Identifying Relationships Among Sentences in Court Case Transcripts Using Discourse Relations

Gathika Ratnayaka[1], Thejan Rupasinghe, Nisansa de Silva, Menuka Warushavithana, Viraj Gamage, Amal Shehan Perera

Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
[1]gathika.14@cse.mrt.ac.lk

Abstract— **Case Law has a significant impact on the proceedings of legal cases. Therefore, the information that can be obtained from previous court cases is valuable to lawyers and other legal officials when performing their duties. This paper describes a methodology of applying discourse relations between sentences when processing text documents related to the legal domain. In this study, we developed a mechanism to classify the relationships that can be observed among sentences in transcripts of United States court cases. First, we defined relationship types that can be observed between sentences in court case transcripts. Then we classified pairs of sentences according to the relationship type by combining a machine learning model and a rule-based approach. The results obtained through our system were evaluated using human judges. To the best of our knowledge, this is the first study where discourse relationships between sentences have been used to determine relationships among sentences in legal court case transcripts.**

Keywords— **discourse relations, natural language processing, machine learning, support vector machine**

I. INTRODUCTION

Case Law can be described as a part of common law, consisting of judgments given by higher (appellate) courts in interpreting the statutes (or the provisions of a constitution) applicable in cases brought before them [1]. In order to make use of the case law, lawyers and other legal officials have to manually go through related court cases to find relevant information. This task requires a significant amount of effort and time. Therefore, automatic extraction of Information from legal court case transcripts would generate numerous benefits to the people working in the legal domain.

From this point onwards, we are referring to the *court case transcripts* as *court cases*. In the process of extracting information from legal court cases, it is important to identify how arguments and facts are related to one another. The objective of this study is to automatically determine the relationships between sentences which can be found in documents related to previous court cases of the United States Supreme Court. Transcripts of U.S. court cases were obtained from FindLaw [2] following a method similar to numerous other artificial intelligence applications in the legal domain [3]–[6].

When a sentence in a court case is considered, it may provide details on arguments or facts related to a particular legal situation. Some sentences may elaborate on the details provided in the previous sentence. It is also possible that the following sentence may not have any relationship with the details in the previous sentence and may provide details about a completely new topic. Another type of relationship is observed when a sentence provides contradictory details to the details provided in the previous sentence. Determining these relationships among sentences is vital to identifying the information flow within a court case. To that end, it is important to consider the way in which clauses, phrases, and text are related to each other. It can be argued that identifying relationships between sentences would make the process of Information Extraction from court cases more systematic given that it will provide a better picture of the information flow of a particular court case. To achieve this objective, we used discourse relations based approach to determine the relationships between sentences in legal documents.

Several theories related to discourse structures have been proposed in recent years. Cross-document Structure Theory (CST) [7], Penn Discourse Tree Bank (PDTB) [8], Rhetorical Structure Theory (RST) [9], [10] and Discourse Graph Bank [11] can be considered as prominent discourse structures. The main difference that can be observed between each of these discourse structures is they have defined the relation types in a different manner. This is mainly due to the fact that different discourse structures are intended for different purposes. In this study, we have based the discourse structure on the discourse structure proposed by CST.

A sentence in a court case transcript can contain different types of details such as descriptions of a scenario, legal arguments, legal facts or legal conditions. The main objective of identifying relationships between sentences is to determine which sentences are connected together within a single flow. If there is a weak or no relation between two sentences, it would probably infer that those two sentences provide details on different topics. Consider the following sentence pair taken from *Lee v. United States* [12] shown in Example 1.

Example 1:
- Sentence 1.1: *The Government makes two errors in urging the adoption of a per se rule that a defendant with no viable defense cannot show prejudice from the denial of his right to trial.*
- Sentence 1.2: *First, it forgets that categorical rules are ill suited to an inquiry that demands a "case-by-case examination" of the "totality of the evidence".*

It can be seen that sentence 1.2 elaborates further on the details provided by sentence 1.1 to give a more comprehensive idea on the topic which is discussed in sentence 1.1. These two sentences are connected to each other within a same flow of information. This can be considered as *Elaboration* relationship, which is a relation type described in CST. Now, Consider the following sentence pair which was also taken from *Lee v. United States* [12]:

Example 2:
- Sentence 2.1: *Courts should not upset a plea solely because of post hoc assertions from a defendant about how he would have pleaded but for his attorney's deficiencies*.
- Sentence 2.2: *Rather, they should look to contemporaneous evidence to substantiate a defendant's expressed preferences.*

In this example, it can be seen that the two sentences have the *Follow Up* relationship as defined in CST. But still, these two sentences are connected together within the same information flow in a court case. There are also situations where we can see sentences are showing characteristics which are common to multiple discourse relations. Therefore, several discourse relations can be grouped together based on their properties to make the process of determining relationships between sentences in court case transcripts more systematic.

