



UNIVERSIDAD AUTÓNOMA DE YUCATÁN  
FACULTAD DE MATEMÁTICAS

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

UN SISTEMA DE ÍNDICES PARA VALORAR LA RELEVANCIA DE LA  
PRODUCCIÓN ACADÉMICA Y LA INVESTIGACIÓN A PARTIR DE  
REPOSITORIOS DIGITALES Y METADATOS

TESIS

PRESENTADA POR:

ING. JARED DAVID TADEO GUERRERO SOSA

EN OPCIÓN AL GRADO DE:

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

DIRECTORES DE TESIS:

DR. VÍCTOR HUGO MENÉNDEZ DOMÍNGUEZ  
M.G.T.I. MARÍA ENRIQUETA CASTELLANOS BOLAÑOS

MÉRIDA, YUCATÁN, MÉXICO

AGOSTO, 2019

## **DEDICATORIA**

*A Dios por todo.*

*A mis padres. Gracias por siempre guiarme y ayudarme a ser quien soy. No tengo palabras para agradecer todo lo que han hecho por mí.*

*A mis hermanos por su apoyo durante toda mi vida.*

*A mis amigos que me han motivado a dar más de mí siempre. A los de toda la vida y a los que conocí en este camino.*

*A las personas que he conocido en este trayecto y alguna vez me dieron palabras de aliento.*

## **AGRADECIMIENTOS**

A mi asesor Víctor Hugo. Gracias por apoyar y difundir la importancia de este proyecto, por guiarme a través de su conocimiento y, sobre todo, por su amistad.

A María Enriqueta por su orientación y valiosa contribución en este proyecto.

A Salvador por compartir sus conocimientos para los estudios de correlación de los indicadores.

A la Coordinación General de Posgrado, Investigación y Vinculación de la Universidad Autónoma de Yucatán por las facilidades otorgadas.

A los profesores de la Facultad de Matemáticas.

A Arturo, del repositorio nacional del CONACYT, por el apoyo e interés en este proyecto.

A Francisco, José Ángel y Jesús de la Escuela Superior de Informática de la Universidad de Castilla-La Mancha por permitirme vivir la experiencia de una estancia en el extranjero para esta investigación.

Al CONACYT por la beca otorgada.

A la Universidad Autónoma de Yucatán y a la Facultad de Matemáticas.

## RESUMEN

Los repositorios digitales de producción científica permiten el acceso a la información de las diversas publicaciones científicas. Por otra parte, los repositorios de acceso abierto brindan los contenidos en texto completo. Cada recurso almacenado en un repositorio es descrito, localizado y referenciado por medio de los metadatos.

Sin embargo, los metadatos suelen adolecer de integridad, completitud, exactitud y consistencia, principalmente aquellos que contienen el nombre de los autores de la publicación.

Diversas instancias, como el Sistema Nacional de Investigadores del Consejo Nacional de Ciencia y Tecnología, y el Programa para el Desarrollo Profesional Docente para el Tipo Superior, ambas en México, se encargan de evaluar y reconocer a los investigadores y profesores que realizan labores de investigación con base en la producción almacenada en diversos repositorios.

Considerando que los valores de los metadatos suelen ser erróneos, una de las posibles consecuencias es una evaluación incorrecta a los investigadores por parte de las instancias evaluadoras.

Es así como en esta tesis se propone (1) un modelo de indicadores para la relevancia científica (2) un sistema software que mide la relevancia de la producción académica y la investigación a partir de repositorios digitales y metadatos con base en el modelo de indicadores propuesto.

Se plantea el uso de servicios web y del protocolo para la interoperabilidad entre repositorios OAI-PMH para la recuperación de la producción científica, utilizando diversas condiciones que validan si la producción pertenece a un investigador en específico.

La arquitectura del sistema está organizada por capas, facilitando la comunicación entre los componentes. Además, es independiente de cualquier herramienta tecnológica, permitiendo que sea adaptable a necesidades específicas.

Se hicieron estudios de correlación entre los indicadores de la relevancia del investigador en diversos escenarios de la Facultad de Matemáticas de la Universidad Autónoma de Yucatán (UADY), observando los diferentes comportamientos en la investigación. También se analizó el estado actual de la investigación en la UADY por medio de técnicas de minería de datos, con la finalidad de caracterizar a los diversos grupos de cuerpos académicos y de investigadores de acuerdo con su relevancia; así como reafirmar que en la UADY se cumplen los lineamientos de las instancias evaluadoras.

Los experimentos desarrollados muestran que esta propuesta puede mejorar significativamente muchos de los procesos asociados con la evaluación de la producción científica que generalmente se realizan de forma manual pero que resultan complejos por la cantidad de información a examinar y además, con posibles inconsistencias.

***Palabras clave:** Repositorios digitales; metadatos; interoperabilidad; OAI-PMH; índices de relevancia científica; producción indizada.*

## TABLA DE CONTENIDO

DEDICATORIA.....	i
AGRADECIMIENTOS.....	ii
RESUMEN.....	iii
LISTA DE TABLAS.....	ix
LISTA DE FIGURAS.....	xi
<b>1 INTRODUCCIÓN.....</b>	<b>1</b>
1.1 ENTORNO DE LA INVESTIGACIÓN.....	1
1.2 HIPÓTESIS Y OBJETIVO.....	2
1.3 METODOLOGÍA DE LA INVESTIGACIÓN.....	4
1.4 TRABAJOS RELACIONADOS.....	5
1.5 MARCO DE TRABAJO.....	8
1.6 APORTACIONES.....	9
1.6.1 APORTACIONES TEÓRICAS.....	9
1.6.2 APORTACIONES PRÁCTICAS.....	10
1.7 ORGANIZACIÓN DEL DOCUMENTO.....	10
<b>2 ESTADO DEL ARTE.....</b>	<b>12</b>
2.1 REPOSITORIO DIGITAL.....	12
2.1.1 DEFINICIÓN.....	12
2.1.2 CARACTERÍSTICAS.....	13
2.1.3 HERRAMIENTAS PARA EL DESARROLLO DE REPOSITORIOS.....	13
2.1.4 INTEROPERABILIDAD.....	15
2.1.5 IMPORTANCIA DE LOS REPOSITORIOS DE DOCUMENTACIÓN CIENTÍFICA.....	17
2.1.6 SISTEMAS CRIS.....	19
2.2 METADATOS.....	20
2.2.1 DEFINICIÓN.....	20
2.2.2 ESTÁNDARES.....	20
2.2.3 DUBLIN CORE.....	21
2.2.4 CALIDAD DE LOS METADATOS.....	23

2.2.5	PROBLEMAS EXISTENTES CON LOS METADATOS .....	25
2.3	RELEVANCIA EN LAS INVESTIGACIONES CIENTÍFICAS .....	27
2.3.1	PUBLICACIONES INDIZADAS .....	27
2.3.2	INDICADORES CUANTITATIVOS PARA LA PRODUCTIVIDAD CIENTÍFICA ...	28
2.4	INSTANCIAS EVALUADORAS DE LA PRODUCTIVIDAD CIENTÍFICA .....	32
2.4.1	SISTEMA NACIONAL DE INVESTIGADORES.....	33
2.4.2	PROGRAMA PARA EL DESARROLLO PROFESIONAL DOCENTE PARA EL TIPO SUPERIOR .....	37
2.4.3	OTROS PAÍSES .....	39
2.5	REPOSITORIOS INSTITUCIONALES DE ACCESO ABIERTO .....	39
2.5.1	SCOPUS.....	39
2.5.2	REPOSITORIO NACIONAL.....	40
2.5.3	REPOSITORIOS INSTITUCIONALES .....	41
2.5.4	REMERI.....	42
2.5.5	REDALYC.....	42
2.5.6	REDI-UADY .....	43
2.6	NORMATIVAS PARA LA CIENCIA ABIERTA EN MÉXICO.....	45
2.6.1	LINEAMIENTOS RELATIVOS AL PROGRAMA DE REPOSITORIOS .....	45
2.6.2	LINEAMIENTOS TÉCNICOS PARA LOS REPOSITORIOS MEXICANOS .....	46
2.6.3	OTROS PAÍSES .....	46
2.7	CUERPOS ACADÉMICOS.....	47
2.8	CONCLUSIONES.....	49
<b>3</b>	<b><i>MODELO DE INDICADORES PARA LA RELEVANCIA CIENTÍFICA.....</i></b>	<b>50</b>
3.1	RELEVANCIA DE UN CUERPO ACADÉMICO .....	50
3.1.1	CÁLCULO DE LA PRODUCCIÓN INDIZADA POR SCOPUS.....	50
3.1.2	CÁLCULO DEL GRADO DE COLABORACIÓN DEL CUERPO ACADÉMICO .....	54
3.1.3	CÁLCULO DE LA PRODUCCIÓN NO INDIZADA DEL CUERPO ACADÉMICO..	55
3.1.4	CÁLCULO DE LA RELEVANCIA DE UN CUERPO ACADÉMICO .....	55
3.2	RELEVANCIA DE UN INVESTIGADOR.....	56
3.2.1	RELEVANCIA EN CUERPO ACADÉMICO.....	56

3.2.2	CÁLCULO DE LA RELEVANCIA DE LA COLABORACIÓN EXTERNA AL CUERPO ACADÉMICO.....	58
3.2.3	CÁLCULO DE LA RELEVANCIA DE LA PRODUCTIVIDAD INDIVIDUAL .....	59
3.2.4	CÁLCULO DE LA PRODUCTIVIDAD NO INDIZADA DEL INVESTIGADOR .....	60
3.2.5	CÁLCULO DE LA RELEVANCIA DE UN INVESTIGADOR.....	60
3.3	CONCLUSIONES.....	61
<b>4</b>	<b>SISTEMA GENERADOR DE INDICADORES DE RELEVANCIA CIENTÍFICA.....</b>	<b>62</b>
4.1	MARCO ARQUITECTÓNICO .....	62
4.2	ESPECIFICACIONES TÉCNICAS.....	65
4.2.1	TECNOLOGÍAS DE DESARROLLO.....	65
4.2.2	ESTRUCTURA DE LA BASE DE DATOS NOSQL.....	66
4.3	INTERFAZ WEB.....	68
4.4	FUNCIONALIDAD DEL SISTEMA.....	72
4.4.1	CONSULTA Y ALMACENAMIENTO DE LAS PUBLICACIONES CIENTÍFICAS .	72
4.4.2	GENERACIÓN DEL INDICADOR DE LA PRODUCTIVIDAD DE CUERPOS ACADÉMICOS .....	87
4.4.3	GENERACIÓN DEL INDICADOR DE LA PRODUCTIVIDAD DE LOS INVESTIGADORES .....	89
4.5	REPRESENTACIÓN DEL CONOCIMIENTO .....	91
4.6	CONCLUSIONES.....	95
<b>5</b>	<b>RESULTADOS.....</b>	<b>96</b>
5.1	CORRELACIÓN DE LOS INDICADORES PARA LOS CUERPOS ACADÉMICOS DE LA FMAT.....	104
5.2	CORRELACIÓN DE LOS INDICADORES PARA LA FMAT POR ÁREAS.....	105
5.3	CORRELACIÓN DE LOS INDICADORES PARA LA FMAT PARA INVESTIGADORES CON CA.....	111
5.4	CORRELACIÓN DE LOS INDICADORES PARA LA FMAT PARA INVESTIGADORES SIN CA.....	117
5.5	CORRELACIÓN DE LOS INDICADORES EN CADA CA DE COMPUTACIÓN.....	120

5.6	CORRELACIÓN DE LOS INDICADORES EN CADA CA DE MATEMÁTICAS.....	125
5.7	ANÁLISIS DE LA RELEVANCIA CIENTÍFICA A PARTIR DE TÉCNICAS DE MINERÍA DE DATOS .....	135
5.8	DISCUSIÓN DE LOS RESULTADOS.....	148
5.9	CONCLUSIONES.....	151
<b>6</b>	<b>CONCLUSIONES Y TRABAJO FUTURO.....</b>	<b>152</b>
6.1	INTRODUCCIÓN.....	152
6.2	ANÁLISIS DE LA CONSECUCCIÓN DE LOS OBJETIVOS .....	153
6.3	ANÁLISIS DE LAS APORTACIONES .....	157
6.4	CONTRASTE DE RESULTADOS .....	159
6.4.1	ARTÍCULOS EN REVISTAS.....	159
6.4.2	PONENCIAS EN CONGRESOS INTERNACIONALES.....	159
6.5	TRABAJO FUTURO .....	160
6.5.1	MEJORAS EN LAS TÉCNICAS UTILIZADAS .....	160
6.5.2	MEJORAS EN EL SISTEMA .....	161
<b>7</b>	<b>BIBLIOGRAFÍA.....</b>	<b>164</b>

## LISTA DE TABLAS

TABLA 2.1 ESTÁNDARES DE METADATOS .....	21
TABLA 2.2 COMPARACIÓN DE LOS CRITERIOS DE EVALUACIÓN DE CADA ÁREA DEL SNI .....	36
TABLA 4.1 PROPIEDADES DEL MODELO ONTOLÓGICO .....	92
TABLA 5.1 CUERPOS ACADÉMICOS DE LA UADY POR CAMPUS Y GRADO DE CONSOLIDACIÓN .....	97
TABLA 5.2 PROFESORES DE LA UADY POR CAMPUS Y GÉNERO .....	97
TABLA 5.3 PROFESORES DE TIEMPO COMPLETO DE LA UADY DESDE LA PERSPECTIVA DEL PRODEP Y DEL SNI.....	98
TABLA 5.4 PROFESORES DE LA UADY CON PRODUCCIÓN CIENTÍFICA DESDE LA PERSPECTIVA DEL PRODEP Y DEL SNI .....	99
TABLA 5.5 PRODUCCIÓN CIENTÍFICA DE LA UADY ALMACENADA EN SCOPUS Y EN EL REPOSITORIO NACIONAL .....	100
TABLA 5.6 INSTITUCIONES MEXICANAS CON MAYOR COLABORACIÓN CON LA UADY.....	102
TABLA 5.7 PAÍSES CON MAYOR COLABORACIÓN CON LA UADY .....	102
TABLA 5.8 INSTITUCIONES EXTRANJERAS CON MAYOR COLABORACIÓN CON LA UADY .....	103
TABLA 5.9 NOTACIÓN DE LOS INDICADORES UTILIZADOS PARA LOS ESTUDIOS DE CORRELACIÓN .....	104
TABLA 5.10 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE LOS CUERPOS ACADÉMICOS DE LA FMAT (N=8) .....	105
TABLA 5.11 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE LOS INVESTIGADORES DEL ÁREA DE COMPUTACIÓN DE LA FMAT (N=42) .....	106
TABLA 5.12 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL ÁREA DE MATEMÁTICAS DE LA FMAT (N=53).....	107
TABLA 5.13 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES CON CUERPO ACADÉMICO DEL ÁREA DE COMPUTACIÓN DE LA FMAT (N=19).....	112
TABLA 5.14 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES CON CUERPO ACADÉMICO DEL ÁREA DE MATEMÁTICAS DE LA FMAT (N=19).....	113
TABLA 5.15 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES SIN CUERPO ACADÉMICO DEL ÁREA DE COMPUTACIÓN DE LA FMAT (N=23).....	117
TABLA 5.16 COEFICIENTES DE CORRELACIÓN DE SPEARMAN ENTRE LAS MÉTRICAS DE PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES SIN CUERPO ACADÉMICO DEL ÁREA DE MATEMÁTICAS DE LA FMAT (N=34).....	118

TABLA 5.17 ESTADO DE LA PRODUCCIÓN Y RELEVANCIA DE LOS CUERPOS ACADÉMICOS.....	136
TABLA 5.18 ESTADO DE LA PRODUCCIÓN Y RELEVANCIA DE LOS INVESTIGADORES.....	143

## LISTA DE FIGURAS

FIGURA 1.1 INDICADORES PRESENTADOS POR SCOPUS EN LOS PERFILES DE AUTORES .....	6
FIGURA 1.2 INDICADORES PRESENTADOS POR SCOPUS EN LAS REVISTAS Y LIBROS SERIALIZADOS .....	7
FIGURA 2.1 ESTADÍSTICA POR ÁREA DEL NÚMERO DE INVESTIGADORES PERTENECIENTES AL SNI.....	37
FIGURA 2.2 ARQUITECTURA DE REDI-UADY .....	44
FIGURA 3.1 EJEMPLO DE UNA PUBLICACIÓN PERTENECIENTE A UN CUERPO ACADÉMICO.....	57
FIGURA 4.1 DISEÑO DE LA ARQUITECTURA PROPUESTA .....	63
FIGURA 4.2 INTERACCIÓN DE LAS TECNOLOGÍAS INVOLUCRADAS EN EL SISTEMA.....	66
FIGURA 4.3 ESQUEMA DE RELACIONES ENTRE BASES DE DATOS Y SUS COLECCIONES .....	67
FIGURA 4.4 PÁGINA PRINCIPAL DE LA INTERFAZ DEL USUARIO .....	68
FIGURA 4.5 VISUALIZACIÓN DE LOS RESULTADOS DE LA RELEVANCIA DE CUERPOS ACADÉMICOS .....	69
FIGURA 4.6 VISUALIZACIÓN DE LOS RESULTADOS DE LA RELEVANCIA DE INVESTIGADORES .....	70
FIGURA 4.7 GRAFO DE COLABORACIÓN ENTRE CUERPOS ACADÉMICOS EN LA INTERFAZ DEL USUARIO .....	71
FIGURA 4.8 GRAFO DE COLABORACIÓN ENTRE INVESTIGADORES EN LA INTERFAZ DEL USUARIO .....	71
FIGURA 4.9 EJEMPLO DE DICCIONARIO DE LAS FACULTADES .....	73
FIGURA 4.10 PRIMERA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	74
FIGURA 4.11 SEGUNDA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	75
FIGURA 4.12 TERCERA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	76
FIGURA 4.13 CUARTA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	77
FIGURA 4.14 QUINTA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	78
FIGURA 4.15 SEXTA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	79
FIGURA 4.16 SÉPTIMA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR.....	81
FIGURA 4.17 OCTAVA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	82
FIGURA 4.18 NOVENA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	84
FIGURA 4.19 DÉCIMA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR.....	85
FIGURA 4.20 UNDÉCIMA CONDICIÓN PARA LA SELECCIÓN DE ID DE SCOPUS DEL INVESTIGADOR .....	86
FIGURA 4.21 PROCEDIMIENTO DE CÁLCULO DE INDICADORES PARA CUERPOS ACADÉMICOS .....	89
FIGURA 4.22 EJEMPLO DEL REGISTRO DE LAS AFILIACIONES DE LOS INVESTIGADORES EN PUBLICACIONES CIENTÍFICAS .....	89
FIGURA 4.23 PROCEDIMIENTO DE CÁLCULO DE INDICADORES PARA INVESTIGADORES .....	90
FIGURA 4.24 CLASES Y SUBCLASES RELACIONADAS CON PUBLICACIONES Y SUS PROPIEDADES.....	95
FIGURA 5.1 PRODUCCIÓN CIENTÍFICA DE LOS PROFESORES DE LA UADY ALMACENADA EN DIVERSOS REPOSITORIOS .....	100
FIGURA 5.2 PRODUCCIÓN CIENTÍFICA DE LA UADY POR AÑOS.....	101

FIGURA 5.3 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAC1 DE LA FMAT .....	121
FIGURA 5.4 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAC2 DE LA FMAT .....	122
FIGURA 5.5 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAC3 DE LA FMAT .....	124
FIGURA 5.6 RELEVANCIA Y PRODUCCIÓN TOTAL DE LOS CUERPOS ACADÉMICOS DEL ÁREA DE COMPUTACIÓN DE LA FMAT.....	125
FIGURA 5.7 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAM1 DE LA FMAT ....	127
FIGURA 5.8 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAM2 DE LA FMAT ....	129
FIGURA 5.9 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAM3 DE LA FMAT ....	130
FIGURA 5.10 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAM4 DE LA FMAT ..	132
FIGURA 5.11 MÉTRICAS DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES DEL CAM5 DE LA FMAT ..	133
FIGURA 5.12 RELEVANCIA Y PRODUCCIÓN TOTAL DE LOS CUERPOS ACADÉMICOS DEL ÁREA DE MATEMÁTICAS DE LA FMAT.....	134
FIGURA 5.13 CONJUNTOS BORROSOS DEL NÚMERO DE INTEGRANTES DE LOS CUERPOS ACADÉMICOS .....	139
FIGURA 5.14 CONJUNTOS BORROSOS DEL PTS DE LOS CUERPOS ACADÉMICOS .....	139
FIGURA 5.15 CONJUNTOS BORROSOS DEL GCC DE LOS CUERPOS ACADÉMICOS .....	140
FIGURA 5.16 CONJUNTOS BORROSOS DEL PTPNI DE LOS CUERPOS ACADÉMICOS .....	140
FIGURA 5.17 CONJUNTOS BORROSOS DE LA PRODUCCIÓN TOTAL DE LOS CUERPOS ACADÉMICOS.....	141
FIGURA 5.18 CORRELACIÓN DE CADA VARIABLE EN LOS DOS COMPONENTES PRINCIPALES CONSIDERADOS PARA EL CLUSTERING DE LOS INVESTIGADORES .....	144
FIGURA 5.19 CONJUNTOS BORROSOS DEL PC1 DE LOS INVESTIGADORES .....	147
FIGURA 5.20 CONJUNTOS BORROSOS DEL PC2 DE LOS INVESTIGADORES .....	147
FIGURA 5.21 CONJUNTOS BORROSOS DE LA PRODUCCIÓN TOTAL DE LOS INVESTIGADORES.....	148

# 1 INTRODUCCIÓN

## 1.1 ENTORNO DE LA INVESTIGACIÓN

La tecnología ha crecido a pasos agigantados desde la década anterior. Diferentes disciplinas gozan de los beneficios y facilidades que los avances tecnológicos brindan, y efectivamente, la ciencia no es la excepción. La investigación científica se ha vuelto más accesible gracias a los repositorios digitales, los cuales se encargan de almacenar, preservar y ser una herramienta de difusión de contenido y además, son entidades de naturaleza digital o no digital, y pueden ser utilizados, reutilizados o referenciados a lo largo del aprendizaje apoyado por la tecnología (Ip, Morrison y Currie, 2001). Existe la necesidad de describir y localizar los recursos de un repositorio, por lo que se hace uso de los metadatos. Éstos son datos descriptivos sobre los datos, es decir, proporcionan la información mínima necesaria para identificar un recurso (Senso y Rosa Piñero, 2003). Entre otras cualidades de los metadatos se encuentran: el incremento al acceso, disminución de tráfico en la red, expansión de la utilización de la información, control de versiones, aspectos legales y precisión en los procesos de búsqueda y recuperación (Agudelo, 2009).

En México, diversas instituciones cuentan con uno o varios repositorios digitales. El repositorio nacional es una plataforma digital la cual contiene más de 80 repositorios institucionales y proporciona acceso abierto en texto completo a una amplia gama de recursos de conocimiento académico, científico y tecnológico. Entre los materiales a consultar se encuentran artículos de revistas científicas, tesis de instituciones de educación superior (licenciatura, maestría y doctorado), protocolos de investigación, memorias de congresos y patentes, es decir, documentos académicos/científicos producidos en la República Mexicana con fondos públicos (CONACYT, 2017m). Otro caso de acceso abierto en México es la Red Mexicana de Repositorios Institucionales (REMEDI), compuesta por más de 60 instituciones de educación superior y 100 repositorios. La REMEDI se enfoca en crear una red de interconexión de repositorios digitales de instituciones de educación superior del país para la integración,

difusión, preservación y visibilidad de la producción científica, académica y documental generada en dichas instituciones y además, anexarse a redes o directorios de repositorios internacionales para el fomento del trabajo mutuo y apoyo al acceso y difusión de contenidos de acceso abierto (REMERI, 2017b).

La problemática que se ha encontrado en común con los aspectos mencionados anteriormente es la siguiente: un investigador genera investigaciones científicas (*productos*, desde el enfoque de los repositorios digitales), las cuales son descritas por medio de los metadatos, elementos que permiten al repositorio la localización de cada recurso. Si los metadatos no se encuentran llenados de forma adecuada, es decir, no cumplen con la calidad necesaria, pueden surgir dos anomalías: (1) el repositorio no contará con la capacidad de intercambiar y compartir datos con otros repositorios que contienen al recurso que se desea localizar, es decir, no contará con la interoperabilidad, o (2) la referencia al producto en cuestión se hará erróneamente. Estos dos aspectos dan lugar a estas consecuencias: no sería posible calcular de forma adecuada por medio de indicadores cuantitativos la relevancia de un investigador a través de sus investigaciones científicas, afectando su nivel de producción, el cual, al no ser el correcto, puede ocasionar que un investigador no fuera distinguido adecuadamente como investigador de calidad, tal es el caso en México del Sistema Nacional de Investigadores (SNI) y del Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP), los cuales cuentan con criterios cualitativos propios para determinar la calidad y relevancia en los investigadores.

## 1.2 HIPÓTESIS Y OBJETIVO

En esta subsección se presenta la hipótesis, el objetivo general y los objetivos específicos. Para ello, se plantea la siguiente pregunta de investigación:

**¿Es posible valorar la relevancia de un investigador a través de su producción científica almacenada en repositorios digitales y de los metadatos asociados a la misma, aún cuando éstos tienen inconsistencias?**

Se pretende probar que a través de una herramienta tecnológica es posible medir la relevancia de un investigador científico tomando como base sus investigaciones científicas, las cuales se consultan en uno o más repositorios digitales, aun cuando los metadatos pueden adolecer de

- Integridad
- Completitud
- Exactitud
- Consistencia

Tomando en cuenta la información disponible, se plantea la siguiente hipótesis:

**La relevancia de un investigador puede medirse mediante una herramienta tecnológica que considere la producción almacenada en uno o más repositorios digitales; aún cuando los metadatos asociados a dicha producción posiblemente adolecen de integridad, completitud, exactitud y consistencia.**

A partir de la hipótesis que se ha planteado y del estudio de los temas relacionados a la problemática, se presenta el objetivo general del trabajo.

**Desarrollar una solución tecnológica que valore la relevancia de un investigador a través de la evaluación de los metadatos de su producción científica, los cuales se pueden consultar en uno o más repositorios digitales.**

Este objetivo general se detalla a través de los siguientes objetivos específicos:

1. Realizar una revisión de las herramientas existentes que den a conocer los indicadores cuantitativos estadísticos de la producción científica de un investigador.
2. Establecer las herramientas y técnicas que permitan recuperar la producción científica almacenada en diversos repositorios.

3. Utilizar técnicas de extracción de datos que permitan obtener los nombres de los investigadores y su producción desde los repositorios digitales, aún cuando sus registros incluyan inconsistencias.
4. Definir métricas para la evaluación de los indicadores propuestos para la relevancia científica.
5. Implementar una solución tecnológica que calcule por indicadores cuantitativos estadísticos la relevancia de la producción académica y científica de los investigadores.
6. Analizar experimentalmente el funcionamiento de la solución tecnológica con el Repositorio Digital de la Universidad Autónoma de Yucatán.

