# LexDrafter: Terminology Drafting for Legislative Documents using Retrieval Augmented Generation

**Ashish Chouhan and Michael Gertz**
Institute of Computer Science, Heidelberg University
Im Neuenheimer Feld 205, Heidelberg, Germany
{chouhan, gertz}@informatik.uni-heidelberg.de

### Abstract

With the increase in legislative documents at the EU, the number of new terms and their definitions is increasing as well. As per the Joint Practical Guide of the European Parliament, the Council and the Commission, terms used in legal documents shall be consistent, and identical concepts shall be expressed without departing from their meaning in ordinary, legal, or technical language. Thus, while drafting a new legislative document, having a framework that provides insights about existing definitions and helps define new terms based on a document's context will support such harmonized legal definitions across different regulations and thus avoid ambiguities. In this paper, we present *LexDrafter*, a framework that assists in drafting *Definitions* articles for legislative documents using retrieval augmented generation (RAG) and existing term definitions present in different legislative documents. For this, definition elements are built by extracting definitions from existing documents. Using definition elements and RAG, a *Definitions* article can be suggested on demand for a legislative document that is being drafted. We demonstrate and evaluate the functionality of *LexDrafter* using a collection of EU documents from the energy domain. The code for *LexDrafter* framework is available at https://github.com/achouhan93/LexDrafter.

**Keywords:** legal, EU legislative documents, EUR-Lex, retrieval-augmented generation (RAG), large language models (LLMs), text generation

## 1. Introduction

Around 24k regulatory documents are accessible via the EUR-Lex platform[1] for the year 2022. 15% of these documents are legislative documents (also called legal acts), comprising *regulations*, *directives*, *decisions*, *recommendations*, and *opinions*.

Legal acts are considered a vital resource for protecting and safeguarding individuals and organizations. However, the domain-specific terminologies used in these documents are complex, thus complicating the process of reviewing texts and leading to different interpretations of the same content by different users (Sai et al., 2023). Similarly, drafting a legal document requires manual effort and domain knowledge to capture the relevant context and details for the document (Lam et al., 2023). As legal documents are usually highly structured and standardized, the manual steps involved in document drafting are time-consuming, resource-intensive, and prone to human error, resulting in a significant potential to automate the document drafting process (Achachlouei et al., 2021).

Gauci (2016) pointed out that during document drafting, there are situations where different legal documents assign different definitions to a legal term. In general, having harmonized legal definitions is an essential part of legal drafting practice for precise and effective communication (Amaludin et al., 2021). As per the Joint Practical Guide (JPG) of the European Parliament, the Council and the Commission, for people involved in drafting European Union legislation (European Commission and Legal service, 2015), guidelines 6 and 14 provide information about drafting of the terminologies mentioned in a document. Considering these guidelines and the steady increase in the number of documents containing definitions, as shown in Figure 1, a framework that extracts or "generates" definition elements from existing documents might serve as an important supporting tool in drafting legal documents and definitions sections in particular. This

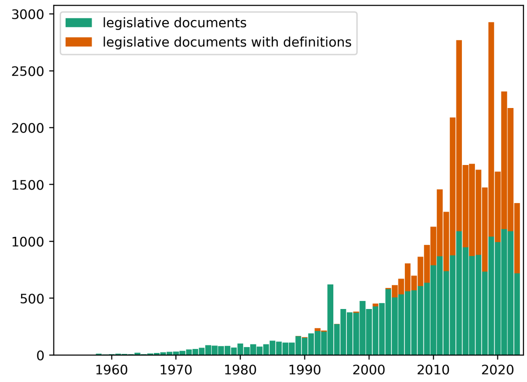

Figure 1: Number of legal documents (with definitions) per year.

[1] https://eur-lex.europa.eu/statistics/2022/eu-law-statistics.html?locale=en, accessed on 11th September 2023

*LREC-COLING 2024*, pages 10448–10458
20-25 May, 2024.

need is further justified by looking at some statistics. For example, on average, 20 definitions are present in legal acts for the domain *Energy*, which is one of 20 domains[2]. But there are also outliers, e.g., the document on "establishing a guideline on electricity transmission system…" (Celex ID 32017R1485) has 159 definitions.

As technology advances, AI-powered tools can assist legal professionals and expedite the writing process, e.g., LegalSifter[3], Kira Systems[4], or LawGeex[5] are AI-powered tools that assist in drafting contracts, identifying and extracting information from contracts, and perform contract analysis. With the advent of Chatbots, legal services platform such as LawDroid[6] or Law ChatGPT[7] use templates and questionnaires to assist drafting contracts, wills, or agreements. Another product, Lexis+ AI[8] also uses Generative Pre-Trained Transformer (GPT) for generating drafts of documents such as demand letters or client emails. But to the best of our knowledge, there is no tool that helps in drafting definitions and terminology in the context of legal documents.

The paper's main contribution is the *LexDrafter* framework that assists in terminology drafting, i.e., the drafting of *Definitions* articles in legal acts. *LexDrafter* is based on retrieval augmented generation (RAG), where for terms to be defined, relevant text fragments are retrieved and a definition is generated using large language models (LLMs). The advantages of *LexDrafter* are three-fold: first, it aims at harmonizing legal definitions across different legal acts by automating the drafting of *Definitions* articles. Second, human errors are reduced by providing existing term definitions (e.g., for citation) or generating new term definitions, and third, time-consumption is reduced for drafting documents, because users do not have to manually browse through large corpora to ensure consistency of term definitions.