The two sentences for Example 3 below were also taken from *Lee v. United States* [12]:

Example 3:
- Sentence 3.1: *The question is whether Lee can show he was prejudiced by that erroneous advice*.
- Sentence 3.2: *A claim of ineffective assistance of counsel will often involve a claim of attorney error "during the course of a legal proceeding"–for example, that counsel failed to raise an objection at trial or to present an argument on appeal*.

The sentence 3.2 follows sentence 3.1. A significant connection between these two sentences cannot be observed. It can also be seen that sentence 3.2 starts a new flow by deviating from the topic discussed in sentence 3.1. These observations which were provided by analyzing court cases emphasize the importance of identifying relationships between sentences.

In this study, we defined the relationship types that are important to be considered when it comes to information extraction from court cases. Next, for each of the relationship type we defined, we identified the relevant CST relations [7]. Finally, we developed a system to predict the relationship between given two sentences of a court case transcript by combining a machine learning model and a rule-based component.

The next section provides an overview on applications of discourse relations related to different domains including the legal domain. Section III describes the methodology which was followed when developing our system. Section IV describes the approaches we took to evaluate the system. The results obtained by evaluating the system are analyzed in Section V. Finally, we conclude our discussion in Section VI.

II. BACKGROUND

Understanding how information is related to each other in machine-readable text has always been a challenge when it comes to Natural Language Processing. Determining the way in which two textual units are connected to each other is helpful in different applications such as text classification, text summarization, understanding the context, evaluating answers provided for a question. Analyzing of discourse relationships or rhetorical relationships between sentences can be considered as an effective approach to understanding the way how two textual units are connected with each other.

Discourse relations have been applied in different application domains related to NLP. [13] describes CST [7] based text summarization approach which involves mechanisms such as identifying and removing redundancy in a text by analyzing discourse relations among sentences. [14] compares and evaluates different methods of text summarizations which are based on RST [10]. In another study [15], text summarization has been carried out by ranking sentences based on the number of discourse relations existing between sentences. [16]–[18] are some other studies where discourse analysis has been used for text summarization. These studies related to text summarization suggest that discourse relationships are useful when it comes to identifying information that discusses on a same topic or entity and also to capture information redundancy. Analysis of discourse relations has also been used for question answering systems [19], [20] and for natural language generation [21].

In the study [22], discourse relations existing between sentences are used to generate clusters of similar sentences from document sets. This study shows that a pair of sentences can show properties of multiple relation types which are defined in CST [7]. In order to facilitate text clustering process, discourse relations have been redefined in this study by categorizing overlapping or closely related CST relations together. In [23], the discourse relationships which are defined in [22] have been used for text summarization based on text clustering. The studies [22], [23] emphasize how discourse relationships can be defined according to the purpose and objective of the study in order to enhance the effectiveness.

When it comes to the legal domain, [24] discusses the potential of discourse analysis for extracting information from legal texts. [25] describes a classifier which determines the rhetorical status of a sentence from a corpus of legal judgments. In this study, a rhetorical annotation scheme was defined for legal judgments. The study [26] provides details on summarization of legal texts using rhetorical annotation schemes. The studies [25], [26] focus mainly on the rhetorical status in a sentence, but not on the relationships between sentences. An approach which can be used to detect the arguments in legal text using lexical, syntactic, semantic and discourse properties of the text is described in [27].

In contrast to other studies, this study is intended to identify relationships among sentences in court case transcripts by analyzing discourse relationships between sentences. Identifying relationships among sentences will be useful in the task of determining how information flows within a court case.

III. METHODOLOGY

A. Defining Discourse Relationships observed in Court Cases

Five major relationship types were defined by examining the nature of relationships that can be observed between sentences in court case transcripts.

- Elaboration - One sentence adds more details to the information provided in the preceding sentence or

one sentence develops further on the topic discussed in the previous sentence.

- Redundancy - Two sentences provide the same information without any difference or additional information.
- Citation - A sentence provides references relevant to the details provided in the previous sentence.
- Shift in View - Two sentences are providing conflicting information or different opinions on the same topic or entity.
- No Relation - No relationship can be observed between the two sentences. One sentence discusses a topic which is different from the topic discussed in the other sentence.

After defining these relationships, we adopted the rhetorical relations provided by CST [7] to align with our definitions as shown in Table I.

