{\text{Legal}}$, perform similarly good with a minor advantage by fastText$_{\text{Legal}}$. Their similar performance aligns with the large overlap among their recommendations. Despite a small corpus of 65,635 documents, the in-domain training generally improves the performance as the gap between the out-of-domain fastText and fastText$_{\text{Legal}}$ shows. Given that the training of custom word vectors is feasible on commodity hardware, in-domain training is advised. More significant than the gap between in- and out-of-domain word vectors is the effect of limited document lengths. For Open Case Book, the fastText$_{\text{Legal}}$ variation limited to the first 512 tokens has only 52% of the MAP of the full-text method. For Wikisource, the performance decline exists as well but is less significant. This effect highlights the advantage of the word vector-based methods that they derive meaningful representations of documents with arbitrary length.

The evaluated Transformers cannot process documents of arbitrary length but are either limited to 512 or 4096 tokens. This limitation contributes to Transformers' low performance. For instance, Longformer-large's MAP is almost twice as high as BERT-large's MAP on Open Case Book. However, for Wikisource both models yield the same MAP scores. For Wikisource, the in-domain pretraining as a larger effect than the token limit since Legal-AUEB-BERT achieves the best results among the Transformers. Regarding the Transformer pretraining, the difference between Legal-JHU-BERT and Legal-AUEB-BERT shows the effect between two pretraining approaches. The corpora and the hyperparameter settings used during pretraining are crucial. Even though Legal-JHU-BERT was exclusively pretrained on the CAP corpus, which has a high overlap with Open Case Book, Legal-AUEB-BERT still outperforms Legal-JHU-BERT on Open Case Book. Given these findings, we expect the performance of Transformers could be improved by increasing the token limit beyond the 4096 tokens and by additional in-domain pretraining. Such improvements are technically possible but add significant computational effort. In contrast to word vectors,

Transformers are not trained on commodity hardware but on GPUs. Especially long-sequence Transformers such as the Longformer require GPUs with large memory. Such hardware may not be available in production deployments. Moreover, the computational effort must be seen in relation to the other methods. Put differently, even fastText$_{Legal}$ limited to 512 tokens outperforms all Transformers.

Concerning the citation embeddings, we consider Poincaré, closely followed by Walklets, as the best method. In particular, the two methods outperform the other citation methods for documents even when only a few citations are available, which makes them attractive for legal research. Poincaré also provides the highest coverage for Open Case Book, emphasizing its quality for literature recommendations. For Wikisource, DeepWalk has the highest coverage despite yielding generally low accuracy scores. As Figure 2 shows, DeepWalk's MAP score improves substantially as the number of citations increases. Therefore, we expect that DeepWalk but also the other citation methods would perform even better when applied on larger citation graph. The analysis of recommendation similarity also shows little overlap between the citation-based methods and the text-based methods (Figure 3). This indicates that the two approaches complement each other and motivates the use of hybrid methods.

Related work has already shown the benefit of hybrid methods for literature recommendations [6, 52]. Our experiments confirm these findings. The simple approaches of score summation or vector concatenation can improve the results. In particular, Poincaré + fastText$_{Legal}$ never leads to a decline in performance. Instead, it increases the performance for corner cases in which one of the sub-methods performs poorly. Vector concatenation has mixed effects on the performance, e.g., positive effect for Wikisource and negative effect for Open Case Book. Using citations as training data in Sentence Transformers can also be considered as a hybrid method that improves the performance. However, this requires additional effort for training a new Sentence Transformer model.

As we discuss in Section 3.1, we consider Open Case Book and Wikisource more of silver than gold standards. With the qualitative evaluation, we mitigate the risk of misinterpreting the quantitative results, whereby we acknowledge our small sample size. The overall agreement with the domain expert is high. The expert tends to classify more recommendations as relevant than the silver standards, i.e., relevant recommendations are missed. This explains the relatively low recall from the quantitative evaluation. In a user study, we would expect only minor changes in the ranking of methods with similar scores, e.g., fastText$_{Legal}$ and GloVe$_{Legal}$. The overall ranking among the method categories would remain the same. The benefit of our silver standards is the number of available relevance annotations. The number of annotations in related user studies is with up to 50 annotations rather low. Instead, our silver standards provide a magnitude more relevance annotations. Almost 3,000 relevance annotations enable evaluations regarding text length, citation count, or other properties that would be otherwise magnitudes more difficult. Similarly, the user studies are difficult to reproduce as their data is mostly unavailable. This leads to reproducibility being an issue in recommender system research [4]. The open license of the silver standards allows the sharing of all evaluation data and, therefore, contributes to more reproducibility. In summary, the proposed datasets bring great value to the field, overcoming eventual shortcomings.

