Technical report of the ERC project HyperModeLex

Hyperdimensional Modelling of the Legal System in Digital Society

Michele Corazza, Monica Palmirani, University of Bologna

Chapter 1

Introduction

The subject of this technical report is the discussion of our progress in creating tools to aid the legislative process in the context of European institutions and European Member States. The legislative process is a complex task, which involves a multitude of goals that need to be addressed before and during the drafting and discussion of any piece of legislation. Typically, one of the more demanding aspects is the discovery of existing regulation on a given subject, which in the case of a European country involves an examination of documents from a multitude of institutions: the existing national legislation, European legislation, the constitution and judgments from the national Constitutional Court or equivalent institutions. With respect to the heterogeneous sources, there are legal theory principles that need to be considered, which also create a hierarchy of legal sources:

- *Lex superior derogat inferiori*: there is a hierarchy between the legal sources. For example, an EU regulation is directly enforceable in Member states, and a new law should not violate constitutional principles;
- *Lex specialis derogat legi generali*: a process of increased specification can happen, so that a more specific law is approved on a subject that already has some legislation about it. In this case, the more specific provisions will alter the existing law in the relevant portions.
- *Lex posterior derogat legi priori*: there is a temporal component that needs to be considered when a new law is approved. In general, new laws can supersede previous ones.

Another aspect of the discovery of existing legislation is an analysis of policies, or goals that a specific institutions aims to achieve with his legislative efforts. In this context, one of the subjects examined in this report regards the Sustainable Development Goals program from the United Nations. These are 17 global goals adopted by all the United Nation members in 2015, and their goal is to promote "peace and prosperity for people and the planet". With these goals

in mind, there is a necessity to find correspondences between SDGs and existing legislation, since this information can be used to improve future legislative efforts and to produce better results in terms of sustainable development.

Another goal of this report is to describe the challenges and techniques that we are in the process of creating in order to create a generative model that can be used for the automatic generation of preambles in the context of European legislation. The choice of preambles as our first goal is motivated by the idea that these are less normative portions of regulations and directives, which allows us to attempt an automatic generation of preambles without compromising the autonomy of the EU parliament and institutions. This effort is also constrained by the fact that we need to create an approach that is in line with the theory of law, which involves the previously mentioned hierarchy of legal sources, the temporal aspects of the law, the need to consider normative references. For this reason, one of the main challenge of this endeavor is to incorporate knowledge about various other documents in the prompt used to generate new preambles.

Chapter 2

Large Language Model for the legislative process

One of the goals of our research is the creation of a large language model (LLM) for the legislative process. While it would be possible to define many different tasks for such a model, our first use case is the automatic generation of preambles in European legislative documents. The choice of preambles is motivated by the idea that these are portions of documents that are less normative than others, therefore they can be generated by an automatic process with a smaller impact on the autonomy of the European institutions.

In general, the naif application of a LLM to generate any portion of a legislative documents is not a sound approach. This is because, even in the presence of a very accurate model, the documents that are in force change over time, with amendments, abrogations, new norms entering into force all the time. This is motivated by the fact that LLMs are only aware of documents that they have previously been trained on. For this reason, it is crucial to devise a method of Retrieval Augmented Generation, often abbreviated to RAG [1], which is an approach where relevant information is retrieved before prompting the model, allowing the LLM to leverage relevant information that is provided before generating an answer. This would allow the model to be aware of all the recent changes to the existing legislation, even when said changes are not part of its training set. For this reason, a preliminary goal of our approach is to create a Transformer model which is able to retrieve relevant legislative documents to allow an LLM to generate preambles.

2.1 Objectives and Challenges

In order to generate preambles, the most relevant objectives of a retrieval model should be the normative references that are included in citations, which are needed in order to allow the model to include them in the generated document without generating spurious or out-of-date references (hallucinations). Since

the application of transformer models to large collections of data is expensive, it is not feasible to apply a model to the existing documents each time the user asks for relevant references for a new normative document, since this would require a lot of computational power to compare the query to tens of thousands of documents. For this reason, our goal is to create a model that creates vector representations for documents which support the application of a distance or similarity metric, so that the comparison between a user query and the existing documents is an operation on a vector database, not involving expensive models. The results can then be improved by using a cross-encoder for re-ranking [2] to further improve the performance of the model.

To create a retrieval model for legislative documents, there are a number of challenges that need to be addressed, which involve a multitude of peculiarities of legislative documents. These challenges do not allow the straightforward application of an existing model, and they require different solutions. One of the challenges of treating legislative documents is the fact that there are information about said documents which are not included when considering only a plain text representation of their content. In particular, legislative documents are structured in a hierarchical way, where for example and article is composed of multiple paragraphs, and some of them might include lists of points or subparagraphs. In addition to these structural aspects of the document, another challenge is the fact that information about context is also a challenging aspect of these types of documents: the jurisdiction and temporal parameters of a normative document are generally represented in text, but they might be missed by an encoder model. Finally, the usage of normative references is also problematic from the point of view of any language model that aims to perform some tasks on legislative document. As an example, in European directives and regulations it is frequent to find definitions that have the form: "(50) 'personal data' means personal data as defined in Article 4, point (1), of Regulation (EU) 2016/679;". In these instances, the definition of 'personal data' is defined from a normative reference (in this case the GDPR), meaning that we are not able to represent the actual definition without resolving the normative reference first. For these reasons, our aim is to create a model which leverages the information contained in the Akoma Ntoso XML standard [3, 4], which has been adopted by a multitude of international institution [5, 6, 7, 8, 9] and which is able to represent all the relevant information about legal documents (temporal aspects, normative references, structure of the document, etc). For these reasons, both the retrieval model and the LLM should operate on Akoma Ntoso documents and not plain text ones.

Another challenge that emerges when applying any LLM or transformer model to normative documents is the fact that most models operate on relatively short sequences of text, since the attention mechanism behind transformers has a memory cost that is quadratic in the length of the input. With longer sequences, the naif application of the quadratic multi-head attention which is present both in encoder and decoder models becomes unfeasible. This is especially critical for encoder model, which typically operate on very short sequences of 512 tokens maximum. In this context, a visualization of the length of normative documents

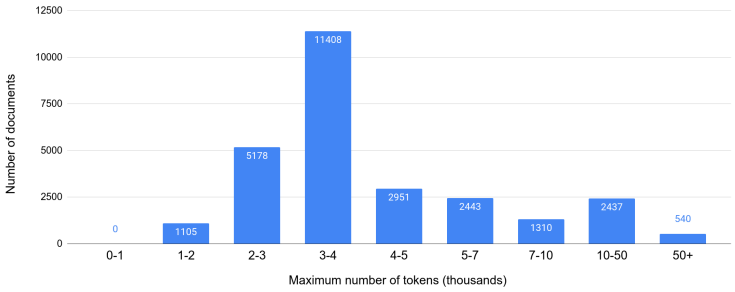

Figure 2.1: The frequency of documents having a given length in terms of the number of tokens

from the EU is shown in Figure 2.1, where we show the frequency of documents having a given length in terms of tokens. For this experiment, we used the RoBERTa [10] BPE tokenizer. The results show that no document has less than one thousand tokens, with most of them having more than three thousands tokens. In this context, it would still be possible to encode these documents using models that have been devised to treat at least some documents such as the Longformer [11]. This approach, however, does not consider the fact that information about normative references should be included in any representation of legislative documents. This is also valid for state-of-the-art generative models which are able to treat 128k tokens [12]. With the requirement to generate long documents, and a number of documents being included in the prompt from the retrieval model the context would quickly be used, preventing the generation of the preamble.