### 1.3 METODOLOGÍA DE LA INVESTIGACIÓN

Se seguirá la metodología de la Investigación en acción, la cual, en síntesis, es única, ya que asocia la investigación con la práctica, y la investigación informa la práctica y la práctica se encarga de informar la investigación de modo cooperativo (Avison, Lau, Myers y Nielsen, 1999). Las fases a seguir son:

1. **Definir el escenario de la problemática y análisis del estado de arte.** Se refiere a mejorar la recolección de la información acerca de la producción académica y científica de un profesor almacenada en repositorios digitales. Esto origina la necesidad de definir una herramienta tecnológica que permita por medio de los metadatos asociados a los recursos generados por un académico, medir la producción científica por medio de indicadores cuantitativos estadísticos. Se realiza el análisis del estado del arte, el cual consiste en hacer una revisión sistemática de las temáticas afines a los repositorios digitales, los metadatos (incluyendo su calidad y estándares), la interoperabilidad entre repositorios, los problemas existentes en las comunidades científicas en relación con los metadatos, así como los indicadores cuantitativos estadísticos para medir la producción

científica, además del análisis de las evaluaciones a los investigadores por parte de las instancias, las cuales requieren de los datos de producción científica.

2. **Proponer las herramientas y las métricas que ayudarán a la solución de la problemática planteada.** Consiste en definir el modelado de la solución tecnológica por medio de herramientas especializadas, establecer las herramientas de software que permiten realizar de manera eficaz y persistente la solución planteada, así como la definición de las métricas para la evaluación del desempeño de la misma.
3. **Implementar un prototipo para solucionar la problemática.** A partir del modelado, desarrollar una herramienta de software que permita obtener los metadatos de toda la producción científica de cada profesor en cuestión para su posterior análisis y obtención de los indicadores de producción, así como diseñar la interfaz de usuario que presenta los resultados de forma comprensible al usuario.
4. **Realización de pruebas.** Por medio de las métricas establecidas en la fase 2, se realizan estudios que determinan el nivel de desempeño de la herramienta. Se evalúa la efectividad y la eficiencia del uso de la herramienta en escenarios reales.
5. **Documentación y difusión.** Se comparten los resultados obtenidos en las fases anteriores a la comunidad investigadora por medio de la presentación de trabajos en revistas científicas y en foros nacionales e internacionales.

#### 1.4 TRABAJOS RELACIONADOS

Google Scholar (o Google Académico en español) es un buscador que, a diferencia de Google tradicional, está enfocado en la literatura del ámbito científico. En el buscador se pueden consultar títulos y autores. Los autores pueden crear su perfil donde, entre otros datos, mencionan la afiliación a la que pertenecen, áreas de interés, etcétera. De hecho, Google Scholar permite a cualquier usuario visualizar los perfiles de los autores, donde se pueden observar las citas, el Índice h, Índice i10 y los coautores con los que ha trabajado (Google, 2017).

En México existe el repositorio nacional, el cual se define como una plataforma digital encargada de proporcionar acceso abierto (es decir, sin requerimientos de suscripción, registro o pago) en texto completo a una amplia variedad de recursos de información de tipo académica, científica y tecnológica. Permite realizar búsqueda simple y avanzada. La segunda puede ser parametrizada o por área de conocimiento. Sin embargo, por medio de las búsquedas, únicamente presenta los productos del autor, el tipo de producto (artículo, patente, tesis, etc.) y el área de conocimiento al que pertenece cada producto (CONACYT, 2017). No presenta indicadores de citas, como Google Scholar.

Por su parte, la REMERI, la cual interconecta los repositorios digitales de instituciones de educación superior en México, cuenta con un buscador, el cual localiza productos en las colecciones. Las búsquedas únicamente permiten visualizar productos por su tipo (tesis, artículo, etcétera), título, autor, adscripción y la URL del producto. No permite visualizar indicadores cuantitativos del autor, ni algún otro dato relevante (REMERI, 2017b).

Scopus es una base de datos de citas y resúmenes de literatura enfocada a la investigación (Elsevier, 2019). Su buscador permite buscar por datos del autor (apellido, nombre y afiliación) (Elsevier, 2018b) o bien, por ORCID, el cual es un identificador único para cada autor (ORCID, 2017). Cada perfil de autor de Scopus presenta los indicadores *h-index*, total de documentos del autor y total de citas, además de una gráfica de la tendencia entre el número de publicaciones y de citas por año (Figura 1.1) (Elsevier, 2018b).

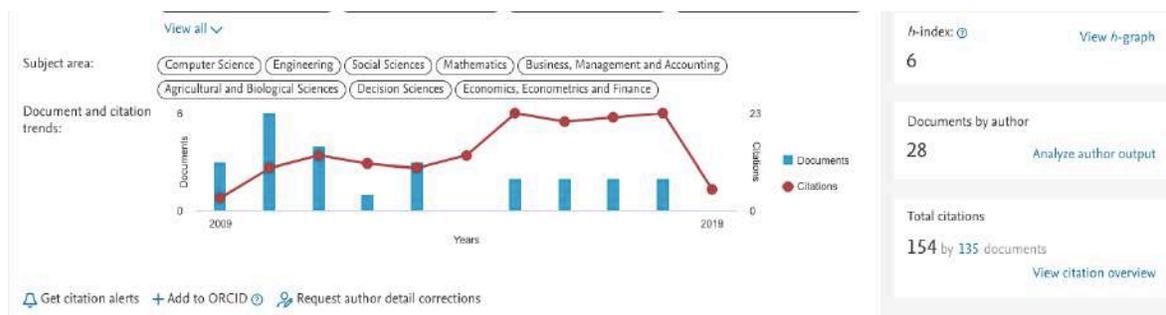


Figura 1.1 Indicadores presentados por Scopus en los perfiles de autores

Respecto con las revistas científicas y libros indizados por Scopus, se presentan los indicadores y se presenta un ejemplo de su visualización en Scopus en la Figura 1.2:

- CiteScore: Promedio de citas recibidas en un año determinado.
- SJR: Mide las citas ponderadas recibidas, involucrando el prestigio de la publicación y el campo de conocimiento, en un año determinado.
- SNIP: Mide las citas reales recibidas, a comparación del número esperado, en un año determinado.
- CiteScore rank: Posicionamiento en el ranking de los campos de conocimiento al que pertenece la publicación.



Figura 1.2 Indicadores presentados por Scopus en las revistas y libros serializados

En esta propuesta se plantea la creación de un modelo y una herramienta tecnológica que permita calcular desde los productos y sus metadatos, los indicadores cuantitativos estadísticos de un investigador y de comunidades científicas, los cuales se puedan obtener por lapsos, tipos de investigación (artículos, ponencias, tesis, entre otros). También será posible identificar a cada autor una única vez, ya que los metadatos tienden a ser inconsistentes.

## 1.5 MARCO DE TRABAJO

El presente trabajo de tesis es un proyecto del grupo académico de Ingeniería de Software e Informática Educativa, que tiene como objetivo la investigación e innovación de los procesos de desarrollo, operación y mantenimiento de software, así como asistir los procesos de enseñanza y de aprendizaje mediante aplicaciones educativas. Las líneas de investigación son:

- Ingeniería de Software. Su planteamiento se dirige al estudio y aplicación de enfoques sistemáticos, disciplinados y cuantificables para el desarrollo, operación y mantenimiento de Software.
- Informática Educativa. Su enfoque es el estudio, uso, efectos y consecuencias de las Tecnologías de la Información en el proceso educativo.

Esta investigación ha contribuido al cumplimiento de los objetivos en distintos proyectos de la Universidad Autónoma de Yucatán (UADY). A continuación se presentan y describen brevemente los proyectos en los cuales este trabajo ha tenido relevancia.

- **Proyecto SISTPROY FMAT-2018-0014. Generación de indicadores de productividad de cuerpos académicos y profesores.** El proyecto se orientó a la creación de una herramienta tecnológica que determine la calidad de un académico a través de la evaluación de los metadatos de su producción académica y científica a partir de repositorios digitales por medio de los indicadores cuantitativos estadísticos y de esta forma obtener datos relevantes que ayuden a definir si el académico puede pertenecer a alguna instancia que lo reconozca como investigador de calidad, así como su nivel de producción interna (colaboraciones con académicos de la misma institución) y externa (colaboraciones con académicos de otras instituciones).
- **Proyecto SISTPROY FMAT-2016-0017.** Creación del repositorio digital institucional de la UADY. La finalidad del proyecto fue el desarrollo e implementación de un repositorio digital para la UADY, de acceso abierto, poblado mediante un proceso de importación que recupera e integra la producción académica, científica, tecnológica y de innovación de la UADY. Su infraestructura se basa en estándares de almacenamiento,

búsqueda y recuperación, buscando favorecer la interconexión del repositorio con aplicaciones externas y su integración al repositorio nacional del CONACYT. La información que se encuentra en el repositorio está disponible para su consulta por medio de un portal Web, ofreciendo servicios de búsqueda y recuperación. Se utilizaron tecnologías estándares de distribución libre que garantizan la portabilidad y el nulo costo en su uso.

## 1.6 APORTACIONES

La consecución de este trabajo de investigación se refleja en una serie de aportaciones relativas a la medición de la relevancia de un investigador a través de su producción científica almacenada en diversos repositorios, que están ligadas con los objetivos que se propusieron para esta tesis. Las contribuciones se presentan a continuación, ennumeradas en dos grupos: aportaciones teóricas, en la forma de modelos, métricas y metodologías; y aportaciones prácticas, dadas mediante componentes de software desarrollados para la implementación de las aportaciones teóricas.

### 1.6.1 APORTACIONES TEÓRICAS

- Una serie de indicadores para medir la relevancia de los cuerpos académicos y de los investigadores con base en su producción almacenada en diversos repositorios, considerando el impacto de la investigación, el grado de colaboración y la importancia de las publicaciones de acuerdo con las instancias evaluadoras mexicanas.
- Una metodología para el cálculo de los indicadores de relevancia científica.
- Un modelo para la selección de perfiles de autor en un repositorio de publicaciones de alto impacto.
- Un modelo para la selección y extracción de los metadatos de las publicaciones asociadas a un investigador aún cuando los metadatos contengan inconsistencias.

## 1.6.2 APORTACIONES PRÁCTICAS

- Un componente de software que consulta, extrae y depura los metadatos de la producción asociada a un investigador almacenada en un repositorio de publicaciones de alto impacto.
- Un componente de software que cosecha por medio del protocolo OAI-PMH, extrae y depura la producción asociada a un investigador almacenada en diversos repositorios de acceso abierto.
- Un componente de software que calcula los diversos indicadores para la relevancia científica con base en la producción recuperada y almacenada.
- Un componente de software que se encarga de localizar y presentar las diversas colaboraciones internas de los cuerpos académicos y de los investigadores.
- Un componente de software que prepara el formato para la visualización por medio de mapas de árboles de los valores de los diversos indicadores.
- Un componente de software que almacena todos los datos en un modelo ontológico.
- Un componente de software que realiza consultas al modelo ontológico.

## 1.7 ORGANIZACIÓN DEL DOCUMENTO

A continuación, se presentan los capítulos contenidos en este trabajo de tesis.

**Capítulo 1. Introducción.** Se define el entorno de la investigación, la hipótesis, así como los objetivos. Se establece la metodología de la investigación y se presentan trabajos relacionados al tema de la tesis.

**Capítulo 2. Estado del arte.** Presenta las temáticas afines para la comprensión del desarrollo del trabajo de tesis: repositorios digitales, metadatos (estándares, calidad y problemas), calidad en las investigaciones científicas, repositorios institucionales de producción científica y las normativas para la Ciencia Abierta en México.

**Capítulo 3. Modelo de indicadores para la relevancia científica.** Contiene una propuesta de indicadores cuantitativos estadísticos que permiten medir la relevancia de la producción grupal

(cuerpos académicos) considerando las publicaciones indizadas, el grado de colaboración y las publicaciones no indizadas; así como la relevancia individual de un investigador, tomando en cuenta el grado de colaboración en su cuerpo académico, sus colaboraciones externas, sus publicaciones individuales y producción no indizada.

**Capítulo 4. Sistema generador de indicadores de relevancia científica.** Implementa la metodología descrita en el capítulo 3. Se describe la aplicación de software que mide la relevancia científica desde la perspectiva de las Ciencias de la Computación.

**Capítulo 5. Experimentación.** Describe los experimentos realizados para validar la propuesta presentada en el trabajo de tesis. Se realizan diversos análisis de correlación entre los indicadores propuestos desde diversas perspectivas en una facultad de la UADY. También se estudia el estado actual de la investigación en dicha universidad haciendo uso de técnicas de la minería de datos. Finalmente, se analizan y describen los resultados obtenidos.

**Capítulo 6. Conclusiones y trabajo futuro.** Contiene una descripción de las conclusiones y del trabajo futuro a realizar. Se presentan las publicaciones y contribuciones derivadas de esta investigación.

## 2 ESTADO DEL ARTE

Este capítulo tiene como finalidad dar a conocer los antecedentes y temáticas afines al trabajo de tesis. Primero se describen los repositorios digitales, características y tendencias. Luego, se definen los metadatos, estándares y problemas relacionados. Posteriormente se abordan los aspectos relacionados con la relevancia de las publicaciones científicas. También se presentan las instancias evaluadoras de la productividad científica en México. Luego se mencionan algunos repositorios institucionales de acceso abierto, las normativas para la Ciencia Abierta en México y finalmente, una descripción de los distintos tipos de cuerpos académicos.

### 2.1 REPOSITORIO DIGITAL

#### 2.1.1 DEFINICIÓN

Durante este siglo, tanto el aprendizaje móvil como a distancia se han convertido en una necesidad gracias a que las personas se han vuelto más ansiosas por obtener conocimiento en cualquier instante o lugar gracias a los avances de Internet y los dispositivos electrónicos asociados (Yen, Shih, Chao y Jin, 2010), dando lugar al e-learning, el cual, es el proceso de formación a distancia con base en el uso de las tecnología de la información y las telecomunicaciones, haciendo posible un aprendizaje interactivo, flexible y accesible (Azcorra, Bernardos, Gallego y Soto, 2001). Desde la perspectiva del e-learning, las tecnologías incluyen herramientas de autoría, sistemas de administración de aprendizaje y los repositorios digitales (Yen et al., 2010). Un repositorio digital es una plataforma que se encarga de almacenar, preservar y ser una herramienta de difusión de contenido y éstos pueden clasificarse de la siguiente manera: como Recursos Educativos Abiertos (REA), como referencias de REA, como iniciativas de OpenCourseware y como sistemas de gestión del aprendizaje (Ochoa y Duval, 2009). Los repositorios son una entidad de naturaleza digital o no digital, y pueden ser utilizados, reutilizados o referenciados a lo largo del aprendizaje apoyado por la tecnología (Ip et al., 2001). De igual manera, se entiende como repositorio digital a un depósito de archivos digitales, los cuales son de diferentes clasificaciones y pueden ser accedidos, difundidos y preservados (Texier, De Giusti, Oviedo, Villarreal y Lira, 2012).

### 2.1.2 CARACTERÍSTICAS

Los repositorios no deben considerarse únicamente como sitios de almacenamiento y divulgación, sino que se consideran ciertas características a cumplir (Wolf Iszaevich, Uriarte Santillán, Galina Russell y Miranda Quevedo, 2014), y se clasifican de la siguiente manera:

- **Estándares de catalogación.** Para poder describir a cada uno de los recursos de un repositorio, se utiliza un estándar de metadatos.
- **Protocolos para la interoperabilidad.** Es gracias a los metadatos que la información se organiza de manera consistente. Y para que un repositorio pueda compartir la información de sus recursos con otros, se emplean protocolos para la interoperabilidad.
- **Estabilidad de identificadores.** Es la capacidad de la independencia de la localización de los recursos con el sitio web, siendo que los primeros se almacenan en el repositorio, y el sitio web únicamente enlaza a la dirección correspondiente. De esta manera, la reestructuración de los enlaces del sitio web no afecta a los recursos.

De igual forma, existen otras características, las cuales pueden variar de una institución a otra (Keefer, 2008):

- La finalidad de su creación
- El material admitido
- Los formatos mantenidos
- Procesos aplicados para la evaluación y aceptación de los recursos
- Control de acceso
- Duración de la preservación de los recursos

### 2.1.3 HERRAMIENTAS PARA EL DESARROLLO DE REPOSITORIOS

Para el desarrollo de un repositorio de recursos digitales abiertos, no es necesario crear todos los componentes de autoría propia. Existen diversas herramientas de software que facilitan esta tarea, siendo el beneficio la incorporación de diversos módulos que facilitan las tareas básicas. Algunas de las más conocidas se presentan a continuación.

- **DSpace.** Esta herramienta es de código abierto, facilitando el uso, modificación e incorporación del código en una aplicación comercial de forma gratuita (DuraSpace,

2018). Entre los formatos de archivo que puede administrar se encuentran PDF, archivos de Word, JPEG, MPEG y TIFF, además de proporcionar un registro de formato de archivo simple para el registro de extensiones de archivo no reconocidos para uso futuro. Entre sus características se encuentran la arquitectura de la aplicación (base de datos, administrador de almacenamiento y una interfaz de usuario), flujos de trabajo incorporados, motor de búsqueda, esquema de metadatos basado en Dublin Qualified, recuperación de errores, entre otras. Es utilizado por instituciones educativas, gubernamentales, privadas y comerciales.

- **EPrints.** Herramienta de código abierto. La idea original era el almacenamiento de artículos científicos, tesis y material educativo, aunque actualmente acepta cualquier tipo de documento. Utiliza el esquema de metadatos Dublin Core, con la posibilidad de exportar a otros formatos (ejemplo: METS y MODS). Para la interoperabilidad, utiliza los protocolos OAI-PMH y SWORD. EPrints permite exportar los metadatos y la estructura de los directorios del repositorio en formato XML, así como generar las estadísticas de la actividad en el repositorio, como descargas. En lo que respecta a la búsqueda, EPrints puede buscar en los campos de metadatos, con operadores booleanos y admite el uso del motor Xapian, el cual se encarga de presentar los resultados por orden de relevancia (Castagné, 2013).
- **Digital Commons.** Herramienta para la gestión de revistas, artículos de conferencias y revistas. Únicamente utiliza el esquema de metadatos Dublin Core. En lo que respecta a la interoperabilidad, utiliza el protocolo OAI-PMH pero sólo como proveedor de datos. Permite la generación de estadísticas, importa en cualquier formato de archivo, cuenta con un interfaz amigable para el usuario, y su motor de búsqueda permite la indexación de texto completo, además de realizar búsquedas de forma interinstitucional, las cuales consisten en localizar recursos en las instituciones que utilizan Digital Commons (Castagné, 2013).

#### 2.1.4 INTEROPERABILIDAD

Es la capacidad de intercambio de contenido entre dos o más sistemas. Se ha analizado este concepto desde cuatro aspectos fundamentales: el sintáctico, el semántico, el de infraestructura y el estructural (Gómez Dueñas, 2010), siendo el último el más importante en este trabajo, ya que se refiere a la comunicación entre repositorios heterogéneos por medio del uso de protocolos especializados, de los cuales, algunos se mencionan a continuación:

- **OAI-PMH.** Es un mecanismo de baja barrera para la interoperabilidad del repositorio (NSF, 2018). Los proveedores de datos son repositorios que exponen metadatos estructurados a través de OAI- PMH. Dichos proveedores de servicios son los encargados de la realización de solicitudes de servicio OAI- PMH para recolectar metadatos. Prácticamente, es un conjunto de seis verbos (servicios) que son invocados dentro de HTTP, los cuales pueden hacer uso de argumentos requeridos, opcionales o exclusivos. OAI-PMH utiliza seis verbos (Nelson y Warner, 2015), los cuales se enlistan a continuación con su respectiva descripción y argumentos necesarios, es decir, parámetros que permiten realizar especificaciones a las consultas.
  1. *GetRecord.* Recupera un registro de metadatos de forma individual de un repositorio. Argumentos: *identifier* (requerido) y *metadataPrefix* (requerido).
  2. *Identify.* Recupera la información de un repositorio. Los repositorios pueden utilizar este verbo para devolver información descriptiva adicional. Argumentos: ninguno.
  3. *ListIdentifiers.* Forma abreviada del verbo ListRecords. Recupera únicamente los encabezados, en vez de los registros. Argumentos: *from* (opcional), *until* (opcional), *metadataPrefix* (requerido), *set* (opcional) y *resumptionToken* (exclusivo).
  4. *ListMetadataFormats.* Recupera los formatos de metadatos disponibles de un repositorio. Argumentos: *identifier* (opcional).
  5. *ListSets.* Recupera la estructura establecida de un repositorio. Argumentos: *resumptionToken* (exclusivo).
  6. *ListRecords.* Recupera la información de cada uno de los recursos de un repositorio. Argumentos: *from* (opcional), *until* (opcional), *set* (opcional), *resumptionToken* (exclusivo) y *metadataPrefix* (requerido).

Cada argumento tiene una función específica (Nelson y Warner, 2015), y se presentan en la siguiente lista:

- *identifier*. Especifica un identificador único asociado a un recurso que se encuentra en el repositorio
  - *metadataPrefix*. Especifica el prefijo de metadatos del formato que debe incluirse en la parte de metadatos del registro devuelto. Los formatos de metadatos se pueden recuperar con el verbo *ListMetadataFormats*
  - *from*. Ayuda a especificar un límite inferior para la recolección selectiva con base en la fecha
  - *until*. Especifica un límite superior para la cosecha selectiva basada en la fecha
  - *set*. Especifica criterios de cosecha selectiva basada en conjuntos (*sets*). Éstos pueden recuperarse con el verbo *ListSets*
  - *resumptionToken*. Cuenta con un valor que es el token de control de flujo devuelto por una solicitud hecha con anterioridad que dio como resultado una lista incompleta
- 
- **OpenAIRE**. Es un proyecto que incita el uso de la política de acceso abierto en el continente europeo, brindando los medios para la promoción y realización de la adopción generalizada de las políticas de acceso abierto (RECOLECTA, 2018). Los objetivos de OpenAIRE son, en resumen, la construcción de estructuras de apoyo a los investigadores para el depósito de las publicaciones de investigación, establecimiento de un portal web para el acceso a la información científica y fomentar el depósito en acceso abierto de toda la producción científica. Para la identificación de proyectos, financiadores, publicaciones referenciadas y conjuntos de datos, OpenAIRE se basa en la sintaxis de Dublin Core, la cual toma forma de URI y se define como el espacio de nombres *info:eu-repo*. El uso de los campos en OpenAIRE son, de acuerdo con OpenAIRE (2015): Obligatorio (M), Obligatorio cuando sea aplicable (MA), Recomendado (R), Opcional (O) (RECOLECTA, 2018).
  
  - **IMS Digital Repositories Interoperability**. Estándar creado por el IMS Global Learning Consortium. Su finalidad es facilitar el acceso a los contenidos en repositorios

que cuenten con contexto educativo, pero igualmente con otro tipo de sistemas, como los sitios web de búsqueda (López Guzmán, 2005). Es una propuesta para la interoperabilidad entre aplicaciones (igualmente servicios) que cuenten con las funciones de búsqueda, exposición, colecta, envío, almacenamiento, pedido, entrega y alerta. Las actividades principales de este estándar son: Buscar/Exponer, Colectar/Exponer, Enviar/Almacenar, Pedir/Entregar y Alertar/Exponer.

#### 2.1.5 IMPORTANCIA DE LOS REPOSITARIOS DE DOCUMENTACIÓN CIENTÍFICA

En la actualidad, las universidades son un pilar importante para el crecimiento de la sociedad, debido a las dos tareas principales que realizan: formar recursos humanos de calidad y la promoción de las nuevas tecnologías, y bajo esta idea, las universidades deben establecer y utilizar prácticas para el impulso de la innovación industrial, mejorando la gestión universitaria tradicional bajo tres premisas: gestión de la calidad, gestión de la información y gestión del conocimiento, siendo la tercera la principal (Vergara González, 2008). Se entiende por gestión del conocimiento a la colaboración en la identificación de problemas locales y específicos que requieren del conocimiento para su solución y la identificación de las personas e instituciones que pueden realizar verdaderos aportes para la creación de redes de conocimiento (Núñez Jover, Félix Montalvo y Pérez Ones, 2006).

Por otra parte, el conocimiento se divide en dos: el conocimiento explícito, que es aquel que se encuentra a disposición de las personas a través de diversos medios, enfocado en la teoría; y el conocimiento tácito, que se genera por medio de la práctica, aún siendo de forma inconsciente (E. A. Smith, 2001). El modelo SECI (Socialización, Externalización, Combinación, Internalización) considera ambos tipos de conocimiento, abarcando la transferencia del conocimiento explícito a las personas y del conocimiento tácito de una persona a varias personas; además de la transformación del conocimiento tácito en conocimiento explícito (Nonaka y Takeuchi, 1995).

Existe modelo para el aprendizaje organizacional, el cual consiste en cuatro premisas (Crossan, Lane y White, 1999):

1. El aprendizaje conlleva una relación entre la asimilación del nuevo aprendizaje y el uso de lo que se ha aprendido
2. El aprendizaje organizacional se encuentra en tres ámbitos: individuo, grupo y organización
3. Dichos ámbitos se relacionan por medio de la intuición, interpretación, integración e institucionalización
4. La cognición afecta a la acción, y viceversa

Los dos modelos mencionados anteriormente representan el proceso de conversión de conocimiento más utilizado en las universidades tanto para el proceso de enseñanza y en la investigación (López G., Cabrales G. y Schmal S., 2005). Sin embargo, se considera que la transferencia del conocimiento relacionado con la información de publicaciones científicas representa un problema cuando éstas son de acceso restringido y es necesario realizar un pago o pertenecer a una institución afiliada a la editorial en cuestión. Como se considera que la gestión del conocimiento se relaciona fuertemente con los sistemas de información y las Tecnologías de la Información, y más para las universidades, principalmente por los dos motivos expuestos al principio de esta sección, forma parte de las buenas prácticas tecnológicas hacer uso de repositorios de documentos que recopilen los trabajos de investigación científica, tecnológica, académica y de innovación producida en la institución y por las personas adscritas, considerando profesores y estudiantes, es decir, la comunidad universitaria.

La tendencia de los repositorios institucionales es el libre acceso, es decir, que cualquier persona pueda acceder a sus recursos sin necesidad de licencias o permisos adicionales. De esta manera, la utilidad y el impacto de los trabajos puede crecer de forma exponencial, comparado al uso y acceso por unas cuantas personas. Para esto es necesario adoptar el esquema de acceso abierto, o ciencia abierta, que consiste en dos elementos, donde se elige alguno (Vergara González, 2008):

1. El *camino de oro*, donde las revistas permiten el acceso libre a sus publicaciones
2. El *camino verde*, que consiste en el permiso otorgado por el autor para acceder libremente a sus publicaciones

Gran parte de estos repositorios utilizan el protocolo OAI-PMH, facilitando la interconexión entre diversos sistemas y con ello, la creación de comunidades de colaboración y el aprovechamiento de los resultados de trabajos de investigación gracias a la existencia de una mayor divulgación de la ciencia y como consecuencia, mayor desarrollo en los países, por lo que las instituciones correspondientes pueden asignar recursos para la realización de más proyectos de investigación considerando que éstos tienen un gran impacto y uso.

#### 2.1.6 SISTEMAS CRIS

Una tendencia actual que va a la par con el concepto de repositorio digital, pero que no debe confundirse con él, es el uso de los Sistemas de Información de Investigación Actual (CRIS, por sus siglas en inglés, *Current Research Information System*). Los sistemas CRIS consisten en gestionar la información relevante de la investigación en una institución, abarcando las oportunidades de financiamiento, redacción de publicaciones, presentación de propuestas, ofertas para su posterior conversión en proyectos, así como las subvenciones, personas, organizaciones, productos e instalaciones y equipos de investigación (Gottschling, 2017). Es decir, los sistemas CRIS se encargan de ver por cada uno de los procesos involucrados con la investigación, e incluso, a través de la interoperabilidad, puede conectarse directamente a un repositorio.