6 CONCLUSION

We present an extensive empirical evaluation of 27 document representation methods in the context of legal literature recommendations. In contrast to previous small-scale studies, we evaluate the methods over two document corpora containing 2,964 documents (1,601 from Open Case Book and 1,363 from Wikisource). We underpin our findings with a sample-based qualitative evaluation. Our analysis of the results reveals fastText$_{Legal}$ (averaged fastText word vectors trained on our corpora) as the overall best performing method. Moreover, we find that all methods have a low overlap between their recommendations and are vulnerable to certain dataset characteristics like text length and number of citations available. To mitigate the weakness of single methods and to increase recommendation diversity, we propose hybrid methods like score summation of fastText$_{Legal}$ and Poincaré that outperforms all other methods on both datasets. Although there are limitations in the experimental evaluation due to the lack of openly available ground truth data, we are able to draw meaningful conclusions for the behavior of text-based and citation-based document embeddings in the context of legal document recommendation. Our source code, trained models, and datasets are openly available to encourage further research[9].

ACKNOWLEDGMENTS

REFERENCES

[1] Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2017. A simple but though Baseline for Sentence Embeddings. In *5th International Conference on Learning Representations (ICLR 2017)*, Vol. 15. 416–424.

[2] Elliott Ash and Daniel L. Chen. 2018. Case Vectors: Spatial Representations of the Law Using Document Embeddings. *SSRN Electronic Journal* 11, 2017 (may 2018), 313–337. https://doi.org/10.2139/ssrn.3204926

[3] Xiaomei Bai, Mengyang Wang, Ivan Lee, Zhuo Yang, Xiangjie Kong, and Feng Xia. 2019. Scientific paper recommendation: A survey. *IEEE Access* 7 (2019), 9324–9339.

[4] Joeran Beel, Corinna Breitinger, Stefan Langer, Andreas Lommatzsch, and Bela Gipp. 2016. Towards reproducibility in recommender-systems research. *User Modeling and User-Adapted Interaction (UMAI)* 26 (2016).

[5] Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. (2020). arXiv:2004.05150

[6] Paheli Bhattacharya, Kripabandhu Ghosh, Arindam Pal, and Saptarshi Ghosh. 2020. Methods for Computing Legal Document Similarity: A Comparative Study. (2020). arXiv:2004.12307

[7] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[8] Guido Boella, Luigi Di Caro, Llio Humphreys, Livio Robaldo, Piercarlo Rossi, and Leendert van der Torre. 2016. Eunomos, a legal document and knowledge management system for the Web to provide relevant, reliable and up-to-date information on the law. *Artificial Intelligence and Law* 24, 3 (2016), 245–283.

[9] Alexander Boer and Radboud Winkels. 2016. Making a cold start in legal recommendation: An experiment. *Frontiers in Artificial Intelligence and Applications* 294 (2016), 131–136. https://doi.org/10.3233/978-1-61499-726-9-131

[10] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.

---

[9]GitHub repository: https://github.com/malteos/legal-document-similarity

[11] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. *Proceedings of EMNLP* (2015), 632–642.

[12] Jane Bromley, J.W. Bentz, Leon Bottou, I. Guyon, Yann Lecun, C. Moore, Eduard Sackinger, and R. Shah. 1993. Signature verification using a Siamese time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence* 7, 4 (1993).

[13] Daniel Cer, Mona Diab, Eneko Agirre, Iñigo Lopez-Gazpio, and Lucia Specia. 2017. SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. In *Proc. of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. ACL, Vancouver, Canada, 1–14.

[14] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The Muppets straight out of Law School. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. ACL, Stroudsburg, PA, USA, 2898–2904.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of the 2019 Conf. of the NAACL*. ACL, Minneapolis, Minnesota, 4171–4186.

[16] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. 2010. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*. ACM Press, New York, New York, USA, 257.

[17] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. In *Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining - KDD '16*. ACM Press, New York, New York, USA, 855–864.

[18] Nils Holzenberger, Andrew Blair-Stanek, and Benjamin Van Durme. 2020. A dataset for statutory reasoning in tax law entailment and question answering. In *Proceedings of the 2020 Natural Legal Language Processing Workshop*. 31–38.

