Journal of Engineering Research and Reports

9(3): 1-12, 2019; Article no.JERR.53416
ISSN: 2582-2926

Evaluation the Software Tools Quality to Thesaurus in the Implementation of a Controlled Media Art Vocabulary

Miguel Angel Marzal[1*], Beatriz Ríos[2] and Diego Ferreyra[3]

[1]Faculty of Communication and Library Science, Carlos III University, C/ Madrid, 126 (Decanato Building), C.P. 28903, Getafe, Madrid, Spain.
[2]Department of Computer Systems, Tecnológico Nacional de México en San Luis Potosí, Av. Tecnológico, U.P.A. C.P. 78437, Soledad de G. Sánchez, S.L.P, México.
[3]Quirós 2781 (C1427DAC), Ciudad de Buenos Aires, Argentina.

Authors' contributions

This work was carried out in collaboration among all authors. Authors MAM and BR designed the study and manage the analysis and wrote the first draft of the manuscript. Author DF developed and advice on the use of the TemaTres tool and managed the analyses of the study. All authors perform the sample collection and analysis, managed the bibliographic search in collaboration with IT team. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JERR/2019/v9i317019
Editor(s):
(1) Dr. S. Selva Nidhyananthan, Associate Professor, Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi - 626005, India.
(2) Dr. Anuj Kumar Goel, Professor, CMR Engineering College, Kandlakoya (V), India.
(3) Dr. Leandro Antonio Pasa, Professor, Campus Medianeira da Universidade Tecnologica Federal do Parana, Brazil.
Reviewers:
(1) Cristina Álvarez de Morales Mercado, University of Granada, Spain.
(2) Rohit L. Vekariya, Sardar Patel University, Gujarat, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/53416

Short Research Article

Received 26 October 2019
Accepted 30 December 2019
Published 07 January 2020

ABSTRACT

In this project, he develops the evaluation of seven thesaurus software tools, in the implementation of a controlled vocabulary, for the CAAC (Collections and Archives of Contemporary Art of the Faculty of Fine Arts of Cuenca, UCLM) in collaboration with UC3M; framed in the "Media Art vocabulary" project. For the conservation of artistic products as digital objects of the museum of Cuenca, Spain. The Construction and application of the thesaurus, Web implementation, publication of its terms as Linked Data, using TemaTres Software. A customized method is used combining quality, usability and semantic web techniques. A method was designed to evaluate the

**Corresponding author: Email: mmarzal@bib.uc3m.es;*

characteristics of each of the tools. Such as the availability of download, easy of access, use, learning, deployment on the server, sharing, and exchange of data, and free use. This study is a part project to Vocabularies for a Media Art Archives and Collections Network. This project establishes the scientific and methodological convergence of two research teams, one from Humanities, Art and Fine Arts, the other from Documentation, with a support of Computer Science and Audiovisual Communication. The importance of project progress lies in metadata labeling, identification by descriptors of the specific thesaurus, guarantee of semantic reuse of content, interoperability for a network of Media Art archives and collections, its digital continuity and dissemination in virtual spaces Wide projection. Winning the TemaTres tool, the thesaurus was implemented in the period from April to July 2019.

Keywords: Media art; thesaurus; vocabulary; semantic web.

1. INTRODUCTION

This article presents the selection, evaluation and analysis of seven software tools options for vocabulary implementation.

Internet is undoubtedly one of the greatest inventions of human success. The evolution of the static internet to the dynamic internet, the combination of technologies that support what today constitutes the public Internet has proven its effectiveness during two decades of explosive growth [1].

To understand the semantic Web 3.0, we will start explaining its evolution.

Web 1.0 of personal computer, was the first, static Web, characterized by showing only the information, did not allow the exchange between the users, the information was only the deployment of mostly texts, neither did the devices exist as at present. The information searches were only syntactic[1]. In this sense, it does mean that the search is only about syntax, the response will be about the word.

The Web 2.0 called collective intelligence, the dynamic, interactive, collaborative, participatory Web, the Web is used as a platform, taking advantage of collective intelligence, the data is as follows: “Intel inside”, end of the version cycles of software, lightweight programming models, software not limited to a single device.

It was characterized by allowing the exchange of information, through social networks, thus allowing you to upload files such as photos, text, press a “like”, in the social communities of academic exchange, upload files with content on educational topics, in distribution lists, send mass advertising by email, etc.

[1]*Of or pertainig to syntax, [10].*

Its principles date back to the first interactive applications, 2006, thanks to the Ajax language, such as social networks, Wikis, Mashups, folksonomies, etc. The different mail providers, also updated allowing access to their own networks, using contact lists, became more efficient. The Gmail mail engine showed the various options of the semantic Web, according to the profile of the connected person, ads or purchase recommendations could appear [1].

