ISSN 2007-9737

Semantic MOCIBA 2021: A Vocabulary for Cyberbullying based on Open Data Analysis

María Auxilio Medina Nieto[1,*], Jorge de la Calleja Mora[1], Eduardo López Domínguez[2], Yesenia Hernández Velázquez[3], Delia Arrieta Díaz[4]

[1] Universidad Politécnica de Puebla,
Departamento de Posgrado,
Mexico

[2] Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional,
Departamento de Computación,
Mexico

[3] Universidad Veracruzana,
Facultad de Estadística e Informática,
Mexico

[4] Universidad Juárez del Estado de Durango,
Facultad de Economía,
Mexico

{maria.medina, jorge.delacalleja}@uppuebla.edu.mx, eduardo.lopez.dom@cinvestav.mx,
zS19019679@estudiantes.uv.mx, dad@ujed.mx

Abstract. Information and Communication Technologies are present in homes, cultural, work and academic environments to improve quality of life, however, they are also means for *harassment* or *cyberbullying*, a form of violence that affects the mental and physical health of internet users. Annually, staff of the National Institute of Statistics and Geography conducts interviews to collect anonymous data on the prevalence of cyberbullying via a survey called *MOCIBA*. Until now, the six applications of this survey have had a distinct thematic coverage and questionnaire, as a consequence, heterogeneos open datasets for results are produced making analysis over time difficult. This paper presents Semantic MOCIBA 2021, an original ontology and vocabulary dedicated to the exploitation of MOCIBA 2021 dataset. The goal is to significantly improve data reuse by providing a standardized vocabulary using Semantic Web technologies and ontologies. The paper describes the development process of this vocabulary from scratch to form enriched datasets where concepts and relationships are formalized to represent and reason via linked data. The Semantic MOCIBA 2021 vocabulary can serve as a reference resource and practical tool for students and practitioners for information systems communities and to support the decision-making process and the generation of actions against cyberbullying by individuals or organizations in the academic or social sector based on the evidence distributed as open data.

Keywords. Open data, vocabularies, ontologies, cyberbullying, semantic web, linked open data.

1 Introduction

Information and Communication Technologies (ICTs) are present in homes, cultural, work and academic environments. The International Telecommunication Union (ITU) estimates that approximately 5.3 billion people (66% of the world's population) used the Internet in 2022, while 2.7 billion people were still offline [15]. In Mexico, the

percentage of internet users was 75.63% of the population of 2021 [11].

In many environments, ICTs improve quality of life. However, they are also means for *harassment* or *cyberbullying*. This phenomenon with negative impacts on society is defined in [14] as "an intentional act, either by an individual or a group, aimed at harming or annoying a person by means of ICTs, in particular, the internet".

Cyberbullying is a form of violence that affects the mental and physical health of people. Annually, staff of the National Institute of Statistics and Geography (INEGI) collect anonymous data on the prevalence of cyberbullying via a survey called *MOCIBA*, by its name in Spanish: *"Módulo sobre Ciberacoso"* [12], the Cyberbullying Module under study in this research. The purpose of MOCIBA is to generate statistical information to know the prevalence of cyberbullying among the population aged 12 years and older who are internet users, as well as the characterization of those who experienced a cyberbullying situation in the last 12 months, including the identity and sex of the stalker, frequency of cyberbullying and consequences for the victim.

The staff of INEGI conducts interviews and collects the MOCIBA data using a questionnaire formed by closed and open questions. The results are presented at local and national level and they are disseminated from the INEGI web site [12] in two formats. On one hand, in a report entitled *main results* such as [14] of 2021, a PDF file that also includes a comparison of 2020 and 2021 data. On the other hand, raw data are available as *open data*, a set of files in CSV[1] format.

Until now, the six applications of MOCIBA have had a distinct thematic coverage and questionnaire, as a consequence, heterogeneous datasets for results are produced making analysis over time difficult. In order to address the difficulties in the management and meaning of MOCIBA data, this paper presents a resource named *Semantic MOCIBA 2021*, a dedicated ontology and vocabulary to represent and organize concepts of the MOCIBA applied during 2021. The goal is to significantly improve reusability by providing a standardized vocabulary using Semantic Web technologies and ontologies.

The vocabulary emerged on the basis of raw data analysis gathered from August 2020 to September 2021, the most recent data available at INEGI web site [12]. The vocabulary was designed to the general public with focus on information systems communities, this is publicly available online[2].

The paper is organized as follows. Section 2 deals with related work. Section 3 presents the development process of the Semantic MOCIBA 2021 vocabulary, this includes the description of main concepts, relationships and relevant instances. Section 4 provides examples of how to use the vocabulary and discuss its potential. Finally, Section 5 contains the conclusions and the future work.

2 Related Work

Around the world, information from public administrations at local, regional, national or international level use open data portals. To begin with related works, the project described in [6] deals with cyberbullying and bullying as forms of violence against women, reflecting unequal power relations between women and men based on current national and Slovenian activities. The purpose was to determine the incidence of cyberbullying since a gender perspective. Although the theme in this project and our research is similar, the dataset are no longer accessible and therefore a comparison cannot be made.

Secondly, we found two datasets available since *data.europa.eu* web site, the official portal for European data. The first dataset is described in [5], this refers to an annual population survey on safety, quality of life and victimisation, collecting data on neighborhood nuisance, disrespectful behavior, prevention measures, police performance and municipal safety policy for the Netherlands. As a result, security figures are reported at national, regional and local levels. The latest data for this survey is 2012, only the statistical

[1] The acronym CVS refers to Comma Separated Values

[2] http://www.mauxmedina.com/vocabularies/

information is distributed in a tabular format but in dutch language.