[19] Paul Jaccard. 1912. The Distribution of the Flora in the Alpine Zone. *New Phytologist* 11, 2 (feb 1912), 37–50.

[20] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. ACL, Stroudsburg, PA, USA, 427–431.

[21] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. 2010. Hyperbolic geometry of complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 82, 3 (2010), 1–18.

[22] Sushanta Kumar, P. Krishna Reddy, V. Balakista Reddy, and Aditya Singh. 2011. Similarity analysis of legal judgments. *Compute 2011 - 4th Annual ACM Bangalore Conference* (2011). https://doi.org/10.1145/1980422.1980439

[23] Jörg Landthaler, Bernhard Waltl, Patrick Holl, and Florian Matthes. 2016. Extending full text search for legal document collections using word embeddings. *Frontiers in Artificial Intelligence and Applications* 294 (2016), 73–82.

[24] Steven A. Lastres. 2013. *Rebooting Legal Research in a Digital Age*. Technical Report. LexisNexis. https://www.lexisnexis.com/documents/pdf/20130806061418_large.pdf

[25] J. H. Lau and T. Baldwin. 2016. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In *Proceedings Workshop on Representation Learning for NLP*. https://doi.org/10.18653/v1/w16-1609

[26] Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. *Int. Conf. on Machine Learning* 32 (2014), 1188–1196.

[27] Jundong Li, Liang Wu, Ruocheng Guo, Chenghao Liu, and Huan Liu. 2019. Multi-level network embedding with boosted low-rank matrix approximation. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, New York, NY, USA, 49–56.

[28] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. (2019). arXiv:1907.11692

[29] Arpan Mandal, Raktim Chaki, Sarbajit Saha, Kripabandhu Ghosh, Arindam Pal, and Saptarshi Ghosh. 2017. Measuring Similarity among Legal Court Case Documents. In *Proceedings of the 10th Annual ACM India Compute Conference on ZZZ - Compute '17*. 1–9. https://doi.org/10.1145/3140107.3140119

[30] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. 2008. *Introduction to Information Retrieval*. Vol. 16. Cambridge University Press, Cambridge. 100–103 pages. https://doi.org/10.1017/CBO9780511809071

[31] David Mellinkoff. 1963. The language of the law. *Boston: Little Brown and Company* (1963).

[32] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. (2013), 1–12. arXiv:1301.3781

[33] Akshay Minocha, Navjyoti Singh, and Arjit Srivastava. 2015. Finding Relevant Indian Judgments using Dispersion of Citation Network. In *Proceedings of the 24th International Conference on World Wide Web - WWW '15 Companion*. ACM Press, New York, New York, USA, 1085–1088.

[34] Rohan Nanda, Giovanni Siragusa, Luigi Di Caro, Guido Boella, Lorenzo Grossio, Marco Gerbaudo, and Francesco Costamagna. 2019. Unsupervised and supervised text similarity systems for automated identification of national implementing measures of European directives. *Artificial Intelligence and Law* 27, 2 (2019), 199–225. https://doi.org/10.1007/s10506-018-9236-y

[35] Maximilian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. *Advances in Neural Information Processing Systems* 2017-Decem, Nips (2017), 6339–6348. arXiv:1705.08039

[36] Malte Ostendorff, Till Blume, and Saskia Ostendorff. 2020. Towards an Open Platform for Legal Information. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. ACM, New York, NY, USA, 385–388.

[37] Malte Ostendorff, Peter Bourgonje, Maria Berger, Julian Moreno-Schneider, Georg Rehm, and Bela Gipp. 2019. Enriching BERT with Knowledge Graph Embeddings for Document Classification. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*. GSCL, Erlangen, Germany, 305–312.

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[39] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ACL, Stroudsburg, PA, USA, 1532–1543. https://doi.org/10.3115/v1/D14-1162

[40] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*. ACM Press, New York, New York, USA, 701–710.

[41] Bryan Perozzi, Vivek Kulkarni, Haochen Chen, and Steven Skiena. 2017. Don't Walk, Skip!: Online Learning of Multi-scale Network Embeddings. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM, New York, NY, USA, 258–265.