The remainder of this paper is structured as follows. In Section 2, work related to text generation in the legal domain is discussed, followed by a description of the conceptual components underlying the framework in Section 3. Section 4 details the two workflows of the *LexDrafter* framework, (1) creation of document and definition corpus, and (2) identification and generation of definitions. In Section 5, we present the experiments on generating definitions, and finally, Section 6 concludes the paper and gives a brief outline of our ongoing work.

## 2. Related Work

Our work is mostly related to text generation in the legal domain, more specifically EU legislative documents, and there the drafting of *Definitions* articles in respective legislative documents in particular.

### 2.1. Extraction of Legal Definitions

Using text segmentation and Part-of-Speech (POS) tagging, Ferneda et al. (2012) and Hwang et al. (2018) investigated Brazilian Portuguese texts and Chinese law texts, respectively, to extract legal terms and to build a law ontology. However, manual verification is still required due to a large number of false positives. Rule-based approaches focusing on German laws and cases are used by Waltl et al. (2017) for extracting legal definitions and relevant semantic information, such as the year of a dispute. The authors created a taxonomy differentiating between legal definitions, context-extending definitions, and interpretation of legal terms, and further state that legal definition extraction is a difficult task due to additional work required in the field of legal theory.

Addressing the challenge of manually analyzing European regulatory documents, Sai et al. (2023) focus on semi-automating the analysis by extracting legal definitions and their semantic relations. The author investigates the structure of legal acts and focuses on a single article, i.e., *Definitions* article that specifies legal terms. As punctuation plays an important role in understanding the syntax of legal definitions, using such punctuations, definition terms are extracted, and POS tags are used to extract the explanation of the terms.

### 2.2. Document Drafting

Document creation, document analysis, and document management are the three main branches of legal Document Automation. Different architectures for document generation are reviewed by Achachlouei et al. (2021). For example, Hotdocs[9], DocuPlanner (Branting et al., 1997), Virtual Court

[2] https://eur-lex.europa.eu/browse/directories/legislation.html, accessed on 1st March 2024

[3] https://www.legalsifter.com, accessed on 24th August 2023

[4] https://www.kirasystems.com, accessed on 24th August 2023

[5] https://www.lawgeex.com, accessed on 24th August 2023

[6] https://lawdroid.com/, accessed on 24th August 2023

[7] https://lawchatgpt.com/, accessed on 24th August 2023

[8] https://www.lexisnexis.com/en-us/products/lexis-plus-ai.page, accessed on 25th August 2023

[9] https://www.hotdocs.com, accessed on 17th August 2023

Action (Barton and McKellar, 1998), ToXgene (Barbosa et al., 2002), and FreeMarker[10] are template-based tools that allow converting regularly used documents into templates that assist in automated document creation.

The work by Palmirani and Vitali (2011) presents essential elements of the Akoma Ntoso XML standard. Consuming the Akoma Ntoso format, Marković and Gostojić (2022) produced legal documents using a set of legal rules. Furthermore, Palmirani and Governatori (2018) used a combination of Akoma Ntoso for marking up legal texts and legal concepts for checking GDPR-compliant public cloud computing services. Supporting the Document Product Line (DPL) methodology, Gómez et al. (2014) introduced the DPLFW framework for multi-user, variable content, and reuse-based document generation. Marković and Gostojić (2022) propose a knowledge-based document assembly by explicitly formulating the legal norms prescribing the content and the form of service contracts.

In addition to the traditional and knowledge-based document assembly methods discussed, one notable application of RAG proposed by Lewis et al. (2020) is document drafting. Markey et al. (2024) demonstrated the application of RAG in the medical domain to automatically generate documents essential for clinical trials, and Ranade and Joshi (2024) propose FABULA, a framework that employs RAG to facilitate the report generation process regarding a news event.

Compared to traditional legal NLP tasks, such as classification, information retrieval, and information extraction, currently only a few research is focusing on text generation, i.e., automated drafting of legal documents (Katz et al., 2023). Contract clause generation is a very prominent text generation task, as discussed by Aggarwal et al. (2021) and Joshi et al. (2022). Also, Lam et al. (2023) propose combining traditional and generative AI techniques to enhance contract drafting. In their work, the authors investigate the performance of LLMs in contract clause drafting and propose an approach to evaluate the generated clauses by retrieving similar clauses using sentence transformers and then performing content similarity analysis.

Unlike the above approaches, our work focuses on the task of terminology drafting, i.e., *Definitions* articles for EUR-Lex legislative documents. For this, EUR-Lex legal acts are extracted from the EUR-Lex platform, preprocessed, and stored in an Information Retrieval (IR) system. Similar to the approach by Sai et al. (2023), existing definitions are extracted from these documents. However, apart from definition extraction itself, citation resolution is additionally carried out for definitions with external references, eventually resulting in so-called definition elements. The stored documents together with definition elements then build the basis for drafting, identifying existing definitions or generating new definitions for selected terms using RAG.

## 3. Document Model and Definition Elements

Basis for our *LexDrafter* framework is a corpus of EUR-Lex legal documents. These documents, typically represented in the form of HTML documents, are extracted from the EU platform and preprocessed, following a particular document model. Furthermore, (legal) term definitions in the documents as well as citations related to legal terms are extracted and mapped to definition elements. Both these conceptual components underlying our framework are briefly described in the following.

**Document Model.** We assume a corpus $D = \{d_1, \dots, d_n\}$ of (legal) documents. Each document $d_i \in D$ is composed of a sequence of sections $S_i = [s_{i1}, \dots, s_{ik}]$, with a section providing the coarsest level of (text) granularity. Each section can be decomposed further, e.g., into paragraphs (based on newline elements) or even sentences. Given the natural hierarchical structure of (HTML) documents, the order of elements and their position, respectively, are always preserved and are a property of sections and more fine-granular document components.