[7] describes the second dataset, this is associated with two categories: 1) population and society, and 2) health. English and Spanish data are available in formats such as CSV, HTML[3] and Resource Description Framework (RDF) [24], a standard of the World Wide Web Consortium (W3C) originally designed as a data model for metadata that is also used as a general method for description and exchange of graph data [16]. The dataset is a CSV file composed of statistical information represented in 3 columns and 241 rows; a graphical user interface (GUI) allow users to query data by community or autonomous city, genre and frequency of discrimination, the GUI is available at: `https://ine.es/jaxi/Tabla.htm?tpx=51503&L=1`.

Thirdly, we carried out a systematic review of the Linked Open Vocabularies web site [21], a widely-used catalog of vocabularies available for reuse with the aim of describing data on the web. As of the date of January 6th 2023, any of the 782 vocabularies deals with the cyberbullying domain.

Finally, we focused our efforts on review literature related with the process of working with messy data, ETL[4] tools and transformation of tabular data into RDF tuples. Although the domain of the guide described in [26] is biodiversity, its application allow users to estimate and improve the quality of datasets.

A domain-independent and detailed description of the use of vocabularies and ontologies to represent data content and links can be found in [2]. Furthermore, available at the Spanish government portal, the guide [4] includes best practices, tips and workflows for the efficient and sustainable creation over time of datasets that according with the author's point of view, bring greater economic and social value to citizens.

Related works mentioned above are focused on datasets, vocabularies and linked open data. In contrast, our research analyzed a large dataset of raw data, with the added value of following a domain-independent ontology design approach.

[3] HyperText Markup Language

[4] ETL is the acronym for Extract, Transform and Load data from multiple sources to a data warehouse or other unified data repository

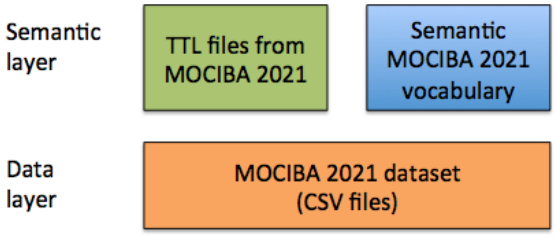

Fig. 1. Framework for the proposed vocabulary

3 The Development Process of the Semantic MOCIBA 2021 Vocabulary

Besides usefulness and reusability, the motivation for developing Semantic MOCIBA 2021 is based on the understanding, in the information systems communities, that consuming open data requires to take into consideration the relations among the perspective of data producers and different types of consumers. Therefore, we executed our research taking into account a multi-disciplinary approach of participants and authors, (as is described in Section 3.1), as well as the framework of Figure 1, where "TTL files" refer to RDF files that uses the sintaxis of TURTLE.

3.1 Open Data Analysis

The open data analysis carried out to design the proposed vocabulary is presented according to the information of six applications of MOCIBA by identifying terms and their relationships, with special emphasis on MOCIBA 2021 dataset, that constitute the most recent data.

3.1.1 Applications of MOCIBA

As it was mentioned on Section 1, the MOCIBA raw data for each year are available as open data and distributed in a compressed file (*.zip format) that includes the following 5 folders: 1) *metadata*, 2) *data dictionary*, 3) *catalogues*, 4) *model_entity_relationship* and 5) the *dataset* itself, a large CSV file.

Table 1. Themes directly related with cyberbullying in MOCIBA's questionnaires

ID	Theme
T1	Cibyerbullying situations experienced
T2	Effects on the victim
T3	Frecuency of cyberbullying
T4	Measures against cyberbullying
T5	Media used for cyberbullying

Table 2. Number of possible answers per theme

ID	2015	2016	2017	2019	2020	**2021**
T1	4	10	10	10	10	**12**
T2	0	9	9	9	10	**15**
T3	5	5	3	4	5	**5**
T4	0	0	0	9	10	**13**
T5	6	7	2	4	4	**4**

Thematic coverage and the questionnaires have been different over time as is summarized in the tables 1 and 2; the numbers in Table 2 refer to the possible answers per theme. Furthermore, take into account that the questionnaire of 2015 had 10 questions, 12 questions were included for the questionnaires of 2016, 2017, 2019 and 2020, while the latest has 14 questions (2021); the survey was not conducted in 2018.

Table 3 shows the size of the datasets that integrate the results per year, note that the 2015 corresponding dataset is not available. As a consequence, the data analysis over time is not a simple task and requires valuable human efforts and high processing capacity for computers.

Our research is focused on MOCIBA 2021 dataset, the most recent data as is described in the following section.

3.1.2 The MOCIBA 2021 Dataset

The MOCIBA 2021 dataset gathers geographic and statistical information about cyberbullying considering a spacial coverage at national and local level that covers the following key themes:

Table 3. Size of the MOCIBA datasets per year

Year	Columns	Rows
2016	208	91,675
2017	172	33,566
2019	177	10,145
2020	228	37,867
2021	**277**	**40,491**

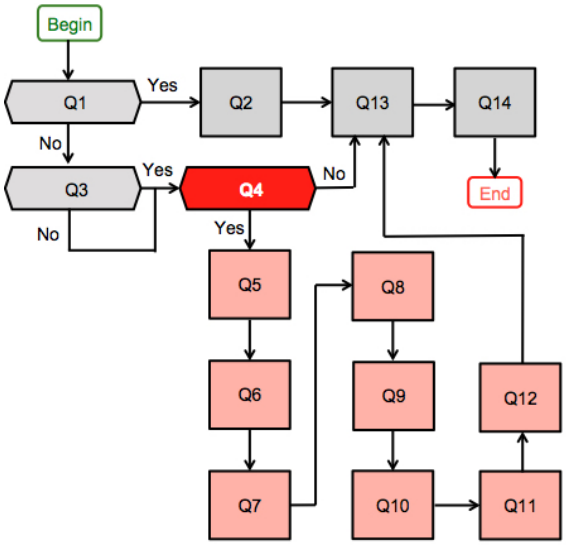

Fig. 2. Sequence of questions

— Condition of use of security measures,

— Implemented security measures, and

— Characterization of cyberbullying situations experienced.

