applied sciences

MDPI

Article

Integroly: Automatic Knowledge Graph Population from Social Big Data in the Political Marketing Domain

Héctor Hiram Guedea-Noriega [1] and Francisco García-Sánchez [2,*]

[1] Escuela Internacional de Doctorado, University of Murcia, 30100 Murcia, Spain
[2] Departamento de Informática y Sistemas, Faculty of Computer Science, University of Murcia, 30100 Murcia, Spain
* Correspondence: frgarcia@um.es; Tel.: +34-868-88-8107

Featured Application: Knowledge system and data analysis for political marketing.

Abstract: Social media sites have become platforms for conversation and channels to share experiences and opinions, promoting public discourse. In particular, their use has increased in political topics, such as citizen participation, proselytism, or political discussions. Political marketing involves collecting, monitoring, processing, and analyzing large amounts of voters' data. However, the extraction, integration, processing, and storage of these torrents of relevant data in the political domain is a very challenging endeavor. In the recent years, the semantic technologies as ontologies and knowledge graphs (KGs) have proven effective in supporting knowledge extraction and management, providing solutions in heterogeneous data sources integration and the complexity of finding meaningful relationships. This work focuses on providing an automated solution for the population of a political marketing-related KG from Spanish texts through Natural Language Processing (NLP) techniques. The aim of the proposed framework is to gather significant data from semi-structured and unstructured digital media sources to feed a KG previously defined sustained by an ontological model in the political marketing domain. Twitter and political news sites were used to test the usefulness of the automatic KG population approach. The resulting KG was evaluated through 18 quality requirements, which ensure the optimal integration of political knowledge.

Keywords: political marketing; knowledge graph; social big data; knowledge graph population; political marketing ontology; natural language processing; named entity recognition

Citation: Guedea-Noriega, H.H.; García-Sánchez, F. Integroly: Automatic Knowledge Graph Population from Social Big Data in the Political Marketing Domain. *Appl. Sci.* **2022**, *12*, 8116. https://doi.org/10.3390/app12168116

Academic Editors: Huan Wang and Wen Zhang

Received: 11 July 2022
Accepted: 9 August 2022
Published: 13 August 2022

1. Introduction

The definition of marketing has evolved to the extent that its impact not only brings together promotion and communication techniques for products or services but acquires connotations within non-commercial organizations [1]. Political parties and their candidates are not exempt from taking advantage of these strategies to communicate better their political products to seek the electoral victory [2]. Political marketing is defined as the application of different marketing techniques and concepts by political actors to achieve the political organizations' goals [3]. In the last few years, political marketing has been receiving increased attention in different disciplines such as voter behavior, social media marketing, political branding, political microtargeting, among others [4].

Contemporary political campaigns are characterized by a series of proselytizing activities through a number of digital media platforms to influence the electors' vote decisions [5]. With the main goal of influencing voters' perceptions in a positive manner, political actors design, plan, and execute propaganda strategies accompanied by mainly political campaign elements, including investigation, management, and communication techniques [6]. Political marketing comprises four large study groups to generate collective knowledge [7]: candidate, adversaries, electorate, and election. Candidates (and adversaries) are applicants

for a public position and their success depends on the votes cast by citizens (the electorate) in an election; during the political campaign, citizens are supposed to get acquainted with the political, economic, and social views as well as the personal background of candidates.

Political marketing is increasingly relying on the use of data-driven voter research with personalized political advertising [8]. The focus is mainly set on public opinions and data collection, occasionally enriched with other related data, to make people-targeted political advertisements and formulate the correct message facilitating the communication between the candidate and the voter. This big amount of user data collected online and integrated with other data coming from different sources entails significant challenges in processing, storage, and knowledge inference, due to many factors, especially data heterogeneity. Ontologies, defined as "*an explicit and formal specification of a shared conceptualization*" [9] and knowledge graphs (KGs), defined as "*real world entities and their interrelations, organized in a graph*" [10], could serve to overcome this obstacle. Ontologies enable the storage of data in a machine-readable format. In the context of the semantic web, ontologies have been used to support the discovery, integration, representation, and management of knowledge [11]. A KG is conceptually represented as a combination of terminologies and definitions (T-Box) and assertions (A-Box). Together A-Box and T-Box statements make up a knowledge base or a KG [10]. Throughout the manuscript we make use of the term 'ontology' to refer to the formal schema (T-Box) and 'KG' ('populated ontology' or 'instantiated ontology') to refer to the combination of the ontology schema and its instances (T-Box and A-Box).

Knowledge engineers and domain experts are the main individuals responsible for ontology and KG development. However, the manual construction of ontologies and their population are error-prone, time-consuming tasks [12]. In order to make more effective, functional, and solve common problems in ontology tasks, research fields such as Ontology Learning (OL), Ontology Population (OP), and Ontology Evolution/Enrichment (OE) have been developed that define the methods and activities that enable the automation of these processes. On the one hand, OL is defined as the process of acquiring, constructing, or integrating an ontology in a (semi-) automatic fashion [13,14]. On the other hand, OP is the task of adding new instances of concepts to the ontology [15]. In other words, while OL deals with T-BOX's conceptual model, OP focuses on adding new assertions to the A-BOX. Then, as the domain of discourse under consideration evolves so should the ontology; the task of extending and adapting an existing ontology with additional concepts, semantic relations, and rules is usually referred to as OE [16,17]. OE encompasses the processes that involve the increase, maintenance, and scalability of the ontology from its initial structure [18].

OP identifies the instances, relationships, and properties of an ontology with discovered knowledge. Unstructured sources such as text documents are the dominant online knowledge resources that contain a large number of facts expressed either explicitly or implicitly [15]. A fast and low-cost OP process becomes extremely important for the development of knowledge systems [19]. In this work, we propose Integroly, an automatic KG population framework to gather voter insights from both semi-structured and unstructured sources. The framework is built upon a previously defined ontology for the political marketing domain [20]. Semantic technologies were integrated to provide a formal representation of political marketing concepts such as candidate proposals, electorate opinion, and political news. The KG population process encompasses three main tasks, namely, (i) source connection and data collection, (ii) information extraction, and finally, (iii) population process and ontology validation. Natural Language Processing (NLP) and Machine Learning (ML) techniques were leveraged to process the gathered data and to fulfil the instantiation of the KG. All these components constitute a novel tool for the discovery and management of new knowledge within the highly competitive and dynamic politics domain. With all, the main contributions of this work are as follows:

- We present the Political Marketing Ontology (PMont), a shareable conceptual model involving the most relevant elements in the political marketing domain, including

three main operative knowledge items from political campaigns: candidate, electorate, and media. The ontology is an evolution of our previous work [20].
- Different unstructured, semi-structured, and structured data sources are exploited to populate the knowledge base by using a variety of information extraction techniques. New data sources can be added seamlessly.
- Data enrichment and analysis in semantic and linguistic levels for Spanish texts through NLP and ML techniques contributed to the validation and classification of the opinions, proposals, and news media publications.
- We propose an automatic KG population framework that homogenizes, collects, and relates the data of political marketing concepts with the possibility of inferring new knowledge for decision making.
- As a result of the work, a validated and populated KG for political marketing purposes was obtained.

The rest of the manuscript is organized as follows. In Section 2, background information about political marketing is provided along with a comprehensive analysis of KG enrichment through OP. The proposed KG population framework is described in detail in Section 3. Quality properties of the KG produced by the proposed framework are evaluated and discussed in Section 4. Finally, conclusions and future work are put forward in Section 5.

2. Literature Review

In this paper, we describe an automatic system for ontology population and KG construction in the political marketing domain. Different and heterogenous semi-structured and unstructured sources are exploited to gather the data with which to populate the knowledge base. In the last few years, researchers in the NLP field have proposed a significant number of tools to automate the ontologies instantiation from text. In this section, political marketing underpinnings are discussed and the most representative works in ontology population from semi-structured and unstructured data sources are listed.

2.1. Political Marketing

Political marketing emerges to understand the new ways of making modern politics using a cross-disciplinary approach that encompasses marketing, analysis of the behavior of political parties and voters, communication, and the effective use of different mass media (traditional and digital) as promotional channels [21]. The main objective is the creation and development of political concepts related to specific parties or candidates to satisfy certain groups of voters in exchange for their votes [6]. Political marketing involves three fundamental concepts, namely, (i) the political product, (ii) the political organization, and (iii) the electoral market [22]. Electoral campaigns are carried out on the basis of evolutionary processes with the purpose of building a profitable political strategy for the electorate. Voters can be defined as consumers of political assets; therefore, a candidate or political party has the need to satisfy said demand through a detailed study of the electorate and correctly communicate the political offer.

