A Text Similarity Approach for Automated Transposition Detection of European Union Directives

Rohan NANDA [a,1], Luigi DI CARO [a] and Guido BOELLA [a]

[a] *Department of Computer Science, University of Turin*

Abstract. This paper investigates the application of text similarity techniques to automatically detect the transposition of European Union (EU) directives into the national law. Currently, the European Commission (EC) resorts to time consuming and expensive manual methods like conformity checking studies and legal analysis for identifying national transposition measures. We utilize both lexical and semantic similarity techniques and supplement them with knowledge from EuroVoc to identify transpositions. We then evaluate our approach by comparing the results with the correlation tables (gold standard). Our results indicate that both similarity techniques proved to be effective in detecting transposition and such a system could be useful as an information retrieval system to identify the transposed provisions for both EC and legal professionals.

Keywords. Transposition, text similarity, EU legislation

1. Introduction

The effective application of European Union (EU) Law at the national level is important to achieve the objectives of the Treaties and smooth functioning of the EU. Member States are responsible for the correct and timely implementation of EU law. The European Commission (EC) is responsible for monitoring the national implementations to ensure their compliance with EU law. The Commission also has the responsibility to examine the application of EU law under the control of the Court of Justice of the European Union (CJEU) [5].

Among the three major EU legal instruments, we are interested to study the transposition of directives into the national law. This is because directives are not directly applicable and Member States need to pass legislations to implement them into national law. A directive comes into effect only after it has been transposed into national law by the Member States [7]. Transposition is therefore quite important for effective implementation of EU policies across the Member States. Delayed or incorrect transposition of directives hinder the EU policy objectives and the potential benefits they bring with them for European citizens [5]. Each directive is associated with a deadline by which Member States must implement national transposition measures which take into account

[1]Corresponding Author

the obligations of the Directive. In this paper, we will refer these transposition measures as national implementing measures (NIM).

Member States send the texts of NIMs to the Commission. The Commission then examines these texts to ensure that Member States have taken appropriate measures to achieve the objectives of the directive. The main goal of the Commission is to ensure that the NIMs are compliant with the directive. The Commission outsources the monitoring of NIMs to subcontractors and legal consulting firms. For instance, Milieu, a legal consultancy firm based in Brussels has been carrying out conformity checking studies of NIMs in different Member States since 2003 to study the transposition of several directives [1]. These studies carried out by a team of competent legal experts, comprise of legal analysis and concordance tables for studying the transposition of directive into the national law. Specifically, the concordance tables identify the specific provisions of NIMs which implement a particular article of the directive. Each row represents the transposition of a particular section of a directive into a specific provision of the NIM.

These legal measures undertaken by the Commission to monitor NIMs are time consuming and expensive [2]. For instance, to make a concordance table lawyers need to read several NIMs for each directive and then understand which provision of a particular NIM implements a particular article of the directive. This becomes more cumbersome for the Commission and lawyers doing cross-border or comparative legal research because they need to study the transposition in several Member States. Therefore, there is a need for a technological approach which utilizes text mining and natural language processing (NLP) techniques, to assist the Commission and legal professionals in studying and evaluating the transposition of directives at the national level.

In this paper, we investigate the application of cosine similarity (CS) and latent semantic analysis (LSA) to detect the transposition of EU directives into NIMs. The objective is to identify the specific provisions of NIMs which transpose a particular article of the directive. We further supplement both CS and LSA techniques by enriching directive and NIM provisions with knowledge from EuroVoc thesaurus. We study and compare the results from these techniques on 5 Directives and their corresponding NIMs by evaluating them with a gold standard (correlation tables, see Section 2). Our results indicate that cosine similarity technique achieves the best F-Score in all five directives.

We were restricted to study only five directives as we could find correlation tables for only certain NIMs in English (due to our lack of competency in other EU languages). We are aware of the United Kingdom's exit from the EU. However, as per Article 50 of the Treaty of European Union, the national acts adopted to transpose EU law would still remain valid until they are amended or repealed. This is a valid reason to include NIMs from the United Kingdom (UK) in our research.

The rest of the paper is organized as follows. In the next section, we will discuss the transposition process of EU directives briefly. Section 3 describes the our approach for automated transposition detection process. Section 4 discusses the results and analysis. Section 5 presents the conclusion and future work.

