
Climate Policy Tracker: Pipeline for automated analysis of public climate policies

Artur Żółkowski

Warsaw University of Technology
artur.zolkowski.stud@pw.edu.pl

Mateusz Krzyżński

Warsaw University of Technology
mateusz.krzyzinski.stud@pw.edu.pl

Piotr Wilczyński

Warsaw University of Technology
piotr.wilczynski.stud@pw.edu.pl

Stanisław Giziński

University of Warsaw
s.gizinski@student.uw.edu.pl

Emilia Wiśnios*

University of Warsaw
e.wisnios@student.uw.edu.pl

Bartosz Pieliński

University of Warsaw
b.pielinski@uw.edu.pl

Julian Sienkiewicz

Warsaw University of Technology
julian.sienkiewicz@pw.edu.pl

Przemysław Biecek

Warsaw University of Technology
przemyslaw.biecek@pw.edu.pl

Abstract

The number of standardized policy documents regarding climate policy and their publication frequency is significantly increasing. The documents are long and tedious for manual analysis, especially for policy experts, lawmakers, and citizens who lack access or domain expertise to utilize data analytics tools. Potential consequences of such a situation include reduced citizen governance and involvement in climate policies and an overall surge in analytics costs, rendering less accessibility for the public. In this work, we use a Latent Dirichlet Allocation-based pipeline for the automatic summarization and analysis of 10-years of national energy and climate plans (NECPs) for the period from 2021 to 2030, established by 27 Member States of the European Union. We focus on analyzing policy framing, the language used to describe specific issues, to detect essential nuances in the way governments frame their climate policies and achieve climate goals. The methods leverage topic modeling and clustering for the comparative analysis of policy documents across different countries. It allows for easier integration in potential user-friendly applications for the development of theories and processes of climate policy. This would further lead to better citizen governance and engagement over climate policies and public policy research.

1 Introduction

Machine learning (ML) systems have already been used to analyze climate change [15]. They are able to follow climate change itself as well as develop and evaluate solutions preventing or even reversing the process. ML in the form of Natural Language Processing (NLP) has also started to be implemented to study discussions around climate change and climate policies [18]. It has also been

*Corresponding author

used to research how governments around the world reacted to climate change [11, 19]. The system presented in this paper is a part of the latter strand of research. It aims to develop NLP methods dedicated to analyzing specific types of policy documents – highly standardized national climate strategies and reports.

In recent years, as the danger of climate change has become acknowledged by public opinion, policymakers, and politicians, the number of policy documents dedicated to climate policy has been steadily increasing (Appendix A). However, after the introduction of the Paris Agreement, a specific group of documents has started to be produced by national governments of the Member States of the European Union (EU) – climate strategies and reports on implementing these strategies. The Agreement (Article 13) established an Enhanced Transparency Framework design to provide guidance for regular reporting on national climate policies. As a consequence, the Agreement has stimulated the development of policy documents having similar structures as they address previously agreed issues. Those documents have been analyzed so far mainly through the perspective of climate policy indicators showing commitments of national governments to limit climate change [8, 6].

In this study, we propose the pipeline (Figure 3) that focuses on different aspects of these documents – the way they frame climate policy. The issue framing has been found to influence policy dynamics over the long run [5]. The similar structure of the documents allows us to compare the policy framing in different dimensions separately, which allows for a more fine-grained analysis of policy frames.

The pipeline was created based on two assumptions: (1) the number and publishing frequency of standardized documents regarding climate policy will soon increase significantly; (2) NLP methods allow a relatively quick and comprehensive summary of these documents and their comparison. They can also reveal important nuances in how governments frame their climate policies.

The main contributions of our work are as follows:

1. We propose a pipeline for analyzing similarly structured documents regarding climate public policy, leveraging Natural Language Processing.
2. We provide an easy-to-use web application for exploring structured results of our pipeline.

2 Data

We analyze the 10-year national energy and climate plans (NECPs) for the period from 2021 to 2030 established by 27 Member States of the EU. These documents contain information on how individual countries want to relate to the EU's critical energy and climate policy dimensions. Each plan was published in English and the country's national language on the EU's official website. Only documents written in English were considered in this study. All of the data and information used in the paper, either as a subject of the analysis or to support its claims, are publicly available on this website.

To ensure the parallelism of the plans between the Member States, NECPs had to follow a unified structure facilitating comparison and aggregation. According to the template, five Energy Union dimensions and five sections should be distinguished. See Appendix C for more details.

We performed preprocessing of all documents; a detailed description of this process is available in Appendix D.

3 Methodology

Topic modeling For modeling topics, we use the Latent Dirichlet Allocation (LDA) model [1] as it assumes each document is a finite mixture over an underlying set of topics, which is in line with the assumptions of our study.

Modeling was carried out using the Gensim package [13] separately on seven sub-datasets – five defined by Energy Union dimensions and two related to subsections where these dimensions were not distinguished. In addition, words were filtered to include only those appearing in at least four analyzed texts, which translates into the presence of words in documents submitted by at least two different countries. Such filtering ensures that texts are not grouped based upon expressions unique to each country yet unrelated to the document's topic.

We assumed that prior over topic-word distribution may be asymmetric and it was matched automatically based on the subcorpora. For selecting the parameters describing the symmetric prior over document-topic distribution (α parameter in Gensim LDA), we performed two grid searches. We observed that choosing the correct number of topics is more important than choosing the optimal α . However, models gave more intuitive and interpretable results for $\alpha \in \{50, 100, 150, 200, 250\}$. The performance of such parameter values is probably explained by the fact that considered documents are similar in terms of topics and differ only in nuances. We assessed the quality of topic modeling using two coherence measures, CV [14] and UMass [9]. Moreover, we manually analyzed and validated the results, which is necessary to obtain semantically meaningful topics [3].

