LexDrafter: Terminology Drafting for Legislative Documents using Retrieval Augmented Generation

Ashish Chouhan and Michael Gertz
Heidelberg University
Institute of Computer Science
Data Science Research Group
chouhan@informatik.uni-heidelberg.de

LREC-COLING 2024

Outline

- Introduction

- LexDrafter Framework

- Experiments

- Conclusion and Future Work

Chouhan and Gertz: "LexDrafter"

Introduction

Introduction

About EUR-Lex

What is EUR-Lex?

EUR-Lex is your online gateway to EU Law. It provides the official and most comprehensive access to EU legal documents. It is available in all of the EU's 24 official languages and is updated daily.

Who runs EUR-Lex?

EUR-Lex is run by the Publications Office of the European Union ☑️.

What can you find on EUR-Lex?

You can access the following EU legal documents:

- Treaties
- Legal acts from EU institutions
- Preparatory documents related to EU legislation

EUR-Lex platform: https://eur-lex.europa.eu/homepage.html

- **Around 24k** regulatory documents for the year 2022
- **15%** are legislative documents
  - They comprise regulations, directives, decisions, and opinions

Introduction

6.7.2011 — EN — Official Journal of the European Union — L 178/1

COMMISSION DELEGATED REGULATION (EU) No 626/2011

of 4 May 2011

supplementing Directive 2010/30/EU of the European Parliament and of the Council with regard to energy labelling of air conditioners

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Directive 2010/30/EU of 19 May 2010 of the European Parliament and of the Council on the indication by labelling and standard product information of the consumption of energy and other resources energy-related products ([1]), and in particular Article 10 thereof,

Whereas:

(1) Directive 2010/30/EU requires the Commission to adopt delegated acts as regards the labelling of energy-related products representing significant potential for energy savings and having a wide disparity in performance levels with equivalent functionality.

(2) Provisions for the energy labelling of air conditioners were established by Commission Directive 2002/31/EC of 22 March 2002 implementing Council Directive 92/75/EEC with regard to energy labelling of household air-conditioners ([2]). The implementing Directive establishes different labelling scales for air conditioners using different technologies and the determination of energy efficiency is based on full load operation only.

Commission Delegated Regulation (EU) 626/2011 with regard to energy labelling of air conditioners

Introduction

Article 2

Definitions

In addition to the definitions set out in Article 2 of Directive 2010/30/EU of the European Parliament and of the Council $(^5)$, the following definitions shall apply:

(1) '*air conditioner*' means a device capable of cooling or heating, or both, indoor air, using a vapour compression cycle driven by an electric compressor, including air conditioners that provide additional functionalities such as dehumidification, air-purification, ventilation or supplemental air-heating by means of electric resistance heating and appliances that may use water (either condensate water that is formed on the evaporator side or externally added water) for evaporation on the condensor, provided that the device is also able to function without the use of additional water, using air only;

(2) '*double duct air conditioner*' means an air conditioner in which, during cooling or heating, the condensor or evaporator intake air is introduced from the outdoor

Commission Delegated Regulation (EU) 626/2011 with regard to energy labelling of air conditioners

Introduction

- As per Joint Practical Guide (JPG) of European Parliament (*European Commission and Legal service, 2015*)
  - Terminology be consistent
  - Terms used in the act are defined together

Introduction

- As per Joint Practical Guide (JPG) of European Parliament (*European Commission and Legal service, 2015*)
    - Terminology be consistent
    - Terms used in the act are defined together