La diferencia entre un repositorio y un sistema CRIS se basa en el enfoque con el que tratan a los documentos: mientras que el primero considera a los documentos como los elementos que se encarga de divulgar, el segundo los visualiza como el resultado de un conjunto de procesos que conlleva la investigación (De Castro, Shearer y Summann, 2014). Una buena práctica para las instituciones que disponen de un repositorio es utilizar el sistema CRIS y el repositorio de manera independiente, pero consistente. Es decir, que los valores de los metadatos de los recursos se describan desde el sistema CRIS y el recurso se aloje en el repositorio. Respecto con los estándares de metadatos de sistemas CRIS, generalmente son complejos, a comparación de los utilizados por los repositorios (Gottschling, 2017). Uno de estos estándares es el Formato Europeo Común de Información de Investigación (*Common European Research Information Format*, CERIF) que conceptualiza, describe y formaliza las entidades de la investigación (proyectos, personas, equipos, financiamiento, etcétera) y sus relaciones (IOSR, 2019). A pesar

de eso, CERIF se ha convertido en el estándar más extendido, permitiendo la interoperabilidad con repositorios y las instituciones encargadas de financiar proyectos de investigación (Gottschling, 2017).

## 2.2 METADATOS

### 2.2.1 DEFINICIÓN

Un metadato es un dato con una estructura que está por encima de la información, es decir, son los datos que describen a los datos (a analizar) (Lamarca Lapuente, 2013). Otra definición de metadato es la siguiente: es un registro que consiste de información estructurada acerca de un recurso; se puede definir como información sobre información o datos sobre datos. Se estructura de una manera que facilita la administración, descubrimiento y recuperación de los recursos (Al-Khalifa y Davis, 2006). El uso de los metadatos debe considerarse primordial para el acceso de nuevos conocimientos, ya que éstos van creciendo de forma exponencial año tras año, por lo que no utilizar metadatos causaría una búsqueda ineficiente de información, y por consecuencia, falta de aprovechamiento del conocimiento que se encuentra al alcance de cualquier persona (J. R. Smith y Schirling, 2006).

### 2.2.2 ESTÁNDARES

Para el uso y asignación de valores a los metadatos se debe seguir un esquema, esto con la finalidad de conseguir la interoperabilidad, es decir, que varios sistemas sean capaces de intercambiar contenido entre ellos (COAR, 2011). Cada esquema de metadatos va de acuerdo al tipo de contenido. Algunos de los más populares que están siendo desarrollados como estándares se presentan en la Tabla 2.1.

Tabla 2.1 Estándares de metadatos

Estándar	Características
DDI	Estándar internacional para la descripción de datos producidos por estudios y métodos de observación en el ámbito social, psicológico y ciencias de la salud (DDI, 2017).
IEEE LOM	Para la descripción de REA, los cuales son una entidad digital o no, que es utilizado para el ámbito educacional (Barker, 2005).
Dublin Core	Estándar de propósito general, independiente de los dominios informativos. Es adaptable tanto para la web y otros esquemas de metadatos (E. Méndez Rodríguez, 2006).
ONIX	Estándar internacional para la representación y difusión de libros electrónicos (Editeur, 2009).
ISO/IEC 19506	Estándar para la representación de datos de software, asociaciones y entornos de operación. La interoperabilidad, junto con el intercambio de información son dos de los fines de este estándar (ISO, 2012).
RDF	Su uso es para la descripción de recursos en la web para conformar la interoperabilidad, haciéndolo posible por medio de XML (E. M. Méndez Rodríguez, 1999).
EML	Enfocada a la ecología. Está basada en los trabajos previos de la Ecological Society of America. Facilita la documentación de investigación de campo por medio de archivos XML (KNB, 2017).

### 2.2.3 DUBLIN CORE

El estándar de metadatos más utilizado para la descripción de cualquier documento digital es Dublin Core (DC). Incluso se ha adoptado en varios países en sus respectivos idiomas (incluyendo a nivel gubernamental), así como el trabajo en conjunto con las iniciativas de acceso a los archivos abiertos, mejor conocidos como OAI (E. Méndez Rodríguez, 2006).

A continuación se presenta una descripción detallada de Dublin Core y sus características (Prieto Méndez, Menéndez-Domínguez y Vidal Castro, 2014). El estándar Dublin Core tiene como objetivos:

- Ser fácil de crear y mantener
- Permitir un entendimiento común en la semántica de los metadatos
- Considerar un ámbito internacional para permitir una mejor representación de la información considerando su naturaleza multilingüe y multicultural
- Extender el conjunto base según las necesidades particulares mediante perfiles de aplicación para una mejor representación de las necesidades de cierta comunidad o dominio del conocimiento

El estándar DC define un conjunto base de 15 metadatos que constituyen el nivel simple. Estos metadatos describen el contenido del documento, la propiedad intelectual y la información para la creación de instancias.

El nivel cualificado incorpora tres nuevos metadatos (*audience*, *provenance* y *rightHolder*) al nivel simple. También define un conjunto de atributos que clarifican la representación semántica de los recursos descritos para una fácil recuperación. Estas propiedades refinan las descripciones de metadatos y están limitadas a ciertos elementos, pudiendo utilizar vocabularios asociados o descripciones en lenguaje natural. Los atributos pueden ser de dos tipos:

1. **Elementos de refinamiento.** Estos atributos aclaran el significado de un elemento. Por ejemplo, el metadato *title* puede refinarse con el elemento *alternative* para indicar un subtítulo.
2. **Esquemas de codificación.** Estos atributos definen esquemas que permiten una mejor interpretación del valor del elemento. Generalmente, estos esquemas representan vocabularios controlados y notaciones formales (como la ISO 3166 para la representación de nombres de países con dos caracteres).

Dublin Core cuenta con un conjunto de propiedades para describir contenido (DCMI, 2005). Los 15 principales se enlistan a continuación:

- *Title* (Título)
- *Subject* (para palabras clave o frases que clasifican el tema del que trata el recurso)

- *Description* (Puede ser el resumen del recurso, tablas de contenido, referencia a una representación gráfica, etcétera)
- *Type* (Tipo del recurso, por ejemplo, imagen, sonido, texto)
- *Source* (Origen del recurso)
- *Relation* (Para la referencia de un recurso relacionado)
- *Coverage* (El alcance del contenido del recurso, en términos geográficos o de lapsos de tiempo)
- *Creator* (El o los responsables de la creación del recurso)
- *Publisher* (La entidad responsable de poner a disposición el recurso)
- *Contributor* (Entidad responsable de realizar aportaciones al recurso)
- *Rights* (Para derechos de autor)
- *Date* (Fechas asociadas a eventos de la vida del recurso)
- *Format* (Especificación del formato físico o digital)
- *Identifier* (Identificador del recurso, por ejemplo, DOI, ISSN, ISBN)
- *Language* (Idioma en que se encuentra el contenido del recurso)
- *Audience* (Audiencia para el que se dirige el recurso)

En Dublin Core, cada elemento es opcional y puede repetirse. No existe un orden específico de aparición. Su contenido puede regirse por esquemas de codificación, vocabularios controlados o estar abierto a las descripciones de texto en lenguaje natural.

El estándar proporciona un conjunto de directrices para su implementación en los formatos XHTML, XML y RDF, lo que permite la interoperabilidad de los recursos.

#### 2.2.4 CALIDAD DE LOS METADATOS

Los metadatos requieren calidad para que su labor sea aprovechada al máximo y no sean sólo datos que describen, pero sin trascendencia alguna. Por eso es requerida la calidad. En metadatos, tenemos dos tipos de calidad (Palavitsinis, 2013):

1. **Calidad de los metadatos.** Hace hincapié a la evaluación de desempeño y garantía de que los metadatos cumplen con la calidad que se pretende, o hasta qué nivel
2. **Calidad en los metadatos.** Se refiere a la calidad de los recursos que se describen

La más importante para esta sección es la calidad de los metadatos, la cual evalúa los siguientes aspectos (Myrseth, Stang y Dalberg, 2011; Palavitsinis, 2013):

- Datos completos
- Integridad
- Exactitud
- Consistencia
- Oportunidades (respecto al aprovechamiento en tiempo)

La calidad de los metadatos es un terreno que no ha sido explotado, pero las investigaciones respecto al tema continúan al día de hoy. Para lograr la calidad de los metadatos es vital evaluar la etapa de creación de los metadatos (asignar valores a los metadatos) (Park, 2009) la cual, al tener resultados negativos, sufre consecuencias, como la recuperación y precisión deficientes, resultados fuera de contexto, incluso hasta ambigüedades (Barton, Currier y Hey, 2003). Existen metadatos con valores que no corresponden al objeto a describir, causando un incumplimiento en los aspectos correspondientes a los datos completos, integridad, exactitud y consistencia de los metadatos. La calidad de los metadatos es un aspecto descuidado a pesar de que se han creado estándares como Dublin Core, ONIX, entre otros.

Se ha demostrado que el llenado de los metadatos presenta ineficiencias, siendo estos incompletos e inconsistentes y clasifica los factores que influyen en la calidad de los metadatos de la siguiente manera (Brasher y McAndrew, 2004):

- **Motivación de los productores.** Se refiere a la utilidad que tienen los metadatos para los que los producen
- **Precisión.** Similar a como se ha expuesto anteriormente en este artículo. Si los metadatos describen completamente y de acuerdo a lo establecido por los diseñadores de los sistemas que utilizarán al máximo los metadatos
- **Consistencia.** Se cuestiona si los metadatos llenados por diferentes personas serán interpretados de la misma forma por los sistemas que harán uso de ellos

La creación de los metadatos, al ser lenta y dada la amplia flexibilidad de los estándares de metadatos da lugar a las inconsistencias en el llenado de metadatos. Existe un estudio (Sicilia,

García, Pages, Martínez y Gutiérrez, 2005) sobre la integridad de los metadatos de objetos de aprendizaje a partir de muestras obtenidas de los repositorios MERLOT y CAREO, analizando desde el punto de vista de la estructura inadecuada de los metadatos, utilizando el estándar IEEE-LOM como marco de referencia, teniendo como resultados los siguientes aspectos:

- Únicamente el 4.7% de los REA analizados ha pasado por un proceso de revisión de calidad para su posterior publicación. La cantidad de información de los metadatos es baja. El 60% de los REA analizados se describe únicamente con datos básicos
- Incoherencias en contenido
- Los metadatos, al ser la mayoría no estructurados, requieren una comprensión humana, lo cual impide a los sistemas que explotan a los metadatos realizar su trabajo eficientemente

En resumen, es necesario comprender la función de los metadatos, conocer qué es lo que van a describir, y de acuerdo a ello, elegir el esquema de metadatos adecuado para permitir la interoperabilidad entre repositorios. Independientemente del esquema elegido, los metadatos siempre deben ser completos, íntegros, exactos y consistentes para que se puedan considerar de alta calidad.

#### 2.2.5 PROBLEMAS EXISTENTES CON LOS METADATOS

Algo que caracteriza a la ciencia contemporánea (a partir de mediados del siglo XX) es la constante colaboración entre científicos de conocimientos homogéneos y heterogéneos, siendo una ventaja para complementar conceptos y concebir otros. A este grupo de personas se le conoce como comunidad científica, y no necesariamente tienen que conocerse frente a frente.

Medios como el teléfono e Internet ayudan a la interacción entre ellos, facilitando el trabajo colaborativo. Incluso no hay investigador que su forma de trabajo más conveniente sea la individual, siendo de vital importancia las alianzas con otros especialistas para lograr metas, debido a que, en muchas situaciones, el trabajo colaborativo es el único camino para lograrlo (González Alcaide y Gómez Ferri, 2014). Los investigadores generan diversos tipos de artículos científicos, entre ellos destacan (Peh y Ng, 2008):

- Artículos originales
- Reporte de caso
- Nota técnica
- Ensayo pictórico
- Reseña científica
- Comentario
- Carta al editor
- Editorial
- Material no científico
- Otros

Estos tipos de documentos científicos se pueden encontrar en uno o varios repositorios digitales, los cuales son propiedad de instituciones donde el investigador se encuentra adscrito. Como se mencionó anteriormente, para localizar estos recursos es necesario hacer una búsqueda. Los repositorios permiten hacer búsquedas simples o avanzadas. Pero desafortunadamente no todos los recursos tienen los valores correctos en los campos de sus metadatos. De aquí surge el problema con las comunidades científicas: los valores incorrectos en los metadatos hacen en muchas ocasiones nula la colaboración de una comunidad científica. Los problemas más comunes con el llenado de los metadatos, los cuales afectan a los artículos producidos por las comunidades científicas (Barton et al., 2003) se presentan en

- La ortografía, así como el uso de abreviaturas
- El campo de autor y los relacionados a las colaboraciones
- Título
- Tema (respecto con las palabras clave y la clasificación)
- Fecha (respecto al formato)

Otros problemas con los metadatos que pueden perjudicar a las investigaciones de las comunidades científicas son los relacionados con los metadatos de derechos (copyright). Es posible que instituciones no quieran saber acerca de este tipo de metadatos ya que lo consideran complejo o es algo donde tienen que invertir. Una idea colectiva acerca de este tipo de derechos es que sólo es para cuestiones de copyright y no es necesario, pero no se está considerando que

la importancia de proteger los derechos es para mayor responsabilidad de las obras y/o publicaciones en un repositorio digital. En esta era de las Tecnología de la Información, preocuparse por los derechos legales debe ser primordial. Al no utilizar este tipo de metadatos, las publicaciones de las comunidades científicas son más vulnerables. Un ejemplo simple es: si un investigador utiliza información que no es propia (existe en otra publicación), no la cita y la hace pública, puede no reconocerse el valor de la investigación original y los autores no pueden proceder legalmente (Whalen, 2008).

En resumen, las comunidades científicas producen investigaciones a través de diversos tipos de documentos. Éstos se pueden encontrar en repositorios, los cuales hacen uso de los metadatos para localizar dichos productos (artículos). Pero el uso de abreviaturas en los datos descriptivos del producto, no mencionar a todos los autores que trabajaron en colaboración, etc., son inconsistencias en los metadatos que afectan a los niveles de producción de las comunidades científicas y al momento de realizar estadísticas sobre la producción científica por comunidad o de forma individual, los datos obtenidos no serán confiables.

## 2.3 RELEVANCIA EN LAS INVESTIGACIONES CIENTÍFICAS

### 2.3.1 PUBLICACIONES INDIZADAS

Son todas aquellas publicaciones que se pueden localizar en una base de datos que se encarga de recopilar las citas, incluso a través de evaluaciones periódicas. Utilizan indicadores cuantitativos estadísticos para medir la productividad científica y con ellos determinan la calidad de las publicaciones, generalmente revistas científicas y libros serializados. Internacionalmente, las principales son dos: Scopus y Web Of Science. Sin embargo, se ha considerado Google Scholar como una alternativa más debido al uso de sus propios indicadores para la productividad científica.

Existen estudios que comparan características, fortalezas y debilidades de bases de datos de citas, utilizando como referencia Scopus, Web of Science y Google Scholar (Falagas, Pitsouni, Malietzis y Pappas, 2007) (Bakkalbasi, Bauer, Glover y Wang, 2006) (Jacso, 2005) (Harzing y Alakangas, 2016). Sin embargo, las dos primeras tienen mayor credibilidad para las

instituciones para considerar la calidad y relevancia de las publicaciones de sus investigadores evaluados, tal es el caso del CONACYT, que para el ingreso al SNI considera de mayor peso las publicaciones que se encuentran indizadas en Web of Science (CONACYT, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g).

Por otra parte, existen otras instituciones que indexan publicaciones, tal es el caso del CONACYT con su Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología, dividido en 8 categorías (CONACYT, 2016):

1. Física, Matemáticas y Ciencias de la Tierra
2. Biología y Química
3. Medicina y Ciencias de la Salud
4. Humanidades y Ciencias de la Conducta
5. Ciencias Sociales
6. Biotecnología y Ciencias Agropecuarias
7. Ingenierías
8. Multidisciplinarias

Otras bases de datos de citas, pero enfocadas a Latinoamérica y a hispanohablantes son:

- **Redalyc.** Red de Revistas Científicas de América Latina y el Caribe, España y Portugal. Contiene más de 1,302 revistas científicas, 48,356 fascículos y 625,671 artículos a texto completo (UAEM, 2019)
- **SciELO.** Modelo para la publicación electrónica en colaboración entre países de desarrollo, enfocado en Latinoamérica y el Caribe (SciELO, 2018)
- **Latindex.** Contiene revistas de investigación científica editadas en Latinoamérica, el Caribe, España y Portugal. Contiene más de 1,508,784 artículos indizados (UNAM, 2018)

### 2.3.2 INDICADORES CUANTITATIVOS PARA LA PRODUCTIVIDAD CIENTÍFICA

Existen tres indicadores para la calidad de la investigación científica, los cuales son los más comunes.

1. **Factor de Impacto de la Revista (FIR).** En inglés se le conoce como *Journal Impact Factors* (Conocida como *JIF*, pero en este capítulo se nombrará como *FIR*). Consiste en la proporción de citas en el año actual referentes a los artículos publicados en la revista en los últimos dos años dividido entre el número de artículos publicados en los mismos dos años. Toma en cuenta las citas de los artículos que se publican en una revista. La finalidad de *FIR* es elevar el nivel de calidad de la investigación científica, y por consecuencia, el de la revista a evaluar (Sharma, 2012).
2. **Citas.** El reconocimiento de un trabajo en otro es satisfactorio para el autor y/o comunidad científica porque eleva el nivel de calidad siendo positivo para el esfuerzo y tiempo invertido. Pero no ser citado tiene consecuencias negativas. Se puede entender que la investigación que desea ser citada y no lo es tiene una calidad baja o no es comprensible. Sharma (2012) recomienda hacer uso de las siguientes fórmulas para calcular el rendimiento del trabajo de investigación de una revista científica. Sin embargo, son aplicables para calcular el rendimiento en una comunidad científica (Sharma, 2012).

$$\sum_{i=1}^N FIR_i$$

es la suma de todos los *FIR* de todas las publicaciones.

$$\sum_{i=1}^N Cn_i$$

es la suma de todas las citas que se han hecho de las publicaciones de la comunidad científica o revista.

$$\sum_{i=1}^N FIR_i(1 + Cn_i)$$

es el Impacto de la Investigación (RI por sus siglas en inglés), donde  $N$  es el número total de publicaciones de la comunidad científica o revista,  $FIR_i$  representa el factor de impacto de la revista que contiene la  $i$ -ésima publicación, mientras que  $Cn_i$  corresponde al número total de citas de la  $i$ -ésima publicación.

3. **H index.** Es el índice más importante para investigadores y personas dedicadas a la ciencia. Aplica para la evaluación individual de un investigador, más que de una comunidad científica. Un investigador tiene como índice el valor  $h$  si  $h$  de sus  $N_p$  trabajos cuenta con, por los menos,  $h$  citas. Los trabajos  $N_p - h$  cuenta con  $h$  citas como máximo (Bornmann y Marx, 2011).

Además del *FIR*, es necesario hacer uso del Factor de Prestigio (*FP*), el cual no incluye las citas que vienen de los artículos de revisiones (Buela-Casal, 2003). Para calcularla, se sigue la siguiente fórmula:

$$FP = \frac{NTCR}{NAOP}$$

Es decir, divide el número total de citas que reciben en un año los artículos de contenido original publicados en una revista en el año en curso y en los dos anteriores (*NTCR*) entre el número de artículos originales publicados en la misma revista en los mismos años (*NAOP*). Los resultados se reflejan en un rango entre 0 y 1.000. El factor de prestigio puede utilizar más de una base de datos, cada una recoge más de seis mil revistas y éstas, a su vez, están clasificadas en casi mil disciplinas.

El factor de impacto y el de prestigio no son comparables, ya que cada una utiliza rangos de años diferentes. Para solucionar este problema en cuanto a la calidad de los artículos científicos y la revista que lo publica, Buela-Casal (2003) propone el uso de los siguientes indicadores.

- **Factor de Impacto Medio de las Revistas donde se producen las Citas (*FIMRC*).** Es el factor de impacto ponderado de las revistas que citan a la revista en cuestión (Buela-Casal, 2003). Se obtiene a través de la siguiente fórmula.

$$FIMRC = \frac{(FIR * NAC)_1 + (FIR * NAC)_2 + \dots + (FIR * NAC)_i}{NTAC}$$

La fórmula consiste en la suma de la multiplicación del factor de impacto de ese año de cada revista las cuales han citado artículos de los años anteriores de la revista analizada (*FIR*) por el número de artículos que han sido citados en cada revista (*NAC*). El resultado

se divide por el número total de artículos que han sido citados (*NTAC*). *i* es el número de revistas a analizar. Este factor permite conocer si las citas a una revista vienen de otras con un impacto mayor.

- **Factor de Impacto Ponderado (*FIP*)**. Se obtiene con una fórmula simple, la cual consiste en la suma del *FIMRC* y el *FIR* de una revista dividido entre 2, es decir, un promedio (Buela-Casal, 2003).

$$FIP = \frac{FIMRC + FIR}{2}$$

Calcular el *FIP* permite conocer un dato concreto sobre las citas que una revista recibe y el factor de impacto que tienen las revistas que realizan la cita.

- **Factor de Prestigio Medio de las Revistas donde se producen las Citas (*FPMRC*)**. Se calcula a través de la siguiente fórmula (Buela-Casal, 2003).

$$FPMRC = \frac{(FP * NAC)_1 + (FP * NAC)_2 + \dots + (FP * NAC)_i}{NTAC}$$

Se obtiene por medio de la suma de la multiplicación del *FP* de cada revista las cuales citan artículos del año actual y anteriores de la revista a analizar por el número de artículos citados en cada revista (*NAC*). Una vez obtenido el resultado, se divide entre el número total de artículos citados (*NTAC*). *i* es el número de revistas a analizar.

- **Factor de Prestigio Ponderado (*FPP*)**. Se obtiene a través del promedio del *FP* de una revista y el *FPMRC* (Buela-Casal, 2003).

$$FPP = \frac{FPMRC + FP}{2}$$

- **Porcentaje de Interacción Parcial de Citas (*PIPC*).** Ayuda a permitir el impacto que tiene una revista sobre otra por medio de sus citas (Buela-Casal, 2003). Se calcula por medio del total de citas que contiene una o más revistas en un año y calculando el porcentaje en función del número de citas recibidas por esa o esas revistas en el mismo año. Para facilitar la comprensión, se presenta un ejemplo.

Citas de la revista 1 en el año 2016: 150

Citas de la revista 2 en el año 2016: 1853

Citas de la revista 1 en la revista 2: 41

PIPC de revista 1 sobre revista 2: 2.21%

- **Porcentaje de Interacción Mutua de Citas (*PIMC*).** Se refiere al porcentaje de los artículos citados entre dos revistas, sin considerar las auto- citas (Buela-Casal, 2003). Para calcular el *PIMC* se realiza obteniendo el total de citas comunes entre dos o más revistas en un año y calculando el porcentaje en relación del número de citas que se han recibido en el mismo año por las revistas evaluadas (Buela-Casal, 2003).

En resumen, los indicadores cuantitativos estadísticos de la calidad de las investigaciones científicas son datos que generan impacto en el investigador, y si para obtenerlos se analiza su producción almacenada en un repositorio digital, es necesario que los metadatos tengan los valores correctos que describan adecuadamente los productos del investigador, ya que de eso depende que un repositorio pueda realizar referencias adecuadamente. Puede suceder que las investigaciones realmente sean de calidad, pero si los metadatos son incorrectos, al momento de realizar un estudio sobre la calidad de las investigaciones, los resultados no serán favorables y no se aprovechará el contenido de los productos, reduciendo drásticamente el índice de citación del investigador.

#### 2.4 INSTANCIAS EVALUADORAS DE LA PRODUCTIVIDAD CIENTÍFICA

A continuación, se presentan los lineamientos o criterios para medir la calidad de un investigador de acuerdo a su producción científica. Primero, se menciona lo establecido por el

Sistema Nacional de Investigadores, posteriormente el Programa para el Desarrollo Profesional Docente para el Tipo Superior, siendo ambas, instancias para México. Finalmente, se menciona brevemente cómo se mide la calidad del investigador en otros países.

#### 2.4.1 SISTEMA NACIONAL DE INVESTIGADORES

En México existe el Sistema Nacional de Investigadores (SNI), el cual fue creado para reconocer los diversos trabajos hechos por personas dedicadas al ámbito científico y tecnológico. Incluso existe un título llamado investigador de calidad, el cual incluye beneficios económicos, según el nivel alcanzado. Pueden ser miembros las personas que se dedican a la investigación científica y tecnológica de forma habitual. Deben presentar los trabajos completamente documentados. Entre los requisitos que destacan para pertenecer al SNI se encuentran (CONACYT, 2017n; DOF, 2017):

- Verificar por medio de documentación oficial que la persona presta por lo menos 20 horas a la semana a la investigación tecnológica y científica
- Realizar investigación tecnológica y científica en dependencias, entidades, instituciones de educación superior o centros de investigador en el sector público, privado o social de la República Mexicana
- Desempeñarse en México, sin importar la nacionalidad. Puede ser mexicano realizando actividades de ciencia y tecnología en el extranjero

Existen tres distinciones que un miembro del SNI puede obtener. Cada uno debe cumplir con los lineamientos establecidos por el SNI y tener el grado de doctor (CONACYT, 2017n).

- **Candidato a Investigador Nacional.** Debe demostrar la capacidad para realizar investigación en ámbitos científicos o tecnológicos a través de sus productos generados en investigación o desarrollo tecnológico
- **Investigador Nacional.** Se divide en tres niveles:
  1. Nivel I. Debe contar con material original de calidad de investigación científica o tecnológica, así como contar con participación en la dirección de tesis de licenciatura

- o posgrado, o haber impartido asignaturas, así como haber participado en otras actividades docentes. Es importante que apoye a la divulgación de la investigación
2. Nivel II. Además de lo necesario para pertenecer al nivel I, se requiere que demuestre su participación en colaboración con otros investigadores por medio de material de investigación original y de calidad para demostrar desarrollo en alguna línea de investigación, así como su participación en la dirección de tesis de posgrado y formación de recursos humanos de alta calidad
  3. Nivel III. Además de lo necesario para pertenecer al nivel II, es necesario que el investigador tenga en su historial de investigaciones alguna que cause impacto en la actualidad, realizar actividades de liderazgo nacional en lo que respecta a ciencia o tecnología y tener reconocimientos nacionales e internacionales por su labor
- **Investigador Nacional Emérito.** Para pertenecer a esta distinción es necesario que, al cierre de la convocatoria, el candidato tenga 65 años. Debe contar con por lo menos 15 años y tres evaluaciones ininterrumpidas obteniendo la distinción Investigador Nacional Nivel III (DOF, 2017)

Cada investigador pertenece a un área de conocimiento.