2.2 The retrieval module

To fulfill the requirements outlined in the previous section, our goal is to create a retrieval model starting from an encoder transformer that can deal with the length of normative documents and normative references. For this reason, we took inspiration from hierarchical BERT [13], which is a model that uses a pre-trained RoBERTa to produce embeddings for the sentence in a document. Then, a global model uses the output vectors from RoBERTa to produce a global representation of the entire document, which is trained to perform the desired task. Our approach is similar and it starts from a reference encoder model, which is tasked with producing a vector representation for the references of a given document d.

In order to operate on the document, we first replace the references from the document by finding the "ref" elements in the Akoma Ntoso document and replacing them with a special token [REF] which is added to the model tokenizer so that it is always represented as a single token. We then chunk the document in fragments, which we name $f_1, \ldots, f_n$. For each fragment, we denote as R_{ij}

the j-th reference present in fragment i. We also denote as t_{ij} the position of the [REF] token corresponding to R_{ij} in the sequence of tokens of fragment f_i.

The referenced document is then chunked in different fragments, which we denote as $r_{ij1}, \ldots, r_{ijn}$. The local references model is applied to the various fragments as follows:

$$e_{r_{ijh}} = L_R(r_{ijh}) \tag{2.1}$$

Where L_r is an encoder model (in our experiments we are using RoBERTa) which we call the **local references model**, which returns a vector representation for the sequence. We denote as $e_{R_{ij}}$ the concatenation of all the vector representations of the reference fragments:

$$e_{R_{ij}} = \bigoplus_{h=1}^{n} e_{r_{ijh}} \quad e_{rijh} \in \mathbb{R}^{1 \times e_s} \tag{2.2}$$

Where $\bigoplus$ denotes the concatenation of vectors along their first dimension. In order to use the resulting vectors in a global model, we first need to insert the [SEP] and [CLS] tokens that are used at the beginning and end of the sequences in BERT and other encoders, which we denote using the BERT nomenclature for clarity. The presence of these special tokens is crucial, as it allows the model to understand where sequences begin and end. These tokens are represented using their embedding, which are derived from the embedding layer of the global references model G_R. Given the sequence of local embeddings e_{Rij} we produce the inputs for the global model as follows:

$$I_{Rij} = e_g([CLS]) \oplus e_{Rij} \oplus e_g([SEP]) \tag{2.3}$$

Where e_g represents an embedding layer, and $[CLS]$, $[SEP]$ represent the index assigned to the [CLS] and [SEP] tokens used in encoders. Then, it is possible to apply the global model to the vector to obtain a vector representation for an entire document:

$$D_{Rij} = G_R(I_{Rij}) \tag{2.4}$$

Where G_R is an encoder model, the **global references model**, which does not use an embedding layer, instead it only applies positional embeddings to the existing embeddings (I_{Rij}). The next step for the encoding of entire document is to incorporate information about normative references in the overall representation of the document. For this purpose, we devised a method that uses the vector representation for the references D_{Rij} as one of the inputs for the representation of the document. For a given fragment f_i we first produce the embeddings for its tokens using an embedding layer:

$$E_{f_i} = e_l(f_i) \tag{2.5}$$

In order to correctly position the references, we move each reference in the position of the corresponding [REF] token in a matrix which we call E_R:

$$E_R = (h_{ij}) \tag{2.6}$$

$$h_{ij} = \begin{cases} D_{Rik} & \exists k \; t_{ik} = j \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

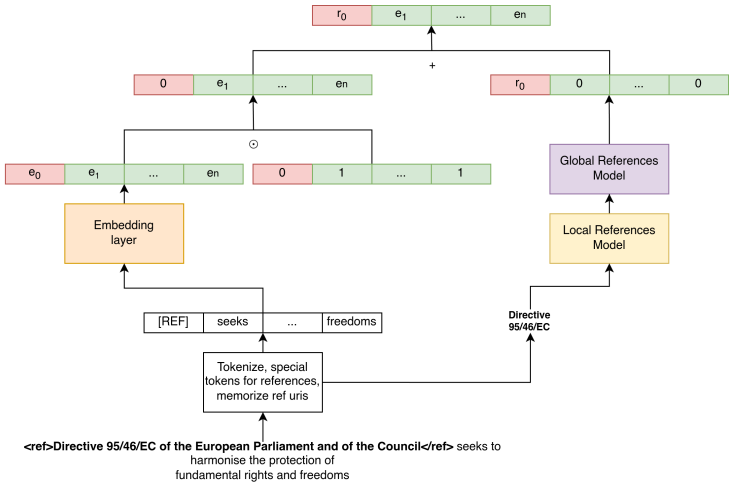

Figure 2.2: The procedure used to include information about references in the vector representation of documents

Furthermore, we produce a mask M_R which contains 0 only when the corresponding token is a reference, 1 elsewhere:

$$M_R = (m_{ij}) \tag{2.8}$$

$$m_{ij} = \begin{cases} 0, & \exists h \; t_{ih} = j \\ 1 & \text{otherwise} \end{cases} \tag{2.9}$$

Then, it is possible to combine the embeddings E_{fi} with the matrix as follows:

$$\hat{E}_{fi} = E_{f_i} \odot M_R + E_R \tag{2.10}$$

Where $\odot$ denotes the Hadamard product (element-wise product) of the two matrices. The resulting embeddings $\hat{E}_{fi}$ have the desired properties: they encode normative references as inputs to the model and their position in the text is still considered. The overall procedure used to represent references inside the input is shown in Figure 2.2.

From the embeddings $\hat{E}_{fi}$ we can then apply a local model as follows:

$$e_{F_i} = L_D(\hat{E}_{fi}) \tag{2.11}$$

Where L_D is an encoder model which does not apply an embedding layer and which returns a vector representation of the entire sequence. The resulting representations for the fragments are then combined as follows, in the same way that reference fragment embeddings were combined in equation 2.3:

$$I_d = e_G([CLS]) \oplus e_{F_i} \oplus e_G([SEP]) \tag{2.12}$$

Where e_G is another embedding layer, distinct from the one used for references (e_g). Finally, a global model is used to obtain the overall representation of the documents:

$$D = G_D(I_d) \tag{2.13}$$

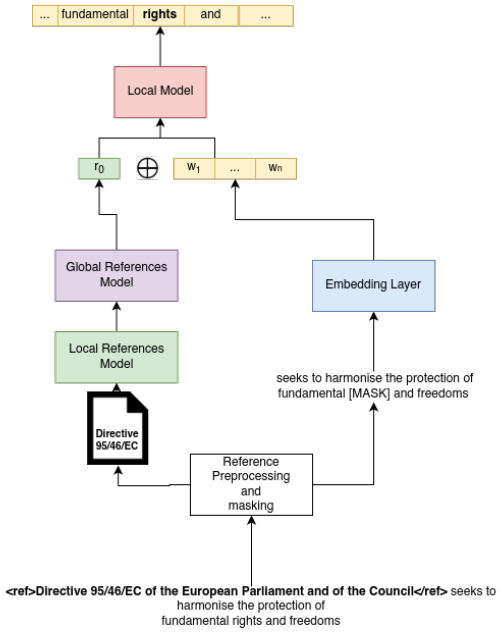

Figure 2.3: The pretraining procedure using the Masked Language Modeling training objective

In order to train this model, we adopt a two-phase approach. The first goal is to adapt a pre-trained mdoel to the vocabulary of Akoma Ntoso XML and to the legal domain. For this reason, we continue its pre-training by using Masked Language Modeling (MLM) as the training objective, with the usual categorical crossentropy loss function between the predicted masked words and the real ones which is used in RoBERTa. In particular, we operate on the local model L_D and use its The procedure is shown in Figure 2.3 and it allows us to train both the local model and the local and global references model, since they are both used during this phase. The global model G_D is not involved in this phase.

