Clause-Level Similarity and Network Analysis of Standard Service Contract Documents Used in Public Procurement across Japanese National Universities

Nobushige Doi
The University of Tokyo
Tokyo, Japan
n-doi@g.ecc.u-tokyo.ac.jp

Takehito Shimada
The University of Tokyo
Tokyo, Japan
shimada.takehito@mail.u-tokyo.ac.jp

Takehisa Yairi
The University of Tokyo
Tokyo, Japan
yairi@g.ecc.u-tokyo.ac.jp

Abstract

This study examines publicly available standard contract documents for service procurement used by Japanese national universities and identifies both shared standard provisions and university-specific variations through clause-level comparison. We collected documents from 41 national universities, segmented them into 1,432 clauses, aligned comparable clauses using 55 manually assigned and adjudicated label types, and computed inter-clause similarity using character 3-grams and BM25. We further analyzed both verbatim reuse and minor wording adjustments by constructing label-specific clause-similarity networks and an aggregate inter-university similarity measure. The results show that the documents do not converge on a single template; rather, they share a common structural core overlaid with university-specific wording adjustments. Verbatim reuse was observed across clauses in multiple categories, whereas termination clauses shared similar expressions while diverging into several wording clusters. A supplementary geographic analysis found a significant negative association between textual similarity and geographic distance among universities, suggesting that these contract documents are shaped not only by nationwide institutional commonalities but also by regionally shared templates and practical conventions. This study organizes similar contracts issued by different institutions into comparable clause-level data and presents a method for capturing standard provisions and local variations within a unified framework.

CCS Concepts

• **Information systems** → **Similarity measures**; • **Computing methodologies** → *Natural language processing*; • **Applied computing** → Law.

Keywords

national university, contract, public procurement, text analysis, natural language processing, document similarity

1 Introduction

Public procurement is not merely an auxiliary administrative function. According to the OECD [17], public procurement expenditure in OECD countries accounted, on average, for 12.7% of GDP and 29.9% of total government expenditure in 2023, indicating that public procurement is a major governmental activity that supports a wide range of policy objectives. Grandia and Meehan [5] position public procurement as a policy instrument for achieving socially desirable outcomes. In contrast, Malacina et al. [12] propose a framework that understands the value of public procurement not merely in terms of cost reduction but as multidimensional public value spanning buyers, suppliers, and users. In this sense, public procurement should be understood not simply as a purchasing process but as a means of creating public value.

This perspective is likely to apply to public procurement by national universities as well. However, research on public procurement by national universities remains limited compared with the broader public procurement literature. Quayle and Quayle [18] investigated strategic procurement in the further and higher education sectors in the United Kingdom, and Glock and Broens [4] empirically examined the organization of purchasing functions in German universities. Nevertheless, Leal Filho et al. [10] pointed out the scarcity of research on sustainable procurement in higher education institutions. Furthermore, the systematic review by Bhandari et al. [2] shows that research on green public procurement in higher education still rests on a limited body of literature. In other words, although national universities are important actors in public procurement, their role has not yet been sufficiently clarified.

Japan is one country in which national universities tend to disclose a relatively large number of public procurement-related documents. Many Japanese national universities publish regulations governing contract administration and contract terms (*keiyaku kijun*) as foundational documents for public procurement. In this paper, "contract terms" is used as the English rendering of *keiyaku kijun* and refers to a category of standard-form contractual documents used for individual procurement agreements rather than to contract clauses in the generic sense. Many Japanese national universities publish separate contract terms for different procurement types, such as construction, manufacturing, goods supply, and services. Among these, service contract terms are designed to cover a wide variety of procurements. According to our survey, as of December 31, 2025, 41 of Japan's 82 national universities had published service contract terms on their websites. For national university organizations formed through the integration of multiple national university corporations, each organization was counted as one national university.

As noted above, a substantial number of foundational public procurement documents are publicly available from Japanese national universities. Although these universities operate under a common institutional framework, each university independently develops its own regulations and standards. Comparing these documents, therefore, makes it possible to capture inter-university differences in contractual emphasis. Accordingly, analyzing public procurement-related documents from Japanese national universities is meaningful both for addressing the shortage of research on public procurement in higher education and for revealing institutional characteristics and tendencies across universities.

Against this backdrop, this study focuses on service contract terms published by Japanese national universities and compares their contents at the clause level using manually assigned labels to clarify contractual differences across universities. Specifically, we construct a dataset based on labels manually assigned to each clause and measure textual similarity among clauses within each label. We further examine how inter-university similarity relates to geographic proximity as a possible factor underlying that similarity. On this basis, we use inter-university similarity measures and network analysis to extract both standard provisions shared under Japan's common institutional framework and tendencies specific to individual universities.

This paper makes three contributions. First, it constructs a clause-level, label-adjudicated dataset of publicly available Japanese national university service contract terms and releases the coded data and analysis scripts as supplementary materials. Second, it proposes a label-constrained similarity framework that combines BM25-based clause matching, university-level aggregation, and label-specific network analysis. Third, it shows empirically that these documents exhibit both template-like reuse and clustered local variation, with an exploratory association between textual similarity and geographic proximity.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the documents analyzed in this study. Section 4 presents the proposed method. Section 5 reports the experiments, and Section 6 concludes the paper and discusses future work.

2 Related Work

Research on the text analysis of public procurement and legal documents has expanded in recent years. Grandia and Kruyen [6] analyzed more than 140,000 Belgian public procurement notices using text mining. Cao et al. [3] quantitatively evaluated the implementation of sustainable public procurement using a large-scale corpus of tender documents in China. Furthermore, Cifuentes et al. [1] coded government procurement laws in 141 countries and measured the extent of protectionist provisions through comparative analysis of legal texts. With regard to the analysis of legal documents, Linder et al. [11] presented a framework for capturing the proximity among bill texts using text similarity and text reuse, while Nelson et al. [16] showed that combining manual coding with automated methods is effective for capturing complex concepts.

In legal NLP for contracts, English-language corpora such as LEDGAR [21], CUAD [7], and ContractNLI [9] have been developed. However, these resources are primarily intended for tasks such as automatic classification and contract review support, and thus differ from research that compares standard contract documents for public procurement in order to extract differences across organizations. This study aims to bridge the gap between research on university procurement and text analysis in the legal domain.