2. The EU Directive Transposition Process

In this section, we will discuss the transposition process of a directive briefly. We have identified the following three steps in this process:

1. Steps taken by the Commission to ensure effective transposition
2. Steps taken by the Commission to monitor NIMs
3. Pre-infringement and Infringement steps taken by the Commission

2.1. Steps taken by EC to ensure effective transposition

The Commission has developed a policy to promote compliance measures for effective transposition of directives. The Commission facilitates informal groups, called Networks which comprise of representatives from Member States responsible for implementation of specific EU laws. Networks intend to achieve correct implementation of directives by enhancing cooperation between representatives of Commission and Member States. The Commission also prepares Transposition and Implementation Plans (TIPs) to assist Member States in transposing and implementing the directive. These plans identify the risks to correct and timely transposition of the directive and also provide appropriate measures to counter those risks. Member states are obliged to plan appropriate measures to achieve the objectives of the directive and communicate them to Commission [5]. The Commission also provides Guidelines (non-legally binding documents) which clarify the Commission's stand on the implementation and interpretation requirements of specific provisions of directives'. Guidelines are also specified by the Commission when the implementations of specific provisions of a directive vary quite a lot across Member States [5].

2.2. Steps taken by EC to monitor NIMs

After Member States have adopted the NIMs, the Commission starts monitoring them to ensure the correct transposition of the directive [3]. Some of the steps taken by the Commission to monitor NIMs are discussed in this section.

Correlation Tables: Correlation tables identify the specific provisions of NIMs for each article of a directive. Correlation tables are made by Member States to ensure that the directive is fully transposed. These tables also assist the Commission to monitor the transposition of each provision of the directive. Correlation tables are generally not available to public as they are sent by Member States to the Commission as part of a confidential bilateral exchange. There is no agreed format or compulsory content for correlation tables. Therefore some Member States may provide detailed tables while others may just submit a list of some provisions [5]. In this paper, we evaluate our results by comparing them with the available correlation tables as a gold standard.

Conformity Checking: Conformity Checking studies are discussed in 3rd paragraph of Section 1. Some Member States have called for conformity checking to be made more transparent and also involvement of Member States when studies are drafted.

2.3. Pre-infringement and Infringement steps taken by Commission

Despite the steps taken by Commission to ensure effective transposition of Directives there are several instances of non-compliance. The Commission itself acknowledges that the compliance of EU law in Member States is still an unresolved issue [5]. The instances of non-compliance are identified while monitoring of NIMs (by Commission) and also by complaints from public, businesses and petitions from the European parliament [5]. When the Commission detects a possible infringement of the EU directive, it resorts to pre-infringement tools to achieve out-of-court settlements by establishing a partnership with the Member States. EU Pilot is a pre-infringement tool developed by the Commission for resolving issues of non-compliance of EU law by carrying out informal bilateral discussion with the Member States. If the pre-infringement EU pilot dialogue is unsuccessful, the Commission may launch formal infringement proceedings against the Member State under Article 258 of the Treat on the Functioning of the European Union (TFEU) [3] . The infringement may be launched against Member States under the following three conditions: failure to notify NIMs to Commission on time; non-conformity or non-compliance of national legislation with the requirements of EU directive; incorrect or no application of the directive.

3. Automated Transposition Detection of EU Directives

In this section, we describe our approach for automated transposition detection of EU directives (Figure 1). While transposing EU directives the Member States have a certain amount of discretion in the choice of methods. For example, in the United Kingdom, the national legislators generally use two broad approaches for transposing a directive. The first approach is called 'copy-out', where the NIM provision uses similar wording as that of the directive. In this approach, the NIM may also cross refer to the relevant directive provision. This motivated us to choose cosine similarity approach to identify such kind of transpositions. In the second approach, called 'elaboration', the provisions of NIM use language different from the wordings of the directive. This is done to clarify the meaning of NIM provision for legal or domestic policy reasons [4]. We decided to apply latent semantic analysis (LSA) to identify cases of 'elaboration' transposition as lexical methods like cosine similarity are deficient to capture semantic similarity of texts.

First of all, each group of directive and NIMs are stored in a format to adhere to the structure of their particular correlation table. This would enable us to compare our results with that of the correlation tables. This was carried out for each group of directive and NIM because correlation tables have no standard way of structuring the directives and NIMs. Sometimes they mention the transposed provisions for a complete article of the directive and sometimes only for a paragraph of the article. Thus, the articles and provisions of each Directive-NIM group are stored as a corpus.

