Major Minors – Ontological Representation of Minorities by Newspapers

Paulo Jorge Pereira Martins
University of Minho, Braga, Portugal

Leandro José Abreu Dias Costa
University of Minho, Braga, Portugal

José Carlos Ramalho
Department of Informatics, University of Minho, Braga, Portugal

Abstract

The stigma associated with certain minorities has changed throughout the years, yet there's no central data repository that enables a concrete tracking of this representation. Published articles on renowned newspapers are a way of determining the public perception on this subject, mainly digital newspapers, being it through the media representation (text and photo illustrations) or user comments. The present paper seeks to showcase a project that attempts to fulfill that shortage of data by providing a repository in the form of an ontology: RDF triplestores composing a semantic database (W3C standards for Semantic Web). This open-source project aims to be a research tool for mapping and studying the representation of minority groups in a Portuguese journalistic context over the course of two decades.

2012 ACM Subject Classification Information systems → Web Ontology Language (OWL); Information systems → Ontologies; Computing methodologies → Ontology engineering; Information systems → Graph-based database models

Keywords and phrases RDF, OWL, Ontologies, Knowledge Representation, Minorities

Digital Object Identifier 10.4230/OASIcs.SLATE.2021.3

Supplementary Material

Software (Source Code): `https://github.com/Paulo-Jorge-PM/crawler-majorminors`
archived at `swh:1:dir:595c4e87b736bf4c8b3ae59d50f79e4c921cc4c8`

Dataset: `https://github.com/Paulo-Jorge-PM/datasets-majorminors`
archived at `swh:1:dir:053b6e2c5e2cd710fcb8e769db46479d836965aa`

Software (Source Code): `https://github.com/Paulo-Jorge-PM/ontology-assembler-majorminors`; archived at `swh:1:dir:49e46f6b20ac99a03db72f9121ceb2760606c8c7`

Software (Source Code): `https://github.com/leandrocosta16/scraper-MajorMinors`
archived at `swh:1:dir:7396d982c983ecdd8d8e1c2ffd169fba37bd9ad4`

1 Introduction

Tim Berners-Lee, director of the World Wide Web Consortium, established the term Semantic Web and promotes the concept of converting the Web into a big collection of semantic databases: "People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics – everything rippling and folding and looking misty-on Web 2.0 and access to a semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource" [9].

"Web 1.0 started as a Read only medium; the next version Web 2.0 established itself as a Read/Write medium. Now the currently evolving version of web, Web 3.0, is said to be a technologically advanced medium which allows the users to Read/Write/Execute and also allows the machines to carry out some of the thinking, so far, expected only from the humans" [8], introducing the concept of Semantic Web and Linked Open Data.

Ontologies, the backbone of the Semantic Web 3.0, which contain the vocabulary, semantic relationships, and simple rules of inference and logic for a specific domain, are accessed by software agents. These agents locate and combine data from many sources to deliver meaningful information to the user [5].

This project explores this concept, trying to contribute to the Semantic Web initiative, by building a web of relational semantic data related to a subject, instead of a traditional web of documents. Our subject of study is the media representation of minorities in Portuguese newspapers. For this study 8 minority groups were focused (based on the research fields of the teams associated with this project): "refugees", "women", "Africans", "Asians", "homosexuals", "migrants", "gypsies" and "animals". By "minorities" we refer to macro groups with some kind of social stigma and/or struggle for equality/rights, in a broad sense. In the future we would like to add more specific minority groups according to the scientific community feedback.

Figure 1 Minorities extracted.

Survey data is a common way of studying the public perception of minorities, but "they are only conducted sporadically and they seldom use a comparable set of questions, limiting their usefulness for comparing across time, space or minority group" [2]. With this project, we propose to use ontologies built with the content of digital newspapers, from 1996 to 2019, to analyze the evolution of the representation of identity groups. Ontologies can be a powerful research tool because not only they store a corpus, but they can also be enriched with immense semantic information and relationships with additional entities.

An ontology of press clippings from Portuguese newspapers was built with references to minorities and dozens of defined entities, resulting in a corpus of articles, comments, and images featuring semantic relationships. The corpus covers the first two decades of the 21st century. The outcome of this project is an ontological database, that is, a tree of semantic relations and hierarchical references, which aims to contribute to a portrait of the representation of minorities by Portuguese media. Currently, our main corpus is the "Público" newspaper (digital version) from 1996 to 2018 (and the first months of 2019). This data was crawled through the "Arquivo.pt" national archive which collects the history of Portuguese World Wide Web (".pt" domains). This specific newspaper was used due to its popularity and the fact that it was one of the first journalistic sources to have a digital presence in Portugal.

