

Internship results: BiDO ontology extension for measuring scientific production

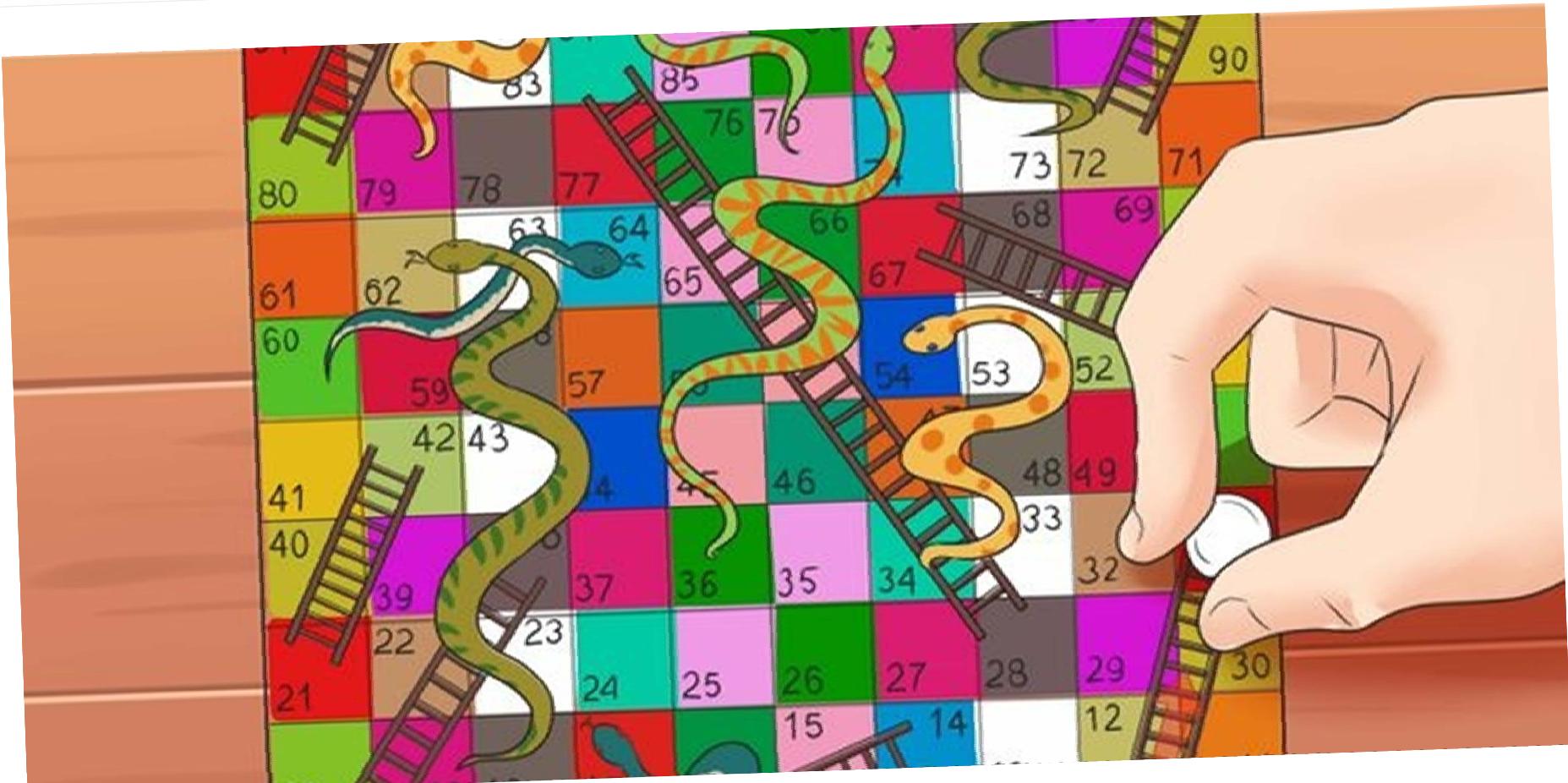
Mariela Tapia León

Agenda

- What was I going to do here?
- What did I really do?
- But first, how did I feel?
- Suggestions
- Your suggestions

Agenda

- What was I going to do here?
- What did I really do?
- **But first, how did I feel?**
- Suggestions
- Your suggestions



Agenda

- What was I going to do here?
- What did I really do?
- But first, how did I feel?
- Suggestions
- Your suggestions

ACTIVIDADES	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Sem 7	Sem 8	Sem 9	Sem 10	Sem 11	Sem 12	Sem 13
1. Inserción al grupo de investigación													
2. Introducción al desarrollo de ontologías													
3. Análisis de las ontologías para publicaciones científicas													
4. Creación de la ontología para publicaciones científicas													
5. Análisis de las ontologías para los proyectos de investigación													
6. Creación de la ontología para proyectos de investigación													
7. Redacción de los artículos generados													

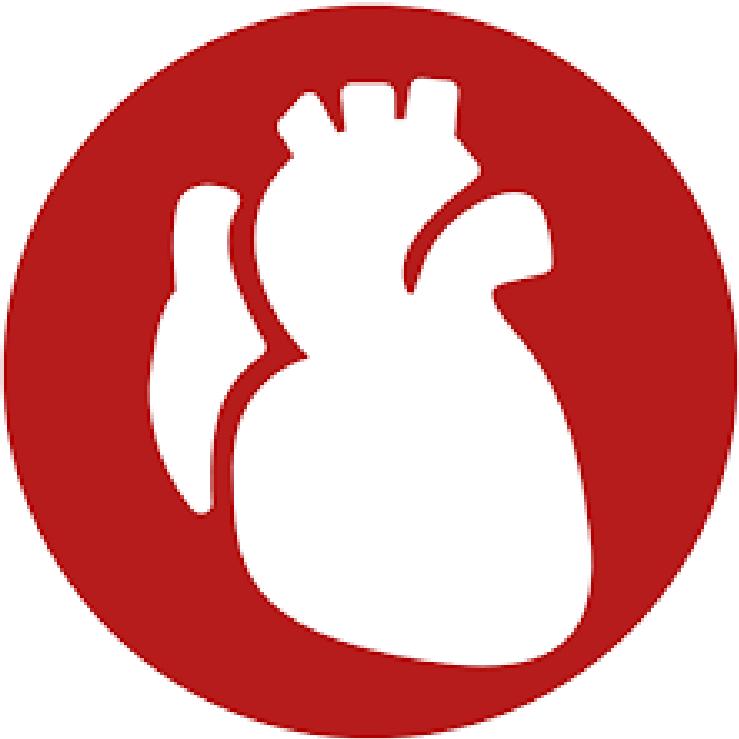
Agenda

- What was I going to do here?
- **What did I really do?**
- But first, how did I feel?
- Suggestions
- Your suggestions

¿Se puede planificar la investigación?