Topic interpretation One of the most challenging aspects of topic modeling is the characterization of the topics learned by the used model [12]. Topics obtained using LDA do not have an interpretable description, but usually, the most common words are used to describe topics. We used the 25 most popular words for automatic topic naming using GPT-3 [2].

However, such a representation does not give the complete picture and can even be misleading, as each topic is a distribution over *the entire* vocabulary. Therefore, we used the LDAvis tool [17] to characterize and interpret the topics. This interactive method provides a global overview of the topics and illustrates how they differ. At the same time, it allows for a scrutinous examination of the terms most associated with each topic. Precisely, two metrics were used in the topic analysis and description process: the saliency metric [4] used to identify the most informative words, and the relevance metric [17] used to rank words within topics and identify the most relevant terms. Based on this exploration and automatic suggestions, we made the final naming of topics.

Comparison of policy frames mixture among countries In the analysis, we used similarity measures computed in terms of similarity of distributions described directly for the LDA model [16]. We select an information radius measure (also known as Jensen-Shannon divergence).

We examined the dependencies between the topic mixtures of individual countries by grouping them with a hierarchical clustering algorithm. An agglomerative approach with an average linkage criterion was used. The choice of this clustering procedure was due to the fact that it produces a complete dendrogram showing the range of levels at which countries are related. Thus, it presents a more holistic picture of the agendas' underlying structure than other methods that return a single version of clustering. In addition, we used the t-distributed stochastic neighbor embedding (t-SNE) [21] to visualize the results of clustering.

We examined how the clusters of countries differ across dimensions. This has a twofold purpose. First, it allows us to validate the clustering and topics against the expert knowledge of well-established countries' alliances, such as the Visegrad Group. Secondly, it allows us to pinpoint other, non-obvious groups of countries with matching or similar policy framings. This creates an opportunity to provide new hypotheses for further geopolitical and climate studies.

4 Results

Web application for exploration of the results The result of topic modeling are the distributions of topics in each analyzed text. Matrices composed of these vectors for each of the seven models are the direct result of the modeling process, and they can be analyzed in a variety of ways. Thus, we have prepared a demo version of the web application that enables in-depth, interactive exploration on many levels, see Appendix E. First, it is possible to examine the distribution of topics for the selected country and compare them between sections (Figure 5) or juxtapose them with the

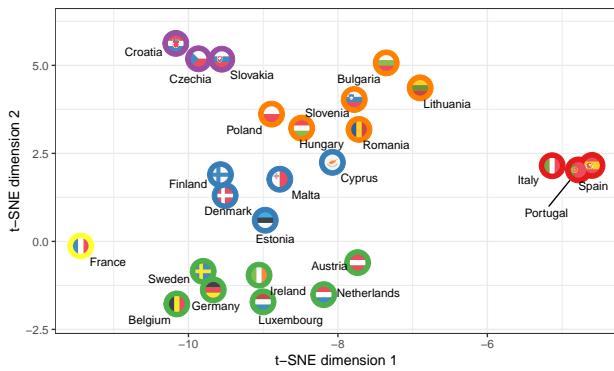


Figure 1: Visualization of selected country clustering for the *Decarbonisation* dimension. Positions in two-dimensional space are based on t-SNE embedding.

overall results for other countries (Figure 6). Further, the application allows for clustering based on the found representation (Figure 1). The interpretability of the results is ensured by analysis that allows the exploration of particular topics through their keywords (Table 1). Another feature is the ability to compare the values for topics with external variables (Figure 8).

Differences between sections across dimensions The modeling results can also be used for other, more sophisticated analyses, such as comparisons between *National objective and targets* and *Policies and measures* sections to detect inconsistencies in agendas. It turns out that some states have more discrepancies and that this phenomenon is more typical for some dimensions, see Appendix G.

Policy frames focus for countries We find there are 5 policy frames for *Decarbonization* and *Energy efficiency*, 3 policy frames for *Energy security*, *Internal market*, and *R&I and Competitiveness*. Overall, our system detects 19 policy frames across all dimensions. A full list of policy frames is available in Appendix F. By analyzing the topics' distribution, we conclude which country focuses on which policy frame in each of the dimensions. As an example, we can see that Bulgaria is the country that devotes the most parts of the *Decarbonization* section to the *Greenhouse gas emission*, (.65 topic probability). Netherlands is an outlier in the topic *Sustainable transport*, with .45 topic probability (Figure 9).

Co-occurrence in clusters across dimensions We have analyzed countries' co-occurrences in the same clusters. Three pairs of countries are in the same groups in all five dimensions mentioned earlier: Netherlands and Austria, Malta and Cyprus, Estonia and Finland. It may suggest that these countries frame issues similarly in all areas. France seems to be the most isolated country, co-occurrence only once with Belgium, Spain, Romania, and Slovenia. One reason explaining this could be that France has the most nuclear energy in their energy mix among other countries.

We have also checked some insights regarding countries' alliances, e.g. the Visegrad Group, consisting of Czechia, Hungary, Poland, and Slovakia (Figure 12a). In our analysis, there is one dimension, *Energy efficiency*, in which all countries belong to the same cluster. In others, they are split into two clusters. What is worth seeing is that Poland and Hungary jointly appear in *Decarbonization*, *Energy efficiency*, and *Energy security* dimensions. Czechia and Slovakia are together in *Decarbonization*, *Energy efficiency*, *R&I and Competitiveness*.

Another interesting group is the Nordic countries (Figure 12b). In *Internal market* all countries are in one cluster. In other categories, Denmark and Finland differ only in *R&I and Competitiveness*. Figure 11 shows the full co-occurrence matrix.

5 Conclusion

Climate change is one of the most critical global threats. Creating public policies regarding this matter is essential in counteracting this problem. However, the creation of the policy itself is insufficient without the actions undertaken by the citizens and policymakers as a result. This is why our system allows for quick analysis of the focus of each national plan, comparing multiple countries across different dimensions, and detecting alignment between countries in policy framing. We believe that this could benefit the public understanding of the landscape of climate plans in the EU.