The political marketing history has its origins with the term "political propaganda" in the mid-20th century. It comes from the Latin "*propagare*" which means "to spread, to extend, to be disclosed", where the main objective was the massive dissemination of arbitrary ideas that seek to influence the value systems and behaviors of citizens. The evolution of the process of influencing the behavior of citizens with the democratization of political procedures gave rise to political marketing as a new and independent branch of political science whose activity is not only to spread a message, but also to study the psychology of the masses, corporate ethics, statistics, mass media, among others [23].

The constant evolution of political marketing is linked to the development of temporary human values such as civil liberty and democracy, changing from a non-critical representation, to an actively critical one, where public scrutiny is increasingly meticulous, and social media have fostered the active participation of citizens during electoral pro-

cesses [24]. Political marketing has proven to be a science with an efficient set of practices, strategies, and tools for the political influence of candidates in order to maximize profits [8]. However, the techniques of communication and electorate knowledge have become more complex due, in part, to the appearance of the internet and the explosion of social networking sites and services (SNS) [6,25]. Additionally, to know the needs from voters through the information from different sources has become a hard task but important given that it is one of the main political marketing goals [26]. The needs for novel information extraction, classification, and storage strategies for enabling big data analysis are fundamental tasks, and one of the main challenges to be solved when dealing with information coming from various sources is that of data heterogeneity. Semantic technologies have proven useful in the management and analysis of data at the knowledge level [27]. In particular, ontologies provide a shared understanding of knowledge about a specific domain [28] and can facilitate the integration of data from heterogeneous sources [29].

The use of semantic technologies in the new digital marketing era is one of the approaches outstanding given their usefulness in decision making [30,31], data analysis [32–34], data integration [35–37], and knowledge extraction [38,39]. In decision making, semantic technologies provided the support for data integration with reasoning methods for efficiently querying the knowledge base, strategies to separate the relevant dimensions (technical, financial, political), and means to communicate the comprehensible decision options to stakeholders [30]. The key for decision making based on ontologies is the availability of relevant knowledge in a machine-readable way to ensure processing in a timely manner [31]. In data analysis, the usage of semantically enhanced formats provides characteristics of enrichment techniques, semantic analytics, and linked data, which will benefit marketing in its needs for collecting, processing, relating, and analyzing large amounts of data quickly and optimally to obtain knowledge aimed at developing the campaign strategy [32]. The main advantages of implementing semantic technologies are (i) performing the analysis of big data at the knowledge level, and (ii) its broad visualization with linked data techniques [33]. The combination of semantic technologies and NLP potentiates the inference and real-time access to unstructured data with Named Entity Recognition (NER) techniques for its processing and analysis [34]. The integration of unstructured data, especially from sources such as SNS, is one of the fundamental and most challenging tasks in current knowledge systems. It comprises the construction of data mapping and modeling as support for data unification using ontologies with the aim to generate a knowledge base on a specific domain [35]. Governments use semantic technologies (ontologies, data integration) as the basis for the accessibility, reusability, and easy consumption of their data by systems and applications of all kinds improving the reliability and transparency of the political system to promote the Open Government [36].

Big data is frequently attributed with three V's main properties [40]: volume, velocity, and variety. Volume refers to the amount of data that is generated every second, minute, and day in the system environment; variety is related to the heterogeneity of the data generated from different kinds of data sources and formats; and velocity refers to the speed at which data is generated. The big data management involves three main dimensions [41]: (i) people, (ii) process, and (iii) technology. People refers to technical skills to manage and understand the management of the big data technologies; process is related to all actions, from data acquisition to reflection or visualization, performed in the technological and business environments; and, finally, technology is about the several techniques and technologies for collecting, storing, processing, and analyzing the data. In recent years, the challenges associated to big data analysis have increased due to a variety of reasons including the difficulty of managing such amount of data with traditional data processing applications, the variability of data sources, and the complexity of reusing application systems [37]. Semantic web technologies provide an easier way to search, reuse, combine, and share information, which brings added value and benefits helping to counter the challenges of big data [33]. In conclusion, ontologies and semantic technologies could be seen as supporting tools for the knowledge integration, structuring, and sharing in the

political marketing domain. In this work, we propose a framework to gather knowledge from heterogeneous sources (including SNS) and to instantiate a political marketing KG; thus, alleviating the problems associated with this time-consuming, error-prone task and enabling its subsequent analysis to define a successful political marketing campaign.

2.2. Ontology Instantiation: Populating Knowledge Graphs

The web is a complex scenario for information extraction due to the enormous amount of data available, the heterogeneity of the sources, and their technological variety. Ontologies have become the cornerstone technology of the semantic web allowing to represent knowledge in a machine-readable fashion and to organize the information in a homogeneous structure [42]. Some methodologies have been conceived that guide the development of ontologies, offering considerations such as determining the specific domain and ontological scope, the reuse of existing ontologies, T-BOX design (classes, concepts, properties, and relationships), and instantiation (A-BOX) [20]. The manual construction of ontologies turns into a rigorous group of tasks, which are time-consuming, human effort-demanding, error-prone, and require some technical skills [13]. It is possible to solve these problems by implementing automatic or semi-automatic tools for ontology creation, also referred to as OL [43]. OL comprises methods and techniques for the extraction, generation, or acquisition of ontologies from natural language texts, involving data pre-processing, term/concept and relation extraction, axioms acquisition, and evaluation [43]. In synthesis, OL is the approach for automating the knowledge acquisition process.

Conversely, OP is the task of adding new instances of concepts or relationships to the ontology, also including processes to eliminate redundant instances [14]. OP does not change the ontology structure; therefore, the concept hierarchy and non-taxonomic relations are not modified. OP requires an existing and specific domain ontology with defined structure. OP systems are related to ontology-based information extraction systems, for the similarities in the procedures used to associate data with ontology concepts and instances [44]. The manual OP process is time consuming and susceptible to human errors; therefore, it is highly advisable to use automatic or semi-automatic processes leveraging various techniques such as NLP and information extraction to acquire and classify ontology instances [19].

In recent years, KGs have become popular because of the high demand for graph technology to make sense of data [45]. KGs are a specific type of graph with an emphasis on contextual understanding. KGs are interlinked sets of facts that describe real-world entities, events, or things and their interrelations in a human and machine-understandable format [46]. In a simple and clear definition, KGs represent words as nodes and the relationships as edges between words [47]. With all, KGs allow people and machines to improve the connections in their datasets [48]. The property graph model is the most popular method to create KGs, consisting basically of nodes representing entities in the domain and relationships representing how entities interrelate [46]. Ontologies allow for more complex types of relationships than property graph-based KGs; for example, properties such as *part_of*, *compatible_with*, or *depends_on* can be established between categories. They also enable the definition of hierarchical relationships and the further characterization of relationships (e.g., whether they are transitive, symmetric, reflexive, etc.) [46,49]. A KG built on top of an ontology schema acquires and integrates information into an ontology and can apply a reasoner to derive new knowledge [48].

Unstructured data sources are the most dominant knowledge resources online and, consequently, the most common source format exploited in OP systems. Typically, the first process in an OP pipeline is data acquisition including the identification of the type of source (i.e., image, video, text), which influences the information extraction techniques to be used [15]. There are many information extraction techniques such as Application Programming Interfaces (APIs), web scraping, and the use of specialized libraries to extract heterogeneous data from social media and news websites. The popularity and growth of APIs over the past few years was mainly due to the existence of open platforms,

standardized cloud information access points, and third-party applications. However, the current limitations imposed by the strict privacy policies that apply to companies such as Facebook, Twitter, and Instagram in terms of data exposure, prevent the access to sensitive users' information and increase the security caused by the social engineering attacks [50].

Web scraping is one of the possible solutions to the APIs problem in data access mentioned above [51]. It refers to the practice of automatically extracting content from webpages and other digital files. In fact, it is old technology prior to APIs, more complicated in terms of usability but very flexible and powerful. One of the practical disadvantages of web scraping solutions is that they are more difficult to learn and apply than API-based tools. Web scraping involves examining and understanding the document object model (DOM) structure from pages. The DOM elements containing the desired data must be inserted into the chosen tool's functions and the output captured in a file. It is not only a more complex workflow than APIs typically require, but it also entails a custom-made solution for every website or social media platform [52] and is more susceptible to change. In recent years, the use of web scraping techniques has increased due to the lack of API implementation and the terms of service restrictions of many social media sites. Libraries such as Facebook Scraper (https://pypi.org/project/facebook-scraper/ (accessed on 12 August 2022)) and Twitter Scraper (https://pypi.org/project/twitter-scraper/ (accessed on 12 August 2022)) are very popular and often used as SNS scraping solutions [53].