[42] Georg Rehm, Peter Bourgonje, Stefanie Hegele, Florian Kintzel, Julian Moreno Schneider, Malte Ostendor, Karolina Zaczynska, Armin Berger, Stefan Grill, Soren Rauchle, Jens Rauenbusch, Lisa Rutenburg, Andre Schmidt, Mikka Wild, Henry Homann, Julian Fink, Sarah Schulz, Jurica Seva, Joachim Quantz, Joachim Bottger, Josene Matthey, Rolf Fricke, Jan Thomsen, Adrian Paschke, Jamal Al Qundus, Thomas Hoppe, Naouel Karam, Frauke Weichhardt, Christian Fillies, Clemens Neudecker, Mike Gerber, Kai Labusch, Vahid Rezanezhad, Robin Schaefer, David Zellhofer, Daniel Siewert, Patrick Bunk, Julia Katharina Schlichting, Lydia Pintscher, Elena Aleynikova, and Franziska Heine. 2020. QURATOR: Innovative technologies for content and data curation. In *Proceedings of the Conference on Digital Curation Technologies (Qurator 2020)*. arXiv:2004.12195

[43] Georg Rehm, Julian Moreno-Schneider, Jorge Gracia, Artem Revenko, Victor Mireles, Maria Khvalchik, Ilan Kernerman, Andis Lagzdins, Marcis Pinnis, Artus Vasilevskis, Elena Leitner, Jan Milde, and Pia WeiÃŸenhorn. 2019. Developing and Orchestrating a Portfolio of Natural Legal Language Processing and Document Curation Services. In *Proceedings of Workshop on Natural Legal Language Processing (NLLP 2019)*, Nikolaos Aletras, Elliott Ash, Leslie Barrett, Daniel Chen, Adam Meyers, Daniel Preotiuc-Pietro, David Rosenberg, and Amanda Stent (Eds.). Minneapolis, USA, 55–66. Co-located with NAACL 2019. 7 June 2019.

[44] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *The 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019)*. arXiv:1908.10084

[45] Benedek Rozemberczki, Oliver Kiss, and Rik Sarkar. 2020. An API Oriented Open-source Python Framework for Unsupervised Learning on Graphs. (2020). arXiv:2003.04819

[46] G. Salton, A. Wong, and C. S. Yang. 1975. Vector Space Model for Automatic Indexing. Information Retrieval and Language Processing. *Commun. ACM* 18, 11 (1975), 613–620.

[47] Malte Schwarzer, Moritz Schubotz, Norman Meuschke, and Corinna Breitinger. 2016. Evaluating Link-based Recommendations for Wikipedia. *Proc. of the 16th ACM/IEEE Joint Conference on Digital Libraries (JCDL'16)* (2016), 191–200.

[48] Marc van Opijnen and Cristiana Santos. 2017. On the concept of relevance in legal information retrieval. *Artificial Intelligence and Law* 25, 1 (2017), 65–87.

[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems* 30 (Jun 2017), 5998–6008.

[50] Rupali S. Wagh and Deepa Anand. 2020. Legal document similarity: A multicriteria decision-making perspective. *PeerJ Computer Science* 2020, 3 (2020), 1–20. https://doi.org/10.7717/peerj-cs.262

[51] Lidan Wang, Ming Tan, and Jiawei Han. 2016. FastHybrid: A hybrid model for efficient answer selection. *Proceedings of the 26th International Conference on Computational Linguistics* (2016), 2378–2388.

[52] Gineke Wiggers and Suzan Verberne. 2019. Citation Metrics for Legal Information Retrieval Systems. In *BIR@ECIR*. 39–50.

[53] Wikisource. 2020. United States Supreme Court decisions by topic. https://en.wikisource.org/wiki/Category:United_States_Supreme_Court_decisions_by_topic

[54] Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. (2018), 1112–1122. https://doi.org/10.18653/v1/n18-1101

[55] Radboud Winkels, Alexander Boer, Bart Vredebregt, and Alexander Van Someren. 2014. Towards a Legal Recommender System. In *Frontiers in Artificial Intelligence and Applications*, Vol. 271. 169–178.

A DETAILED CASE DESCRIPTIONS AND COMPARISONS

Seed decision: Mugler v Kansas. A new Kansas law prohibited the sale and manufacture of intoxicating liquor. Prior to the passage of the Kansas law, Mugler built a brewery. Mugler was indicted for violating the law and having manufactured intoxicating liquors without a permit. The main issue is if the Kansas law violated the Due Process Clause of the Fourteenth Amendment. More specifically, does prohibiting the sale and manufacture of intoxicating liquors, subsequently lowering the economic value of property, deprive the owner of that property and as articulated in the Due Process Clause of the Fourteenth Amendment?