- Área I: Físico- Matemáticas y Ciencias de la tierra. Abarca a todos aquellos que estudian Astronomía, Ciencias de materiales, Física, Matemáticas, Ciencias de la tierra y Ciencias del mar, en sus vertientes básica y aplicada (CONACYT, 2017a)
- Área II: Biología, Química y Ciencias de la vida (CONACYT, 2017b)
- Área III: Medicina y Ciencias de la salud. Abarca a todos aquellos que estudian Ciencias de la salud, Farmacia, Medicina y Odontología (CONACYT, 2017c)
- Área IV: Humanidades y Ciencias de la conducta. Incluye las siguientes disciplinas: Antropología, Antropología física, Arqueología, Artes y Letras, Bibliotecología, Diseño,

Filosofía, Historia, Historia del Arte, Lingüística, Matemáticas educativas y enseñanza de las ciencias, Pedagogía y Psicología (CONACYT, 2017d)

- Área V: Ciencias Sociales. Abarca a todos aquellos que estudian Administración, Ciencia Política y Administración Pública, Comunicación, Contabilidad, Demografía, Derecho y Jurisprudencia, Economía, Geografía Humana, Sociología y Prospectiva (CONACYT, 2017e)
- Área VI: Biotecnología y Ciencias Agropecuarias. Comprende a los investigadores cuyos trabajos se ubican en Acuicultura, Agronomía, Alimentos, Biotecnología, Veterinaria y Zootecnia en sus vertientes básica y aplicada (CONACYT, 2017f)
- Área VII: Ingenierías. Abarca a todos aquellos que estudian la Ingeniería en sus diferentes disciplinas entre otras: Aeronáutica, Ambiental, Civil, de Comunicaciones, Electrónica y Control, Eléctrica, Computación, Industrial de Materiales, Marina y Portuaria, Mecánica, Minera, Nuclear, Petrolera, Química y Textil (CONACYT, 2017g)

El CONACYT establece los criterios específicos de evaluación para ser investigador en todas de las distinciones del SNI en cada una de las áreas de conocimiento. En la Tabla 2.2 se presenta una comparación, de elaboración propia, entre el contenido de cada uno de los criterios por área de conocimiento.

Tabla 2.2 Comparación de los criterios de evaluación de cada área del SNI

	Área 1	Área 2	Área 3	Área 4	Área 5	Área 6	Área 7
Objetivo general	No	No	No	Sí	Sí	No	No
Marco general de la evaluación	Sí	Sí (no especificado con este título)	Sí	Sí	Sí	Sí	Sí
Del procedimiento de evaluación	Sí	Sí	Sí	No	No	Sí	Sí
Elementos de la evaluación 1. Productos de investigación 2. Impacto 3. Criterios, características y calidad de los productos 4. Participación en actividades educativas 5. Liderazgo 6. Divulgación de la Ciencia 7. Infraestructura	2, 7 (Se encuentran en la siguiente sección, no en ésta)	1, 2, 4, 5, 6	1, 2, 3, 4	Todas	1, 4, 5, 6, 7	1, 2, 3, 4, 5, 6,	1 (añade desarrollo tecnológico), 2, 5, 7
Requisitos de ingreso y reingreso al SNI por nivel	Nuevo ingreso, reingreso vigente y reingreso no vigente	Nuevo ingreso y reingreso vigente	Ingreso y reingreso	Ingreso y reingreso no vigente y reingreso vigente	Ingreso y reingreso vigente (con prórroga para candidatos)	Ingreso, reingreso vigente (con prórroga para candidatos) y reingreso no vigente	Candidato y reingreso vigente
Criterios cualitativos	No	No	No	No	No	No	Sí
Actividades de plusvalía	No	No	No	No	No	No	Sí

Se puede observar que no se especifican todos los lineamientos en cada una de las áreas, lo cual representa un problema para las personas que aspiran ser miembros del SNI, e incluso para los que ya pertenecen, porque no conocen exactamente qué se requiere de forma cuantitativa para lograr una distinción de tal tipo.

El CONACYT pone a disposición un enlace en el sitio oficial del Sistema Integrado de Información sobre Investigación Científica y Tecnológica (CONACYT, 2017i), el cual direcciona a las estadísticas de datos públicos, entre ellos el número de investigadores SNI (CONACYT, 2017k). A partir de esos datos se puede observar el incremento en la investigación en todos los estados de la República Mexicana. Un ejemplo es el presentado en la Figura 2.1 (CONACYT, 2017k), donde se puede observar el incremento de investigadores en cada una de las áreas de conocimiento del SNI desde 1991 a 2014. Se puede observar que hasta este último año el área con mayor número de investigadores era Biología y Química (Área 2), mientras que la de menor número de investigadores era la que pertenece a Ingenierías (Área 7).

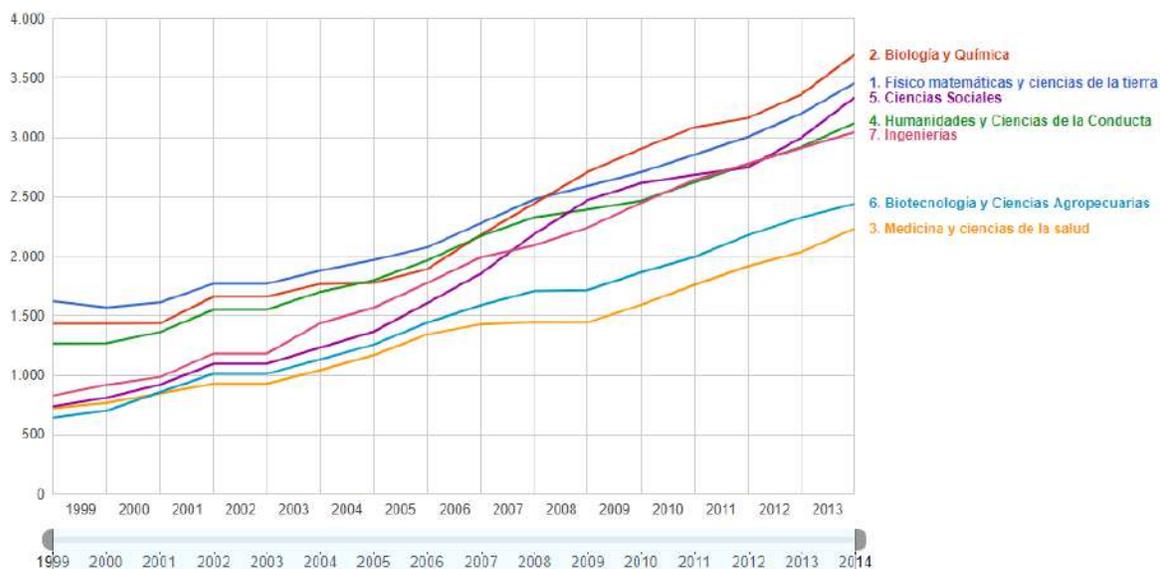


Figura 2.1 Estadística por área del número de investigadores pertenecientes al SNI

#### 2.4.2 PROGRAMA PARA EL DESARROLLO PROFESIONAL DOCENTE PARA EL TIPO SUPERIOR

Además del SNI, existe el Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP). Éste tiene como objetivo profesionalizar a los Profesores de Tiempo Completo (PTC) para que formen cuerpos académicos y realicen actividades de investigación-docencia y desarrollo de la tecnología e innovación haciendo uso de la responsabilidad social. Actualmente cuenta con 730 instituciones de Educación Superior en su cobertura, incluyendo universidades politécnicas, universidades tecnológicas, institutos tecnológicos federales, entre otros (DGESU, 2017).

El PRODEP mide la calidad de sus PTC por medio del perfil deseable. Si cumple con los siguientes aspectos, se considera un PTC de calidad (DGESU, 2017).

- Tener nombramiento de PTC
- Contar con doctorado (preferente) o maestría (mínimo). En caso de ser médico, alguna especialidad clínica
- Haber impartido un curso frente a grupo al año, durante los tres años anteriores a la fecha de presentar solicitud para pertenecer al PRODEP o durante el tiempo que ha

transcurrido desde el primer nombramiento como PTC, o bien, desde que obtuvo su último grado

- Comprobar un producto de investigación de buena calidad al año durante los últimos tres años anteriores a presentar solicitud para pertenecer al PRODEP o durante el tiempo que ha transcurrido desde el primer nombramiento como PTC o bien, desde que obtuvo su último grado
- Haber impartido tutorías durante el último año inmediato a la fecha de presentación de solicitud para pertenecer al PRODEP. Las tutorías deben ir dirigidas a estudiantes, grupos o dirigir al menos una tesis
- Durante el último año anterior a la presentación de solicitud, haber participado en cuerpos colegiales formales, comisiones para el diseño, evaluación y operación de programas educativos y planes de estudio, comisiones para evaluar proyectos relacionados con la investigación, vinculación o difusión; así como en la dirección, coordinación y supervisión de programas relacionados con la educación, investigación, vinculación o difusión

PRODEP tiene registrados más de 26710 Profesores de Tiempo Completo en todo el país, distribuidos en Universidades Públicas Estatales (UPE), UPE de Apoyo Solidario, Institutos de Educación Superior (IES) Federales, Universidades Politécnicas, Universidades Tecnológicas, Institutos Tecnológicos Federales, Escuelas Normales, Institutos Tecnológicos Descentralizados y Universidades Interculturales, siendo un programa eficiente para crear investigación y formar a través de la docencia a las futuras generaciones de científicos e investigadores (PRODEP, 2017).

En resumen, las instancias mexicanas expuestas, SNI y PRODEP son opciones que los investigadores tienen para conseguir una distinción por su labor y que su investigación sea aprovechada por la comunidad en general. Cada instancia cuenta con sus propios indicadores o lineamientos para determinar la calidad del investigador y es necesario que sean específicos y objetivos para que todo aquel que aspire a una distinción conozca si cumple con lo mínimo necesario, basándose, en parte, en la calidad de sus investigaciones científicas.

### 2.4.3 OTROS PAÍSES

A diferencia de los indicadores para la definición de la calidad de las investigaciones científicas que se pueden aplicar en cualquier país y en cualquier disciplina científica, los indicadores para definir la calidad del investigador varían en cada país y los establecen diversas instancias dedicadas a la ciencia. La Agencia Nacional de Evaluación de la Calidad y la Acreditación dispone enlaces a los criterios de evaluación para diversas disciplinas de los investigadores en España a través de la Agencia Nacional de Evaluación de la Calidad y Acreditación, siendo indicadores cuantitativos y objetivos (ANECA, 2017).

Estados Unidos cuenta con la National Science Foundation, la cual igualmente establece sus indicadores de calidad para sus investigadores (NSF, 2017).

Los indicadores de la calidad del investigador, tanto en España, como en Estados Unidos, son cuantitativos y específicos, mientras que las instancias mexicanas disponen de indicadores cualitativos, los cuales no son del todo claros.

## 2.5 REPOSITARIOS INSTITUCIONALES DE ACCESO ABIERTO

### 2.5.1 SCOPUS

Es una base de datos de citas y resúmenes de literatura enfocada a la investigación. Cuenta con más de 70 millones de recursos, 70 mil perfiles de instituciones y 16 millones de perfiles de autores (Elsevier, 2019). Su buscador permite buscar por datos del autor (apellido, nombre y afiliación) (Elsevier, 2018b) o bien, por ORCID, el cual es un identificador único para cada autor (ORCID, 2017). Entre las características particulares que Scopus ofrece se encuentran (Elsevier, 2018a):

- Ser una guía para investigadores para identificar y analizar revistas donde pueden enviar sus publicaciones
- Conocer a través de sus indicadores la productividad de los investigadores
- Localizar a los autores potenciales para realizar colaboraciones
- Seguimiento de las investigaciones de impacto
- Localizar temas existentes en el ámbito científico

Diversas instituciones y organizaciones utilizan Scopus, entre ellas se encuentran instituciones académicas, gobierno y agencias de financiación, centros de investigación y desarrollo, y organizaciones de clasificación (Elsevier, 2018c). Además, Scopus cuenta con diversas APIs que permiten al usuario, a través de una clave personal, recuperar información acerca de instituciones, autores, recursos, citas y otros datos relacionados (Elsevier, 2017).

### 2.5.2 REPOSITORIO NACIONAL

En México existe el repositorio nacional, el cual se define como una plataforma digital encargada de proporcionar acceso abierto (es decir, sin requerimientos de suscripción, registro o pago) en texto completo a una amplia variedad de recursos de información de tipo académica, científica y tecnológica (CONACYT, 2017m).

Entre los materiales que se pueden consultar se encuentran artículos de revistas científicas, tesis elaboradas en instituciones de educación superior, protocolos de investigación, memorias de congresos y patentes, al igual que otros documentos producidos en la República Mexicana con el soporte de fondos públicos.

El repositorio nacional, está compuesto por más de 80 repositorios institucionales, 61,405 recursos de información y 3,942,826 consultas. El repositorio nacional permite realizar búsqueda simple y avanzada. La segunda puede ser parametrizada o por área de conocimiento.

La parametrizada puede contener:

- Palabras
- Título
- Autor
- Colaborador
- Materia
- Tipo
- Editor
- Nivel de acceso
- Repositorio
- Fecha inicial

Por área de conocimiento se refiere a las siguientes:

- Biología y Química
- Ciencias Agropecuarias y Biotecnología
- Ciencias Físico- Matemáticas y Ciencias de la tierra
- Ciencias Sociales.
- Humanidades y Ciencias de la conducta
- Ingeniería y Tecnología
- Medicina y Ciencias de la salud

En cuanto a consultar los datos de un autor, el repositorio digital, por medio de las búsquedas, únicamente presenta los productos del autor, el tipo de producto (artículo, patente, tesis, etc.) y al área de conocimiento que pertenece cada producto (CONACYT, 2017l). No presenta indicadores de citas.

De acuerdo con la Declaración de protección de los Derechos de Autor y otros Derechos de Propiedad Intelectual, toda persona que consulte el repositorio nacional podrá realizar tareas sobre el contenido e información, como analizar, imprimir y utilizar, todo de forma gratuita y libre, siempre y cuando se siga lo establecido en la Ley de Ciencia y Tecnología en materia de Ciencia Abierta, a excepción de los contenidos que expresen lo contrario, o el contenido no pueda ser accedido de forma abierta.

Se prohíbe utilizar los contenidos del repositorio nacional con fines comerciales o de lucro sin el previo consentimiento del titular o fuera de los casos en que los tratados, las leyes o las disposiciones así lo consientan (CONACYT, 2017h). La información disponible en el repositorio nacional se puede recuperar a través del protocolo para la interoperabilidad OAI-PMH.

### 2.5.3 REPOSITARIOS INSTITUCIONALES

Tienen como finalidad el acopio, preservación, gestión y acceso electrónico de información y contenidos de calidad, incluyendo aquellos de interés social y cultural que produzcan las instituciones, centros de investigación y demás integrantes del Sistema Nacional de Ciencia,

Tecnología e Innovación, que hayan sido financiadas de forma total o parcial con recursos públicos o bien, hayan utilizado infraestructura pública (CONACYT, 2014). Existen lineamientos técnicos para integrar los repositorios institucionales al repositorio nacional. Entre los beneficios se encuentran (CONACYT, 2014):

- Mayor visibilidad de los trabajos de investigación financiados con recursos públicos
- Eliminación de barreras físicas, económicas y legales al facilitar el acceso
- Aprovechamiento de la ciencia, acelerando los procesos afines
- Proporcionar el material para la realización de minería de datos y de texto para la generación de nuevos resultados a partir de lo existente

Los repositorios institucionales pueden cosechar sus metadatos por medio del protocolo para la interoperabilidad OAI-PMH.

#### 2.5.4 REMERI

Por su parte, la Red Mexicana de Repositorios Institucionales (REMEDI), la cual interconecta los repositorios digitales de instituciones de educación superior en México, cuenta con un buscador, el cual localiza productos en las colecciones. Las búsquedas únicamente permiten visualizar productos con su tipo (tesis, artículo, etc.), título, autor, adscripción y la URL del producto. No permite visualizar indicadores cuantitativos del autor, ni algún otro dato relevante (REMEDI, 2017b).

Para la integración de más repositorios a la REMEDI se utiliza un servidor de metadatos en el estándar OAI-PMH. Utiliza el sistema de consultas Indix, mismo que devuelve los resultados ordenados por relevancia de los términos o bien, con la aplicación de los filtros correspondientes, se pueden ordenar por fecha, institución, autor, entre otros (REMEDI, 2017a).

#### 2.5.5 REDALYC

Es la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal. Cuenta con revistas científicas, fascículos y artículos a texto completo. Pertenece a la Universidad Autónoma del Estado de México (UAEM, 2019). Es descrito como una hemeroteca científica y un sistema de información en línea de acceso libre con la finalidad de apoyar el análisis de la

producción, la difusión y el consumo de la literatura científica. Sigue la filosofía del acceso abierto. Entre sus principales características se encuentran (UAEM, 2017):

- Uso de su propia metodología para la incorporación de revistas, basada en la calidad editorial internacional y revisión por pares
- Contiene publicaciones provenientes de Latinoamérica, el Caribe, España y Portugal. Sin embargo, considera a todas aquellas que incluyan trabajos de investigación acerca de las zonas geográficas mencionadas
- Las publicaciones pueden localizarse en otros sitios, como Google, Google Scholar, DOAJ y JournalTOCS
- Sus dos principales áreas del conocimiento (divididas en sub áreas) son las Ciencias Naturales y las Ciencias Sociales

#### 2.5.6 REDI-UADY

Es el Repositorio Digital Institucional de la Universidad Autónoma de Yucatán (UADY). Sigue la filosofía de los repositorios institucionales, de dar acceso abierto en texto completo a diversos recursos de información científica, académica y tecnológica.

Entre sus colecciones se encuentran artículos de revistas científicas, tesis, protocolos de investigación, entre otros (UADY, 2017a). Para su creación, se integró un grupo multidisciplinario de investigadores relacionados con el área, mismos que llevaron a cabo las siguientes actividades:

- Implementación del repositorio institucional, mismo que utiliza el protocolo para la interoperabilidad OAI-PMH
- Uso de una arquitectura de tres capas: interfaz de uso, comportamiento del sistema y las operaciones para la gestión de la información
- Programación de una colección de herramientas, encargadas de facilitar el proceso de importación de los datos de los recursos, utilizando tres sistemas diferentes de la UADY:

1. SISBIUADY. Es el Sistema Bibliotecario de la UADY. Permite consultar los metadatos de los recursos almacenados de manera física en las bibliotecas de los diversos campos de conocimiento de la UADY.
  2. SIGA. Sistema de Informes y Gestión de los Académicos. cual tiene como fin recolectar las evidencias de la producción o constancias de la actividad académica por medio de archivos electrónicos (UADY, 2017b).
  3. SISTPROY. Sistema de Registro de Proyectos de Investigación.
- Implementación de una herramienta para la curación de metadatos, con la finalidad de asegurar que los metadatos de los recursos del repositorio sean los correctos
  - Poblar el repositorio con la información existente en los diversos sistemas de la UADY
  - Definir la infraestructura tecnológica que permita conectar a REDI-UADY con el repositorio nacional, utilizando el protocolo OAI-PMH
  - Impulsar el acceso abierto a los recursos, producto de diversos trabajos de enfoque científico, tecnológico, académico y de innovación

Los tres sistemas mencionados anteriormente cuentan con sus respectivas bases de datos. Cada uno de ellos contiene información estructurada de forma diferente. El middleware se encarga de recolectar, procesar y transferir al repositorio los metadatos necesarios después de ser validados y reestructurados (Figura 2.2).

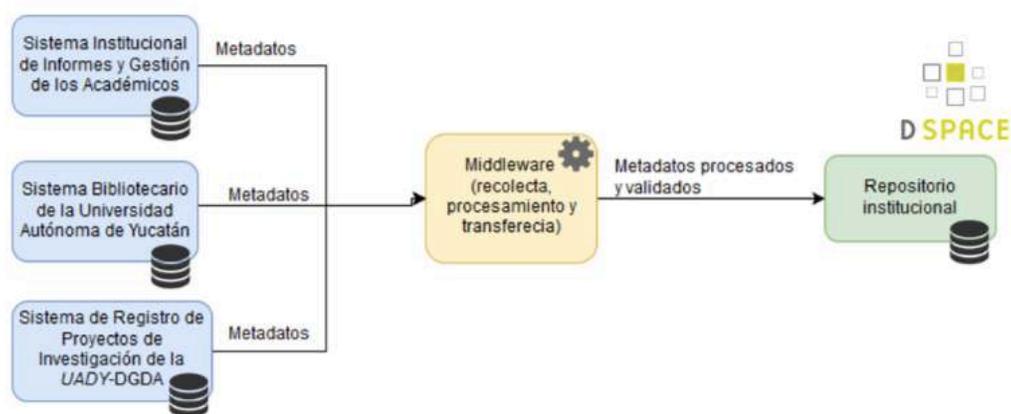


Figura 2.2 Arquitectura de REDI-UADY

## 2.6 NORMATIVAS PARA LA CIENCIA ABIERTA EN MÉXICO

La Ciencia Abierta es una política de Estado que tiene como objetivo incrementar la accesibilidad de las investigaciones científicas (financiadas con recursos públicos) para todos los ciudadanos a través de la diseminación máxima del conocimiento científico, la tecnología y la innovación. Los Lineamientos Jurídicos de Ciencia Abierta establecen reglas sobre la política de Ciencia Abierta, sobre los derechos de propiedad intelectual de forma general y en los repositorios, cosecha y diseminación, sobre los depósitos de obras ya divulgadas, uso de los recursos depositados, sobre las cosechas del repositorio nacional, autorizaciones para el depósito, entre otros (CONACYT, 2017j).

### 2.6.1 LINEAMIENTOS RELATIVOS AL PROGRAMA DE REPOSITORIOS

Los lineamientos generales de Ciencia Abierta tienen un capítulo dedicado al Programa de repositorios, el cual tiene por objetivo el acopio, la preservación y el aseguramiento del acceso abierto a los recursos de la Información Científica, Tecnológica y de Innovación producidos generalmente con recursos públicos. Para conseguir este fin específico, el Programa de Repositorios se encarga de asegurar el óptimo funcionamiento del repositorio nacional y del desarrollo de Repositorios Institucionales. Tanto el repositorio nacional como los Repositorios Institucionales deben contener los siguientes tres tipos de información:

1. Publicaciones científicas. Hace referencia a todo lo que es resultado de la investigación. En esta categoría se consideran artículos, libros, capítulos de libros, tesis o trabajos para la obtención de grados, entre otros.
2. Productos del desarrollo tecnológico y la innovación, tales como patentes, desarrollos del tipo tecnológico, innovaciones, transferencias tecnológicas, entre otros.
3. Datos primarios de las investigaciones. Es toda información recolectada y utilizada para la investigación académica, científica, tecnológica y de innovación.

Para la homologación de los contenidos, cada publicación debe cumplir con las directrices que establece CONACYT para el llenado de los metadatos para que la explotación a través de la vinculación de datos sea efectiva y adecuada. Cada directriz, característica y requisito que debe

contener la información (o producto) son determinados por el CONACYT por medio de los Lineamientos Técnicos y de las convocatorias (CONACYT, 2017j).

### 2.6.2 LINEAMIENTOS TÉCNICOS PARA LOS REPOSITORIOS MEXICANOS

La creación de políticas y lineamientos para archivar de forma efectiva en los Repositorios Institucionales y para publicar en revistas de Acceso Abierto permite la conservación del patrimonio académico, científico, tecnológico y de innovación. Estas políticas igual permiten (CONACYT, 2014):

- El aseguramiento de la difusión y visibilidad de la investigación que se ha logrado gracias al financiamiento público, y así lograr un mayor impacto
- Favorecer la eliminación de las barreras de tipo física, económica, legal y tecnológica al acceso a la investigación para conseguir la aceleración de la ciencia
- Agilizar los procesos de creación, uso, reutilización, intercambio y disseminación de las investigaciones científicas
- Mejorar la administración y valoración del patrimonio de investigación en las instituciones
- Prevenir la duplicidad, así como la promoción de la transferencia tecnológica y del conocimiento, así como la optimización de la inversión en Ciencia y Tecnología.
- Mayor transparencia en las producciones científicas, académicas, tecnológicas y de innovación, así como en sus resultados
- A partir de los resultados que se generan gracias a las investigaciones, proveer el material para la realización de minería de datos y de texto y de este modo generar nuevos conocimientos
- Más intercambio y menos desigualdad en el intercambio de información entre instituciones, entidades y países
- Asegurar el respeto a las patentes y todo lo relacionado con los derechos de autor, así como a los elementos de tipo confidencial

### 2.6.3 OTROS PAÍSES

En España existe la Ley de la Ciencia, la Tecnología y la Innovación del año 2011. Entre sus objetivos se encuentran (AEBOE, 2019):

- Fomentar la investigación científica y técnica en todas las áreas de conocimiento para progresar como sociedad utilizando como base el conocimiento
- Estimular la transferencia del conocimiento para la creación de grupos de colaboración
- Fomentar la innovación en diversos sectores y la sociedad en general
- Coordinar las políticas afines a la investigación científica para el establecimiento de objetivos, indicadores y prioridades para la asignación de recursos
- Apoyar a la internacionalización de la investigación científica, principalmente con la Unión Europea

En Estados Unidos, el acceso abierto ha sido un elemento importante para el progreso de la sociedad y las investigaciones científicas. El pilar fue ARPANET en 1966, que fue creado con la finalidad de compartir investigaciones científicas sin límite alguno (Suber, 2006). Actualmente, existe la Ley de Investigación de Acceso Justo a la Ciencia y la Tecnología (Fair Access to Science and Technology Research Act), un proyecto de ley para la ciencia abierta. Se ha propuesto en los años 2013, 2015 y 2017. Entre sus enfoques se encuentra la mejora en la calidad de vida de los estadounidenses y del resto del mundo al difundir de manera pública los resultados obtenidos por las nuevas ideas y descubrimientos resultantes de la investigación financiada por el gobierno. Cada trabajo de investigación realizado con los recursos del gobierno, debe cumplir con ciertas características, entre ellas se encuentran (Corryn, 2017):

- Ser enviado a un repositorio digital indicado o mantenido por el gobierno, en su versión final, aceptado en alguna revista científica
- Indicar los cambios realizados por el proceso de revisión por pares
- Ser de libre acceso en un periodo no mayor a 12 meses en la revista en cuestión
- Proporcionar el artículo en formatos y bajo condiciones que permitan la reutilización de la investigación y técnicas computacionales para el análisis
- Su preservación debe ser a largo plazo

## 2.7 CUERPOS ACADÉMICOS

Un cuerpo académico no es únicamente una fracción de la comunidad científica, ya que tiene ciertas características específicas. De forma general, es un conjunto de profesores-

investigadores que tienen en común una o varias líneas de estudio y tienen como objetivo tanto la aplicación como la generación de nuevos conocimientos a raíz del trabajo en conjunto. En México, los cuerpos académicos se clasifican en tres grupos y cada uno tiene características específicas para las universidades estatales y afines, institutos tecnológicos y universidades tecnológicas (DGESU, 2018). Las características que se presentan a continuación corresponden a las universidades estatales y afines, debido a que el caso de estudio se aplica con cuerpos académicos de una universidad perteneciente a ese grupo.