After data acquisition, the next steps of a typical OP system are the knowledge extraction process (i.e., semantic annotation, NLP, and semantic extraction), followed by the population process (i.e., redundancy elimination, consistency checking, and ontology instantiation) and, finally, the storage stage (usually in a TripleStore) [13–15,54,55]. Semantic annotation is the process of adding metadata about concepts, entities, and relationships over unstructured documents and data. Semantic annotations are usually used by machines for context understanding. Semantically tagged documents are easier to find, interpret, combine, and reuse [13]. NLP tools allow finding linguistic annotations on unstructured texts and documents, including sentence limits and tokens, parts of speech (PoS), named entities, numerical and time values, dependency and circumscription analysis, coreference, sentiment, and relationships. The NLP process also allows obtaining semantic elements to incorporate them into semantic structures such as ontologies [56]. There are a wide variety of applications and tools dedicated to performing NLP tasks such as SpaCY (https://spacy.io/ (accessed on 12 August 2022)), Stanford CoreNLP (https://stanfordnlp.github.io/CoreNLP/ (accessed on 12 August 2022)), NLTK (https://www.nltk.org/ (accessed on 12 August 2022)), and Textrazor (https://www.textrazor.com/ (accessed on 12 August 2022)) [57].

In general, three main approaches can be distinguished when dealing with OP for extracting domain-specific terms [55]: (i) rule-based ontology population systems [15,58], (ii) ontology population systems using ML or NLP [54,59], and (iii) ontology population systems that use statistical approaches [57]. In rule-based systems (i), the rule group is designed and developed for the location, classification, and extraction of information in predefined categories such as people, organizations, time expressions, places, etc., known as named entities. NER are tools and techniques used for these tasks' automatization. In [15], the authors describe the use of NER, a set of predefined predicates and a lexical database (WordNet) to generate the instance classification rules that will be used to populate the ontology. Another strategy for OP (ii) makes use of a predefined ontological schema for the term's comparison, along with learning algorithms using ML and NLP tasks for information extraction from unstructured data [54]. The main disadvantage is the supervision in the selection of the initial concepts and the schema construction. In the same line, a blend of techniques is described in [59] using association rules, clustering, classification based on associations through advanced analytics employing NLP and data mining techniques. The system uses federated data analysis, knowledge discovery, information retrieval, and new techniques to deal with semantic web and KG representation. The processing process

integrates data from multiple sources virtually by creating virtual databases. Finally, in statistics-based systems (iii) is the measurement of the semantic similarity between the terms extracted from the unstructured data and the ontology instances. However, this approach is not appropriate for all situations. For example, it cannot treat terms not covered by synonym dictionaries. Similarly, this approach is not suitable for understanding abbreviations. In [57], a k-means algorithm is proposed for topic identification and data classification. The algorithm compares and obtains named entities from unstructured data to later transform them into Linked Open Data (LOD) enriched with DBpedia URIs.

Our approach employs a classification model to identify candidate instances through a semi-supervised process using NLP and our own pattern learning algorithm of political marketing concepts. The algorithm filters the candidate terms and NLP tools facilitate the extraction of instances from unstructured text to populate the PMont. Our work presents improvements in the ontology and KG population process, implementing source connection and data collection, data enrichment with sentiment analysis, and redundancies elimination, producing as a result a populated political-marketing-related KG.

3. Integroly: A Framework for Political Marketing Knowledge Graph Enrichment

The functional architecture of the proposed system is shown in Figure 1 and comprises five main components: (i) the source connection and data collection module, (ii) the information extraction procedure, (iii) the PMont ontology, (iv) the ontology population and validation subsystem, and (v) the resulting KG. In this characterization, both the ontology and the KG are differentiated to distinguish between the ontology schema (i.e., PMont) and the populated ontology (i.e., KG). In brief, the system works as follows. The input of the system is heterogeneous data from social media and news sites. First, data is retrieved from its source with the use of web scraping techniques and APIs and stored in an intermediate database. Then, instance candidates of the PMont ontology are extracted from the gathered data with a combination of NLP, NER, and ML methods and algorithms. Next, during the ontology population stage new instances of entities and relationships are added into the knowledge base, and an automatic validation process is carried out to ensure the semantic consistency of the graph with the new axioms. Finally, the augmented KG is available for visualization or analysis with the aim to assist in the definition of successful political marketing campaigns.

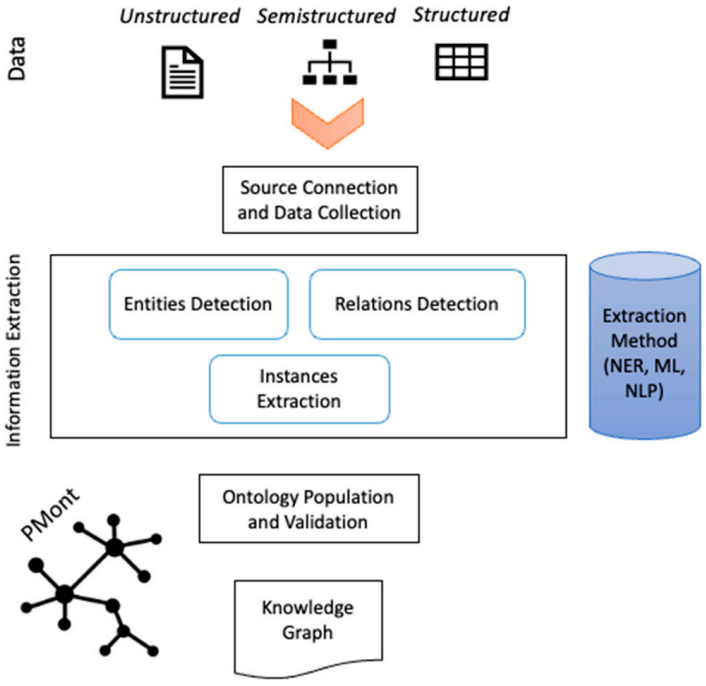

Figure 1. Integroly functional architecture.

Next, a detailed description of the main components of the framework is put forward.

3.1. PMont: Political Marketing Ontology

The most important purpose of gathering political-marketing-related knowledge in the form of an ontology is to later be able to analyze the available knowledge in order to obtain insights useful for the proper definition of political strategies. The adequate design of this ontology will provide the answers needed to improve decision making in this scenario, obtaining four operative knowledge items: what the candidate is like, what the adversaries are like, what the voters are like, and what the election is like in general.

The PMont ontology [20] was created from scratch by following the ontology development methodology named Ontology 101 [60]. PMont was conceived to serve as an essential guide for our knowledge-based system, and to answer the main questions of political marketing typical in an election campaign including:

- What does the electorate demand?
- What does public opinion say about the candidate (profile, proposals)?
- What does public opinion say about the political party to which the candidate belongs?
- What kind of message and political language should the candidate establish to positively impact citizenship?
- What kind of campaign proposals should be designed?
- What are the target groups of the electorate?
- In which group has greater acceptance?
- What comparative does the candidate have with his adversaries?
- In what mass media does the candidate have the greatest impact?
- Which mass media are partial (identify candidate/party of your liking) and which are impartial?
- What part of the electorate has decided its vote (identify individual voters and their candidate/chosen political party) and what part of the electorate is undecided (identify individuals)?

Based on these research questions, we defined the important terms of the ontology which are shown in Table 1.

Table 1. Ontology political marketing terms.

- Campaign
- CandidateCampaign
- Organization
  - Political Party
  - Mass Media
- Person
  - Eligible voter
- Candidate
  - Not eligible voter
- Publication
  - Types of publication (opinion, proposal, and news)
  - Features of publication (polarity, degree of sentiment, and topic)
- Reputation
- Trustworthiness
- Ideology
- Geographic impact
- Source and format
  - Structured, semi-structured, and not structured

According to the main concepts listed above, the ontology development process was started with a *top-down* approach to present and define the hierarchy of classes, subclasses, and object properties more easily. The ontology has been defined in OWL2 [61] using Protégé (https://protege.stanford.edu/ (accessed on 12 August 2022)). It contains seven main concepts: *Campaign*, *Candidate*, *CandidateCampaign*, *Organization*, *Person*, *Publication*, and *Source*. There are also other related concepts to cover the aims suggested in political marketing. These main concepts directly impact the research questions and can be defined as follows:

- *Campaign*: this class defines the dates of the campaign and political election. It is related to political parties and candidates.