Between 2006 and 2008, the suggestions began to appear based on the profile, for example, if the word “models” was being searched, Web 3.0 [2], was not only syntactic, but also semantic. For a person with a dentist profile, he cast dental models among the search results; for a dressmaker, clothing models; for a software engineer, software models, for an architect, construction models, for an electronic engineer, circuit models, for a gateway model, people modeling, for chemists, Darwing's atomic models, and so on. The searches surprised users a lot, it was the era of Technology available to all people, including laws such as the Internet for everyone, electronic devices such as PDAs, e-Books, tablets and smartphones; And with these, the ubiquity phase, the Web 4.0. With these mobile devices, access was available from anywhere, and the WiFi connection was facilitated almost from any public place, which the government provides freely and in cafeterias [1].

The semantic Web works through of: Ontology in computer science and the semantic web is a formal representation of knowledge.

For those involved in metadata development, when you have defined all of your data elements, your controlled vocabularies, how they fit together and anything else that is needed to make your metadata work, and then you have an ontology for your metadata domain [3].

"An ontology is an explicit specification of a conceptualization" [2].

Ontologies provide a way to represent knowledge (especially on the web) and are an important approach to capture semanthics.

Thesauri: "Thesaurus is a dynamic documentary language that contains semantically related terms, according to a domain of knowledge. It is a classification system for organizing knowledge" [4].

According to UNESCO's definition, "a thesaurus is an instrument of terminological control used to translate the natural language used in documents and by indexers into a more strict language" [5].

Metadata is machine understandable information about web resources or other things.

Semantic relations: They are all those relations that are established between the terms of a thesaurus.

The semantic Web uses descriptions that are classified into semantics, metadata and ontologies; as well as automatic language, logic and semantic inference engines. The Web allows, identify, describe, classify and locate a resource [6].

From this collaboration, you can relate the different contexts, situations, states, schemes and participating entities, providing more accurate results and with less sensitivity to the vocabulary used in the search [7].

With the advent of cloud services, the different websites became more efficient, and with artificial intelligence, the effectiveness between search results is even better. You can say they are almost intuitive.

Since the semantic Web [8] studies the meaning of words, sentences, and changes over time. For Tim Berners-Lee, the Semantic Web, it must be fed with data to make available to users, large volumes of data in repositories, which are accessible, through data languages: Resource Description Framework (RDF), language SPARQL, OWL and SKOS query; all of them, form the Web of the linked data (linked data). The Web Ontology Language (OWL), [9] is used in applications that need to process the information content; it facilitates a better mechanism for interpreting Web content, than the mechanisms supported by XML and RDF. For the creation of controlled vocabulary, taxonomy and thesaurus, SKOS (Simple Knowledge Organization System) is used, [6] which allows the creation and publication of concepts on the Web, and links them with other concept schemes or with other data or their reuse by from third parties. With the SKOS-XL extension "more data can be obtained than on a label" [10].

We participated in the creation of controlled vocabulary for Media-Art, of the CAAC (Collections and Archives of Contemporary Art of the Faculty of Fine Arts of Cuenca, UCLM); With the collaboration of the work team of the Agustin Millares Institute of the Carlos III University, in the Department of Library and Communication on the Getafe campus in Madrid, Spain; In a Research stay, authorized by the Carolina Foundation and the Ministry of Foreign Affairs of Mexico [11].

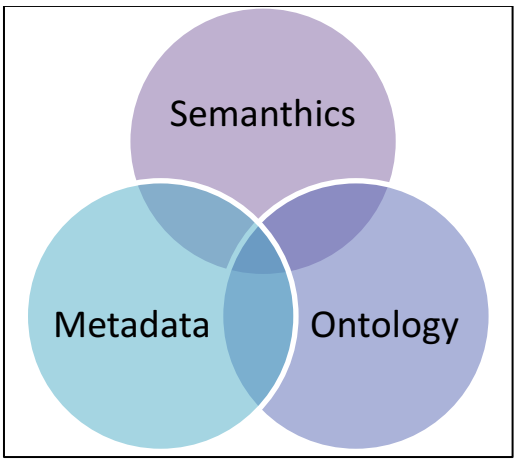

Illustration 1. Web semantics descriptions, based on [12]

2. MATERIALS AND METHODS

The methodology used is based on the scientific method, it consists of the following phases [13]:

1. Definition and detection of opotunities:

The first phase consisted of the documentary part:

- With the work team of the UC3M of the faculty of communication, the guidelines and collaboration of the communication team of the University of Cuenca, it was identified and documented:

  a) The lists of terms (labels),
  b) Taxonomies (classification),
  c) Definition and documentation of the thesauri
  d) Definition of ontologies.

- At UC3M, with the IT team, a sample selection, collection and bibliographic search and analysis were carried out to determine the indicators to be evaluated and the implementation.