Our method and pipeline can also be used to analyze other kinds of similarly structured policy documents, not only related to climate change. One of the documents that can be run through our framework in the future is a progress report produced by each Member State every two years regarding the changes they introduce as a result of their NECP.

This work can be seen as a first step in developing methods for automated analysis of climate policies. In the future, such methods could allow tracking of how those policies change over time and how they relate to changes in the economic factors of countries, such as the energy mix.

Acknowledgments and Disclosure of Funding

We would like to thank Zuzanna Kwiatkowska for valuable discussions on the structure of the pipeline and the way of communicating the results and Ibrahim El-chami for mentoring and helpful comments.

Research was funded by (POB Cybersecurity and Data Science) of Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme. This research was carried out with the support of the Laboratory of Bioinformatics and Computational Genomics and the High Performance Computing Center of the Faculty of Mathematics and Information Science Warsaw University of Technology under computational grant number A-22-10.

References

- [1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 2003.
- [2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- [3] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-graber, and David Blei. Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc., 2009.
- [4] Jason Chuang, Christopher D Manning, and Jeffrey Heer. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the international working conference on advanced visual interfaces*, pages 74–77, 2012.
- [5] Falk Daviter. Introduction. In Falk Daviter, editor, *Policy Framing in the European Union*, Palgrave Studies in European Union Politics, pages 1–25. Palgrave Macmillan UK, London, 2011.
- [6] Matthias Duwe, Eike K Velten, Nick Evans, Mona Freundt, Julien Pestiaux, Benoit Martin c, and Pascal Vermeulen. Planning for net-zero: Assessing the draft national energy and climate plans. Technical report, Ecologic Institute and Climact, European Climate Foundation, 2019.
- [7] Matthew Honnibal and Ines Montani. spaCy 3: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. 2017.
- [8] Harriet Mackaill-Hill, Dora Petroula, Markus Trilling, Veerle Dossche, Elif Gunduzyeli, Joerg Muehlenhoff, Goksen Sahin, Bárbara Mauricio, and Filipa Alves. Opportunities and gaps in the final national energy and climate plans. Technical report, Climate Action Network (CAN) Europe, ZERO – Association for the Sustainability of the Earth System, 2020.
- [9] David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, page 262–272, USA, 2011. Association for Computational Linguistics.
- [10] European Parliament. Regulation on the Governance of the Energy Union and Climate Action (EU) 2018/1999. *OJ*, 2018.
- [11] Jordi Planas, Daniel Firebanks-Quevedo, Galina Naydenova, Ramansh Sharma, Cristina Taylor, Kathleen Buckingham, and Rong Fang. Beyond modeling: Nlp pipeline for efficient environmental policy analysis, 2022.
- [12] Daniel Ramage, Evan Rosen, Jason Chuang, Christopher D. Manning, and Daniel A. McFarland. Topic modeling for the social sciences. In *Workshop on Applications for Topic Models, NIPS*, 2009.
- [13] Radim Rehurek and Petr Sojka. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.
- [14] Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. *WSDM '15*, page 399–408, New York, NY, USA, 2015. Association for Computing Machinery.

- [15] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Sasha Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla P. Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning. *ACM Comput. Surv.*, 55(2), feb 2022.
- [16] Vasile Rus, Nobal Niraula, and Rajendra Banjade. Similarity measures based on latent dirichlet allocation. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 459–470, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [17] Carson Sievert and Kenneth Shirley. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 63–70, Baltimore, Maryland, USA, June 2014. Association for Computational Linguistics.
- [18] Manfred Stede and Ronny Patz. The climate change debate and natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 8–18, Online, August 2021. Association for Computational Linguistics.
- [19] Pradip Swarnakar and Ashutosh Modi. Nlp for climate policy: Creating a knowledge platform for holistic and effective climate action, 2021.
- [20] Martin Szomszor and Euan Adie. Overton – a bibliometric database of policy document citations, 2022.
- [21] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

A Number of public policies regarding climate change

There are multiple databases containing public policies. The biggest of them is *Overton.io* database [20]. We have used query *climate change* to check the number of public policies available. We got 322012 documents from 171 countries. Most of them was published by USA (81643), Intergovernmental Organizations (61884), United Kingdom (55393), and EU (40480). We have counted policies published each year, the results are shown in Figure 2. As we can see, the number of documents is rapidly growing.

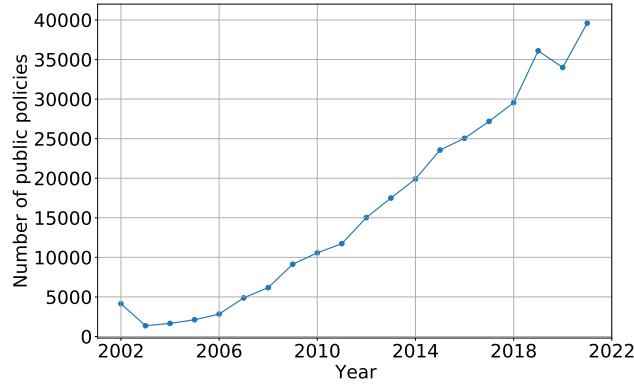


Figure 2: Number of public policies regarding climate change in the *Overton.io* database [20].

B Pipeline structure

We define the proposed pipeline as a set of procedures and methods that are used to generate results, ranging from text extraction from documents to processing of data from trained models that can be explored using domain knowledge and analyzed using a web application. A simplified diagram of our *Climate Policy Radar* pipeline is presented in Figure 3.