- Cronograma de actividades SEMANA a SEMANA
- Resultados esperados
- ¿Se puede indicar con certeza qué resultados de investigación entregar antes de ponerse investigar?

Serendipia



12/14/2019

10

Un equipo de investigadores de Pfizer estaba estudiando los resultados de un nuevo fármaco para el tratamiento de la enfermedad coronaria, sin éxito



"Los pacientes no quieren renunciar al tratamiento. Insisten en seguir tomando la medicación"



12/14/2019

12

Así comenzó el descubrimiento de la cápsula 'milagrosa'.



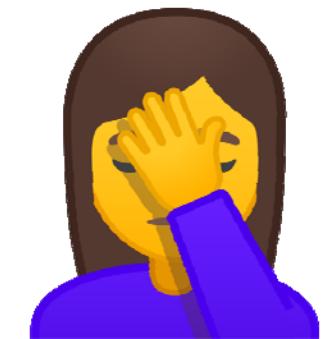
-
-
-
-
-
-
-

ACTIVIDADES	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Sem 7	Sem 8	Sem 9	Sem 10	Sem 11	Sem 12	Sem 13
1. Inserción al grupo de investigación													
2. Introducción al desarrollo de ontologías													
3. Análisis de las ontologías para publicaciones científicas													
4. Creación de la ontología para publicaciones científicas													
5. Análisis de las ontologías para los proyectos de investigación													
6. Creación de la ontología para proyectos de investigación													
7. Redacción de los artículos generados													

producción

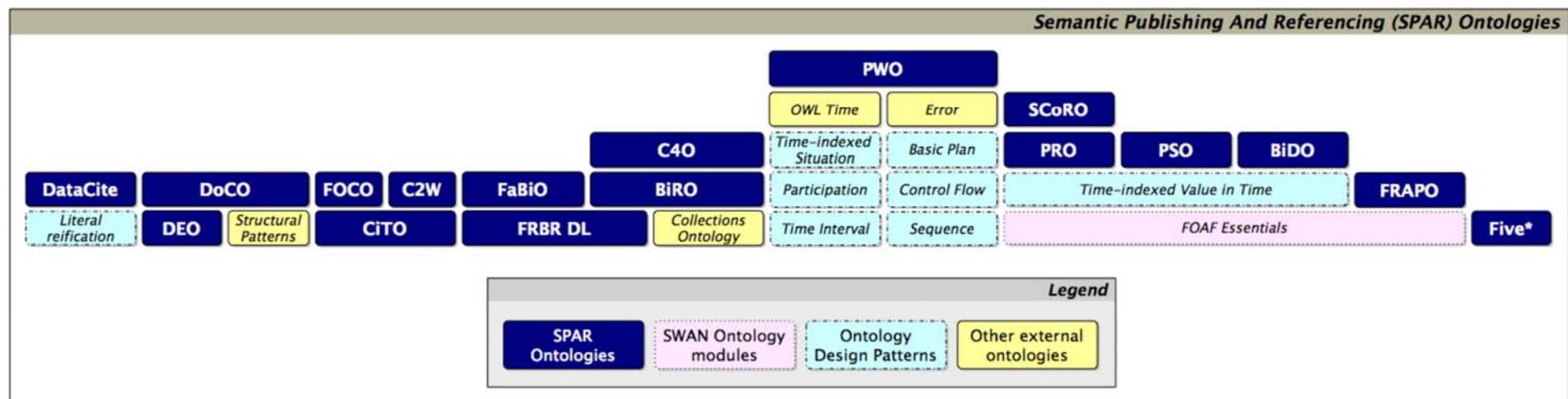
extensión

Redacción de **dos artículos científicos**. Uno se enviará a una publicación indexada en Scopus o WoS y el otro a una indexada a una base de datos regional.



SPAR Ontologies

Semantic Publishing and Referencing Ontologies



Glossary

- Semantic Publishing
- Scholarly Communication
- Scholarly Publishing
- Scientific Production

Semantic Publishing

- The semantic publishing is the use of web technologies as ontologies to enrich the meaning of scholarly documents in order to facilitate its automatic discovery and integration of data.

