Legal Question Answering System using FrameNet

Ryosuke Taniguchi[1], Reina Hoshino[1] and Yoshinobu Kano[1]

[1] Informatics Course, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu City, Japan
rtaniguchi@kanolab.net rhoshino@kanolab.net
kano@inf.shizuoka.ac.jp

Abstract. A central issue of yes/no question answering is the usage of knowledge source given a question. While yes/no question answering has been studied for a long time, legal yes/no question answering largely differs from other domains. The most distinguishing characteristic is that legal issues require precise analysis of predicate argument structures and semantical abstraction in these sentences. We have developed a yes/no question answering system for answering questions for a statute legal domain. Our system uses a semantic database based on FrameNet, which works with a predicate argument structure analyzer, in order to recognize semantic correspondences rather than surface strings between given problem sentences and knowledge source sentences. We applied our system to the COLIEE (Competition on Legal Information Extraction/Entailment) 2018 task. Our frame based system achieved better scores on average than our previous system in COLIEE 2017, and was the second best score among participants of Task 4. We confirmed effectiveness of the frame information with the COLIEE training dataset. Our result shows the importance of the points described above, revealing opportunities to continue further work on improving our system's accuracy.

Keywords: COLIEE, Question Answering, Legal Bar Exam, Legal Information Extraction, FrameNet.

1 Introduction

Automatic question answering is attracting more interests recently. Due to the increasing expectation to the Artificial Intelligence (AI) technologies, people tend to regard question answering systems as a brand new technology emerged today. However, most successful systems employ rather traditional techniques of question answering which have decades of history [1] [2] [3] [4] [5] [6] [7], including series of shared tasks such as TREC [8], NTCIR [9] and CLEF [10]. This paper describes our challenge to the COLIEE 2018 legal bar exam, which asks participants to answer true or not based on the civil law Articles, given text drawn from the Japanese legal bar exam.

A variety of algorithms and systems has been proposed for question answering. Typically, these question answering systems used *big data* for answering questions [11] [12] [13] [14]. For example, Dumais et al. [15] focused on the redundancy available in large corpora as an important resource. They used this redundancy to simplify their algorithm and to support answer mining from returned snippets. Their system performed quite well given the simplicity of the techniques being utilized.

The now widely known IBM Watson system [16] would be considered as a typical example of such a question answering system of the big data approach. The IBM Watson system won in the Jeoperdy! Quiz TV program competing with human quiz winners. The core Watson system employed a couple of open source libraries, including the traditionally well-designed DeepQA system [17] as its skeleton of question answering processing. Because their target domain, the Jeoperdy! Quiz, could ask broad range of questions, they collected a huge amount of knowledge sources from the Internet, etc., extracting relevant knowledge by combining a couple of different natural language processing (NLP) techniques.

Answering university examinations is another example. The Todai Robot project [18] is a challenge to solve Japanese university examinations, focusing towards attaining a high score in the National Center Test for University Admissions by 2016, and passing the entrance exam of the University of Tokyo (Todai) in 2021 [19]. Although the Todai Robot project tries to achieve higher scores, their aim is rather to reveal the current performance and limitation of the existing AI technologies, using the examinations as its benchmark, similar to the COLIEE's legal bar exam task. In contrast to the COLIEE task, the challenge of Todai Robot project includes variety of subjects including Mathematics, English, Japanese, Physics, History, etc. all written in Japanese language. While solving any problem of these subjects could be considered as question answering, some problems require special technologies. For example, Mathematics and Physics require to process formula; Japanese requires to infer emotions of story characters. Solving the History subjects might be considered as rather an extension of the existing question answering issues. The Todai Robot project achieved better scores than the average of the real human applicants in their Mock Exam challenges.

Recognition of textual entailments (RTE or RITE) is another related issue. RTE has been intensively studied for recent days, including shared tasks such as RTE tasks of PASCAL [20][21], SemEval-2012 Cross-lingual Textual Entailment (CLTE) [22], NTCIR RITE tasks [23][24][25], etc. In the third PASCAL RTE-3 task, contradiction relations are included in addition to entailment relations [21]. In the RTE-6 task, given a corpus and a set of candidate sentences retrieved by a search engine from that corpus, systems are required to identify all the sentences from among the candidate sentences that entail a given hypothesis. NTCIR-9 RITE, NTCIR-10 RITE2, and NTCIR-11 RITEVal Exam Search tasks [25] required participants to find an evidence in source documents and to answer a given proposition by yes or no. Research of RTE normally tries to employ logical processing.