2. Analysis and determination of the factors to review.

3. Bibliographic research and selection of tools. [14].

4. Download, use and tests.

5. Obtaining results.

2.1 Methods

A relational coding was designed, combining quality, usability, and semantic Web techniques to represent the results. In illustration 2 you can see the indicators that are evaluated in usability. [15].

The seven software tools for the administration of the thesauri and the construction of the Media-Art vocabulary were chosen based on bibliographic study, advanced knowledge and mastery of literature, and empirical knowledge.

The bibliographic search was carried out in its description, a table was created to relate them to the indicators that show the degree of effectiveness that each one has. Considering the metrics of quality [16], usability [17] and interaction with the semantic Web, the evaluation method was designed.

The seven tools selected are: ThManager [18], TopBraid [19], PoolParty [20], TemaTres [21], VoBench [22], Skosed [23], e iQvoc [24].

Within the context of Software Engineering, the definition of quality in the software proposed by the international organization of standards (ISO / IEC DEC 9126): The totality of characteristics of a software product that have the ability to meet explicit or implicit needs [16].

Based on the concepts of software quality assurance [16], the following metrics were considered:

1. Functionality
2. Efficiency
3. Reliability
4. Maintenance
5. Accessibility / usability
6. Portability
7. Security

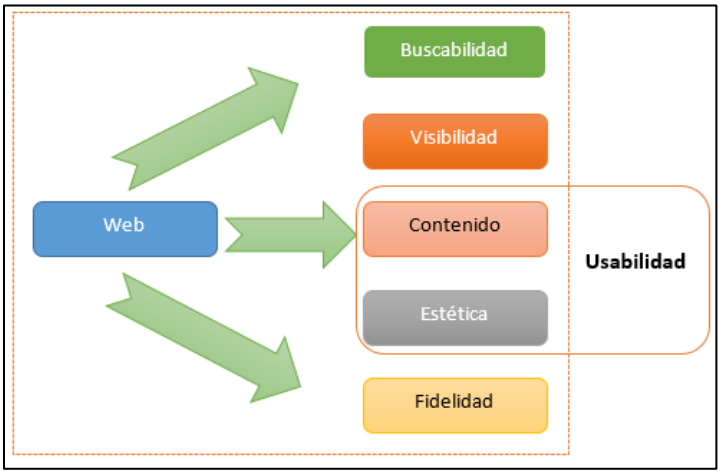

Illustration 2. Indicators to assess usability based on [15]

Table 1. Software tool description and characteristics

Software tool	Description	Characteristic
ThManager	ThManager is an open source tool to create and visualize SKOS RDF vocabularies, a W3C initiative for the representation of knowledge organization systems such as thesauri, classification schemes, subject headings lists, taxonomies and other types of controlled vocabulary. Open source (under the GNU Lesser General Public License (LGPL)) and developed by the University of Zaragoza.	• Multiplatform (Windows, Unix) • Multilingual • Selection and filtering of the thesauri stored in the local repository • Description of the thesaurus through metadata according to an application profile based on Dublin Core for the thesaurus • Thesaurus concept visualization • Alphabetic viewer • Hierarchical viewer • Concept Viewer • Thesaurus content edition • Thesaurus exchange according to the SKOS format
VocBench	This tool is developed mainly by the University of Rome Tor Vergata, and is closely related to the thesaurus AGROVOC, since its management was the reason for the creation of VocBench. The AGROVOC thesaurus was first published (in English, Spanish and French) at the beginning of the 1980s by the Food and Agriculture Organization (FAO)	• It is an open source web platform (under Mozilla Public License MPL) • Allows the editing of SKOS multilingual thesauri (XL) collaboratively. • Native support for SKOS (XL), role management and workflows or multilingualism
TemaTres	It is a web tool, open source and distributed under the GNU Public License (GPL), for the management and exploitation of controlled vocabularies, thesauri, taxonomies and other models of formal representation of knowledge.	• It is an easy to use tool and a possible starting point. • TemaTres, which is a controlled vocabulary server. • This free application for web use facilitates continuous access to documentary sets of specific topics for study, research and decision-making.
TopBraid Enterprise Data Governance	The first data governance solution that uses real knowledge graphics technology based on W3C standard RDF graphics.	• The EDG, can accelerate a data governance program, different organizations • Use EDG for business glossaries or reference data or metadata management only • Provides focused packages, available as an initial EDG configuration
PoolParty Semantic Suite	It is an artificial intelligence platform based on semantic technologies and machine learning. It helps organizations create and manage Enterprise Knowledge Graphs as the basis for various AI applications	• The most complete semantic middleware • Latest generation machine learning. • Based on W3C standards since 2009 • ISO / IEC 27001 Certificate