One of the main goals of the Semantic Web initiative (Web 3.0) is to achieve a general Natural-language Interface, but it suffers from inconsistency, vastness, and vagueness [4]. Linked Open Data attempts to make this ideal come true, but it is still a work in progress. An endeavor to contribute to this ideal was conducted, following the W3C recommendations and most recent standards, intuitive interfaces were built in order to achieve an accessible product. Mainly two ontologies were developed with inferences, using OWL (RDF, Turtle, etc.) to store the data in a GraphDB database, SPARQL to query it and a VUE interface generated dynamically and having multiple endpoints to interact with it.

The ontology was built and enriched by automatically identifying thousands of entities (public figures, political parties, brands, religions, etc.) in different parts of the content (title, body, preview, tags, etc.) as well as a network of relations between them. Besides the newspaper articles themselves, two additional corpus were built: one with the photos used to illustrate the articles and another with the user comments inside the articles.

This data repository, which aims to compile and map the representation of minorities in the Portuguese press using ontologies, was specially designed to serve as study and research material for different research groups from different areas, such as Computer Engineering, Social Sciences, Linguistics, Humanities, Statistics, etc. Some research groups at the University of Minho are working on specific questions on the topics covered, such as, NLP and analysis of the evolution of written styles by the Portuguese cyber users during the last 20 years (user comments); gender studies (articles related to "women" and "sexuality" representations); post-colonial studies ("racial" representation); G2i is studying the portrait and image representation (photos used to illustrate the articles reflect the evolution of media representation of some minorities during the last 20 years); etc.

We invite the scientific community to freely conduct research and new approaches to this data. Every output was released as open-source.

Like Gillespie said, in her study related to media, minorities, and cultural change, the "globalization of communications and cultures articulates new kinds of temporal and spatial relations, it transforms the modes of identification available within societies. Media are being used by productive consumers to maintain and strengthen boundaries, but also to create new, shared spaces in which syncretic cultural forms, such as 'new ethnicities' can emerge. These processes are uneven and their consequences unforeseeable, though likely to be weighty. But they cannot be examined in the abstract or at a distance" [3].

This project tries to bring factual data to this subject, attempting to become a useful research tool. Its main roadmap can be summarized in Figure 2.

In Section 2 a preview of the project outputs and objectives will be summarized. The remaining sections will cover particular aspects concerning the implementation and development stages, namely the Section 3 will focus the data mining steps, the Section 4 will discuss the data treatment, and the Section 5 will present the ontology results and implementation. Finally, Section 6 will discuss the results and future objectives.

2 Project outcomes

From this project resulted a website with different services, two ontologies (ontology a) all the newspaper corpus; ontology b) only the corpus referring minorities) and a public API (GraphDB / SPARQL). Namely, these are some of the main endpoints/URLs:

- **Website:** `http://minors.ilch.uminho.pt`
- **GraphDB (SPARQL public API):** `http://sparql.ilch.uminho.pt`
- **Visual interface:** `http://minors.ilch.uminho.pt/search`
- **SPARQL search:** `http://minors.ilch.uminho.pt/sparql`

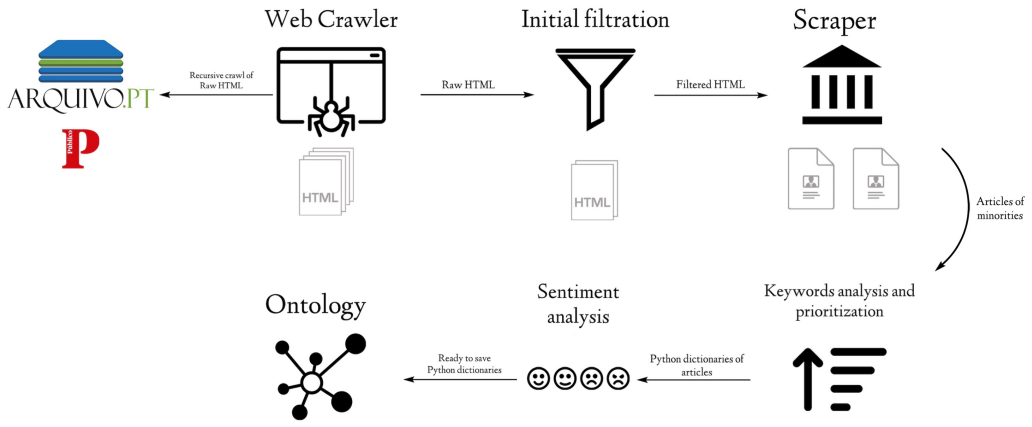

Figure 2 Data collection and processing flow of events.