The next step is pre-processing of the data. This consists of a number of steps to remove noise from the text. It is important to select only suitable and relevant terms as the performance of information retrieval systems is dependent on these pre-processing methods. The punctuation was removed and the text was converted to lowercase. Then tokenization was carried out to extract single words from the text. The stop words were removed using NLTK's corpus of stopwords for English. We used NLTK's part-of-speech tagger (POS tagger) to filter out nouns, verbs and adjectives from the remaining set of tokens. The manual reading of several directive and NIM provisions suggested that nouns, verbs and adjectives contained the most informative features.

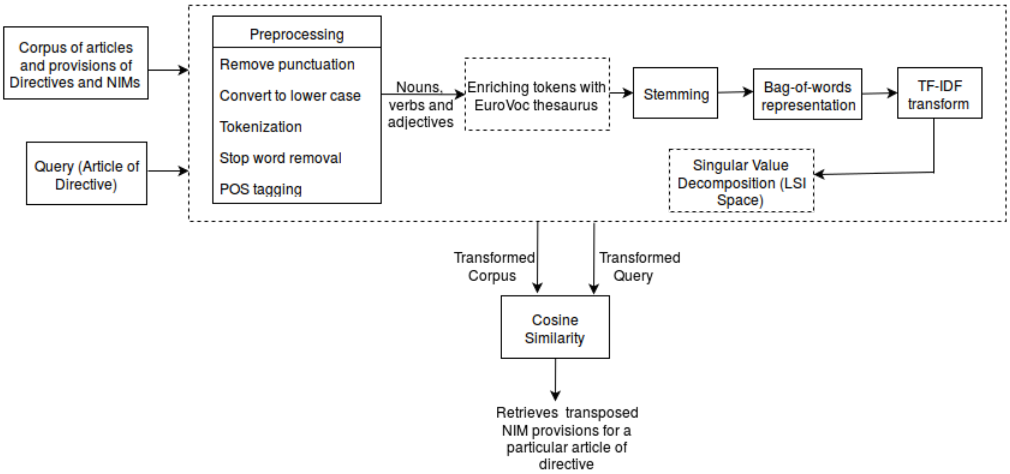

Figure 1. System Architecture for automated transposition detection

In the next step, the tokens obtained after pre-processing were enriched with the knowledge from EuroVoc [cite]. EuroVoc seemed to be an ideal choice because it not only contains documentary information of the EU institutions but also covers a wide range of fields representing both Community and national points of view. We utilized all the 127 microthesauri of EuroVoc to cover all possible domains for our corpus. We made use of the equivalence relationship between preferred and non-preferred terms in EuroVoc [6]. This relationship defines the indexing term when more than one terms represent the same concept. The EuroVoc microthesauri contains several instances of equivalence relationship. We used this relationship for synonyms and near-synonyms terms. For the implementation purpose, the 127 microthesauri was consolidated as a python dataframe. The list of tokens obtained after pre-processing was compared with each element of the microthesauri. In case of a match the token in our corpus was replaced by the preferred indexing term as per equivalence relationship of EuroVoc. Afterwards, the set of new tokens are stemmed to reduce the inflectional forms of words.

Each article (of directive) and provision (of NIM) of the corpus are then represented in a bag-of-words format. It is basically a list of each token and its count in a particular article or provision. Further, we applied Term Frequency-Inverse Document Frequency (tf-idf) weighting scheme to all the provisions and articles. The tf-idf measure evaluates the importance of each token, by offsetting its frequency in the provision with its frequency in the entire corpus.

We implemented latent semantic analysis (LSA) by applying Singular Value Decomposition (SVD) to the tf-idf provision-token matrix. SVD decomposes the tf-idf matrix into separate matrices which capture the similarity between tokens and between provisions across different dimensions in space. The relationship between tokens is represented in a reduced dimensional space to reduce noise and find latent relations between tokens and documents. LSA extracts the meaning of terms by analyzing the patterns in term usage among different provisions. This implies that two terms have similar meaning if they occur with same terms within different provisions. This property is helpful for retrieving information which maybe missed by other information retrieval systems.

The query (specific article of directive) is also transformed through the above steps. Since we want to evaluate the influence of adding knowledge from EuroVoc and also compare the performance of CS and LSA, therefore we divide the evaluation into four cases : (i) Cosine similarity (CS), (ii)Cosine similarity with EuroVoc, (iii) Latent semantic analysis (LSA), (iv) Latent semantic analysis with EuroVoc. It is important to note that dotted box of EuroVoc in Figure 2 is considered only for case (ii) and (iv). Similarly, the dotted box of SVD is considered only for case (iii) and (iv). For case (i) and (ii), cosine similarity is calculated as cosine of the angle between the transformed query vector (in tf-idf representation) and each provision vector in the corpus (also in tf-idf representation). The matching NIM provisions with similarity value greater or equal to the threshold value are retrieved by the system. Similarly for case (iii) and (iv), the similarity is measured by the cosine of the angle between the query vector and each provision vector in the reduced-dimensional space.