Figure 2 shows the sequence of the questions that form the questionnaire of 2021. The notation is as follows: the rhomboids represent the dichotomous questions and the squares the multiple choice questions. The questions from Q5 to Q12 refer to one or more cyberbullying situation experienced (Q4), a maximum of three situations is recorded per each surveyed person.

The MOCIBA 2021 dataset is accesible from the INEGI website [12] and allow users to verify the

content of the report [14]; some of its features are the following:

— The dataset is formed by 277 columns and 40,491 rows (see Table 3),

— The interpretation of the data is based on catalogs, there is one catalog for each column,

— The data dictionary has 2,593 elements,

— The information gathered in each of 'Other' option of the questionnaire is not part of the dataset,

— The image of the *model_entity_relationship* folder only shows the name and data types of 4 columns.

The analysis and interpretation of the data requires potential users to simultaneous manage the information, that represents a high cognitive load prior to its reuse. Some tasks carried out by the vocabulary development team are summarized as follows:

1. Content review of the folders *catalogues*, *dataset*, *dictionary*, *metadata* and *entity-relationship model*,
2. Study of the dictionaries,
3. Column grouping of the dataset by thematic coverage,
4. Transformation of the rearranged dataset by integrating information from catalogues and dictionaries.

The tasks 3 and 4 were implemented using the OpenRefine software tool [8], the version 3.7 for Mac OS computers and 3.6.3 for Windows.

3.1.3 Identification of General Concepts and Basic Relationships

As a result of the tasks described in the Section 3.1.1 and 3.1.2 was the identification of general concepts and basic relationships about the characterization of cyberbullying. The development team built a graphical representation of these items to introduce the vocabulary to the general public, and asked the participation of 20 persons (male (♂):10, female (♀):10) for feedback; there was no answer of 6 persons (♂).

The profile of the 14 participants was as follows: 8 higher education teachers from 5 Mexican universities (♂:3, ♀:5), 3 masters and 1 Ph.D. student (♂:1, ♀:3) and one person with an administrative role (♀:1). After four versions and a consensus-based negotiation phase, the representation called *the conceptual model of Semantic MOCIBA 2021 vocabulary* was produced, this is illustrated in Figure 3, note that this combines elements of sets theory, flow charts and class diagrams which are widely-used in information systems communities.

3.2 Vocabulary Building

The Semantic MOCIBA 2021 vocabulary is built under the next key concepts: persons, cyberbullying and digital media. At present, there are two versions of this vocabulary, the first one uses labels only in Spanish language [18] while the second one also has labels in English language [19], both are available on the following web site: `https://www.mauxmedina.com/vocabularies/`. From now on, the figures and tables refer to the second version.

The vocabulary formalizes and specifies the conceptual model and reuse the MOCIBA 2021 dataset by implementing the steps proposed in [9], a domain-independent ontology design approach. The steps are the following:

1. Consider reusing existing ontologies,
2. Enumerate important terms in the ontology,
3. Define the classes and the class hierarchy,
4. Define the properties of classes - slots,