As described above, question answering techniques could include logic, reasoning, syntactic and semantic analysis. Many previous related works tried to employ such deeper analyses. However, required techniques more or less differ depending on a target domain. Another issue is whether the knowledge source needs to be "big data" or not. Regarding the COLIEE's legal problems, required knowledge source can be limited.

In this paper, we suggest using a semantical corpus based on a Rule-based predicate argument structure analyzer in a precise way, rather than to use any machine learning methods. Due to this small data issue, supervised machine learning methods would suffer from insufficient training data. In addition, there are no “similar” problems for most of the legal bar exam problems. Therefore, a solver needs to “comprehend” the contents of the knowledge sources. Moreover, it is difficult to analyze why machine learning answers so, due to their black box architecture. Rule-based methods would make analyses less difficult and are especially effective in a limited domain like legal documents.

Based on these thoughts, we built our yes/no question answering system. Our system does not employ any machine learning. The main method of our system is a predicate argument structure analyzer using FrameNet. We integrated them and applied to COLIEE 2018 Task 4. Our frame based system achieved the second best score among participants. We compared our frame based system with our previous system as baselines, confirming effectiveness of the frame information. There are still many difficult issues remained to be solved though.

We explain about previous works and FrameNet in Section 2. Section 3 describes our design of the yes/no question answering system especially using FrameNet. Section 4 shows our experimental results for this COLIEE task and the comparison with previous system. We discuss our achievements and limitations comparing with previous system in Section 5, mentioning possible future works in Section 6. We conclude our paper in Section 7.

2 Background

2.1 COLIEE

The COLIEE shared task series is held in association with the JURISIN (Juris-informatics) workshop. The first one was the COLIEE 2014 shared task [26]. Following this, COLIEE 2015 shared task [27], COLIEE2016 shared task [28], and COLIEE 2017 [29] shared task (this time in conjunction with ICAIL) were held. This paper mainly describes our participation to the COLIEE 2018 shared task. We call COLIEE 2018 simply as COLIEE in this paper.

The COLIEE shared task consists of four tasks. Task 1 is the legal case retrieval task which involves reading a new case and extracting related cases. Task 2 is the legal case entailment task which compares the new case with related cases given by Task 1.

Task 3 of this legal question answering task involves reading a Japanese legal bar exam question and extracting a subset of Japanese Civil Code Articles. Task 4 is a legal question answering task which requires both of the legal information retrieval system and textual entailment system. Given a set of legal yes/no questions, a participant’s system will retrieve relevant civil law articles. Then, answer yes/no entailment relationship between input yes/no question and the retrieved articles.

2.2 Previous Work

In COLIEE 2016 [30] , our yes/no question answering system was based on case-role analyses using JUMAN [31] and KNP [32]. JUMAN is a Japanese morphological analyzer where we added a custom dictionary for legal technical terms based on a Japanese legal term dictionary ("有斐閣法律用語辞典第4版"). KNP is a Japanese dependency case structure analyzer, works on top of JUMAN. Using results of these tools, we obtained a subject and an end-of-sentence expression for each sentence. A subjective case is normally specified by particles "が" or "は", which are subjective case markers in Japanese. We regarded these cases as subjective cases. When we analyzed the civil law articles, we removed each header part "X条 (Article X)", which includes an article name and numbers. We compared the pairs of the subject and the end-of-expression between the civil law articles and the legal bar exams.

Our COLIEE 2017 [33] system was based on our COLIEE 2016 system above. We defined our own *clause* unit ("節") in order to recognize condition clauses and proposition clauses precisely, which are included in a single sentence. After recognizing condition clauses and proposition clauses in a sentence, we compared corresponding clauses between a given question and civil law articles. A clause should include a predicate as a core element of that clause. We applied a dependency parser that makes chunks ("文節") of a couple of morphemes. Starting form a chunk that includes a predicate, we aggregated neighboring chunks when a neighboring chunk does not include any predicate.

As comparing clauses, we used three modules, a precise match, a loose match and a rough match. The precise match performed exact matches for its predicate, its subject and its object. When we could not find any subject nor any object, we skip that sentence. We outputted yes if everything matched, else outputted no. When there was any negation either in problem clause or in article clause, we reversed the yes/no output. The loose match was looser version of the precise match. When comparing proposition clauses and condition clauses, we outputted yes if either subjects or objects were match in addition to matching predicates. The rough match was the loosest match. We only compared predicates of proposition clauses.