TABLE I
ADOPTING CST RELATIONS

Definition	CST Relationships
Elaboration	Paraphrase, Modality, Subsumption, Elaboration, Indirect Speech, Follow-up, Overlap, Fulfillment, Description, Historical Background, Reader Profile, Attribution
Redundancy	Identity
Citation	Citation
Shift in View	Change of Perspective, Contradiction
No Relation	-

It is very difficult to observe the same sentence appearing more than once within nearby sentences in court case transcripts. However, we have included *Redundancy* as a relationship type in order to identify redundant information in a case where the two sentences in a sentence pair are the same.

B. Expanding the Dataset

A Machine Learning model was developed in order to determine the relationship between two sentences in court cases. We used the publicly available dataset of CST bank [28] to learn the Model. The dataset obtained from CST bank contains sentence pairs which are annotated according to the CST relation types. Since we have a labeled dataset [28], we performed supervised learning to develop the machine learning model. Support Vector Machine (SVM) was used as SVMs have shown promising results in previous studies where discourse relations have been used to identify relationships between sentences [22], [23].

Table II provides details on the number of sentence pairs in the data set for each relationship type.

TABLE II
NUMBER OF SENTENCE PAIRS FOR EACH RELATIONSHIP TYPE

CST Relationship	Number of Sentence Pairs
Identity	99
Equivalence	101
Subsumption	590
Contradiction	48
Historical Background	245
Modality	17
Attribution	134
Summary	11
Follow-up	159
Indirect Speech	4
Elaboration	305
Fulfillment	10
Description	244
Overlap (Partial Equivalence)	429

By examining the CST relationship types available in the dataset as shown in Table II, it can be observed that a relationship type which suggests that there is no relationship between sentences cannot be found. But *No Relation* is a fundamental relation type that can be observed between two sentences in court case transcripts. Therefore, we expanded the data set by manually annotating 50 pairs of sentences where a relationship between two sentences cannot be found. This new class was named as *No Relation*. The 50 sentence pairs which were annotated were obtained from previous court case transcripts.

A sentence pair is made up of a source sentence and a target sentence. The source sentence is compared with the target sentence when determining the relationship that is present in the sentence pair. For example, if the source sentence contains all the information in target sentence with some additional information, the sentence pair is said to have the *Subsumption* relationship. Similarly, if the source sentence elaborates the target sentence, the sentence pair is said to have the *Elaboration* relationship.

C. Determining the relationship between sentences using SVM Model

In order to train the SVM model with annotated data, features based on the properties that can be observed in a pair of sentences were defined. Before calculating the features related to words, we removed stop words in sentences to eliminate the effect of less significant words. Also, coreferencing was performed on a given pair of sentences using Stanford CoreNLP CorefAnnotator (*"coref"*) [29] in order to make feature calculations more effective. The two sentences for Example 4 are also taken from *Lee v. United States* [12],

---

Example 4:
- Sentence 4.1 (Target): *Petitioner Jae Lee moved to the United States from South Korea with his parents when he was 13.*
- Sentence 4.2 (Source): *In the 35 years he has spent in this country, he has never returned to South Korea, nor has he become a U.S. citizen, living instead as a lawful permanent resident.*

Here the "Petitioner Jae Lee" in the target sentence, is referred using the pronouns "he" and "his" in both sentences. As all these words are referring to the same person, the system replaces "he" and "his" with their *representative mention* "Petitioner Jae Lee". Then the sentences in Example 4 are changed as shown below.

Example 4 (updated):
- Sentence 4.1 (Target): *Petitioner Jae Lee moved to the United States from South Korea with Petitioner Jae Lee parents when Petitioner Jae Lee was 13.*
- Sentence 4.2 (Source): *In the 35 years Petitioner Jae Lee has spent in this country, Petitioner Jae Lee has never returned to South Korea, nor has Petitioner Jae Lee become a U. S. citizen, living instead as a lawful permanent resident.*

By resolving co-references calculating Noun Similarity, Verb Similarity, Adjective Similarity, Subject Overlap Ratio, Object Overlap Ratio, Subject Noun Overlap Ratio and Semantic Similarity features between two sentences are made more effective.

All the features are calculated and normalized such that their values fall into [0,1] range. We have defined 9 feature categories based on the properties that can be observed in a pair of sentences.

Following 5 feature categories were adopted mainly from [22] though we have done changes in implementation such as use of co-referencing.