In information retrieval for legal documents, sparse representations based on lexical matching remain highly effective. In particular, BM25 [20] has been reported to serve as a strong baseline for both case law retrieval and statutory retrieval [14]. In the retrieval of provisions from the Japanese Civil Code, BM25-based methods were also reported to achieve top-level performance in the Competition on Legal Information Extraction and Entailment (COLIEE) 2021 [8, 19]. Therefore, the use of sparse representations such as BM25 is also a reasonable approach in this study.

3 Dataset

The documents analyzed in this study are Japanese-language service contract terms that were publicly available from 41 Japanese national universities as of December 31, 2025. These documents serve as standard-form contracts used by each national university for public procurement of services. The subject matter covered by service contracts is broad and may include, depending on the characteristics of the work, information systems, research support, facility management, and other services. Accordingly, the presence or absence of particular clauses and the degree of detail in service contract terms are likely to differ across national universities, making these documents especially likely to reflect inter-university differences in characteristics and tendencies.

In this paper, to simplify the notation used in figures and tables, we assign an institution ID to each institution under analysis, as shown in Table 1. Even when the entity publishing the contract terms is a university system, such as the Hokkaido National Higher Education and Research System (A1) or the Tokai National Higher Education and Research System (A2), it is treated as a single institution for the purposes of analysis and is referred to collectively as a "university" below for convenience. To construct the dataset, we collected from the official websites of each university or university system documents explicitly identified as service contract terms or as standard contract documents with equivalent functions. When multiple versions were available, we adopted the latest version accessible as of December 31, 2025.

In this study, we constructed a dataset by segmenting these documents into clauses and assigning one or more legal-functional labels to each clause. The labels were developed as a codebook rather than generated automatically: the authors first read the collected service contract terms, grouped clauses by their legal and procurement function, and then refined the groups so that clauses serving the same role could be compared across universities. Multiple labels were allowed because one article sometimes combines several functions, such as a general-provisions clause that also contains a confidentiality obligation.

Two annotators, each with more than five years of experience in public procurement or legal affairs, independently assigned labels to 1,353 annotation cases. They agreed in 1,041 cases, corresponding to a raw agreement rate of 76.9%; the remaining 312 cases were resolved through discussion, with the discussion focused primarily on the points of disagreement. As a result, a total of 1,966 final label assignments were made for 1,432 clauses. The average number of labels per clause was 1.37, while the number of clauses per university ranged from 20 to 43 and the number of distinct labels used ranged from 26 to 46.

Table 1: List of institution IDs

ID	Institution	ID	Institution
01	Hokkaido University	52	Toyohashi University of Technology
02	Hokkaido University of Education	53	Mie University
03	Muroran Institute of Technology	55	Shiga University of Medical Science
04	Otaru University of Commerce	56	Kyoto University
06	Asahikawa Medical University	61	Osaka Kyoiku University
12	Akita University	63	Kobe University
13	Yamagata University	71	Okayama University
15	Ibaraki University	73	Yamaguchi University
17	University of Tsukuba	75	Naruto University of Education
19	Gunma University	81	University of Teacher Education Fukuoka
20	Saitama University	82	Kyushu University
21	Chiba University	84	Kyushu Institute of Technology
22	The University of Tokyo	85	Saga University
23	Institute of Science Tokyo	87	Nagasaki University
34	Yokohama National University	88	Kumamoto University
35	Niigata University	89	Oita University
36	Nagaoka University of Technology	91	University of Miyazaki
45	Shinshu University	96	Japan Advanced Institute of Science and Technology
47	Shizuoka University	A1	Hokkaido National Higher Education and Research System
48	Hamamatsu University School of Medicine	A2	Tokai National Higher Education and Research System
51	Nagoya Institute of Technology		

The final label set comprises 55 label types organized into the seven category groups shown in Table 2. The most frequent labels were Recovery of Liquidated Damages and Damages (130 assignments), Purchaser's Right of Termination (General) (111 assignments), Management of Supplied Materials and Loaned Items (80 assignments), Contractor's Right of Termination (General) (66 assignments), and Delay Damages and Interest (66 assignments). These counts show that the dataset contains both widely recurring core provisions and more specialized provisions that appear only in a small number of universities.

Although all seven category groups are observed in the service contract terms of all national universities, substantial differences can be seen in the scope of adoption and the frequency of individual labels. We also confirmed cases in which exactly identical clause text was reused across multiple universities. Accordingly, this dataset is well suited to comparing, under a common institutional framework, which issues each university makes explicit, to what extent, and in what formulations.

The annotated dataset and the Python scripts used to generate the reported tables and figures are included in the supplementary materials.

4 Proposed Methods

In this study, we use the dataset to quantify similarities between clauses and between universities in the service contract terms of national universities. Because each clause is assigned one or more legal-functional labels, comparisons are first conducted within each label and then aggregated at the university level. The labels are used as an alignment layer: comparing all clauses in one undifferentiated pool would allow general legal phrases and boilerplate expressions to dominate the matching, whereas label-constrained comparison asks whether universities formulate the same procurement issue in similar wording.

Because Japanese text is not segmented by spaces, the design of preprocessing and tokenization substantially affects performance in the analysis of Japanese legal texts [15]. In Japanese information retrieval, character 2-grams and 3-grams are known to yield stable performance [13]. Accordingly, we concatenate the heading and body of each clause, apply Unicode normalization, normalize major full-width and half-width symbols, and remove whitespace. We then represent the resulting text as character 3-grams and use BM25 to retrieve similar clauses within the same label. Below, we describe the method for measuring clause-level similarity, the method for measuring university-level similarity, and the label-specific network analysis.

4.1 Similarity between clauses

To measure similarity between clauses, we use label-specific BM25 retrieval scores. Because BM25 is a directional query–document measure, and the absolute value of the score may vary across queries, we treat each clause as a query, retrieve clauses from other universities within the same label, and then symmetrize the directed scores after normalizing them by the maximum value for each query.