**Definition Elements.** For a document $d_i \in D$ and a document section $s_{ij}$, a fragment is simply a sequence of tokens in that section, in most cases a sentence or sequence of sentences in that section. We are in particular interested in fragments that define a legal term, and the process of identifying such fragments is detailed in the next section. Eventually, a set $F_{def} = \{f_1, \dots, f_l\}$ of definition elements is to be determined for a document collection $D$ such that each definition element $f_i \in F_{def}$ is a tuple $f_i = \langle id_i, t_i, e_i, r_i \rangle$ with

- $id_i$ is a unique identifier of the definition element,
- $t_i$ being a term, which is a single word or sequence of words (phrase), that is defined (e.g., “energy from renewable sources”),
- $e_i$ an explanation of the term (i.e., the definition), and
- $r_i$ is a (possibly empty) list of references to definition elements (or rather their $ids$).

If an explanation $e_i$ of a term $t_i$ provides a unique explanation, then $r_i$ is empty, otherwise, $r_i$ is a list of respective definition elements ($ids$).

[10] https://freemarker.apache.org, accessed on 17th August 2023

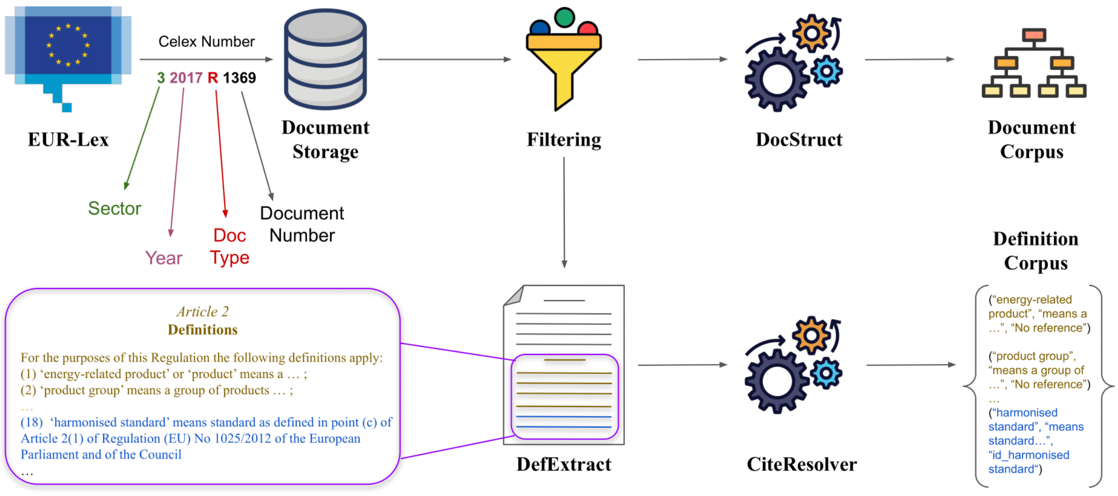
Celex Number
3 2017 R 1369
EUR-Lex
Document Storage
Filtering
DocStruct
Document Corpus
Sector
Year
Doc Type
Document Number
Definition Corpus
Article 2
Definitions
For the purposes of this Regulation the following definitions apply:
(1) 'energy-related product' or 'product' means a … ;
(2) 'product group' means a group of products … ;
…
(18) 'harmonised standard' means standard as defined in point (c) of Article 2(1) of Regulation (EU) No 1025/2012 of the European Parliament and of the Council
…
DefExtract
CiteResolver
("energy-related product", "means a …", "No reference")
("product group", "means a group of …", "No reference")
…
("harmonised standard", "means standard…", "id_harmonised standard")

Figure 2: Overview of workflow to build the document and definition corpus, with an example document on "setting a framework for energy labelling…" (Celex ID 32017R1369). Legal acts based on Celex ID are extracted from the EUR-Lex platform and filtered to consider only legal acts in HTML format. The *DocStruct* component extracts all information from legal acts and stores it in the document corpus; the *DefExtract* component identifies and extracts definitions from a document. The *CiteResolver* component resolves citations in explanations that have references.

## 4. LexDrafter Framework

*LexDrafter* functions help users when drafting a legal act, in particular when drafting document sections has been completed, but the section with terminology definitions (the *Definitions* article) is missing. The *fragments* that include a given *term* in the drafted sections are the key components required by *LexDrafter*, as such fragments provide contextual information for the definition of a term using our RAG approach. Note that automatically identifying terms that need to be defined requires an understanding of legal theory. Therefore, in our work, terms for which definitions need to be determined are selected by the user.

In the following, we detail the two workflows our *LexDrafter* framework realizes. The first workflow (see also Figure 2) takes EUR-Lex legal acts (here for the *Energy* domain) as input, preprocesses them, and stores them in an IR system. The second workflow (see Figure 3) takes a term selected by the user and either determines existing definitions or generates a definition for that term. The data acquisition process in this work is similar to the one employed by Aumiller et al. (2022), where a particular legal act web page is crawled to store the text and metadata in an OpenSearch instance.

### 4.1. Creating Document and Definition Corpus

As illustrated in Figure 2, from the EUR-Lex website, two data sources are built, (1) the document corpus, which primarily consists of the (properly fragmented) texts of legal acts, including metadata, and (2) term definitions extracted from respective parts of these documents.