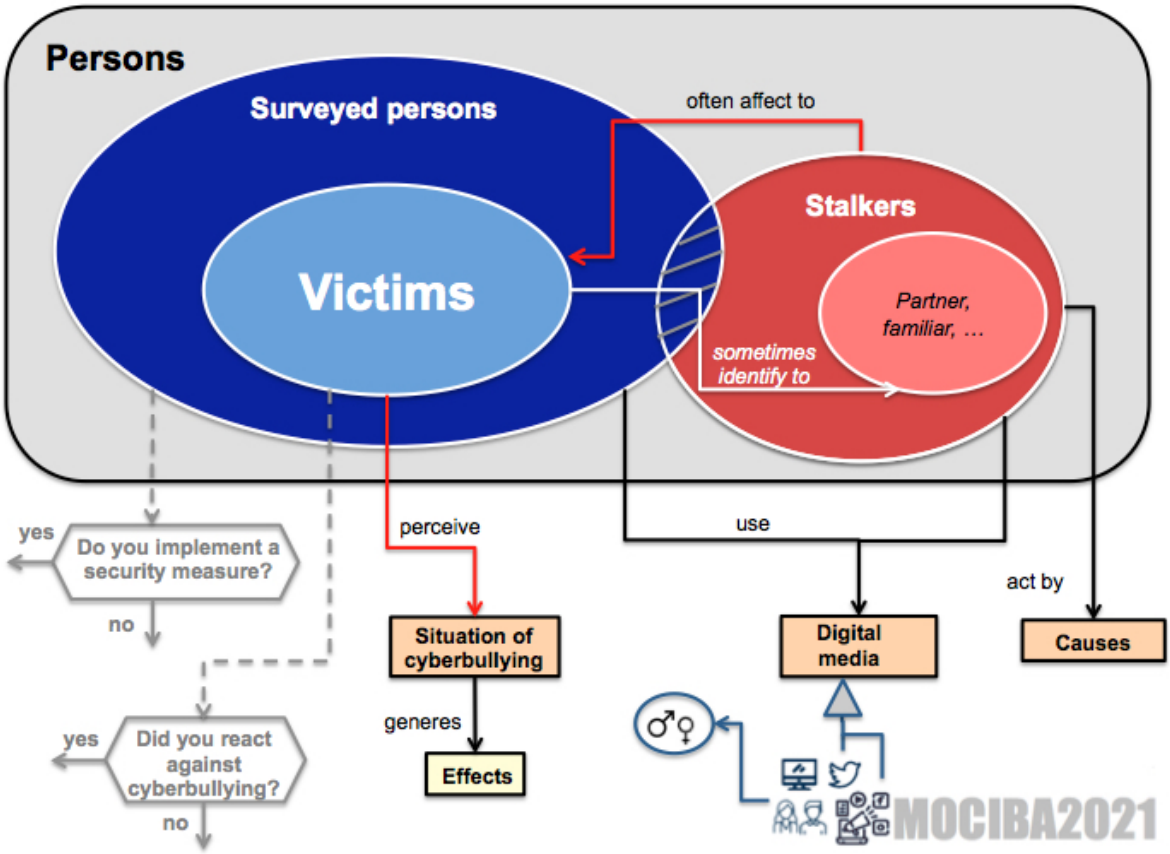

Fig. 3. The conceptual model of Semantic MOCIBA 2021 vocabulary

5. Define the facets of the slots,
6. Create instances.

The next sections describe those steps, they were implemented by using the Protégé editor [20].

3.2.1 Domain, Scope and Reuse of Ontologies

The domain of Semantic MOCIBA 2021 vocabulary is cyberbullying. The scope is the representation of terms and relationships found on the INEGI website [12], this includes the dataset itself, the report [14] and the questionnaire [13].

Table 4 shows the *competency questions* (CQs) that drive the vocabulary building. More information about CQs can be found in [10]. The reused ontologies and their prefixes are illustrated in Figure 4.

3.2.2 Terms, Classes and Hierarchy of Classes

After a systematic review of the documents enumerated in Section 3.2.1, the terms were extracted to define and construct the class hierarchy showed in Figure 5. The two main classes are: `Person` and `Semantic MOCIBA 2021`. Note that the name of the classes are in English language except the `AcosoCibernetico` and its equivalent class `Ciberacoso`, the purpose is to include *acoso cibernético* as synonym of cyberbullying and to preserve its definition according to MOCIBA 2021 report [14] in both versions of the vocabulary.

In the first place, the `Person` class is defined by the FOAF ontology [3], actually extended as Vocabulary of a Friend (VOAF[5]), this is used

[5]https://lov.linkeddata.es/vocommons/voaf/v2.3/

Table 4. Competency questions

ID	Competency question
CQ1	What are the cyberbullying situations experienced?
CQ2	What security measures do surveyed persons implement?
CQ3	How people relate to each other in cyberbullying situations?
CQ4	What digital media do stalkers use?
CQ5	What are the effects of cyberbullying?

Ontology prefixes:

Prefix	Value
	http://www.mauxmedina.com/vocabularies/mociba2022
cc	http://creativecommons.org/ns#
data-view	http://www.w3.org/2003/g/data-view#
dc	http://purl.org/dc/elements/1.1/
foaf	http://xmlns.com/foaf/0.1/
ns	http://www.w3.org/2003/06/sw-vocab-status/ns#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
vann	http://purl.org/vocab/vann/
voaf	http://purl.org/vocommons/voaf#
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#

Fig. 4. Prefixes and values of reused ontologies

to represent three types of persons: stalker, surveyed persons or victims; a victim is defined as a person who has suffered a cyberbullying situation. Due to MOCIBA 2021 dataset is formed by anonymous data, a different use for the Person class is reserved for future scenarios. In the second place, the `Semantic MOCIBA 2021` class groups most of the thematic coverage by including the following subclasses: `Action against cyberbullying`, `Digital media`, `Effect on the victim`, `Harassment situation experienced`, `Identity of stalker`, `Motivation of the stalker` and `Security measure`. These classes are associated with the questions `Q2`, `Q4`, `Q5`, `Q8`, `Q10`, `Q11` and `Q12` of Figure 2 and are populated by instances that represent their possible answers, (see Figure 6). The instances whose name ends with `html` are metadata.

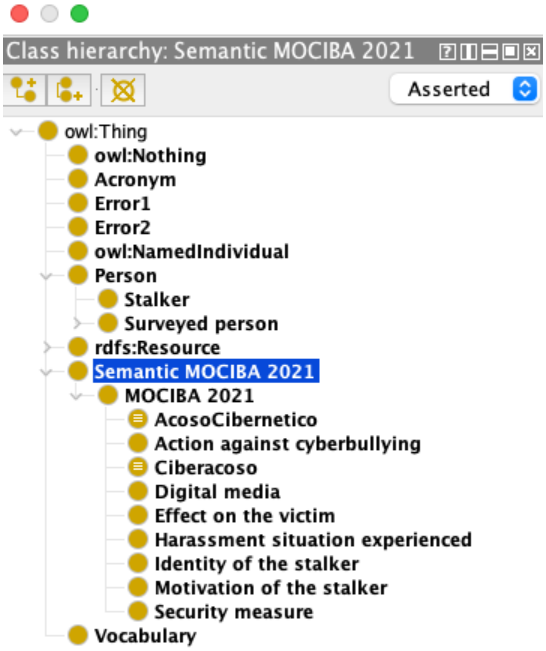

Fig. 5. Class hierarchy of Semantic MOCIBA 2021 vocabulary

3.2.3 Define the Properties of Classes and the Facets

Table 5 contains the properties of classes, also known as *object properties*, they represent the binary relationships between instances of the `Person` class and the instances of some subclass of the `Semantic MOCIBA 2021` class. The second and third column refer to the restrictions on the domain and range.