- **Articles:** `http://minors.ilch.uminho.pt/articles`
- **Images:** `http://minors.ilch.uminho.pt/images`
- **Comments:** `http://minors.ilch.uminho.pt/comments`
- **WebVOWL:** `http://minors.ilch.uminho.pt/ontology`

The main objective of this project is to provide a powerful research tool for the scientific community. The project compiles and makes complex data publicly available, which would otherwise be inaccessible to the average user due to the technical complexity involved. In addition to the project website, which counts with reactive search interfaces, and repositories, a SPARQL endpoint that allows complex queries and direct download of large amounts of data is provided.

In a first instance, one can simply navigate the compiled galleries of articles, images, and commentaries in the main website. Its cached data (JSON), previously extracted through the query interfaces, with every possible result for each class, ordered by minority and priority (scoring system).

With the reactive user interface ("Visual navigation") it is possible to filter and extract more complex results: the interface dynamically builds the SPARQL queries and requests the data. One can easily search and download more complex data without technical knowledge of SPARQL or Ontologies, incorporating three possible levels of data filtering: classes, individuals, and some relationships (Figure 3).

But the potential of the platform shines when dealing directly with SPARQL. As an example, with a simple SPARQL query, it is possible to obtain data (XML, JSON, or CSV) with a set of complex and extensive criteria and deep semantic relationships, for example: "display all images that appear associated with articles published in the first quarter of 2016, but only those which are related to "women" minorities and refers in the main text the public figure "António Costa", with the profession of politician, but only if it also includes the keywords "ban" and "abortion" in the body of the text, in addition to mentioning cities in Portugal, excluding the Algarve (etc.)" – the query could be extended to infinity. In a traditional database, with simple text matches, this ease in manipulating data and entity relations, depth, and richness of results would not be possible with the same degree of precision.

Figure 3 Visual navigation.

Figure 4 shows a sample of the ontology, with one of the many articles resulting from a query that extracted articles related to a specific minority but also mentioning a particular company brand ("Disney").

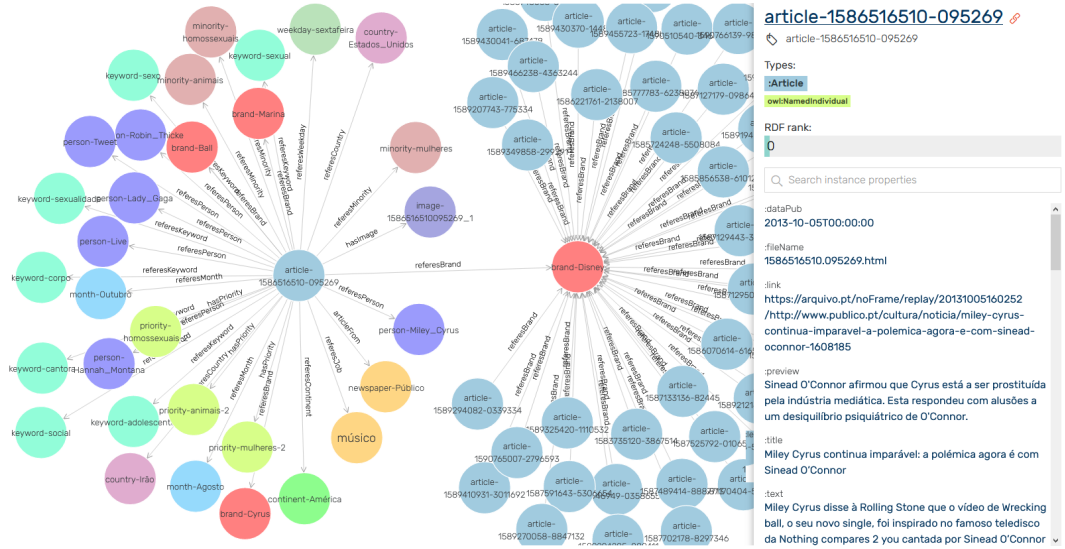

Figure 4 Example of queried data and relationships extracted from the ontology (only two classes are visually expanded: an Article and the mentioned brand "Disney").