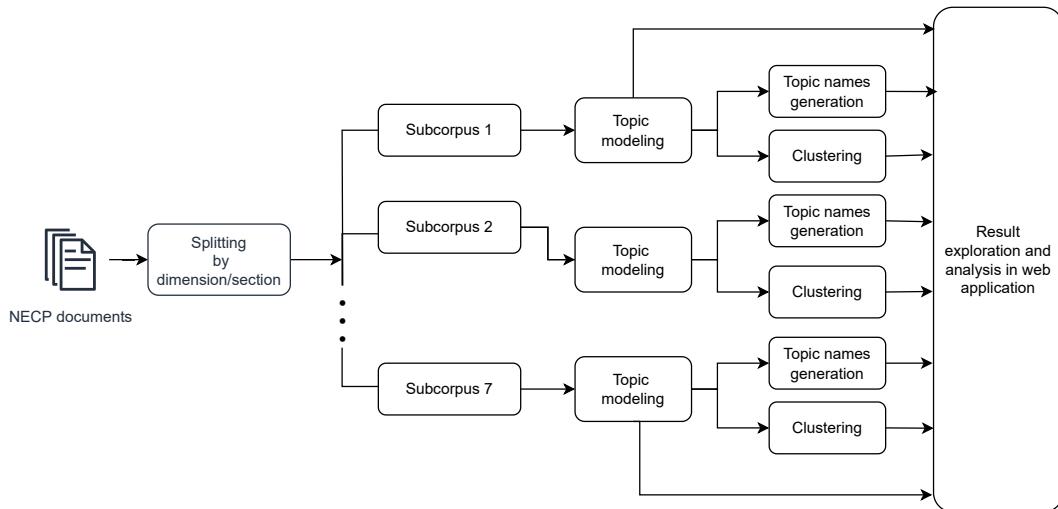


Figure 3: Diagram of the proposed pipeline for automated analysis of national energy and climate plans with the use of Natural Language Processing tools.

C Structure of National energy and climate plans

The creation of a template for National energy and climate plans has been guaranteed by the Regulation of the European Parliament and of the Council on the Governance of the Energy Union and Climate Action [10]. The diagram of the structure of the documents is presented in Figure 4.

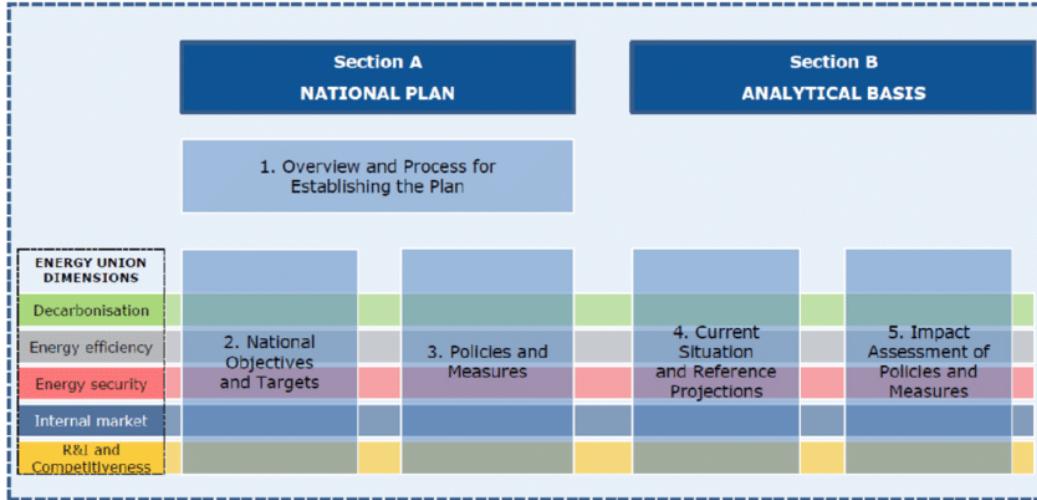


Figure 4: Structure of National Energy and Climate Plans provided by the European Commission.

There are five Energy Union dimensions:

1. *Decarbonisation*,
2. *Energy efficiency*,
3. *Energy security*,
4. *Internal market*,
5. *R&I and Competitiveness*.

There are also five sections:

1. *Overview and Process for Establishing the Plan*,
2. *National Objectives and Targets*,
3. *Policies and Measures*,
4. *Current Situation and Reference Projections*,
5. *Impact Assessment of Planned Policies and Measures*.

The dimensions are well separated in sections (2)-(4), while the remaining sections constitute separate entities. It should be noted that there are discrepancies in the documents in relation to this structure. In particular, in the *Impact Assessment of Policies and Measures* section, dimensions have not been explicitly separated.

D Text preparation

We used a well-defined structure of documents to create the corpus. First, we created a script that extracts the texts based on the table of contents. Then, since not every country strictly adhered to the imposed structure, we manually validated the results and corrected errors in problematic files. Ultimately, by creating separate records for each dimension in each subsection, the corpus of 453 documents were constructed. We applied a set of preprocessing steps to the prepared corpus with tools and models available in the spaCy package [7]. We performed tokenization, lemmatization and added bigrams and trigrams to unigrams with the aim to improve the topic modeling and streamline topic interpretation.

E Web application for exploration of the results

We present the process of examining the results obtained for *Decarbonisation* dimension since it is, on average, the longest and most comprehensive part of the documents. For this dimension, we obtained five notably contrasting main topics presented in Table 1 along with the corresponding keywords.

Table 1: Topics and keywords for the *Decarbonisation* dimension.

Topic name	Keywords
<i>Sustainable transport</i>	support, plan, transport, vehicle, government, reduce, public transport, car, green, mobility
<i>National energy policy</i>	promote, plan, national, system, integrated national energy, policy measure, objective
<i>All developments</i>	development, electricity, preparation, production, implementation, research development
<i>GHG emissions</i>	total, GHG emission, projection, gross final consumption, share, GHG, decrease
<i>Renewable energy sources</i>	target, biomass, wind, solar, heat pump, bioenergy, geothermal, trajectory

The web application enables a detailed analysis of the results for a single selected Member State. Examples of visualizations facilitating the analysis are depicted in Figures 5 and 6 for Finland.