The score for a query clause q and a candidate clause d belonging to label l is defined as follows.

$$\mathrm{BM25}_l(q,d) = \sum_{g \in G_3(q)} \mathrm{tf}(g,q)\,\mathrm{idf}_l(g) \times \frac{\mathrm{tf}(g,d)(k_1+1)}{\mathrm{tf}(g,d) + k_1\left(1 - b + b\frac{|d|}{L_l}\right)} \tag{1}$$

Table 2: Overview of the label codebook

Category group	Label types	Label assignments	Representative labels
Basic provisions	6	182	Purpose, General Provisions, and General Conditions; Submission Documents, Cost Breakdowns, and Plans; Supervising and Inspection Officials
Performance management	15	417	Management of Supplied Materials and Loaned Items; Specification Changes; Inspection and Delivery
Rights and information management	7	209	Confidentiality; Restrictions on Assignment of Rights and Obligations; Restrictions on Subcontracting
Payment and guarantees	6	162	Request and Payment of Contract Price; Contract Guarantee; Change of Contract Amount
Liability and damages	7	396	Recovery of Liquidated Damages and Damages; Delay Damages and Interest; Liability for Contract Non-conformity / Defect Warranty
Termination and misconduct response	11	556	Purchaser's Right of Termination (General); Contractor's Right of Termination (General); Bid-rigging and Misconduct
Supplementary provisions	3	44	Supplementary Provisions and Consultation; Deleted or Missing Articles; Exclusions

$$\mathrm{idf}_l(g) = \log\left(1 + \frac{n_l - \mathrm{df}_l(g) + 0.5}{\mathrm{df}_l(g) + 0.5}\right).$$

Here, $G_3(q)$ denotes the set of distinct character 3-grams appearing in clause q, $\mathrm{tf}(g, q)$ and $\mathrm{tf}(g, d)$ denote the frequency of g in q and d, respectively, $|d|$ is the total number of character 3-grams in candidate clause d, L_l is the average total number of character 3-grams among clauses belonging to label l, n_l is the number of clauses belonging to label l, and $\mathrm{df}_l(g)$ is the number of clauses containing g within label l.

Next, let $u(c)$ denote the university to which clause c belongs, and let $\mathcal{N}_K(c_i, l)$ denote the set of the top K clauses with label l from universities other than $u(c_i)$ for which $\mathrm{BM25}_l(c_i, \cdot)$ is positive. Then, the normalized directed score $\tilde{b}_l(c_i, c_j)$ from query clause c_i to candidate clause c_j is defined as

$$\tilde{b}_l(c_i, c_j) = \begin{cases} \dfrac{\mathrm{BM25}_l(c_i, c_j)}{\max_{c \in \mathcal{N}_K(c_i, l)} \mathrm{BM25}_l(c_i, c)}, & c_j \in \mathcal{N}_K(c_i, l), \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\tilde{b}_l(c_i, c_j) = 0$ when $\mathcal{N}_K(c_i, l) = \emptyset$.

Based on the above, the symmetric clause similarity $s_l(c_i, c_j)$ is defined as

$$s_l(c_i, c_j) = \frac{1}{2}\left(\tilde{b}_l(c_i, c_j) + \tilde{b}_l(c_j, c_i)\right). \tag{3}$$

This definition suppresses differences in the score scale across queries while assigning high similarity to clause pairs that are strongly matched to each other within the same label. The resulting matrix is a cross-university similarity matrix rather than a general identity kernel. Therefore, pairs belonging to the same university, including the diagonal entries, are set to zero to avoid self-loops and within-university matches in the network analysis. Identical clauses appearing in different universities can still receive high similarity and are separately counted by the exact-match indicator introduced below.

4.2 Similarity between universities

Similarity between universities is measured from two perspectives: (i) similarity in the set of issues addressed, represented by the labels they adopt, and (ii) for labels adopted in common, the extent to which the wording of the corresponding clauses is similar. In this subsection, let $\mathcal{L}_u$ denote the set of labels adopted by university u.

First, the label-adoption similarity $A(u, v)$ between universities u and v is defined as the Jaccard similarity of their adopted label sets:

$$A(u, v) = \frac{|\mathcal{L}_u \cap \mathcal{L}_v|}{|\mathcal{L}_u \cup \mathcal{L}_v|}. \tag{4}$$

Next, we define similarity in wording within shared labels. Because the same issue may, depending on the university, be described by multiple clauses, comparison within a label is performed between sets of clauses rather than individual clauses. Let $C_{u,\ell}$ denote the set of clauses with label ℓ in university u, and let $s_\ell(c, c')$ denote the clause similarity defined in the previous subsection. Then, the university-level similarity for label ℓ, $S_\ell(u, v)$, is defined as the bidirectional average in which each clause takes its most similar counterpart on the other side:

$$S_\ell(u, v) = \frac{1}{2}\left(\frac{1}{|C_{u,\ell}|}\sum_{c \in C_{u,\ell}} \max_{c' \in C_{v,\ell}} s_\ell(c, c') + \frac{1}{|C_{v,\ell}|}\sum_{c' \in C_{v,\ell}} \max_{c \in C_{u,\ell}} s_\ell(c, c')\right). \tag{5}$$

Textual similarity between universities, $T(u, v)$, is then defined by taking a weighted average of $S_\ell(u, v)$ over the labels adopted in common. Let $\mathcal{U}$ denote the set of all universities, let the total number of universities be $N = |\mathcal{U}|$, and let the number of universities adopting label ℓ be

$$df_\ell = |\{u \in \mathcal{U} \mid \ell \in \mathcal{L}_u\}|.$$

Then, $T(u, v)$ is given by

$$T(u, v) = \frac{\sum_{\ell \in \mathcal{L}_u \cap \mathcal{L}_v} w_\ell S_\ell(u, v)}{\sum_{\ell \in \mathcal{L}_u \cap \mathcal{L}_v} w_\ell}, \tag{6}$$

$$w_\ell = \log\left(\frac{N + 1}{df_\ell + 1}\right) + 1. \tag{7}$$

This weighting assigns slightly greater weight to agreement on labels adopted by fewer universities than to labels adopted uniformly by many universities. However, when $\mathcal{L}_u \cap \mathcal{L}_v = \emptyset$, we set $T(u, v) = 0$.

Finally, the overall similarity between universities, $M(u,v)$, is defined as the convex combination of the label-adoption similarity $A(u,v)$ and the textual similarity $T(u,v)$:

$$M(u,v) = \alpha A(u,v) + (1-\alpha)T(u,v). \tag{8}$$

Here, $\alpha \in [0,1]$ is the weight assigned to label-adoption similarity. $A(u,v)$ represents the extent to which the two universities make the same issues explicit, whereas $T(u,v)$ represents how similarly they formulate shared issues. Therefore, $M(u,v)$ is an overall measure of proximity between universities that integrates similarity in issue coverage and similarity in wording. For self-comparison, we set $A(u,u) = T(u,u) = M(u,u) = 1$.