- *Candidate*: it represents a person affiliated to a *Political Party*, with the intention of being elected to a public office.
- *CandidateCampaign*: this class is used to define ternary relationships between the classes *Campaign*, *Candidate*, and Political Party.
- *Organization*: it is a superclass containing two subclasses, namely, *Mass Media* and *Political Party*, with the purpose of defining structures and groups of people.
- *Person*: it is a superclass containing two subclasses, namely, *EligibleVoter* and *NotEligibleVoter*. *EligibleVoter* is the person who can vote for a candidate in a political election, while *NotEligibleVoter* cannot vote under the circumstances demanded by the law where the election is located (for example, minor).
- *Publication*: it represents different types of publications. In particular, three subclasses are considered, namely, *NewsItem* (news article created and shared by mass media), *Proposal* (a political proposal by a candidate), and *Opinion* (public opinion by a citizen shared on social media).
- *Source*: it indicates the source of the publications. It is a superclass containing *Structured*, *Semistructured*, and *Unstructured* as subclasses.

As suggested in the ontology development methodology mentioned above, the properties of classes or slots were also defined, which aim to relate the different classes and subclasses and add data for the inference of knowledge. Figure 2 shows the high-level classes and their relationships.

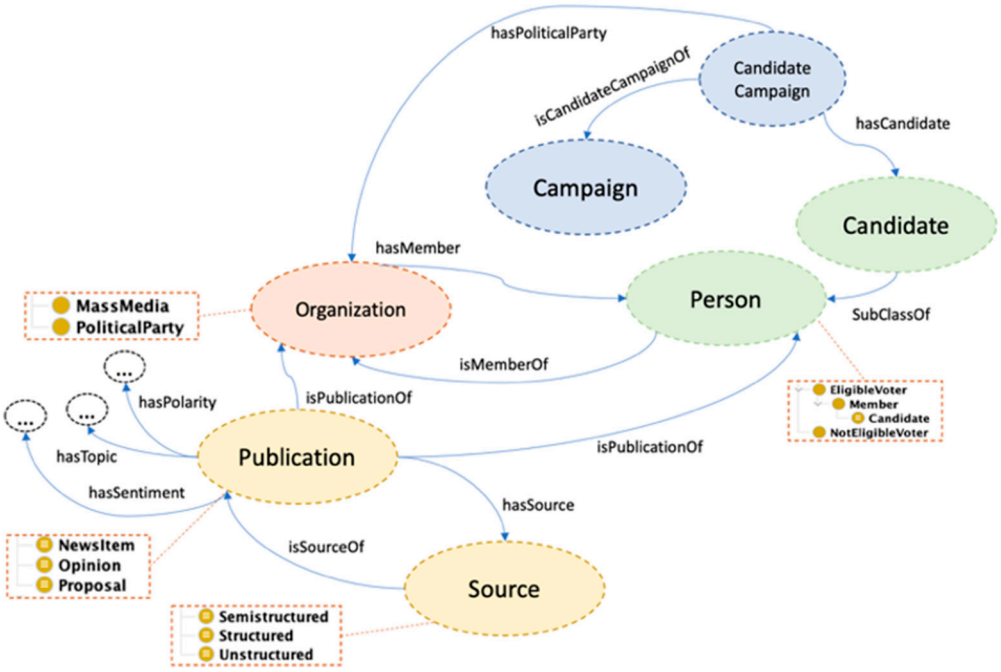

Figure 2. Ontology's high-level classes and their relationships.

The ontology is available as a Github repository (https://github.com/hectorguedea/pmont (accessed on 12 August 2022)), using the OnToology system (http://ontoology.linkeddata.es/ (accessed on 12 August 2022)) for online documentation with Widoco (https://github.com/dgarijo/Widoco (accessed on 12 August 2022)). It has been validated through the OOPS! OntOlogy Pitfall Scanner (http://oops.linkeddata.es/ (accessed on 12 August 2022)); the result was positive without minimal or critical errors in all the evaluations divided into the structure (model, relationships, inference), functionality, usability, and consistency (https://hectorguedea.github.io/pmont/OnToology/Political-Marketing-Ontology.owl/evaluation/oops.html (accessed on 12 August 2022)).

3.2. Source Connection and Data Collection

The main aim of the 'source connection and data collection' component of the system is to gather data from a variety of selected raw sources (structured, semi-structured, and unstructured) using a number of mechanisms. In particular, three techniques are used to connect with the chosen sources, namely, APIs, web crawling/scraping, and local databases. For social media, specifically Twitter, we used the Twitter API to request real-time and streamed data through authentication mapping of OAuth 2.0 and token access. We implemented Twittered (https://github.com/redouane59/twittered (accessed on 12 August 2022)), a Java library to consume the API and additional Simple Logging Facade for Java, better known as SLF4J (https://www.slf4j.org (accessed on 12 August 2022)), dependency used as a generic API making the logging independent of the current implementation. The 'search tweets' function in the API extracts the data with fields as JSON objects and values, some of them being: tweet id, displayed name (username), name, location, text, likes-retweets-and-replies count. As result, a bucket of raw data with easily accessibility to manage is retrieved.

Web crawling is implemented as a technique for extracting data from news sites and other user-selected websites. We used TextRazor (https://www.textrazor.com (accessed on 12 August 2022)) as a complete cloud and self-hosted REST infrastructure for web scraping. It also assists in the information extraction process with features such as entity extraction, key phrase extraction, and automatic topic tagging and classification. Our system allows users to add a single website or a list of websites for extraction. Additionally, the proposed framework supports the access to a local database (e.g., CSV file) to extract all the website links from the rows and columns, and at the end, gets the cleaned content from every website.

All the raw data obtained from the considered sources is inserted into an intermediary relational database with a structure compatible with the information required to populate the PMont ontology classes. Figure 3 shows the tables created in an MySQL database, which presents *tweets*, *pages*, and *texts* as source types, and an index table used as the IDs dictionary. The data retrieved from the sources is thus separated from the information produced in later stages of the framework process pipeline. That is, the results of the data analysis process using NLP, NER, relation extraction, or classification techniques and tools as described in the next section are not included in the database but otherwise managed in-memory to enable further processing.

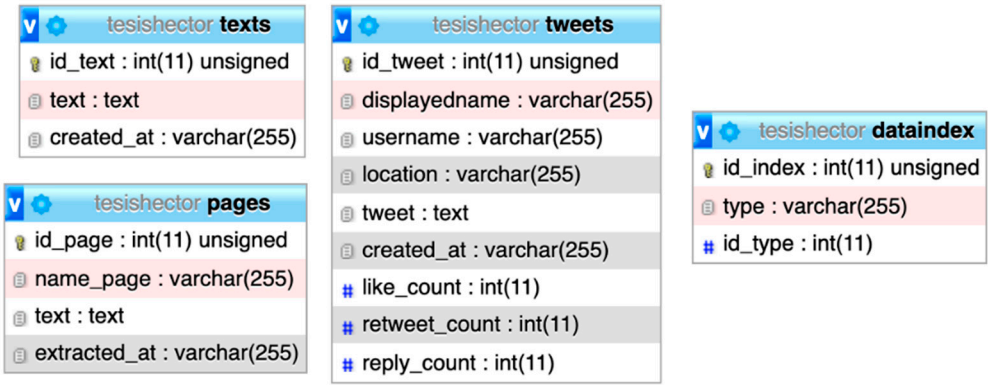

Figure 3. Relational database schema (raw data collection).

3.3. Information Extraction

To gather semantic knowledge from the retrieved data, specific for the concerns of our political-marketing-related use case, we used a number of different techniques including NER, ML, and NLP. The combination of these techniques facilitates the detection and extraction of the entities, relations, and attributes required to the instantiation process in the knowledge base. The TextRazor (https://www.textrazor.com (accessed on 12 August 2022)) REST API was the tool chosen for text analysis based on an NER and ML approach,

while Stanford CoreNLP (https://stanfordnlp.github.io/CoreNLP/ (accessed on 12 August 2022)), a Java library that supports a number of NLP tasks, was selected for performing sentiment analysis on text data determining the polarity of the text (either positive, negative, or neutral) using an NLP approach.

The method for knowledge items extraction according to the PMont ontology model is as follows:

1. For NER, we used the TextRazor API. This API works by leveraging a huge knowledge base of entities extracted from various web sources, including Wikipedia, DBPedia, and Wikidata. A matching engine is in charge of finding relationships between the text content and the possible entities grouped as dictionaries. Additionally, a statistics-based tagger is used to identify people, places, and companies among other concepts, and regular expressions are applied to spot the less ambiguous elements including email addresses and websites.
2. As regards relation extraction, we produced linked data using the TextRazor API, first, disambiguating the terms and, second, linking the entities to canonical IDs in the linked web (Wikipedia, DBPedia, and Wikidata IDs). In this process, we converted raw text into a set of well-structured subject–predicate–object expressions that related entities, phrases, and concepts. The approach of linking entities to canonical IDs offers great accuracy and recall, providing the flexibility to an unlimited number of relationship types.
3. As mentioned above, sentiment and polarity extraction are completed using the Stanford CoreNLP toolkit, an approach with a good understanding of the relevant metadata, deploying existing sentiment lexicons (including misspellings, morphological variants, slang, and social media markup) and the basic feature extraction (tokenization, stemming, POS-tagging). The whole features combination provided us with a robust tool for sentiment and polarity analysis.