Scholarly Communication

involves receivers and senders

12/14/2019



Scholarly Publishing

scholarly writings are created

12/14/2019

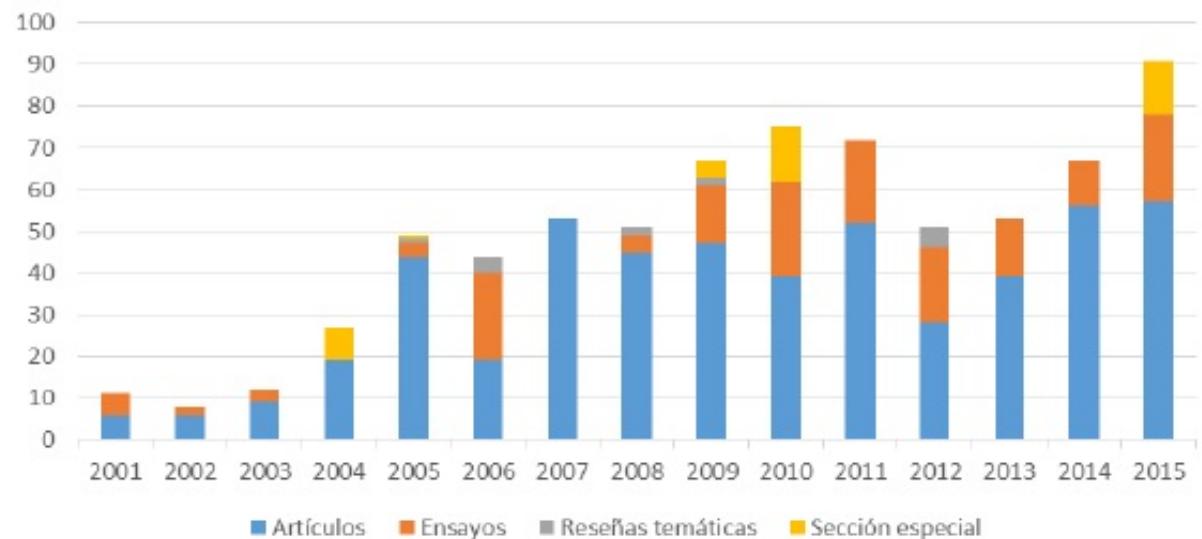


Scientific production

the amount of research produced by an author, an institution or a given country

12/14/2019

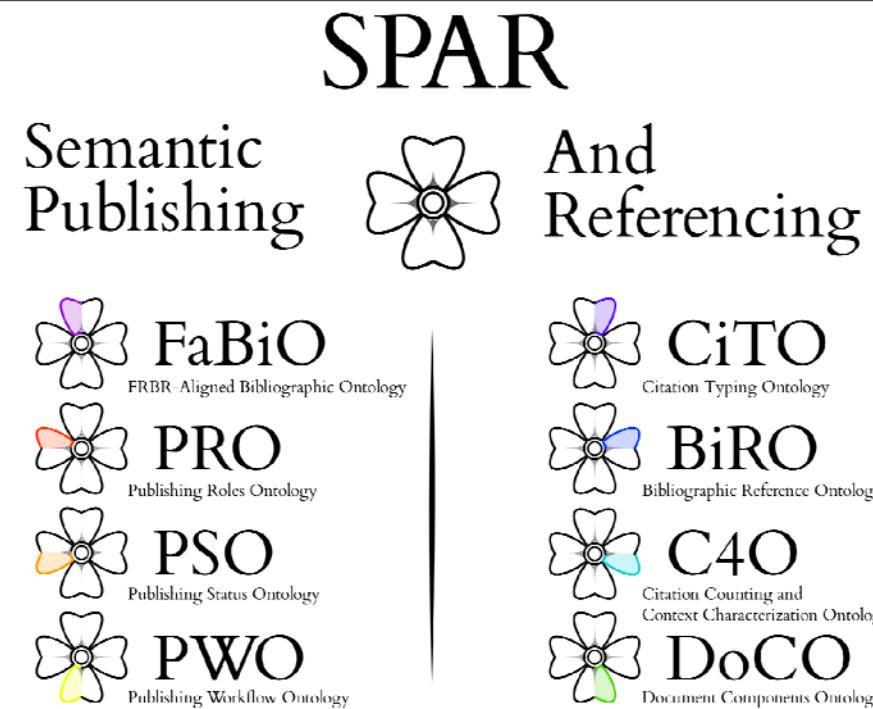
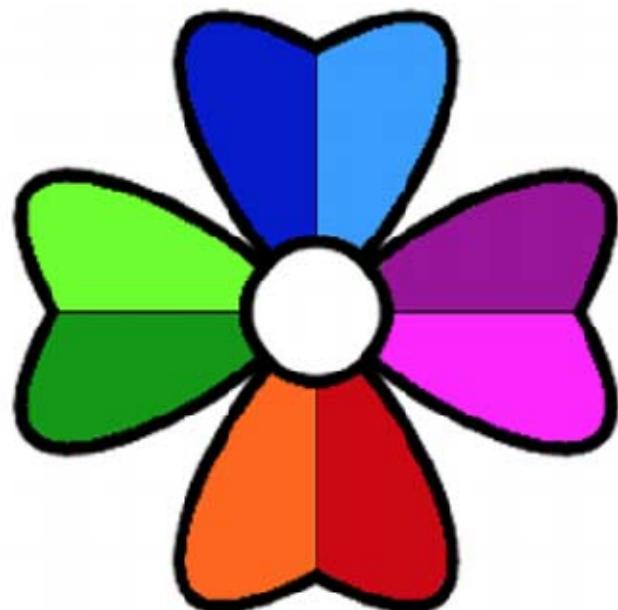
Total de documentos publicación por año según tipología
2001-2015



Related works

- Dublin Core Metadata Terms
- Functional Requirements for Bibliographic Records (FRBR)
- Publishing Requirements for Industry Standard Metadata (PRISM)
- Bibliographic Ontology Specification (BIBO - bibliontology.com)

SPAR Ontologies



SPAR Ontologies

Ontologies for describing bibliographic resources and their parts	FaBiO FRBR-DL DoCO DEO DataCite	Ontologies for describing citations of scholarly resources	CiTO BiRO C4O FOCO C2W	Ontologies for describing the publishing workflow	PRO PSO PWO SCoRO FRAPO FR	Ontologies for describing metrics and statistics for bibliographic resources	BiDO FiveStars
---	---	--	------------------------------------	---	---	--	-------------------