In addition, to capture the overall pattern of inter-university similarity, we construct a heatmap of similarities. To determine the display order in the heatmap, we define the distance

$$D(u,v) = 1 - M(u,v)$$

and perform hierarchical clustering using average linkage. The clustering here is an auxiliary procedure for visualization and is not used in the computation of the statistics reported later.

4.3 Geographic association with textual similarity

As a supplementary analysis, we compute the inter-university geographic distance $g(u,v)$ in kilometers from the latitudes and longitudes of representative university locations and examine its relationship with the inter-university textual similarity $T(u,v)$. Because we focus on textual closeness, we use T rather than the overall similarity M. Two analyses are conducted here. First, for each university, one counterpart university with the highest $T(u,v)$ is selected, and if the same university pair is selected reciprocally, the pair is merged into a single point to avoid double-counting in the scatter plot. Second, for a given university, a scatter plot is produced against each of the other 40 universities. Because the distribution of geographic distance is right-skewed, $\log_{10}(1+g(u,v))$ is used on the horizontal axis. This is a supplementary descriptive analysis of the relationship between geographic proximity and textual similarity and does not by itself establish a causal relationship.

4.4 Network analysis by labels

Although the overall similarity between universities is useful for grasping the closeness of the documents as a whole, it does not reveal which issues share template-like wording and which issues exhibit divergence in wording. Therefore, in this study, we construct a clause-similarity network for each label and analyze its structure to distinguish standard provision groups from local variations.

Specifically, nodes correspond to clauses with label ℓ, and an edge is placed between a pair of clauses belonging to different universities when their similarity $s_\ell(c_i, c_j)$ is at least the threshold τ_ℓ. In this paper,

$$\tau_\ell = \max\{Q_{0.85}(\mathcal{S}_\ell),\ 0.25\}$$

where $\mathcal{S}_\ell = \{s_\ell(c_i,c_j) \mid u(c_i) \neq u(c_j),\ s_\ell(c_i,c_j) > 0\}$ and $Q_{0.85}(\mathcal{S}_\ell)$ is its 85th percentile. For the thresholded undirected graph, the statistics in Table 5 and Table 6 are computed over communities identified by weighted greedy modularity community detection. By contrast, in Fig. 3, to directly show separated clusters in the overall network, we use the connected components of the thresholded graph and display the top three components in terms of the number of nodes. For each community g, let $U(g)$ denote the set of universities included in g, and let E_g denote its set of internal edges.

Then, coverage $\mathrm{Cov}(g)$, average edge weight $\mathrm{Coh}(g)$, and exact-match rate $\mathrm{Ex}(g)$ are defined as

$$\begin{aligned} \mathrm{Cov}(g) &= \frac{|U(g)|}{N}, \\ \mathrm{Coh}(g) &= \frac{1}{|E_g|} \sum_{(c_i,c_j)\in E_g} s_\ell(c_i,c_j), \\ \mathrm{Ex}(g) &= \frac{1}{|E_g|} \sum_{(c_i,c_j)\in E_g} \mathbf{1}[z(c_i) = z(c_j)] \end{aligned} \tag{9}$$

where $z(c)$ denotes the clause string obtained by concatenating the heading and body after preprocessing. A community with high $\mathrm{Cov}(g)$ represents a group of clauses shared by many universities, a community with high $\mathrm{Coh}(g)$ represents a group of clauses with strong textual cohesion, and a community with high $\mathrm{Ex}(g)$ represents a group of clauses with substantial reuse of identical wording. Accordingly, a community with both high $\mathrm{Cov}(g)$ and high $\mathrm{Ex}(g)$ is interpreted as a standard provision group, whereas a community with high $\mathrm{Cov}(g)$ and high $\mathrm{Coh}(g)$ but low $\mathrm{Ex}(g)$ is interpreted as a local variation in which each university adjusts the wording under the same issue.

5 Experiments

This section first describes the experimental setup and then presents the results and discussion.

5.1 Experimental Setup

In these experiments, we applied the proposed method described in Section 4 to the dataset constructed in Section 3. For each clause in the dataset, the heading and body were concatenated and, after preprocessing, converted into character 3-grams.

The weight for the label-adoption similarity A was set to $\alpha = 0.5$. For BM25, we fixed the parameters at $k_1 = 1.2$ and $b = 0.75$, following the documented default settings of Elasticsearch[1] and Lucene BM25Similarity[2]. For each query clause, the top $K = 10$ clauses with the same label from other universities were retrieved as candidates. This top-K step is used only to retain plausible cross-university counterparts and to prevent weak lexical overlaps from becoming edges in later aggregation.

For the evaluation of university pairs, we used the overall inter-university similarity M and the textual similarity T. In the label-specific network analysis, we used coverage Cov, average edge weight Coh, and exact-match rate Ex for each community.

5.2 Experimental Results

The mean overall inter-university similarity M was 0.491, with a standard deviation of 0.137. The mean textual similarity T was 0.214, with a standard deviation of 0.216. Figure 1 shows a heatmap of the matrix of overall inter-university similarity M. The figure

[1] Elasticsearch similarity settings, accessed 2026-04-11.
[2] Lucene BM25Similarity API, accessed 2026-04-11.

indicates that proximity among universities is not uniform and that several high-similarity blocks can be observed near the diagonal. This suggests that similar service contract terms are shared not only by a few highly similar university pairs but also at the level of clusters consisting of multiple universities.

Moreover, four of the top five university pairs in terms of overall inter-university similarity M exceeded 0.90, and the fifth-ranked pair also showed a high value of 0.896. These top pairs were well above average both in the structure of the issues addressed and in textual similarity. Four of the top five pairs were either located within the same prefecture or geographically close to each other. Table 3 presents a summary of the top five pairs by overall inter-university similarity.

Figure 2 shows the relationship between geographic distance and inter-university textual similarity $T(u, v)$. In the scatter plot of 27 university pairs, constructed by selecting, for each university, the counterpart with the highest textual similarity and removing duplicate reciprocal selections (Fig. 2(a)), a significant negative correlation was observed between $\log_{10}(1 + \text{distance})$ and $T(u, v)$ ($r = -0.4518$, $p = 0.017986$, $R^2 = 0.2041$). In addition, for Kyushu University (ID: 82), which belongs to one of the university pairs with the highest overall inter-university similarity M, a comparison with the other 40 universities (Fig. 2(b)) also showed a significant negative correlation ($r = -0.3876$, $p = 0.0134577$, $R^2 = 0.1503$).