In order to train the overall model, then, we adopt a contrastive loss to force the model to produce "semantic" embeddings which can be used to retrieve the documents that should be referenced from a given title. In particular, we use a batch of titles of documents T. We define a batch of documents $P = (p_1, \ldots, p_{bs})$ where p_i is cited by the document having title T_i in its preamble (positive samples). We denote a series of documents $N = N_1, \ldots, N_{bs}$, where document N_i is not cited by the document having title T_i (negative samples). First, we produce the concatenation of the negative and positive sample embeddings along

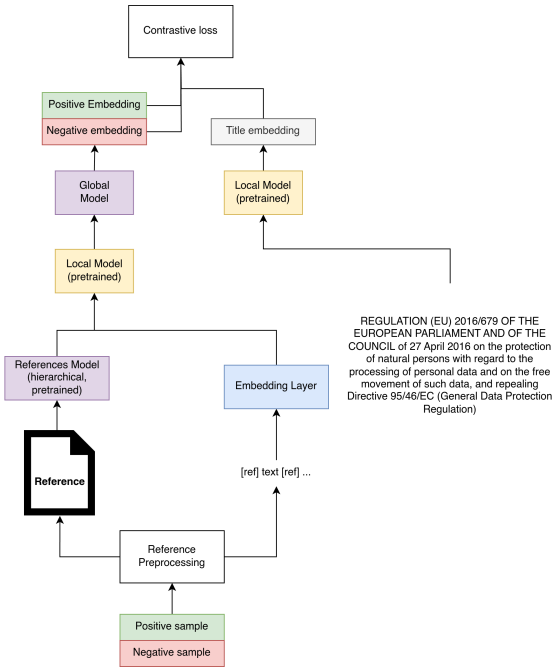

Figure 2.4: The fine-tuning procedure using the Masked Language Modeling training objective

the first dimension:

$$C = (L_D(p_1), \ldots, L_d(p_{bs})) \oplus (L_D(n_1), \ldots, L_d(n_{bs})) \tag{2.14}$$

We also compute the title embedding from the global model:

$$\hat{T} = G_D(T) \tag{2.15}$$

We can then compute the cosine similarity matrix between the title embeddings and the positive and negative samples as follows:

$$S = \lambda \frac{\hat{T} C^T}{|\hat{T}||C|} \tag{2.16}$$

Where $|T|$ denotes the l2 norm of each row of a matrix and the fraction is also applied row-wise, λ is an integer value used to scale the cosine similarity. Since the positive samples appear before the negative samples, we can calculate some labels which inform us of where to find the relevant document for each title:

$$y = \text{diag}(1, 2, \ldots, b_s) \tag{2.17}$$

Which represents a matrix where only the diagonal contains positive numbers, while the rest of the matrix is set to 0. Finally, we can define the loss function

in terms of cross-entropy as follows:

$$\mathcal{L}(T, S) = \frac{1}{b_s} \sum_{i=0}^{b_s} \log \frac{\exp S_i}{\sum_{j=0}^{b_s} \exp S_j} y \tag{2.18}$$

Which attempts to condition the model to produce title vectors which are most similar with positive sample embeddings (cosine similarity of 1), while title embeddings and negative sample embeddings should be orthogonal (cosine similarity of 0). The procedure used to train the model is shown in Figure 2.4. Finally by applying this procedure, we are able to train a model that can handle very long sequences of tokens ($512^2 = 262k$), which can include information about normative references in its representation of documents, and that is suitable to perform information retrieval for the generation of preambles.

2.3 Document Chunking Strategy

In order to perform the chunking of documents required for the model, we could use a simple sentence-based approach, where we would segment the document in sentences and use those as fragments for the local models. However, thanks to the structural information provided by the Akoma Ntoso XML format, it is possible to leverage this information to produce fragments that follow the structure of the document itself.

Figure 2.5: An example of the application of our Akoma-Ntoso informed chunking procedure.

In order to achieve this goal, we adopted a greedy algorithm, which operates by traversing the XML tree using a depth-first approach (which coincides with processing the XML elements in sequence). First, we start to segment from the beginning of the document (the first token) and establish a token limit of 512 for our approach, which is the maximum allowed length for BERT and similar

models. While adopting a pre-trained tokenizer, we force it to consider ">" as a special token, meaning that the end of an element is kept as an isolated token without being joined with other textual content.

We traverse the elements of the XML tree and measure the number of tokens that are comprised between the first one and the token at the end of the current element. We continue traversing the tree until we find an element for which the sequence of tokens is longer than 512. We then use the last valid element which had a shorter sequence to populate the first chunk. The procedure is then repeated starting from the first token after the last chunk until all the document is consumed. Crucially, while traversing the tree, we avoid nodes that are inline for the Akoma Ntoso standard such as dates, references, modifications in order to avoid splitting a sentence in two parts. An example of the application of this procedure is provided in Figure 2.5.

2.4 Critical Assessment

While the procedure described in this chapter is already implemented, we faced a number of challenges during its development, which is still ongoing. Our first experiments used a pre-trained RoBERTa model to initialize both the local references model and the local model. This approach, however, can led to a very high number of tokens for each document, since the RoBERTa tokenizer was not trained on the Akoma Ntoso vocabulary. For this reason, we manually inserted the opening and closing of each element as special tokens, in an effort to reduce the length of the sequences (eg <ref and </ref are special tokens). Furthermore, the hierarchical nature of the model is very computationally expensive, especially in terms of memory. In the first experiments, this did not allow us to start from a pre-trained RoBERTa model due to the high memory requirements of the model and the high number of tokens from a pre-trained BPE tokenizer, even using an A100 GPU from Nvidia.

Another crucial question that we do not yet have an answer for is whether the usage of a single embedding for an entire document is sufficient, and the same question can be asked for the fragments. This compression of information in a relatively low dimensionality vector could also create problems in terms of the flow of gradients, since we are asking the outputs to be potentially conditione on a single word of a reference, which is filtered through three separate transformer models before influencing the output.

On the other hand, our hierarchical approach fulfills all the requirements for the encoding of normative references. First of all, it operates on Akoma Ntoso XML documents, meaning that it has access to metadata and structural information about the document which would otherwise be lost. Furthermore, it can handle documents with a very high number of tokens, meaning that the vast majority of documents can be encoded using it without any truncation. The usage of Akoma Ntoso also allows our approach to discard abrogated documents that are not useful references for a new preamble, reducing the computational cost of the operation and improving its accuracy. The model is also reference-

aware, meaning that it is able to consider the content of normative references when producing the vector representation of a model. Finally, the result of the application of the model is a vector representation for each document, which can be compared with the title of the norm being created by just applying cosine similarity to their vector representations. This approach has relatively small requirements in terms of computational cost.