2.3 FrameNet

FrameNet [34] [35] is an English semantical lexical database based on a theory of meaning called frame semantics [36]. Basic idea of frame semantics is that people understand the meaning of words largely by frames which they evoke. Frames have some semantic roles called Frame Elements (FEs). Frame evoking words are called Lexical Units (LUs). For example, a typical situation of shopping involves *buyers, sellers, goods, money, means, rate,* and *unit*. FrameNet has ten kinds of relations (*Inheritance, Using, Perspective_on, Subframe, Precedes, Inchoactive_of, Causative_of, Metaphor, See_also*, and *ReFraming_mapping*) within frames (called Frame Relations). Fig. 1 shows an example of the "Commerce-transaction" frame evoked by the shopping concept in FrameNet. There is a Japanese version of FrameNet [37]. We use LUs of Japanese FrameNet, in addition to the English version of FrameNet.

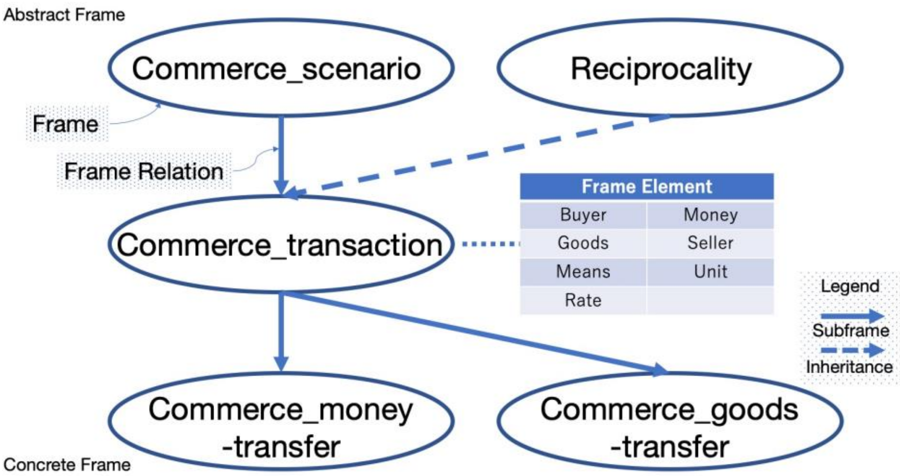

Fig. 1. An example shows a "Commerce-transaction" frame. Each node shows a frame, and an arrow between two frames shows a frame relation.

We use Japanese WordNet [38], in addition to FrameNet. Japanese WordNet is a lexical database for Japanese, where synonyms, hypernyms, hyponyms and English translation words are defined. We use WordNet to expand lexical units of FrameNet.

3 Proposed Method

We use FrameNet in addition to the previous rule-based system. A reason is that structures of civil law articles are clean. The civil law articles use only one place (snippet) for one topic. While our previous system performed textual entailments in a superficial layer, our proposed method using FrameNet could perform in a deeper level. Another reason is that we need precise analyses to solve legal issues, rather than statistically calculate rough estimate values in a superficial way. We took an unsupervised approach for the same reasons.

3.1 Previous Rule-based System

We define two types of clauses in our previous system: proposition clauses and condition clauses. Before using FrameNet, we obtain these clauses from the previous system. We apply a Japanese dependency parser KNP with JUMAN to make the clauses including a set of a predicate, a subject, and an object. When comparing a pair of sets between civil law articles and legal bar exams, we use a *precise match* and a *loose match*. The precise match performs exact matches of strings for its predicate, its subject and its object. The loose match compares either a pair of subjects or a pair of objects, in addition to matching predicates. When our systems could not output any answer, our system answer *yes* as a default output. Additionally, when any negation appears in clauses, we reverse yes/no output.

3.2 Frame-Evoking Words

Our frame based system works like a part of semantic role labeling. Semantic Role Labeling (SRL) is a representative NLP task using FrameNet. The SRL has four processes: (i) identify a frame-evoking word, (ii) identify a frame from the frame-evoking word (frame disambiguation), (iii) estimate words, phrases, or clauses which we have to give FEs, (iv) labeling the FEs. Our frame based system corresponds to these (i) and (ii) processes.

To identify a frame-evoking word, we use a predicate in a proposition clause set which is given by our previous system. Next, we add a candidate of frame-evoking words from the predicate using Japanese WordNet to connect with a specific LU. This is because the number of frame evoking words contained in a LU is small.