3 Crawling Phase – Data Extraction

As a starting point, a mean was needed to obtain the raw data that would compose the ontological database. As mentioned before, the reference platform for data extraction was "Arquivo.pt". For that purpose, an open-source crawler for data-mining[1] was built.

"Arquivo.pt", the public repository used for data acquisition, provides a public API for filtering and data search (exact text matches) in their repository, but after some tests, we found it very inefficient for our objective. For example, it wasn't possible to simply search for "women minorities" or the archive would only return a very limited few pages with this exact same match, ignoring the ones with, for example, "feminist protest for the right to abortion" or "young girl was killed by her boyfriend". There was a need to develop a specific algorithm to find pages referring to semantic associations with minorities. But first, full access to the source code of each archived web page was needed in order to be able to interact with the HTML structure and optimize search criteria (title, description, body, tags, etc.). We opted to build semantic datasets, with entities related to different subjects, and searching for them in specific parts of the documents, scoring the results according to an algorithm aimed at establishing a hierarchy of relevant associations useful to identify relevant results and navigate them, for this reason, it was important to have a local corpus of the source-code.

For this reason, a personalized crawler was developed from the ground up, downloading all the "Público" newspaper web pages source-code directly from the "Arquivo.pt", crawling their repository itself directly by building a multidimensional array of inner links and recursively extracting new ones. We started the crawling process from different versions of the "publico.pt" homepage using different snapshots from the archive (we selected the last snapshot of each year and/or each website's new rebranding first edition). For each collected web page, its source HTML code was extracted and stored, and through the use of regular expressions, XPATH, "selenium" and the Python library "requests"[2], URLs found inside the HTML code were recursively collected and processed, avoiding any possibility of blind areas.

The crawler was developed using Python, for which there are multiple auxiliary similar projects (e.g.: Scrappy[3]). However, a custom web crawler was developed from scratch. This decision was based on the freedom intended in some aspects, which is maximized when custom code is used instead of libraries. Thus, full customization on the gathering process and data manipulation were achieved. The crawler is composed of a recursive algorithm that dynamically collects pages and linked pages of a specific pre-defined customizable domain. As mentioned, the used domain was the "Público" newspaper but the algorithm was developed to operate with any defined domain, so other newspapers can be used in the future. Moreover, since millions of pages may be collected from a specific domain, a save logs system was introduced so whenever the crawler stops, the collection can be resumed at any time.

Initially, we ended up with 6 million HTML files, but the majority were repeated content. The reason for this is related to the snapshot system of "Arquivo.pt", we extracted annual snapshots identified by timestamps ending up repeating past content for each one (on purpose, with the objective of covering non-archived versions from different snapshots). "Público" newspaper had half a dozen of different website remakes and more than one domain in its

[1] Crawler's GitHub: `https://github.com/Paulo-Jorge-PM/crawler-majorminors`
[2] `https://docs.python-requests.org/en/master/`
[3] `https://scrapy.org/`

history, this also increased the number of the repeated snapshots. A filtering step was needed to eliminate duplicated content. First, we eliminated non-relevant files (such as marketing campaigns, technical error pages,etc.), selecting only files possibly with articles. In a second step, we used two criteria for isolating individual non duplicated articles: filtered HTML files by their individual unique URIs (but because of the different domains used by the newspaper in the past this was not enough) and later filtered these by individual non repeated titles inside the articles.

After this initial filtering process we ended up with 94.309 individual pages URIs and 48.949 individual articles filtered by duplicated titles (see Table 1). We should add that we ended up with some discrepancies: some years had more extracted content than others (see Figure 5 for total results referring minorities). For example, the newspaper's first year is 1996, but "Arquivo.pt" only has a snapshot of half a dozen pages from that time. Also, during its first years, the newspaper had not much digital content created. This low volume of articles continued until 2004. After this period the "Público" newspaper started many rebranding processes with its digital image and website template, including different domains ("publico.pt", "publico.clix.pt", etc.). This resulted in different HTML source codes, some of them not fully accessible, others dynamically generated and harder to crawl or not collected by the snapshot. Also, the newspaper has a subscription system, blocking access to non-paying users, which nowadays it gives full access to the "Arquivo.pt", but in past years many snapshots instead of the full content showed the blocked access message or cut text. These, and other variables, resulted in different numbers of collected articles for different years and for this reason, the number of minority references found for each year should be treated with caution. Also, the comment system changed many times and limited the archiving, for example after 2010 a login system blocked the "Arquivo.pt" snapshots and we could not retrieve the most recent user comments.