In terms of the current progress, first we managed to collect a corpus of approximately 40k EU legislative documents in the Akoma Ntoso format, which in terms of files size corresponds to 3.8 GiB. In an effort to reduce the number of tokens per document and to discard less relevant information we proceeded to exclude many different types of data and metadata for our experiments (tables, annexes, attachments, lifecycles, proprietary, restrictions, etc). This resulted in a marked reduction in the overall size of the dataset, which is now of approximately 2.0 GiB. The entire model is ready, but we are still testing whether it behaves as expected by applying it to a standard Information Retrieval dataset, The Stanford Question Answering Dataset 2.0 [14, 15], before reintroducing all the complexities which are connected with the legal domain. Meanwhile, we are investigating ways to reduce the memory requirements of the model such as reducing the vocabulary size of the global models to only include [CLS] and [SEP]. For the local models, we are also investigating more advanced vocabulary transfer techniques [16] in order to reduce the number of tokens for our documents.

Chapter 3

Hybrid classification of Sustainable Development Goals in European legislative documents

The Sustainable Development Goals (SDGs) were adopted by the UN assembly in 2015 and they are composed by 17 goals and 169 targets, typically 12 to 15 for each goal. Our goal is to adopt an unsupervised hybrid classification framework to match existing legislative documents from the EU with SDG targets, producing a fine-grained classification of articles. This approach allows an *ex-ante* analysis of drafts, which can help the legislative efforts by checking whether new legislation is in line with the policies outlined by the SDGs. Crucially, the same analysis can also be applied *ex-post*, allowing a more long-term analysis of the merits and shortcomings of existing legislation with respect to the SDGs.

The Joint Research Center (JRC) of the European Commission recognized the importance of matching JRC with existing legislation, and for this reason a manual annotation of legislative documents has been performed starting in 2017. This annotation effort was applied to actions from the Juncker commission (2014-2019) so that all document were clearly annotated with one or more SDG goals [17]. While the effort to map EU legislation with SDG targets, continues, the von der Leyen commission opted to perform an automatic annotation of the documents, which allowed the creation of a second, manually annotated, matching between SDG and EU legislative documents. Both annotations can be found at `https://knowsdgs.jrc.ec.europa.eu/policies-sdgs`, where "previous initiatives" corresponds to the Juncker commission's manual annotation, while "current initiatives" is the current effort to annotate documents automatically.

Crucially, both the manual and automatic annotations only examined the

document-level match between provisions and legislation, while it would be useful to provide a more fine-grained classification of normative documents, which can be useful in both *ex-ante* and *ex-post* analyses to assess the effectiveness of specific portions of the legislative document (i.e. articles).

Our goal in this context is the creation of an automatic, unsupervised method that, given a legislative document, matches articles and recitals with the relevant SDGs. Furthermore, we aim to reconstruct the document-level annotations from the fine-grained ones.

3.1 The dataset

In order to be able to assess the effectiveness of the method, we used the documents that were manually annotated during Juncker Commission's mandate from 2014 to 2019. Our dataset, on the other hand, is a collection of documents extracted from the EUR-LEX portal from the period 2010-2021. These documents were automatically converted from FORMEX to the Akoma Ntoso XML format, allowing us to leverage information about normative references, the temporal aspects of the law, the structure of the documents (i.e. articles, lists, points, paragraphs, etc). By intersecting these documents, we obtained a first dataset composed of 2791 documents, which was used in the first experiments involving EU legislation and SDG targets. Later, unlike the SDG annotations from the JRC, we opted to include in our approach the consolidated versions of documents as well, assuming that they retain the same relevant SDGs as the original versions, obtaining a second dataset composed of 3846 annotated documents. Since our approach is unsupervised, the fact that we include successive consolidated versions of documents allows us to examine the trends that emerge over time when considering the relation between SDG targets and EU legislation.

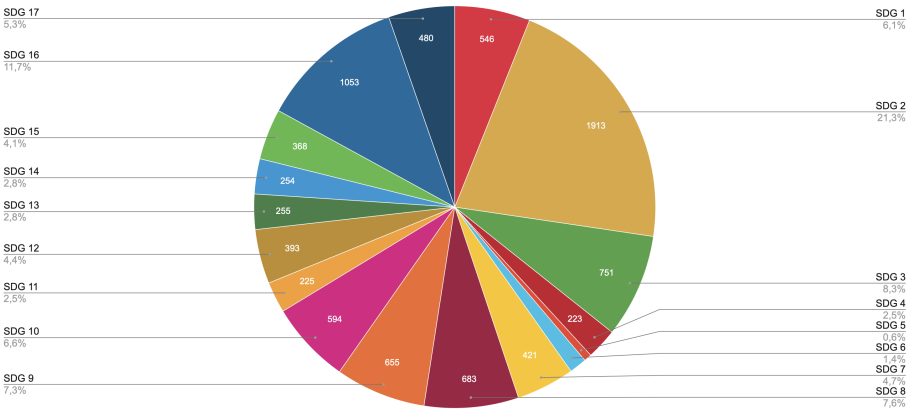

Figure 3.1: Number of documents per each SDG in our dataset.

In Figure 3.1, we show the number of documents that have been annotated with each one of the 17 SDG goals. From this image, it is clear that all goals are represented in our dataset, but there are strong differences in their prevalence in our dataset. In particular, the second goal ("end hunger, achieve food security and improved nutrition and promote sustainable agriculture") is very prominent in EU legislative documents, appearing in more than 20% of the documents, and goal 16 ("promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels") is also relatively prominent, encompassing more than 10% of the documents. On the other hand, the least attested goal is number 5 ("achieve gender equality and empower all women and girls"), found in less than 1% of the documents, and goal number 6 ("ensure availability and sustainable management of water and sanitation for all") is also relatively rare.

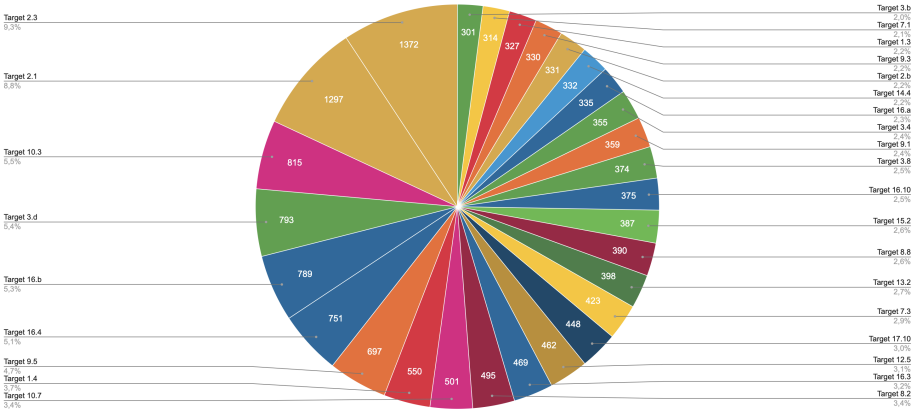

Figure 3.2: The 27 most featured targets in our dataset.

While examining the relative frequency of overarching goals can already provide useful information about the legislative efforts and it is crucial to consider the existing imbalances between goals for any automatic annotation method, it is also interesting to consider whether there are meaningful differences with respect to SDG targets, which are the main objective of our hybrid classification. For this purpose, we can observe Figure 3.2, which shows the 25 most frequent targets in our dataset. Unsurprisingly, the targets belonging to goal 2 appear frequently, with 3 different targets in the 25 most frequent (targets 2.3, 2.1 and 2.b). Targets 2.3 and 2.1 are also the ones that are overall more frequent in the entire dataset. The third most frequent target is 10.3 from goal 10, whose aim is to "Reduce inequality within and among countries". Another prominent goal in the most frequent targets is goal 16, represented by 4 targets (16.b, 16.4, 16.3, 16.10,16.a), the most out of any other goal.

From this preliminary analysis it is already clear that the behavior of targets and SDG goals can be different from one goal to another. As an example, goal 16 appears more infrequently than goal 2, but it is represented by more targets.