3. Criterios de evaluación de carreras

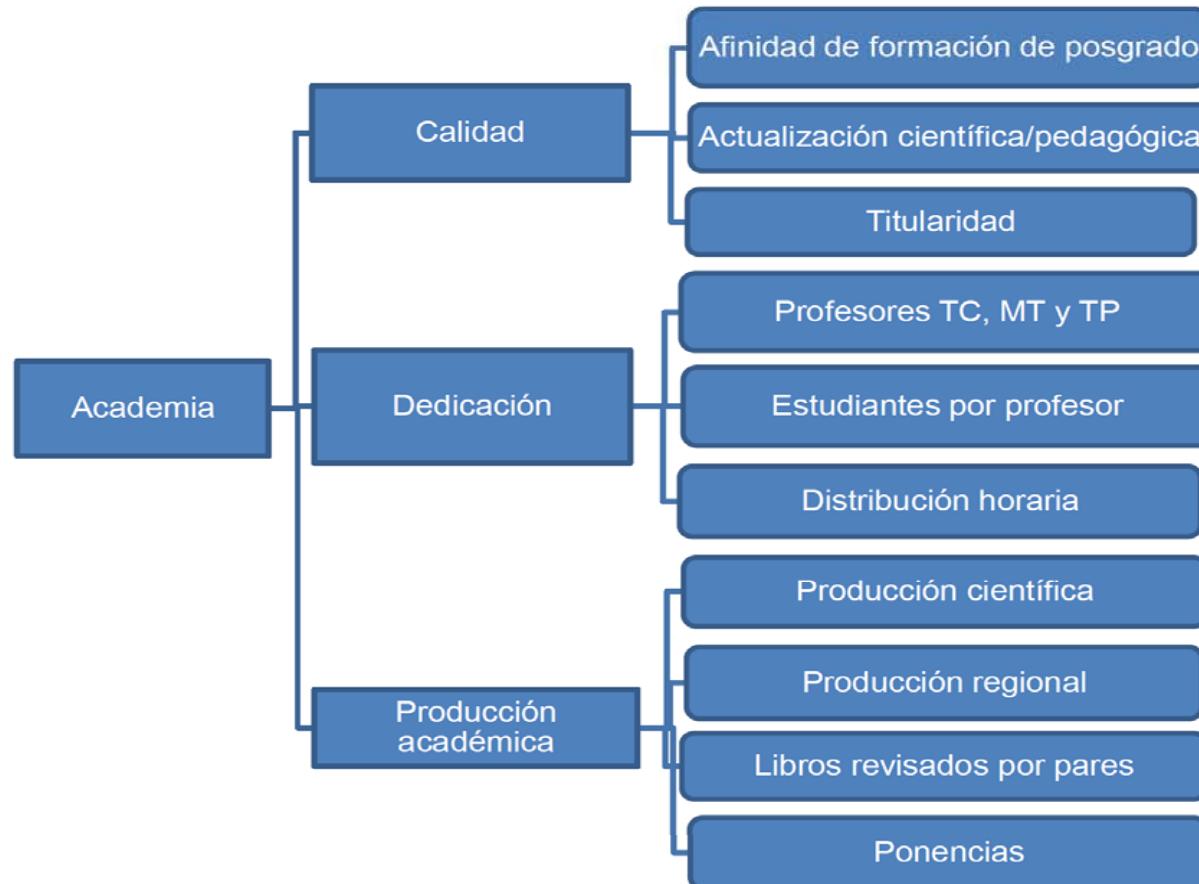
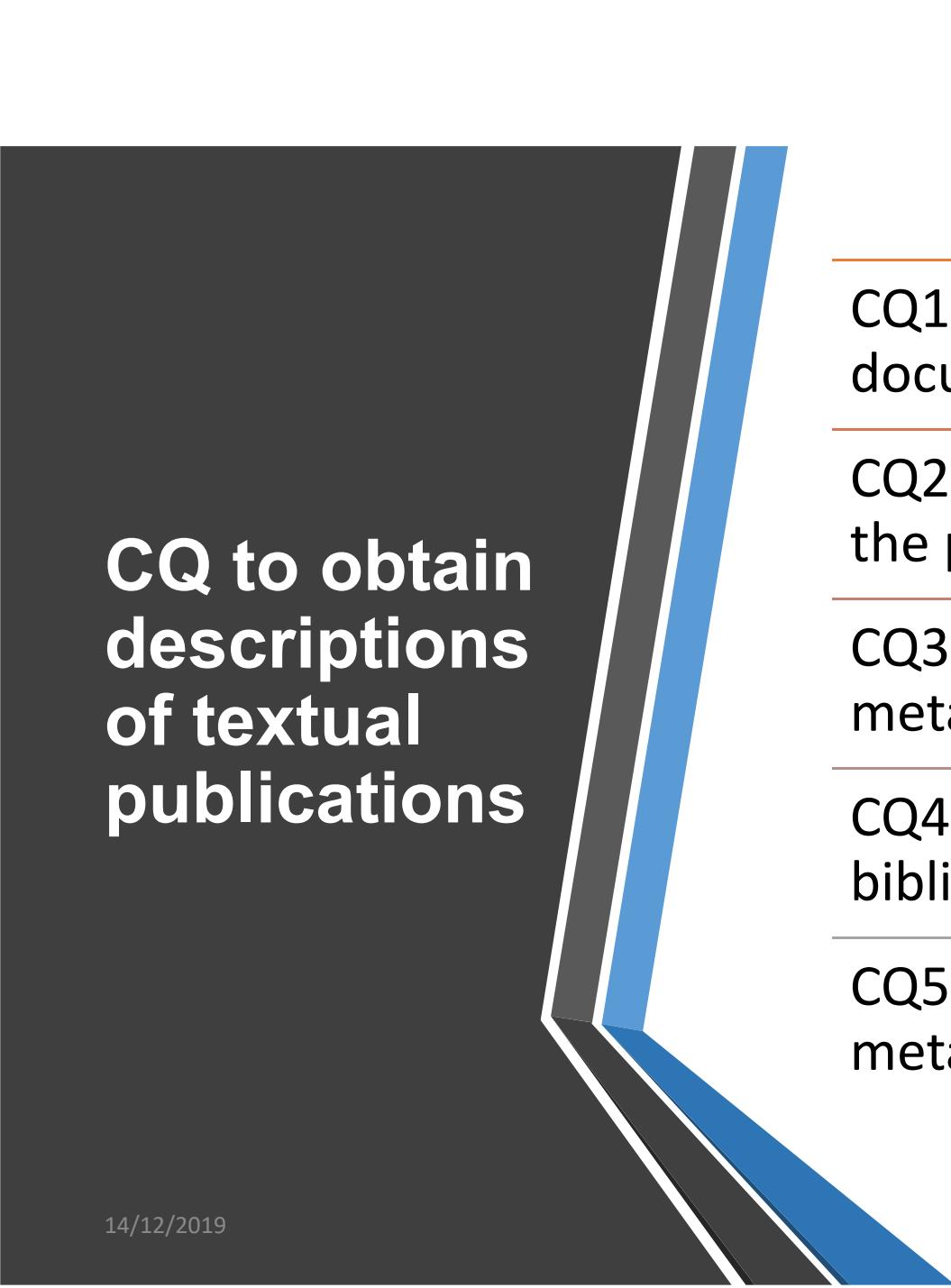