Table 1 Crawled output.

Type	Number
Newspapers articles (individual URIs)	94.309
Newspapers articles (non repeating titles)	48.949
Articles referring minorities (non repeating)	11.496
User comments related with minorities	3.348
Newspapers illustrations related with minorities	9.658

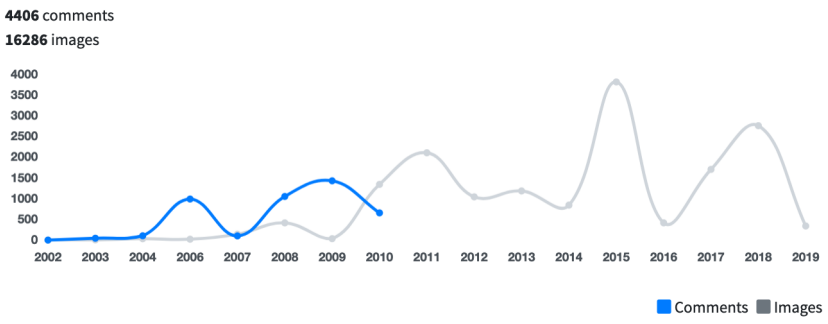

Figure 5 Total number of images and comments filtered per year.

Listing 1 Time intervals of structural changes in Público's articles.

```
end_date_publico2001_2002 = datetime.datetime(2003, 2, 14)
end_date_publico2003_2004 = datetime.datetime(2005, 2, 22)
end_date_publico2005_2006_2007_Jun = datetime.datetime(2007, 5, 18)
end_date_publico2007Jun_Nov = datetime.datetime(2007, 11, 30)
end_date_publico2007Nov_2008_2009Set = datetime.datetime(2009, 9, 30)
end_date_publico2009Set_16Nov2012 = datetime.datetime(2012, 11, 17)
end_date_publico16Nov2012_Out2016 = datetime.datetime(2016, 11, 21)
```

4 Scraping Phase – Data Mining

Once the raw HTML pages were collected by the Crawler, further filtration proceeded. For that purpose we built an open-source scraper for data-extraction[4].

Beforehand, in order to perform the filtering, Python dictionaries were created containing a collection of keywords often associated with the minority groups within the defined scope. These keywords were the result of a study of word occurrences and textual patterns in minority-related articles, which can assist with the identification of a possible minority reference.

In an initial phase, before any cleaning and segregation of elements were preceded, every HTML file which did not contain any of the selected keywords within its structure were excluded. This initial filtration avoided unnecessary post-processing of dispensable pages.

Following the removal of unwanted pages, it was noticed that a considerable part of the collected pages were not article pages, being most of them landing pages, which are also part of the "Público" newspaper domain. The "Arquivo.pt" APIs[1] from which pages were collected, does not feature specific page types filtering, resulting in a mixture of multiple different types of pages. However, the only intended pages were articles, which had a specific pattern of HTML elements. By checking for article-exclusive patterns of HTML elements it was possible to perform another filtration process.

As a result of both filtration processes, all the HTML files left were articles that mentioned minorities-specific keywords. This initial filtering took place before any intensive scraping due to the posterior use of complex regular expressions which are CPU intensive tasks.

Finally, the main scraping algorithm was executed. This algorithm was developed by analyzing the structure evolution of the articles pages from Público. Their structure changed periodically, and a different use case was added to the algorithm for each iteration of the structural changes.

Before entering the main scraping algorithm the date of the page was verified against the defined time intervals (see Listing 1) which then invokes the corresponding function. Each one of the functions start by looking for specific HTML elements, using a combination of the beautifulsoup [5] library and regular expressions. Depending on the date of the article, several elements were collected and stored in separate variables. The elements were Main Title, SubTitle, Preview, Corpus, Tags, Comments and Image URLs.

Once all elements were segregated, each one of them go through a prioritization algorithm. This algorithm made use of the keyword dictionaries and assigned a priority value to the article, depending on the element where the keywords were found. Priority values were assigned as in Table 2.

[4] Scraper's GitHub: `https://github.com/leandrocosta16/scraper-MajorMinors`
[5] `https://pypi.org/project/beautifulsoup4/`

Table 2 Priority points associated with each article elements.