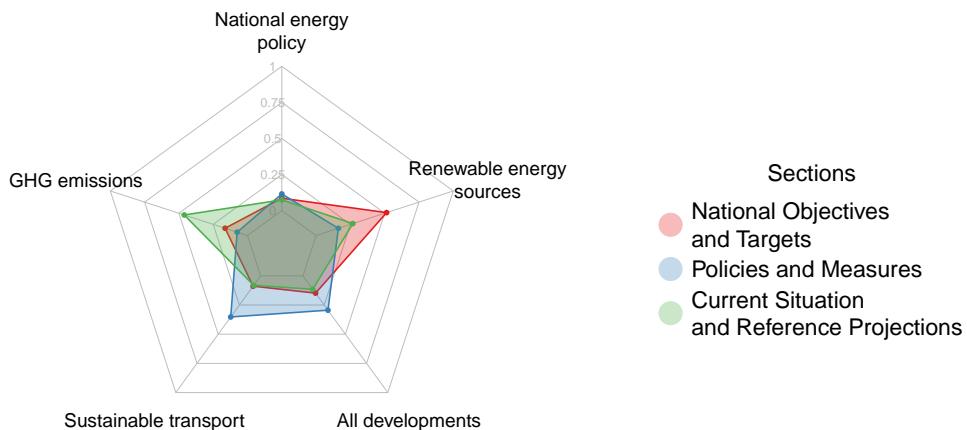


Figure 5: Topic distributions by section in the *Decarbonisation* dimension for Finland.

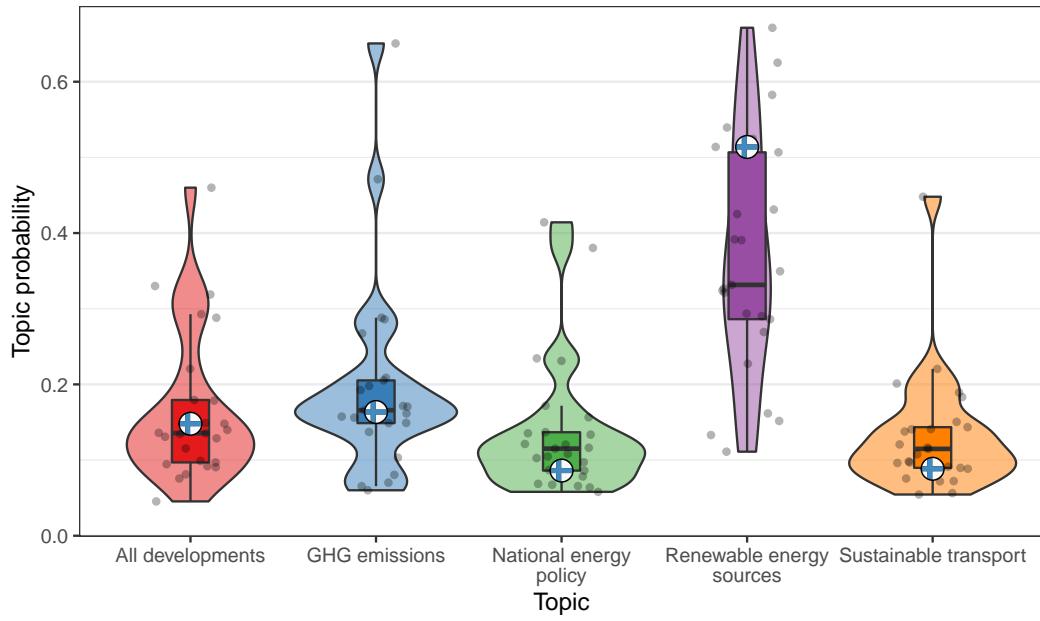


Figure 6: Finland's topic distribution compared to other countries in the *Decarbonisation* dimension in *National Objectives and Targets* section.

The web application also allows for grouping countries with similar topic distributions in a selected dimension or section. Hierarchical clustering (selected for analysis in this study), HDBSCAN, or K-means algorithm can be used for this purpose. For dimensions (covered in three sections), clusterings are based on bound and normalized topic distributions. The result for the *Decarbonization* dimension in the form country grouping is shown in Figure 1. Distance threshold implying specific grouping was selected on the basis of the analysis of dendrogram and matrix of similarities between countries' agendas which are presented in Figure 7.