In order to demonstrate the functionality of the different techniques in use by the information extraction module (i.e., NER, relation extraction, and NLP) a running example is presented next:

1. The framework interface allows adding unstructured data (e.g., text) to the database, and then, extracting it with a query statement. The text below is a real record from the *texts* table in the database (an excerpt of the text in natural language written in Spanish gathered from a news website):

   "De acuerdo a la evaluación promedio registrada por MITOFSKY en diciembre de 2021 para El Economista, el presidente López Obrador registró una aprobación promedio de 66% durante el último mes del 2021, la más alta desde febrero de 2019 cuando alcanzó 67%; en diciembre pasado la desaprobación presidencial promedio fue de 34 por ciento. Se identificó que por tipo de ocupación, los campesinos fueron quienes tuvieron el mayor porcentaje a favor del tabasqueño con 83.7 por ciento. Respecto de la 'percepción de seguridad', 41% de los encuestados opinó que está 'igual', 28.3% que está 'peor', 24.7% que está 'mejor', y 6% 'no contestó'".

2. The first stage is to implement NER and relation extraction analysis with the use of the TextRazor API. The results depicted in Figure 4 show the entities and the semantic relations. In Figure 5, the categorization and classification of the text based on the semantic identification of prominent key phrases and concepts related to the domain is shown. The number beside each category and topic is the matching score.

```
Entidad:  Tabasco -- [Place, PopulatedPlace, Region, AdministrativeRegion]
Entidad:  2021 -- [Number]
Entidad:  Andrés Manuel López Obrador -- [Agent, Person]
Entidad:  2019 -- [Number]
Entidad:  67 -- [Number]
Entidad:  34 -- [Number]
Entidad:  83.69999999999999903 -- [Number]
Entidad:  2021 -- [Number]
Entidad:  66 -- [Number]
```

Figure 4. NER and relation extraction results (partially in Spanish; 'Entidad' stands for 'Entity').

```
Categoría:20000593 politics>government 0.8575
Categoría:11000000 politics 0.8497
Categoría:20000610 politics>government>heads of state 0.5618
Categoría:20000654 politics>political process>political system>democracy 0.4533
Categoría:20000651 politics>political process>political parties and movements 0.4067
Categoría:20000606 politics>government>executive (government) 0.3895
Categoría:20000745 science and technology>social sciences>economics 0.3534
Categoría:20000574 politics>election 0.3384
Categoría:20000592 politics>fundamental rights>human rights 0.3225
Categoría:20000614 politics>government>national government 0.3221

Tema: Politics WikiID: Q7163 Score: 1.0
Tema: Government WikiID: Q7188 Score: 1.0
Tema: Andrés Manuel López Obrador WikiID: Q318508 Score: 0.7553
Tema: Presidencies WikiID: Q11708087 Score: 0.7128
Tema: Latin America WikiID: Q12585 Score: 0.6534
Tema: Mexico WikiID: Q96 Score: 0.644
Tema: South America WikiID: Q18 Score: 0.6053
Tema: Middle America (Americas) WikiID: Q29876 Score: 0.6019
Tema: Human activities WikiID: Q61788060 Score: 0.5981
Tema: Government of Mexico WikiID: Q2182191 Score: 0.591
Tema: Presidents WikiID: Q30461 Score: 0.574
Tema: Politics of Mexico WikiID: Q1155509 Score: 0.5691
```

Figure 5. Classification and categorization results (partially in Spanish; 'Categoría' stands for 'Category' and 'Tema' stands for 'Topic').

3. Finally, sentiment analysis is applied with the Standford CoreNLP library obtaining three possible values: positive, neutral, or negative. For the text presented above, the content analysis result is "neutral" (see Figure 6).

```
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator tokenize
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator ssplit
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator pos
[main] INFO edu.stanford.nlp.tagger.maxent.MaxentTagger - Loading POS tagger from e
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator parse
[main] INFO edu.stanford.nlp.parser.common.ParserGrammar - Loading parser from seri
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator sentiment
[main] INFO edu.stanford.nlp.sentiment.SentimentModel - Loading sentiment model edu
NEUTRAL
```

Figure 6. Sentiment analysis result.

At the end of the whole information extraction process, the data is enriched and augmented with semantic annotations with the purpose of being related to the PMont schema. Database records that were not processed for instantiation are tagged as instance candidates; each entry is enriched with semantic metadata using the aforementioned extraction techniques.

3.4. Ontology Population and Validation

For the instantiation of the ontology (i.e., KG population) we made use of OWL API (https://github.com/owlcs/owlapi (accessed on 12 August 2022)), a Java API for creating, manipulating, and serializing the domain ontology or knowledge base. The API allows to connect, read, and write over the RDF/XML and OWL/XML. It also includes reasoner interfaces such as FaCT++ and HermiT, to assist in deriving implicit facts from a set of given explicit facts; in summary, the capacity to infer logical consequences.

The OWL API starts the instantiation process of the records called instance *candidates*, adds the aforementioned semantic annotations for each one, and also adds a source reference that is the ID of the database entry (*hasValue*). This latter step can be useful for the validation of the formulated instances by an expert; thus, carrying out a semi-supervised process. The result is an ontological model populated with new instances.

Our validation method establishes the rules to clean the processed data previous to the ontology population. The goal is to search for duplicate, ambiguous, and incomplete data from the queue, and ensure the consistency of the ontology when the data is inserted as a new instance. For that, we use HermIT (http://www.hermit-reasoner.com (accessed on 12 August 2022)) as a reasoner to determine the ontology consistency and identify subsumption relationships between classes. HermIT is based on hypertableau calculus which provides much more efficient reasoning than any previously known algorithm. Moreover, it is fully compliant with the OWL 2 Direct Semantics as standardized by the World Wide Web Consortium (W3C).

3.5. Resulting Knowledge Graph

Finally, the KG is presented as a representation of the recently populated ontology (knowledge base). Visually, the KG generated could be presented using tools such as OntoGraph (https://github.com/NinePts/OntoGraph (accessed on 12 August 2022)), WebVOWL (https://service.tib.eu/webvowl/ (accessed on 12 August 2022)), and OVisualizer (https://github.com/yWorks/ontology-visualizer (accessed on 12 August 2022)). The new instances are inserted in the PMont and later can be visualized through a graph representation (see Figure 7). The resulting populated ontology is publicly available at https://github.com/hectorguedea/pmont (accessed on 12 August 2022).

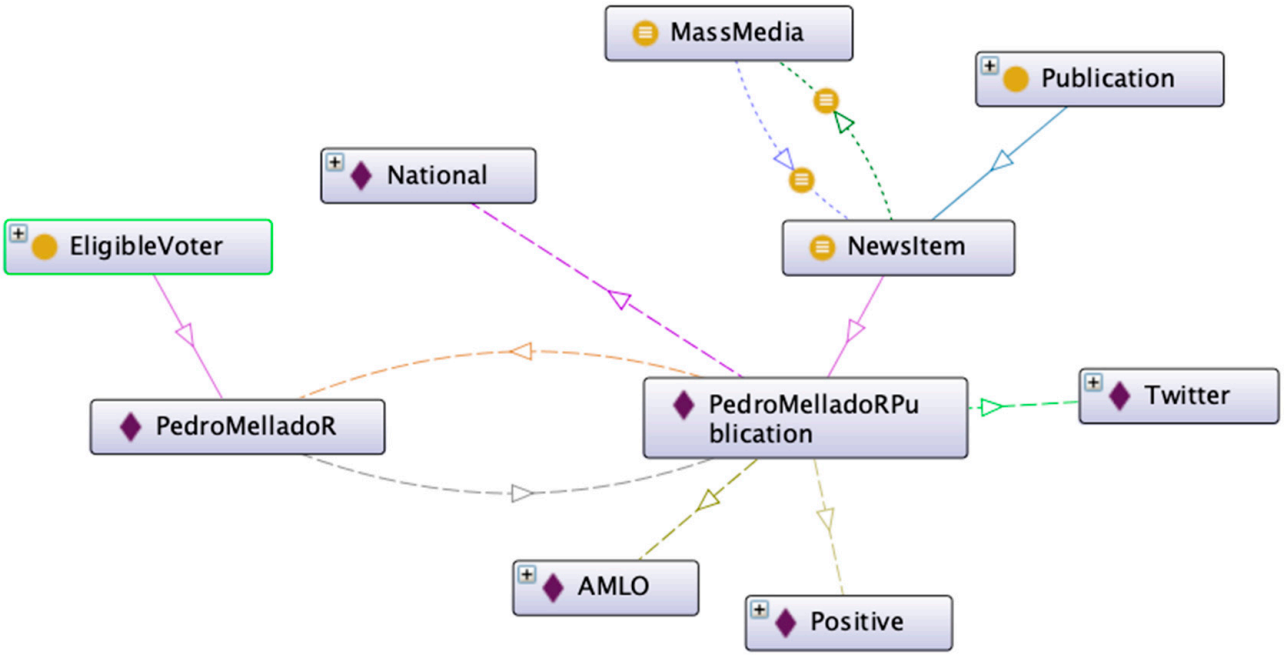

Figure 7. New instance displayed in OntoGraf.