Table 4 first gives a direct label-level view of wording differences before thresholding the networks. The table reports the number of universities adopting a label, the number of label assignments, the mean pairwise university-level wording similarity S_ℓ, its interquartile range, and the share of university pairs whose reference-normalized clauses are exact matches. The contrast between labels is substantial: for example, Contractor's Termination after Notice to Cure has a mean S_ℓ of 0.380 and a wide interquartile range of 0.908, while Submission Documents, Cost Breakdowns, and Plans has a mean of only 0.162 and an interquartile range of 0.208. Thus, the same method distinguishes labels with a few highly similar wording groups from labels whose wording is more diffusely distributed.

Tables 5 and 6 should be read as the thresholded-network counterpart to this direct similarity table. Table 5 summarizes the scale of the label-wise networks, while Table 6 lists representative labels and the largest community found for each. An exact-match rate of 1.000 means that every edge inside the relevant community connects clauses with identical preprocessed wording; an exact-match rate of 0.000 means that none of those edges is verbatim identical, even though the BM25-based edge weights are high. With this interpretation, the first two rows of Table 6 identify template-like reuse, whereas the lower rows identify high-cohesion but non-verbatim variants. In particular, large-scale verbatim reuse was observed for Notification of Subcontractors and Lower-tier Contractors and Supplementary Provisions and Consultation. By contrast, Purpose, General Provisions, and General Conditions, Purchaser's Right of Termination (General), Measures upon Termination, Recovery of Liquidated Damages and Damages, and Management of Supplied Materials and Loaned Items form large communities with high cohesion but no exact matches, indicating that the same issues are shared while the wording diverges.

At the label level, this tendency reveals two contrasting patterns. First, some labels form communities involving many universities with an exact-match rate of 1.000; these can be interpreted as template-like standard clauses. Second, other labels form communities involving many universities with high cohesion but an exact-match rate of 0.000; these can be regarded as local variations in which each university adjusted the wording under a common issue.

Termination clauses are examined in greater detail as a representative example of wording divergence, because they were adopted by many universities and clearly exhibited the characteristics of high cohesion without verbatim agreement. As shown in Fig. 3, for Purchaser's Right of Termination (General), the top three connected components by number of nodes all showed high cohesion, while their exact-match rates were all 0.000. The largest component included 20 universities, the second included 14 universities, and the third included 8 universities. This indicates that provisions on termination rights do not converge to a single verbatim template, but instead branch into several wording groups sharing similar expressions.

To confirm this tendency across termination clauses as a whole, we conducted an additional comparative analysis of termination clauses. Table 7 shows the comparison of termination clauses. The mean similarities for Purchaser's Right of Termination (General) and Contractor's Right of Termination (General) were 0.208 and 0.215, respectively, which were close to each other; however, on the contractor side, eight exact-match communities were identified, indicating a stronger concentration on specific templates. For termination after notice to cure, the contractor side showed both a larger mean similarity and a wider distribution: Contractor's Termination after Notice to Cure had a mean of 0.380 and an IQR of 0.908, whereas Purchaser's Termination after Notice to Cure had a mean of 0.254 and an IQR of 0.499. For termination without notice to cure, the difference between the two sides was relatively small, at 0.196 for the purchaser side and 0.185 for the contractor side. Furthermore, when Purchaser-related Termination Rights and Contractor-related Termination Rights were viewed as groups, Contractor-related Termination Rights showed a high proportion of zero-similarity pairs, at 65.7%; however, when limited to university pairs with positive similarity, their mean similarity was 0.607, exceeding the purchaser-side value of 0.407.

5.3 Discussion

These results indicate that the service contract terms of Japanese national universities are neither based on a single verbatim template nor necessarily drafted independently by each university.

The geographic analysis based on Fig. 2 is consistent with the observation that geographically proximate universities tend to have more similar wording. One possible explanation is that, when drafting legal documents, the responsible staff at each university may refer to documents used by geographically nearby universities. Another possibility is that geographically proximate universities may deal with businesses sharing similar commercial practices, and may therefore modify contract content to align with those practices. However, the coefficients of determination remain only $R^2 = 0.2041$ in the left panel and $R^2 = 0.1503$ in the right panel,

Table 3: Summary of the top five university pairs by overall inter-university similarity.

University pair	M	T
University of Teacher Education Fukuoka (Fukuoka) – Kyushu University (Fukuoka)	0.974	0.974
Nagoya Institute of Technology (Aichi) – Tokai National Higher Education and Research System (Aichi)	0.969	0.960
Hokkaido University (Hokkaido) – Asahikawa Medical University (Hokkaido)	0.953	0.906
University of Tsukuba (Ibaraki) – Kyoto University (Kyoto)	0.949	0.898
University of Teacher Education Fukuoka (Fukuoka) – Saga University (Saga)	0.896	0.843

Note: The prefecture in which each university is located is shown in parentheses. M means the overall similarity and T means the textual similarity.

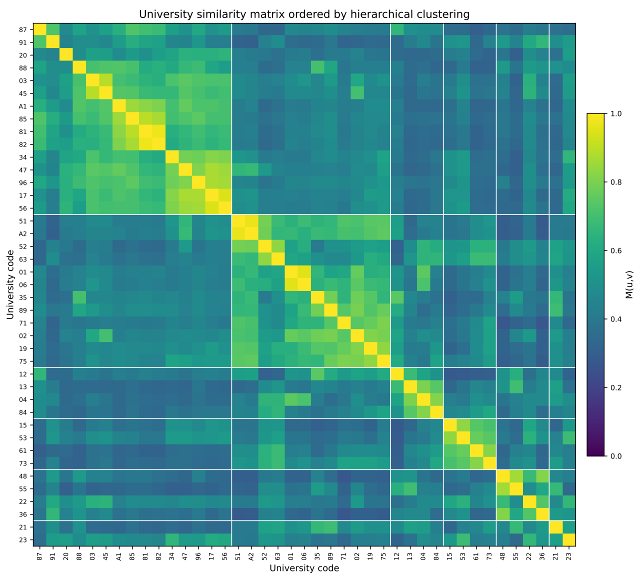


Figure 1: Heatmap of overall inter-university similarity $M(u, v)$, reordered by average-linkage hierarchical clustering. White lines indicate the cluster boundaries obtained by cutting the hierarchy based on the distance $1 - M(u, v)$ at the silhouette-optimal number of clusters.

meaning that geographic proximity alone cannot explain most of the textual similarity. Accordingly, while these results suggest the existence of regional reference relationships or the circulation of templates, it is reasonable to interpret their influence as coexisting with institutional arrangements and practical conventions shared nationwide. The block structure observed in the heatmap can also be understood as the result of such overlapping factors.