Figura 3: Criterio Academia. Ceaaces, 2015.



Producción académico - científico (C3.1)	<p>Este indicador evalúa la producción per cápita de artículos académico-científicos de los profesores/investigadores de la carrera, publicados o aceptados para su publicación en revistas que figuran en las bases de datos SCIMAGO (Scopus), o en las bases del ISI Web of Knowledge.</p> <p>Cada publicación recibirá una valoración adicional a la excelencia, en base al índice de SCIMAGO SJR, calculado en función del impacto de la revista donde ha sido publicada. En las publicaciones debe constar que el autor es profesor de la carrera o IES.</p> <p>El periodo de evaluación corresponde a los cinco últimos años previos a la evaluación.</p> <p>De existir publicaciones anteriores a los cinco últimos años, deben ser registradas con fines estadísticos.</p>	<p><i>El estándar de este indicador se fijará a través de la función de utilidad, considerando parámetros de calidad con la información obtenida del proceso de evaluación.</i></p>	$IPAC = \frac{1}{N} \left(\sum_{i=1}^M 1 + 3.61 SJR_i \right)$ <p>Dónde:</p> <p>IPAC: Indicador de producción académica científica.</p> <p>N: Número de profesores de la carrera durante el semestre en el que se efectúa la evaluación.</p> <p>M: Número de artículos académicos publicados por los profesores/investigadores de la carrera durante el periodo de evaluación.</p> <p>SJR: Índice de impacto calculado por Scimago Journal Rank.</p>	<ol style="list-style-type: none"> 1. Artículos académico-científicos publicados y/o aceptados para publicación. Archivo digital conforme a las exigencias de las revistas indexadas. 2. Notificación de aceptación del artículo para ser publicado en una revista de la base SCIMAGO (Scopus) o ISI Web. 3. Certificado de aceptación del trabajo. Puede ser un correo electrónico del editor de la revista. 4. Ficha catalográfica con: <ol style="list-style-type: none"> a. Nombre del artículo b. Nombre de la revista. c. ISSN de la revista. d. DOI del artículo. e. Volumen, número, páginas del artículo. f. Fecha de publicación
---	--	---	---	---

CQ to obtain descriptions of textual publications



CQ1: What kind of publication is the document?

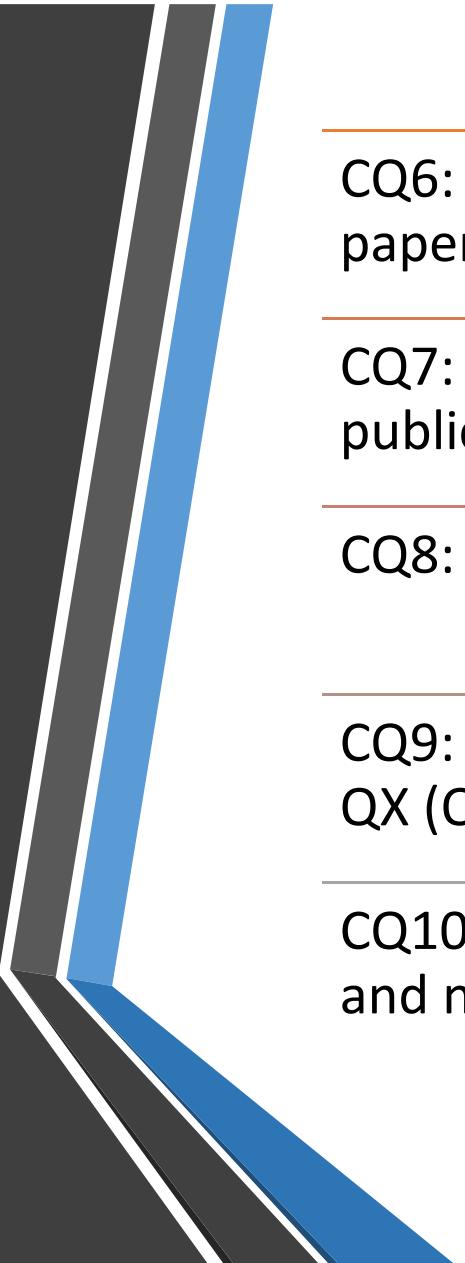
CQ2: With what other organizations were the publication made?

CQ3: What is the article's bibliographic metadata?

CQ4: What is the conference paper's bibliographic metadata?

CQ5: What is the book's bibliographic metadata?

CQ to obtain quantitative measures



CQ6: How many articles, books and conference papers have researchers published?

CQ7: How many citations has a researcher's publication received?

CQ8: How long has a researcher published?