The object properties answer the question CQ3, they are also represented in the conceptual model. The facets for all of them are `irreflexive` and `asymmetric`. When the vocabulary is published in RDF, in particular using a TURTLE syntax file, this information can be easily retrieved by using a SPARQL query [23] as the following:

```
##### SPARQL QUERY 1

SELECT distinct ?objectProperty
WHERE { ?objectProperty rdf:type
            owl:ObjectProperty,
```

Table 5. Domain and range for object properties

Property	Domain	Range
acts_for	Stalker	Motivation
affects_to	Stalker	Victim
identifies	Victim	Identity of the stalker
perceives	Victim	Harassment situation experienced situation experienced
uses	Stalker or Victim	Digital media

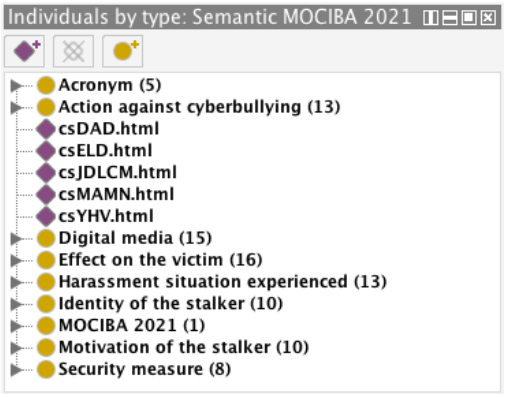

Fig. 6. Instances per class

```
            owl:AsymmetricProperty,
            owl:IrreflexiveProperty .
} ORDER by ?objectProperty
```

Besides object properties, the vocabulary has data properties that model the information of the questions `Q1`, `Q3`, `Q6`, `Q7`, `Q9`, `Q13` and `Q14`, as well as metadata from the MOCIBA 2021 dataset. Figure 7 shows these properties in bold type.

Fig. 7. Data properties

3.2.4 Create Instances

According to [9], the creation of instances is the last step of the vocabulary building. However, in this paper, they were introduced in the Figure 6.

The instances of Figure 8 and 9 constitute the answers to CQ1 and CQ2, respectively. Figure 8 shows a DL[6] query that retrieves the superclasses and the instances of the class `Harassment situation experienced`, the 13 instances represent the same number of situations of cyberbullying gathered by the MOCIBA 2021 questionnaire [13]. Likewise, Figure 9 shows the instances of the `Security measure` class. It is worth to notice that a reasoner needs to classify and validate the logical consistency of a vocabulary and ontology before the execution of DL queries.

The digital media that answer to CQ4 are enumerated in Figure 10 as part of a comment, although they are also modeled as instances. The properties `rdfs:label` and `rdfs:comment` are used as the main annotation properties for all the elements of the vocabulary.

The module Ontograph of the Protégé editor support the construction of graphs with instances as in Figure 11, this shows the effects that the

[6] DL is the acronym of Description Logics

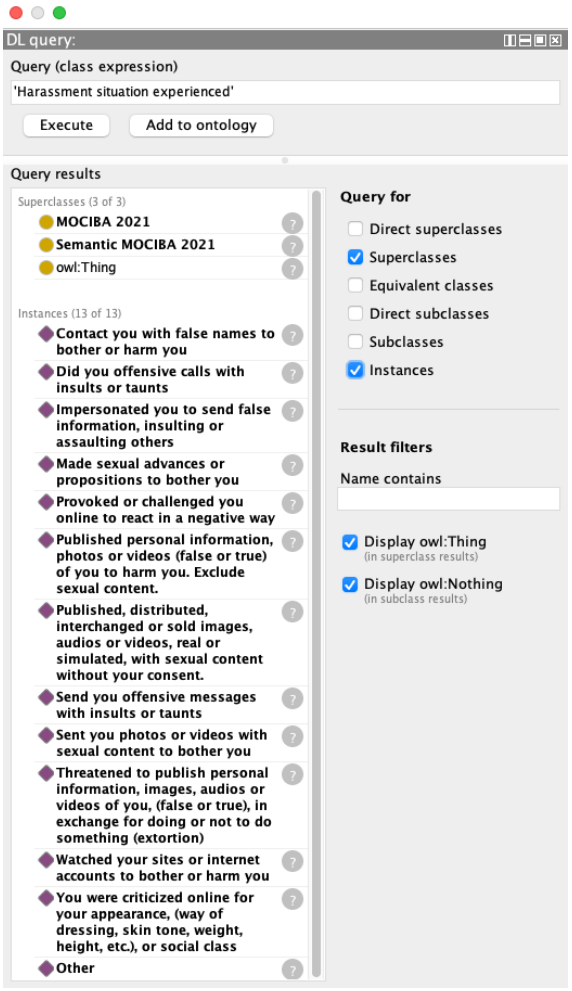

Fig. 8. Cyberbullying situations experienced

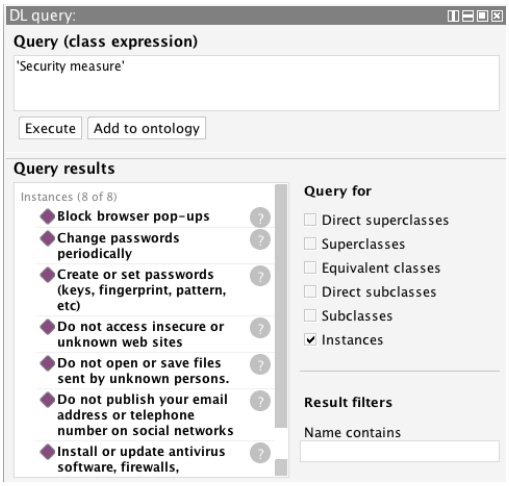

Fig. 9. Security measures

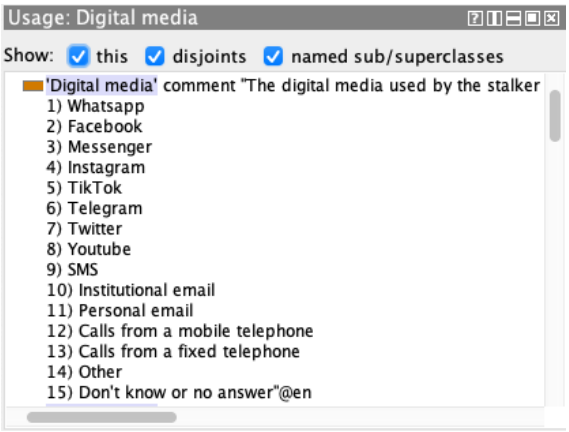

Fig. 10. Digital media used by stalkers

cyberbullying cause on the victims. The graph answers to CQ5.