Focusing on the label-specific network analysis, the service contract terms of Japanese national universities exhibit broad proximity based on a common institutional skeleton, but this sharing is not necessarily verbatim. The mean cohesion of multi-node communities was high at 0.982, whereas the number of communities with an exact-match rate of 1.000 was limited to 65. Looking at individual labels, Notification of Subcontractors and Lower-tier Contractors and Supplementary Provisions and Consultation showed strong sharing of verbatim templates, whereas Purchaser's Right of Termination (General) and Recovery of Liquidated Damages and Damages suggest that each university adjusted the wording while addressing the same issue. This suggests that, although common reference sources or model templates exist for many issues, each university makes minor revisions or rephrasings when drafting its contract documents.

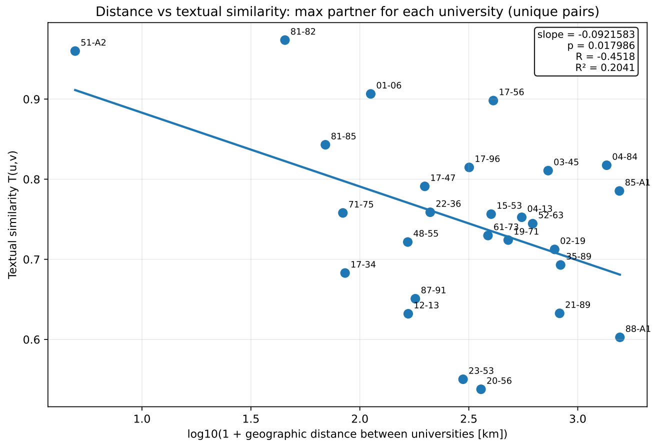
With regard to termination clauses, the wording differs between purchaser-side and contractor-side provisions. Purchaser-side termination rights maintain relatively broad commonality, whereas contractor-side termination rights show both clusters concentrated around specific templates and clauses that are less readily connected across universities. In particular, the large dispersion observed in Contractor's Termination after Notice to Cure suggests that differences in universities' views are more likely to be reflected in the design of contractor-side termination conditions. However, what this study directly measures is not the substantive strength of termination rights themselves, but rather which issues are stipulated and in what wording.

Taken together, these findings suggest that the analytical method proposed in this study is effective for organizing similar contracts issued by different entities into comparable clause-level data and for capturing both standard clauses and local variations within a unified framework. Several limitations should nevertheless be noted. First, this study intentionally emphasizes lexical similarity because the research question concerns template reuse and wording variation, but semantic similarity is also important for detecting clauses that differ in wording while serving an equivalent legal function. Future work should therefore compare the present BM25-based approach with dense embedding methods. Second, the 55-label codebook requires manual legal-functional annotation. The raw agreement rate and adjudication procedure reported in Section 3 mitigate this concern, but future work should examine whether unsupervised or weakly supervised clustering can recover similar clause groups and how much the manual labels improve interpretability. Third, because the analysis is limited to publicly available documents, it does not directly evaluate the practices of universities whose documents are not publicly available, nor the gap between published documents and actual operational practice. This is a limitation arising from sample selection, and caution is therefore required in generalizing the results.

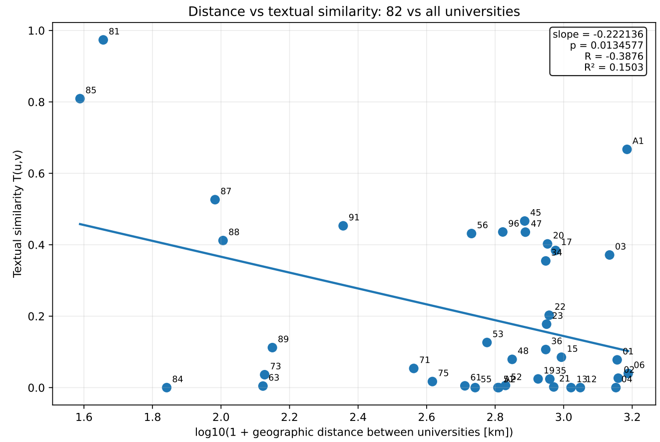
6 Conclusion

In this study, focusing on the 41 Japanese national universities for which service contract terms were confirmed to be publicly available, we captured within a single framework both shared standard provisions and university-specific differences through clause-level comparison. Using 55 manually assigned labels to align comparable clauses, we computed inter-clause similarity based on character 3-grams and BM25, and aggregated the results into an overall inter-university similarity measure and label-specific networks.

The results showed that service contract terms do not conform to a single template; rather, they exhibit a structure in which



(a) For each university, one counterpart university with the maximum textual similarity $T(u, v)$ was selected, and duplicate reciprocal selections were removed.



(b) Comparison between Kyushu University (ID: 82) and the other universities.

Figure 2: Relationship between geographic distance and inter-university textual similarity $T(u, v)$. The horizontal axis is $\log_{10}(1 + \text{distance [km]})$, and the solid line represents the ordinary least squares regression line.

Table 4: Direct label-level wording similarity for representative labels

Label	Adopting universities	Label assignments	Mean S_ℓ	IQR of S_ℓ	Exact reuse
Contractor's Termination after Notice to Cure	25	25	0.380	0.908	0.167
Notification of Subcontractors and Lower-tier Contractors	30	30	0.321	0.983	0.257
Liability for Contract Non-conformity / Defect Warranty	33	58	0.264	0.649	0.001
Purchaser's Right of Termination (General)	41	111	0.207	0.390	0.002
Supplementary Provisions and Consultation	41	41	0.201	0.251	0.134
Inspection and Delivery	41	41	0.171	0.275	0.001
Submission Documents, Cost Breakdowns, and Plans	40	40	0.162	0.208	0.008

Note: S_ℓ is the university-pair wording similarity for label ℓ defined in Eq. 5. Exact reuse is the mean reference-normalized exact-match rate across adopting university pairs for the label. Labels shown here were selected to illustrate high, medium, and low similarity patterns among commonly adopted labels.