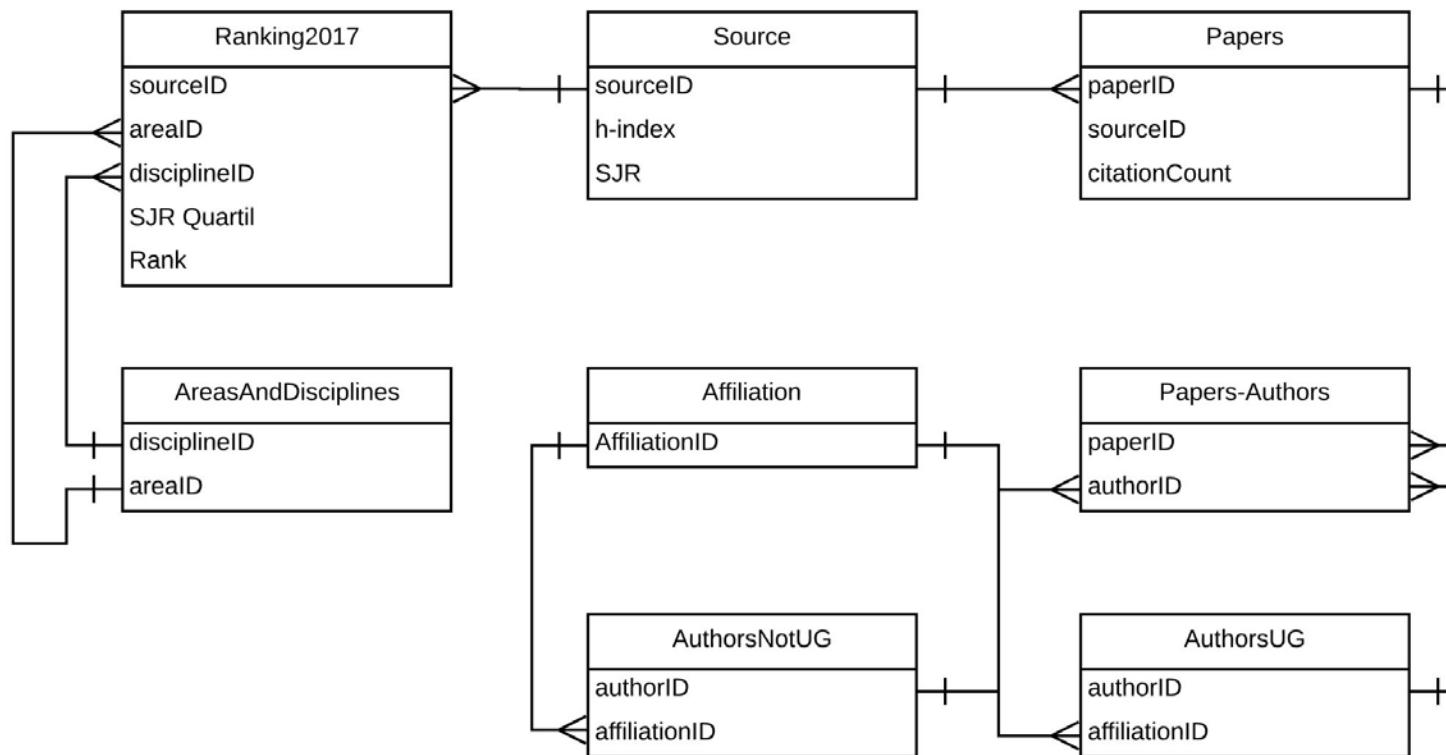
CQ9: How many researchers have published in QX (Quartile X) journals and in which area?

CQ10: What is the H-index, number of citations and number of publications of a researcher?

University of Guayaquil: Case Study

A. Obtaining Datasets	B. Relating and Cleaning Datasets	C. Mapping Datasets
 Scopus		 Refine OPEN Available Prefixes: affiliationID URI foaf:Organization add rdf:type RDF Skeleton
D. Transforming to RDF	E. Publishing RDF	F. Quering with SPARQL
	 Triplestore Virtuoso	 SPARQL

DataSets relationships



SPARQL RESULTS

- CQ1: What kind of publication is the document?
- CQ2: With what other organizations were the publication made?
- CQ3: What is the article's bibliographic metadata?
- CQ4: What is the conference paper's bibliographic metadata?
- CQ5: What is the book's bibliographic metadata?

CQ to obtain quantitative measures

- ✓ • CQ6: How many articles, books and conference papers have researchers published?
- ✗ • CQ7: How many citations have a researcher's publication received?
- ✓ • CQ8: How long has a researcher published?
- ✗ • CQ9: How many researchers have published in QX (Quartile X) journals and in which area?
- CQ10: ✓ What is the H-index, ✓ number of citations and ✗ number of publications of a researcher?

WWW.WORLDCIST.COM

- 7th World Conference on Information Systems and Technologies
- La Toja, Spain
- 16 al 19 de abril del 2019
- EI-Compendex, SCOPUS, DBLP and Google Scholar

Using the SPAR Ontology Network to represent the Scientific Production of a University: a Case Study

Mariela Tapia-Leon^{1[0000-0002-2609-5955]} Janneth Chicaiza Espinosa^{2[0000-0003-3439-3618]}
Paola Espinoza Arias^{3[0000-0002-3938-2064]} Idafen Santana-Perez^{3[0000-0001-8296-8629]} Oscar
Corcho^{3[0000-0002-9260-0753]}

¹ Universidad de Guayaquil, Guayaquil, Ecuador

² Universidad Técnica Particular de Loja, Loja, Ecuador

³ Universidad Politécnica de Madrid, Madrid, Spain
mariela.tapial@ug.edu.ec

Abstract. Research is commonly used to measure the prestige of universities. Currently, many universities register some of their scientific production (such as thesis, articles) in open access repositories using technologies like DSpace or ePrints. Likewise, scientific production is available in different and overlapping databases such as Scopus, Web of Science, Google Scholar, Crossref. Connecting these datasets, with the application of ontologies, will potentially increase their value and help discover interesting relationships amongst them. The present study aims to use the SPAR Ontology Network, which allows representing scholarly publishing, in order to check whether it is possible to represent the scientific production of universities. For that, we propose competency questions with the purpose of measuring scientific production. Likewise, we obtained data from Scopus regarding an Ecuadorian university (University of Guayaquil) and transformed it to RDF using the SPAR Ontology Network and other well-known on-