1) Cosine Similarities

Following cosine similarity values are calculated for a given sentence pair,

- Word Similarity
- Noun Similarity
- Verb Similarity
- Adjective Similarity

Following equation is used to calculate the above-mentioned cosine similarities.

$$Cosine\ Similarity = \frac{\sum_{i=1}^{n} FV_{S,i} * FV_{T,i}}{\sqrt{\sum_{i=1}^{n}(FV_{S,i})^2} * \sqrt{\sum_{i=1}^{n}(FV_{T,i})^2}}$$

Here $FV_{S,i}$ and $FV_{T,i}$ represents frequency vectors of source sentence and target sentence respectively. Stanford CoreNLP POS Tagger (*"pos"*) [30] is used to identify nouns, verbs and adjectives in sentences.

In calculating the Noun Similarity feature, singular and plural nouns, proper nouns, personal pronouns and possessive pronouns are considered. Both superlative and comparative adjectives are considered when calculating the Adjective Similarity. The system ignores verbs that are lemmatized into "be", "do", "has" verbs when calculating Verb Similarity feature as the priority should be given to effective verbs in sentences.

2) Word Overlap Ratios

Two ratios are considered based on the word overlapping property. One ratio is measured in relation to the target sentence. Other ratio is measured in relation to the source sentence. These ratios provide an indication of the equivalence of two sentences. For example, when it comes to a relationship like *Subsumption*, source sentence usually contains all the words in the target sentence. This property will be also useful in determining relations such as *Identity*, *Overlap (Partial Equivalence)* which are based on the equivalence of two sentences.

$$WOR(T) = \frac{Comm(T,S)}{Distinct(T)}$$

$$WOR(S) = \frac{Comm(T,S)}{Distinct(S)}$$

WOR(S), *WOR(T)* represents the word overlap ratios measured in relation to source and target sentences respectively. *Distinct(S)*, *Distinct(T)* represents the number of distinct words in source sentence and target sentence respectively. The number of distinct common words between two sentences are shown by *Comm(T, S)*.

3) Grammatical Relationship Overlap Ratios

Three ratios which represent the grammatical relationship between target and source sentences are considered.

- Subject Overlap Ratio

$$SubjOverlap = \frac{Comm(Subj(S),\ Subj(T))}{Subj(S)}$$

- Object Overlap Ratio

$$ObjOverlap = \frac{Comm(Obj(S),\ Obj(T))}{Obj(S)}$$

- Subject Noun Overlap Ratio

$$SubjNounOverlap = \frac{Comm(Subj(S),\ Noun(T))}{Subj(S)}$$

All these features are calculated with respect to the source sentence. *Subj, Obj, Noun* represents the number of subjects, objects and nouns respectively. *Comm* gives the number of common elements.

Stanford CoreNLP DependencyParseAnnotator (*"depparse"*) [31] is used here to identify subjects and objects. All the subject types including nominal subject, clausal subject, their passive forms and controlling subjects are taken into account in calculating the number of subjects. Direct and indirect objects are considered when calculating the number of objects. All subject and object types are referred from *Stanford typed dependencies manual* [32].

4) Longest Common Substring Ratio

Longest Common Substring is the maximum length word sequence which is common to both sentences. When the number of characters in longest common substring is taken as n(LCS) and the number of characters in source sentence is taken as n(S), Longest Common Substring Ratio (LCSR) can be calculated as,

$$LCSR = \frac{n(LCS)}{n(S)}$$

This value indicates the part of the target sentence which is present in the source sentence as a fraction. Thus, this will be useful especially in determining discourse relations such as *Overlap*, *Attribution* and *Paraphrase*.

5) Number of Entities

The ratio between the number of named entities can be used as a measurement of the relationship between two sentences.

$$NERatio = \frac{NE(S)}{Max(NE(S),\ NE(T))}$$

NE represents the number of named entities in a given sentence. Stanford CoreNLP Named Entity Recognizer (NER)

---

[33] was used to identify named entities which belong to 7 types; PERSON, ORGANIZATION, LOCATION, MONEY, PERCENT, DATE and TIME.

In addition to the features mentioned above, following features have been introduced to the system.