An error was fixed after the record of the vocabulary in the Copyrights National Institute, (its Spanish name is *Instituto Nacional de Derechos de Autor* (IndAutor), see [18] and [19]), the instance effect 16 was added with the label “Don’t know or no answer” and the label of effect 15 was changed by “Nothing”. Figure 12 shows how to retrieve these effects by using a SPARQL query within the Protégé editor itself.

The Ontology Web Language (OWL) is used to formalize the Semantic MOCIBA 2021 vocabulary. An overview of this language is available at [17]. In summary, the following modeling artifacts were used:

— Classes that model main concepts

— Instances associated with the answers to multiple choice questions

— Object properties to establish relationships between instances based on the conceptual model

— Data properties that refer to dichotomous questions and metadata

ISSN 2007-9737

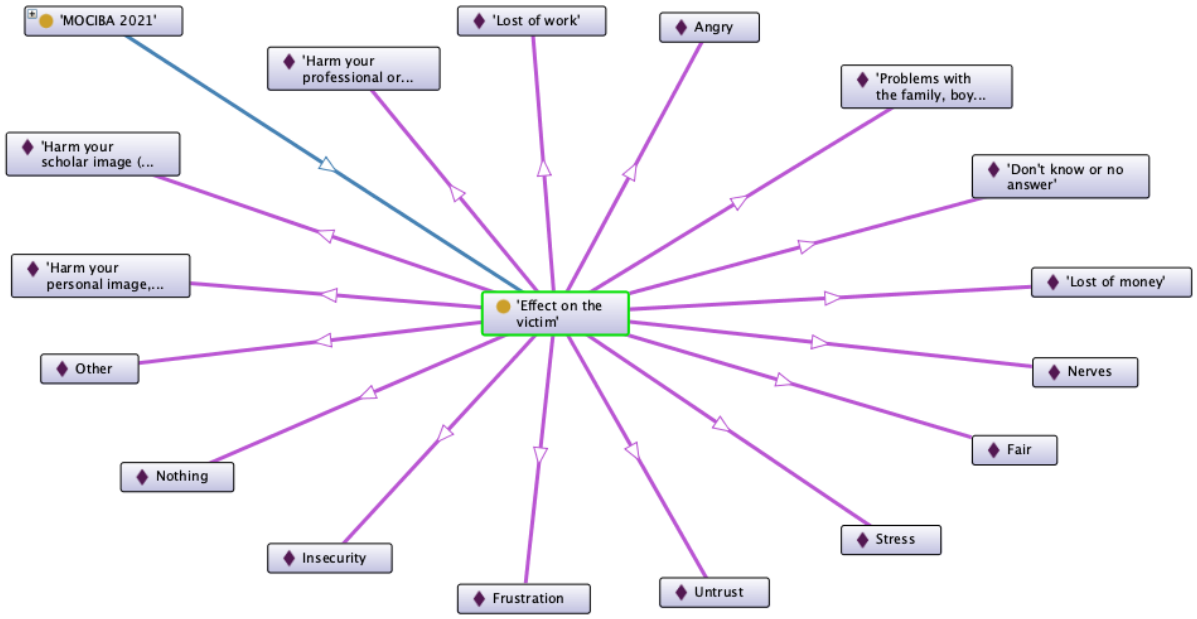

Fig. 11. Effects of cyberbullying on the victim

— Tagging of ontology elements by using the `rdfs:label` and `rdfs:comment` properties

— Capture of English and Spanish language specifics on the labels and comments

After following the recommendations for Linked Open Data vocabularies to improve reusability [25], the metrics of Semantic MOCIBA 2021 vocabulary are illustrated in Figure 13.

4 How to Use the Semantic MOCIBA 2021 Vocabulary

Semantic MOCIBA 2021 is an original ontology and vocabulary dedicated to the exploitation of MOCIBA 2021 dataset that provides context for data analysis. On one hand, a key use case for this vocabulary is to serve as a reference resource and practical tool for students and practitioners of Semantic Web technologies and for information systems communities.

On the other hand, the vocabulary can be used for exposing the complete MOCIBA 2021 dataset or some subsets as “Linked Data”, this enables the construction of enriched and interconnected data. For example, we propose to rename and arrange the columns of the MOCIBA 2021 dataset as follows:

1. Geographical data, (state, population, type of population, primary sampling unit, primary sampling unit_design),
2. Data of surveyed persons, (age and genre),
3. Answers per question

Figure 14 shows the geographical data used to identify the data for each questionnaire, the acronym UPM refers to the Spanish expression “Unidad Primaria de Muestreo”, *primary sampling unit* and UPM_DIS for “Unidad Primaria de Muestreo_diseño”, *primary sampling unit_design*.

We used the terms of the vocabulary to rename the columns of interest that will be used to construct new datasets, in other words, we designed a mapping of MOCIBA 2021 columns and transform some cell values based on linked data principles.