We use either English LUs or Japanese LUs. When we use English LUs, we add English translation words in Japanese WordNet to the candidate of word-evoking words. When we use Japanese LUs, we add synonyms, hypernyms and hyponyms in Japanese WordNet to the candidate of word-evoking words. We select LUs which contain one of the word-evoking words. Finally, we make a candidate of frames from the selected LUs. Fig. 2 shows an example of this process.

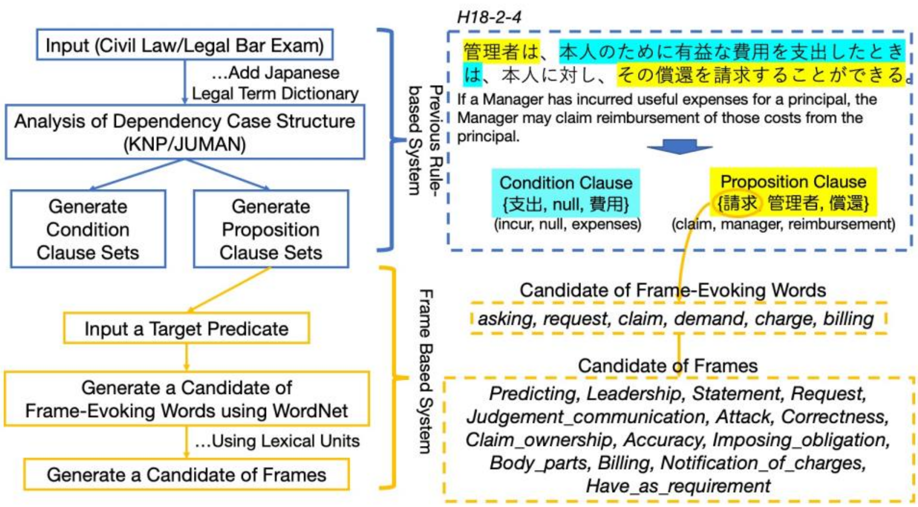

Fig. 2. An example of our frame detection process, using English LUs.

3.3 Frame Disambiguation and Metrics of Frame Confidence

We compare a pair of frame candidates in round robin. We take a pair of frames which confidence value is highest. To calculate the confidence between two frames, we use a shortest path determined by the Dijkstra Algorithm [39] from the entire graph of the frame relations. We assigned a weight value to all of the frame relation types (Table 1). These weights are determined by heuristics. When a weight value is higher between a pair of frames, then we regard this pair as more similar. The following four examples are some of the typical relations.

Inheritance is the strongest relation between frames, corresponding to is-a relationship. So, each frame element in a parent frame should correspond to a frame element in its child frame. Therefore, we set the highest value to this Frame Relation. *Using* is used in a part of a scene evoked by a child frame that refers to its parent frame; some parent frame elements might not have corresponding child frame elements. *Perspective_on* is similar to *Using*. While *Perspective_on* could treat at least two perspectivized frames (e.g. the Commercial_transaction frame specifies a complex scheme involving an exchange of subjects between a seller and a buyer). *Subframe* aggregates frames that form a complex sequence as a whole.

We define the confidence value as a multiplication by the weights of frame relations on the path (Fig. 3). We set threshold for confidence values. When a confidence value is beyond the threshold, we regard the corresponding pair of predicates as similar. Therefore, the lower threshold we set, the more pairs our system could compare. Then

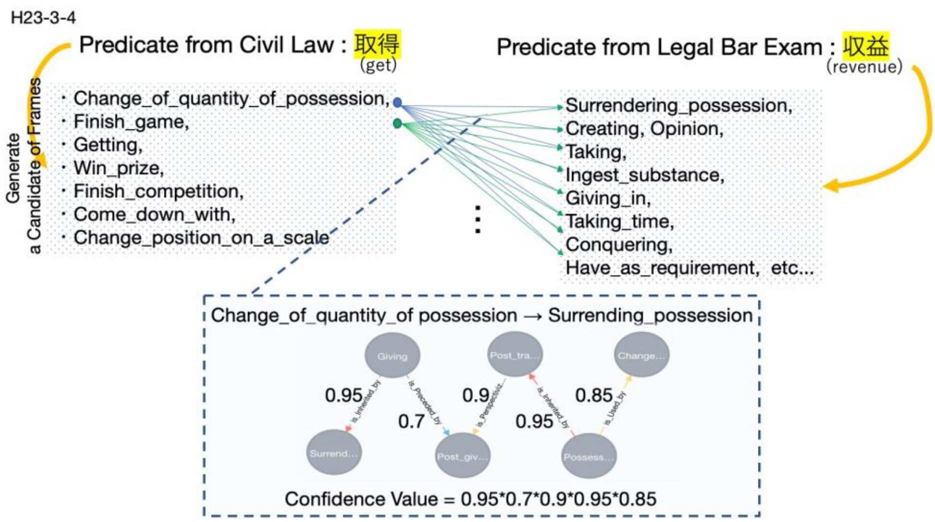

Fig. 3. An example of the confidence value, which is calculated by a multiplication by the weights of frame relations on the path.

we compare the corresponding clauses of civil law articles and legal bar exams extracted by our rule based system as same as our previous system, assuming the corresponding pair of predicates is identical.