- On average, **20 definitions** are present in each legal act for the Energy domain
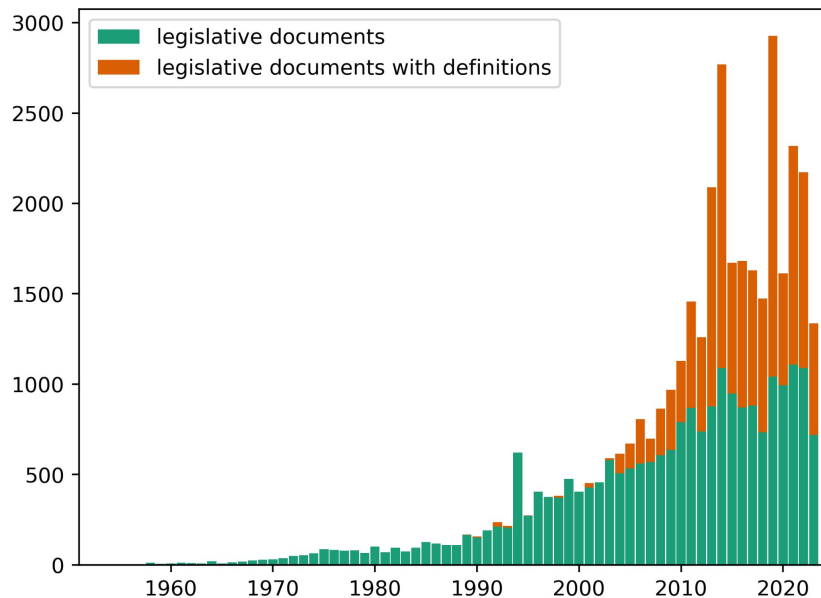
Chouhan and Gertz: "LexDrafter"

Contribution

- LexDrafter framework **assists in terminology drafting** (*Definitions articles*)
  - Extracts definitions from documents
  - Generates definitions based on contextual information

Contribution

- LexDrafter framework **assists in terminology drafting** (*Definitions articles*)
  - Extracts definitions from documents
  - Generates definitions based on contextual information

- Advantages of LexDrafter
  - Harmonizing legal definitions
  - Human errors are reduced by pointing to existing term definitions
  - Less time spent on ensuring term definitions consistency

LexDrafter Framework
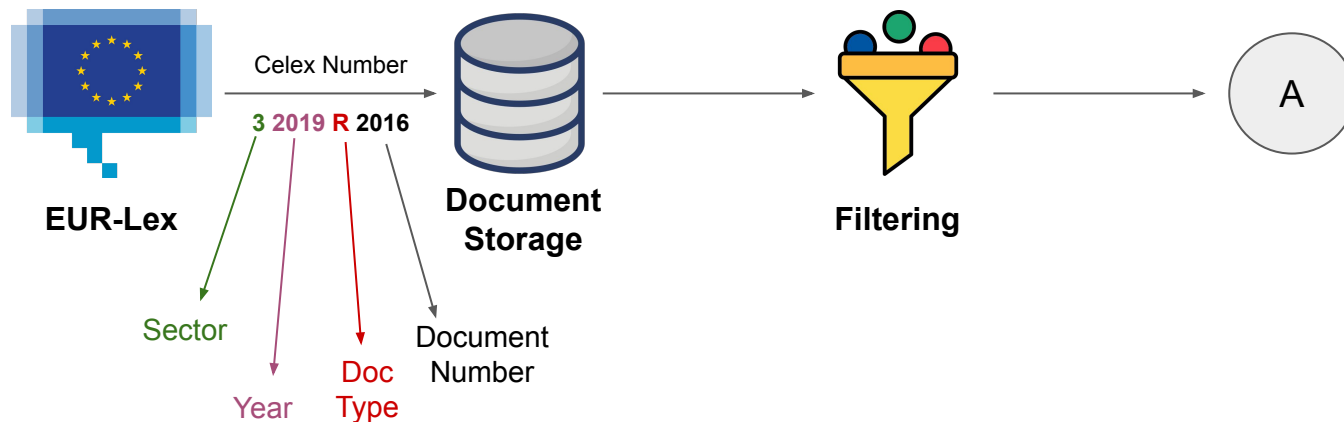
LexDrafter Framework

- Retrieval Augmented Generation (RAG)