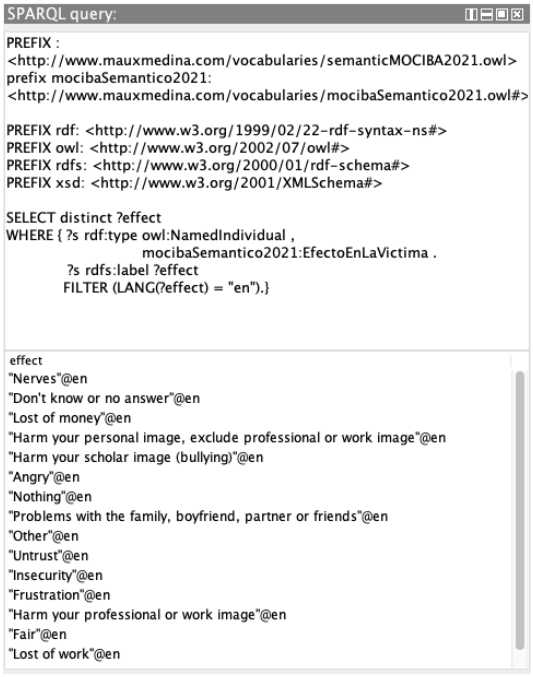

Fig. 12. Effects of cyberbullying as a result set

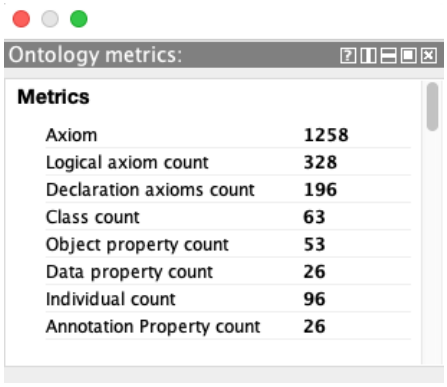

Fig. 13. Metrics of Semantic MOCIBA 2021 vocabulary

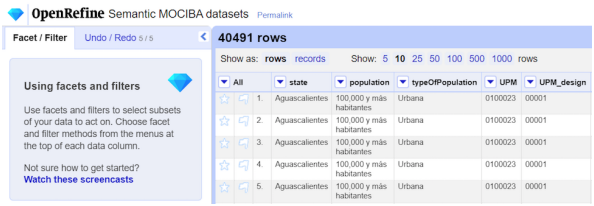

Fig. 14. Geographical data to construct new datasets

Finally, we exported the new datasets into RDF tuples using the OpenRefine extension called `RDF Transform` [22]. Thus, data are expressed as RDF serialized in TURTLE. We suggest to review the recommendations of the guide [4] in order to construct valuable, useful and interoperable datasets. As a proof of concept, consider the following SPARQL query:

```
##### SPARQL QUERY 2

SELECT ?o (count(?s) AS ?total)
WHERE
{
  ?s :carryOutSecurityMeasure ?o .
}
GROUP BY?o
```

Figure 15 shows the execution of the SPARQL query 2 on a new dataset called `Question1.ttl`, a file with RDF tuples in TURTLE sintax that was constructed by integrating geographical data and information about security measures implemented by surveyed persons, (see Q1 of [13]). The code of the SPARQL query 2 was stored in the file called `testQuestion1.rq`. The `arc` software tool [1] was used to execute this query on a computer with Windows Operating System. In spite that question 2 is relatively simple, this shows that it is possible to perform complex queries that exploit the new datasets.

5 Conclusion and Future Work

This paper presented the Semantic MOCIBA 2021 vocabulary, an original ontology and standardized vocabulary dedicated to the exploitation of MOCIBA 2021 dataset. The vocabulary emerged as an alternative to significantly improve reusability after understanding how data related to cyberbullying has been collected and disseminated according to the documentation distributed by INEGI; the difficulties of analyzing those heterogeneous data over time from a data management perspective were exposed.

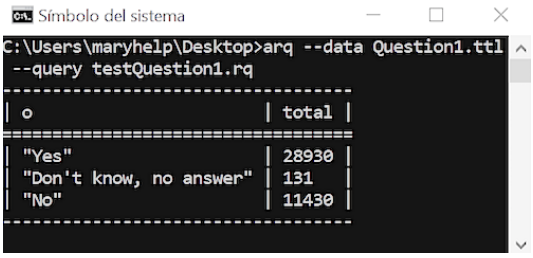

Fig. 15. Use of arc to query a new dataset

The development process of the Semantic MOCIBA 2021 vocabulary was described from scratch to create new and enriched datasets where concepts and relationships are formalized to represent and reason via linked data; the process is useful to explore other sources of valuable datasets commonly distributed by public administrations since open data portals.

According to the author's point of view, the contributions of this work are the following: the creation of a conceptual model to introduce the vocabulary to the general public, the construction of the Semantic MOCIBA 2021 vocabulary and its representation in the OWL language, the transformation of the vocabulary into RDF tuples in order to explore and retrieve information using SPARQL queries, and the exemplification to create new and enriched datasets using the proposed vocabulary.

The Semantic MOCIBA 2021 vocabulary serves as a reference resource and practical tool for students and practitioners for information systems communities, this was designed to support the decision-making process that can result in some actions against cyberbullying by individuals or organizations in the academic or social sector based on the evidence distributed as open data. We expect that its reuse will contribute to make visible the incidence of the cyberbullying phenomenon.

As future work, we plan to continue with the creation and exploration of new datasets to retrieve cyberbullying situations experienced according to specific features of the persons and the digital media involved.

Acknowledgments

We would like to thank to the participants involved in the construction of the Semantic MOCIBA 2021 vocabulary, their feedback was very useful to understand and integrate valuable points of view about the perception of cyberbullying.