**Building Document Corpus.** Chalkidis et al. (2019) detail the structure of legal documents extracted from the EUR-Lex platform, stating that each legal act is structured into four major zones: *header*, comprising title and name of a legal body enforcing the legal act; *recitals*, consisting of references to legal background of decisions; *main body*, organized as a sequence of articles; and *attachments*, which include appendices and annexes.

In our framework, according to our document model, each header, recital, attachment, and article present in the main body is referred to as a *section*. The *DocStruct* component shown in Figure 2 builds the document corpus by extracting respective components from retrieved legal acts, preprocessing and storing them in the IR system.

**Building Definition Corpus.** As per JPG (European Commission and Legal service, 2015), term consistency must be maintained, and the terms should be defined in a single article in a legal act, called "*Definitions*" to avoid wrong interpretations. Therefore, building a definition corpus starts with identifying the *Definitions* article in a document, followed by extracting the definitions, similar to the approach proposed by Sai et al. (2023).

Fragments defining a legal term are identified by the *DefExtract* component in Figure 2 using patterns as per "*Wording Laying Down Definitions*" (Section C.7 in JPG):

- '…' means …; *[static definition]*

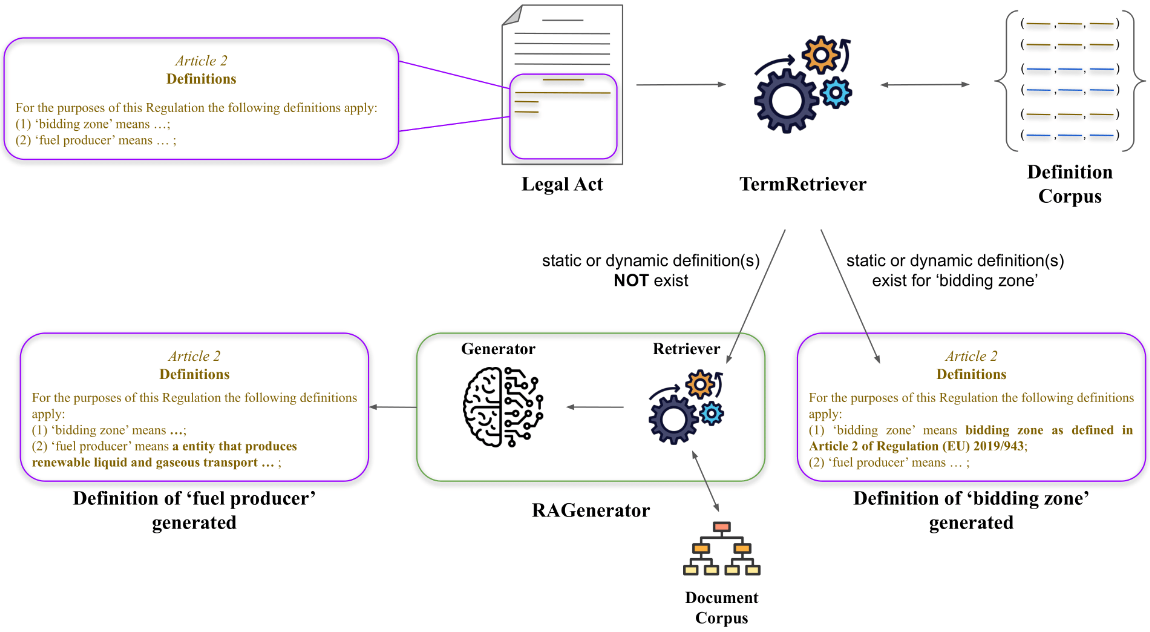
Article 2
Definitions
For the purposes of this Regulation the following definitions apply:
(1) 'bidding zone' means …;
(2) 'fuel producer' means … ;
Legal Act
TermRetriever
Definition Corpus
static or dynamic definition(s) NOT exist
static or dynamic definition(s) exist for 'bidding zone'
Article 2
Definitions
For the purposes of this Regulation the following definitions apply:
(1) 'bidding zone' means …;
(2) 'fuel producer' means a entity that produces renewable liquid and gaseous transport … ;
Definition of 'fuel producer' generated
Generator
Retriever
RAGenerator
Document Corpus
Article 2
Definitions
For the purposes of this Regulation the following definitions apply:
(1) 'bidding zone' means bidding zone as defined in Article 2 of Regulation (EU) 2019/943;
(2) 'fuel producer' means … ;
Definition of 'bidding zone' generated

Figure 3: Overview of the Definition Generation workflow with an example document on "…production of renewable liquid and gaseous transport fuels…" (Celex ID 32023R1184). Terms to be defined are passed to the *TermRetriever* component to retrieve matching definition elements. '*bidding zone*' is already defined in another legal act; therefore, the definition simply cites this legal act. For '*fuel producer*', a definition does not exist and needs to be generated by the *RAG* component. The *Retriever* subcomponent retrieves fragments and passes these fragments along with the term to be defined to the generator to generate a definition for that term.

- '…' means …as defined in [reference to the static definition]; *[dynamic definition]*

Based on this, the legal term within apostrophes and its explanation are extracted. For example, consider the definition of the term '*energy from renewable sources*' or '*renewable energy*' present in the document on "common rules for the internal market for electricity…" (Celex ID 32019L0944):

"*'energy from renewable sources'* or *'renewable energy'* means energy from renewable non-fossil sources, namely wind, solar (solar thermal and solar photovoltaic) and geothermal energy, ambient energy, tide, wave and other ocean energy, hydropower, biomass, landfill gas, sewage treatment plant gas, and biogas;" (Celex ID 32019L0944, Article 2, point 31)

The definition corpus comprises two definition elements for the terms '*energy from renewable sources*' and '*renewable energy*', each with its own id. Both have the same explanation $e$, and as the explanation is unique, the list $r$ of references is empty for both terms.