1) Semantic Similarity between Sentences

This feature is useful in determining the closeness between two sentences. This feature will provide the semantic closeness between two sentences. A method described in [34] is adopted when calculating the semantic similarity between two sentences. Semantic similarity score for a pair of words is calculated using WordNet::Similarity [35].

$$Score = Average\left(\sum_{i=1}^{n} NounScore + \sum_{i=1}^{n} VerbScore\right)$$

2) Transition Words and Phrases

Availability of a transition word or a transition phrase at the start of a sentence indicates that there is a high probability of having a strong relationship with the previous sentence. For example, sentences beginning with transition words like "And", "Thus" usually elaborates the previous sentence. Phrases like "To make that", "In addition" at the beginning of a sentence also implies that the sentence is elaborating on the details provided in the previous sentence. Considering these linguistic properties two Boolean features were defined.

i. Elaboration Transition

If the first word of the source sentence is a transition word which implies elaboration such as "and", "thus", "therefore" or if a transition phrase is found within first six words of the source sentence, this feature will output 1. If both of the above two conditions are false, the feature will return 0. Two lists containing 59 transition words and 91 transition phrases which imply elaboration were maintained. Though it is difficult to include all transition phrases in the English language which implies elaboration relationship, we can clearly say that if these phrases are present at the beginning of a sentence, the sentence is more than likely to elaborate the previous sentence.

ii. Follow-up Transition

If the source sentence begins with words like "however", "although" or phrases like "in contrast", "on the contrary" which implies that the source sentence is following up the target sentence, this feature will output 1. Otherwise, the feature will output 0.

3) Length Difference Ratio

This feature considers the difference of lengths between the source sentence and the target sentence. When *length(S)*, *length(T)* represents the number of words in source sentence and target sentence respectively, Length Difference Ratio (LDR) is calculated as shown below.

$$LDR = 0.5 + \frac{length(S) - length(T)}{2 * Max(length(S), length(T))}$$

In a relationship like *Subsumption*, the length of the source sentence has to be more than the length of the target sentence. In *Identity* relationship, both sentences are usually of the same length. These properties can be identified using this feature.

4) Attribution

This feature checks whether one sentence describes a detail in another sentence in a more descriptive manner. Within this feature, we check whether a word or phrase in one sentence is cited in the other sentence. This is also a Boolean feature. The source sentence and target sentence for Example 5 were obtained from *Turner v. United States* [36]:

Example 5:
- Sentence 5.1 (Target): *Such evidence is 'material' . . . when there is a reasonable probability that, had the evidence been disclosed, the result of the proceeding would have been different.*
- Sentence 5.2 (Source): *A 'reasonable probability' of a different result is one in which the suppressed evidence 'undermines confidence in the outcome of the trial.*

It can be seen that source sentence defines or provides more details on what is meant by "reasonable probability" in the target sentence. Such properties can be identified using this feature.

D. Determining Explicit Citation Relationships in Court Case Transcripts

In legal court case documents, several standard ways are used to point out whence a particular fact or condition was obtained. The target sentence and source sentence in Example 6 are obtained from *Lee v. United States* [12].

Example 6:
- Sentence 6.1 (Target): *The decision whether to plead guilty also involves assessing the respective consequences of a conviction after trial and by plea.*
- Sentence 6.2 (Source): *See INS v. St. Cyr, 533 U. S. 289, 322-323 (2001).*

The two sentences given in Example 6 are adjacent to each other. It can be clearly seen that the source sentence provides a citation for the target sentence. This is only one of the many ways of providing citations in court case transcripts.

After observing different ways of providing citations in court case transcripts, a rule-based mechanism to detect such citations was developed. If this rule-based system detects that there is citation relationship, the pair of sentences will be assigned with the citation relationship. Such a pair of sentences will not be inputted to the SVM model for further processing.

IV. EXPERIMENTS

In order to determine the effectiveness of our system, it is important to carry out evaluations using legal court case transcripts, as it is the domain this system is intended to be used. Court case transcripts related to the United States Supreme Court were obtained from FindLaw [2]. Then the transcripts were pre-processed in order to remove unnecessary data and text. Court case title, section titles are some examples of details which were removed in the pre-processing process. Those details are irrelevant when it comes to determining relationships between sentences.

The relationship types of sentence pairs were assigned using the system. First, the pairs were checked for citation relationship using the rule-based approach. The relationship

types of the sentence pairs where citation relationship could not be detected using the rule-based approach were determined using the Support Vector Machine model.

The results obtained using the system for the sentence pairs extracted from the court case transcripts were then stored in a database. From those sentence pairs, 200 were selected to be annotated by human judges. Before selecting the 200 sentence pairs, they were shuffled to eliminate the potential bias that could have been existent due to a particular court case. Shuffling was helpful in making sure that the sentence pairs to be annotated by human judges were related to different court case transcripts.

Then the selected 200 pairs of sentences to be annotated were grouped together as clusters of five sentence pairs. Each cluster was annotated by two human judges who were trained to identify the relationships between sentence pairs as defined in this study.