Software tool	Description	Characteristic
SKOSEd - Editor thesauri to the Semantic Web	Is a complement to Protege that allows you to create and edit thesauri (or similar artifacts) represented in the Simple Knowledge Organization System (SKOS).	• Allows you to configure the order and language of label representation.
iQvoc	It was originally created and is maintained by innoQ Deutschland GmbH (2011)	• It admits vocabularies that are common to many knowledge organization systems, such as: • Thesaurus, Taxonomies, Classification schemes, Subject heading systems. • Supports several customizable masks to align with any type of user-specific style guide. • The iQvoc web interface provides unrestricted multilingualism and meets the guidelines of the Web Accessibility Initiative (WAI). • Editorial workflow • It provides unique and stable identifiers for each concept according to the W3C guidelines "Great URI for the semantic web", and "How to publish linked data on the web".

Since usability is evaluated within quality, usability will be analyzed separately, so quality point 5 is changed to accessibility[2].

For the evaluation of Usability, we consider the 10 heuristic principles of Nielsen [17] and we have adapted the method, eliminating those that were not necessary, for our study, and leaving the following:

1. Visibility of the system status
2. Adequacy between the system and the real world
3. Freedom and user side control
4. Consistency and standards
5. Aesthetic and minimalist design
6. Flexibility and efficiency in use
7. Help the user to recognize, diagnose and recover from errors

The elements for the semantic Web interaction were also considered:

1. Availability} Download and Access- (Bad, good, excellent)
2. Server
3. Open Source
4. SKOS-XL
5. Support provider
6. Documentation
7. Import
8. Export
9. Workflow
10. Information Security Management Certification ISO / IEC 27001
11. Language settings (Accessibility)

In these variables, the installation, learning curve, support, maintenance and errors of the data repository, as well as others that will be considered for future research work, remained pending.

The sum of the characteristics of quality, usability and interaction and semantic Web was calculated, obtaining the average of each category.

Individual and general data were plotted for the 3 categories evaluated.

[2] *Web accessibility means that people with some type of disability will be able to use the Web. Specifically, when talking about Web accessibility, reference is being made to a Web design that will allow these people to perceive, understand, navigate and interact with the Web, contributing content. Web accessibility also benefits other people, including elderly people who have seen their ability diminished as a result of age [11].*

3. RESULTS AND DISCUSSION

A coding was used for the final grade, between 1 and 3, with 3 being the highest grade, and 1 the lowest, for the characteristics: Bad, good, excellent.

The tests that were carried out on the documentation were awarded a rating of 2 and according to the interface shown, usability, interaction and semantic Web were evaluated.

The graphs for each of the seven tools are shown below.

In Fig. 1, iQvoc, shows the metrics considered with an average of 1,857 in quality, 2 in usability, 2,363 in semantic Web interaction and 2,120 in the general average.

In Fig. 2, the metrics considered for VoBench with an average of 2 in quality, 2,571 in usability, 2,727 in semantic Web interaction and 2,432 in the general average are observed, obtained higher values than iQvoc.

Fig. 3, shows the results of SkosEditor with an average of 2,500 in quality, 2,750 in usability, 2 in semantic Web interaction and 2,416 in the general average. Greater than iQvoc, but slightly below VoBench.

Fig. 4, shows the results for ThManager with an average of 1,714 in quality, 2 in usability, 2,818 in semantic Web interaction and 2,177 in the general average, above iQvoc, but below SkosEditor and VoBench.

The results for TopBraid Governance are shown in Fig. 5, with an average of 1,428 in quality, 1,428 in usability, 1 in semantic Web interaction and 1,285 in the general average. Resulting so far, rated as the lowest.

The results for TemaTres are shown in Fig. 6, with an average of 2.5 in quality, 2.75 in usability, 2 in semantic Web interaction and 2.41 in the general average.

Fig. 7 shows the results for PoolParty, with an average of 2,428 in quality, 2,571 in usability, 2,818 in semantic Web interaction and 2,606 in the general average. Being so far the highest valued.

Once the tests were carried out, the results obtained are shown in Table 1, with a score between 1 and 3. The best rated thesaurus management software tool was: 1. PoolParty, followed by tool 2. TemaTres. followed by 3.

VoBench, 4. Skosed, 5. ThManager, 6. iQvoc and 7. TopBraid. As shown in the Table 2, they were rated with a significance level of 0.126 the smallest, and 0.187 the largest.

The level of significance between the values obtained between the 3 categories can be seen in Fig. 8.

The interface on each of the tools was quite different. For example, VoBench, Skosed and Thmanager tools have the appearance of Windows 2007 windows, Web 1.0. However, the rest, have a better interface and appearance as in the current Web. The worst-rated tool was TopBraid, because it is a tool for the government, so it follows that it is not accessible, nor is it open access, so it has many restrictions even for evaluation, the requirements could not be verified, because The documentation could not be accessed.