SPAR Ontologies

Ontologies for describing bibliographic resources and their parts	FaBiO FRBR-DL DoCO DEO DataCite	Ontologies for describing citations of scholarly resources	CiTO BiRO C4O FOCO C2W	Ontologies for describing the publishing workflow	PRO PSO PWO SCoRO FRAPO FR	Ontologies for describing metrics and statistics for bibliographic resources	BiDO FiveStars
---	---	--	------------------------------------	---	---	--	-------------------

BIDO Ontology

BiDO Core Module (BiDO-Core)

BiDO Standard Bibliometric
Measures module (BiDO-
Standard-Bibliometric-Measures)

BiDO Review Measures module
(BiDO-Review-Measures)

BiDO Research Career Categories
module (BiDO-Research-Career-
Categories)

BIDO Ontology

- BiDO Core Module (BiDO-Core)
- BiDO Standard Bibliometric Measures module (BiDO-Standard-Bibliometric-Measures)
- BiDO Review Measures module (BiDO-Review-Measures)
- BiDO Research Career Categories module (BiDO-Research-Career-Categories)

Existing Terms

author
citation count

h-index

e-index

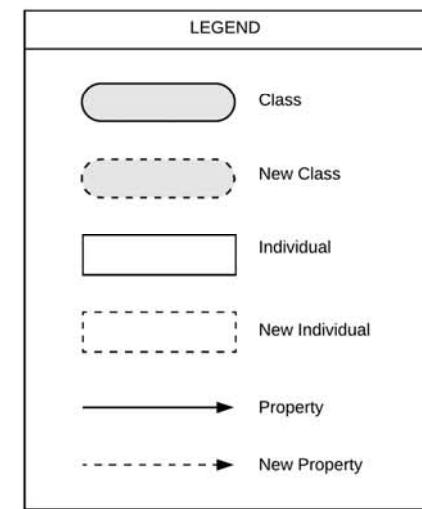
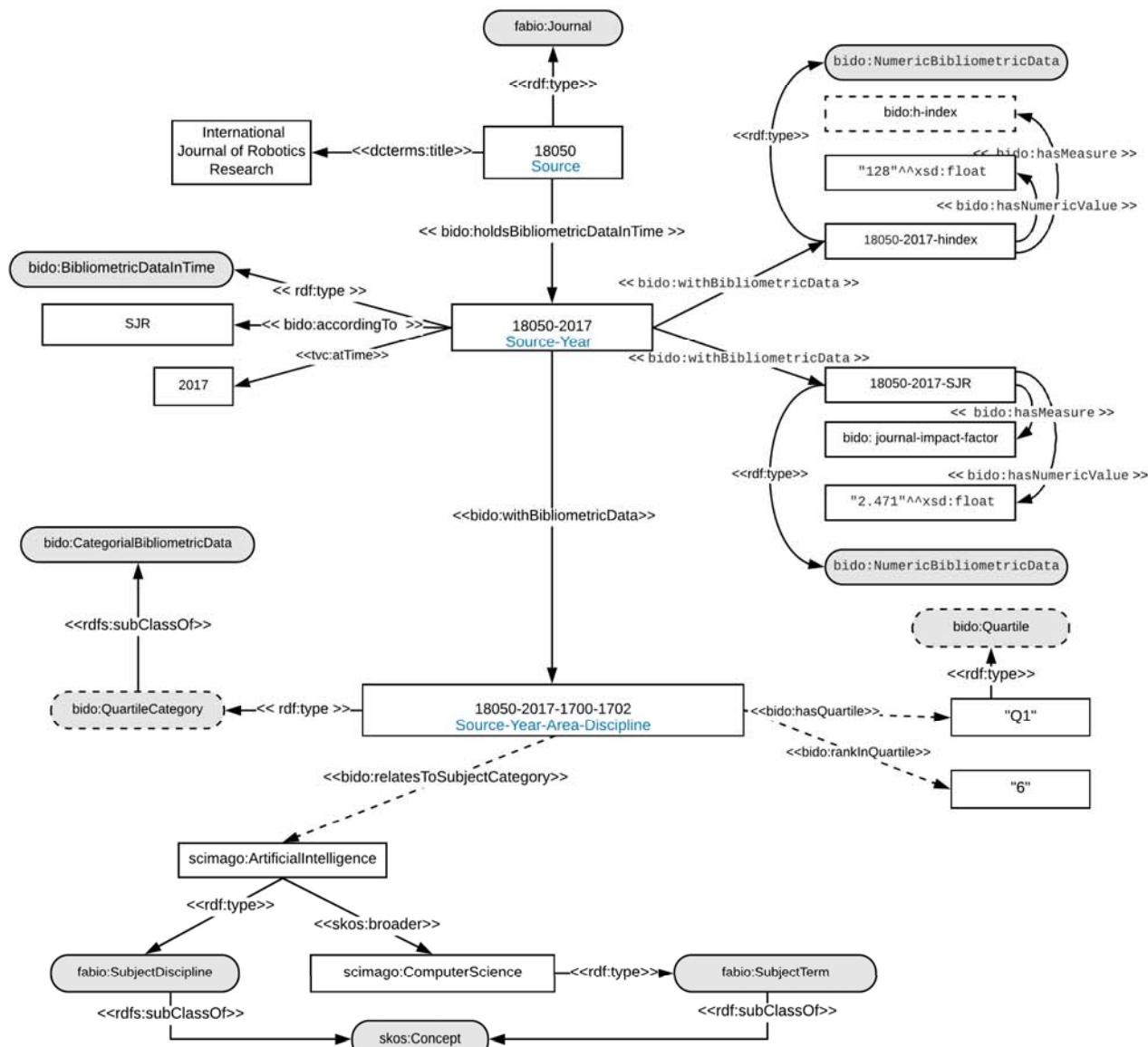
journal
impact factor

Missing Terms

- T1. Author number of documents
- T2. Paper citation count
- T3. Journal h-index
- T4. SJR or JCR (or similar) **rank**
- T5. Quartile in its correspondent area and discipline
- T6. Rank according to a specific discipline