4. Knowledge Graph Evaluation

Despite the validation step during the information retrieval process, the graph may contain syntax and semantic issues that impact its quality. A low-quality KG produces low-quality applications; therefore, an exhaustive evaluation becomes a necessary task to build high-quality KGs [62–64].

In [64], the authors highlight 'accuracy', 'consistency', 'timeliness', 'completeness', 'trustworthiness', and 'availability' as the main evaluation dimensions of the KG quality, where 'accuracy' is associated to error detection in relations, entity types, and attributes, 'timeliness' refers to global update and local update of KG, and 'completeness' means completion of relations, entity types, and attributes. Nonetheless, in our work, and based on the research [62], 18 requirements are considered when assessing the quality of a KG (see Table 2).

Table 2. Requirements of KG quality.

1. Triples should be concise	8. Data for constructing a knowledge graph should be different types and from different resources	13. Knowledge graph should be publicly available and proprietary
2. Contextual information of entities should be captured	9. Synonyms should be mapped and ambiguities should be eliminated to ensure reconcilable expressions	14. Knowledge graph should be authority
3. Knowledge graph does not contain redundant triples	10. Knowledge graph should be organized in structured triples for easy processing by machine	15. Knowledge graph should be concentrated
4. Knowledge graph can be updated dynamically	11. The scalability with respect to the KG size	16. The triples should not contradict with each other
5. Entities should be densely connected	12. The attributes of the entities should not be missed	17. For domain specific tasks, the knowledge graph should be related to that field
6. Relations among different types of entities should be included		18. Knowledge graph should contain the latest resources to guarantee freshness
7. Data source should be multi-field		

Table 2 shows a summary of the KG quality requirements and Figure 8 presents the classification of these quality requirements into four high-level dimensions. Table 3 gathers the requirement numbers to be evaluated in the KG as identified in Table 2 and Figure 8;

each cell that is filled with light grey indicates a positive response to its fulfillment; on the other hand, dark grey designates the non-compliance with such quality requirement.

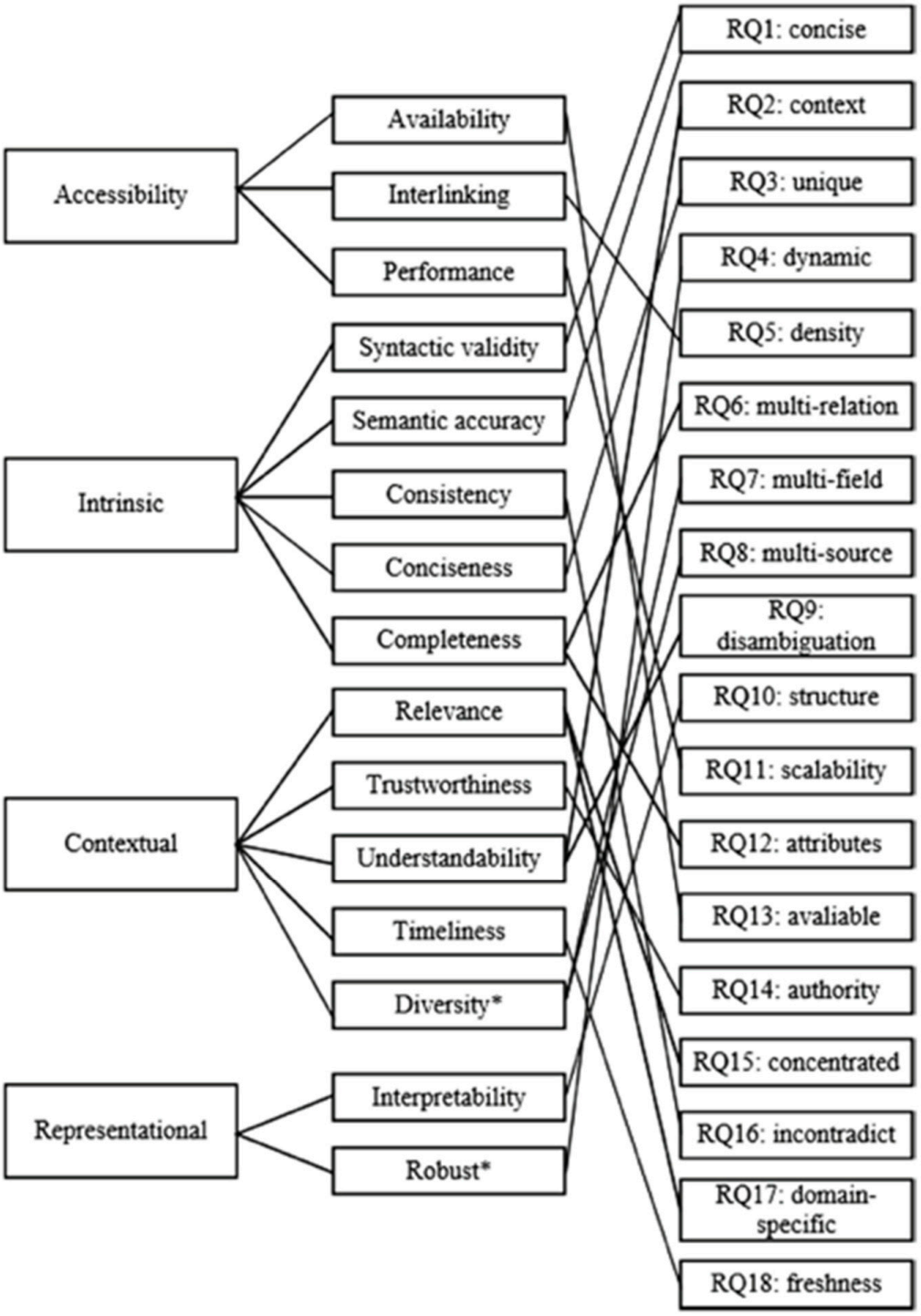

Figure 8. Dimensions of quality requirements in KGs [62].

Table 3. Evaluation results of the graph.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18		

In conclusion, the evaluation showed that the resulting graph is accurate, consistent, timely, complete, reliable, and available as dimensions of the KG quality assessment. However, the points to improve are those related to:

Quality requirement (5)—Entities must be densely connected: given the automatic nature of the KG population process suggested in this work, not all the attributes and properties associated to each entity are found in the considered data sources. Consequently, the resulting KG is only sparsely connected.

Quality requirement (9)—Synonyms must be mapped and ambiguities must be removed to ensure reconcilable expressions: Spanish content gathered from heterogeneous sources is prone to containing synonyms and ambiguous terms. It is yet to be conceived an automatic mechanism capable of dealing with such challenge with total effectiveness.

5. Conclusions and Future Work

In recent years, one of the main challenges of political marketing discussed in the literature review is the access to and management of massive social data, whose difficulty is mainly associated to the heterogeneity of the data sources caused by the massive use of social networking sites and many other digital platforms. This fact produces late (and less informed) decision making, having to spend a large amount of time in the information management and extraction processes. Moreover, it also symbolizes ignorance of the electorate, human cost, and economic cost for political marketers. To contribute to the solution of the above-mentioned challenges, the research work described in this work presents an ontological model (with the main concepts involved in the political marketing domain) which sustains a KG envisioned to facilitate the integration and reconciliation of data coming from different sources. Then, a framework to automatically populate the KG with data gathered from social media sites and news sites is proposed.