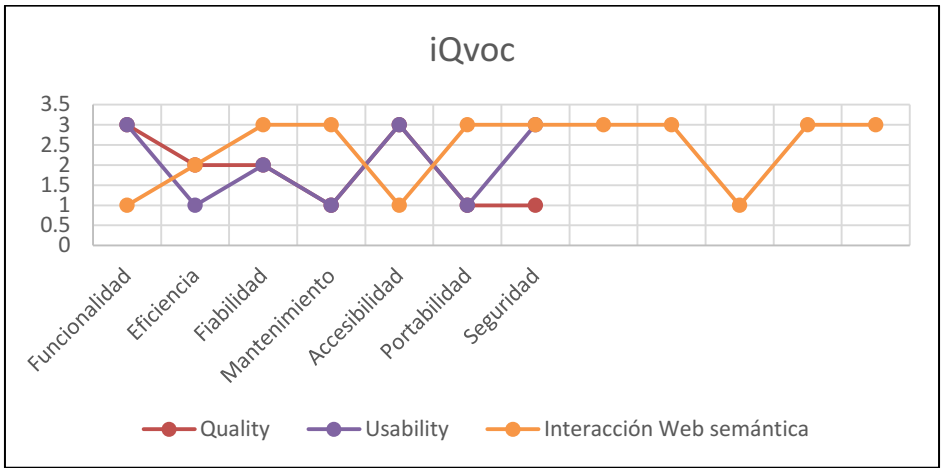

Fig. 1. Graph of the metrics for iQvoc

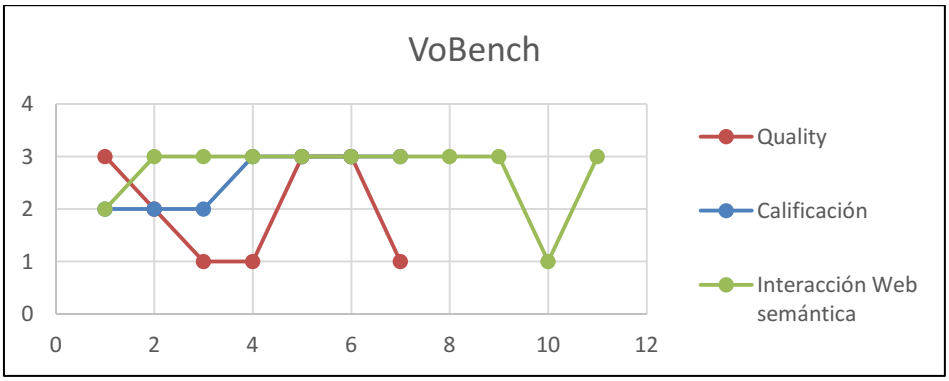

Fig. 2. Metrics graph for VoBench

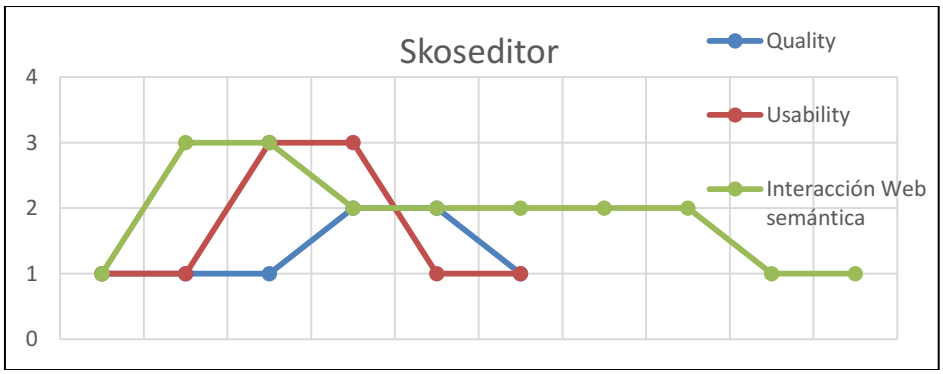

Fig. 3. Skoseditor results

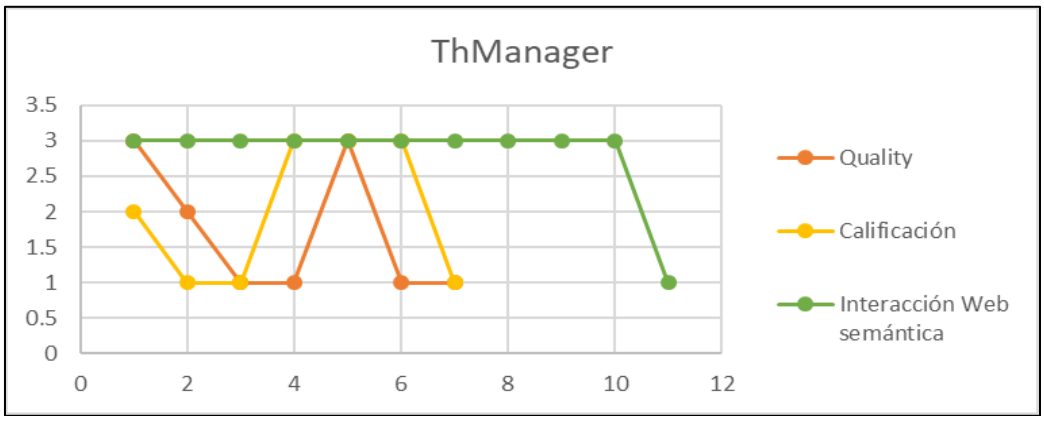

Fig. 4. ThManager results

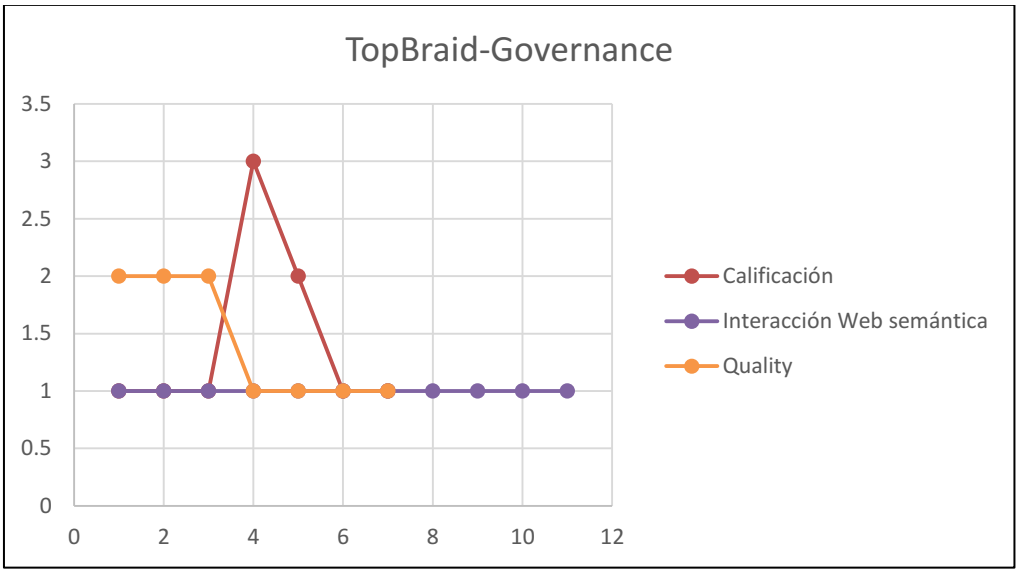

Fig. 5. Results of top braid-Governance

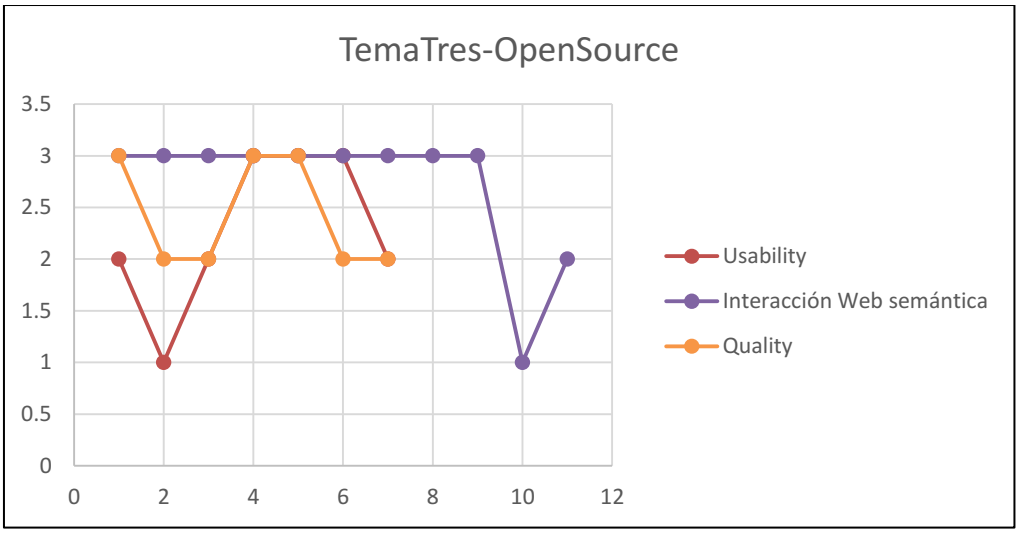

Fig. 6. Results of TemaTres

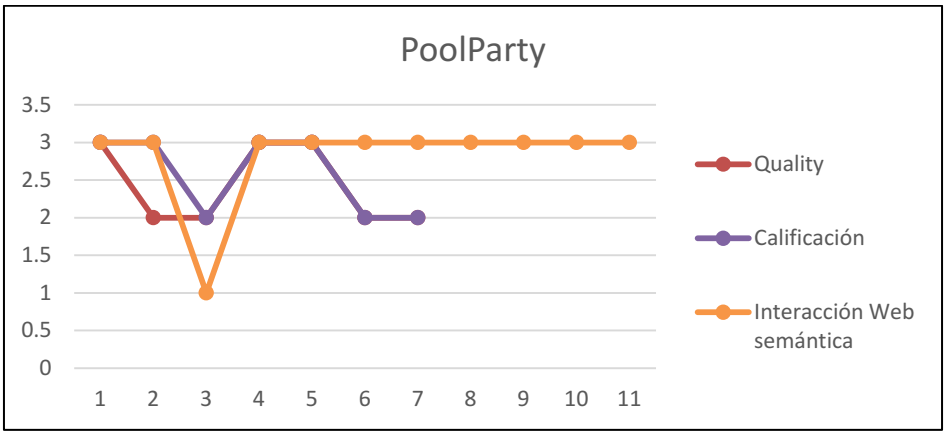

Fig. 7. Pool party results

Table 2. Global results of the seven tools

Name	Quality	Usability	Web semantic interaction	Media	Level of significance
PooParty	2.430	2.571	2.810	2.604	0.000