Elements	Priority points
Main tile	7
Tags/Topics	5
Literal mention of the minority name	4
Subtitle/Headline	3
Preview	3
Corpus	2

Firstly, minorities are detected individually for each collected element of the article. Once the minorities are identified for a given element, a data structure containing the identified minorities is iterated and for each one of them the associated element priority value is added cumulatively. Moreover, if a given article references two or more different minority groups it will assign a priority value for each one of them separately.

Although the same list of keywords is used to classify the priority rate, there are some which directly relate to a given entity (e.g.: the literal name of a minority), and those must hold a higher priority value. For this reason, a separate list containing top priority keywords was used and articles which mention them would acquire extra priority points.

This points-based prioritization system provided a way of ordering articles by relevance, with further filtering options in perspective. Higher priority articles are likely to be highly related to the minority in question.

5 Ontology

The main outcomes of this project are the developed ontologies and interfaces to interact with them.

"An ontology is a formal description of knowledge as a set of concepts within a domain and the relationships that hold between them. To enable such a description, we need to formally specify components such as individuals (instances of objects), classes, attributes, and relations as well as restrictions, rules, and axioms. As a result, ontologies do not only introduce a shareable and reusable knowledge representation but can also add new knowledge about the domain" [6]. For example, the YAGO (Yet Another Great Ontology) [7] is one of the largest ontologies in the Linked Open Data cloud, being in itself a knowledge base. This project also aims in being a smaller knowledge base focused on a particular domain of study.

The developed interfaces to interact with the resulted ontologies can be divided in three layers: a basic UI gallery through the main website; a reactive interface that generates SPARQL queries automatically in three layers of filtration (general classes, specific individuals and basic relationships); and a complete SPARQL API and/or visual navigation graph (through GraphDB).

This project was built upon W3C official recommendations: RDF, OWL, and SPARQL. RDF represents information using semantic triples: subject, predicate, and object. We opted for the Turtle syntax for this representation. For storing and managing our triplestores we opted for the GraphDB database[6]. From their official documentation: "GraphDB is an enterprise-ready Semantic Graph Database, compliant with W3C Standards. Semantic

[6] GraphDb: `https://graphdb.ontotext.com/`

graph databases (also called RDF triplestores) provide the core infrastructure for solutions where modeling agility, data integration, relationship exploration, and cross-enterprise data publishing and consumption are important".

GraphDB has free and commercial versions. We implemented the first one, but it is limited by a maximum of two queries in parallel. This added to the fact that our project runs on an older limited server makes the performance not ideal for dealing with many concurrent connections. We created a caching system for improving this limitation, but even so, performance is not our target at the moment, due to the lack of budget, our main focus was accessibility.

In Figure 6 we can see the architecture and inner sequence of the blocks that compose this platform.