FrameRelationType	RelevanceWeignt	FrameRelationType	RelevanceWeight
Inheritance	0.95	Inchoactive_of	0.65
Perspective_on	0.9	Causative_of	0.65
Using	0.85	Metaphor	0.5
Subframe	0.8	See_also	0.5
Precedes	0.7	ReFrameing_Mapping	0.5

Table 1. Weight values of the frame relation types.

4 Experiments and Results

Experiments were conducted on the COLIEE 2018 statute law competition data corpus (Task 4). We did not use the training data except for evaluations, because our frame based system does not use any machine learning method i.e. unsupervised.

4.1 COLIEE Datasets

In this paper, we focused on Task 4. Training data of Task 4's legal questions is drawn from the Japanese legal bar exams. Relevant Japanese civil law articles were also provided. While there was an English translation version of the dataset provided, we only used the original Japanese version. Fig. 4 shows an example of the COLIEE statue law competition data.

t1: （留置権の行使と債権の消滅時効）
第三百条　留置権の行使は、債権の消滅時効の進行を妨げない。
(Exercise of Rights of Retention and Extinctive Prescription of Claims)Article 300
The exercise of a right of retention shall not preclude the running of extinctive prescription of claims.
t2: 留置権者が留置物の占有を継続している間であっても，その被担保債権についての消滅時効は進行する。
Even while the holder of a right to retention continues the possession of the retained property, extinctive prescription runs for its secured claim.

Fig. 4. An example of COLIEE legal bar problem which asks to answer whether t1 entails t2 or not.

4.2 Performance Experiments

In order to investigate our frame based system performance, we used the COLIEE training dataset which includes the past legal bar exam problems and answers. We performed textual entailment part of Task 4, given the gold standard answer of Task 3. We compared a couple of combinations of our modules, in order to observe effects of the frame information. Because the COLIEE dataset is unbalanced, i.e. the number of yes answers and no answers are not equal. When our systems could not output any answers, we tried to fill with either all yes or all no answers as default output to normalize this unbalance. Table 2 shows the result of these performance experiments, using Japanese LUs.

Default Output	Score	PreciseMatch	PreciseMatch + FrameNet	LooseMatch	LooseMatch + FrameNet
N	Average	0.577	0.574	0.595	0.591
	CorrectNum	378	376	390	387
Y	Average	0.522	0.525	0.533	0.542
	CorrectNum	342	344	349	355

Table 2. Results of performance experiments. Our FrameNet system uses Japanese LUs with 0.9 as its threshold.

4.3 Formal Run Experiments

Table 3 shows our formal run results in COLIEE 2018 Task 4. The differences from the performance experiments above. The results of KIS_Frame based on the loose match, which uses English LUs and the threshold is 0.99. The number of comparable

Team	Language	# Correct Answers (total 69 answers)	Accuracy
YA	?	44	0.6388
KIS_Frame	J	39	0.5652
KIS_mo3	J	38	0.5507
KIS_dict	J	37	0.5362
KIS_SVM	J	36	0.5217
KIS_Frame2	J	35	0.5072
UE	E	33	0.4783

Table 3. The COLIEE 2018 formal run results. “J” is using Japanese test datasets, and “E” is English version.

pairs increases by using FrameNet, which becomes too many when comparing all of the civil law articles. We restrict the possible number of the comparisons by setting larger threshold in Task 4. We changed the threshold to 0.7 from 0.99 in the performance experiments, because we use the training datasets which include gold standard civil law articles to be compared with.

KIS_Frame2 is different from KIS_Frame in that the loose, rough and precise match modules are used together. In KIS_Frame2, we use English LUs and the threshold is 0.7.

5 Discussion

Firstly, we observed similar score distribution between the baselines and our frame based systems, among examination years from H18 to H28. Our system should have added new results, rather than changing the entire answer set drastically.