4. Results and Analysis

In this section, we study the results of transposition detection of 5 Directives using the techniques discussed in the previous section. We also observed that there were many cases where a particular article of a directive is transposed by multiple provisions of a NIM. In this situation, we must also consider the cases where the provisions retrieved by our system are a subset of the transposed provisions as per the correlation tables. For instance, let us consider that according to the correlation table, article A_n of Directive D is transposed by a set $P_m = \{p_1, p_3, p_6\}$ of provisions of NIM N_d. If our system retrieves all three provisions of set P_m then it is considered as an exact match. If our system retrieves a strict subset of provisions say just p_1, then it is considered as a partial match. However, if the system retrieves, say p_1 and also p_9(which is not a subset of set P_m), then it is not considered as a match. We evaluate our system for both exact and partial matches. We believe in legal information retrieval systems, even such partial matches of transposition are important because it is quite challenging for any information retrieval system to match the entire correct set. The real usefulness of these matches maybe be left to the lawyers and legal professionals using such a system.

We evaluate our system by computing the metrics: Precision, Recall and F-score (harmonic mean of precision and recall) for both exact and partial matches (partial matches are considered correct while computing precision and recall). We did not consider accuracy as we have very different number of true positives and true negatives resulting in an unbalanced dataset. In such cases, accuracy is not a fair metric for evaluation. We model and evaluate the system by considering these four cases for both partial and exact matches: (i) Cosine similarity (CS), (ii)Cosine similarity with EuroVoc, (iii) Latent semantic analysis (LSA), (iv) Latent semantic analysis with EuroVoc. In case (i) and case (iii) the tokens obtained after pre-processing are not enriched with EuroVoc.

Figure 2 shows the results of the transposition detection of all 5 directives. Directive1, Directive2, Directive3 and Directive4 are each transposed by 1 NIM. Directive 5 is transposed by 4 NIMs. We carried out experiments to determine the appropriate threshold level at which relevant transpositions were detected. A threshold of 0.35 and 0.40 was selected for cosine similarity and LSA respectively, for first four directives. For Directive5 threshold values are 0.25 (for cosine similarity) and 0.30 for LSA. This is due to the fact that Directive 5 has much higher number of total provisions (Table 1) as compared to other four directives. The same thresholds are used when both CS and LSA models are supplemented with knowledge from EuroVoc. The results in Figure 2 indicate no clear winner in terms of performance. However, we do make a few interesting observations. In terms of F-Score, CS achieves the best performance across all 5 cases of directive transposition. The performance of LSA was similar to CS in Directive1 and Directive2. However, it was outperformed by CS in Directive3, Directive4 and Directive5. This is because, LSA has been shown to perform well when a large corpus is available to extract the latent relationships between different terms with same meaning in different documents. LSA needs a large corpus to derive the semantics of a word by analyzing its relationship with other words[referphdthesis]. In a small corpus(like in our case),there is not enough text to extract the relationships between different words. Also the application of SVD causes some important features (needed for text similarity) to be lost, which results in higher false negatives (system is unable to detect the transposition, even though its present). This results in LSA systems achieving lower recall as compared to CS systems (as recall depends on false negatives). The same is observed through the graphs of Figure 2. In Directive3, Directive 4 and Directive 5 the recall of LSA is always lower than CS due to these higher false negatives. In Directive1 and Directive2 CS has the same number of false negatives as LSA resulting in similar recall. The low recall of LSA systems is compensated by the higher precision due to the trade-off. The precision values of LSA were equal to or higher than CS in Directive1, Directive2, Directive3 and Directive5. However, the precision values of CS are quite close to LSA. Overall in terms of all three metrics CS has the best performance due to higher recall and F-score and decent precision in all the directives.

In case of Directive5, there were several cases where an article was transposed by multiple provisions of different NIMs (NIM5, NIM6, NIM7 and NIM8). This made it pretty challenging for both CS and LSA to retrieve all the matching provisions from different NIMs, thus resulting in very few exact matches. Also the total number of NIM provisions in this case was 306 (Table 1), much higher than the other four cases. We believe handling such cases of multiple transposition is a major challenge for our research ahead.