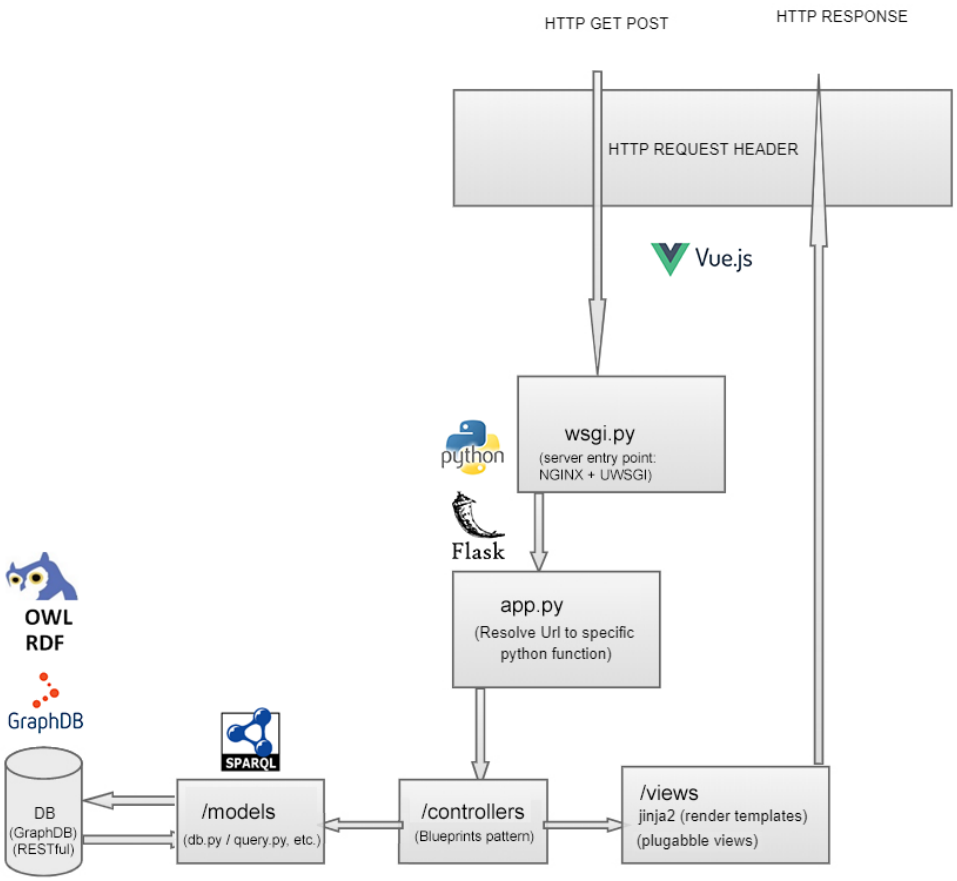

Figure 6 Website: sequence diagram.

Beforehand, preceding the ontology development, datasets were created for each individual set of entities that we aimed to search and build relationships within the articles (see Table 3). These entities (political parties, public figures, brands, etc.) were presented in the previous sections. These datasets were extracted using crawlers, Regular Expressions, XPATH (etc.) from dozens of sources, and Linked Open Data. They are open source and documented in GitHub[7].

[7] Dataset's GitHub: `https://github.com/Paulo-Jorge-PM/datasets-majorminors`

Table 3 Entities extracted.

Type	No. of References
Public Figures	32.648
Political Parties	8.525
Minorities	11.530
Entities	11.473
Brands	46.629
Religions	260
Sports	4.333
Ethnicities	1.345
Car brands	969
Animals	6.199
TV Channels	3.799
Continents	7.119
Countries	28.794
Cities	28.933
Other Places	1.067
Newspapers defined tags	10.186
Weekdays	20.575
Months	24.113
Minorities related keywords	36.897

These entities datasets were used to build the final ontology and the scraper provided inner metadata related to each article (sections, structure, minorities references, etc.). From this point, it was time to build a tree of relationships between these articles (and their images and comments) and the multiple entities referenced in their texts.

For this purpose we built an open-source ontology assembler, that dynamically searched and compared the dataset files with the article's metadata (using Regular Expressions, ontology skeletons, and structured dictionaries) to build the final ontology triplestores and automatically expand the graph tree. It is open-source and its code can be found in GitHub[8].

In Figure 7 a summary of the main classes defined for the ontology is represented and Table 4 is a summary of the ontology outputs. At `http://minors.ilch.uminho.pt/ontology` it is possible to find a more detailed visualization converted to WebVowl with the classes and their properties.

6 Conclusions and Future Work

The intent behind the development of this project is to provide a powerful research tool for the scientific community. The project compiles and makes complex data publicly available, which would normally be inaccessible to the average user due to the technical barrier and complexity overheads.

One of the goals of the Semantic Web is to achieve a general natural-language interface that is both intuitive and universal, fed by a global network of semantic databases. W3C recommended standards like Resource Description Framework (RDF), SPARQL, XML, Web

[8] Ontology assembler: `https://github.com/Paulo-Jorge-PM/ontology-assembler-majorminors`

