PODIO: A Political Discourse Ontology

Ibai Guillén-Pacho[1,*], Ana Iglesias-Molina[1], Carlos Badenes-Olmedo[1,2] and Oscar Corcho[1]

[1]*Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

[2]*Computer Science Department, Universidad Politécnica de Madrid, Spain*

Abstract

In this study, we present the POlitical DIscourse Ontology (PODIO) and the accompanying Knowledge Graph, both designed to enhance the formalization and accessibility of political discourse in digital media. The core contribution of this work is the development of a comprehensive ontological framework and its instantiation in a Knowledge Graph that systematically organizes and represents political discourse. This framework provides structured insights to better understand and analyze political debates on various digital platforms. PODIO is specifically engineered to address the complexities of political communication, enabling detailed analysis of political proposals, ideological foundations, target audiences, and the temporal context of discourse. Through the integration of diverse existing datasets into the Knowledge Graph, we demonstrate PODIO's effectiveness in encapsulating a broad spectrum of semantically annotated political discourses

Keywords
ontology, political discourse, knowledge graph

1. Introduction

Political discourse, a strategic and dynamic form of communication in democratic societies, is crafted by political agents to adapt to changing sociopolitical contexts and interacts with various audiences [1]. Platforms like TikTok, X (formerly Twitter), and Facebook have transformed political discourse by adding immediacy and interactivity, facilitating real-time engagement and broad dissemination of political messages. In contrast, traditional media such as electoral programs and government bulletins offer a more formal, less interactive communication style.

Research on political discourse incorporates a range of methodologies that focus on different elements, such as national and regional party manifestos [2, 3], politician activities on social media [4], state head speeches [5], and enacted legislation [6]. Traditional research has often been siloed by communication channels, limiting data integration across diverse sources. To address these challenges and improve the interoperability of data, we have developed an ontology that redefines traditional classifications, facilitating a more nuanced representation of political discourse across various media. This innovation extends beyond existing models like the BBC Politics ontology,[1] which lacks the detail needed for complex political analyses, by offering a more versatile *Discourse* class. Additionally, the integration with the Legal Knowledge Graph[2] via federated queries enhances our capability to combine different data sources, moving toward a more integrated and comprehensive approach to modeling political discourse.

This paper introduces the Political Discourse Ontology (PODIO)[3] and Knowledge Graph,[4] focusing on advancing political discourse analysis. PODIO integrates diverse political communications, such as

[1]https://www.bbc.com/ontologies/politics-ontology/
[2]https://lynx-project.eu/doc/lkg/
[3]https://w3id.org/podio/v1.0
[4]https://w3id.org/podio/sparql

speeches, tweets, and manifestos, into a comprehensive Knowledge Graph. This integration not only includes social media activities of political figures but also legislative outputs, significantly improving the interpretation and interaction analysis of political strategies across various media. This innovative approach leverages the existing Political Marketing Ontology and extends its capabilities by focusing on political discourse representation rather than just the dissemination.

2. The PODIO Ontology

The PODIO ontology has been developed using the Linked Open Terms (LOT) methodology [7], which guides the creation of ontologies through four main stages: Requirements Specification, Implementation, Publication, and Maintenance. Initially, the ontology's purpose and competency questions are defined,[5] specifying how it should represent political discourse across various channels and the types of political information it must handle, such as data from social media, political manifestos, and party proposals. Following the specification, the ontology is built and evaluated using tools like Protégé and OOPS! [8] for model validation, ensuring no inconsistencies before publication. Once completed, the ontology is documented, published, and continuously updated through community feedback on GitHub.[6]