V. RESULTS

As expected, the redundancy relationships between sentences could not be observed within the sentence pairs which were annotated using human judges. From the 200 sentence pairs that were observed, our system did not predict *Redundancy* relationship for any sentence pair. Similarly, human judges did not assign the *Redundancy* relationship to any sentence pair.

The confusion matrix which was generated according to the results obtained is given in Table III. The details provided in the matrix are based only on the sentence pairs that were agreed by two human judges to have the same relationship type. The reasoning behind this approach is to eliminate sentence pairs where there are ambiguities of the relationship type between them.

The same approach was used to obtain the results which are presented in Table IV. In contrast, Table V contains results obtained by considering sentence pairs where at least one of the two judges who annotated the pair agrees upon a particular relationship type.

The *Recall* results given in Table IV has a significant importance as all the sentence pairs contained in that results set are annotated with a relationship type which was agreed by two human judges. The *Precision* results provided in Table V indicate the probability of at least one human judge agreeing with the system's predictions in relation to each relationship type. Evaluation results from Table IV, Table V shows that the system works well when identifying *Elaboration*, *No Relation* and *Citation* relationship types where F-measure values are above 75% in all cases. *Shift in View* relationship type was not assigned by the system to any of the 200 sentence pairs which were considered in the evaluation.

Human vs Human correlation and Human vs System correlation when it comes to identifying these relationship types were also analyzed. First, we calculated these correlations without considering the relationship type using the following approach. For a given sentence pair P, $m(P)$ is the value assigned to the pair. n is the number of sentence pairs.

1) Human vs Human Correlation (Cor(H,H))

When both human judges are agreeing on a single relationship type for the pair P, we assign $m(P) = 1$. Otherwise, we assign $m(P) = 0$.

$$Cor(H,H) = \frac{\sum_{P=1}^{n} m(P)}{n}$$

TABLE III
CONFUSION MATRIX

Predicted / Actual	Elaboration	No Relation	Citation	Shift in View	Σ
Elaboration	93.9	6.1	0.0	0.0	99
No Relation	11.9	88.1	0.0	0.0	42
Citation	0.0	4.8	95.2	0.0	21
Shift in View	100	0.0	0.0	0.0	3
Σ	101	44	20	0	165

TABLE IV
RESULTS COMPARISON OF PAIRS WHERE BOTH JUDGES AGREE

Discourse Class	Precision	Recall	F-Measure
Elaboration	0.921	0.939	0.930
No Relation	0.841	0.881	0.861
Citation	1.000	0.952	0.975
Shift in View	-	0	-

TABLE V
RESULTS COMPARISON OF PAIRS WHERE AT LEAST ONE JUDGE AGREES

Discourse Class	Precision	Recall	F-Measure
Elaboration	0.930	0.902	0.916
No Relation	0.846	0.677	0.752
Citation	1.000	0.910	0.953
Shift in View	-	0	-

2) Human vs System Correlation (Cor(H,S))

When both human judges are agreeing with the relationship type predicted by the system for the sentences pair P, we assign $m(P) = 1$. If only one human judge is agreeing with the relationship type predicted by the system for P, we assign $m(P) = 0.5$. If both human judges disagree with the relationship type predicted by the system for P, we assign $m(P) = 0$.

$$Cor(H,S) = \frac{\sum_{P=1}^{n} m(P)}{n}$$

The following results could be observed after calculating the correlations,

- The correlation between a human judge and another human judge = 0.805
- The correlation between a human judge and the system = 0.813

When analyzing these two correlations, it can be seen that our system performs with a capability which is near to the human capability.

The results obtained by calculating Human vs. Human and Human vs. System correlations in relation to each relationship type are given in Table VI. The following

approach was used to calculate these two correlations for each relationship type.

Consider the relationship type R, Let,

- S = The set containing all the sentences pairs which are predicted by the system as having the relationship type R
- U = The set containing all the sentences pairs which were annotated by at least one human judge as having the relationship type R.
- V = The set containing all the sentences pairs which were annotated by two human judges as having the relationship type R.

Corr(H,H) represents Human vs Human correlation and *Corr(H,S)* represents Human vs System correlation. For a given set *A*, *n(A)* indicates the number of elements in set *A*.

$$Corr(H, H) = \frac{n(V)}{n(U)}$$

$$Corr(H, H) = \frac{n(S \cap U)}{n(S \cup U)}$$

The results obtained using this approach is provided in Table VI.