Chouhan and Gertz: "LexDrafter"
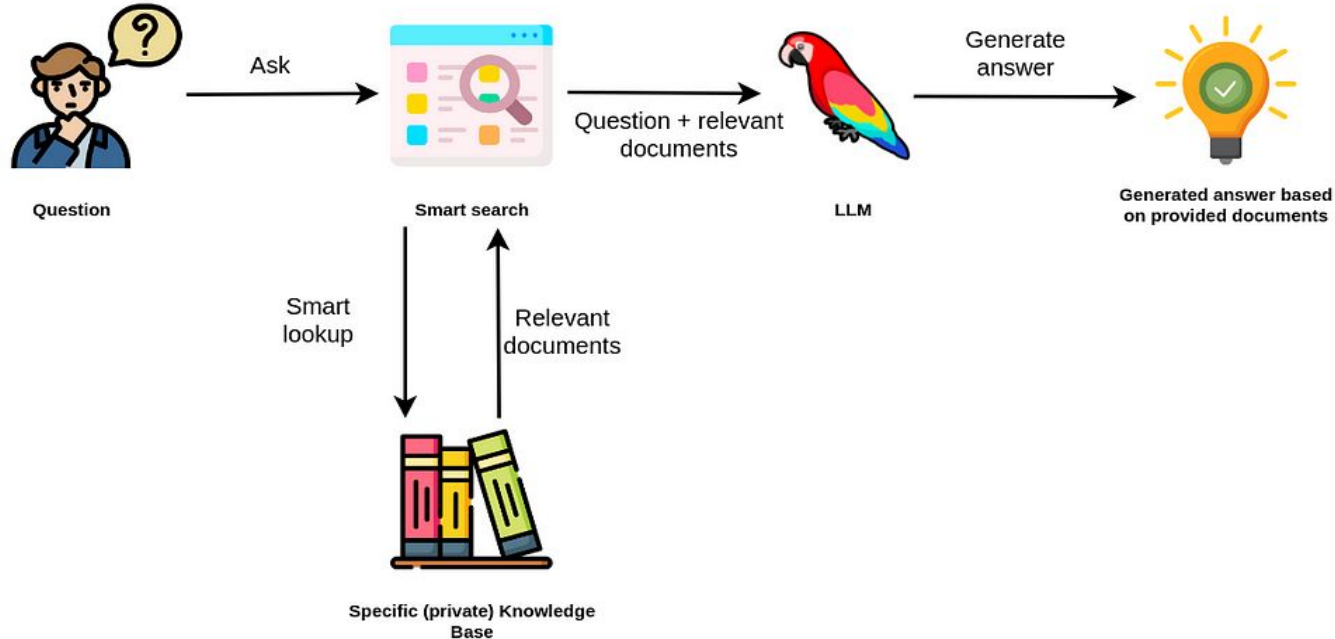
LexDrafter Framework

- Retrieval Augmented Generation (RAG)
- Two workflows
  - Workflow 1: building document and definition corpus
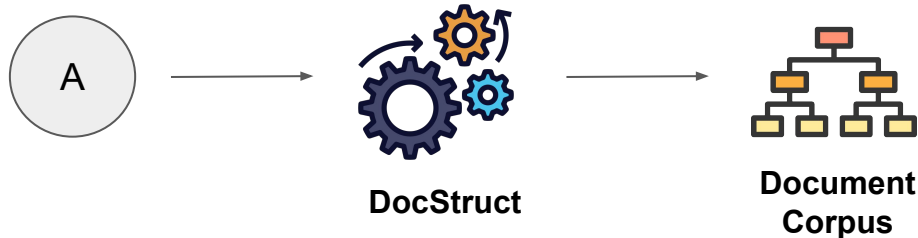 - Workflow 2: identifying and generating definitions

Chouhan and Gertz: "LexDrafter"

LexDrafter Framework: Workflow 1

- EUR-Lex legal acts (Energy domain)
- Data acquisition process (Aumiller et al., 2022)

Chouhan and Gertz: "LexDrafter"

LexDrafter Framework: Workflow 1/2

Commission Delegated Regulation (EU) 2019/2016
(Celex Number: 32019R2016)

