

# XÂY DỰNG ĐỒ THỊ TRI THỨC TỪ VĂN BẢN PHÁP LÝ VỚI NHẬN DIỆN THỰC THỂ DỰA TRÊN LLM

Phan Doãn Thái Bình - 240101003

# Tóm tắt

- Lớp: CS2205.CH183
- Link Github của nhóm:  
<https://github.com/binh120702/CS2205.CH183>
- Link YouTube video: <https://youtu.be/TAWZ4LSmgRo>
- Phan Doãn Thái Bình - 240101003



# Giới thiệu

**S1:** Bà Dương Thị N khởi kiện yêu cầu tuyên bố hợp đồng đặt cọc ngày 2/6 giữa bà và bà Đặng Thị P vô hiệu

← PER →

← DATE →

← PER →

**S2:** Quy định tại các điều 6 và điều 7 Luật Thi hành án dân sự...

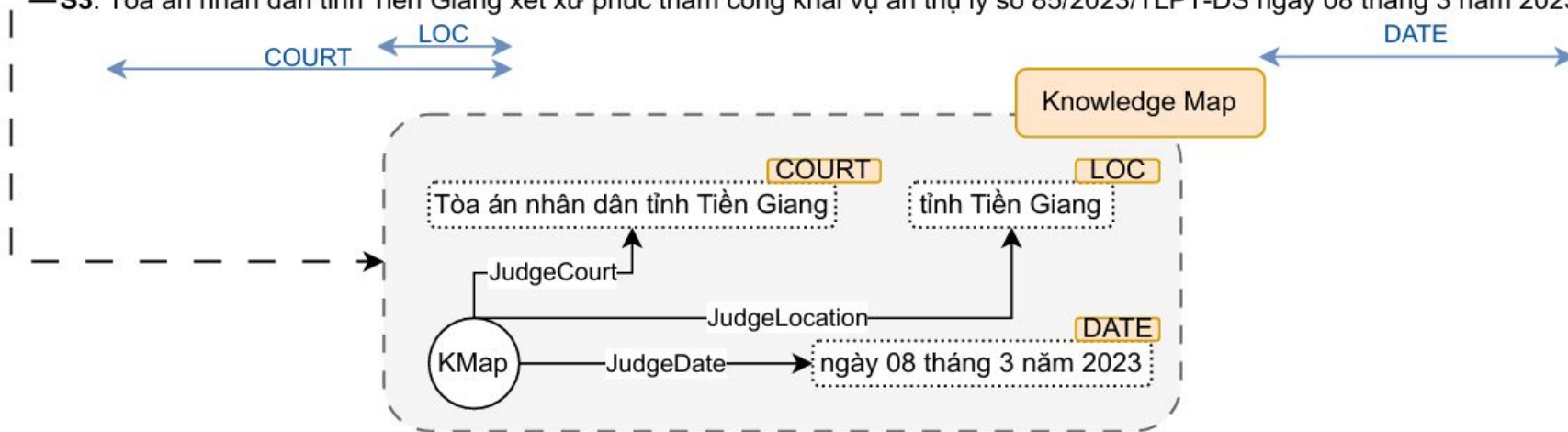
← PROVISION →

← PROVISION →

— **S3:** Tòa án nhân dân tỉnh Tiền Giang xét xử phúc thẩm công khai vụ án thụ lý số 85/2023/TLPT-DS ngày 08 tháng 3 năm 2023

← COURT ← LOC →

← DATE →



# Giới thiệu

- Hệ thống pháp lý chứa khối lượng lớn văn bản phức tạp
- Tài liệu pháp lý có ngôn ngữ chuyên ngành, cấu trúc lồng nhau
- Việc tìm kiếm và phân tích thông tin pháp lý còn nhiều hạn chế
- Giải pháp: Ứng dụng mô hình ngôn ngữ lớn (LLMs) để nhận diện thực thể và xây dựng đồ thị tri thức pháp lý



# Mục tiêu

- Xây dựng hệ thống nhận diện thực thể pháp lý bằng mô hình LLM
- Tổ chức thông tin thành đồ thị tri thức, liên kết thực thể pháp lý
- Đánh giá hệ thống trên bộ dữ liệu pháp lý tiếng Việt

# Nội dung và Phương pháp

- Thu thập và tiền xử lý dữ liệu: thuvienphapluat
- Nhận diện thực thể pháp lý (Legal NER):
  - Xác định các thực thể trong văn bản (tòa án, luật, điều khoản...)
  - Đặc thù thực thể lồng nhau trong tài liệu pháp lý
- Đồ thị tri thức (Knowledge Graphs):
  - Mô hình hóa thông tin dưới dạng quan hệ giữa các thực thể
  - Ứng dụng trong tìm kiếm, phân tích tri thức
- Đánh giá và kiểm thử hệ thống