Table 5: Overall summary of the label-specific networks

Item	Value
Number of edges (after thresholding)	2,047
Number of communities	750
Multi-node communities	491
Singleton nodes	259
Mean cohesion of multi-node communities	0.982
Median of multi-node communities	0.993
Interquartile range	0.968–1.000
Number of communities with at least 5 universities	102
Number of communities with at least 9 universities	23
Number of communities with exact-match rate = 1.000	65
Number of communities with exact-match rate ≥ 0.500	79

university-specific wording adjustments are layered onto a common framework. In particular, strong verbatim reuse was observed for Notification of Subcontractors and Lower-tier Contractors and Supplementary Provisions and Consultation, whereas termination clauses diverged into multiple wording groups under the same issue. A supplementary geographic analysis also found that geographically proximate universities tended to have more similar contract wording, suggesting that the proximity of contract terms across universities may be shaped not only by nationwide institutional commonalities but also partly by regional reference relationships. It should be noted, however, that what this study directly measures is the closeness of issues and wording across clauses, rather than the substantive strictness or leniency of the provisions themselves. Future work will include a more detailed substantive analysis of termination clauses, comparison with semantic embedding and clustering approaches, an examination of the effect of sample selection, including the presence or absence of publicly available documents, and an extension of the comparison to other contract types and university attributes.

Table 6: Results of network analysis for representative labels

Label	Number of adopting universities	Number of communities	Nodes in the largest community	Cohesion of the largest community	Exact-match rate of the largest community
Notification of Subcontractors and Lower-tier Contractors	30	9	12	1.000	1.000
Supplementary Provisions and Consultation	41	15	11	1.000	1.000
Purpose, General Provisions, and General Conditions	41	14	10	0.964	0.000
Purchaser's Right of Termination (General)	41	43	14	0.974	0.000
Measures upon Termination	41	16	11	0.943	0.000
Recovery of Liquidated Damages and Damages	41	53	12	0.980	0.000
Management of Supplied Materials and Loaned Items	40	35	11	0.970	0.000
Liability for Contract Non-conformity / Defect Warranty	33	21	9	0.984	0.000

Note: "Number of communities" includes singleton nodes. The rightmost three columns show the values for the community with the largest number of nodes for each label.

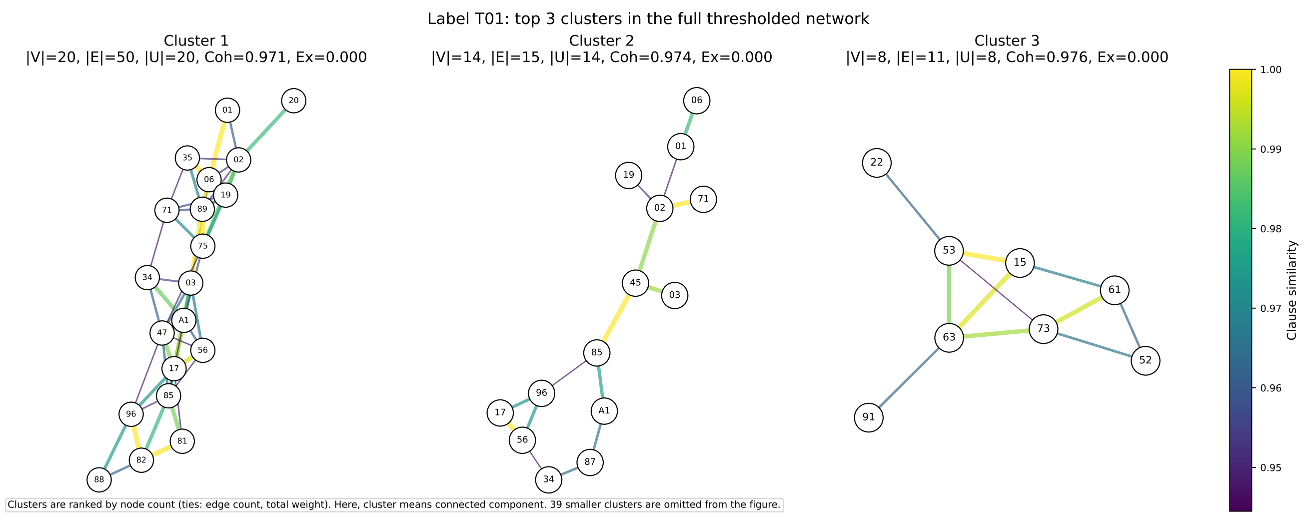


Figure 3: Top three connected components by number of nodes in the thresholded network for the label Purchaser's Right of Termination (General).

Note: Each node represents a clause with a university ID, and edge weights represent clause similarity. Clusters 1–3 in the figure are the connected components of the thresholded graph, displayed in descending order of the number of nodes.

Table 7: Comparison of similarities for termination clauses

Label	n	Mean	Standard deviation	IQR	Exact-match communities
Purchaser's Right of Termination (General)	820	0.208	0.292	0.388	2
Contractor's Right of Termination (General)	820	0.215	0.341	0.414	8
Purchaser's Termination after Notice to Cure	351	0.254	0.324	0.499	0
Contractor's Termination after Notice to Cure	300	0.380	0.447	0.908	3
Purchaser's Termination without Notice to Cure	820	0.196	0.323	0.352	0
Contractor's Termination without Notice to Cure	820	0.185	0.319	0.300	3

Group	Mean similarity	Proportion of zero-similarity pairs	Mean positive similarity
Purchaser-related Termination Rights	0.196	52.0%	0.407
Contractor-related Termination Rights	0.208	65.7%	0.607

Note: "Exact-match communities" denotes the number of multi-node communities within the same label whose exact-match rate is 1.000. In the group rows below, Purchaser-related Termination Rights = {Purchaser's Right of Termination (General), Purchaser's Termination after Notice to Cure, Purchaser's Termination without Notice to Cure}, and Contractor-related Termination Rights = {Contractor's Right of Termination (General), Contractor's Termination after Notice to Cure, Contractor's Termination without Notice to Cure}. For each university pair, similarities were averaged using only the labels adopted by both universities and then aggregated.