LexDrafter Framework: Workflow 1/3

Chouhan and Gertz: "LexDrafter"
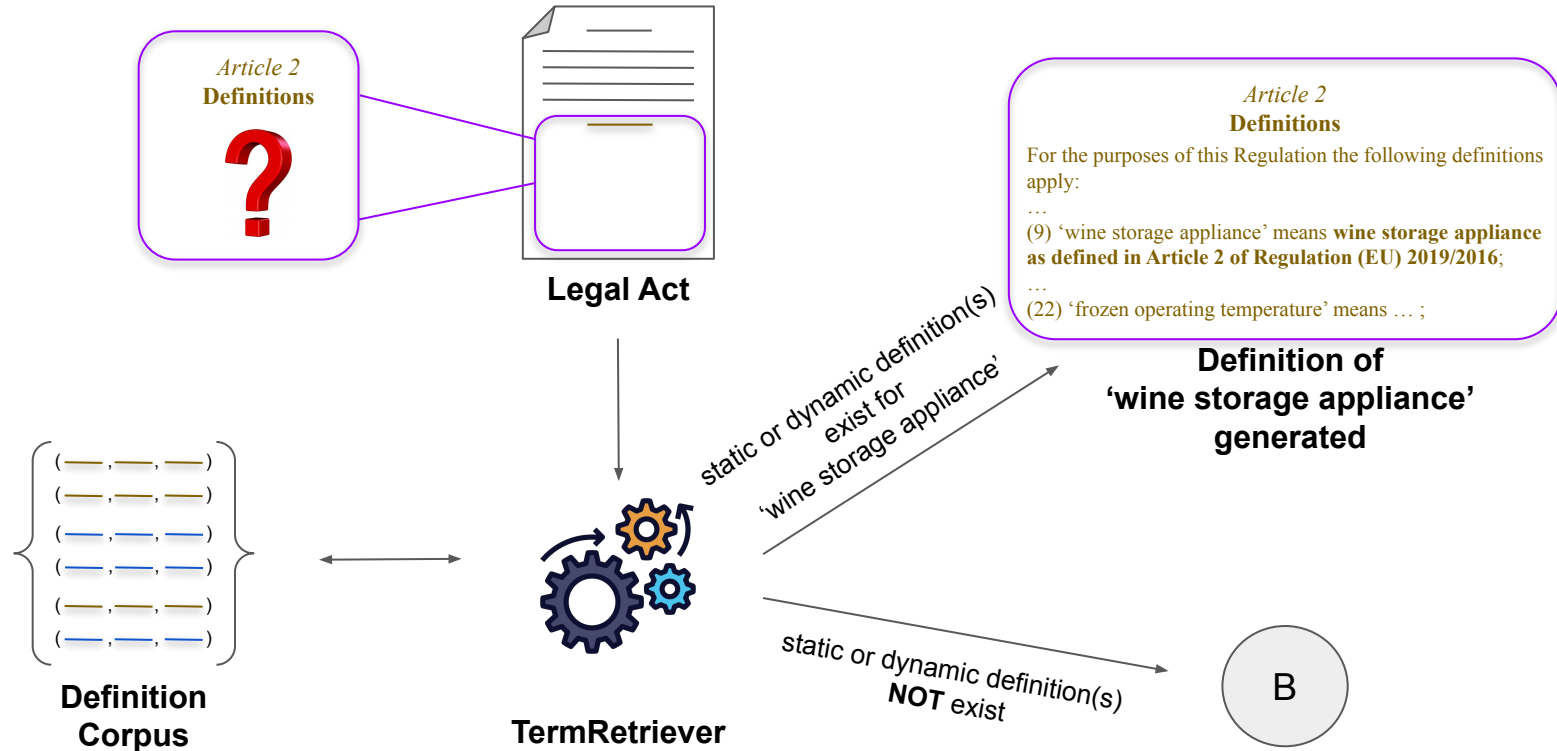
LexDrafter Framework: Workflow 1/4

EUR-Lex — Celex Number — 3 2019 R 2016 → **Document Storage** → **Filtering** → **DocStruct** → **Document Corpus**

Filtering → **DefExtract** → **CiteResolver** → **Definition Corpus**

Article 2
Definitions

For the purposes of this Regulation the following definitions apply:
(1) 'mains' or 'electric mains' means a … ;
…
(8) 'professional refrigerated storage cabinet' means an insulated refrigerating appliance ….. as defined in Commission Regulation (EU) 2015/1095;
…
(24) 'wine storage appliance' means a dedicated refrigerating …

Static definition

("mains", "means a …", "No reference")
…
("professional refrigerated storage cabinet", "means an insulated…", "32015R1095")
…
("wine storage appliance", "means a dedicate …", "No reference")

Dynamic definition

Chouhan and Gertz: "LexDrafter"

15

LexDrafter Framework: Workflow 2

Definition of
'wine storage appliance'
generated

Chouhan and Gertz: "LexDrafter"

LexDrafter Framework: Workflow 2/2

Definition Generator using RAG

Document Corpus

Article 2
Definitions

For the purposes of this Regulation the following definitions apply:

…

(9) 'wine storage appliance' means **…**;

