

# Interactive Knowledge Graph for Case Laws

Team 26: Zichen Bai, Yun-Hsuan Chen, Tien-Chi Hsin, Ching-Chieh Huang, Ming-Ying Li, and Ting Liao



## INTRODUCTION

Law texts, albeit precise and descriptive, can sometimes be tedious or overwhelming to the public or even legal professionals. The purpose of our project is to increase the efficiency of information extraction from case texts for law professionals or general public, particularly in the following perspectives:

1. Visualize each case using knowledge graph (KG).
2. Present a collection of cases in an intuitive format.
3. Provide an easy-way of accessing information

### Why is it important?

Using our interface, users will save tremendous amount of time and resources from scrutinizing every case text to get their desired information. Law information will be delivered swifter and clearer to professionals and public, and such is our hope.

## DATA

Data was .txt files crawled from casetext.com. Three sets of keywords were used: "Samsung v. Apple," "Microsoft v. United States," and "Oracle v. Google." A total of 1,179 cases (Avg.6012.73 words each) were collected for topic classification (LDA), and the background section of a total of 510 cases (Avg. 25,803.31 words each) were collected.

## RESULT

After applying the LDA model, we found three topics with 141, 685, and 353 cases, respectively. We therefore named the three topics as Competition Disputes, Procedural Disputes, and Substantive Disputes (Algorithm, Patent). After conducting relation extraction, we acquired the subject-relation-object triples for 510 cases that are ready for visualization.

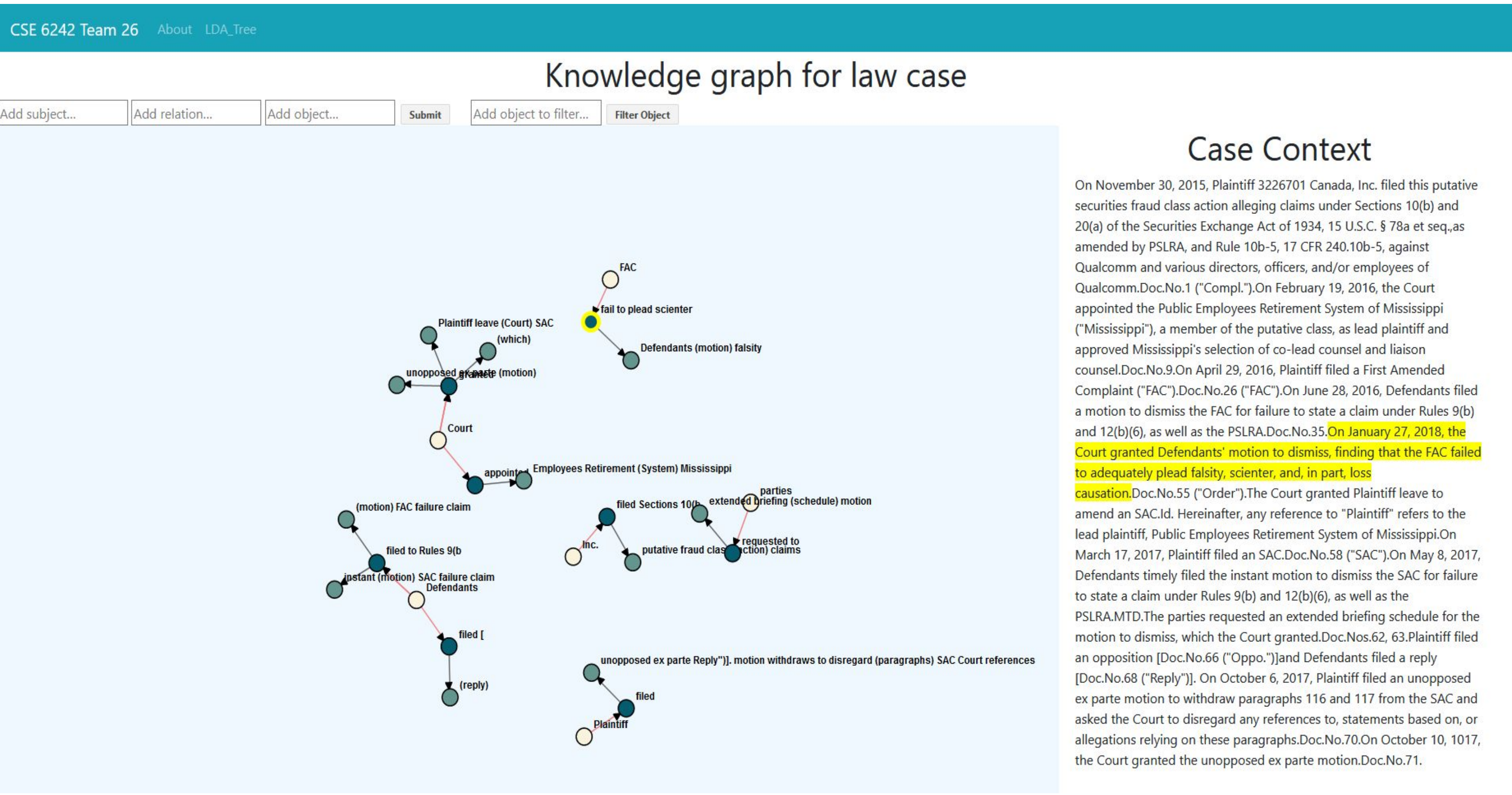
## METHODS

Data Collection	Case texts was crawled and manually collected from casetext.com. They were then processed into full text and background section text to feed into case clustering and relation extraction methods mentioned below.
Case Clustering	Implemented Latent Dirichlet Allocation (LDA) to make the case clustering. The key factor is that the algorithm uses the matching rate to classify the cases into the highest matching topic.
Relation Extraction	Used SpaCy language processing features and additional grammer matching rules to extract the subject-relation-object triples from the background section of case texts into JSON format files.
Visualization	With the JSON outputs from clustering and relation extraction, we used D3 to create an interactable interface of knowledge graph and collapsible tree for users to visualize and access the case information.

### Designed Use Flow:

When users enter the interface, cases are first presented under topic clusters represented by a tree. Users may explore the topics and cases underneath. When an interesting case is found, users can click on the case to bring up the KG for that case. In the KG, users will see the extracted relation triples as well as the full case text.

## VISUALIZATION



## EVALUATION

- (1) Case Clustering: The perplexity of our LDA model is 2677.43, which is reasonable as our model copes with 1,179 cases (Avg. 6,012.73 words). In human evaluation, we record the probability of the classified topic for each document and observe whether the probability and document are reasonable.
- (2) Relation Extraction: our current system is able to extract a majority of them correctly, and the replacement of pronouns helped enhance the readability and the correctness of the results.
- (3) Visualization: our interface showed advantages in the interaction ability and information presentation. The topic tree help users easily find cases, and the KG assist users to comprehend case laws.

## CONCLUSION

In this project, we created a web crawler to obtain case text from webpages as our data. Cases were organized into 3 topics identified by LDA. We then implemented new rule-based algorithms for relation extraction to build up our KG of each case background. We mainly evaluate both performances in human evaluation. We perform gridsearch and perplexity to obtain a reasonable LDA model. The extracted results suggested a 35% improvement from a simple information extraction method. The word clouds show highly-related words in each topic. Collapsible topic trees and relation KGs were constructed for law case visualization. An interactive interface website was created to access the case information. The interface also allows users to customize the KG by adding nodes and edges, as well as applying filters.