This heterogeneous behavior must be considered when examining the results, as the prominence of some SDGs can be higher than those of others by two orders of magnitude, making any automatic classification more challenging.

3.2 Method

In order to produce an unsupervised model to classify legislative documents with its SDG targets, we opted to use a Sentence Transformer model [18]. These models have a crucial advantage over other types of transformer models, and it has to do with how the embeddings obtained from them behave. These embeddings are trained using an approach similar to Siamese Networks, with a contrastive loss, which allows the direct application of a similarity or distance metric (typically cosine similarity) in order to assess the semantic similarity between two different sentences. Between the various pretrained models, we chose *all-distilroberta-v1* since it shows good performance and it supports a sequence length of 512 tokens.

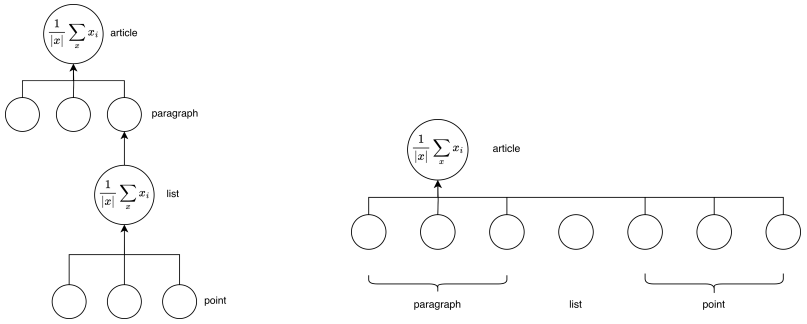

Figure 3.3: From left to right, the tree and flat strategies to average the embeddings, applied to an article with three paragraphs, one containing a list of points. Each node in the XML tree that directly contains text is associated with its own embedding vector from the model (the list, which contains no text, is ignored in the flat strategy). The vectors are then aggregated using the two strategies, and the sum in some nodes denotes the mean between vectors. The arrows denote the flow of the (aggregated) vectors in the tree.

Unfortunately, while we opted to maximize the number of tokens that the model supports, some articles were still longer than the allowed limit, which meant that some strategy for chunking and arggregating vectors needed to be adopted. Since our documents are in the Akoma Ntoso XML format, we opted to leverage information about the structural components of the documents (articles, points, paragraphs, etc) to inform the chunking procedure. In particular, we devised two different aggregation strategies for the chunking procedure, which we dubbed **flat** and **tree** (see Figure 3.3). The flat procedure is the simplest of the two, and it works by obtaining the representation of a portion of the document (like an article) from a flat average of its children element in

Akoma Ntoso. During this process, we excluded the inline elements of Akoma Ntoso (eg references, dates, etc) in order to obtain only elements that indicate structural components of the document (paragraphs, points in lists, etc). Formally, given a leaf element l (an element with no non-inline children) and the Sentence Transformer model M, we obtain its vector representation using

$$v(l) = M(t(l)) \tag{3.1}$$

Then, we can obtain the flat vector representation of an arbitrary element by applying the following procedure:

$$F(e) = \frac{1}{1 + |f(e)|} \left(M(t(e)) + \sum_i v(f_i(e)) \right) \tag{3.2}$$

Where $c(e)$ is a function returning all the non-inline children of e and all its children and $t(e)$ returns the textual content of node e.

The second strategy, named tree, leverages the hierarchical structure of the XML in order to aggregate the components from the bottom up. In particular, if an article contains a list of points, we first obtain the vector representation of the list, which is then averaged with the vector representation of all the other children of the article element. Formally, we obtain the vector representation of an element using the following recursive function:

$$T(e) = \frac{1}{1 + |c(e)|} \left(M(t(e)) + \sum_i T(c_i(e)) \right) \tag{3.3}$$

Where $c(e)$ is a function that returns the children of element e.

While this approach was used for one of the publication on the subject [19], we also extended the tree aggregation approach to consider normative references as well. In particular, when an element contains a normative reference, its vector representation is obtained by the average of its textual content, the vector obtained from its children elements and the average of the vector representations of all the references that are included in its text. For the purposes of this report, we well name this aggregation strategy "tree+". In order to represent normative references, we differentiated between punctual references (to a specific article, section, point) from generic references (to an entire document). Formally, we obtain their vector representations from:

$$R(i) = \begin{cases} v(i) & \text{if } i \text{ is a punctual reference} \\ \frac{1}{2} M(title(i)) + v(article_1(i)) & \text{otherwise} \end{cases} \tag{3.4}$$

Where $title(i)$ and $article_1(i)$ represent the title and first article of a document, respectively. With this approach, we obtain the vector representation of punctual references like we would for any given element of the XML tree, while we obtain the vector representation of an entire document for the purposes of the references from the average of its title and the vector representation of its first

article from the tree strategy. Finally, it is possible to incorporate the normative references in our vector representation:

$$v(e) = \frac{1}{2 + |c(e)|} \left(M(t(e)) + \sum_i v(c_i(e)) \frac{1}{r(e)} \sum_j R(r_j(e)) \right) \tag{3.5}$$

Where $r(e)$ is a tuple containing all the references in the text of the node e, while $r_j(e)$ represents the *j-th* reference.

The two approaches we described (plus the addition of the references) were used to produce a vector representation for each article of the documents. Then, the vector representation of all the SDG targets descriptions from the official UN documentation were obtained by simply applying the sentence transformer model to them. Finally, a comparison between these two groups could be performed by calculating the cosine similarity between them, where a higher similarity corresponds to a higher degree of relatedness between an article and an SDG target

3.3 Experiments and evaluation

Using the methods described in the previous section, we performed two distinct but related experiments, which have been published in two different conferences. In the first experiment, we wanted to validate the overall approach and to assess which aggregation strategy showed more promise between the flat and tree ones. This experiment was performed using the smaller dataset, which did not yet include the consolidated documents. The tree+ strategy had also not been developed yet.

In order to assess the validity of our approach, the first step was to perform a statistical test to see whether the obtained results could be explained by a random choice of SDGs. Unfortunately, we did not have a full annotation of all the documents at the article level, instead we used only annotations at the document level. To assess the effectiveness of our methods, we made a reasonable assumption about the nature of the relation between SDG targets and articles: that the SDGs that are relevant for a given document should be considered relevant for at least one article. Using this assumption, it is possible to calculate the expected number of articles that would be found even by random chance through a Poisson Binomial Distribution to obtain a random baseline and compare it with the two splitting strategies, which allows us to obtain a p-value for the results as well. The results, shown in Table 3.1, show that our results can't be due to random chance (p-values that are less than 10^{-19}). Unfortunately, the formula for the cumulative distribution of the Poisson Binomial Distribution involves a sum over a large number of probabilities, which led to numeric cancellation, meaning that the p-values can only be expressed as inequalities. This also means that the p-values of the two different strategies are the same, but we do see that the tree strategy produces more correct matches than the flat one.

Split Stategy	No. Matches	P-value	Correct matches
Flat	3613	$< 2.71 * 10^{-19}$	0.165
Tree	3738	$< 2.71 * 10^{-19}$	0.171
Random Baseline	826	0.51	0.038

Table 3.1: Comparison between the two splitting strategies and a random baseline, including p-values for the right tailed tests.

Another test involves the comparison of the semantic similarity between the distances between articles and SDG targets that are relevant for the document containing them (gold standard SDGs) and those which are not (non gold standard SDGs).