We also observe from the results that the addition of knowledge from EuroVoc does not improve the performance of both CS and LSA. In most cases, the precision and recall values of CS with EuroVoc and LSA with EuroVoc are same as those of CS and LSA respectively. We found that in our corpus there were several provisions of both directives and NIMs where some terms were enriched from EuroVoc thesaurus. However, the terms added from EuroVoc to a particular article of a directive did not match any terms present in the transposing provision and vice versa. This is why the knowledge from EuroVoc does not help to improve the existing CS and LSA results. For instance, in an article of directive, the terms 'worker', 'evaluation','health' and 'risk' are replaced by EuroVoc terms, 'labour force', 'evaluation method', 'health policy' and 'insured risk' respectively. The provision transposing this article does not contain any of these EuroVoc terms. Also, in the provision, the terms 'employee' and 'evaluation' are replaced by 'wage earner' and 'evaluation method' respectively. The term 'evaluation' was already common between both article and provision. Its replacement by 'evaluation method' does not make any difference with regard to text similarity. The term 'wage earner' is not present in the article. Thus, the knowledge from EuroVoc did not help to improve the

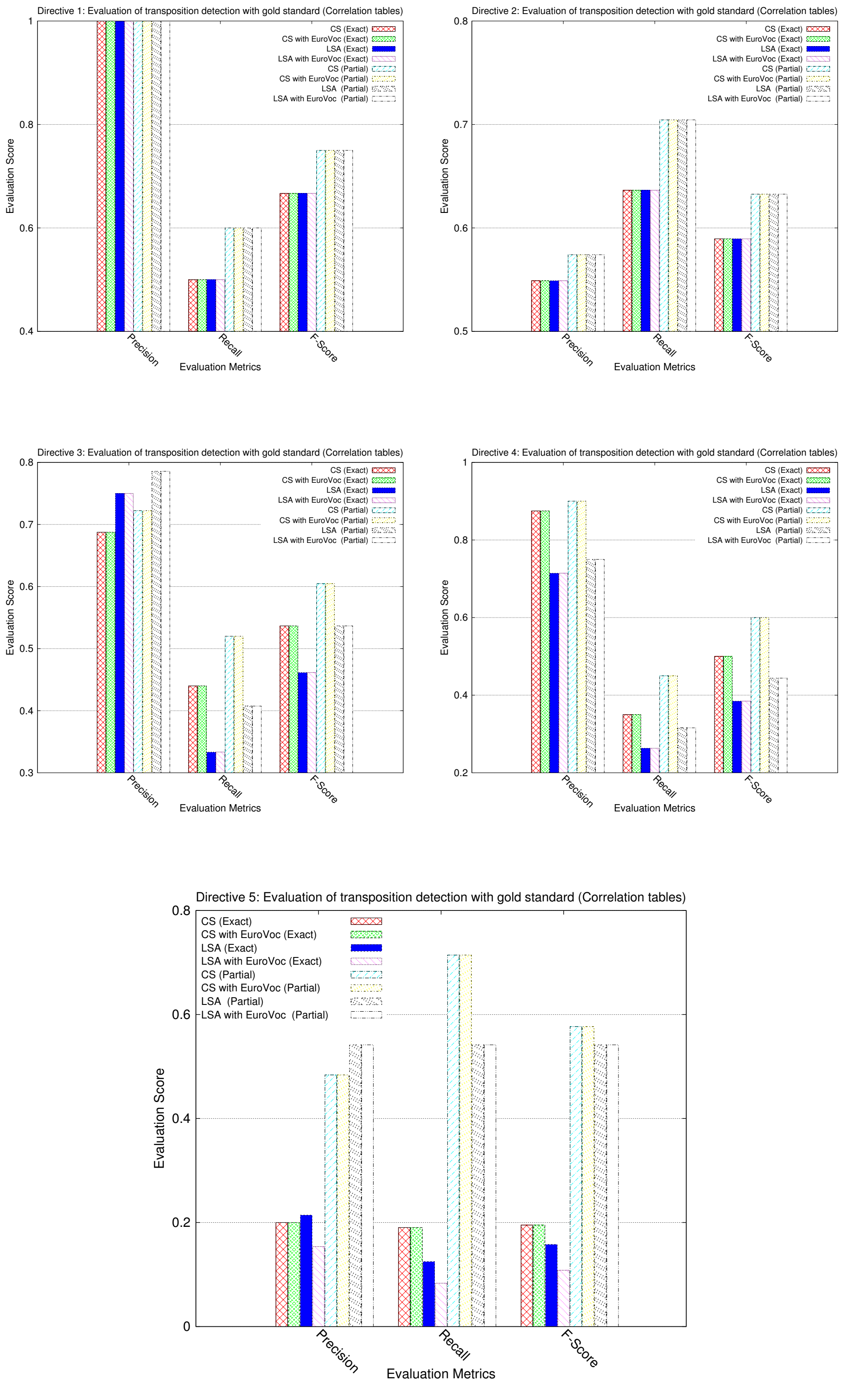

Figure 2. Evaluation of transposition detection with gold standard (Correlation tables)

Table 1. Statistics of Directives and NIMs under consideration

Directive-NIM	Number of provisions of Directive	Number of provisions of NIM/s	Total number of provisions
(Directive1, NIM1)	31	17	48
(Directive2, NIM2)	68	88	156
(Directive3, NIM3)	37	67	104
(Directive4, NIM4)	24	50	74
(Directive5, NIM5, NIM6, NIM7, NIM8)	41	306	347

results of text similarity in the present corpus of directives and NIMs. In the future, it would be interesting to investigate the knowledge from EuroVoc on other directive and NIMS.

In summary, our results indicate the fact that both CS and LSA similarity techniques are effective in detecting transposition of EU directives. There was no notable performance improvement by incorporating the knowledge from EuroVoC. The performance of the system is subject to the requirements of legal professionals studying the transposition. In majority of the cases, LSA achieves higher precision (except in Directive4). While CS always achieves higher recall (except Directive1 and Directive2, where they have same recall). In terms of F-score, CS outperforms LSA (except Directive1 and Directive2, where they have same F-score).

5. Conclusion and Future Work