The ontological model was built by following the well-known Ontology Development 101 methodology taking into account the criteria and needs of political marketing in political campaign scenarios. Therefore, once populated, the ontology, which was baptized as PMont (Political Marketing Ontology), provides answers for each of the key questions raised, starting with the most important one: what does the electorate demand? PMont enabled the semantic storage and processing of available information about the electorate and the candidates through the possibility of consuming data from different data sources.

In order to have a gateway to integrate new data to the ontological model in the form of a graph, an automated solution was created based on ML and NLP techniques to process Spanish texts enabling the possibility to (i) extract and collect meaningful data from semi-structured and unstructured digital media sources, (ii) deal with massive data, and, finally, (iii) populate the KG previously defined upon the ontological model of the political marketing domain. The proposed automation system foresees the following components for its optimal implementation: (i) source connection and data collection, (ii) information extraction module, (iii) ontology population and validation, and (iv) the resulting KG.

The final result is a KG populated with validated, accurate, consistent, and reliable information. The KG was evaluated through 18 quality requirements on the dimensions of accessibility, contextual quality, intrinsic quality, and representational quality, which ensure the optimal integration of knowledge to be used as market intelligence for defining political campaign strategies.

There are possibilities for further development and management of this resulting knowledge graph, starting with a more robust validation process. A large corpus of text from news sites and social media sites is to be gathered and manually annotated to check the performance of the proposed automatic population framework in terms of traditional information retrieval metrics such as precision, recall, and F1-measure. On the other hand, the application of machine and deep learning techniques can be leveraged for semantic analysis of information to infer new knowledge and make better decisions during political campaigns. Another important improvement that can be applied as future work is the elimination of the relational database layer prior to the ontology population

stage; this will help to consolidate a real-time system. The end-user usefulness of the proposed system lays on the development and implementation of a visualization tool for the resulting political marketing KG and GUI (Graphical User Interface), a user interface for the system. These graphic elements will be important for defining a complete decision-making dashboard, benefiting the political campaign manager in rapid response to voter demand or crisis management.

Author Contributions: Conceptualization, H.H.G.-N. and F.G.-S.; methodology, F.G.-S.; software, H.H.G.-N.; validation, H.H.G.-N.; formal analysis, F.G.-S.; investigation, H.H.G.-N.; resources, F.G.-S.; data curation, H.H.G.-N.; writing—original draft preparation, H.H.G.-N.; writing—review and editing, F.G.-S.; visualization, H.H.G.-N.; supervision, F.G.-S.; project administration, F.G.-S.; funding acquisition, F.G.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the research project PDC2021-121112-I00 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. This work is also part of the research project LaTe4PSP (PID2019-107652RB-I00) funded by MCIN/AEI/10.13039/501100011033. Moreover, it was partially supported by the Seneca Foundation—the Regional Agency for Science and Technology of Murcia (Spain)-through project 20963/PI/18.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kotler, P.; Levy, S.J. Broadening the Concept of Marketing. *J. Mark.* **1969**, *33*, 10–15. [CrossRef] [PubMed]
2. Katz, R.S.; Mair, P. (Eds.) *How Parties Organize: Change and Adaptation in Party Organizations in Western Democracies*; Sage: New York, NY, USA, 1994; Volume 528, p. 375.
3. Ingram, P.; Lees-Marshment, J. The anglicisation of political marketing: How Blair 'out-marketed' Clinton. *J. Public Aff.* **2002**, *2*, 44–56. [CrossRef]
4. Perannagari, K.T.; Chakrabarti, S. Analysis of the literature on political marketing using a bibliometric approach. *J. Public Aff.* **2019**, *20*, e2019. [CrossRef]
5. Trent, J.S.; Friedenberg, R.V.; Denton, R.E., Jr. *Political Campaign Communication: Principles and Practices*, 8th ed; Rowman & Littlefield Publishers: Lanham, MD, USA, 2015.
6. Coto, M.A.A.; Adell, Á. *Marketing Político 2.0: Lo que todo Candidato Necesita Saber para Ganar las Elecciones*; Gestión 2000: Barcelona, Spain, 2011.
7. Kirchner, A.E.L.; Juárez, S.B.; Vite, L. *Marketing Político*, 1st ed.; Cengage Learning: Queretaro, Mexico, 2011.
8. Borgesius, F.J.Z.; Möller, J.; Kruikemeier, S.; Fathaigh, R.; Irion, K.; Dobber, T.; Bodo, B.; De Vreese, C. Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Rev.* **2018**, *14*, 82. [CrossRef]
9. Studer, R.; Benjamins, V.; Fensel, D. Knowledge engineering: Principles and methods. *Data Knowl. Eng.* **1998**, *25*, 161–197. [CrossRef]
10. Ehrlinger, L.; Wöß, W. Towards a Definition of Knowledge Graphs. CEUR Workshop Proceedings; CEUR-WS. 2016, Volume 1695. Available online: http://ceur-ws.org/Vol-1695/paper4.pdf (accessed on 23 May 2022).
11. Shadbolt, N.; Berners-Lee, T.; Hall, W. The Semantic Web Revisited. *IEEE Intell. Syst.* **2006**, *21*, 96–101. [CrossRef]
12. García-Sánchez, F.; García-Díaz, J.A.; Gómez-Berbís, J.M.; Valencia-García, R. Financial Knowledge Instantiation from Semi-structured, Heterogeneous Data Sources. In *Computer Science On-line Conference*; Springer: Cham, Switzerland, 2019; Volume 764, pp. 103–110. [CrossRef]
13. Somodevilla, M.J.; Ayala, D.V.; Pineda, I. An Overview on Ontology Learning Tasks. In *Computación y Sistemas*; Instituto Politecnico Nacional: Mexico City, Mexico, 2018; pp. 137–146. [CrossRef]
14. Petasis, G.; Karkaletsis, V.; Paliouras, G.; Krithara, A.; Zavitsanos, E. Ontology Population and Enrichment: State of the Art. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6050, pp. 134–166. [CrossRef]
15. Lubani, M.; Noah, S.A.M.; Mahmud, R. Ontology population: Approaches and design aspects. *J. Inf. Sci.* **2018**, *45*, 502–515. [CrossRef]
16. Iyer, V.; Mohan, L.; Bhatia, M.; Reddy, Y.R. A Survey on Ontology Enrichment from Text. In Proceedings of the 16th International Conference on Natural Language Processing, Hyderabad, India, 18–21 December 2019; pp. 1–10.
17. Drumond, L.; Girardi, R. A survey of ontology learning procedures. *CEUR Workshop Proc.* **2008**, *427*, 1–13.
18. Kondylakis, H.; Papadakis, N. EvoRDF: Evolving the exploration of ontology evolution. *Knowl. Eng. Rev.* **2018**, *33*, e12. [CrossRef]