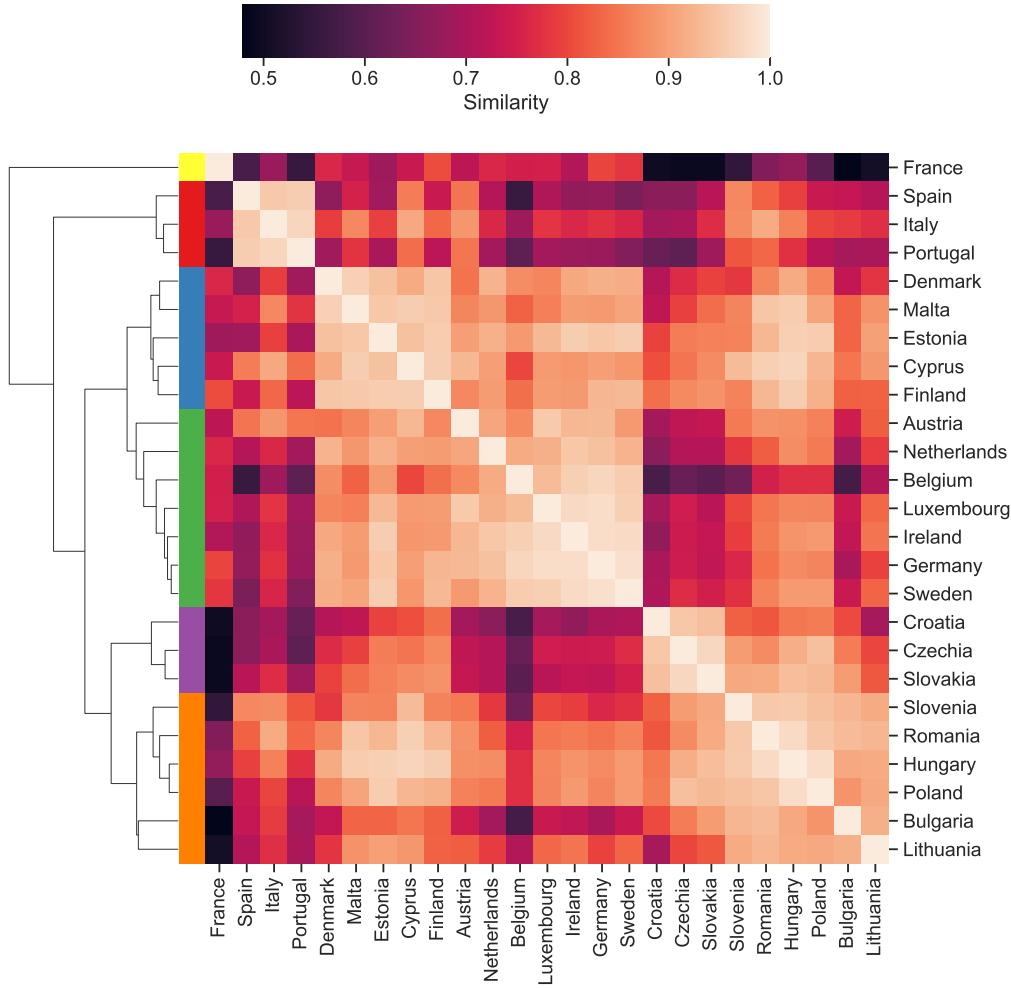


Figure 7: Matrix of similarities between the bound distributions of topics for the *Decarbonization* dimension. On the left side of the matrix, there is a dendrogram that visualizes the result of a hierarchical clustering calculation. The colors in the column preceding the matrix indicate the clusters in which individual countries appear on the basis of the selected distance threshold (country names are in the rows), see Figure 1.

One of the key system features is the possibility of comparing the obtained topics (policy frames) with external variables in order to detect/clarify the dependencies. We examined the correlations with objectives, targets, and contributions under the Governance Regulation [10].

The values were drawn from assessments of NECPs made by the European Commission. National targets and contributions values are precisely specified for 2020 and 2030. There are five different distinguished categories:

- **greenhouse** – a binding target for greenhouse gas emissions reductions compared to 2005 under the Effort Sharing Regulation (ESR);
- **renewable energy** – national target/contribution for renewable energy: share of energy from renewable sources in the gross final consumption of energy;
- **primary energy** – national contribution for energy efficiency: primary energy consumption;
- **final energy** – national contribution for energy efficiency: final energy consumption;
- **electricity** – level of electricity interconnectivity.

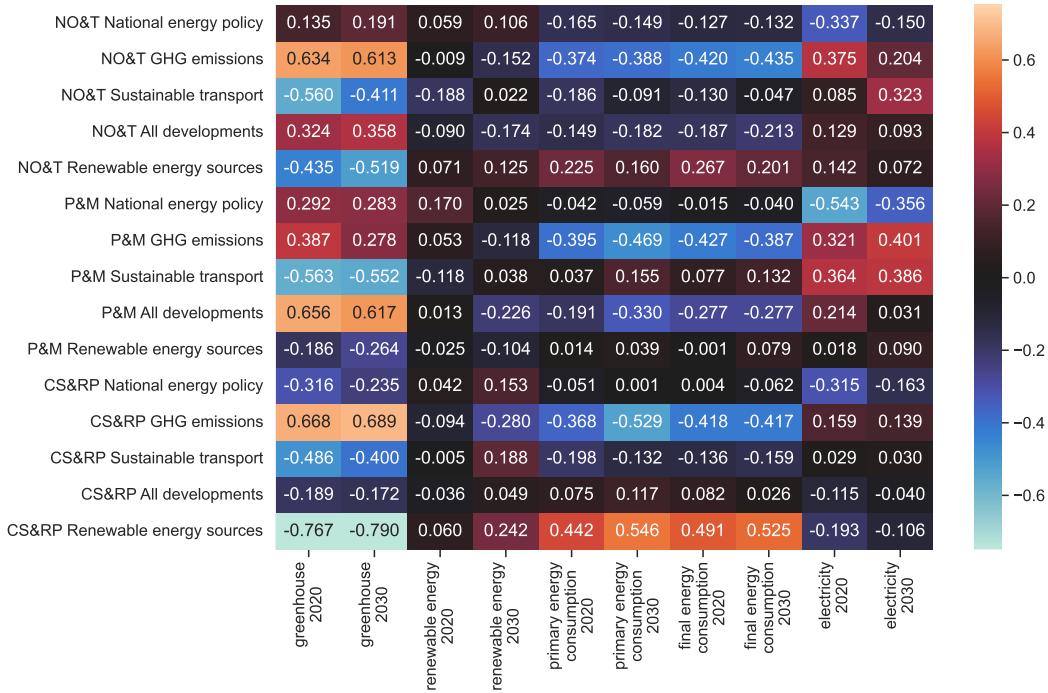


Figure 8: Spearman correlations between topic distributions for the *Decarbonisation* dimension in three sections (vertical axis; the section names are abbreviated for simplicity) and objectives extracted from the European Commission's individual assessments of NECPs (horizontal axis).

F Policy frames

In the process of topic modeling, we have obtained the topics that may be seen as policy frames for Energy Union dimensions (Table 2).

Table 2: Topics obtained for each Energy Union dimension.