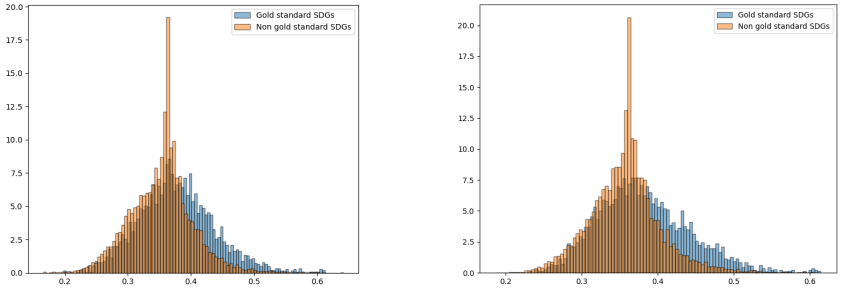

Figure 3.4: The histograms of the similarities from SDG targets in the gold standard and those that are not for the flat split strategy (on the left) and the tree split strategy (on the right).

The first comparison is a visual one and it is applied to both strategies, and is shown in Figure 3.4, which shows that the gold standard SDGs appear to have a higher similarity with the relevant documents when compared with non gold standard SDGs. In order to test this claim, we perform a Welsh t-test on the two distributions, in order to assess whether the observed difference between them is statistically significant.

Split Strategy	Gold Standard SDGs	Non-Gold Standard SDGs	P-value
Flat	0.377 ± 0.06	0.353 ± 0.05	$6.84 * 10^{-105}$
Tree	0.381 ± 0.06	0.357 ± 0.05	$5.61 * 10^{-113}$

Table 3.2: Welsh t-test comparing the distributions of top ranking SDG targets that are and are not in the gold standard, respectively, for both splitting strategies.

The result of this evaluation, which is shown in Table 3.2, shows that the

difference in distributions is indeed significant, and that the difference seems to be more pronounced when using the Tree strategy, meaning that this strategy should be favored when choosing between the two alternatives.

The second evaluation [20], which used the tree+ aggregation strategy and the addition of the consolidated documents, involved the question of whether our approach could be used to reconstruct document-level annotations from the more fine-grained ones that we get from the model. First, we define as "predicted matchin" those SDG targets that appear in the top 5 highest similarities of a given article. In order to reconstruct the document-level correspondence with SDG targets, we experimented with four different methods:

- **All articles**: we use the union of the predicted matching SDGs for all articles of a document;
- **First four articles**: we use the union of the predicted matching SDGs for the first four articles of a document.
- **First four articles + recitals**: we use the union of the predicted matching SDGs for the first four articles of the document, as well as the predicted matching SDGs for all the recitals.

The validation of this approach, then, was performed on the entire dataset, including the consolidated versions of documents (which were considered as having the same associated SDG targets as the original version of the document). A random baseline was also included, by selecting 5 random SDGs for each article.

Strategy	Average	Precision	Recall	F1 Score
All articles	Macro	**0.11**	0.16	0.07
	Weighted	**0.37**	0.22	0.14
Random (All articles)	Macro	0.03 ± 0.0003	0.33 ± 0.008	0.06 ± 0.0004
	Weighted	0.12 ± 0.001	0.41 ± 0.003	0.17 ± 0.001
First four	Macro	0.09	0.09	0.04
	Weighted	0.29	0.12	0.06
Random (First four)	Macro	0.02 ± 0.0006	0.08 ± 0.007	0.03 ± 0.0008
	Weighted	0.09 ± 0.003	0.1 ± 0.003	0.08 ± 0.002
Recitals + first four	Macro	0.09	0.36	**0.10**
	Weighted	0.27	0.44	**0.22**
Random (Recitals + first four)	Macro	0.03 ± 0.0001	$\mathbf{0.52} \pm 0.008$	0.05 ± 0.0002
	Weighted	0.10 ± 0.0008	$\mathbf{0.65} \pm 0.003$	0.16 ± 0.001

Table 3.3: Precision, recall and F1 score for the four strategies, obtained from Macro and Weighted averages over individual classes. In bold, the best values for each metric. Underlined, the higher metric when comparing each strategy with its baseline. For each baseline we report the means and standard deviations of the metrics over 100 runs.

The results, shown in Table 3.3, show that, while it is possible to reconstruct the annotation for the entire document from more fine-grained ones even in an

unsupervised setting, the results suffer especially in terms of precision. This is due to the fact that, by selecting the top 5 SDG targets for each article (or recital) we probably select targets for articles that are not associated with any SDG or for those that are associated with less than 5. While assessing the performance of the strategies, we can see that the recitals + first four approach is the one that obtains the best results, and that it is the only one that is better than the random baseline for both F1 scores.

The last experimental results obtained from the second set of experiments regards how the relation between SDG goals and EU legislative documents changed over time. This comparative analysis could be performed thanks to the inclusion of consolidated versions of the original documents in the second dataset.

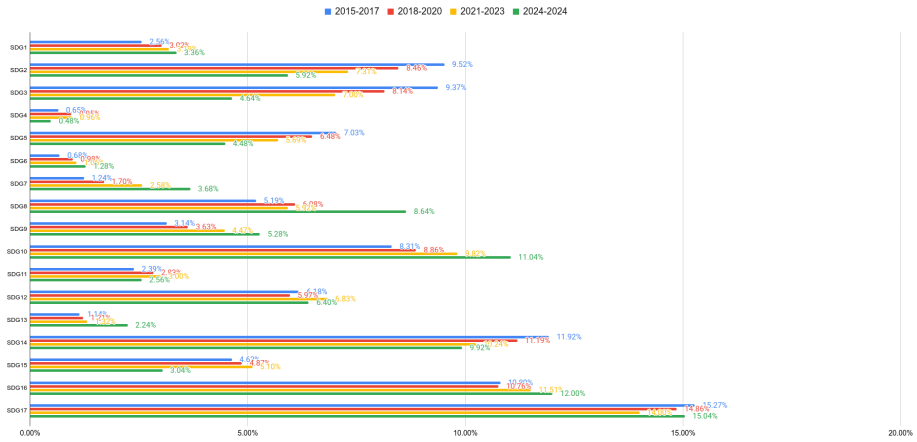

Figure 3.5: Percentage distribution of EU legislation across Sustainable Development Goals (SDGs) from 2015 to 2024 - 3-year range

Figure 3.5 shows the relative prominence of SDG goals in ranges spanning 3 years. One notable discovery regard SDG02 (Zero Hunger), which went from being associated with approximately 10% of documents in 2015-2017 to being associated with just 6% in 2024, with a steady decline over the examined period.

3.4 Conclusions

The experiments described in this chapter show that applying an unsupervised method to the classification of SDG targets that are relevant for EU legislation is a worthwhile endeavor, and that it is possible to produce a fine-grained classification by applying Sentence Transformers to a structure-aware representation of the documents in Akoma Ntoso. The proposed technique is also able to consider normative references, which are often overlooked when using Natural Language Processing techniques on legislative documents. Despite these

results, the reliability of this methodology is still not high, meaning that more experiments would need to be performed, using better transformer models or more sophisticated techniques to classify the relevant SDG targets.

Chapter 4

Hybrid classification of European legislative documents using EUR-LEX

One of the phases of the legislative process is the discovery phase, during which the existing legislation on a given subject is examined by experts, in order to avoid conflicts and redundancies of the new norm. In the context of the member states of the EU, this is further complicated by the fact that this discovery phase involving European legislation needs to be performed from the state's institution, meaning that a multitude of languages is involved in this discovery. In order to assess whether it is possible to partially automate the discovery of relevant documents, we describe an automatic, unsupervised classification of EU legislative documents using one of the official thesauri of the EU. This classification is then used to retrieve the most relevant class for a given user query, as well as the most relevant documents belonging to this class.