TABLE VI
CORRELATIONS BY TYPE

Discourse Class	Human-Human	Human-System	Human-System / Human-Human
Elaboration	0.750	0.843	1.124
No Relation	0.646	0.603	0.933
Citation	1.000	0.955	0.955
Shift in View	0.188	0.000	0.000

The results which are in Table VI suggest that the system performs with a capability which is near to the human capability when it comes to identifying relationships such as *Elaboration*, *No Relation* and *Citation*. Enhancing system's ability to identify *Shift in View* relationship is one of the major future challenges. At the same time, Human vs Human correlation when it comes to identifying *Shift in View* relationship type is 0.188. This indicates that humans are also having ambiguities when identifying *Shift in View* relationships between sentences.

Either *Elaboration* or *Shift in View* relationship occurs when two sentences are discussing the same topic or entity. *Shift in View* relationship occurs over *Elaboration* when two sentences are providing different views or conflicting facts on the same topic or entity. The *No Relation* relationship can be observed between two sentences when two sentences are no longer discussing the same topic or entity. In other words, *No Relation* relationship suggests that there is a shift in the information flow. As shown in Table III, the sentence pairs with *Shift in View* relationship are always predicted as having *Elaboration* relationship by the system. By observing these results, it can be seen that in most of the cases the system is able to identify whether the sentences are discussing the same topic or not.

VI. CONCLUSIONS

The primary research contribution of this study was the use of discourse relationships between sentences to identify the relationships among sentences in transcripts of United States court cases. Five discourse relationship types were defined in this study in order to automatically identify the flow of information within a court case transcript. This study describes how a machine learning model and a rule-based system can be combined together to enhance the accuracy of identifying relationships between sentences in court case transcripts. Features based on the properties that can be observed between sentences have been introduced to enhance the accuracy of the machine learning model.

The proposed methodology can be successfully applied to identify the sentences which develop on the same discussion topic. In addition, the system is capable of identifying situations in court cases where the discussion topic changes. The system is highly successful in the identification of legal citations. These outcomes demonstrate that the approach described in this study has a promising potential to be applied in tasks related to systematic information extraction from court case transcripts. One such task is the identification of supporting facts, citations which are related to a particular legal argument. Another is the identification of changes in discussion topics within a court case.

The system has difficulties in detecting the occasions where the two sentences are providing different opinions on the same discussion topic. Enhancing this capability in the system can be considered as the major future work.

REFERENCES

[1] WebFinance Inc, "What is case law? definition and meaning - businessdictionary.com," http://www.businessdictionary.com/definition/case-law.html, (Accessed on 05/17/2018)

[2] "Caselaw: Cases and codes – findlaw caselaw," https://caselaw.findlaw.com/, (Accessed on 05/20/2018).

[3] K. Sugathadasa, B. Ayesha, N. de Silva, A. S. Perera, V. Jayawardana, D. Lakmal, and M. Perera, "Synergistic union of word2vec and lexicon for domain specific semantic similarity," in Industrial and Information Systems (ICIIS), 2017 IEEE International Conference on. IEEE, 2017, pp. 1–6

[4] V. Jayawardana, D. Lakmal, N. de Silva, A. S. Perera, K. Sugathadasa, and B. Ayesha, "Deriving a representative vector for ontology classes with instance word vector embeddings," in Innovative Computing Technology (INTECH), 2017 Seventh International Conference on. IEEE, 2017, pp. 79–84.

[5] K. Sugathadasa, B. Ayesha, N. de Silva, A. S. Perera, V. Jayawardana, D. Lakmal, and M. Perera, "Legal document retrieval using document vector embeddings and deep learning," 2018.

[6] V. Jayawardana, D. Lakmal, N. de Silva, A. S. Perera, K. Sugathadasa, B. Ayesha, and M. Perera, "Semi-supervised instance population of an ontology using word vector embedding," in Advances in ICT for Emerging Regions (ICTer), 2017 Seventeenth International Conference on. IEEE, 2017, pp. 1–7

[7] D. R. Radev, "A common theory of information fusion from multiple text sources step one: cross-document structure," in Proceedings of the 1st SIGdial workshop on Discourse and dialogue-Volume 10. Association for Computational Linguistics, 2000, pp. 74–83.

[8] P. Rashmi, D. Nihkil, L. Alan, M. Eleni, R. Livio, J. Aravind, W. Bonnie et al., "The penn discourse treebank 2.0," in Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC08), Marrakech, Morocco, may. European Language Resources Association (ELRA)., 2008

[9] W. C. Mann and S. A. Thompson, Rhetorical structure theory: A theory of text organization. University of Southern California, Information Sciences Institute, 1987.

[10] L. Carlson, M. E. Okurowski, and D. Marcu, RST discourse treebank. Linguistic Data Consortium, University of Pennsylvania, 2002.