- **Cuerpo Académico Consolidado (CAC).** Es el nivel máximo que puede alcanzar un cuerpo académico. Sus características son las siguientes: 1) la mayoría de sus integrantes cuenta con la máxima habilitación académica para generar o aplicar de forma innovadora e independiente el conocimiento (doctorado); 2) cuentan con una amplia experiencia en la docencia y en la formación de recursos humanos; 3) la mayoría de sus integrantes cuentan con el perfil deseable definido por el PRODEP (PRODEP); 4) alto compromiso con la institución por medio de la colaboración y producción científica y académica; 5) demuestran una intensa actividad académica manifiesta en congresos, seminarios, mesas, talleres, etcétera. de forma regular y frecuente y 6) sostienen una intensa participación en redes de intercambio académico (DGESU, 2018).
- **Cuerpo Académico en Consolidación (CAEC).** Es el nivel intermedio en el que se puede clasificar un cuerpo académico. Se caracteriza por los siguientes aspectos: 1) más de la mitad de sus integrantes cuentan con doctorado; 2) cuentan con productos académicos con reconocimiento debido a su buena calidad, derivados de las líneas de investigación consolidadas; 3) al menos la tercera parte de sus integrantes cuentan con el perfil deseable definido por el PRODEP; 4) participan de forma conjunta en líneas de investigación o aplicación innovadora del conocimiento; 5) cuentan con amplia experiencia en docencia y en formación de recursos humanos y 6) colaboran con otros cuerpos académicos (DGESU, 2018).
- **Cuerpo Académico en Formación (CAEF).** Como su nombre lo indica, se refiere a los cuerpos académicos que nacen a partir de una o más líneas de investigación y se encuentra

en una etapa temprana. Sus características son: 1) los integrantes están identificados; 2) al menos la mitad de sus integrantes cuentan con el perfil deseable definido por el PRODEP; 3) tienen definidas las líneas de generación o aplicación del conocimiento que cultivarán y 4) tienen identificados los cuerpos académicos afines al que proponen y de alto nivel para establecer contacto (DGESU, 2018).

## 2.8 CONCLUSIONES

En este capítulo se presentaron las temáticas afines que fundamentan a las líneas de investigación del proyecto de tesis. Se describió el estado del arte, enfocándose principalmente en el uso de los repositorios digitales como medio para la evaluación de la productividad científica por medio de indicadores cuantitativos estadísticos. Se detalló la explicación de dichos indicadores que constituyen la base para la evaluación por parte de las instancias evaluadoras.

En el siguiente capítulo, se proponen una serie de indicadores para medir la relevancia de la productividad de una institución, dividida en dos partes: productividad en colaboración y productividad individual. Esta propuesta considera la obtención de los datos de la producción indizada y no indizada de la institución, considerando como origen diversos repositorios de acceso abierto.

### 3 MODELO DE INDICADORES PARA LA RELEVANCIA CIENTÍFICA

En el presente capítulo se propone una serie de indicadores para medir la relevancia de la producción científica de una institución a partir de dos enfoques: producción en colaboración en cuerpos académicos y producción individual. Se considera de mayor importancia a la producción indizada por Scopus, pero también se toma en cuenta la producción no indizada accesible por medio de repositorios de acceso abierto.

#### 3.1 RELEVANCIA DE UN CUERPO ACADÉMICO

##### 3.1.1 CÁLCULO DE LA PRODUCCIÓN INDIZADA POR SCOPUS

*PTS* es un indicador que representa la producción del cuerpo académico indizada por Scopus. Sin embargo, antes de exponer la fórmula, es necesario presentar los indicadores que se utilizan para obtener su valor.

##### 1. Asignación de peso por tipo de publicación

Cada publicación tiene un peso diferente debido a que las instancias consideran que algunos productos son más importantes que otros. El peso propuesto por tipo de publicación se representa como *PP* (Fórmula 3.1), donde:

$$PP = \begin{cases} 4 & \text{si es un artículo} \\ 4 & \text{si es un libro} \\ 2 & \text{si es un capítulo de libro} \\ 2 & \text{si es otro tipo de publicación} \end{cases} \quad (3.1)$$

##### 2. Cálculo del impacto de la publicación

Se utiliza el indicador *IP*, y su valor depende del tipo de identificador de la publicación. Si la publicación pertenece a una revista, entonces su identificador es ISSN. Si la publicación es un libro, capítulo de libro, artículo de conferencia en libro serializado o cualquier otro que se encuentre en un libro, el identificador es ISBN.

### Identificador ISSN

En caso de que la publicación en cuestión cumpla con la primera condición, se utiliza el indicador *SJR* (Fórmula 3.4), cuyo valor representa las citas ponderadas que ha recibido una revista en un año determinado considerando inicialmente los tres años anteriores (González-Pereira, Guerrero-Bote y Moya-Anegón, 2010).

En este trabajo, se hará referencia al indicador como Factor de impacto de Scopus. Cabe mencionar que este indicador pertenece a Elsevier y las revistas que se consideran para calcularlo se encuentran en la base de datos de Scopus. Su valor depende del área de estudio de la revista. Antes de obtener *SJR* es necesario calcular *PSJR* (*Prestige Scimago Journal Rank*) (Fórmula 3.2), el cual es un indicador que refleja el prestigio general de una revista y considerando tres aspectos:

1. Un valor mínimo de prestigio de la revista por pertenecer a la base de datos Scopus (Fórmula 3.2, 1)
2. Un valor de prestigio dado por el número de artículos almacenados en Scopus (Fórmula 3.2, 2)
3. El prestigio de las citas dado por el número e importancia de las citas recibidas por otras revistas (Fórmula 3.2, 3)

$$PSJR_i = \frac{1-d-e}{N} + e * \frac{Art_i}{\sum_{j=1}^N Art_j} + d * \left[ \underbrace{\sum_{j=1}^N C_{ji} * \frac{PSJR}{C_j} * CF}_a + \underbrace{\frac{Art_i}{\sum_{j=1}^N Art_j} * \sum_{k \in DN} PSJR_k}_b \right] \quad (3.2)$$

donde:

- *d* es una constante, cuyo valor es 0.9
- *e* es una constante, cuyo valor es 0.9999
- *N* es el número de revistas en la base de datos Scopus
- *Art<sub>i</sub>* es el número de publicaciones primarias (artículos, revisiones y artículos de conferencia) de la *i*-ésima revista en los tres años previos al año a evaluar

- $Art_j$  es el número de publicaciones primarias (artículos, revisiones y artículos de conferencia) de la  $j$ -ésima revista en los tres años previos al año a evaluar
- $C_{ji}$  es el número de referencias de la revista  $j$  a la revista  $i$
- $C_j$  es el número de referencias en la revista  $j$
- $CF$  es un factor de corrección que distribuye proporcionalmente el prestigio no considerado dentro de la ventana de tiempo de tres años (Fórmula 3.3)
- $DN$  son las revistas que no citan a otras de forma proporcional al número total de artículos primarios en Scopus

Para comprender completamente la fórmula, es necesario conocer la finalidad del tercer aspecto de  $PSJR$ . La primera parte (Fórmula 3.2, 3a)

$$\sum_{j=1}^N C_{ji} * \frac{PSJR}{C_j} * CF$$

representa el prestigio de la revista en cuestión, obtenido por medio de las citas recibidas desde otras revistas. Debido a que únicamente las citas de la ventana de tres años se consideran para realizar la distribución del prestigio, se utiliza  $CF$  (Fórmula 3.3) para evitar la pérdida del valor del prestigio correspondiente a las citas restantes en cada ciclo, repartiendo el prestigio que no se encuentra distribuido en todas las revistas proporcionalmente a su prestigio acumulado (González-Pereira et al., 2010).

$$CF = \frac{1 - (\sum_{k \in DN} PSJR_k)}{\sum_{h=1}^N \sum_{k=1}^N C_{kh} * (PSJR_k / C_k)} \quad (3.3)$$

donde:

- $DN$  son las revistas que no citan a otras de forma proporcional al número total de artículos primarios en Scopus
- $PSJR_k$  es el valor de  $PSJR$  de la  $k$ -ésima revista
- $C_{kh}$  es el número de referencias de la revista  $k$  a la revista  $h$
- $C_k$  es el número de referencias en la revista  $k$

Finalmente, se calcula el indicador *SJR* (Fórmula 3.4)

$$SJR = c * \frac{PSJR}{Art} \quad (3.4)$$

donde:

- *c* es el número de referencias de la revista
- *PSJR* es el valor de prestigio Scimago Journal Rank
- *Art* es el número de publicaciones primarias de la revista

#### *Identificador ISBN*

Si el identificador del recurso es ISBN, el impacto de la publicación se obtiene por medio de la Fórmula 3.5.

$$IISBN = \frac{A}{B} \quad (3.5)$$

donde:

- *A* es el total de citas recibidas por los libros de las mismas áreas de conocimiento de la publicación en los últimos tres años
- *B* es el total de libros publicados en los últimos tres años, pertenecientes a las mismas áreas de conocimiento de la publicación

Es preciso señalar que el impacto de la publicación (*IP*) depende del identificador de la publicación (Fórmula 3.6)

$$IP = \begin{cases} SJR, & \text{si el identificador es ISSN} \\ IISBN & \text{si el identificador es ISBN} \end{cases} \quad (3.6)$$

### **3. Cálculo del impacto de la investigación**

Para ello, se utiliza el indicador *IdI*, mismo que se calcula a través de la Fórmula 3.7.

$$IdI = IP * (1 + Cn) \quad (3.7)$$

donde:

- *IP* es el impacto de la publicación

- $Cn$  es el total de citas que ha recibido la publicación

Una vez calculados los indicadores expuestos con anterioridad, se procede a obtener el valor de la producción total indizada por Scopus del cuerpo académico utilizando el indicador  $PTS$ , el cual se calcula por medio de la Fórmula 3.8.

$$PTS = \sum_{i=1}^n PP_i + IP_i + IdI_i \quad (3.8)$$

donde:

- $n$  es el total de publicaciones indizadas por Scopus pertenecientes al cuerpo académico
- $PP$  es el peso de la  $i$ -ésima publicación
- $IP$  es el impacto de la  $i$ -ésima publicación (Fórmula 3.6)
- $IdI$  es el impacto de la investigación de la  $i$ -ésima publicación (Fórmula 3.7)

### 3.1.2 CÁLCULO DEL GRADO DE COLABORACIÓN DEL CUERPO ACADÉMICO

El Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP) destaca la importancia de la colaboración entre cuerpos académicos, así como entre instituciones nacionales y extranjeras, principalmente entre los cuerpos académicos que se encuentran en la fase de consolidación y consolidados (DGESU, 2018). A partir de esta afirmación, se define el indicador  $GCC$ , que mide el grado de colaboración por cada publicación indizada por Scopus. Su valor se obtiene a través de la Fórmula 3.9.

$$GCC = \sum_{i=1}^n \sum_{j=1}^m vc_{i,j} \quad (3.9)$$

donde:

- $n$  es el total de publicaciones indizadas por Scopus pertenecientes al cuerpo académico
- $m$  es el número de colaboradores externos en la  $i$ -ésima publicación
- $vc$  es un peso asignado al  $j$ -ésimo colaborador en la  $i$ -ésima publicación

$$vc = \begin{cases} 1 & \text{si el colaborador es interno a la institución y no al cuerpo académico} \\ 2 & \text{si el colaborador pertenece a una institución externa nacional} \\ 4 & \text{si el colaborador pertenece a una institución extranjera} \end{cases}$$

### 3.1.3 CÁLCULO DE LA PRODUCCIÓN NO INDIZADA DEL CUERPO ACADÉMICO

*PTPNI* (Fórmula 3.11) es un indicador que mide el peso total de la producción no indizada por alguna instancia. Para calcular *PTPNI* es necesario primero obtener el valor de *PPNI*, que es el indicador cuyo valor es el peso de la publicación. Cada publicación tiene un peso diferente debido a que las instancias consideran más importantes algunas que otras, y se representa con la Fórmula 3.10:

$$PPNI = \begin{cases} 2 & \text{si es un artículo} \\ 2 & \text{si es un libro} \\ 1 & \text{si es un capítulo de libro} \end{cases} \quad (3.10)$$

*PTPNI* se calcula por medio de la Fórmula 3.11.

$$PTPNI = \sum_{i=1}^n PPNI_i \quad (3.11)$$

donde:

- *n* es el total de publicaciones no indizadas por Scopus pertenecientes al cuerpo académico
- *PPNI* es el peso de la *i*-ésima publicación no indizada (Fórmula 3.10)

### 3.1.4 CÁLCULO DE LA RELEVANCIA DE UN CUERPO ACADÉMICO

Ahora que ya se conoce la producción total indizada por Scopus, el grado de colaboración y el peso total de la producción no indizada del cuerpo académico, es posible calcular la relevancia de un cuerpo académico utilizando la Fórmula 3.12.

$$RCA = PTS + GCC + PTPNI \quad (3.12)$$

donde:

- *PTS* es la producción total indizada por Scopus del cuerpo académico (Fórmula 3.8)
- *GCC* es el grado de colaboración del cuerpo académico (Fórmula 3.9)
- *PTPNI* es el peso total de la producción no indizada del cuerpo académico (Fórmula 3.11)

El valor de *RCA* representa la relevancia del trabajo colaborativo del cuerpo académico. Se puede suponer que el valor de este indicador para un cuerpo académico consolidado es mayor al de un cuerpo académico en consolidación, y a su vez, el valor de éste último es superior al de uno en formación.

## 3.2 RELEVANCIA DE UN INVESTIGADOR

### 3.2.1 RELEVANCIA EN CUERPO ACADÉMICO

*RIC* es un indicador que representa la relevancia de un investigador dentro de su cuerpo académico. Para obtener su valor, es necesario conocer *PPC*, cuyo valor representa el peso por colaboración que tiene el investigador en cada publicación, en este caso, se refiere a las publicaciones en las que ha participado en su cuerpo académico. Para esto, se utiliza como base el orden de los autores en la lista de autoría, considerando que éste depende del grado de contribución de cada autor en la publicación, siendo el primer autor el que más aportó, y así sucesivamente. No se consideran a los investigadores externos al cuerpo académico.

Por ejemplo, si en un artículo participan 5 autores y 2 de ellos pertenecen al cuerpo académico, pero el primero, el tercero y el cuarto autor no pertenecen al cuerpo académico, el segundo autor se considerará como primer autor, y el quinto como el segundo (por ser miembros del cuerpo académico). A esta lista se le denominará *lista de autoría del cuerpo académico*. El valor de *PPC* se obtiene a través de la Fórmula 3.13.

$$PPC = \begin{cases} 1 & \text{si es el primer autor o de correspondencia} \\ 2 * 2^{-n} & \text{si no es primer autor ni de correspondencia} \end{cases} \quad (3.13)$$

donde *n* es la posición del investigador en la *lista de autoría del cuerpo académico*. Lo que se busca con *PPC*, es que su valor sea 1 para el primer autor y el de correspondencia, 0.5 para el segundo, 0.25 para el tercero y así sucesivamente. En la Figura 3.1, se presenta un cuerpo académico junto con sus integrantes a modo de ejemplo, así como un artículo denominado *artículo 1* y la lista de autoría que da lugar a la *lista de autoría del cuerpo académico*, en la que

se ha eliminado a los autores no pertenecientes al cuerpo académico. También se presenta el valor de *PPC* de los miembros participantes en el *artículo 1*.

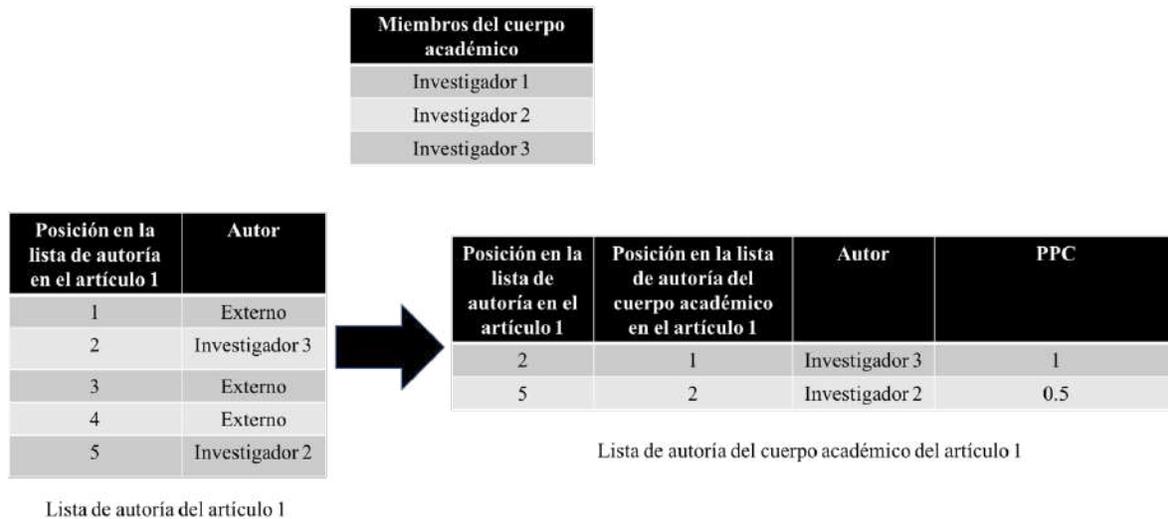


Figura 3.1 Ejemplo de una publicación perteneciente a un cuerpo académico

Una vez obtenido el valor de *PPC*, se puede calcular la relevancia del investigador dentro de su cuerpo académico por medio de la Fórmula 3.14.

$$RIC = RCA * \frac{1}{m} \sum_{i=1}^n PPC_i \quad (3.14)$$

donde:

- *RCA* es la relevancia del cuerpo académico del investigador (Fórmula 3.12)
- *m* es el total de publicaciones del cuerpo académico del investigador
- *n* es el número de publicaciones en las que participa el investigador en su cuerpo académico
- *PPC* es el peso por colaboración en la *i*-ésima publicación (Fórmula 3.13)

Cabe señalar que este indicador refleja la situación más común en la lista de autores de un producto, donde el orden de prelación es el reflejo de la relevancia del trabajo de cada uno de los autores, sin embargo pueden darse otras situaciones (por ejemplo mantener siempre un orden alfabético de los autores), en cuyo caso *PPC* deberá ser ajustado.

### 3.2.2 CÁLCULO DE LA RELEVANCIA DE LA COLABORACIÓN EXTERNA AL CUERPO ACADÉMICO

*RICE* es un indicador que representa la relevancia de la producción indizada por Scopus de un investigador en colaboración de manera externa a su cuerpo académico. Si el investigador no pertenece a un cuerpo académico, todas sus colaboraciones se verán reflejadas en este indicador. Sin embargo, antes es necesario obtener los valores de los indicadores que lo componen para calcular su valor. Dichos indicadores se presentan a continuación.

1. *PCEC* es el indicador que mide la producción en colaboración externa al cuerpo académico del investigador indizada por Scopus. Se calcula por medio de la Fórmula 3.15.

$$PCEC = \sum_{i=1}^n \left( \frac{PP_i}{PPC_i} \right) + IP_i + IdI_i \quad (3.15)$$

donde:

- $n$  es el total de publicaciones indizadas por Scopus en colaboración externa al cuerpo académico
- $PP$  es el peso de la  $i$ -ésima publicación (Fórmula 3.1)
- $PPC$  es el peso por colaboración en la  $i$ -ésima publicación (Fórmula 3.13)
- $IP$  es el impacto de la  $i$ -ésima publicación (Fórmula 3.6)
- $IdI$  es el impacto de la investigación de la  $i$ -ésima publicación (Fórmula 3.7)

2. *GCEC* es una variante del indicador *GCC* (Fórmula 3.9). Representa el grado de colaboración de cada publicación que ha realizado el investigador en colaboración externa al cuerpo académico. Se utiliza la Fórmula 3.16.

$$GCEC = \sum_{i=1}^n \sum_{j=1}^m vc_{i,j} \quad (3.16)$$

donde:

- $n$  es el total de publicaciones indizadas por Scopus en colaboración externa al cuerpo académico
- $m$  es el número de colaboradores externos en la  $i$ -ésima publicación
- $vc$  es un peso asignado al  $j$ -ésimo colaborador en la  $i$ -ésima publicación

$$vc = \begin{cases} 1 & \text{si el colaborador es interno a la institución y no al cuerpo académico} \\ 2 & \text{si el colaborador pertenece a una institución externa nacional} \\ 4 & \text{si el colaborador pertenece a una institución extranjera} \end{cases}$$

Después de obtener los valores de los dos indicadores presentados anteriormente es posible calcular la relevancia del investigador en colaboración externa al cuerpo académico haciendo uso de la Fórmula 3.17.

$$RICE = PCEC + GCEC \quad (3.17)$$

donde:

- *PCEC* es la producción en colaboración externa al cuerpo académico del investigador (Fórmula 3.15)
- *GCEC* es el grado de colaboración externa al cuerpo académico del investigador (Fórmula 3.16)

### 3.2.3 CÁLCULO DE LA RELEVANCIA DE LA PRODUCTIVIDAD INDIVIDUAL

*RPI* es el indicador que representa la relevancia de la producción del investigador realizada de manera individual. Para calcularlo, se considera el indicador *PPI*, que mide el peso de la publicación de acuerdo con su tipo (Variante de la Fórmula 3.1). Se obtiene por medio de la Fórmula 3.18.

$$PPI = \begin{cases} 2 & \text{si es un artículo o libro} \\ 1 & \text{si es otro tipo de publicación} \end{cases} \quad (3.18)$$

Una vez obtenido el indicador presentado anteriormente, se puede calcular la relevancia de la productividad individual con la Fórmula 3.19.

$$RPI = \sum_{i=1}^n PPI_i + IP_i + IdI_i \quad (3.19)$$

donde:

- *n* es el número de publicaciones individuales indizadas por Scopus
- *PPI* es el peso de la *i*-ésima publicación individual (Fórmula 3.18)
- *IP* es el impacto de la *i*-ésima publicación (Fórmula 3.6)

- $IdI$  es el impacto de la investigación en la  $i$ -ésima publicación (Fórmula 3.7)

### 3.2.4 CÁLCULO DE LA PRODUCTIVIDAD NO INDIZADA DEL INVESTIGADOR

$PNI$  es el indicador que representa la productividad no indizada del investigador, externa al cuerpo académico, en colaboración y de forma individual. Se obtiene a través de la Fórmula 3.20.

$$PNI = \sum_{i=1}^n PPNI_i * PPC_i \quad (3.20)$$

donde:

- $n$  es el número de publicaciones no indizadas
- $PPNI$  es el peso de la  $i$ -ésima publicación no indizada (Fórmula 3.10, en lugar de cuerpos académicos, es para la producción del investigador de manera externa)
- $PPC$  es el peso por colaboración en la  $i$ -ésima publicación (Fórmula 3.13)

### 3.2.5 CÁLCULO DE LA RELEVANCIA DE UN INVESTIGADOR

Ahora que ya se conoce la relevancia del investigador en su cuerpo académico, su producción en otras colaboraciones, de forma individual y sus publicaciones no indizadas es posible conocer la relevancia total de un investigador utilizando la Fórmula 3.21.

$$RDI = RIC + RICE + RPI + PNI \quad (3.21)$$

donde:

- $RIC$  es la relevancia del investigador en su cuerpo académico (Fórmula 3.14)
- $RICE$  es la relevancia del investigador en colaboración externa a su cuerpo académico (Fórmula 3.17)
- $RPI$  es la relevancia de la producción individual del investigador (Fórmula 3.19)
- $PNI$  es el peso total de la producción no indizada del investigador y externa al cuerpo académico (Fórmula 3.20)

### 3.3 CONCLUSIONES

En este capítulo se describieron los indicadores que permiten medir la relevancia de la productividad científica en colaboración con cuerpos académicos y de forma individual. Se consideraron aspectos como la producción total indizada por Scopus, el grado de colaboración, la relevancia de las citas y la producción no indizada.

El modelo presentado es flexible para adaptarse a las necesidades institucionales. Los pesos propuestos para cada autor y para cada publicación son arbitrarios, y pueden ser reajustados para futuras evaluaciones. Para el cálculo de los indicadores se ha expuesto utilizar la producción almacenada en Scopus, sin embargo, es posible también recuperar y evaluar a partir de la producción almacenada en otros repositorios basados en estándares abiertos.

En el siguiente capítulo se describirá la implementación del sistema generador de indicadores de relevancia científica, desde la perspectiva de las Ciencias de la Computación.

## **4 SISTEMA GENERADOR DE INDICADORES DE RELEVANCIA CIENTÍFICA**

En este capítulo se presentan los aspectos del Sistema Generador de Indicadores de Relevancia Científica desde la perspectiva computacional. Se muestran el marco arquitectónico en capas, las tecnologías de desarrollo utilizadas, la estructura de la base de datos documental, la interfaz web, los procedimientos para consulta y almacenamiento de las publicaciones científicas, la generación de los indicadores de la relevancia para los cuerpos académicos y los investigadores y finalmente, la estructura de modelo ontológico utilizado para la representación del conocimiento.

### **4.1 MARCO ARQUITECTÓNICO**

A continuación, se describe la arquitectura para la representación de la relevancia científica de los investigadores. La información para realizar el cálculo de los indicadores fue recuperada de diversos repositorios digitales. El diseño de la arquitectura se presenta en la Figura 4.1.

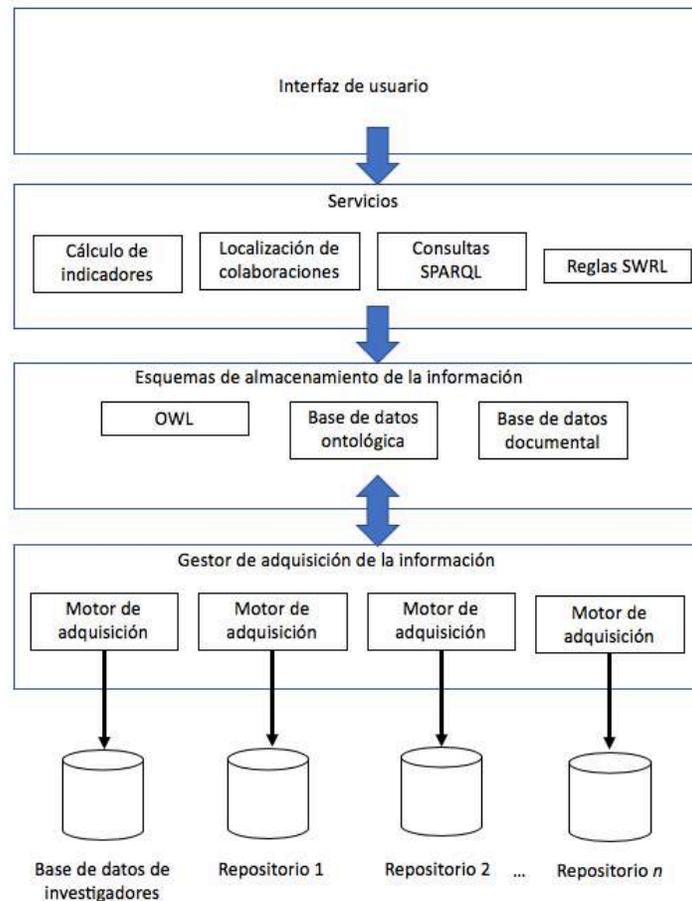


Figura 4.1 Diseño de la arquitectura propuesta

La arquitectura se ha diseñado en capas, de manera ordenada y facilitando el flujo de la información entre ellas y de manera ordenada. Las funciones específicas de cada capa se describen a continuación.

- **Interfaz de usuario.** Contiene la información obtenida por los servicios a partir de los datos obtenidos
- **Servicios.** Se compone de módulos que trabajan en conjunto para la representación de los resultados por medio de la interfaz del usuario.
  - Cálculo de indicadores. Obtiene los valores de los indicadores de la relevancia de los grupos de investigación y los investigadores, a partir de sus publicaciones y los valores de sus metadatos.