The court decided that the Kansas law does not infringe on Fourteenth Amendment rights or privileges. It stated that the principle requiring property holders not to use their property so as to be injurious to the community was compatible with the Fourteenth Amendment. Moreover, the court reasoned that a prohibition on the use of property, by valid legislation, for purposes of protecting the health and safety of the community, cannot be deemed a taking or an appropriation of property for public benefit. Since the legislation did not restrict the owner's control, right to dispose, or ability to use property for lawful purposes, no taking had occurred. If the legislature needs to act due to public safety, it cannot discontinue such activity because individuals suffer inconveniences.

Yick Wo v Hopkins. A San Francisco ordinance required all laundries in wooden buildings to hold a permit issued by the city's Board of Supervisors. The board had total discretion over who would be issued a permit. The majority of laundry businesses were operated by Chinese workers, but not a single Chinese owner was granted a permit. Yick Wo and Wo Lee, who operated a laundry business without a permit, were imprisoned after refusing to pay a fine. They sued for habeas corpus and argued that discriminatory enforcement of the ordinance violated their rights under the Equal Protection Clause of the Fourteenth Amendment.

The Supreme Court of California and the Circuit Court of the United States for the District of California denied the claims. The main problem was if the unequal enforcement of the ordinance violates Yick Wo and Wo Lee's rights under the Equal Protection Clause of the Fourteenth Amendment? The Court concluded that, despite the impartial wording of the law, its biased enforcement violated the Equal Protection Clause and therefore violated the provision of the Fourteenth Amendment. The judgment of the Supreme Court of California and Circuit Court of the United States for the District of California were reversed, and the cases remanded.

Like the seed decision, the main problem of this case is a state law that allegedly infringes the Fourteenth Amendment. While the seed decision focuses more on the Due Process Clause, this case addresses the Equal Protection and Citizenship Clause. So we say it is *not related*.

Munn v Illinois. The legislature of Illinois regulated grain warehouses and elevators by establishing maximum rates that private companies could charge for their use and storage of agricultural products. The grain warehouse firm Munn and Scott was found guilty of violating the law. The company appealed the conviction on the grounds that the law was an unconstitutional deprivation of property without due process of law and that the rates deny the warehouse equal protection that violated the Fourteenth Amendment. The court ruled in favor of the State. It argued that the states can regulate the use of private property when the regulation is necessary for the public good. Moreover, the court declared that even though interstate commerce is the responsibility of Congress, a state could take action in the public interest without impairing the federal control.

Similar to the seed decision, the main problem of this case is a state law that allegedly infringes the Fourteenth Amendment. Like the seed decision, this case addresses the violation of the Due Process Clause of the Fourteenth Amendment. In both cases the court argued that individual interests outweigh public interests which justifies the regulations. This led both cases to be ruled in favor of the state. The case is *related*.

Lifestock Dealers Butchers v Crescent City Lifestock 1870. An act passed by the legislature of the State of Louisiana prohibited all persons and corporations to land, keep, or slaughter any animals at any place within the city and parishes of New Orleans. Only the company created and organized under the new act, the "Crescent City Live-stock Landing and Slaughter Company" was entitled to do the aforementioned. The act was passed on March 1869 and was described as an act to protect the health of the city of New Orleans. A group of excluded butchers sought an injunction against the monopoly on the grounds that they were prevented from practising their trade.

The state courts upheld the law. The appeal was based on the following grounds: the act created an involuntary servitude forbidden by the Thirteenth Amendment, it abridges the privileges and immunities of citizens of the U.S., it denied plaintiffs the equal protection of the laws and deprived them of their property without due process of law, which is all protected under the Fourteenth Amendment. The court stated the involuntary servitude of the Thirteenth Amendment is restricted to personal servitude, not a servitude attached to property and that only privileges and immunities of U.S. citizens are protected by the Fourteenth Amendment so that those of state citizens are unaffected. Moreover, the equal protection clause of the Fourteenth Amendment is primarily intended to prevent discrimination against blacks. The court concluded that the prohibition of the plaintiffs' trade cannot be held to be a deprivation of property with regard to the Fourteenth Amendment. This case was the first case requiring interpretation of the amendments.

Similar to the seed decision, the Court had to interpret and apply due process for a regulation in the public interest. The case is *related*.

Butchers Benevolent Crescent City Lifestock 1872. This is another opinion with the same background as the previous case (see Lifestock Dealers Butchers v Crescent City Lifestock 1870), but with a different plaintiff. Again the decision of the court rules that the Fourteenth Amendment did not forbid Louisiana's use of its police powers to regulate butchers. The Court held that the Fourteenth Amendment's Privileges or Immunities Clause affected only rights of U.S. citizenship. Therefore, according to the court, the butcher's Fourteenth Amendment rights had not been violated.