PODIO incorporates two key concepts: *Conversational Discourse* and *Expository Discourse. Conversational Discourse*, interactive in nature, encompasses content such as texts, videos, and images shared via social media platforms, enriched with metadata such as language and engagement statistics. The following example of a *Conversational Discourse* (Listing 1) shows an instance of `podio:Conversational`, a *Discourse* on *Twitter* that is created by the account of the *PSOE* political party, with the date it was created, the number of likes obtained, the shared content, etc.

```
1   :Post/Twitter/1729783306588180721 a podio:Conversational ;
2     sioc:has_creator :UserAccount/Twitter/psoe_m ;
3     sioc:has_container :Hashtag/viviendas ;
4     terms:created "2023-11-2 13:56:00" ;
5     terms:source <https://twitter.com/i/status/1720062424819040555>;
6     podio:content "El gobierno regional lleva dos legislaturas prometiendo
          [...]" ;
7     schema:sharedContent [ a schema:VideoObject ;
8        schema:contentUrl <https://twitter.com/i/status/1720062424819040555>] ;

9     sioc:mentions :UserAccount/Twitter/CristinaGAlvare ;
10    schema:interactionStatistic [ a schema:InteractionCounter ;
11      schema:interactionType schema:UserLikes ;
12      schema:userInteractionCount 107 ] .
```
Listing 1. Instance of *Conversational Discourse* created from a Tweet.

Expository Discourse, on the other hand, covers non-interactive elements such as policy proposals and approved policies, often embedded in legislative documents and party manifestos. These policies can be of two types: *Approved Policies* (`podio:ApprovedPolicy`), which form or have formed part of the legislation; or *Policy Proposals* (`podio:PolicyProposal`), which can form part of a *Party Manifesto* (`podio:PartyManifesto`). On the one hand, *Approved Policies* are changes that have been introduced into legislation by an authorised entity. To reflect this and import the specification of legislative documents, we have extended the class `lkg:Legislation` of the LKG ontology [9]. In this way, by making *Approved Policies* equivalent to `lkg:Legislation`, we ensure that these policies are part of the political discourse, have a domain expert definition, and both Knowledge Graphs interoperate and are jointly exploitable. The following example (Listing 2) provides an example of a `lkg:Legislation`, which is equivalent to `podio:ApprovedPolicy` and, in addition to existing data in the LKG network, is enriched with more data introduced by PODIO (language, creator, creation date, source, and publisher).

```
1   <https://apis.lynx-project.eu/document-platforms/BOE-A-2018-17773>
2     a lkg:Legislation ;
3     a podio:ApprovedPolicy ;
4     terms:language <http://id.loc.gov/vocabulary/iso639-2/spa> ;
```

---

[5]https://github.com/oeg-upm/PODIO/tree/main/queries
[6]https://github.com/oeg-upm/PODIO

```
5     terms:creator <https://www.Wikidata.org/wiki/Q3113880> ;
6     terms:publisher <https://www.Wikidata.org/wiki/Q5659724> ;
7     terms:created "2018-12-27" ;
8     terms:source <https://www.boe.es/eli/es/rd/2018/12/21/1462 > ;
9     nif:isString "En cumplimiento del mandato al Gobierno para fijar
         anualmente [...]" ;
10    eli:has_part: [ a lkg:LynxDocumentPart ;
11      terms:title "Disposicion final segunda. Entrada en vigor y periodo de
         vigencia." ;
12      nif:beginIndex "10777" ;
13      nif:endIndex "11304" ; ] .
```
Listing 2. Instance of an *Approved Policy*.

Alternatively, *Policy Proposals* are legislative suggestions made by an agent that may not necessarily come to fruition but are often included in *Party Manifestos*. These manifestos are comprehensive documents categorized as *Bibliographic Resource*, detailing language, creation date, source, description, content, ideology, and proposed candidates for political positions. Published by political parties, these documents encapsulate various policy proposals to outline a party's agenda and intentions. This ontology structure effectively captures and analyzes political communication, accommodating both dynamic social interactions and formal content like legislation, crucial for shaping public policy and political narratives.

3. Political Discourse Knowledge Graph

We constructed a Knowledge Graph (KG) based on the PODIO ontology using various data sources categorized into political and legislative domains. In the political domain, we included speeches from politicians and political parties, such as the American Presidency Project's electoral programs [10], social media posts from platforms like Facebook[11] and Twitter [12], and speeches by the Spanish head of state [5]. For the legislative domain, we incorporated data from the official Spanish gazette[13] and a European legislation KG [9]. The resources were carefully downloaded, refined, modified, and manually linked with Wikidata, ensuring all discourse authors are connected to their respective URIs.
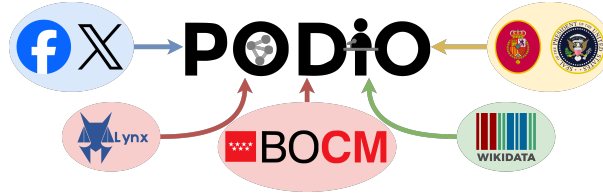

Figure 1: Main systems interoperating in the PODIO KG.