Secondly, there is almost no difference between the precise match results regardless of the FrameNet's effect. These results suggest that there was a little number of clauses in the training set, which the precise match is applicable i.e. a pair of triples (subject, object and predicate) matches. On the other hand, we observed differences in the loose matches. In order to analyze this effect in detail, we focus on H28 (H28-3-5) as shown in Fig. 5. Our previous system cannot compare predicates when their surface strings are different, even though they have comparable similar meanings. In contrast, our frame based system can handle such predicates of similar meanings even if their string forms differ. Our frame based system could answer more problems than our previous system for this reason, which is supported by the actual results.

Thirdly, our frame based system can recognize a pair of predicates which essentially shares a same or similar meaning. However, this is not always the cases. Predicates of abstract meanings, such as "do" and "become", tend to evoke more frames, which results in higher confidence values; used frames sometimes seem not related to the legal

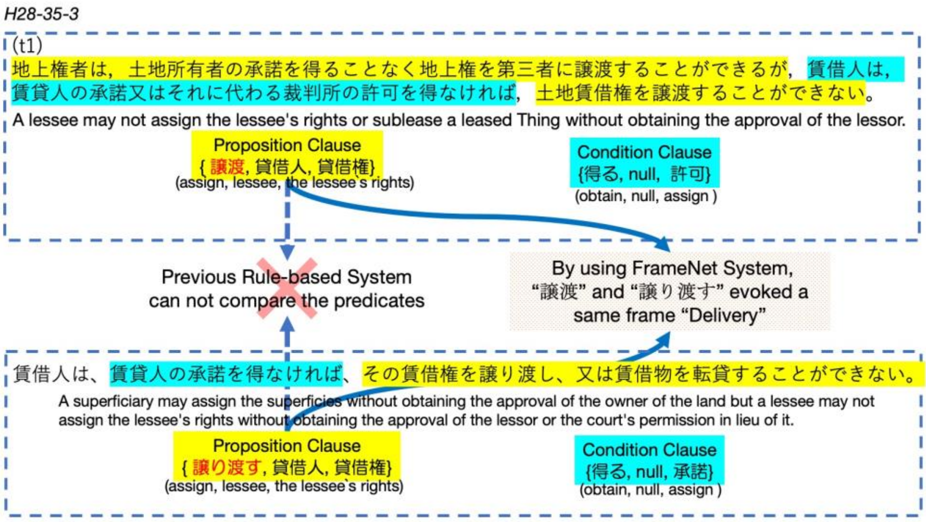

Fig. 5. An example of the effect of FrameNet.

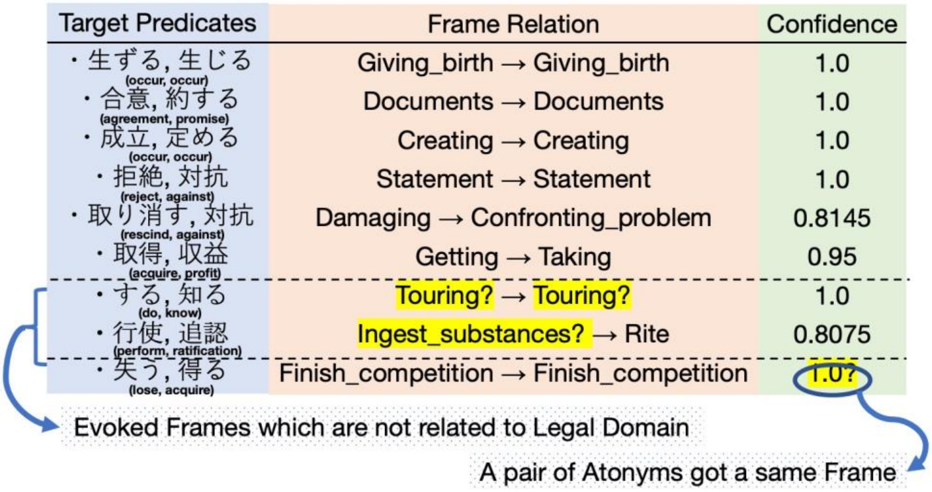

Fig. 6. An example of the analysis results.

domain; a pair of antonyms got the same frame, because they are typically used in the same situation. Fig. 6 shows examples of these cases.

Fourthly, whether using English LUs or Japanese one, we observe different evoked frames. For example, from the predicate "claim (請求)", our system acquired *Predicting, Leadership, Statement, Request, Judgment_communication, Attack, Correctness, Claim_ownership, Accuracy, Imposing_obligation, Body_parts, Billing, Notification_of_charges*, and *Have_as_requirement* Frames by using English LUs, in contrast *Claime_ownership* and *Have_as_requirement* by Japanese LUs. Precise analysis between English LUs and Japanese LUs would be needed to find an effects of FrameNet.