TemaTres	2.428	2.285	2.720	2.478	0.126
VoBench	2.000	2.571	2.727	2.433	0.171
Skosed	2.500	2.750	2.000	2.417	0.187
ThManager	1.710	2.000	2.818	2.176	0.428
iQvoc	1.857	2.000	2.364	2.074	0.530
TopBraid	1.430	1.430	1.000	1.287	1.317
Average	2.214	2.435	2.615	2.421	0.182

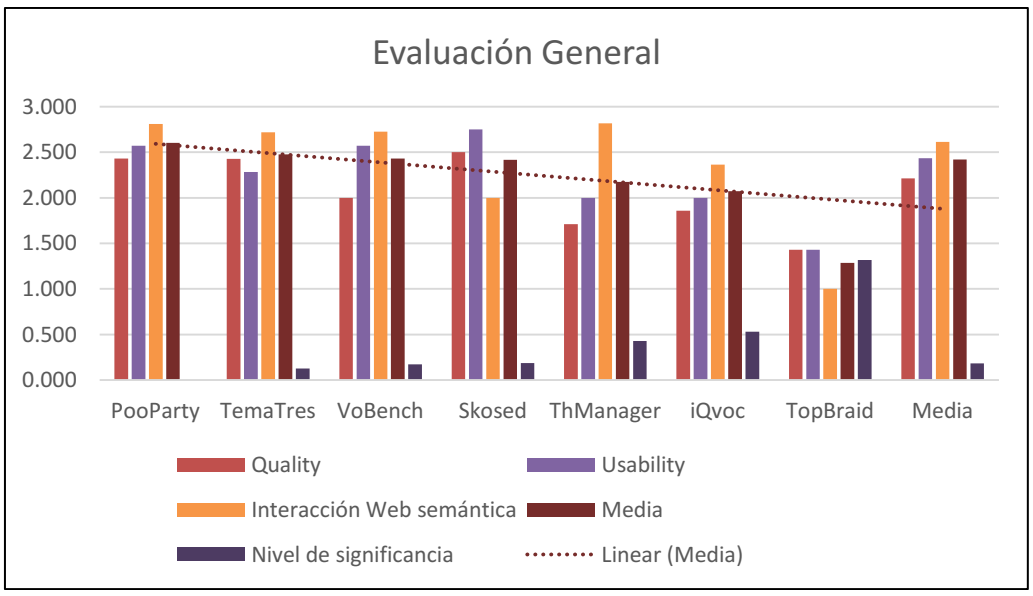

Fig. 8. Global tool data

4. CONCLUSION

Perhaps, the tests applied to these tools can be improved in the sense of semantic Web interaction, because it is difficult to access, download, install, receive support, documentation and access to the server as an administrator has to be requested. To the extent that they can be tested, they will possibly be more efficient.

As it was already shown with the results obtained, although PoolParty is the most valued, it has a paid license, so it was decided to use free access software, so the TemaTres tool was chosen for the creation of the Media Art vocabulary.

Since it is a tool with many benefits, in addition to being open access, there was the advice of the development team and the National Library of Teachers, Ministry of Education, Culture, Science and Technology, Argentina.

There is still work to do and learn about the TemaTres tool, there are some inconsistencies that generate errors, but you must learn to solve them without the help of the administrator providing access to the platform.