Data was transformed into RDF using a three-step approach centered around creating declarative mappings with the RDF Mapping Language (RML). Initially, we utilized YARRRML—a user-friendly serialization of RML—to outline the data source transformations in the YARRRML files, specifying how each data field correlates with elements in the ontology. These mappings were then converted into RML using the yarrrml-parser tool. Finally, the RMLMapper tool processed these mappings to produce RDF triples, forming the KG. This method enhances the KG's interoperability across various data sources, including social networks, external legal KGs, legal documents, institutional discourses, and integrates seamlessly with Wikidata, ensuring diverse information systems within PODIO are interconnected and functional. GraphDB was used to store and manage RDF triples in our Knowledge Graph (KG), accessible through a SPARQL endpoint.[4] This KG is validated using SPARQL queries designed to answer competency questions that address unmet needs within the domain, such as aggregating data across social platforms and enhancing data with external knowledge sources.

To underscore the interoperability and analytical capabilities of our Knowledge Graph, we have

developed an interactive demo environment (figure 2), available online,[7] where users can perform queries that highlight the integration with external legislative resources, enriching political discourse analysis. This platform allows users to see firsthand how queries detailing the latest legislation by jurisdiction and the ideologies and histories of political parties enhance our KG with data from Wikidata. Such integration facilitates complex analyses, like assessing the legislative and propagandistic activities of Spanish leftist parties during various legislative periods, demonstrating the invaluable role of external knowledge in providing comprehensive insights.

PODIO: A POlitical DIscourse Ontology

Query | Results

Competency Questions
Select a competency question to run its SPARQL query

CQ15: How much legislative and propaganda activity have the spanish leftist parties had during their governments in each legislature? ▼