As before, this is *related*.

Lochner v New York. The state of New York enacted the Bakeshop Act, a statute which forbade bakers to work more than 60 hours

a week / 10 hours a day. Lochner was accused of permitting an employee to work more than 60 hours in one week. He was charged with fines. Lochner appealed but lost in state court. He argued that the Fourteenth Amendment should have been interpreted to contain the freedom of contract among the rights emcompassed by substantive due process. In his view the right to purchase or to sell labor should be part of the liberty protected by the amendment.

The question that arises is if the Bakeshop Act violates the liberty protected by the Due Process Clause of the Fourteenth Amendment. The court invalidated the New York statute on the grounds, that it interfered with the freedom of contract and therefore the Fourteenth Amendment's right to liberty afforded employer and employee. Moreover the New York statute failed the rational basis test for determining whether the government action is constitutional. The majority reasoned that the Bakeshop Act had no rational basis because long working hours did not dramatically undermine employees' health and baking is not dangerous per se.

Same as in the seed decision, the court said that the power of the courts to review legislative action in a matter affecting the general welfare exists only when a statute enacted to protect the public health or safety has no real or substantial relation to those objects, or is a plain invasion of rights secured by the Fourteenth Amendment. Different from the seed decision, however, the court found the enacted New York statute to have no rational basis and ruled in favor of the plaintiff. This case is *related*.

Allgeyer v. Louisiana. A Louisiana statute prohibited out-of-state insurance companies from conducting business in Louisiana whithout maintaining at least one place of business and authorized agent within state. The intention behind the implementation of the statute was that it protects citizens from deceitful insurance companies. Allgeyer & Company violated the statute by purchasing insurance from a New-York-based company. The issue was whether the Louisiana statute violates the Fourteenth Amendment's Due Process Clause, which protects companies' liberty to enter in to contracts with businesses of their own choice.

The court ruled in favor of the plaintiff and found that the Louisiana statute deprived Allgeyer & Company of its liberty without Due Process under the Fourteenth Amendment. Moreover, it found that the Fourteenth Amendment extends to protect individuals from restrictions of their freedom to contract in pursuit of one's livelihood or vacation.

Unlike in the seed decision, the Supreme Court of the United States chose to analyze the possible violation of the Fourteenth Amendment from the standpoint of the person rather than the company. The state maintains policing power in relationship to the company, but it cannot legislate in such manner as to deny an individual's liberty. In the seed decision however, the court decided that public health and safety is to prioritize over the individual. This is *related*.

Calder v. Wife. A Connecticut probate court denied Mr. and Mrs. Caleb Bull (the stated beneficiaries of Norman Morrison's will) an inheritance. When the Bulls wanted to appeal the decision more than 1,5 years later, they found that a state law prohibited appeals not made within 18 months of the ruling. The Bulls persuaded the Connecticut legislature to change the restriction, which enabled them to successfully appeal the case. Calder, the initial inheritor of Morrison's estate, took the case to the Supreme Court. The main

issue was, if the Connecticut legislation violates Art. 1 Section 10 of the Constitution, which prohibits ex post facto laws.

The court decided that the Connecticut legislation was not an ex post facto law arguing that restrictions against ex post facto laws were not designed to protect citizens' contract rights but only criminal matters. Moreover, all ex post facto laws are retrospective, but all retrospective laws are not necessarily ex post facto and even vested property rights are subject to retroactive laws.

This case is *not related*.

Davidson v. New Orleans. The city of New Orleans sought to make an assessment on certain real estate within the the parishes of Carroll and Orleans for the purpose of draining swamp lands there. A part of John Davidson's estate was included in the assessment and was assessed for $50,000. The main issue was whether Mrs. Davidson (widow of Mr. Davidson) was being deprived of her property without due process of law clause of the Fourteenth Amendment.

The court ruled against the plaintiff. The court stated that whenever a state takes property for public use, and state laws provide a mode for contesting the charge in the ordinary courts, and if due notice is given to the person, and if there is a full and fair hearing, there is no cause for a suit charging lack of due process of the law. Moreover, the court said that a due process of law does not imply a regular proceeding in a court of justice and that the Fourteenth Amendment was not being infringed.