References

[1] Carlos Sebastian Barreto Cifuentes, Katayoon Beshkardana, Majed M. El-Bayya, and Lorenzo Rotunno. 2024. *Domestic Laws and Protectionism in Government Procurement*. Policy Research Working Paper 10937. World Bank. https://documents1.worldbank.org/curated/en/099352010012430613/pdf/IDU17835873d19f1e1402b184ca12dcc9be7c3a5.pdf

[2] Santosh Bhandari, Myriam Lopes, and Sara Moreno Pires. 2025. A systematic review of factors influencing green public procurement in higher education institutions. *Discover Sustainability* 6, 1 (2025), 720. doi:10.1007/s43621-025-01562-6

[3] Fuguo Cao, Runyu Li, and Xiaoyan Cao. 2022. Implementation of sustainable public procurement in China: An assessment using quantitative text analysis in large-scale tender documents. *Frontiers in Environmental Science* Volume 10 - 2022 (2022). doi:10.3389/fenvs.2022.947962

[4] Christoph H. Glock and Michael G. Broens. 2011. The Organization of Purchasing at German Universities: An Empirical Investigation. *International Journal of Public Administration* 34, 10 (2011), 662–673. arXiv:https://doi.org/10.1080/01900692.2011.583774 doi:10.1080/01900692.2011.583774

[5] Jolien Grandia and Joanne Meehan. 2017. Public procurement as a policy tool: using procurement to reach desired outcomes in society. *International Journal of Public Sector Management* 30, 4 (05 2017), 302–309. arXiv:https://www.emerald.com/ijpsm/article-pdf/30/4/302/1027084/ijpsm-03-2017-0066.pdf doi:10.1108/IJPSM-03-2017-0066

[6] J. (Jolien) Grandia and P.M. (Peter) Kruyen. 2020. Assessing the implementation of sustainable public procurement using quantitative text-analysis tools: A large-scale analysis of Belgian public procurement notices. *Journal of Purchasing and Supply Management* 26, 4 (2020), 100627. doi:10.1016/j.pursup.2020.100627

[7] Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, J. Vanschoren and S. Yeung (Eds.), Vol. 1. https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/6ea9ab1baa0efb9e19094440c317e21b-Paper-round1.pdf

[8] Mi-Young Kim, Juliano Rabelo, Kingsley Okeke, and Randy Goebel. 2022. Legal Information Retrieval and Entailment Based on BM25, Transformer and Semantic Thesaurus Methods. *The Review of Socionetwork Strategies* 16, 1 (2022), 157–174. doi:10.1007/s12626-022-00103-1

[9] Yuta Koreeda and Christopher Manning. 2021. ContractNLI: A Dataset for Document-level Natural Language Inference for Contracts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 1907–1919. doi:10.18653/v1/2021.findings-emnlp.164

[10] Walter Leal Filho, Antonis Skouloudis, Luciana Londero Brandli, Amanda Lange Salvia, Lucas Veiga Avila, and Lez Rayman-Bacchus. 2019. Sustainability and procurement practices in higher education institutions: Barriers and drivers. *Journal of Cleaner Production* 231 (2019), 1267–1280. doi:10.1016/j.jclepro.2019.05.202

[11] Fridolin Linder, Bruce Desmarais, Matthew Burgess, and Eugenia Giraudy. 2020. Text as Policy: Measuring Policy Similarity through Bill Text Reuse. *Policy Studies Journal* 48, 2 (2020), 546–574. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/psj.12257 doi:10.1111/psj.12257

[12] Iryna Malacina, Elina Karttunen, Aki Jääskeläinen, Katrina Lintukangas, Jussi Heikkilä, and Anni-Kaisa Kähkönen. 2022. Capturing the value creation in public procurement: A practice-based view. *Journal of Purchasing and Supply Management* 28, 2 (2022), 100745. doi:10.1016/j.pursup.2021.100745

[13] Paul McNamee. 2001. Experiments in the Retrieval of Unsegmented Japanese Text at the NTCIR-2 Workshop. In *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*. Tokyo, Japan, 5–157–5–162.

[14] Guilherme Moraes, Ruan Rodrigues, Roberto Lotufo, and Rodrigo Nogueira. 2021. Yes, BM25 is a Strong Baseline for Legal Case Retrieval. *Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment* (2021). https://sites.ualberta.ca/~rabelo/COLIEE2021/COLIEE2021proceedings.pdf

[15] Makoto Nakamura, Yasuhiro Ogawa, and Katsuhiko Toyama. 2014. Extraction of Legal Definitions and Their Explanations with Accessible Citations. In *AI Approaches to the Complexity of Legal Systems*, Pompeu Casanovas, Ugo Pagallo, Monica Palmirani, and Giovanni Sartor (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 157–171.

[16] Laura K. Nelson, Derek Burk, Marcel Knudsen, and Leslie McCall. 2021. The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. *Sociological Methods & Research* 50, 1 (2021), 202–237. arXiv:https://doi.org/10.1177/0049124118769114 doi:10.1177/0049124118769114

[17] OECD. 2025. *Government at a Glance 2025*. OECD Publishing, Paris. doi:10.1787/0efd0bcd-en

[18] Michael Quayle and Sharon Quayle. 2000. The impact of strategic procurement in the UK further and higher education sectors. *International Journal of Public Sector Management* 13, 3 (06 2000), 260–284. arXiv:https://www.emerald.com/ijpsm/article-pdf/13/3/260/1021454/09513550010346008.pdf doi:10.1108/09513550010346008

[19] Juliano Rabelo, Randy Goebel, Mi-Young Kim, Yoshinobu Kano, Masaharu Yoshioka, and Ken Satoh. 2022. Overview and Discussion of the Competition on Legal Information Extraction/Entailment (COLIEE) 2021. *The Review of Socionetwork Strategies* 16, 1 (2022), 111–133. doi:10.1007/s12626-022-00105-z

[20] Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. 1994. Okapi at TREC-3. In *Proceedings of the Third Text REtrieval Conference (TREC-3)*. 109–126. https://trec.nist.gov/pubs/trec3/t3_proceedings.html Conference held in Gaithersburg, MD, USA.

[21] Don Tuggener, Pius von Däniken, Thomas Peetz, and Mark Cieliebak. 2020. LEDGAR: A Large-Scale Multi-label Corpus for Text Classification of Legal Provisions in Contracts. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 1235–1241. https://aclanthology.org/2020.lrec-1.155/