If the explanation for a term cites other definitions (elements), then the *CiteResolver* component in Figure 2 resolves this citation and populates the references $r$ for that term with the ids of these definition elements. For example, consider the definition of the terms '*energy from renewable sources*' and '*renewable energy*', respectively, in the document on "the internal market for electricity…" (Celex ID 32019R0943):

"*'energy from renewable sources'* or *'renewable energy'* means energy from renewable sources as defined in point (31) of Article 2 of Directive (EU) 2019/944;" (Celex ID 32019R0943, Article 2, point 50)

For this example, the *DefExtract* component stores two definition elements in the definition corpus: for the term '*energy from renewable sources*' and for '*renewable energy*'. Both have the explanation "*means energy from renewable sources as defined in point (31) of Article 2 of Directive (EU) 2019/944*". Because the explanation cites another definition element (here "*point (31) of Article 2 of Directive (EU) 2019/944*"), the *CiteResolver* component resolves this citation, determines the identifier of the respective definition element that provides the (static) explanation of the terms '*energy from renewable sources*' and '*renewable energy*', respectively, and stores the identifier in the reference list of both terms.

## 4.2. Identifying and Generating Definitions

The task of the second workflow shown in Figure 3 is to identify existing definitions and generate new definitions for terms selected by the user. Existing definitions can easily be identified and retrieved from the definition corpus described above. New definitions are generated using a retrieval augmented generation (RAG) approach. RAG combines pre-trained parametric and non-parametric memory for text generation (Lewis et al., 2020). In the context of *LexDrafter*, the parametric memory is provided by a large language model (LLM), and the non-parametric memory is made up of the fragments and their indexes in the document corpus, respectively. Below we describe (1) how existing term definitions are searched and retrieved, and (2) how a definition for a term is determined using our RAG approach.

**TermRetriever Component.** If the user would like to know whether for a term in a recently drafted text a definition exists, she/he selects the term, which is then passed to the *TermRetriever* component. The task of this component is to determine if there is a (static) definition for that term, based on the definitions in the legal acts that have been imported into *LexDrafter*. There are three cases: (1) if no definition can be found, the system can generate one (see below); (2) if a single definition is found, its information, e.g., in the form of the definition element is shown to the user. In this case, this information can be used to add a citation to the term, if deemed necessary; (3) if there are multiple definitions, those with the same or similar eurovoc descriptors as the document being drafted are ranked higher when shown to the user. In this case, the user eventually decides what definition to cite in the text for that term, if at all.

For example, to draft a definition of the term '*bidding zone*' for the document on "...production of renewable liquid and gaseous transport fuels..." (Celex ID 32023R1184), the *TermRetriever* would retrieve the definition element for the term '*bidding zone*' that is already defined in the document on "the internal market for electricity..." (Celex ID 32019R0943). Thus, the definition of that term could simply be generated as "'*bidding zone*' means bidding zone as defined in Article 2 of Regulation (EU) 2019/943", without any RAG approach.

**Definition Generation Using RAG.** If the TermRetriever component cannot find a definition for a selected term, the user can request that a definition for the term is generated. The selected term and surrounding text fragments are the two essential parts to generate a term definition. The RAG approach proposed by Lewis et al. (2020) and adopted to our framework has two major components, a retriever and a generator. It works as follows: initially, embeddings for the documents (or fragments) are computed and stored, thus building the corpus on which RAG operates. Given a query term and the sections (and fragments) of the document being drafted, the top-$k$ fragments relevant to the query term are determined.

*LexDrafter* also comprises a retriever and a generator. It is known that a dense embedding retriever often introduces noise by considering document fragments that have terms "closer" to the queried term(s). Thus, to reduce such noise in retrieved fragments, we employ a lexical search retriever instead of a dense embedding to retrieve the top-$k$ fragments from the drafted document sections.

As maximum contextual information about a term must be provided to an LLM to generate a term definition, the retrieved fragments are scored based on the term frequency in each fragment, and top-$k$ fragments having the highest term frequency are used. For example, to draft the definition of the term '*fuel producer*' for the document on "...production of renewable liquid and gaseous transport fuels..." (Celex ID 32023R1184), from drafted document sections, the retriever determines those fragments with the highest '*fuel producer*' term frequency.

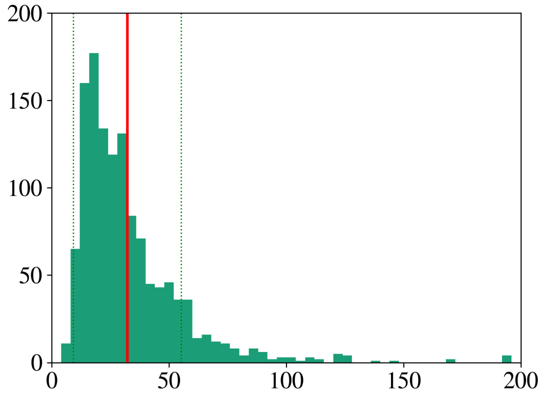

Figure 4: Histogram of the distribution of definition word lengths. Vertical lines show mean length (continuous red) and standard deviation (dotted green lines).