Dimension	Topics
<i>Decarbonisation</i>	1.1 Sustainable transport 1.2 GHG emissions 1.3 National energy policy 1.4 Renewable energy sources 1.5 All developments
<i>Energy efficiency</i>	2.1 Financial aspects 2.2 Energy efficiency scenarios 2.3 Saving energy 2.4 Energy renovation 2.5 Primary and final energy
<i>Energy security</i>	3.1 Fossil fuels supply chain 3.2 Energy import 3.3 Capacity market development
<i>Internal market</i>	4.1 Natural gas 4.2 Development and support 4.3 Energy poverty alleviation
<i>R&I and Competitiveness</i>	5.1 Research innovations 5.2 Financial aspects 5.3 Initiatives and programmes

Topic names for the final analysis were given manually, based on obtained keywords (see Appendix E). Getting those topics was essential to see in which areas countries differ. The plot below shows an example of such analysis for the *Decarbonisation* dimension.

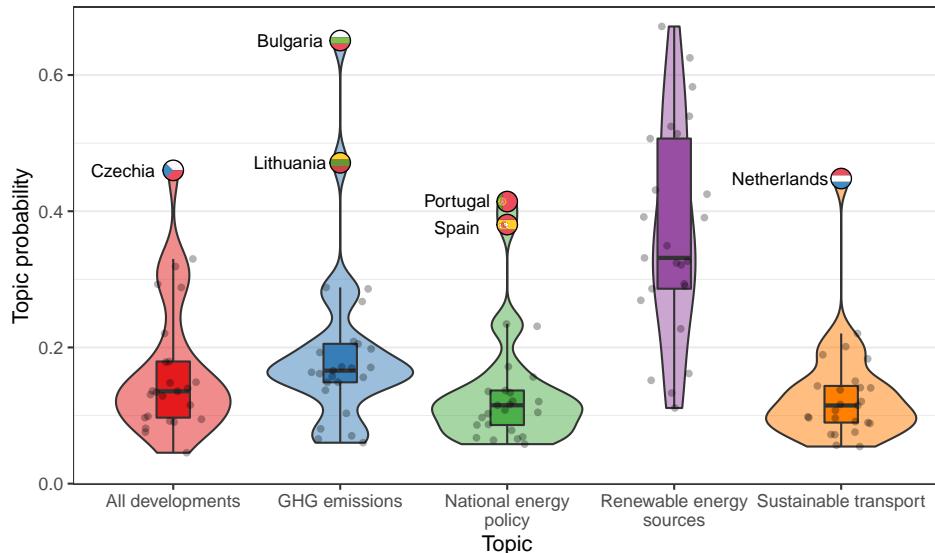


Figure 9: Topic distributions for countries in the *Decarbonisation* dimension in the *National Objectives and Targets* section. Countries that could be recognized as outliers are marked with their flags and captioned.

G Differences between sections across dimensions

For the five Energy Union dimensions mentioned in all sections, we check the differences in the policy agendas for individual countries between the *National Objectives and Targets* and *Policies and Measures* sections. This analysis aims to show that there are notable disparities between the sections describing objectives and policies intended for achieving them. We calculate dissimilarities using the information radius measure applied to pairs of topic distributions for the corresponding documents from the corpus. For each dimension, we divide the obtained measures by the number of policy frames to account for their variable cardinality.

We can see that there are countries where these differences are more pronounced, e.g. dissimilarities for Belgium in 4 out of 5 dimensions are higher than the average for all countries. The dimension with the greatest inconsistencies between the sections is *Decarbonization*. On the other hand, on average, framing policies is the most consistent in the area related to *R&I and Competitiveness*.