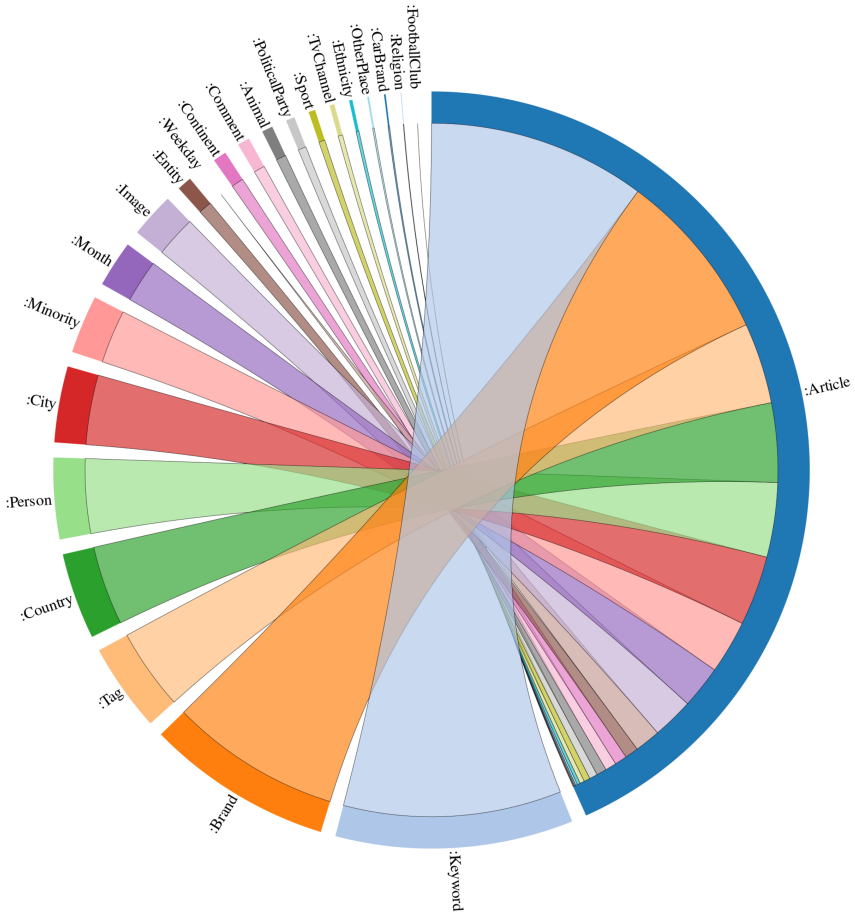

Figure 7 Some of the classes defined in the ontology.

Table 4 Ontology output.

Type	Number
Ontology (All) triples	5.178.169
Infered (All) triples	2.084.435
Ontology (All) size	650,3 Mb
Ontology (Minorities) triples	1.110.411
Infered (Minorities) triples	453.330
Ontology (Minorities) size	125,7 Mb

Ontology Language (OWL) (etc.) have been expanding this area and introducing the concept of Web 3.0, however, the holy grail is still far from being reached: it suffers from inconsistency, vastness and vagueness [4].

This bridge between Semantic Web and Natural-language User Interfaces is one of the main challenges for its growth. With this project, we hope to invite new users to explore the Semantic Web potentialities.

The project currently is in a preliminary stage, with collected data from only one newspaper as a source of data. For future work, the ideal would be to incorporate more sources, extracting data from other main Portuguese newspapers besides the "Público". This requires time and resources, but the platform and tools were built already with this in mind, the ontology structure already incorporates a class for different types of newspapers. In fact, this project was built with flexibility in mind, one could expand it to new languages, countries, or even themes (instead of "minorities" one could clone a similar semantic database and interfaces for extraction of other fields of study).

We invite research groups to freely conduct research upon this tool, they are open-source. On the main website we plan on disclosing research outputs related to the project for future reference, we gladly invite new additions.

References

1 arquivo. pwa-technologies, April 2021. [Online; accessed 19. Apr. 2021]. URL: `https://github.com/arquivo/pwa-technologies/wiki`.

2 Erik Bleich, Hannah Stonebraker, Hasher Nisar, and Rana Abdelhamid. Media portrayals of minorities: Muslims in british newspaper headlines, 2001–2012. *Journal of Ethnic and Migration Studies*, 41(6):942–962, 2015.

3 Marie Gillespie. *Television, ethnicity and cultural change*. Psychology Press, 1995.

4 Catherine C. Marshall and Frank M. Shipman. Which semantic web? *The fourteenth ACM conference*, page 57, 2003. `doi:10.1145/900062.900063`.

5 Robin D Morris. Web 3.0: Implications for online learning. *TechTrends*, 55(1):42–46, 2011.

6 OntoText. What are Ontologies? Ontotext Fundamentals Series, April 2021. [Online; accessed 12. Apr. 2021]. URL: `https://www.ontotext.com/knowledgehub/fundamentals/what-are-ontologies`.

7 Thomas Pellissier-Tanon, Gerhard Weikum, and Fabian Suchanek. Yago 4: A reason-able knowledge base. *ESWC 2020*, 2020. URL: `http://yago-knowledge.org`.

8 Rajiv and Manohar Lal. Web 3.0 in Education. *International Journal of Information Technology*, 3(2):973–5658, 2011.

9 Victoria Shannon. A "more revolutionary" Web - The New York Times, 2006. URL: `https://www.nytimes.com/2006/05/23/technology/23iht-web.html`.