The retrieved fragments and the term to be defined based on the fragments are then passed as a prompt to the generator. The prompt template below is used to generate a definition for the respective term:

```
Act as a Lawyer drafting European
Legislative documents to be pub-
lished on the Eur-Lex website.

Define the term: {term}, based on
the sentences provided between the
triple dashes where new line char-
acters split different sentences.
−−−{sentences}−−−

Provide a clear and concise defini-
tion strictly within 25 to 45 words
that accurately conveys the meaning
within the context of the sentences.

Give your output in the following
```

```
JSON format:
{{
"term":  "{term}"
"definition":  "“‘output text”’"
}}
ONLY return the JSON with the keys:
[term, definition], do not add ANY-
THING, NO INTERPRETATION!
```

In the prompt, we specify a range of 25 to 45 words, considering the average number of tokens in the definitions in our corpus, (approximately 32 words), as shown in Figure 4.

# 5. Experiments

For the creation of our document and definition corpus, 551 legal acts from the *Energy* domain between 1949 and 2023 were crawled[11]. The *DocStruct* component shown in Figure 2 builds the document corpus by extracting respective components from 539 of these legal acts, as only these are in HTML format, preprocesses, and stores them in the IR system. In order to create the definition corpus, those 108 legal acts that contain a *Definitions* article were identified, and a total of 1330 fragments defining 1007 legal terms were detected and stored by the *DefExtract* component as definition elements.

Among the 1007 legal terms, for 787 legal terms, only one definition element exists with static or dynamic definitions, whereas 220 legal terms have multiple definition elements comprising either static or dynamic definitions distributed among 543 definition elements. For example, ‘*equivalent model*’ is often defined as a legal term with 9 definition elements, and surprisingly, all 9 definition elements are static definitions. In order to demonstrate the functionality of *LexDrafter*, the definition elements with static or dynamic definitions for 1007 legal terms are considered.

## 5.1. Large Language Models (LLMs)

For evaluation, we use Google Colab[12], 16GB RAM with V100 GPU, to interact with Vicuna (Zheng et al., 2023) and LLaMA-2 (Touvron et al., 2023) models. The Hugging Face[13] ‘text-generation’ pipeline is used to interact with the Vicuna-7b-v1.5[14] and LLaMA-2-7b[15] as a memory-efficient 16-bit quantized model. The important model configuration parameters are as follows: temperature is 0.2, top_$k$ is 20, top_$p$ is 0.6, repetition penalty is 1.2, and context length is 4096. For all other model configuration parameters, we use the default value provided by the models.

## 5.2. Results

Human-centric metrics, untrained automatic metrics, and machine-learned metrics are the three groups to evaluate text generated with an LLM (Celikyilmaz et al., 2020). The *LexDrafter* framework is evaluated for the definition generation task, where for a legal term, we investigate the quality of a generated definition with respect to a human-written definition present in the legal acts for the same legal term, i.e., we use ground truth definitions. Therefore, instead of human-centric metrics, untrained automatic metric, i.e., bilingual evaluation understudy (BLEU) (Papineni et al., 2002) and machine-learned metrics, i.e., BERTScore (Zhang et al., 2020) and bilingual evaluation understudy with representations from transformers (BLEURT) (Sellam et al., 2020) are considered to evaluate the quality of generated definitions.

BLEU measures $n$-gram word overlap of a generated definition with a ground truth definition, whereas BERTScore measures their semantic similarity. BLURT score captures nuances of human language by comparing a generated definition with a ground truth definition on features comprising $n$-gram overlap, semantic similarity, and fluency. Maximizing the BLEURT score signifies fluent, coherent, and semantically accurate generated definitions. In our experiments, open-source LLMs, i.e., Vicuna and LLaMA-2, are used for generating definitions and ‘distilbert-base-uncased’ model in BERTScore and ‘BLEURT-20’[16] checkpoint in BLEURT is used for evaluation.

Table 1 summarizes BLEU, BERTScore, and BLEURT score of generated term definitions using the prompt template from Section 4.2. The results show that BLEU scores for $n$-gram word overlap where $n$ ranges from 1 to 4 are lower due to different styling and vocabulary used in generated definitions by an LLM when compared to ground truth

[11] https://eur-lex.europa.eu/search.html?type=named&name=browse-by:legislation-in-force&CC_1_CODED=12, accessed on 22nd March 2024

[12] https://colab.google/, accessed on 24th September 2023

[13] https://huggingface.co/, accessed on 24th September 2023

[14] https://huggingface.co/lmsys/vicuna-7b-v1.5, accessed on 24th September 2023

[15] https://huggingface.co/meta-llama/Llama-2-7b-chat-hf, accessed on 24th September 2023

[16] https://github.com/google-research/bleurt/blob/master/checkpoints.md, accessed on 20th October 2023

| Metrics | | LLaMA-2 | Vicuna |
|---|---|---|---|
| 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 | Precision | 0.83 | 0.83 |
| | Recall | 0.81 | 0.81 |
| | F1-score | 0.82 | 0.82 |
| BLEURT | | 0.47 | 0.47 |

Table 1: Evaluation of quality of generated definitions. We report $n$-gram BLEU score where $n$ ranges from 1 to 4 to measure the word overlap, BERTScore to measure semantic similarity, and BLEURT measures features comprising $n$-gram, semantic similarity, and fluency. The best results are highlighted in boldface.

definitions. On the other hand, BERTScores are higher, demonstrating that generated definitions are semantically similar and have content coverage concerning the term definition present in legal acts. BLEURT scores for both LLMs differ marginally; however, the lower scores are due to training of BLEURT on non-legal domain corpus comprising news articles, blog posts, and wikipedia articles. Thus, a solution to improve the BLEURT score for better insights into the fluency, coherence, and semantics of generated definitions is fine-tuning BLEURT with legal domain corpus. As fine-tuning BLEURT on legal corpus is our ongoing work, fine-tuned BLEURT scores are not presented.