Run Query

Open to see the query ▼

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcam: <http://purl.org/dc/dcam/>
PREFIX eli: <http://data.europa.eu/eli/ontology#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX lkg: <http://lkg.lynx-project.eu/def/>
PREFIX nif-core: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX podio: <http://w3id.org/podio#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX schema: <http://schema.org/>
```

Query | **Results**

CQ15: How much legislative and propaganda activity have the spanish leftist parties had during their governments in each legislature?

	ideologyLabel	ppartyLabel	inGovernmentSince	inGovernmentUntil	legislatureLabel	legislatureStartTime	legislatureEndTime	discourses	legislations
0	centre-left	Partido Socialista Obrero Español	2018-06-02T00:00:00Z	2024-11-19T10:01:17.359Z	15th legislature of Spain	2023-08-17T00:00:00Z	2024-11-19T10:01:17.359Z	0	0
1	centre-left	Partido Socialista Obrero Español	2018-06-02T00:00:00Z	2024-11-19T10:01:17.359Z	14th legislature of Spain	2019-12-03T00:00:00Z	2023-05-29T00:00:00Z	0	91
2	centre-left	Partido Socialista Obrero Español	2018-06-02T00:00:00Z	2024-11-19T10:01:17.359Z	13th legislature of Spain	2019-05-21T00:00:00Z	2019-12-03T00:00:00Z	149	9
3	centre-left	Partido Socialista Obrero Español	2018-06-02T00:00:00Z	2024-11-19T10:01:17.359Z	12th legislature of Spain	2016-07-19T00:00:00Z	2019-03-05T00:00:00Z	24	56
4	centre-left	Partido Socialista Obrero Español	2004-04-16T00:00:00Z	2011-12-21T00:00:00Z	10th legislature of Spain	2011-12-13T00:00:00Z	2015-10-27T00:00:00Z	0	1

Figure 2: PODIO demo environment.[8]

4. Conclusions and Future Work

This paper introduces the PODIO ontology, designed to encapsulate political discourse across various communication channels including social media, electoral programs, and legislation from the United States, Spain, and the European legislative domain. PODIO addresses the shortcomings of existing ontologies by providing detailed modeling of political speeches that allow for direct interaction (like social media posts) and those that do not (such as legislative texts). It integrates diverse data types and sources, normalizing them to enable interoperability within a unified Knowledge Graph (KG). This KG, built following the Linked Open Terms (LOT) methodology, has been validated for consistency and

---

[7]https://w3id.org/podio/demo

[8]CQ15 takes the date of when the query is made as `inGovernmentUntil` and as `legislatureEndTime` if they have not yet finished.

error-free structure, ensuring that it meets 33 specific requirements tailored to the political discourse domain. We hope this resource will be of help to experienced sparql users interested in analysing political discourse. Future enhancements of PODIO aim to include multimedia content and extend its geopolitical applicability beyond the initial regions, focusing on expanding both the source variety and the discourse characteristics analyzed, such as intentionality and the presence of offensive content.

Acknowledgments

References

[1] P. De Vos, Discourse theory and the study of ideological (trans-) formations: Analysing social democratic revisionism, Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA) 13 (2003) 163–180.

[2] P. Lehmann, S. Franzmann, T. Burst, S. Regel, F. Riethmüller, A. Volkens, B. Weßels, L. Zehnter, The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2023a, 2023. URL: https://doi.org/10.25522/manifesto.mpds.2023a. doi:10.25522/manifesto.mpds.2023a.

[3] H. Schmitt, A. M. Wüst, Euromanifestos project (emp) 1979 - 2004, GESIS Data Archive, Cologne. ZA4457 Data file Version 1.0.0, https://doi.org/10.4232/1.4457, 2012. doi:10.4232/1.4457.

[4] P. Nulty, Y. Theocharis, S. A. Popa, O. Parnet, K. Benoit, Social media and political communication in the 2014 elections to the european parliament, Electoral Studies 44 (2016) 429–444. URL: http://dx.doi.org/10.1016/j.electstud.2016.04.014. doi:10.1016/j.electstud.2016.04.014.

[5] E. Álvarez-Mellado, A Corpus of Spanish Political Speeches from 1937 to 2019, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 928–932.

[6] C. Borrett, M. Laurer, The CEPS EurLex dataset: 142.036 EU laws from 1952-2019 with full text and 22 variables, 2020. URL: https://doi.org/10.7910/DVN/0EGYWY. doi:10.7910/DVN/0EGYWY.

[7] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, R. García-Castro, LOT: An industrial oriented ontology engineering framework, Engineering Applications of Artificial Intelligence 111 (2022) 104755. doi:https://doi.org/10.1016/j.engappai.2022.104755.

[8] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, International Journal on Semantic Web and Information Systems (IJSWIS) 10 (2014) 7–34. doi:10.4018/ijswis.2014040102.

[9] J. Moreno Schneider, G. Rehm, E. Montiel-Ponsoda, V. Rodríguez-Doncel, P. Martín-Chozas, M. Navas-Loro, M. Kaltenböck, A. Revenko, S. Karampatakis, C. Sageder, J. Gracia, F. Maganza, I. Kernerman, D. Lonke, A. Lagzdins, J. Bosque Gil, P. Verhoeven, E. Gomez Diaz, P. Boil Ballesteros, Lynx: A knowledge-based ai service platform for content processing, enrichment and analysis for the legal domain, Information Systems 106 (2022) 101966. URL: https://www.sciencedirect.com/science/article/pii/S0306437921001563. doi:https://doi.org/10.1016/j.is.2021.101966.

[10] J. T. Woolley, G. Peters, The American Presidency Project, 1999.

[11] T. Baviera Puig, Facebook ads 2019 spanish general elections dataset, 2020. URL: http://dx.doi.org/10.4995/Dataset/10251/146502. doi:10.4995/dataset/10251/146502.

[12] R. Moya, Tweets Política España, 2023. URL: https://www.kaggle.com/datasets/ricardomoya/tweets-poltica-espaa.

[13] Arganda del Rey City Council, Publicaciones BOE / BOCM 1985-2023, 2023. URL: https://datos.gob.es/es/catalogo/l01280148-publicaciones-boe-2023.