6 Future Work

Japanese text requires explicit tokenization process because there is no space between tokens. When this tokenization fails, final result could also be failed. Therefore, we need to refine the tokenization process and following predicate-argument structure analysis process to be optimized with our frame based system. For example, removing an abstract word could be effective.

We heuristically defined the weights of Frame Relations and the metrics of calculating the Frame confidence. Automatic tuning of the weights with some machine learning technique would be our future work. Using relevant graph theory could also improve the system.

The core of FrameNet are frame elements, in other words, semantical roles. By using frame elements, we could identify frames more precisely, capturing deeper semantic structures.

The most difficult issue to solve in legal domain would be the logic and abstraction, and how we approach these problems using FrameNet.

7 Conclusion

Legal document processing requires a variety of issues to be solved compared with other domains. The most distinguishing characteristic is that legal issues require precise analysis of a predicate argument structure and semantical abstraction. Based on this observation, we developed a yes/no question answering system for legal domain. Our system uses a Japanese case structure analyzer and FrameNet. We applied our system to COLIEE 2018 Japanese task (Task 4). Our system achieved the second best score among Task 4 participants, the best among our systems of different module combinations. We analyzed effectiveness of our frame based system by the training dataset, confirming increase of the scores when our frame based system was used.

Acknowledgements

This work was partially supported by MEXT Kakenhi and JST CREST.

References

[1] D. Lin and P. Pantel, "Discovery of Inference Rules for Question-answering," *Nat. Lang. Eng.*, vol. 7, no. 4, pp. 343–360, 2001.

[2] D. Ravichandran and E. Hovy, "Learning Surface Text Patterns for a Question Answering System," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 41–47.

[3] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*, 2003, pp. 129–136.

[4] D. Pinto, A. McCallum, X. Wei, and W. B. Croft, "Table extraction using conditional random fields," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 235–242.

[5] H. Cui, R. Sun, K. Li, M.-Y. Kan, and T.-S. Chua, "Question Answering Passage Retrieval Using Dependency Relations," in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005, pp. 400–407.

[6] X. Xue, J. Jeon, and W. B. Croft, "Retrieval models for question and answer archives," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 475–482.

[7] J. Bian, Y. Liu, E. Agichtein, and H. Zha, "Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media," in *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 467–476.

[8] E. M. Voorhees and D. K. Harman, *TREC: Experiment and Evaluation in Information Retrieval (Digital Libraries and Electronic Publishing)*. The MIT Press, 2005.

[9] N. Kando, K. Kuriyama, and T. Nozue, "NACSIS Test Collection Workshop (NTCIR-1) (Poster Abstract)," in *Proceedings of the 22Nd Annual International ACM SIGIR*

Conference on Research and Development in Information Retrieval, 1999, pp. 299–300.

[10] M. Braschler, "CLEF 2000 - Overview of Results," in *Revised Papers from the Workshop of Cross-Language Evaluation Forum on Cross-Language Information Retrieval and Evaluation*, 2001, pp. 89–101.

[11] C. C. T. Kwok, O. Etzioni, and D. S. Weld, "Scaling Question Answering to the Web," in *Proceedings of the 10th International Conference on World Wide Web*, 2001, pp. 150–161.

[12] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates, "Web-scale Information Extraction in Knowitall: (Preliminary Results)," in *Proceedings of the 13th International Conference on World Wide Web*, 2004, pp. 100–110.

[13] J. Jeon, W. B. Croft, and J. H. Lee, "Finding Similar Questions in Large Question and Answer Archives," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 2005, pp. 84–90.

[14] H. Kanayama, Y. Miyao, and J. Prager, "Answering Yes/No Questions via Question Inversion," in *the 24th International Conference on Computational Linguistics (COLING 2012)*, 2012, pp. 1377–1391.

[15] S. Dumais, M. Banko, E. Brill, J. Lin, and A. Ng, "Web Question Answering: Is More Always Better?," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2002, pp. 291–298.

[16] D. Ferrucci, "Introduction to 'This is Watson,'" *IBM J. Res. Dev.*, vol. 56, no. 3.4, p. 1:1-1:15, 2012.

[17] D. A. Ferrucci, "IBM's Watson/DeepQA," *SIGARCH Comput. Arch. News*, vol. 39, no. 3, 2011.