This paper presented the application of text similarity approaches to detect the transposition of EU directives into the national law of Member States. We discussed the transposition process and identified the need for a technological approach for monitoring NIMs to achieve the objectives of EU directives. We investigated the application of both lexical (cosine similarity) and semantic (latent semantic analysis) similarity techniques in transposition detection. External knowledge from EuroVoc thesaurus was also used to supplement both similarity techniques. We evaluated the techniques by comparing their results with the correlation tables prepared by the respective Member States while transposing the directive. Our results indicate that both cosine similarity and latent semantic analysis were effective in detecting transposition. The overall performance of cosine similarity was superior to LSA in terms of F-score. Our initial experiments indicate that such systems can be useful for legal information retrieval to assist the Commission and legal professionals.

Our future work will comprise of using both n-gram models and quality phrase extraction to improve upon our current work. We believe mining relevant phrases from legal text could be highly useful in detecting transposition. We also intend to study the transposition detection for a particular directive in different Member States. This would help us to characterize and compare how different Member States transpose the same directive with respect to their legal or domestic policy. We are also interested in developing a statistical language-independent model for transposition detection of directives across several Member States. This is possible due to the availability of the directives in all EU languages. However, translation will be an issue for interpretation for cross-border legal research.

Acknowledgements

Research presented in this paper is conducted as a PhD research at the University of Turin, within the Erasmus Mundus Joint International Doctoral (Ph.D.) programme in Law, Science and Technology. Luigi Di Caro has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 690974 for the project "MIREL: MIning and REasoning with Legal texts".

References

[1] Conformity checking, milieu. Electronic, accessed 25 May 2016, Retrieved from http://www.milieu.be/index.php?page=conformity-checking.

[2] Giuseppe Ciavarini Azzi. The slow march of european legislation: The implementation of directives. *European integration after Amsterdam: Institutional dynamics and prospects for democracy*, pages 52–67, 2000.

[3] European Commission. Monitoring the application of union law, 2014.

[4] HM Government. Transposition guidance: How to implement european directives effectively, April 2013.

[5] Mariolina Eliantonio Marta Ballesteros, Rostane Mehdi and Damir Petrovic. Tools for ensuring implementation and application of eu law and evaluation of their effectiveness, July 2013.

[6] Luis Polo Paredes, JM Rodrıguez, and Emilio Rubiera Azcona. Promoting government controlled vocabularies for the semantic web: the eurovoc thesaurus and the cpv product classification system. *Semantic Interoperability in the European Digital Library*, page 111, 2008.

[7] Bernard Steunenberg and Mark Rhinard. The transposition of european law in eu member states: between process and politics. *European Political Science Review*, 2(3):495–520, 2010.