[11] F. Wolf, E. Gibson, A. Fisher, and M. Knight, "Discourse graphbank," Linguistic Data Consortium, Philadelphia, 2004.

[12] "Lee v. United States," in US, vol. 432, no. No. 76-5187. Supreme Court, 1977, p. 23.

[13] Z. Zhang, S. Blair-Goldensohn, and D. R. Radev, "Towards cst-enhanced summarization," in AAAI/IAAI, 2002, pp. 439–446.

[14] V. Uzeda, T. Pardo, and M. Nunes, "A comprehensive summary informaˆtiveness evaluation for rst-based summarization methods," International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM) ISSN, pp. 2150–7988, 2009.

[15] M. L. d. R. Castro Jorge and T. A. S. Pardo, "Experiments with cst-based multidocument summarization," in Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing. Association for Computational Linguistics, 2010, pp. 74–82.

[16] D. Marcu, "From discourse structures to text summaries," Intelligent Scalable Text Summarization, 1997.

[17] D. R. Radev, H. Jing, M. Stys, and D. Tam, "Centroid-based summa-´rization of multiple documents," Information Processing & Management, vol. 40, no. 6, pp. 919–938, 2004.

[18] A. Louis, A. Joshi, and A. Nenkova, "Discourse indicators for content selection in summarization," in Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, 2010, pp. 147–156.

[19] K. C. Litkowski, "Cl research experiments in trec-10 question answering," no. 250. National Institute of Standards & Technology, 2002, pp. 122–131.

[20] S. Verberne, L. W. J. Boves, N. H. J. Oostdijk, and P. A. J. M. Coppen, "Discourse-based answering of why-questions," Traitement Automatique des Langues, vol. 47, pp. 21–41, 2007.

[21] P. Piwek and S. Stoyanchev, "Generating expository dialogue from monologue: motivation, corpus and preliminary rules," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010, pp. 333–336.

[22] N. A. H. Zahri, F. Fukumoto, and S. Matsuyoshi, "Exploiting discourse relations between sentences for text clustering," in 24th International Conference on Computational Linguistics, 2012, p. 17.

[23] N. A. H. Zahri, F. Fukumoto, M. Suguru, and O. B. Lynn, "Exploiting rhetorical relations to multiple documents text summarization," International Journal of Network Security & Its Applications, vol. 7, no. 2, p. 1, 2015.

[24] M.-F. Moens, C. Uyttendaele, and J. Dumortier, "Information extraction from legal texts: the potential of discourse analysis," International Journal of Human-Computer Studies, vol. 51, no. 6, pp. 1155–1171, 1999.

[25] B. Hachey and C. Grover, "A rhetorical status classifier for legal text summarisation," Text Summarization Branches Out, 2004.

[26] B. Hachey and C. Grover, "Extractive summarisation of legal texts," Artificial Intelligence and Law, vol. 14, no. 4, pp. 305–345, 2006.

[27] M.-F. Moens, E. Boiy, R. M. Palau, and C. Reed, "Automatic detection of arguments in legal texts," in Proceedings of the 11th international conference on Artificial intelligence and law. ACM, 2007, pp. 225– 230.

[28] D. Radev, J. Otterbacher, and Z. Zhang, "CSTBank: Cross-document Structure Theory Bank," http://tangra.si.umich.edu/clair/CSTBank, 2003.

[29] K. Clark and C. D. Manning, "Entity-centric coreference resolution with model stacking," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), vol. 1, 2015, pp. 1405–1415.

[30] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Featurerich part-of-speech tagging with a cyclic dependency network," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003, pp. 173–180.

[31] D. Chen and C. Manning, "A fast and accurate dependency parser using neural networks," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 740–750.

[32] M.-C. De Marneffe and C. D. Manning, "Stanford typed dependencies manual," Technical report, Stanford University, Tech. Rep., 2008.

[33] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by gibbs sampling," in Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005, pp. 363–370.

[34] M. A. Tayal, M. Raghuwanshi, and L. Malik, "Word net based method for determining semantic sentence similarity through various word senses," in Proceedings of the 11th International Conference on Natural Language Processing, 2014, pp. 139–145.

[35] T. Pedersen, S. Patwardhan, and J. Michelizzi, "Wordnet:: Similarity: measuring the relatedness of concepts," in Demonstration papers at HLTNAACL 2004. Association for Computational Linguistics, 2004, pp. 38–41.

[36] "Turner v. United States," in US, vol. 396, no. No. 190. Supreme Court, 1970, p. 398.