4.1 Dataset

As discussed in the previous section, our dataset is composed of European legislative documents from the Eur-Lex portal. The resulting corpus is composed of 14305 documents spanning from 2010 to 2021, which have been converted in Akoma Ntoso XML from the FORMEX format. While the documents were annotated with the EuroVoc terms associated with them in the Akoma Ntoso, this information was not used for our unsupervised classification, as we wanted an approach which might generalize to legislation from multiple member states.

The other starting point for our classification is the definition of which classes to consider for a given document. In the context of European legislative documents, it is sensible to adopt the EuroVoc thesaurus, which is a multilingual thesaurus from the Publication Office of the EU, translated in all the official

languages of the Union plus Albanian, Macedonian, Serbian. It is structured as a hierarchy, with top level terms which are more general, while their children are narrower concepts. For a broad classification of EU documents, we selected the top level terms of the thesaurus, excluding the term "European Union" which would apply to all documents. The selected terms are shown in Table 4.1.

Politics	International Relations
European Union	Law
Economics	Trade
Finance	Social Questions
Education and Communications	Science
Business and competition	Employment and Working Conditions
Transport	Environment
Agriculture, Forestry and Fisheries	Agri-foodstuffs
Production, Technology and Research	Energy
Industry	Geography
International organisations	

Table 4.1: Top level EuroVoc terms

4.2 Method

In order to perform an unsupervised classification of the documents, we used an approach which is similar to the one described in Section 3.2, which is based on a Sentence Transformer. In this case, however, we adopted a multilingual Sentence Transformer mode, namely "paraphrase-multilingual-mpnet-base-v2"[21]. These multilingual models have been trained in following a teacher-student paradigm, in order to produce the same vector representation for translations of the same sentence in multiple languages. The choice of a multilingual model is motivated by the fact that this mechanism might be useful for a multitude of Member States, meaning that it should ideally work independently of the language used to query it.

The classification of the documents is performed with the same method that was used for the SDGs in section 3.2. In particular, the articles of a document are represented by a vector obtained from the tree+ aggregation procedure described in Equation 3.5 with the $v(a)$ function. For the EuroVoc terms, we represented using their label followed by a semicolon separated list of the labels associated with their children. Then, since a simple classification is what we are interested in, we just select the more relevant term for a given document as follows:

$$\text{argmax}_j \left[\text{sim}(M(T(d)), M(E_j)) + \frac{1}{|A_d|} \sum_{i=1}^{|A_d|} \text{sim}(v(a_{di}), M(E_j)) \right] \tag{4.1}$$

Where M is the pretrained transformer model, $A_d = \{a_{d1}, \ldots, a_{dn}\}$ is a set that represents all the articles belonging to document d, $\text{sim}(x, y)$ measures the cosine

similarity between two vectors, while E_j represents the j-th EuroVoc term. In other terms, the system selects the more semantically similar EuorVoc using the sum of the similarity between the term and title, and the average similarity between the term and all the document articles.

In order to search for documents, the user can provide a query and a set of keywords. The query vector, then, is produced as follows:

$$Q = \frac{1}{2}\left(M(q) + M(k)\right) \tag{4.2}$$

Where q is the query text, while k represents the set of keywords, joined by semicolons (";"). The query vector can be compared with the EuroVoc terms using cosine similarity, which allows the selection of two relevant arguments for each query. For documents that have been categorized with the two relevant arguments, then, the query vector Q is compared with a document d as follows:

$$D_d = \mathrm{sim}(M(Q), M(T(d))) + \frac{1}{|A_d|}\sum_{i=1}^{|A_d|}\mathrm{sim}(M(Q), v(a_{di})) \tag{4.3}$$

In our system, we select the two most relevant EuroVoc terms and the 10 most relevant documents for each one of them, allowing the user to query European legislative documents and obtaining an argument-based selection of documents.

4.3 Evaluation and conclusions

The results of the unsupervised classification have been evaluated by experts over a randomly selected set of 100 documents. This evaluation resulted in an accuracy figure of 52%, which, while not completely satisfactory, is still higher than the one obtained from assigning a random EuroVoc term to each document, which would result in an accuracy value of $\frac{1}{20} = 5\%$. Furthermore, since our approach is to show more than one suggested reference, more than one suggested definition and more than one pertinent cluster given a user query, the overall usefulness of the approach is in practice higher, since users can select from a multiple of suggestions instead of relying on the first one.

The accuracy of the search was also validated by experts, but we did not yet perform a more quantitative evaluation. The results were promising, but it would be useful to perform a more formal evaluation, perhaps involving stakeholders involved in the European legislative process.

Chapter 5

Framenet-Enhanced RAG for legal definitions

In the context of Retrieval Augmented Generation methods, there is an increased interest in the integration of Knowledge Graphs (KG) and other semantic technologies with LLMs to provide more structural information which can be leveraged to improve the accuracy of the generated text.

In this scenario, we are interested in the application of a KG-based approach for the suggestion of normative definitions in the context of the legislative process. In order to achieve this goal, the first step was an extension of FrameNet [22], a lexical database which is founded on the concepts of Frame Semantics, which postulates that the meaning of a given word can be understood according to the semantic frame it evokes, i.e. the schematic event, relation or entity that is recalled in speakers minds alongside with its participants. The FrameNet annotations are based on frames and their components, which are called Frame Elements (FE), divided in core FE, which are essential for determining the meaning of the frame, and non core FE, which are not. Lexical Units (LU) are also annotated, which consist of words that evoke a specific frame.

The first step of our experimentation with FrameNet annotation is to create a new frame that can be used to annotate normative definitions. While this effort is still ongoing, some core FE have been proposed for this new frame:

- **DEFINENDUM**: The term or expression defined within the definition;
- **DEFINIENS**: The expression that defines the DEFINENDUM.
- **ALIAS_DEFINENDUM**: A term or expression indicated as an alternative to the DEFINENDUM and that can stand for it;

As well as some non-core FE:

- **TIME**: The time in which the Definition can and should be applied.

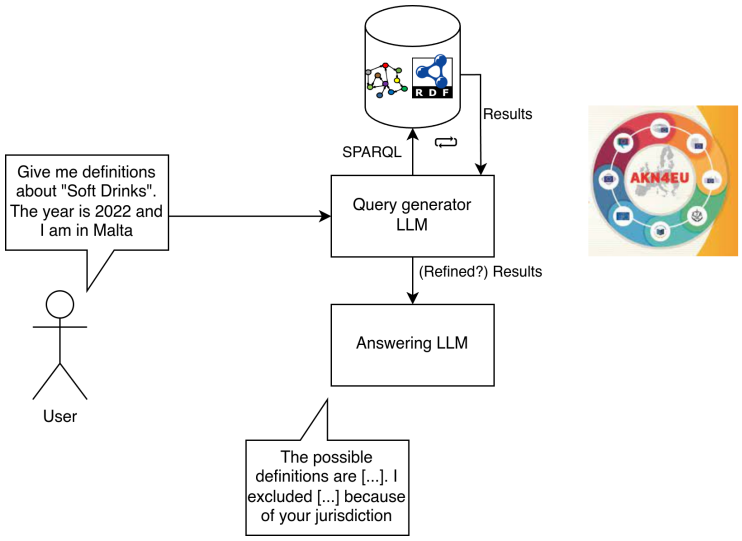

Figure 5.1: A possible visualization of the KG-based RAG for definitions.