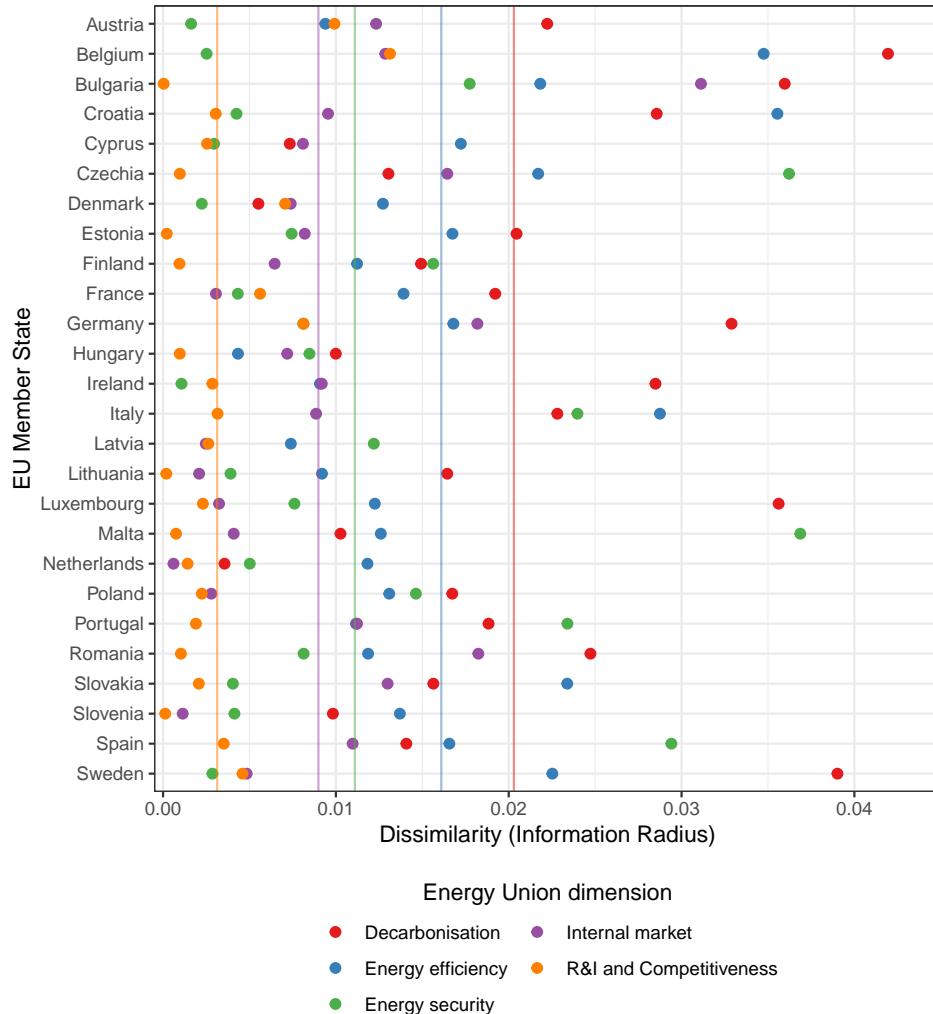


Figure 10: Normalized dissimilarities (expressed as an information radius) between agendas of the *Objectives and Targets* and *Policies and Measures* sections for all countries and Energy Union dimensions. The greater the dissimilarity value, the less homogeneity there is between the topics covered in those sections. The colored lines show the average distance values for each dimension.

H Co-occurrence in clusters across dimensions

For all countries, we have checked which countries are together across all dimensions; by doing that, we wanted to check which countries often group. The result of this analysis is shown in Figure 11.

We have also checked some known groups: the Visegrad Group and Nordic countries, to check if some political alliances or geographical location factor into clustering. As we can see in Figure 12, there are visible similarities in clusters in both cases.

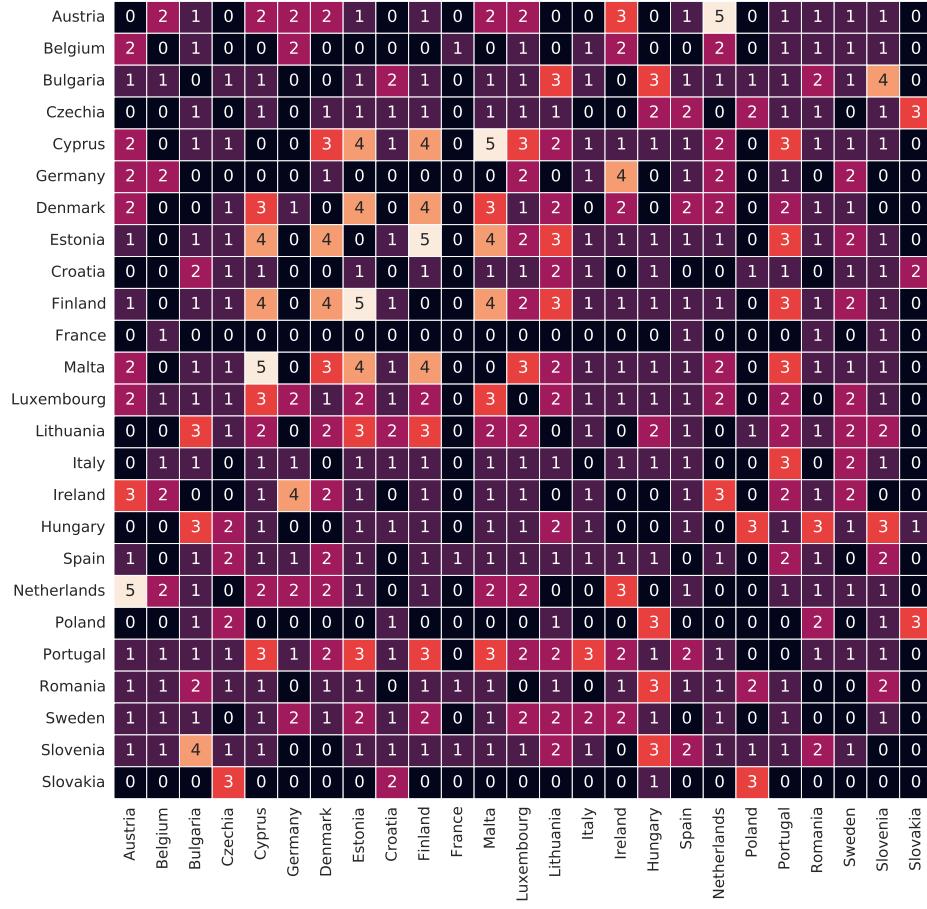


Figure 11: Matrix with a number of co-occurrences of countries in the same clusters across five Energy Union dimensions. We see several Member States being clustered together in all the dimensions, which suggests that they discuss different policy frames similarly.

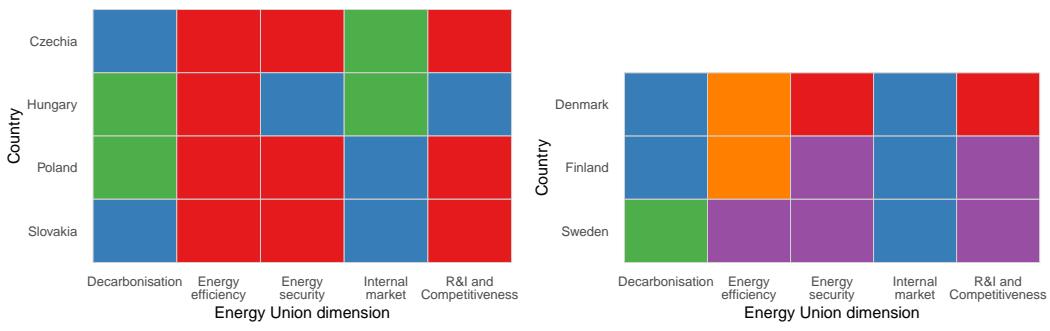


Figure 12: Assignment of countries to clusters in clustering obtained for five Energy Union dimensions. Each color means being together in one cluster.