[18] N. H. Arai, "The impact of AI—can a robot get into the University of Tokyo?," *Natl. Sci. Rev.*, vol. 2, no. 2, pp. 135–136, Jun. 2015.

[19] "The Todai Robot project." [Online]. Available: http://21robot.org/.

[20] I. Dagan, O. Glickman, and B. Magnini, "The PASCAL Recognising Textual Entailment Challenge," in *Proceedings of the First International Conference on Machine Learning Challenges: Evaluating Predictive Uncertainty Visual Object Classification, and Recognizing Textual Entailment*, 2006, pp. 177–190.

[21] D. Giampiccolo, B. Magnini, I. Dagan, and B. Dolan, "The Third PASCAL Recognizing Textual Entailment Challenge," in *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, 2007, pp. 1–9.

[22] M. Negri, A. Marchetti, Y. Mehdad, L. Bentivogli, and D. Giampiccolo, "Semeval-2012 Task 8: Cross-lingual Textual Entailment for Content Synchronization," in *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, 2012, pp. 399–407.

[23] H. Shima, H. Kanayama, C. Lee, C. Lin, T. Mitamura, Y. Miyao, S. Shi, and K. Takeda, "Overview of NTCIR-9 RITE: Recognizing Inference in TExt," in *NTCIR-9 Workshop*, 2011, pp. 291–301.

[24] Y. Watanabe, Y. Miyao, J. Mizuno, T. Shibata, H. Kanayama, C.-W. Lee, C.-J. Lin, S. Shi, T. Mitamura, N. Kando, H. Shima, and K. Takeda, “Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10,” in *the NTCIR-10 Workshop*, 2013, pp. 385–404.

[25] S. Matsuyoshi, Y. Miyao, T. Shibata, C.-J. Lin, C.-W. Shih, Y. Watanabe, and T. Mitamura, “Overview of the NTCIR-11 Recognizing Inference in TExt and Validation (RITE-VAL) Task,” in *the 11th NTCIR (NII Testbeds and Community for information access Research) workshop*, 2014, pp. 223–232.

[26] “Competition on Legal Information Extraction/Entailment (COLIEE-14), Workshop on Juris-informatics (JURISIN) 2014,” 2014. [Online]. Available: http://webdocs.cs.ualberta.ca/~miyoung2/jurisin_task/index.html.

[27] M.-Y. Kim, R. Goebel, and S. Ken, “COLIEE-2015: Evaluation of Legal Question Answering” in *Ninth International Workshop on Juris-informatics (JURISIN 2015)*, 2015.

[28] M.-Y. Kim, R. Goebel, Y. Kano, and S. Ken, “COLIEE-2016: Evaluation of the Comptetion on Legal Information Extraction and Entailment” in *Tenth International Workshop on Juris-informatics (JURISIN 2016)*, 2016.

[29] Y. Kano, M.-Y. Kim, R. Goebel, and S. Ken, “Overview of COLIEE2017" in *International Conference on Artificial Intelligence and Law (ICAIL 2017)*, 2017.

[30] R. Taniguchi, and Y. Kano, “Legal Yes/No Question Answering System using Case-Role Analysis" in *Tenth International Workshop on Juris-informatics (JURISIN 2016)*, 2016.

[31] “JUMAN (a User-Extensible Morphological Analyzer for Japanese).” [Online]. Available: http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?JUMAN.

[32] “Japanese Dependency and Case Structure Analyzer KNP.” [Online]. Available: http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?KNP.

[33] Y. Kano, R. Hoshino, and R. Taniguchi, “Analyzable Legal Yes/No Question Answering System using Linguistic Structures " in *International Conference on Artificial Intelligence and Law (ICAIL 2017)*, 2017.

[34] “Welcome to FrameNet!.” [Online]. Available: https://framenet.icsi.berkeley.edu/fndrupal/.

[35] J. Ruppenhofer, M. Ellsworth, M. R. L Petruck, C. R. Johnson, C. F. Baker, J. Scheffczyk, “FrameNet II: Extended Theory and Practice”, 2016.

[36] C. J. Fillmore, “Frame Semantics”, in *Linguistics in the Morning Calm*, 1982, pp. 111–137.

[37] “Japanese FrameNet.” [Online]. Available: http://jfn.st.hc.keio.ac.jp.

[38] “Japanese WordNet.” [Online]. http://compling.hss.ntu.edu.sg/wnja/index.en.html.

[39] “Graph Compute with Neo4j: Built-in Algorithms, Spark & Extensions.” [Online]. https://neo4j.com/blog/graph-compute-neo4j-algorithms-spark-extensions/.