- Localización de colaboraciones. Analiza las publicaciones y las colaboraciones para encontrar los grupos de colaboración existentes en una institución.
- Consultas SPARQL. Se encarga de obtener información relevante de los datos almacenados en la ontología para ser representados al usuario por medio de la interfaz.
- Reglas SWRL. Con base en la ontología, se validan las instancias que cumplan con una o más condiciones establecidas para la representación del conocimiento.
- **Esquemas de almacenamiento de la información.** Contiene las herramientas que almacenan la información de los investigadores y de su producción. Las herramientas son:
  - OWL. Es un lenguaje de etiquetas para la representación y difusión del conocimiento por medio de ontologías.
  - Base de datos ontológica. Almacena documentos en formato OWL, RDF, entre otros.
  - Base de datos basada en documentos. A diferencia de una base de datos relacional, ésta almacena la información en documentos, facilitando su manejo para transferir sus datos a una ontología o para almacenar datos semi-estructurados.
- **Gestor de adquisición de la información.** Esta capa contiene motores de adquisición que se conectan directamente a diversos repositorios por el medio correspondiente, ya sea a través de la cosecha de metadatos con OAI-PMH o alguna API, si el repositorio así lo requiere.
- **Origen de los datos.** Contiene los repositorios que almacenan la información acerca de la producción académica y científica de los investigadores.

## 4.2 ESPECIFICACIONES TÉCNICAS

En esta sección se explican los aspectos técnicos que tienen relación con el sistema, considerando las tecnologías involucradas en el desarrollo y la estructura de la base de datos.

### 4.2.1 TECNOLOGÍAS DE DESARROLLO

- **Lenguaje de programación Python.** Por medio de este lenguaje se desarrollan diversos scripts, cada uno con una finalidad diferente: recuperar las claves de autores y la producción de los investigadores indizada por Scopus; recuperar la producción almacenada en diversos repositorios de acceso abierto; calcular los indicadores para cuerpos académicos e investigadores y generar los archivos .JSON para la representación gráfica de los resultados.
- **Lenguaje HTML (HyperText Markup Language).** Su uso es primordial para la interfaz del usuario, misma que presenta los resultados obtenidos.
- **Lenguaje CSS (Cascading StyleSheet).** Define la presentación de todos los elementos de cada página web en la interfaz del usuario.
- **Base de datos NoSQL.** Se utiliza una base de datos orientada a documentos debido a la facilidad de integración de los datos. El sistema de base de datos utilizado es MongoDB y almacena los registros en una estructura BSON, que es muy similar a JSON.
- **Base de datos ontológica.** Por medio de clases y propiedades, cada instancia representa un elemento del mundo real, además que permite establecer reglas de inferencia para nuevos individuos.
- **JavaScript.** Brinda funciones adicionales a las páginas web. Se implementa para la visualización de los resultados en la interfaz del usuario, tanto para los grafos y los mapas de árboles.

- **XAMPP.** Es un servidor independiente de plataforma que se compone de un gestor de bases de datos MySQL, el servidor web Apache, PHP y Pearl. La interfaz se encuentra almacenada en el servidor antes mencionado.

En la Figura 4.2, se ilustra la interacción entre las tecnologías involucradas en el sistema, tanto del lado del cliente como del servidor.

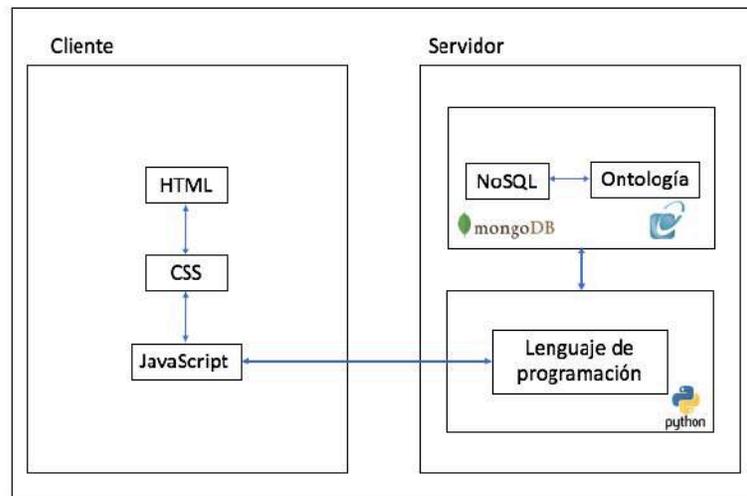


Figura 4.2 Interacción de las tecnologías involucradas en el sistema

#### 4.2.2 ESTRUCTURA DE LA BASE DE DATOS NOSQL

Todos los datos recuperados de diversos repositorios, junto con los indicadores calculados de los cuerpos académicos e investigadores se almacenan en la base de datos MongoDB. A diferencia de una base de datos SQL, el equivalente de una tabla se conoce como *documento* y cada registro recibe el nombre de *colección* o *documento*. Toda la información se almacena en tres bases de datos (Figura 4.3). A pesar de que las colecciones que almacenan la producción cuentan con casi los mismos campos, es útil clasificar cada registro de acuerdo con el tipo de publicación, al ser una base de datos documental.

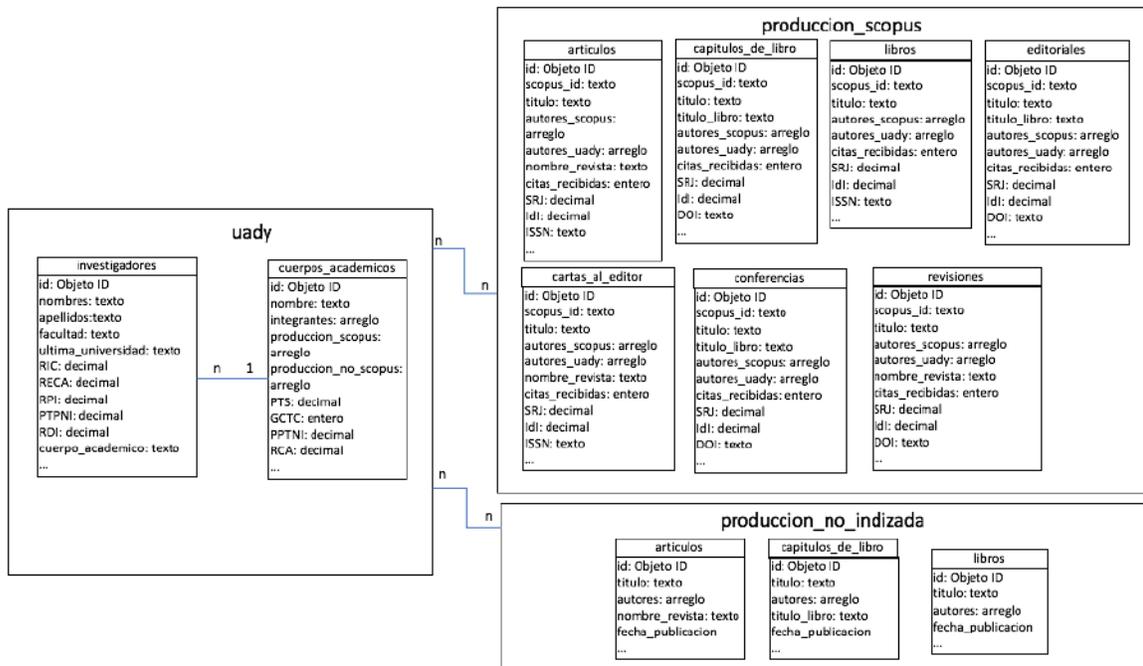


Figura 4.3 Esquema de relaciones entre bases de datos y sus colecciones

El contenido de cada base de datos y sus colecciones se describen a continuación.

- *uady*. Es la base de datos que contiene los registros relacionados con la Universidad Autónoma de Yucatán (UADY). Sus colecciones son:
  - *investigadores*. Contiene el registro de cada profesor de la UADY. Cada registro está conformado por: un identificador único, el nombre completo, la facultad de adscripción, la universidad que le otorgó el último grado, las claves de Scopus y los valores de sus indicadores de relevancia científica.
  - *cuerpos\_academicos*. Almacena el registro de cada grupo de investigación reconocido por el PRODEP. Cada registro está conformado por: un identificador único, el nombre, los integrantes, su producción y los valores de sus indicadores de relevancia científica.
- *produccion\_scopus*. Contiene el registro de toda la producción indizada por Scopus. Todas las colecciones (artículos, capítulos de libro, libros, conferencias, editoriales, cartas al editor y revisiones) cuentan con un identificador único, título, las claves de

Scopus de los autores, las claves de los autores pertenecientes a la UADY y los indicadores SJR y IRI. Los capítulos de libro registran el nombre del libro en donde se encuentra la publicación en cuestión. Las conferencias registran el libro o evento de origen.

- *produccion\_no\_indizada*. Almacena el registro de las publicaciones disponibles en los repositorios de acceso abierto. Cuenta con tres colecciones: artículos, libros y capítulos de libro. Cada registro contiene un identificador único, título, la lista de autores, el nombre de la revista (para artículos) y fecha de publicación.

### 4.3 INTERFAZ WEB

El diseño de la interfaz web se ha realizado con el lenguaje CSS que permite que el usuario visualice de manera estética los resultados de los indicadores. La página principal brinda acceso a: la información sobre el cálculo de los indicadores, los indicadores para los cuerpos académicos y los indicadores para los investigadores (Ver Figura 4.4).



Figura 4.4 Página principal de la interfaz del usuario

Tanto las secciones de los indicadores para cuerpos académicos como la de los investigadores presentan los resultados de cada indicador por medio de mapas de árboles, siendo esto posible con JavaScript, donde el tamaño de cada rectángulo va de acuerdo con el valor de cada indicador. En cada rectángulo se observa qué facultad tiene mayor peso y al seleccionarla, se visualiza el peso por cuerpo académico o investigador. En la Figura 4.5 se presentan los resultados del indicador de relevancia total de los cuerpos académicos y en la Figura 4.6, los de la relevancia total de los investigadores.



Figura 4.5 Visualización de los resultados de la relevancia de cuerpos académicos

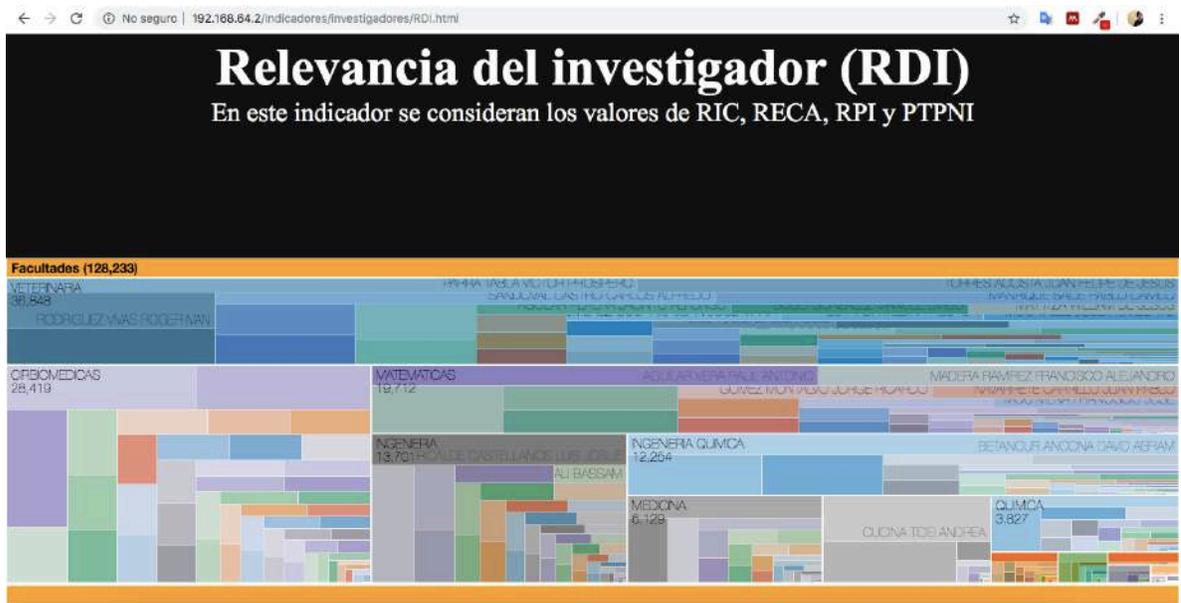


Figura 4.6 Visualización de los resultados de la relevancia de investigadores

Además, es posible apreciar la relación existente entre cuerpos académicos y los investigadores por medio de un grafo, donde cada vértice en color está asociado a la facultad a la que pertenece el cuerpo académico o investigador. Al ser seleccionado cada nodo con el puntero, es posible visualizar las relaciones existentes entre este vértice con los otros. Los grafos pueden ampliarse en la sección deseada para mejor visualización. En la Figura 4.7, se presenta la relación entre cuerpos académicos y en la Figura 4.8, entre investigadores.

← → ↻ No seguro | 192.168.64.2/indicadores/cuerpos\_academicos/colaboracion\_uady2.html

## Colaboración entre cuerpos académicos

En el grafo, cada punto (vértice) representa a un cuerpo académico. Cada color representa a una facultad. Los cuerpos académicos colaboradores son aquellos que han realizado por lo menos una publicación en conjunto



Figura 4.7 Grafo de colaboración entre cuerpos académicos en la interfaz del usuario

← → ↻ No seguro | 192.168.64.2/indicadores/investigadores/colaboracion\_uady2.html

## Colaboración entre investigadores

En el grafo, cada punto (vértice) representa a un investigador. Cada color representa a una facultad. Los investigadores colaboradores son aquellos que han realizado por lo menos una publicación en conjunto

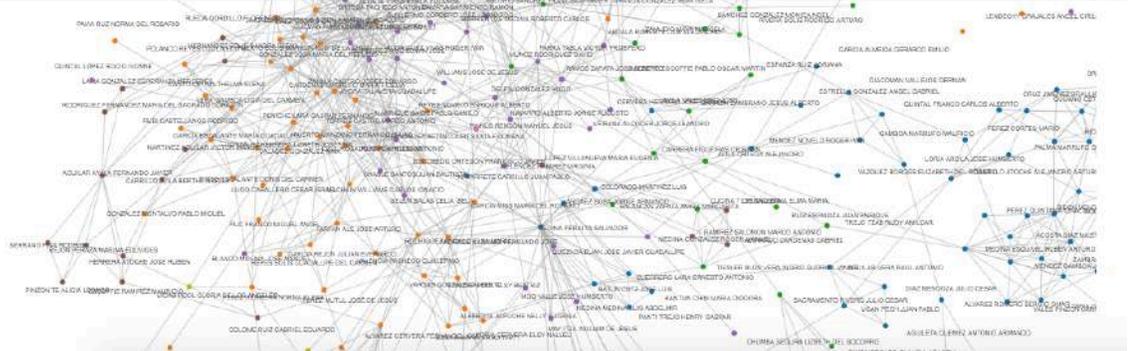


Figura 4.8 Grafo de colaboración entre investigadores en la interfaz del usuario

## 4.4 FUNCIONALIDAD DEL SISTEMA

### 4.4.1 CONSULTA Y ALMACENAMIENTO DE LAS PUBLICACIONES CIENTÍFICAS

Para recuperar la producción científica es necesario mantener una base de datos con los registros de los profesores, considerando como datos importantes:

- Nombre completo
- Facultad o área de estudio
- Universidad que le otorgó el último grado

La recuperación de la producción para el cálculo de los indicadores de la relevancia de los cuerpos académicos y de los investigadores se realiza a través de tres pasos:

1. Recuperación de las claves de Scopus
2. Recuperación de la producción indizada por Scopus
3. Recuperación de las publicaciones no indizadas por Scopus

**1. Recuperación de las claves de Scopus.** Para realizar una búsqueda más específica de las publicaciones asociadas a un investigador, ésta se puede realizar a través de la clave (ID) de autor de Scopus. Es muy común que un investigador tenga más de una clave debido a los diversos formatos con los que se ha registrado su nombre en las publicaciones.

Para recuperar sus claves, se consulta utilizando como parámetros su nombre completo, su facultad de adscripción o área de estudio y la universidad que le otorgó el último grado de estudios. A partir de los resultados retornados en formato JSON, se analiza cada uno de ellos utilizando condiciones para la extracción de los datos. Con base en ellas se seleccionan las claves correctas.

Cada resultado contiene los siguientes metadatos: *preferred-name* (nombre preferente) que incluye *given-name* (nombre), *surname* (apellidos), *initials* (iniciales del nombre), *affiliation-name* (nombre de la institución a la que pertenece), *af-id* (clave de Scopus de la institución, que en este caso, se consideran aquellas pertenecientes a la UADY) y en ocasiones, *Scopus-name-variants* (variantes del nombre del investigador) que incluye los mismos metadatos (nombre,

apellidos e iniciales) y *subject-area*, que indica el o las áreas de conocimiento del investigador. Antes, es necesario eliminar los acentos de los nombres, tanto de los datos obtenidos de la base de datos de investigadores como de los resultados de Scopus para hacer más eficiente la selección de los IDs correctos.

Se considera como clave del investigador si sus metadatos asociados cumplen con por lo menos una de las siguientes condiciones. También se crea un diccionario que asocia cada facultad con las áreas de conocimiento reconocidas por Scopus. Un ejemplo es presentado en la Figura 4.9. En algunas situaciones se considera el diccionario, principalmente en aquellos donde la selección del ID es propensa a ser errónea por la poca coincidencia entre los datos originales y los recuperados.

INGENIERIA: Engineering (all), Materials Science (all), Mathematics (all), Computer Science (all)
INGENIERIA QUIMICA: Chemistry (all), Materials Science (all), Engineering (all), Chemical Engineering (all)
MATEMATICAS: Computer Science (all), Engineering (all), Mathematics (all)
MEDICINA: Medicine (all), Immunology and Microbiology (all), Nursing (all), Agricultural and Biological Sciences (all)

Figura 4.9 Ejemplo de diccionario de las facultades

- **Condición 1.** De acuerdo con la Figura 4.10, si el valor del campo *surname* de *Scopus-preferred-name* coincide con el valor de los apellidos registrados en la base de datos, el valor del campo *given-name* coincide con los nombres del investigador, entonces la clave de autor de Scopus pertenece al investigador. Se considera esta condición como el mejor de los casos porque tanto los nombres como los apellidos del investigador coinciden con los del resultado de la consulta.

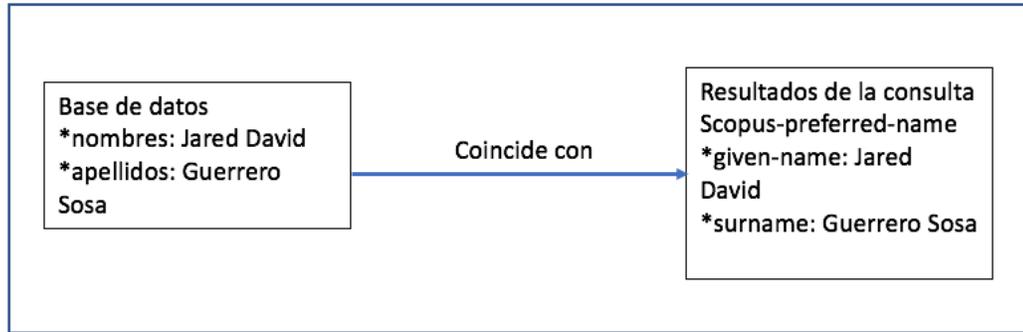


Figura 4.10 Primera condición para la selección de ID de Scopus del investigador

- **Condición 2.** De acuerdo con la Figura 4.11, si el valor del campo *surname* de *Scopus-preferred-name* coincide con el valor de los apellidos registrados en la base de datos pero separados por un guión, entonces la clave de autor de Scopus pertenece al investigador si además cumple con una de las siguientes condiciones:
  - a) Si el valor del campo *given\_name* de *Scopus-preferred-name* es equivalente a una inicial de uno de los nombres del investigador en la base de datos, seguido de un punto
  - b) Si la primera palabra del campo *given\_name* de *Scopus-preferred-name* coincide con el primer nombre del investigador en la base de datos
  - c) Si el valor del campo *initials* de *Scopus-preferred-name* coincide con las iniciales del nombre del investigador como se encuentra almacenado en la base de datos
  - d) Si para que la primera palabra del campo *given-name* de *Scopus-preferred-name* coincida con el primer nombre del investigador registrado en la base de datos es necesario realizarse dos movimientos por medio de la distancia Levenshtein (que indica el número de cambios que se debe realizar en una cadena para convertirse en otra (Levenshtein, 1966))
  - e) Si el valor del campo *given\_name* de *Scopus-preferred-name* coincide con el segundo nombre del investigador (si tiene) en la base de datos
  - f) Si el valor del nombre del investigador en la base de datos contiene la palabra “de” (por ejemplo, María de Jesús) y el valor del campo *given-name* de *Scopus-preferred-name* contiene los nombres del investigador, pero sin la palabra “de”

En esta condición el ID es correcto porque, a pesar de que los apellidos del resultado de la consulta se encuentran separados por un guión, coinciden con los que se encuentran registrados en la base de datos.

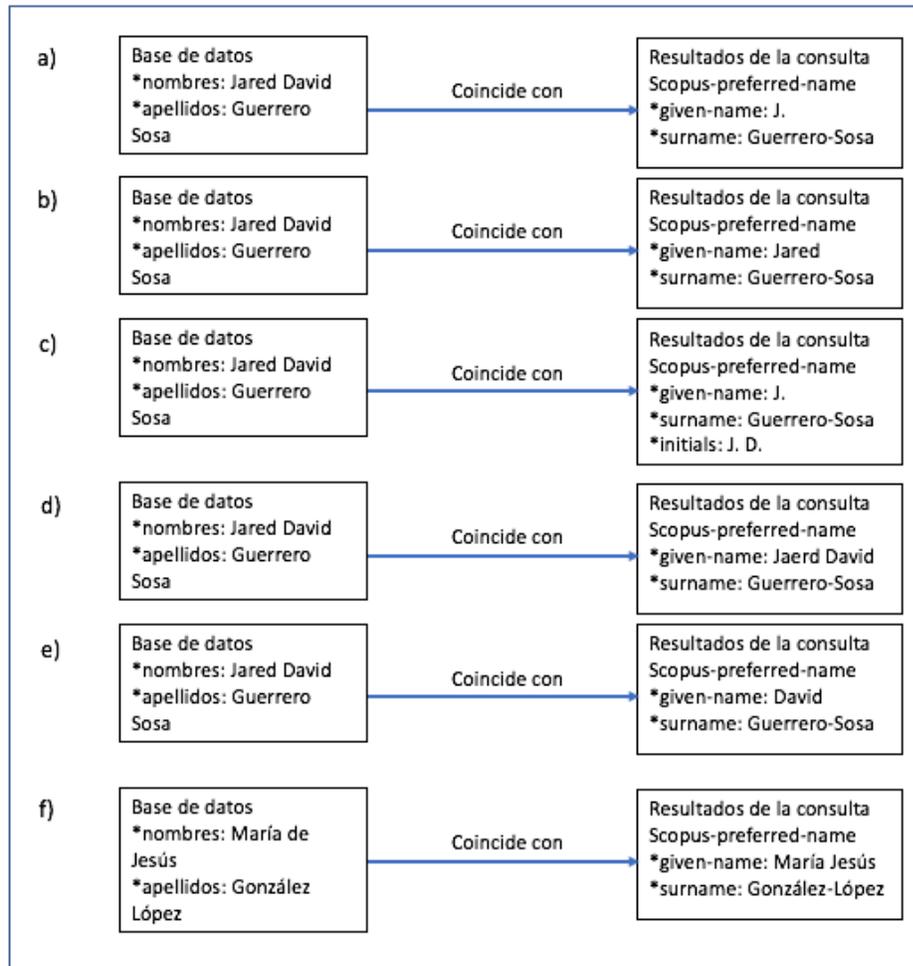


Figura 4.11 Segunda condición para la selección de ID de Scopus del investigador

- Condición 3.** De acuerdo con la Figura 4.12, si los valores de los campos *surname* y *given\_name* de *Scopus-preferred-name* consisten en una sola palabra, los valores del nombre y del apellido de un investigador registrados en la base de datos igualmente consisten en una sola palabra, pero el valor de *given-name* contiene el apellido y el valor de *surname* incluye el nombre, entonces la clave de autor de Scopus pertenece al

investigador. En esta condición el ID es correcto, ya que coincide el nombre del investigador, pero en formato *apellido, nombre*; en lugar de *nombre, apellido*.

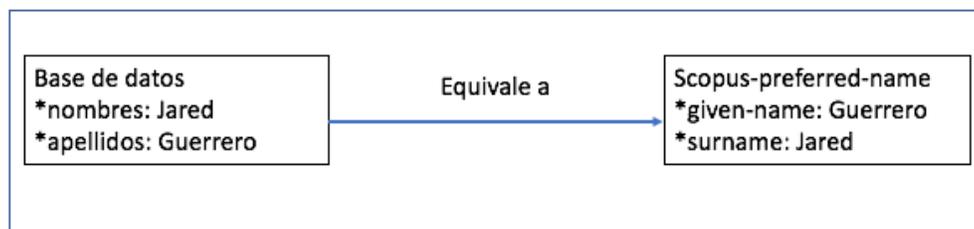


Figura 4.12 Tercera condición para la selección de ID de Scopus del investigador

- **Condición 4.** De acuerdo con la Figura 4.13, si el valor de *given\_name* de *Scopus-preferred-name* tiene longitud de una palabra, el resultado de la consulta no cuenta con *Scopus-name-variants*, el valor de *af-id* coincide con una clave de Scopus de institución de la UADY, al menos uno de los valores de *subject-area* se encuentra en el índice asociado a la facultad del investigador y el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido del investigador en la base de datos, entonces la clave de autor de Scopus pertenece al investigador si además cumple con una de las siguientes condiciones:
  - a) El valor de *given\_name* de *Scopus-preferred-name* coincide con el primer nombre en la base de datos; o
  - b) El valor de *given\_name* de *Scopus-preferred-name* coincide con el segundo nombre del investigador en la base de datos; o
  - c) El valor de *given\_name* de *Scopus-preferred-name* coincide con la inicial del primer nombre del investigador en la base de datos, seguido de un punto.

En esta condición el ID es el correcto porque la coincidencia se encuentra principalmente en el área de conocimiento, la institución, un nombre o inicial, y el primer apellido con los datos del investigador en la base de datos.