Similar to the seed case, this case discusses the Due Process Clause of the Fourteenth Amendment. Like in the seed case the state argued that whenever by the laws of a state, or by state authority a burden is imposed upon property for the public use, with notice to the person and/or adequate compensation, it cannot be said to deprive the owner of this property without due process of law. This is *related*.

Muller v. Oregon. Oregon enacted a law that limited women to 10 hours of work in factories and laundries. Curt Muller, the owner of a laundry business, was fined when he violated the law. Muller appealed the conviction. The main issue was whether the Oregon law violated the Fourteenth Amendment.

The court upheld Oregon law. Even though the case Lochner v. New York dealt with the same issues of limiting work hours, the court distinguished this case because of the existing difference between the sexes. Furthermore, the court reasoned that the child-bearing nature and social role of women provided a strong state interest in reducing their working hours.

Similar to in the seed case, the court found the enacted Oregon law to have rational basis since the law protects public health and therefore does not violate the Fourteenth Amendment. *Related*.

Kidd v Pearson. An Iowa state law made the manufacture of liquor in the state illegal, even when the liquor was for sale and consumption out-of-state. The main issue was whether or not the state law was in conflict with the power of Congress to regulate interstate commerce.

The Court decided that there is no conflict and the state law is valid. The Court erected a distinction between manufacture and commerce. The state law regulated manufacturing only. The justices feared that a broad view of commerce that would embrace manufacturing would also embrace the power to regulate every step

of industry. The court ruled that there was not a conflict between Congress' power to regulate interstate commerce and the state law covering manufacturing within a given state. Therefore, the law was valid.

This case is different from the seed case in its focus on interstate commerce rather than due process. But it discusses similar issues as in the seed case. In the seed case the court decided that a state has the right to prohibit or restrict the manufacture of intoxicating liquors within her limits; to prohibit all sale and traffic in them in said State; to inflict penalties for such manufacture and sale, and to provide regulations for the abatement as a common nuisance of the property used for such forbidden purposes; and that such legislation by a State is a clear exercise of her undisputed police power, which does not abridge the liberties or immunities of citizens of the United States, nor deprive any person of property without due process of law, nor in any way contravenes any provision of the Fourteenth Amendment to the Constitution of the United States. In this case the court agreed with the decision of the lead case and ruled similarly in that matter. It is *related*, mostly on factual grounds but also it is somewhat legally related.

Lawton v Steele. A New York statute preserved fisheries from extractive and exhaustive fishing. It said that nets set upon waters of the state or on the shores of or islands in such waters in violation of the statutes of the state enacted for the protection of fish, may be summarily destroyed by any person and asked certain officers to remove them. Steele, a game and fish protector, removed nets of the alleged value of $525 belonging to the plaintiff.

The taking and destruction of the nets were claimed to have been justifiable under the statutes of the state relating to the protection of game and fish. Plaintiffs claimed there was no justification under the statutes, and if they constituted such justification upon their face, they were unconstitutional. The court decided in favor of the defendant, and held the New York statute to be constitutional.

Similar to the seed case, this case discusses whether or not the Fourteenth Amendment was violated with regard to the Due Process Clause. This case is *related*.

Geer v Connecticut. A Connecticut statute provided that it is prohibited to kill woodcook, ruffled grouse, and quail for conveyance across state borders. Geer was convicted of possessing woodcock, ruffled grouse, and quail with the unlawful intent of transporting them out of state.

The Court concluded that the state had the right to keep the game birds within the state for all purposes and to create and regulate its own internal state commerce with respect to the birds. Therefore the statute did not violate the Constitution. The Court explained that the state had the police power to preserve a food supply that belonged to the people of Connecticut by requiring that the commerce in game birds be kept within the state.

Similar to the seed case, the court decided a due process case with regard to the benefit of the people of the state. *Related*.

Groves v Slaughter. A provision of the Mississippi constitution disallowed bringing slaves into the state for sale. Slaughter took a group of slaves to Mississippi to sell them. He accepted partial payment. The note fell due but remained unpaid. A federal court eventually held that Slaughter was entitled to recover the amount of

the contract. This prohibition was challenged as being an unlawful restriction of interstate commerce in violation of the Commerce Clause. The provision did not become effective until a supporting statute was enacted, but that supporting statute followed the sale in question. Hence, the court decided that the contract was valid.

This is somewhat related to the small part of the seed case related to the Interstate Commerce Clause. But it is unrelated to the main focus on Due Process. *Unrelated*.