Review the laws on access to the thesaurus definitions repository so that they can be linked to universal networks.

An important issue is the publication and implementation of Linked Data for reuse.

ACKNOWLEDGEMENT

The authors thank to: The Minister of Economy and Competitiveness (Spain) for the funding of the project "Vocabularies for a Network of Archives and Media Art Collections and its effects: metaliteracy and knowledge tourism" whose reference is HAR2016-75949-C2-1-R.

The Fundación Carolina (Spain), in the program: “Estancias cortas postdoctorales” in colaboration The Exerior Relations Secretary (México).

PhD.Mercedes Caridad, (UC3M, Communication and Library science Faculty).

Carmen Jorge, Library and documentation (UC3M, Communication and Library science Faculty).

Lic. María Eugenia Dimattía, (National Library of Teachers, Ministry of Education, Culture, Science and Technology. Argentina).

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Ríos B. Models, technologies and tools of the company 2.0 to optimize business management processes. The Spain-Mexico case. Madrid, Madrid: Pontifical University of Salamanca in Madrid; 2012.
2. W3ORG, w3.org.; 2001. Available:https://www.w3.org/2001/sw/wiki/Semantic_Web_terminology. [Accessed: 27 09 2019]
3. Gruber. A translation approach to portable ontologies. Knowledge Acquisition. 1993;5 (2):199-220. Available:http://tomgruber.org/writing/ontolingua-kaj-1993.pdf. [Accessed: 27 09 2019].
4. Carlos Blandón CMZJ. A state-of-the-art review; 2018. Available:http://www.scielo.org.co/pdf/inde/v36n1/2145-9371-inde-36-01-00259.pdf. [Accessed: 30 09 2018]
5. Tesauro de la UNESCO, UNESCO Thesaurus; 2019. Available:http://vocabularies.unesco.org/browser/thesaurus/es/. [Accessed: 28 07 2019].
6. W3C-España W3C-España; 2019. Available: https://www.w3c.es/estandares/.
7. Villaplana R. Cloud TIC Consultoría y desarrollo; 2014. Available:http://www.cloud-tic.com/pasos-interaccion-entre-una-marca-y-los-clientes-en-social-media/.
8. Wordreference, Wordreference, Wordreference; 2019. Available:https://www.wordreference.com/definition/syntactic [Accessed: 05 11 2019]
9. 3WC, «3WC,» 3WC; 2005. Available:https://www.w3c.es/Traducciones/es/WAI/intro/accessibility [Accessed: 2 11 2019]
10. IBM, Dirección, Videocast - Preparing the Work Breakdown Structure (WBS). [Película]. IBM; 2010.
11. Voremetur, Voremetur; 2019. Available: https://voremetur.uc3m.es/
12. Zapater JJS. PTC Technologicall notebooks; 2017. Available:https://www.ptcarretera.es/wp-content/uploads/2018/07/02_2017_lisitt_def.pdf
13. Rios B. Web Tools, Madrid: EUMED; 2012.
14. Datos.gob.es, Datos.gob.es; 2017.

Available:https://datos.gob.es/eu/noticia/herramientas-para-la-gestion-de-vocabularios-controlados-con-soporte-skos [Accessed: 03 02 2019]
15. Almazán F. Library National congress of Chile; 2005. Available:http://www.ciw.cl/material/DMU2005/bcn-almazan.pdf [Accessed: 07 03 2019]
16. Vega C, Rivera L, García A. Best practices for the establishment and assurance of software quality, EUMED; 2008.
17. Ríos B. Usability and user experience in virtual communities and social networks. Software Guru, México; 2010.
18. SourceForget.Net. SourceForget.Net Zaragoza University. SourceForget.Net; 2019. Available:http://thmanager.sourceforge.net/ [Accessed: 01 november 2019]
19. TopQuadrant, TopQuadrant. iQVoc; 2019. Available:https://www.topquadrant.com/products/topbraid-enterprise-data-governance/. [Accessed: 01 november 2019].
20. PoolParty Semantic Suite. PoolParty Semantic Suite. PoolParty Semantic Suite; 2019. Available:https://www.poolparty.biz/ [Accessed: 05 08 2019]
21. TemaTres. TemaTres, TemaTres; 2019. Available:https://r020.com.ar/tematres/demo/ [Accessed: 03 02 2019].
22. Vergatra ART. ART Roma to Vergatra, ART Roma to Vergatra; 2019. Available:http://art.uniroma2.it/ [Accessed: 01 november 2019]
23. SkosEd, «Skosed Google,» Skosed Google; 2019. Available:https://code.google.com/archive/p/skoseditor/ [Accessed: 01 november 2019]
24. iQVoc, «iQVoc,» iQVoc; 2019. Available: http://iqvoc.net/ [Accessed: 01 november 2019]

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/53416