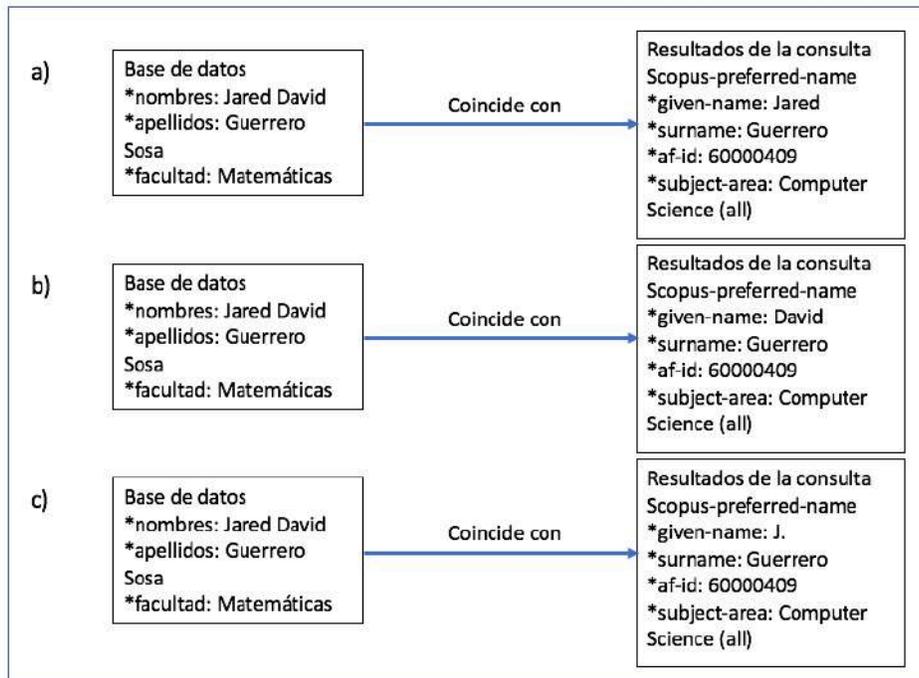


Figura 4.13 Cuarta condición para la selección de ID de Scopus del investigador

- **Condición 5.** De acuerdo con la Figura 4.14, si el resultado de la consulta contiene *Scopus-name-variants*, el valor de *af-id* contiene una clave de Scopus de institución de la UADY, ninguno de los valores de *subject-area* se encuentran en el índice asociado a la facultad del investigador, el valor de *given\_name* de *Scopus-preferred-name* tiene longitud de una palabra y el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido del investigador en la base de datos, entonces la clave de autor de Scopus pertenece al investigador si además cumple con una de las siguientes condiciones:

- Uno de los valores de *given\_name* de *Scopus-name-variants* coincide con el primer nombre del investigador en la base de datos y el valor de su respectivo *surname* de *Scopus-name-variants* coincide con los apellidos del investigador en la base de datos
- Uno de los valores de *given\_name* de *Scopus-name-variants* coincide con un nombre del investigador que no sea el primero en la base de datos y el valor de su respectivo *surname* de *Scopus-name-variants* coincide con los apellidos del investigador en la base de datos

- c) Uno de los valores de *given\_name* de *Scopus-name-variants* coincide con la inicial del primer nombre del investigador en la base de datos y el valor de su respectivo *surname* de *Scopus-name-variants* coincide con los apellidos del investigador en la base de datos

En esta condición, al no coincidir el área de estudio del investigador según Scopus y su facultad, se recurre a validar por medio de *Scopus-name-variants*, encontrando la principal coincidencia entre un nombre o una inicial del investigador junto con sus apellidos.

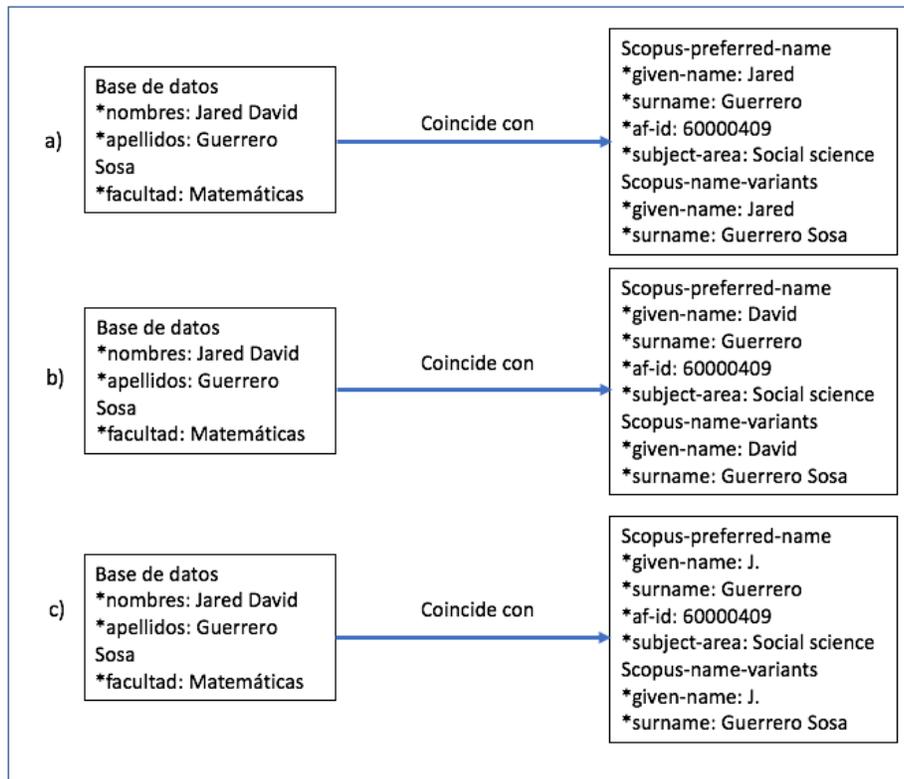


Figura 4.14 Quinta condición para la selección de ID de Scopus del investigador

- **Condición 6.** De acuerdo con la Figura 4.15, si el valor de *surname* de *Scopus-preferred-name* coincide con el valor del primer apellido del investigador en la base de datos, el registro no incluye *Scopus-name-variants*, el valor de *af-id* contiene una clave de Scopus de institución de la UADY, al menos uno de los valores de *subject-area* se encuentra en

el índice asociado a la facultad del investigador y el campo *nombre* de la base de datos contiene más de una palabra, entonces la clave de autor de Scopus pertenece al investigador si además cumple con una de las siguientes condiciones:

- a) El valor de *given\_name* de *Scopus-preferred-name* coincide con el primer nombre del investigador en la base de datos, o
- b) El valor de *given\_name* de *Scopus-preferred-name* coincide con las iniciales del nombre del investigador según el registro de su nombre en la base de datos

En esta condición se identifica que el ID es correcto porque la coincidencia se encuentra entre el resultado con el primer nombre del investigador o sus iniciales, su primer apellido y su facultad.

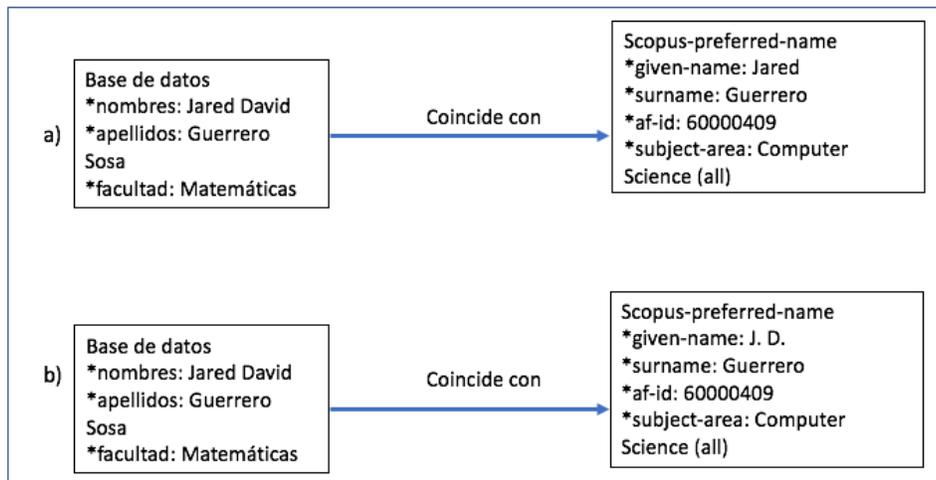


Figura 4.15 Sexta condición para la selección de ID de Scopus del investigador

- **Condición 7.** De acuerdo con la Figura 4.16, si el investigador tiene más de un nombre de acuerdo con su registro en la base de datos, el valor de *given\_name* de *Scopus-preferred-name* coincide con el primer nombre o con las iniciales del nombre del investigador en la base de datos, el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido del investigador en la base de datos, el resultado cuenta con *Scopus-name-variants*, el valor de *af-id* contiene una clave de Scopus de institución de la UADY y ninguno de los valores de *subject-area* se encuentran en el índice asociado

a la facultad del investigador, entonces la clave de autor de Scopus pertenece al investigador si además cumple con una de las siguientes condiciones:

- a) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con uno de los nombres y los apellidos del investigador en la base de datos
- b) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con uno de los nombres del investigador en la base de datos y sus apellidos separados por un guión
- c) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con la inicial del primer nombre y el primer apellido del investigador en la base de datos
- d) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con uno de los nombres y el primer apellido del investigador en la base de datos

En esta condición, el área de estudio del investigador de acuerdo con Scopus no pertenece a su facultad. Entonces se identifica que el ID es correcto validando por medio de *Scopus-name-variants* porque la coincidencia se encuentra entre el resultado con un nombre y los apellidos del investigador, ya sean éstos separados o no por un guión; o bien, entre el resultado y la primera inicial o uno de los nombres con el primer apellido del investigador.

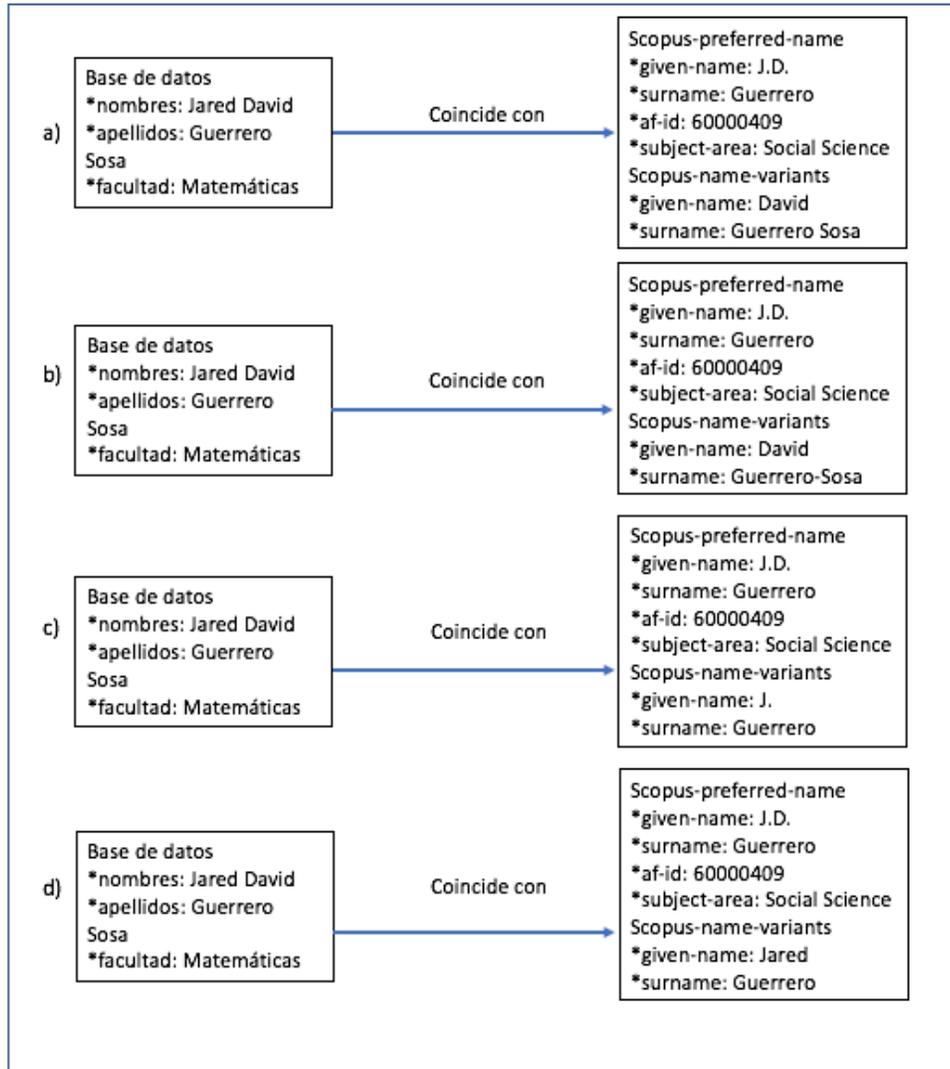


Figura 4.16 Séptima condición para la selección de ID de Scopus del investigador

- Condición 8.** De acuerdo con la Figura 4.17, si el valor de *surname* de *Scopus-preferred-name* no incluye un guión para separar los apellidos, el investigador cuenta con más de un nombre en la base de datos, el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido del investigador en la base de datos, el valor de *given\_name* de *Scopus-preferred-name* coincide con el primer nombre con una inicial de los otros nombres del investigador de acuerdo con su registro en la base de datos, el resultado de la consulta no cuenta con *Scopus-name-variants*, el valor de *af-id* contiene una clave de Scopus de institución de la UADY y al menos uno de los valores de *subject-area* se encuentra en el índice asociado a la facultad del investigador, entonces la clave de autor

de Scopus pertenece al investigador. En esta condición el ID es correcto porque se encuentra la coincidencia entre el resultado con el primer nombre del investigador junto con una inicial de otro nombre y el primer apellido y además, la consulta confirma que pertenece a la UADY y su área de estudio según Scopus pertenece a su facultad.

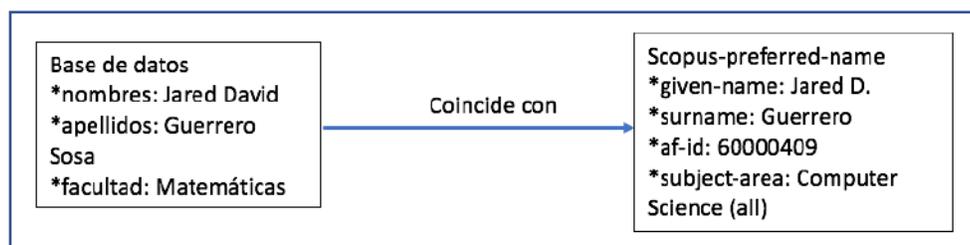


Figura 4.17 Octava condición para la selección de ID de Scopus del investigador

- **Condición 9.** De acuerdo con la Figura 4.18, si el valor de *surname* de *Scopus-preferred-name* no incluye un guión para separar los apellidos, el investigador tiene más de un nombre en la base de datos, el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido del investigador en la base de datos, el valor de *given\_name* de *Scopus-preferred-name* coincide con el primer nombre con una inicial de los otros nombres del investigador de acuerdo con su registro en la base de datos, el valor de *af-id* contiene una clave de Scopus de institución de la UADY y ninguno de los valores de *subject-area* se encuentran en el índice asociado a la facultad del investigador, entonces la clave de autor de Scopus pertenece al investigador si cumple una de las siguientes condiciones:
  - a) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con uno de los nombres del investigador y sus apellidos en la base de datos
  - b) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con uno de los nombres del investigador en la base de datos y sus apellidos separados por un guión
  - c) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con la inicial del primer nombre y el primer apellido del investigador en la base de datos

- d) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con uno de los nombres y el primer apellido del investigador en la base de datos
- e) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants* coinciden con las iniciales del nombre del investigador en la base de datos junto con sus apellidos separados por un guión

En esta condición, el área de estudio del investigador de acuerdo con Scopus no pertenece a su facultad. Entonces se identifica que el ID es correcto validando por medio de *Scopus-name-variants* porque la coincidencia se encuentra entre el resultado con uno de los siguientes casos: 1) un nombre y los apellidos del investigador, ya sean éstos separados o no por un guión, 2) la primera inicial o uno de los nombres con el primer apellido del investigador, y 3) las iniciales de los nombres del investigador y sus apellidos separados por un guión.

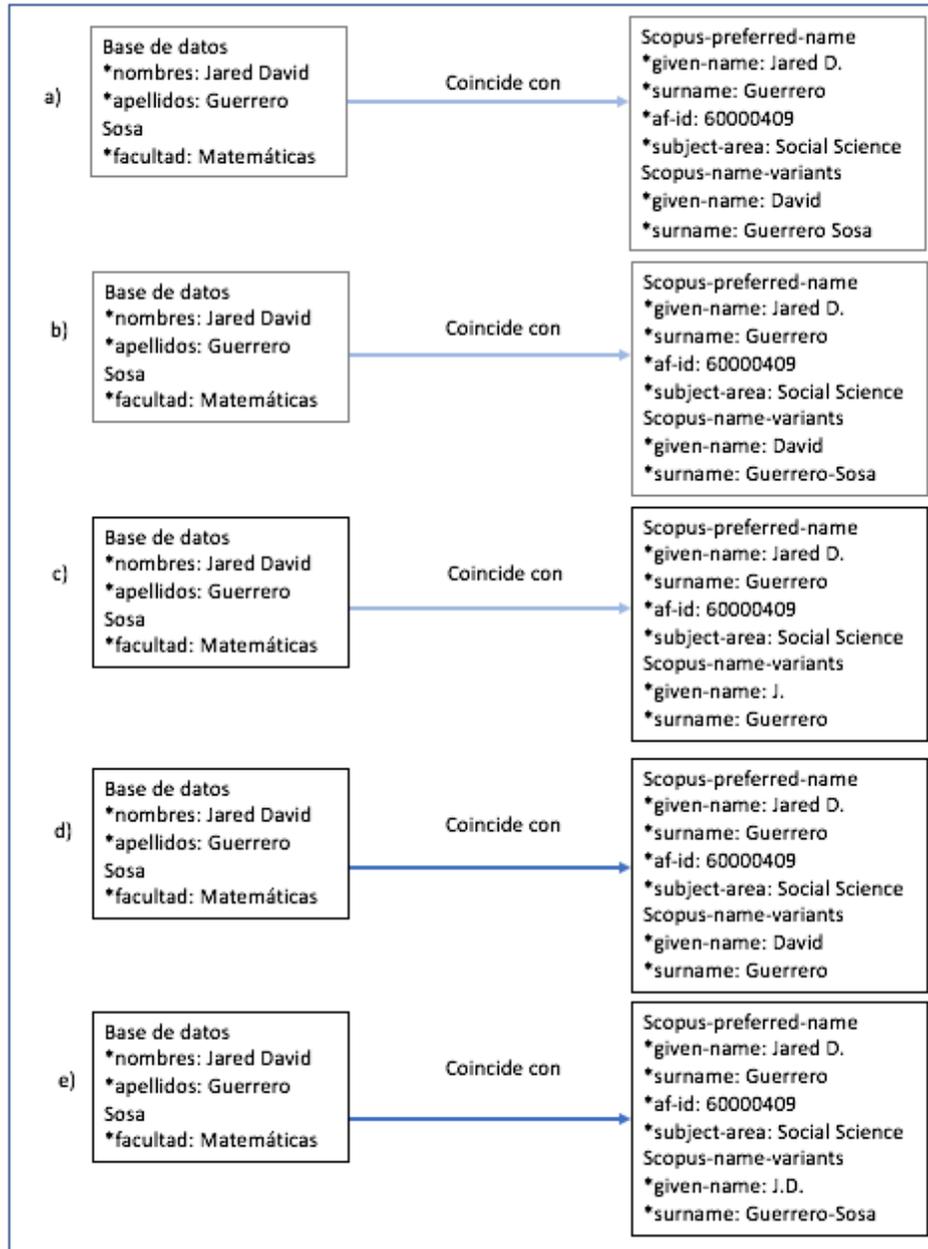


Figura 4.18 Novena condición para la selección de ID de Scopus del investigador

- Condición 10.** De acuerdo con la Figura 4.19, si el valor de *given\_name* de *Scopus-preferred-name* coincide con la inicial del primer nombre y los demás nombres completos del investigador en la base de datos, el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido en la base de datos, el resultado no cuenta con *Scopus-name-variants*, el valor de *af-id* coincide con alguna clave de Scopus asociada a

la UADY y al menos uno de los valores de *subject-area* se encuentra en el índice asociado a la facultad del investigador, entonces la clave de autor de Scopus pertenece al investigador.

En esta condición el ID es correcto porque encuentra la coincidencia entre el resultado con la inicial del primer nombre junto con los demás nombres completos del investigador, su primer apellido y su facultad.

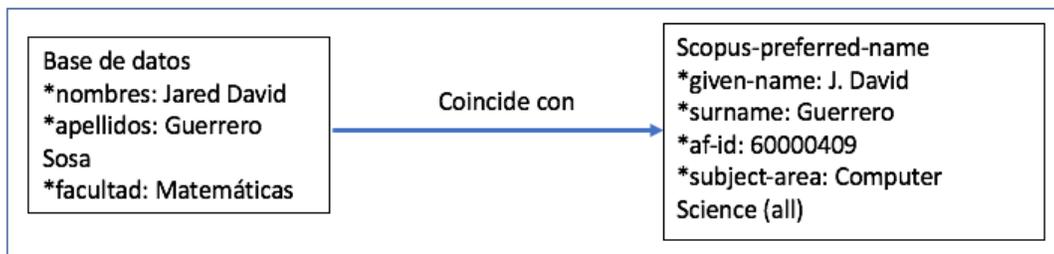


Figura 4.19 Décima condición para la selección de ID de Scopus del investigador

- **Condición 11.** De acuerdo con la Figura 4.20, si el valor de *given\_name* de *Scopus-preferred-name* coincide con la inicial del primer nombre y los demás nombres del investigador en la base de datos, el valor de *surname* de *Scopus-preferred-name* coincide con el primer apellido en la base de datos, el valor de *af-id* coincide con alguna clave de Scopus asociada a la UADY y ninguno de los valores de *subject-area* se encuentran en el índice asociado a la facultad del investigador, entonces la clave de autor de Scopus pertenece al investigador si cumple una de las siguientes condiciones:
  - a) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants*, coinciden con uno de los nombres junto con los apellidos del investigador en la base de datos
  - b) Uno de los valores de *given-name* con su respectivo *surname*, ambos de *Scopus-name-variants*, coinciden con uno de los nombres del investigador en la base de datos junto con los apellidos separados por un guión

En esta condición, el área de estudio del investigador de acuerdo con Scopus no pertenece a su facultad. Entonces se identifica que el ID es correcto validando por medio de *Scopus-name-variants* porque la coincidencia se encuentra entre el resultado con uno de los nombres del investigador junto con sus apellidos, separados o no por un guión.

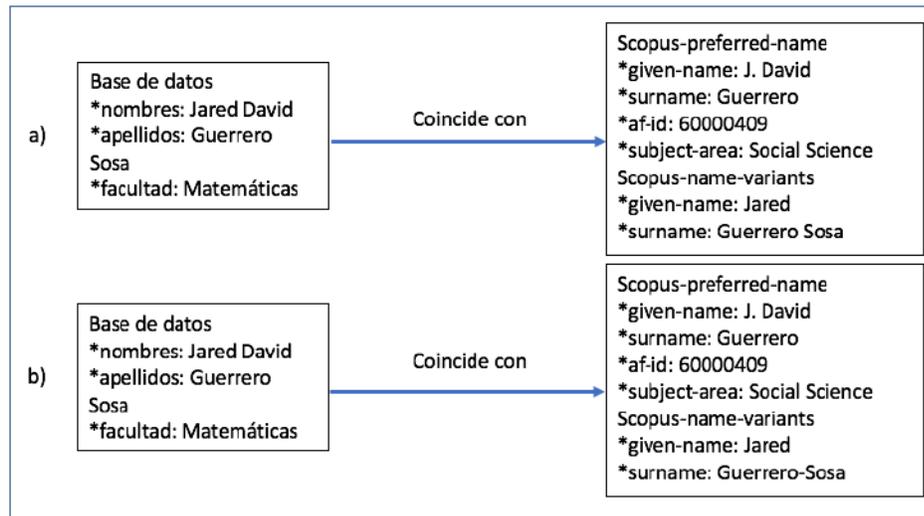


Figura 4.20 Undécima condición para la selección de ID de Scopus del investigador

## 2. Recuperación de la producción indizada por Scopus

Se realiza una consulta en Scopus por medio de su API de búsqueda, utilizando como parámetros las claves de Scopus de autor de cada investigador y las claves institucionales de su institución. De esta manera se limita a recuperar únicamente la producción asociada a la institución correspondiente. Cada resultado contiene diversos metadatos, entre ellos: *dc:identifier* (identificador Scopus del recurso), *dc:title* (título), *citedby-count* (número total de citas) y *prism:pageRange* (páginas). Una vez realizada la consulta se almacenan cada uno de los registros obtenidos.

## 3. Recuperación de las publicaciones no indizadas por Scopus

Para recuperar las publicaciones que se encuentran almacenadas en los repositorios de acceso abierto, se utiliza un cosechador de metadatos OAI-PMH. Esta tarea se lleva a cabo con base en el identificador de OAI. Por medio del verbo *ListRecords* se recuperan todos los registros almacenados del repositorio. Posteriormente, se extraen los valores de los metadatos

*dc:identifier* (identificador del recurso en su repositorio), *dc:title* (título del recurso), *dc:creator* (autores creadores del recurso), *dc:contributor* (autores colaboradores) y *dc:date* (fecha de publicación). Se verifica si en cada recurso localizado se encuentra por lo menos un autor perteneciente a la institución evaluada y en caso de ser verdadero, se almacena en la base de datos de la producción no indizada.

#### 4.4.2 GENERACIÓN DEL INDICADOR DE LA PRODUCTIVIDAD DE CUERPOS ACADÉMICOS

Al contar con la producción almacenada en la base de datos, se procede a realizar las consultas para calcular los indicadores de los cuerpos académicos. Para obtener los indicadores *PTS* (Producción total de Scopus) y *GCC* (Grado de colaboración total) se utiliza la información de la producción indizada. El procedimiento a seguir se presenta en la Figura 4.21 y es el siguiente:

1. Se obtienen los identificadores (ID) de la lista de integrantes del cuerpo académico y se calcula una combinación ordinaria (Fórmula 4.1)

$$C_{n,2} = \frac{n!}{2!(n-2)!} \quad (4.1)$$

donde  $n$  es el número de integrantes en el cuerpo académico.

2. Con cada combinación, se realiza una consulta a la base de datos recuperando la producción perteneciente al par de integrantes del cuerpo académico, debido a que se considera como una publicación del cuerpo académico si pertenece a por lo menos dos integrantes del mismo.

3. Si el registro no se ha tomado en cuenta, debe considerarse como publicación del cuerpo académico.

4. A partir de cada publicación del cuerpo académico, considerar el tipo de publicación, el valor en el indicador SJR (relevancia de la publicación) e IdI (Impacto de la investigación) para obtener el valor del indicador *PTS*.

5. Partiendo de las mismas publicaciones, identificar a los autores del cuerpo académico con el campo *scopus\_id\_autores*. Como este campo es un arreglo (Figura 4.22), se localizan las posiciones de dichos investigadores y se excluyen, dejando únicamente a las personas que no pertenecen al cuerpo académico.

Con las mismas posiciones, localizar las instituciones a las que pertenecen los colaboradores con el campo *afiliaciones\_autores*. Cada elemento del arreglo tiene un vector por cada institución de investigador. Si la posición 2 del vector tiene como valor un país que no es México, se considera como colaborador extranjero. Si en la posición 2 se encuentra registrado México, entonces se verifica el valor en la posición 1, que contiene la clave de Scopus de la institución del investigador. Si no pertenece a una clave de institución de Scopus asociada a la UADY, se considera como colaborador nacional ajeno a la misma institución del cuerpo académico. En caso de que el valor en la posición 1 sea una clave de la UADY, se considera como colaborador de la misma institución pero ajeno al cuerpo académico.

6. De acuerdo con el peso de cada colaborador en cada publicación, se obtiene el valor del indicador *GCC*.

7. Para obtener el indicador *PTPNI* a partir de las combinaciones ordinarias de la Fórmula 4.1 se consulta a la base de datos de producción no indizada con los nombres completos de los investigadores.

8. Si la publicación no se ha considerado como una del cuerpo académico, se toma en cuenta y se asigna el peso correspondiente de acuerdo con el tipo de publicación.