Rast v. Van Deman. A Florida statute of 1913 imposed special license taxes on merchants using profit sharing coupons and trading stamps. A suit was brought to restrain the enforcement of the statue on the ground that it violates the contract and the commerce clauses and the due process and equal protection provisions of the Fourteenth Amendment.

The court decided that the statute does not offend any constitutional provisions but held that the statute showed that the conditions of complainant's business and property engaged therein are such that enforcement of the statute would produce irreparable injury, it furnishes ground for equitable relief.

Like the seed case, this case addresses the violation of the Due Process Clause and Equal Protection Clause of the Fourteenth Amendment. This is about taxes, rather than regulation, though, so it is *unrelated*.

County of Mobile v. Kimball. An act created a board of commissioners for the improvement of the river, harbor, and bay of Mobile, and required the president of the commissioners of revenue of Mobile County to issue bonds to the amount of $1,000,000, and deliver them, when called for, to the board, to meet the expenses of the work directed. The board was authorized to apply the bonds, or their proceeds, to the cleaning out, deepening, and widening of the river, harbor, and bay of Mobile, or to the construction of an artificial harbor in addition to such improvement. The board of commissioners entered into a contract with the complainants, Kimball and Slaughter, to dredge and cut a channel through a designated bar in the bay. The work agreed upon was completed and accepted by the board through its authorized engineer. The amount due to them was not fully paid.

The court decided that the act of the Legislature of Alabama is invalid, as it conflicts with the commercial power vested in Congress.

This case is on an *unrelated* issue.

Brass v. ND. Ex Rel. Stoeser. A North Dakota state law defined persons operating grain elevators as public warehouse men and regulated their fees and charges. Brass, such an operator, refused to receive certain grain at the storage charges provided by the law, alleging they were too low, and a writ of mandate was issued out of the State Court to require him to do so.

The court affirmed, holding that the power of the State to regulate the grain elevator business did not depend upon the fact of a practical monopoly by the elevator owners. The Court held the law to be constitutional under which the elevator operator was required to make contracts at fees and charges under conditions.

The main issue is upon whether or not the Congress may legislate commercial power. The court decided that the harbor board, created by a law of the State, was authorized to make contracts for a public work in which the county was specially interested, and by which it

would be immediately and directly benefited, and to require obligations of the county to meet the expenses incurred. Furthermore, the court argued that it is enough that by force of the law of its creation it could bind the county for work for which it contracted. Having thus bound the county, the contractors are entitled to the bonds stipulated, or their equivalent in money.

Like the seed case, this case addresses the violation of the Due Process Clause and Equal Protection Clause of the Fourteenth Amendment. It is *related*.

Erie R. Co. v. Williams. The contention of plaintiff is that the Labor Law is repugnant to the Fourteenth Amendment because it deprives the company of property and the employees of liberty without due process of law. The court decided that the law operates not only to require the railroads to pay their employees semi-monthly, but prohibits them from making contracts with their employees which shall vary the time of payment.

The court rejected both contentions of plaintiff and sustained the law as an exercise of the power over plaintiff's charter; and that the requirement of semi-monthly payments was an unconstitutional interference with interstate commerce. The Supreme Court affirmed the previous decision.

Similar to the seed case, this case discusses whether or not the Fourteenth Amendment was violated with regard to the Equal Protection Clause and Due Process Clause. This case is *related*.

Hall v. Geiger-Jones Co. The Ohio blue sky law is a restraint upon the disposition of certain property, and requires dealers in securities evidencing title to or interest in such property to obtain a license. Under the blue sky laws, brokers who sold securities within Ohio were to be licensed to do so. To obtain a license, a designated executive officer needed to be satisfied of the good business repute of the applicants and their agents, and licenses, when issued, could be revoked by him upon ascertaining that the licensees were of bad business repute, violated any provision of the act, or engaged in illegitimate business or fraudulent transactions. Appellee Geiger-Jones Co. filed an action seeking to enjoin enforcement of Ohio's blue sky laws. The district granted the injunctive relief. Hall appealed.

The main question was whether or nor the Ohio's blue sky laws were properly enjoined. The Supreme Court reversed the district court's judgment and remanded the matter for further proceedings. The Court ruled that the powers conferred to Hall were not arbitrary or violative of the due process clause of Fourteenth Amendment. Moreover, the blue sky laws did not interfere with interstate commerce and, therefore, did not violate the commerce clause. According to the Court, such regulation affected interstate commerce in securities only incidentally.

Similar to the seed case, this case discusses whether or not the Fourteenth Amendment was violated with regard to the Due Process Clause. It is *related*.