# Nội dung và Phương pháp

- Nhận diện thực thể bằng LLM

Given the following set of labels: {set\_of\_labels}

Based on the provided set of labels,  
identify named entities (which may be nested) in  
the following passage:

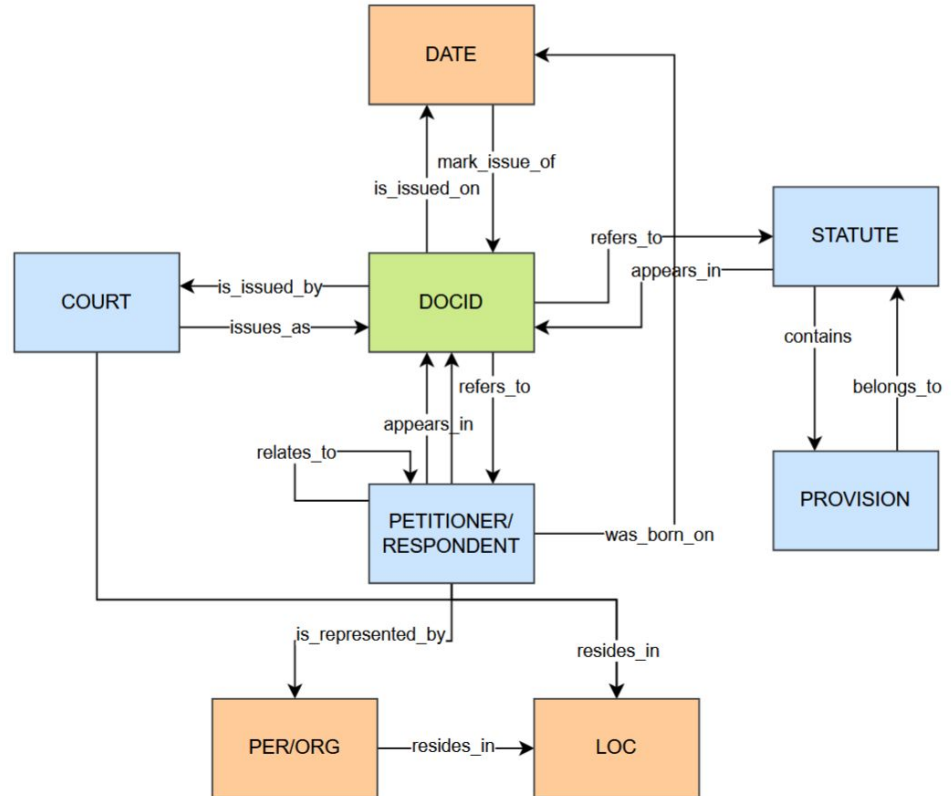
{legal\_document}

Question: Which entities belong to the {target\_label} class  
in the passage above?

Answer in a single line, listing the entities separated by  
"\_\_\_\_", do not add any spare information.

# Nội dung và Phương pháp

- Xây dựng đồ thị tri thức





# Kết quả dự kiến

- Nhận diện chính xác các thực thể pháp lý
- Đồ thị tri thức pháp lý phục vụ tìm kiếm, phân tích dữ liệu
- Bộ dữ liệu gán nhãn hỗ trợ nghiên cứu NLP pháp lý

# Tài liệu tham khảo

- [1]. Erwin Filtz, Sabrina Kirrane, and Axel Polleres. The linked legal data landscape: linking legal data across different countries. *Artif. Intell. Law* 29, 4 (Dec. 2021): 485–539.
- [2]. Hu Zhang, Jiayu Guo, Yujie Wang, Zhen Zhang, and Hansen Zhao. Judicial nested named entity recognition method with MRC framework. *International Journal of Cognitive Computing in Engineering* 4 (2023): 118–126.
- [3]. Ashutosh Modi, Prathamesh Kalamkar, Saurabh Karn, Aman Tiwari, Abhinav Joshi, Sai Kiran Tanikella, Shouvik Kumar Guha, Sachin Malhan, and Vivek Raghavan. SemEval-2023 Task 6: LegalEval - Understanding Legal Texts. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*: 2362–2374.
- [4]. H. Keshavarz *et al.*, "Named Entity Recognition in Long Documents: An End-to-end Case Study in the Legal Domain," *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022: 2024-2033.
- [5]. Hung Q. Ngo, Hien D. Nguyen, and Nhien-An Le-Khac. Ontology Knowledge Map Approach Towards Building Linked Data for Vietnamese Legal Applications. *Vietnam Journal of Computer Science* 11, 02 (2024): 323–342.