References

1. **Apache Jena (2023).** ARQ - A SPARQL Processor for Jena. The Apache Software Foundation, `https://jena.apache.org/documentation/query/index.html`.
2. **Ávila, B. E. (2020).** Los datos enlazados y su uso en bibliotecas. Universidad Nacional Autónoma de México (UNAM). Instituto de Investigaciones Bibliotecológicas y de la Información., Ciudad Universitaria, 04510, Ciudad de México.
3. **Brickley, D., Miller, L. (2004).** FOAF Vocabulary Specification. Copyright ©2000-2004 Dan Brickley and Libby Miller, `http://xmlns.com/foaf/0.1/`.
4. **Chaves-Fraga, D., Corcho, O., Ruckhaus, E. (2022).** Guía práctica para la publicación de datos enlazados. Ontology Engineering Group, https://datos.gob.es/es/documentacion/guia-practica-para-la-publicacion-de-datos-enlazados-en-rdf.
5. **Data.europa.eu (2012).** Victimisation of crime; personal characteristics. `https://data.europa.eu/data/datasets/4117-caribbean-netherlands-victimization-of-crime-personal-characteristics?locale=es`.
6. **Data.europa.eu (2018).** Click here. Liubliana University. [Data set], `http://data.europa.eu/88u/dataset/adp-nadlos18`.
7. **Data.europa.eu (2022).** Harassment or cyberbullying due to their disability at school or study centre by autonomous community and gender. Population aged 6 and over with a disability enrolled in school

or undertaking studies or training courses. Instituto Nacional de Estadística. [Data set], http://data.europa.eu/88u/dataset/urn-ine-es-tabla-px-tpx-sociedad_2589-salud_2590-edad_8494-cap02_8496-mod7r_8543-cap010_8564-0710103.

8. **Delpeuch, A. (2023).** OpenRefine user manual. Open Refine, https://openrefine.org/docs.
9. **Fridman, N. N., D., M. L. (2001).** Ontology development 101: A guide to creating your first ontology. Technical report, KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html.
10. **Gruninger, M., Fox, M. S. (1995).** The Role of Competency Questions in Enterprise Engineering, chapter 3. Springer US. Boston, MA., https://doi.org/10.1007/978-0-387-34847-6_3, pp. 22–31. DOI: 10.1007/978-0-387-34847-6_3.
11. **INEGI (2021).** Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2021. Instituto Nacional de Estadística y Geografía, https://www.inegi.org.mx/programas/dutih/2021/.
12. **INEGI (2022).** Módulo sobre Ciberacoso (MOCIBA) 2021. Instituto Nacional de Estadística y Geografía, https://www.inegi.org.mx/programas/mociba/2021/.
13. **INEGI (2022).** Módulo sobre Ciberacoso (MOCIBA) 2021: Cuestionario. Instituto Nacional de Estadística y Geografía (INEGI), https://www.inegi.org.mx/contenidos/programas/mociba/2021/doc/mociba2021_cuestionario.pdf.
14. **INEGI (2022).** Módulo sobre ciberacoso MOCIBA 2021: principales resultados. Technical Report MEX-INEGI-MOCIBA-2021, Instituto Nacional de Estadística y Geografía (INEGI). Avenida Héroe de Nacozari Sur 2301, Franccionamiento Jardines del Parque, Aguascalientes, Aguascalientes, México, https://www.inegi.org.mx/programas/mociba/2021/.
15. **ITU (2022).** International Telecommunication Union (ITU): Statistics. http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.
16. **Lassila, O., Swick, R. R. (1998).** W3C Resource Description Framework (RDF) model and syntax specification. https://www.w3.org/TR/PR-rdf-syntax/Overview.html.
17. **McGuinness, L. D., Van Harmelen, F. (2004).** OWL Web ontology language overview. W3C Recommendation, https://www.w3.org/TR/owl-features/.
18. **Medina, N. M. A., De la Calleja, M. J., López, D. E., Hernández, V. Y., Arrieta, D. D. (2023).** MOCIBA Semántico 2021. Certificado del Registro Público del Derecho de Autor No. 03-2023-021409255800-01, https://www.mauxmedina.com/vocabularies/MOCIBA_Semantico2021_MAMN.owl.
19. **Medina, N. M. A., De la Calleja, M. J., López, D. E., Hernández, V. Y., Arrieta, D. D. (2023).** Semantic MOCIBA 2021. Certificado del Registro Público del Derecho de Autor No. 03-2023-022209550000-01, https://www.mauxmedina.com/vocabularies/SemanticMOCIBA2021_MAMN.owl.
20. **Musen, M. A. (2015).** The Protégé project: a look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, Vol. 1, No. 4, pp. 1007–1021. DOI: 10.1145/2557001.25757003.
21. **Ontology Engineering Group (2023).** Linked open vocabularies. Universidad Politécnica de Madrid (UPM), https://lov.linkeddata.es/dataset/lov/. Accessed: 2023-01-06.
22. **OpenRefine extensions (2023).** RDF Transform version 2.2.1. Open Refine, https://github.com/AtesComp/rdf-transform.
23. **Prudhommeaux, E., Seaborne, A. (2008).** SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/.

24. RDF Working Group (2004). Resource Description Framework (RDF). World Wide Web Consortium: Semantic web, https://www.w3.org/RDF/.

25. Vandenbussche, P. Y., Vatant, B. (2012). Metadata recommendations for linked open data vocabularies. Linked Open Vocabularies, https://lov.linkeddata.es/Recommendations_Vocabulary_Design.pdf.

26. Zermoglio, P. F., Plata, C. C. A., Wieczorek, R. J., Ortiz, G. R., Buitrago, L. (2022). Guía para la limpieza de datos sobre biodiversidad con OpenRefine. Global Biodiversity Information Facility (GBIF), https://docs.gbif.org/openrefine-guide/3.0/es/.

Article received on 20/03/2023; accepted on 09/12/2024.
Corresponding author is María Auxilio Medina Nieto.