Propose of BIDO extension

Term Label	Type	Description
quartile category	C	A list of journals ordered from highest to lowest impact factor is divided into four equal parts, each of these parts will be a quartile category: Q1, Q2, Q3, Q4.
quartile	C	The quartile is an indicator that serves to evaluate the relative importance of a journal within the total number of journals in its area, e.g., Q1.
has quartile	OP	An object property relating a journal in determined area and discipline with its quartile
relates to subject category	OP	An object property relating a journal with its area and discipline of study
rank in quartile	DP	An object property relating a journal in determined area and discipline with the position it occupies in the quartile, e.g., 2, 2/76, 2nd (of 76).
author number documents	NI	It is the count of all the documents or contributions published by a particular author.
paper citation count	NI	It is the count of all the citation received by one paper of a particular author.
Q1	NI	If a list of 1000 journals is ordered from highest to lowest impact factor, the journals between the number 1 to 250 are Quartile 1 (Q1).
Q2	NI	If a list of 1000 journals is ordered from highest to lowest impact factor, the journals between the number 251 to 500 are Quartile 2 (Q2).
Q3	NI	If a list of 1000 journals is ordered from highest to lowest impact factor, the journals between the number 501 to 750 are Quartile 3 (Q3).
Q4	NI	If a list of 1000 journals is ordered from highest to lowest impact factor, the journals between the number 751 to 1000 are Quartile 4 (Q4).



Competency Questions

CQ1. How many publications have a researcher?

CQ2. How many citations have a publication?

CQ3. Which is the h-index of the journal?

CQ4. How many publications have been published in journals with SJR or JCR?

CQ5. How many publications have published in determinate quartile and in which area and discipline?

CQ6. Which is the rank of the journal?

University of Guayaquil: Case Study

A. Obtaining Datasets	B. Relating and Cleaning Datasets	C. Mapping Datasets
		 Refine ^{OPEN} Available Prefixes: affiliationID URI foaf:Organization add rdf:type
D. Transforming to RDF	E. Publishing RDF	F. Quering with SPARQL
	 Triplestore Virtuoso	

Competency Questions

- ✓ • CQ1. How many publications have a researcher?
- ✓ • CQ2. How many citations have a publication?
- ✓ • CQ3. Which is the h-index of the journal?
- ✓ • CQ4. How many publications have been published in journals with SJR or JCR?
- ✓ • CQ5. How many publications have published in determinate quartile and in which area and discipline?
- ✓ • CQ6. Which is the rank of the journal?

- International Conference Educational information and Technology
- Cambridge, UK
- 2 al 4 de marzo del 2019
- EI-Compendex, SCOPUS,

Extension of BiDO Ontology to Represent the Scientific Production of Researchers

Mariela Tapia-Leon
Universidad de Guayaquil
Guayaquil, Ecuador
mariela.tapial@ug.edu.ec

Paola Espinoza-Arias
Universidad Politécnica de Madrid
Madrid, Spain
pesinoza@fi.upm.es

Idafen Santana-Perez
Universidad Politécnica de Madrid
Madrid, Spain
isantana@fi.upm.es

Janneth Chicaiza
Universidad Técnica Particular de Loja
Loja, Ecuador
jachicaiza@utpl.edu.ec

Maria Poveda-Villalón
Universidad Politécnica de Madrid
Madrid, Spain
mpoveda@fi.upm.es

Oscar Corcho
Universidad Politécnica de Madrid
Madrid, Spain
ocorcho@fi.upm.es

Abstract—The Spar Ontology Network is a suit of complementary ontology modules to describe the scholarly publishing domain. BiDO Standard Bibliometric Measures is part of its set of ontologies. It allows the description of numerical and categorical bibliometric data such as h-index, author citation count, journal impact factor. These measures could be used to evaluate the scientific production of researchers. However, they are not enough. In a previous study, we determined the lack of some terms to provide a more complete representation of scientific production. This study aimed to build an extension to BiDO ontology. For that purpose, we used NeOn Methodology to restructure BiDO ontology and Protégé to build it. After the correspondent validation of the extension using OOPS!, we sent the extension propose to the owners of the SPAR Ontologies for their approval. With this extension, it is possible to represent and measure the number of documents from research, the number of citations from a paper and the number of publications in high impact journals according to its area and discipline.

BiDO, RDF, Ontology, Scientific Production, Scholarly Publishing, SPAR Ontology Network, SPARQL.

bibliographic resources and their parts (FaBiO, FRBL-DL, DoCO, DEO, DataCite), describing citations of scholarly resources (CITO, BiRO, C4O, FOCO, C2W), describing the publishing workflow (PRO, PSO, PWO, SCoRO, FRAPO, FR), and describing metrics and statistics for bibliographic resources (BiDO, FiveStars). They are developed in OWL 2 DL for the creation of comprehensive machine-readable RDF metadata for every aspect of semantic publishing and referencing [2], and it is well documented in www.sparontologies.net.

In a previous study [3] we evaluated the capability of SPAR Ontology Network to cover the publishing domain and some measures that represent the scientific production, such as author h-index, author citations count, and journal impact factor. However, other measures such as author number of documents, paper citations count, journal h-index, rank, quartile and area of study cannot be represented with them. Hence, the objective of this study is proposing an extension to BiDO ontology, specifically BiDO Standard Bibliometric Measures to cover the missing measures. We followed NeOn Methodology for building the ontology and OOPS!¹ for the

Agenda

- What was I going to do here?
- What did I really do?
- But first, how did I feel?
- **Suggestions**
- Your suggestions

Sugerencias para futuros pasantes

- Integración: Actividades para unir al pasante en el grupo.
- Aprendizaje: Involucrar a los pasantes en proyectos en desarrollo.
- Especialización: Aportes específicos del pasante dentro de ese proyecto.
- Experiencia: Participación en talleres, clases, eventos que refuerzen la confianza del pasante.
- Control: Seguimiento de las actividades solicitadas al pasante para su refuerzo.

Agenda

- What was I going to do here?
- What did I really do?
- But first, how did I feel?
- Suggestions
- Your suggestions



MUCHAS GRACIAS