For evaluating the performance of the two LLMs, LLaMA-2 and Vicuna, on the task of generating definitions, our results showed that Vicuna marginally outperformed LLaMA-2 on three metrics: BLEU, BERTScore, and BLEURT. This indicates that Vicuna is better at generating definitions that are fluent, coherent, and semantically similar to the ground truth definitions. A possible explanation for the marginally better performance of Vicuna is the training of Vicuna on a more diverse dataset compared to LLaMA-2, allowing Vicuna to learn a richer representation of the language, enabling it to generate more fluent and coherent definitions. Some examples of generated definitions are shown in Appendix A.

## 6. Conclusions and Ongoing Work

In this paper, we presented *LexDrafter*, a framework that assists in drafting *Definitions* articles for legislative documents using RAG and existing term definitions present in different legislative documents. This work aims to harmonize legal definitions across different legal acts by identifying existing definitions or generating new definition for a term selected by a user, as there are outliers. For example, definitions present in the document on "...protected geographical indications..." (Celex ID 32021R0244) do not follow the pattern of having the legal term within apostrophes and document on "...animal health requirements for movements..." (Celex ID 32020R0688) defines a term 'status free from "disease"' with "disease" as a placeholder.

Our framework realizes two workflows. The first workflow takes EUR-Lex legal acts as input and creates a document corpus and definition corpus. The definition corpus is comprised of extracted term definitions, and the document corpus consists of the (properly fragmented) texts of legal acts. The second workflow takes a term selected by the user as input and either identifies existing definitions or generates a definition for that term using RAG. The RAG approach generates a definition by retrieving fragments using a lexicographic search retriever and passing it to LLMs (Vicuna-7b and LLaMA-7b) as a prompt.

Finally, the functionality of *LexDrafter* is evaluated for a definition generation task with BLEU and BERTScore, where a generated definition is compared to a ground truth definition of a selected term. Table 1 shows high BERTScore and lower 1 to 4-gram BLEU score, signifying that definitions generated using an LLM show low word overlap but high semantic similarity.

In our ongoing work, we are working to automatically identify terms that need to be defined in a drafted document and consider legal acts from all domains on the EUR-Lex platform to harmonize legal definitions by automating the drafting of *Definitions* articles for cross-domain legal acts.

## Broader Impact & Ethical Issues

As the usage of EUR-Lex documents is licensed under CC BY 4.0[17] and EUR-Lex documents are used in this work, we do not see any ethical concerns in a release of the code for the *LexDrafter* framework. To our knowledge, legal acts are drafted using the JPG guidelines provided by the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation, have undergone review within instances of the European Union, leading to no clear concerns in data quality, especially with respect to potential privacy violations or harmful text content.

## 7. Bibliographical References

Mohammad Ahmadi Achachlouei, Omkar Patil, Tarun Joshi, and Vijayan N. Nair. 2021. Doc-

[17] https://eur-lex.europa.eu/content/legal-notice/legal-notice.html, accessed on 20th October 2023

ument Automation Architectures and Technologies: A Survey. *CoRR*, abs/2109.11603.

Vinay Aggarwal, Aparna Garimella, Balaji Vasan Srinivasan, Anandhavelu N, and Rajiv Jain. 2021. ClauseRec: A Clause Recommendation Framework for AI-aided Contract Authoring. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8770–8776. Association for Computational Linguistics.

Bakhtiar Amaludin, Fitria Ratna Wardika, Putu Jasprayana Mudana Putra, and I Gede Yudi Paramartha. 2021. Analyze the Usage of Legal Definitions in Indonesian Regulation Using Text Mining Case Study: Treasury and Budget Law. In *Proceedings of the 34th Conference on Legal Knowledge and Information Systems (JURIX'21)*, volume 346, pages 107–112. IOS Press.

Dennis Aumiller, Ashish Chouhan, and Michael Gertz. 2022. EUR-Lex-Sum: A Multi- and Cross-lingual 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.

Denilson Barbosa, Alberto O. Mendelzon, John Keenleyside, and Kelly A. Lyons. 2002. ToXgene: a template-based data generator for XML. In *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, page 616. ACM.

Karen Barton and Patricia McKellar. 1998. The virtual court action: Procedural facilitation in law. *ALT-J: Research in Learning Technology*, 6(1):87–94.

Karl Branting, James C. Lester, and Charles B. Callaway. 1997. Automated Drafting of Self-Explaining Documents. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law (ICAIL'97)*, pages 72–81. ACM.

Asli Celikyilmaz, Elizabeth Clark, and Jianfeng Gao. 2020. Evaluation of Text Generation: A Survey. *CoRR*, abs/2006.14799.

Ilias Chalkidis, Emmanouil Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2019. Extreme Multi-Label Legal Text Classification: A Case Study in EU Legislation. In *Proceedings of the Natural Legal Language Processing Workshop 2019*, pages 78–87. Association for Computational Linguistics.

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.

Edilson Ferneda, Hércules Antônio do Prado, Augusto Herrmann Batista, and Marcello Sandi Pinheiro. 2012. Extracting Definitions from Brazilian Legal Texts. In *Proceedings of the 12th International Conference on Computational Science and Its Applications (ICCSA'12)*, volume 7335, pages 631–646. Springer.

Gotthard Mark Gauci. 2016. Is it a vessel, a ship or a boat, is it just a craft, or is it merely a contrivance. *Journal of Maritime Law & Commerce*, 47(4):479–499.

Abel Gómez, M. Carmen Penadés, José H. Canós, Marcos R. S. Borges, and Manuel Llavador. 2014. A framework for variable content document generation with multiple actors. *Information and Software Technology*, 56(9):1101–1121.