19. Faria, C.; Serra, I.; Girardi, R. A domain-independent process for automatic ontology population from text. *Sci. Comput. Program.* **2014**, *95*, 26–43. [CrossRef]
20. Guedea-Noriega, H.H.; García-Sánchez, F. Construcción de una Ontología para Marketing Político. *Tecnol. Educ.* **2020**, *7*, 38–44. [CrossRef]
21. Scammell, M. Political Marketing: Lessons for Political Science. *Political Stud.* **1999**, *47*, 718–739. [CrossRef]
22. Juárez, J. Hacia un estudio del marketing político: Limitaciones teóricas y metodológicas. *Espiral* **2003**, *9*, 60–95.
23. Moore, C. *Propaganda Prints: A History of Art in the Service of Social and Political Change*; A&C Black: London, UK, 2010.
24. Ganduri, R.N.; Reddy, E.L.; Reddy, T.N. Social Media as a Marketing Tool for Political Purpose and Its Implications on Political Knowledge, Participation, and Interest. *Int. J. Online Mark.* **2020**, *10*, 21–33. [CrossRef]
25. Jain, A.; Kumar, A.; Dash, M.K. Information technology revolution and transition marketing strategies of political parties: Analysis through AHP. *Int. J. Bus. Inf. Syst.* **2015**, *20*, 71. [CrossRef]
26. Antoniades, N. Political Marketing Communications in Today's Era: Putting People at the Center. *Society* **2020**, *57*, 646–656. [CrossRef]
27. Hoppe, T.; Humm, B.; Reibold, A. *Semantic Applications*; Springer: Berlin/Heidelberg, Germany, 2018. [CrossRef]
28. Pinto, F.M.; Marques, A.; Santos, M.F. Ontology-supported database marketing. *J. Database Mark. Cust. Strat. Manag.* **2009**, *16*, 76–91. [CrossRef]
29. Guedea-Noriega, H.H.; García-Sánchez, F. SePoMa: Semantic-Based Data Analysis for Political Marketing. In *Technologies and Innovation. CITI 2018. Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2018; Volume 883, pp. 199–213. [CrossRef]
30. Fenz, S. Supporting Complex Decision Making by Semantic Technologies. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2020; Volume 12123, pp. 632–647. [CrossRef]
31. Morente-Molinera, J.; Cabrerizo, F.; Mezei, J.; Carlsson, C.; Herrera-Viedma, E. A dynamic group decision making process for high number of alternatives using hesitant Fuzzy Ontologies and sentiment analysis. *Knowl.-Based Syst.* **2020**, *195*, 105657. [CrossRef]
32. Cutrona, V.; De Paoli, F.; Košmerlj, A.; Nikolov, N.; Palmonari, M.; Perales, F.; Roman, D. Semantically-Enabled Optimization of Digital Marketing Campaigns. In *International Semantic Web Conference*; Springer: Cham, Switzerland, 2019; pp. 345–362. [CrossRef]
33. Noriega, H.H.G.; Sanchez, F.G. Semantic (Big) Data Analysis: An Extensive Literature Review. *IEEE Lat. Am. Trans.* **2019**, *17*, 796–806. [CrossRef]
34. Cotfas, L.-A.; Roxin, I.; Delcea, C. Semantic search in social media analysis. In Proceedings of the 18th International Conference on Informatics in Economy, Education, Research and Business Technologies, Bucharest, Romania, 30–31 May 2019; pp. 37–42. [CrossRef]
35. Sebei, H.; Taieb, M.A.H.; Ben Aouicha, M. SNOWL model: Social networks unification-based semantic data integration. *Knowl. Inf. Syst.* **2020**, *62*, 4297–4336. [CrossRef]
36. Milić, P.; Veljković, N.; Stoimenov, L. Semantic Technologies in e-government: Toward Openness and Transparency. In *Smart Technologies for Smart Governments*; Springer: Cham, Switzerland, 2018; pp. 55–66. [CrossRef]
37. Ahmed, J.; Ahmed, M. Ontological Based Approach of Integrating Big Data: Issues and Prospects. In *ICDSMLA 2019*; Springer: Singapore, 2020; pp. 365–378. [CrossRef]
38. Caione, A.; Paiano, R.; Guido, A.L.; Fait, M.; Scorrano, P. Technological tools integration and ontologies for knowledge extraction from unstructured sources: A case of study for marketing in agri-food sector. In Proceedings of the Creating Global Competitive Economies: 2020 Vision Planning and Implementation—Proceedings of the 22nd International Business Information Management Association Conference, IBIMA 2013, Rome, Italy, 13–14 November 2013; pp. 225–236.
39. Alazemi, N.; Al-Shehab, A.J.; Alhakem, H.A. Semantic-Based E-Government Framework Based on Domain Ontologies: A Case of Kuwait Region. *J. Theor. Appl. Inf. Technol.* **2018**, *96*, 2557–2566.
40. Laney, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety Application Delivery Strategies. 2001. Available online: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed on 23 May 2022).
41. Rossi, R.; Hirama, K. Characterizing Big Data Management. In Proceedings of the InSITE 2015: Informing Science + IT Education Conferences, Tampa, FL, USA, 29 June–5 July 2015. [CrossRef]
42. Beydoun, G.; Hoffmann, A.; Breis, J.T.F.; Béjar, R.M.; Valencia-García, R.; Aurum, A. Cooperative Modelling Evaluated. *Int. J. Cooperative Inf. Syst.* **2005**, *14*, 45–71. [CrossRef]
43. Asim, M.N.; Wasim, M.; Khan, M.U.G.; Mahmood, W.; Abbasi, H.M. A survey of ontology learning techniques and applications. *Database* **2018**, *2018*, bay101. [CrossRef] [PubMed]
44. Wimalasuriya, D.C.; Dou, D. Ontology-based information extraction: An introduction and a survey of current approaches. *J. Inf. Sci.* **2010**, *36*, 306–323. [CrossRef]
45. Ontotext, What Is a Knowledge Graph? | Ontotext Fundamentals. 2020. Available online: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/ (accessed on 19 December 2021).
46. Barrasa, J.; Hodler, A.E.; Webber, J. *Knowledge Graphs: Data in Context for Responsive Businesses*, 1st ed.; O'Reilly Media: Newton, MA, USA, 2021.

47. Kertkeidkachorn, N.; Ichise, R. T2KG: A Demonstration of Knowledge Graph Population from Ttext and Its Challenges. CEUR Workshop Proceedings; CEUR-WS. 2018, Volume 2293, pp. 110–113. Available online: http://ceur-ws.org/Vol-2293/jist2018pd_paper9.pdf (accessed on 23 May 2022).
48. Yoo, S.; Jeong, O. Automating the expansion of a knowledge graph. *Expert Syst. Appl.* **2019**, *141*, 112965. [CrossRef]
49. Xu, J.; Kim, S.; Song, M.; Jeong, M.; Kim, D.; Kang, J.; Rousseau, J.F.; Li, X.; Xu, W.; Torvik, V.I.; et al. Building a PubMed knowledge graph. *Sci. Data* **2020**, *7*, 1–15. [CrossRef]
50. Salahdine, F.; Kaabouch, N. Social Engineering Attacks: A Survey. *Futur. Internet* **2019**, *11*, 89. [CrossRef]
51. Guzmán-Guzmán, X.; Núñez-Valdez, E.R.; Vásquez-Reynoso, R.; Asencio, A.; García-Díaz, V. SWQL: A new domain-specific language for mining the social Web. *Sci. Comput. Program.* **2021**, *207*, 102642. [CrossRef]
52. Freelon, D. Computational Research in the Post-API Age. *Political Commun.* **2018**, *35*, 665–668. [CrossRef]
53. Sapountzi, A.; Psannis, K.E. Social networking data analysis tools & challenges. *Futur. Gener. Comput. Syst.* **2018**, *86*, 893–913. [CrossRef]
54. Vargas-Vera, M.; Moreale, E.; Stutt, A.; Motta, E.; Ciravegna, F. MnM: Semi-Automatic Ontology Population from Text. In *Ontologies*; Springer: Boston, MA, USA, 2007; pp. 373–402. [CrossRef]
55. Ayadi, A.; Samet, A.; Beuvron, F.D.B.D.; Zanni-Merk, C. Ontology population with deep learning-based NLP: A case study on the Biomolecular Network Ontology. *Procedia Comput. Sci.* **2019**, *159*, 572–581. [CrossRef]
56. Pech, F.; Martinez, A.; Estrada, H.; Hernandez, Y. Semantic Annotation of Unstructured Documents Using Concepts Similarity. *Sci. Program.* **2017**, *2017*, 7831897. [CrossRef]
57. Achichi, M.; Bellahsene, Z.; Ienco, D.; Todorov, K. Towards Linked Data Extraction from Tweets. In *EGC: Ex-traction et Gestion des Connaissances*; Hermann-Editions: Paris, France, 2015; pp. 383–388.
58. Bereta, K.; Papadakis, G.; Koubarakis, M. Ontop4theWeb: SPARQLing the Web On-the-fly. In Proceedings of the 2021 IEEE 15th International Conference on Semantic Computing, ICSC 2021, Laguna Hills, CA, USA, 27–29 January 2021; pp. 268–271. [CrossRef]
59. Ait-Mlouk, A.; Vu, X.-S.; Jiang, L. WINFRA: A Web-Based Platform for Semantic Data Retrieval and Data Analytics. *Mathematics* **2020**, *8*, 2090. [CrossRef]
60. Noy, N.F.; McGuinness, D.L. Ontology Development 101: A Guide to Creating Your First Ontology. Available online: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (accessed on 29 July 2020).
61. W3C, OWL 2 Web Ontology Language Document Overview (Second Edition). 2012. Available online: https://www.w3.org/TR/owl2-overview/ (accessed on 16 March 2019).
62. Chen, H.; Cao, G.; Chen, J.; Ding, J. A Practical Framework for Evaluating the Quality of Knowledge Graph. In *China Conference on Knowledge Graph and Semantic Computing*; Springer: Singapore, 2019; pp. 111–122. [CrossRef]
63. Gao, J.; Li, X.; Xu, Y.E.; Sisman, B.; Dong, X.L.; Yang, J. Efficient knowledge graph accuracy evaluation. *Proc. VLDB Endow.* **2019**, *12*, 1679–1691. [CrossRef]
64. Wang, X.; Chen, L.; Ban, T.; Usman, M.; Guan, Y.; Liu, S.; Wu, T.; Chen, H. Knowledge graph quality control: A survey. *Fundam. Res.* **2021**, *1*, 607–626. [CrossRef]