…

(22) 'frozen operating temperature' means **the temperature at which a food storage or preparation appliance … ;**

Definition of 'frozen operating temperature' generated

Chouhan and Gertz: "LexDrafter"

Experiments

Experiments

551 legal acts:
Energy domain
(1949 to 2023)

108 legal acts:
HTML format and
Definition article

Definition Corpus
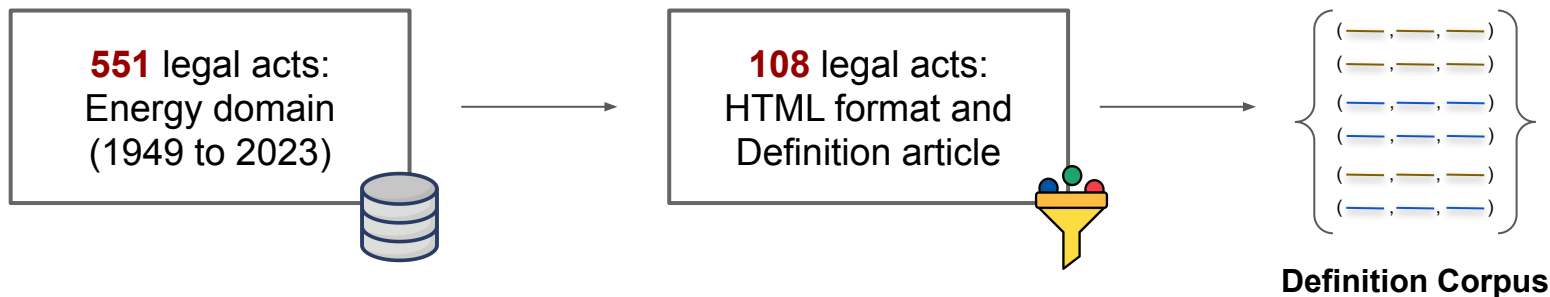
Objective:
- Evaluate for the definition generation task
  - Lexical similarity (BLEU)
  - Semantic similarity (BERTScore and BLEURT)

1007 legal terms
- 787: one definition
- 220: multiple definitions

1330 definitions

Experiments

Metrics		LLaMA-2 (7B)	Vicuna (7B)
BLEU	BLEU-1	0.25	0.28
	BLEU-2	0.13	**0.15**
	BLEU-3	0.07	**0.09**
	BLEU-4	0.04	0.05
BERTScore (F1-score)		**0.82**	**0.82**
BLEURT		**0.47**	**0.47**

Interpretation:

- **Lower BLEU scores** due to different styling and vocabulary

- Semantically similar generated definitions (**higher BERTScore**)

- **Lower BLEURT scores** due to non-legal domain corpus training

Chouhan and Gertz: "LexDrafter"

Conclusion and Future Work

Conclusion and Future Work

Conclusion:

- LexDrafter framework assists in terminology drafting for legal acts
  - Application of RAG
  - Harmonize legal definitions
  - Extracts definitions from documents
  - Generates definitions based on contextual information

Future Work:

- Automatically identify terms to be defined
- Investigate legal acts from all 20 domains (for example: Agriculture, Taxation, External relations, and many more)

Chouhan and Gertz: "LexDrafter"

Resources

European Commission and Legal service. 2015. Joint practical guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation. Publications Office.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rockt.schel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20), volume 793, pages 9459—-9474. ACM.

Dennis Aumiller, Ashish Chouhan, and Michael Gertz. 2022. EUR-Lex-Sum: A Multi- and Crosslingual Dataset for Long-form Summarization in the Legal Domain. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 7626–7639. Association for Computational Linguistics.

Catherine Sai, Anastasiya Damaratskaya, Karolin Winter, and Stefanie Rinderle-Ma. 2023. Identification and Visualization of Legal Definitions and Legal Term Relations. In Advances in Conceptual Modeling, pages 151–161. Springer Nature Switzerland.

Thank you for your attention!
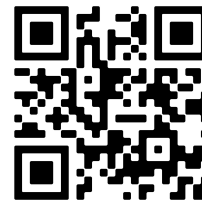
The experimental code is on Github:

https://github.com/achouhan93/LexDrafter

Any questions?

chouhan@informatik.uni-heidelberg.de