Ren-Hung Hwang, Yu-Ling Hsueh, and Yu-Ting Chang. 2018. Building a Taiwan Law Ontology Based on Automatic Legal Definition Extraction. *Applied System Innovation*, 1(3):1–20.

Sagar Joshi, Sumanth Balaji, Jerrin Thomas, Aparna Garimella, and Vasudeva Varma. 2022. Investigating Strategies for Clause Recommendation. In *Proceedings of the 35th Legal Knowledge and Information Systems (JURIX'22)*, volume 362, pages 73–82. IOS Press.

Daniel Martin Katz, Dirk Hartung, Lauritz Gerlach, Abhik Jana, and Michael J. Bommarito II. 2023. Natural Language Processing in the Legal Domain. *CoRR*, abs/2302.12039.

Kwok-Yan Lam, Victor C. W. Cheng, and Zee Kin Yeong. 2023. Applying Large Language Models for Enhancing Contract Drafting. In *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023)*, volume 3423, pages 70–80. CEUR-WS.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, 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.

Nigel Markey, Ilyass El-Mansouri, Gaetan Rensonnet, Casper van Langen, and Christoph Meier. 2024. From RAGs to riches: Using large language models to write documents for clinical trials. *CoRR*, abs/2402.16406.

Marko Marković and Stevan Gostojić. 2022. A knowledge-based document assembly method to support semantic interoperability of enterprise information systems. *Enterprise Information Systems*, 16(5).

Monica Palmirani and Guido Governatori. 2018. Modelling Legal Knowledge for GDPR Compliance Checking. In *Proceedings of the 31st Conference on Legal Knowledge and Information Systems (JURIX'18)*, volume 313, pages 101–110. IOS Press.

Monica Palmirani and Fabio Vitali. 2011. Akoma-Ntoso for Legal Documents. In *Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management*, pages 75–100. Springer.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318. Association for Computational Linguistics.

Priyanka Ranade and Anupam Joshi. 2024. FABULA: Intelligence Report Generation Using Retrieval-Augmented Narrative Construction. In *Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'23)*, pages 603–610. Association for Computing Machinery.

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.

Thibault Sellam, Dipanjan Das, and Ankur Parikh. 2020. BLEURT: Learning Robust Metrics for Text Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892, Online. Association for Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, . . . , Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurélien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. *CoRR*, abs/2307.09288.

Bernhard Waltl, Jörg Landthaler, Elena Scepankova, Florian Matthes, THOMAS Geiger, Christoph Stocker, and Christian Schneider. 2017. Automated extraction of semantic information from german legal documents. In *IRIS: Internationales Rechtsinformatik Symposium, Austria*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *Proceedings of the Eighth International Conference on Learning Representations (ICLR'20)*. OpenReview.net.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *CoRR*, abs/2306.05685.

## A. Examples

Below we show some examples of definitions generated by the RAG approach with LLaMA-2 for a term present in a legal act. *original definition* and *generated definition* state definition present in the legal act, and generated definition respectively.

'**legal term**': 'rules'
'**Celex ID**': '32021Q0618(02)'
'**original definition**': "'rules' means these Rules of the Supply Agency of the European Atomic Energy Community determining the manner in which demand is to be balanced against the supply of ores, source materials and special fissile materials;"
'**generated definition**': "'rules' means procedures and guidelines established by the european agency for the supply of nuclear materials for the purpose of ensuring the regular and equitable supply of nuclear materials to community users;"

'**legal term**': 'generic ecodesign requirement'
'**Celex ID**': '32009L0125'
'**original definition**': "'generic ecodesign requirement' means any ecodesign requirement based on the ecological profile as a whole of a product without set limit values for particular environmental aspects;'
'**generated definition**': "'generic ecodesign requirement' means a method for improving environmental performance of products without setting limit values, focusing on significant environmental aspects;"

'**legal term**': 'refrigerating appliance'
'**Celex ID**': '32019R2016'
'**original definition**': "'refrigerating appliance'

means an insulated cabinet with one or more compartments that are controlled at specific temperatures, cooled by natural or forced convection whereby the cooling is obtained by one or more energy consuming means;'
'**generated definition**': "'refrigerating appliance' means a device that cools or freezes matter by transferring heat from the interior of the appliance to the exterior, using vapor compression or other means, and may include one or more compartments for storing food or drinks;"

'**legal term**': 'abandoned land'
'**Celex ID**': '32019R0807'
'**original definition**': "'abandoned land' means unused land, which was used in the past for the cultivation of food and feed crops but where the cultivation of food and feed crops was stopped due to biophysical or socioeconomic constraints;'
'**generated definition**': "'abandoned land' means land that has been unused or neglected for an extended period of time, often due to economic, environmental, or social reasons;"

'**legal term**': 'adequacy'
'**Celex ID**': '32017R1485'
'**original definition**': "'adequacy' means the ability of in-feeds into an area to meet the load in that area;'
'**generated definition**': "'adequacy' means the absence of a mismatch between the available supply of electricity in a control area and the demand for electricity in that area, as determined through a control area adequacy analysis;"

'**legal term**': 'aggregates'
'**Celex ID**': '32008R1099'
'**original definition**': "'aggregates' means data aggregated at national level on the treatment or use of energy products, namely production, trade, stocks, transformation, consumption, and structural characteristics of the energy system such as installed capacities for electricity generation or production capacities for oil products;'
'**generated definition**': "'aggregates' means quantities of energy products used as raw materials in different sectors and not consumed as a fuel or transformed into another fuel, including quantities declared for electricity and heat production, gross and net calorific values, and consumption in main activity producer plants and energy end-use specification;"