- **GEOGRAPHICAL_SPACE**: The geographical space in which the Definition can and should be applied.
- **JURISDICTION**: The jurisdiction in which the Definition can and should be applied.
- **CONTEXT**: The context in which the Definition can and should be applied.
- **CONDITIONS**: The conditions that should be met in order for the Definiens to be applied to the Definendum.
- **EXCLUDED_CASE**: The exception, a case in which the Definition does not apply.
- **EXCLUDED_JURISDICTION**: The jurisdiction(s) in which the definition does not apply

The second step in our approach is the representation of the new frame and of the annotated documents as a Knowledge Graph, which is allowed by the Framester[23], which acts as "a hub between FrameNet, WordNet, VerbNet, BabelNet, DBpedia, Yago, DOLCE-Zero, as well as other resources". Crucially, it allows us to represent the annotation of frames as a knowledge graph.

The idea behind this conversion is to leverage this semantically informed annotation to perform a KG-based RAG for generating definitions. First, the annotated definitions and the information about the new definition frame, frame elements, etc. In this framework, we would run SPARQL queries on a graph database using an LLM to generate the queries from the user in put in natural

language. The results would then be fed to the same LLM, which can decide to refine the query based on the observed results or to accept the results and provide them to an LLM model which is tasked with giving a candiate definition to the user. Through a careful consideration of some of the annotated FE associated with the proposed definitions it is then possible to explain some of the proposals from the LLM models. As an example, it would possible to justify that one of the candidate definitions was excluded because it does not apply in a certain jurisdiction or at a certain time. These explanations would be provided by the LLM direcly in natural language. Furthermore, the LLM agent tasked with generating the queries might explore the database by trying similar related terms (eg if we are seeking the definition for "soft drink" we might also be interested in a definition for "sugar-based product").

While this endeavor is still in its preliminary stages, such an approach would be a very interesting experiment for the application of hybrid AI methods which combine more knowledge-based methods with statistical models such as LLMs. Furthermore, the semantic information annotated through Frame Semantics is always anchored to the textual content of a given norm, meaning that it can easily be compared against the original document, avoiding the discrepancies that might derive from a completely separate semantic annotation of the definitions.

Bibliography

[1] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.

[2] R. Nogueira and K. Cho, "Passage re-ranking with bert," *arXiv preprint arXiv:1901.04085*, 2019.

[3] M. Palmirani, R. Sperberg, G. Vergottini, and F. Vitali, "Akoma Ntoso Version 1.0 Part 1: XML Vocabulary." OASIS Standard, August 2018.

[4] F. Vitali, M. Palmirani, R. Sperberg, and V. Parisse, "Akoma Ntoso Version 1.0. Part 2: Specifications." OASIS Standard, August 2018.

[5] M. Palmirani, "Lexdatafication: Italian legal knowledge modelling in akoma ntoso," in *AI Approaches to the Complexity of Legal Systems XI-XII - AICOL International Workshops 2018 and 2020: AICOL-XI JURIX 2018, AICOL-XII JURIX 2020, XAILA JURIX 2020, Revised Selected Papers* (V. Rodríguez-Doncel, M. Palmirani, M. Araszkiewicz, P. Casanovas, U. Pagallo, and G. Sartor, eds.), vol. 13048 of *Lecture Notes in Computer Science*, pp. 31–47, Springer, 2020.

[6] M. Palmirani, F. Vitali, A. Bernasconi, and L. Gambazzi, "Swiss federal publication workflow with akoma ntoso," in *Legal Knowledge and Information Systems - JURIX 2014: The Twenty-Seventh Annual Conference, Jagiellonian University, Krakow, Poland, 10-12 December 2014* (R. Hoekstra, ed.), vol. 271 of *Frontiers in Artificial Intelligence and Applications*, pp. 179–184, IOS Press, 2014.

[7] M. Palmirani, "Akoma ntoso for making FAO resolutions accessible," in *Knowledge of the Law in the Big Data Age, Conference 'Law via the Internet 2018', Florence, Italy, 11-12 October 2018* (G. Peruginelli and S. Faro, eds.), vol. 317 of *Frontiers in Artificial Intelligence and Applications*, pp. 159–169, IOS Press, 2018.

[8] A. Cvejić, K.-G. Grujić, A. Cvejić, M. Marković, and S. Gostojić, "Automatic transformation of plain-text legislation into machine-readable for-

mat," in *The 11th international conference on information society, technology and management (ICIST 2021)*, 03 2021.

[9] A. Flatt, A. Langner, and O. Leps, *Model-Driven Development of Akoma Ntoso Application Profiles: A Conceptual Framework for Model-Based Generation of XML Subschemas*. Springer Nature, 2023.

[10] Y. Liu, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[11] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The long-document transformer," *arXiv preprint arXiv:2004.05150*, 2020.

[12] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan, *et al.*, "The llama 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024.

[13] J. Lu, M. Henchion, I. Bacher, and B. M. Namee, "A sentence-level hierarchical bert model for document classification with limited labelled data," in *Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings 24*, pp. 231–241, Springer, 2021.

[14] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (J. Su, K. Duh, and X. Carreras, eds.), (Austin, Texas), pp. 2383–2392, Association for Computational Linguistics, Nov. 2016.

[15] P. Rajpurkar, R. Jia, and P. Liang, "Know what you don't know: Unanswerable questions for SQuAD," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (I. Gurevych and Y. Miyao, eds.), (Melbourne, Australia), pp. 784–789, Association for Computational Linguistics, July 2018.

[16] V. Mosin, I. Samenko, B. Kozlovskii, A. Tikhonov, and I. P. Yamshchikov, "Fine-tuning transformers: Vocabulary transfer," *Artificial Intelligence*, vol. 317, p. 103860, 2023.

[17] European Commission and Joint Research Centre, S. Borchardt, G. Barbero Vignola, D. Buscaglia, M. Maroni, and L. Marelli, *Mapping EU policies with the 2030 agenda and SDGs – Fostering policy coherence through text-based SDG mapping*. Luxembourg: Publications Office of the European Union, 2023.

[18] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (K. Inui, J. Jiang, V. Ng, and X. Wan, eds.), (Hong Kong, China), pp. 3982–3992, Association for Computational Linguistics, Nov. 2019.

[19] M. Corazza, M. Palmirani, F. M. T. Gatti, and S. Sapienza, "Monitoring sustainable development goals in european legislation using hybrid ai (in press)," in *ICEGOV '24: Proceedings of the 17th International Conference on Theory and Practice of Electronic Governance*, 2024.

[20] M. Corazza, F. M. T. Gatti, S. Sapienza, and M. Palmirani, "Hybrid classification of european legislation using sustainable development goals (in press)," in *Proceedings of the 23rd International Conference of the Italian Association for Artificial Intelligence*, 2024.

[21] N. Reimers and I. Gurevych, "Making monolingual sentence embeddings multilingual using knowledge distillation," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 11 2020.

[22] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The Berkeley FrameNet project," in *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, (Montreal, Quebec, Canada), pp. 86–90, Association for Computational Linguistics, Aug. 1998.

[23] A. Gangemi, M. Alam, L. Asprino, V. Presutti, and D. R. Recupero, "Framester: A wide coverage linguistic linked data hub," in *Knowledge Engineering and Knowledge Management* (E. Blomqvist, P. Ciancarini, F. Poggi, and F. Vitali, eds.), (Cham), pp. 239–254, Springer International Publishing, 2016.