9. Sumar los valores de los pesos de cada publicación no indizada y obtener el indicador *PTPNI*.

10. Con los indicadores anteriores se puede calcular el valor de la Relevancia del Cuerpo Académico (*RCA*).

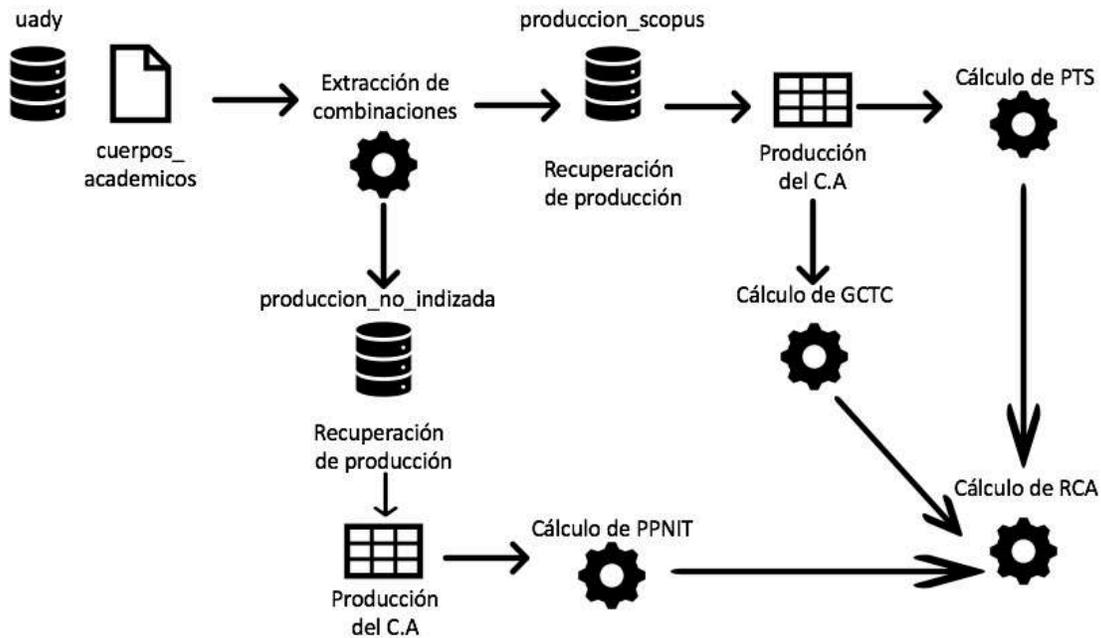


Figura 4.21 Procedimiento de cálculo de indicadores para cuerpos académicos

```

  ▼ afiliaciones_autores: Array
    ▼ 0: Array
      0: 0
      1: "60000409"
      2: "Mexico"
    ▼ 1: Array
      0: 1
      1: "60000409"
      2: "Mexico"
  
```

Figura 4.22 Ejemplo del registro de las afiliaciones de los investigadores en publicaciones científicas

#### 4.4.3 GENERACIÓN DEL INDICADOR DE LA PRODUCTIVIDAD DE LOS INVESTIGADORES

Al contar con la producción almacenada en la base de datos, se realizan las consultas para calcular los indicadores de los investigadores. Para obtener los indicadores RICE (Relevancia Externa al Cuerpo Académico) y *RPI* (Relevancia de la Producción Individual) se utiliza la información de la producción indizada. Para el indicador *PNI* se utiliza la información de la base de datos de la producción no indizada. El procedimiento a seguir se presenta en la Figura 4.23 y es el siguiente:

1. Para obtener el indicador *RIC* se recuperan las publicaciones en las que el investigador ha contribuido en la base de datos indizada y no indizada. También se consulta y recupera el valor del indicador *RCA* de su cuerpo académico y se calcula el indicador *RIC*.
2. Para el cálculo del indicador *RICE* se recupera la producción indizada por Scopus realizada en colaboración externa al cuerpo académico. En caso de no tener, toda la producción en colaboración se refleja en este indicador.
3. El indicador *RPI* se obtiene recuperando la producción realizada de manera individual.
4. Al consultar la producción no indizada en la base de datos de dicha producción, se calcula el indicador *PNI*.
5. Con los indicadores anteriores se procede a calcular la Relevancia Total del Investigador (*RDI*).

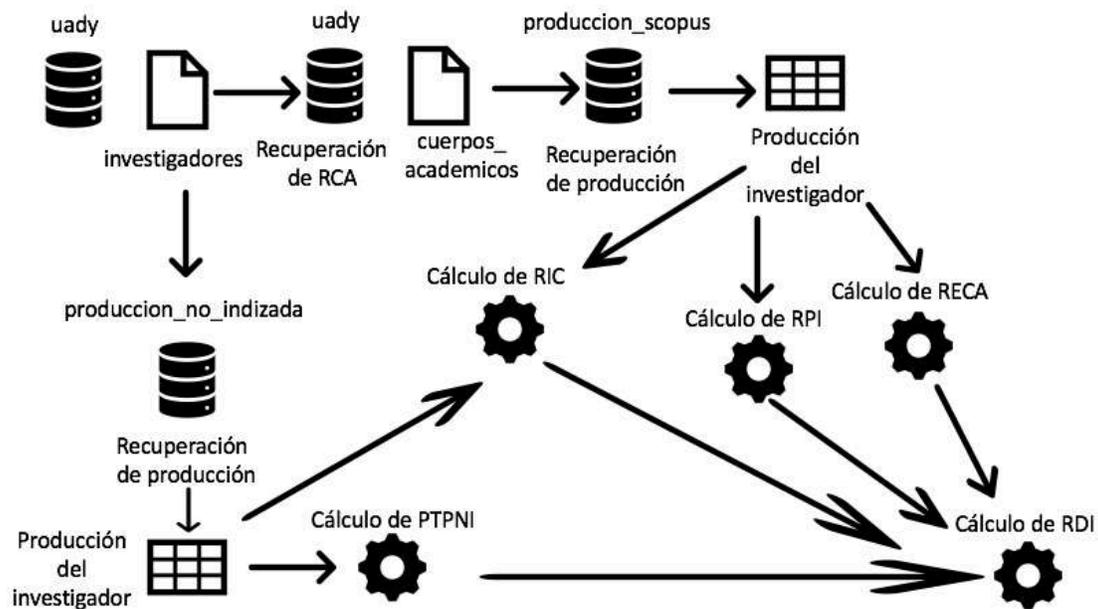


Figura 4.23 Procedimiento de cálculo de indicadores para investigadores

## 4.5 REPRESENTACIÓN DEL CONOCIMIENTO

De acuerdo con la arquitectura del sistema propuesto, los servicios de consultas SPARQL y reglas SWRL se relacionan directamente con la ontología, que es una especificación explícita de una conceptualización (Gruber, 1993), una descripción formal y explícita de conceptos en un dominio del discurso, propiedades de cada concepto describiendo varias características y atributos del concepto y las restricciones en las propiedades (Noy y McGuinness, 2001).

Para que una ontología pueda ser compartida y reutilizada es elemental que se ponga a disposición a través de la web, dando lugar al término Web Semántica, cuya finalidad es la publicación de datos para ser utilizados por diversas aplicaciones por medio de los metadatos semánticos (Berners-Lee y Hendler, 2001).

La ontología que se propone consiste en las clases que se mencionan a continuación. Algunas de ellas son subclases de otras, las cuales, además de tomar las propiedades de su clase padre, son otras las que las distinguen.

- Publicacion
  - ConferenciaPublicacion
  - PublicacionSeriada
    - CapituloDeLibroPublicacion
    - EditorialLibro
    - LibroPublicacion
  - RevistaPublicacion
    - ArticuloPublicacion
    - CartaPublicacion
    - EditorialRevista
    - RevisionPublicacion
- Identificador
  - InvestigadorIdentificador
    - INEIdentificador
    - CVUIdentificador

- CURPIdentificador
  - PublicacionIdentificador
    - ISBNIdentificador
    - ISSNIdentificador
    - DOIIdentificador
    - RecursoIdentificador
- Institucion
- Ciudad
- Estado
- Investigador
- CuerpoAcademico

Las propiedades de las clases de una ontología se clasifican en dos tipos. Las de tipo dato, cuyo valor va de acuerdo con un tipo de dato en específico, y las de tipo objeto, que tienen como valor alguna instancia de una o varias clases. En la Tabla 4.1 se presentan las clases con sus respectivas propiedades. En la Figura 4.24 se presenta una parte de las clases con sus propiedades.

Tabla 4.1 Propiedades del modelo ontológico

Clase	Propiedad	Rango
Publicacion	esPublicacionDeInvestigador	Investigador
	esPublicacionDeCuerpoAcademico	CuerpoAcademico
	publicacionTieneRecursoIdentificador	RecursoIdentificador
	publicacionTieneDOIIdentificador	DOIIdentificador
	publicacionTieneCitasRecibidas	Integer
	publicacionTieneTitulo	String
	publicacionTieneIDI	Float
	publicacionTieneFechaDePublicacion	String
	publicacionTienePaginas	String
	publicacionEsIndizada	String
	publicacionTieneIdI	Float
	publicacionTieneSJR	Float

Publicacion- ConferenciaPublicacion	publicacionTieneISBNIdentificador	ISBNIdentificador
	publicacionTieneISSNIdentificador	ISSNIdentificador
	publicacionTieneOrigen	String
Publicacion- PublicacionSeriada	publicacionTieneISBNIdentificador	ISBNIdentificador
	publicacionTieneOrigen	String
Publicacion- RevistaPublicacion	articuloPublicacionTieneISSNIdentificador	ISSNIdentificador
	publicacionTieneISSNIdentificador	ISSNIdentificador
	revistaPublicacionTieneEjemplar	String
	articuloPublicacionTieneVolumen	String
	revistaPublicacionTieneNombreDeLaRevista	String
Identificador	identificadorTieneValor	String
InvestigadorIdentificador	esInvestigadorIdentificadorDeInvestigador	Investigador
DOIIdentificador	esDOIIdentificadorDePublicacion	Publicacion
ISBNIdentificador	esISBNIdentificadorDePublicacion	ConferenciaPublicacion or PublicacionSeriada
ISSNIdentificador	esISSNIdentificadorDePublicacion	ConferenciaPublicacion or RevistaPublicacion
RecursoIdentificador	esRecursoIdentificadorDePublicacion	Publicacion
Institucion	esInstitucionDeInvestigador	Investigador
	institucionTieneCiudad	Ciudad
	institucionTieneNombre	String
Ciudad	esCiudadDeInstitucion	Institucion
	ciudadTieneEstado	Estado
	ciudadTieneNombre	String
Estado	esEstadoDeCiudad	Ciudad
	estadoTieneNombre	String
Investigador	investigadorTieneInstitucion	Institucion
	investigadorTieneInvestigadorIdentificador	InvestigadorIdentificador
	investigadorTienePublicacion	Publicacion
	investigadorTieneGenero	String
	investigadorTieneCitasRecibidas	Integer
	investigadorTieneTituloAcademico	String
	investigadorTieneApellidos	String
	investigadorTieneNombre	String

	investigadorTieneSNI	String
	investigadorTieneAreaDeConocimiento	String
	investigadorTienePRODEP	Boolean
	investigadorColaboraCon	Investigador
	investigadorTieneRIC	Float
	investigadorTieneRICE	Float
	investigadorTieneRPI	Float
	investigadorTienePNI	Float
	investigadorTieneRDI	Float
	investigadorTieneCuerpoAcademico	CuerpoAcademico
CuerpoAcademico	cuerpoAcademicoTienePublicacion	Publicacion
	esCuerpoAcademicoDeInvestigador	Investigador
	cuerpoAcademicoTieneInstitucion	Institucion
	cuerpoAcademicoTieneFacultad	String
	cuerpoAcademicoTienePTS	Float
	cuerpoAcademicoTieneGCC	Float
	cuerpoAcademicoTiene PTPNI	Float
	cuerpoAcademicoTieneRCA	Float
	cuerpoAcademicoTieneNivel	String
	cuerpoAcademicoTieneNombre	String

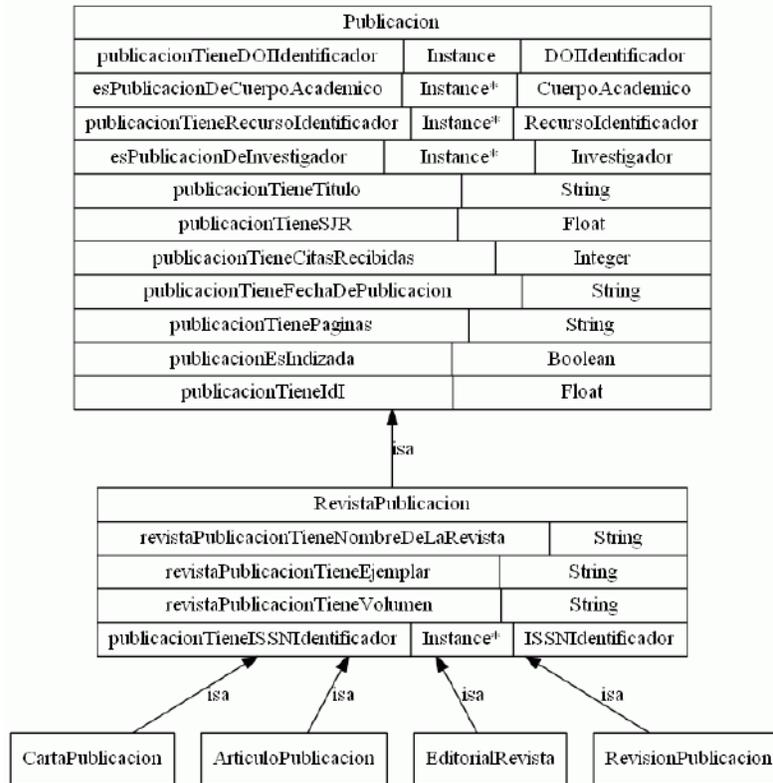


Figura 4.24 Clases y subclases relacionadas con publicaciones y sus propiedades

#### 4.6 CONCLUSIONES

En este capítulo se presentó una descripción completa del sistema de índices para valorar la relevancia de la producción académica y la investigación a partir de repositorios digitales y metadatos. Se han detallado los siguientes aspectos: el marco arquitectónico, las especificaciones técnicas, la interfaz web, la funcionalidad del sistema (consulta y almacenamiento de la producción científica, así como el cálculo de los indicadores propuestos) y finalmente, la estructura del modelo ontológico utilizado para la representación del conocimiento.

En el siguiente capítulo, se presentan los resultados de casos de estudio de la correlación entre los indicadores propuestos para determinar los elementos más importantes para la relevancia científica en cuerpos académicos e investigadores de la Facultad de Matemáticas, así como el análisis del estado de la producción científica a través del uso de técnicas de minería de datos.

## 5 RESULTADOS

Este capítulo tiene como finalidad presentar casos de estudio sobre la correlación entre los indicadores propuestos y las métricas que representan el conteo de las publicaciones, así como el estudio del estado de la investigación a partir de los indicadores propuestos utilizando técnicas de minería de datos.

La Universidad Autónoma de Yucatán (UADY) es la institución educativa más importante en el sureste de México. Cuenta con 15 facultades distribuidas en cinco campus y un centro de investigación enfocado en dos áreas de estudio:

1. Campus de Arquitectura, Hábitat, Arte y Diseño (AHAD)
2. Campus de Ciencias Biológicas y Agropecuarias (CByA)
3. Campus de Ciencias de la Salud (CSalud)
4. Campus de Ciencias Exactas e Ingenierías (CEeI)
5. Campus de Ciencias Sociales, Económico Administrativas y Humanidades (CSociales)
6. Centro de Investigaciones Regionales Dr. Hideyo Noguchi, Unidad de Ciencias Sociales (CIR-Sociales)
7. Centro de Investigaciones Regionales Dr. Hideyo Noguchi, Unidad de Ciencias Biomédicas (CIR-Biomédicas)

Al día 1 de junio del 2018, en la UADY se contaba con 78 cuerpos académicos y 824 profesores de tiempo completo. En la Tabla 5.1 se presentan los cuerpos académicos por campus y por nivel de consolidación, donde se observa que el campus de Ciencias Exactas e Ingenierías y el campus de Ciencias Sociales, Económico Administrativas y Humanidades cuentan con el mayor número de cuerpos académicos, con 21 cada uno.

Tabla 5.1 Cuerpos académicos de la UADY por campus y grado de consolidación

<b>Grado de consolidación</b>	<b>AHAD</b>	<b>CByA</b>	<b>CSalud</b>	<b>CEeI</b>	<b>CSociales</b>	<b>CIR-Sociales</b>	<b>CIR-Biomédicas</b>
En formación	0	0	1	2	3	1	1
En consolidación	2	2	8	9	9	0	2
Consolidado	1	8	4	10	9	2	4
<b>Total</b>	<b>3</b>	<b>10</b>	<b>13</b>	<b>21</b>	<b>21</b>	<b>3</b>	<b>7</b>

En cuanto a los profesores de la UADY, en la Tabla 5.2 se presentan por campus y género, donde se expone que los dos campus con más académicos son el campus de Ciencias Exactas e Ingenierías y el campus de Ciencias Sociales, Económico Administrativas y Humanidades, con 232 y 226, respectivamente.

Tabla 5.2 Profesores de la UADY por campus y género

<b>Género</b>	<b>AHAD</b>	<b>CbyA</b>	<b>Csalud</b>	<b>CEeI</b>	<b>CSociales</b>	<b>CIR-Sociales</b>	<b>CIR-Biomédicas</b>
Hombres	23	73	86	171	107	14	31
Mujeres	11	11	73	61	119	14	30
<b>Total</b>	<b>34</b>	<b>84</b>	<b>159</b>	<b>232</b>	<b>226</b>	<b>28</b>	<b>61</b>

Como se ha mencionado en los capítulos 1 y 2, los profesores pueden ser reconocidos por instancias evaluadoras, como el Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP). En la Tabla 5.3 se muestra, por campus, cuántos profesores pertenecen a estas instancias y sus porcentajes de acuerdo al total de profesores de tiempo completo del campus correspondiente.

Tabla 5.3 Profesores de tiempo completo de la UADY desde la perspectiva del PRODEP y del SNI

	<b>AHAD</b>	<b>CByA</b>	<b>CSalud</b>	<b>CEeI</b>	<b>CSociales</b>	<b>CIR- Sociales</b>	<b>CIR- Biomédicas</b>
Profesores de tiempo completo	34	84	159	232	226	28	61
Doctores	13 (38.2%)	63 (75%)	41 (25.7%)	110 (47.4%)	99 (43.8%)	23 (82.1%)	41 (67.2%)
Profesores con perfil PRODEP	29 (85%)	83 (98.8%)	144 (90.5%)	219 (94.3%)	203 (89.8%)	28 (100%)	58 (95%)
Profesores sin SNI	28 (82%)	46 (54.7%)	138 (86.7%)	159 (68.5%)	163 (72.1%)	8 (28.5%)	30 (49.1%)
Candidato al SNI	0 (0%)	1 (1.1%)	7 (4.4%)	17 (7.32%)	17 (7.5%)	1 (3.5%)	2 (3.27%)
SNI 1	4 (11.7%)	25 (29.7%)	11 (6.9%)	47 (20.2%)	38 (16.8%)	12 (42.85%)	22 (36%)
SNI 2	2 (5.88%)	6 (7.14%)	3 (1.8%)	7 (3%)	7 (3.09%)	6 (21.42%)	7 (11.4%)
SNI 3	0 (0%)	6 (7.14%)	0 (0%)	2 (0.8%)	1 (0.4%)	1 (3.5%)	0 (0%)
Profesores SNI con CA	5 (14.7%)	33 (39.28%)	14 (8.8%)	57 (24.5%)	43 (19.02%)	13 (46.42%)	26 (42.62%)

A partir de la lista de los 824 profesores de tiempo completo, se consultó la producción de cada uno en Scopus y en los repositorios asociados al repositorio nacional, de los cuales 438 (53.15%) cuentan con producción científica almacenada en dichos repositorios. En la Tabla 5.4 se expone, por campus, desde la perspectiva del SNI y del PRODEP, los profesores que cuentan con producción científica y su porcentaje de acuerdo al total de profesores que han realizado investigación en su respectivo campus.

Tabla 5.4 Profesores de la UADY con producción científica desde la perspectiva del PRODEP y del SNI

	<b>CByA</b>	<b>CSalud</b>	<b>CEeI</b>	<b>CSociales</b>	<b>CIR- Sociales</b>	<b>CIR- Biomédicas</b>
Profesores con producción	76	86	159	54	8	55
Doctores	59 (77.6%)	34 (39.5%)	98 (61.6%)	41 (75.9%)	8 (100%)	39 (70.9%)
Profesores con perfil PRODEP	75 (98.6%)	84 (97.6%)	157 (35.8%)	54 (100%)	8 (100%)	53 (96.3%)
Profesores sin SNI	39 (51.3%)	68 (79%)	94 (59.1%)	17 (31.4%)	2 (25%)	25 (45.4%)
Candidato al SNI	0 (0%)	6 (6.9%)	10 (6.28%)	10 (18.5%)	0 (0%)	2 (3.6%)
SNI 1	25 (32.8%)	9 (10.4%)	46 (28.9%)	19 (35.1%)	4 (50%)	21 (38.1%)
SNI 2	6 (7.8%)	3 (3.4%)	7 (4.4%)	7 (12.9%)	2 (25%)	7 (12.7%)
SNI 3	6 (7.8%)	0 (0%)	2 (1.2%)	1 (1.85%)	0 (0%)	0 (0%)
Profesores SNI con CA	33 (43.4%)	13 (15.1%)	54 (33.9%)	23 (42.5%)	4 (50%)	25 (45.4%)

De los 824 profesores, 438 han generado 2858 publicaciones. En la Figura 5.1 se presenta la distribución por cada tipo de publicación. El total por cada tipo de publicación se encuentra en la Tabla 5.5. De acuerdo con estos resultados, los artículos representan la mayor parte de la producción y los libros la menor parte. Cabe destacar que entre los otros tipos de publicaciones se encuentran las cartas al editor, las editoriales y las revisiones.

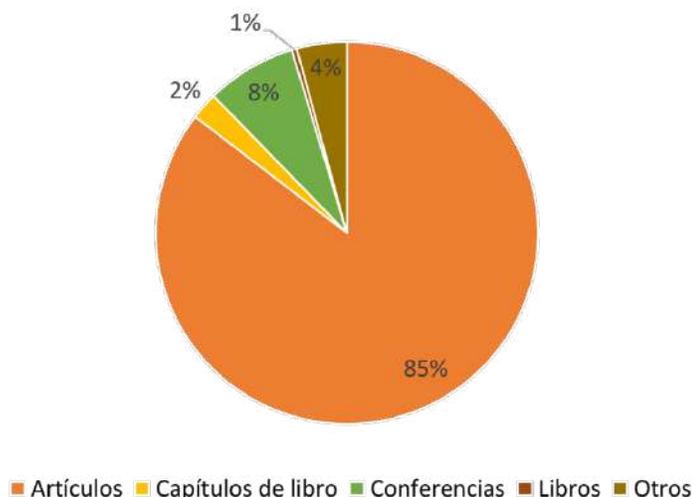


Figura 5.1 Producción científica de los profesores de la UADY almacenada en diversos repositorios

Tabla 5.5 Producción científica de la UADY almacenada en Scopus y en el repositorio nacional

Tipo de publicación	Total	Porcentaje
Artículos	2441	85%
Capítulos de libro	66	2%
Conferencias	218	8%
Libros	12	1%
Otros	121	4%
Producción total	2858	100%

Se encontró que, de las publicaciones recuperadas, las más antiguas pertenecen al año 1979 y las más recientes al año 2019. En la Figura 5.2 se presenta la producción científica de la UADY por años, donde cabe resaltar que las publicaciones en el año 2019 (52) abarcan hasta febrero de 2019, mes en que se realizó la última consulta.

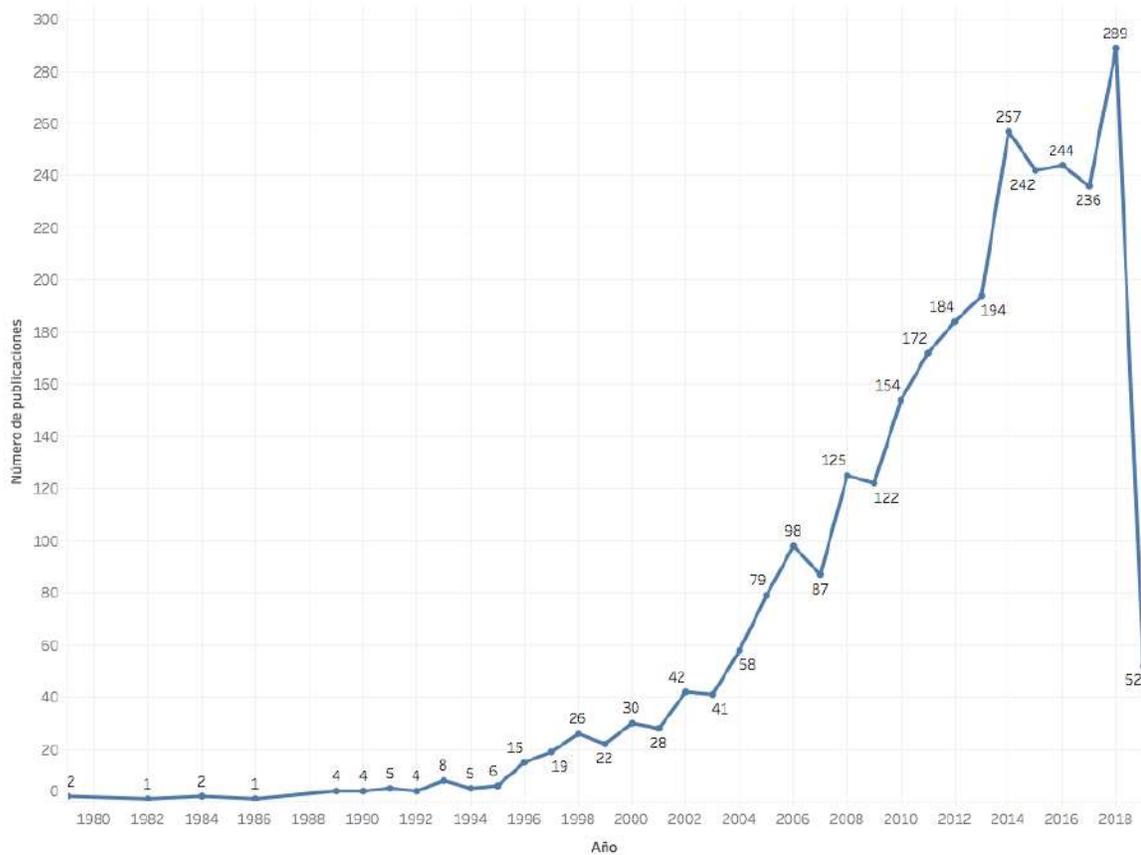


Figura 5.2 Producción científica de la UADY por años

Los repositorios de producción científica ofrecen información de utilidad, como es la institución de adscripción de cada autor en una publicación. A partir de esto se obtuvieron las cinco instituciones mexicanas (Tabla 5.6), los cinco países (Tabla 5.7) y las cinco instituciones extranjeras (Tabla 5.8) que más colaboraciones han realizado con la UADY (se define como colaboración a cada participación de un autor externo en una publicación) y las publicaciones, producto de dichas colaboraciones.

Tabla 5.6 Instituciones mexicanas con mayor colaboración con la UADY

<b>Posición</b>	<b>Institución</b>	<b>Colaboraciones</b>	<b>Publicaciones</b>
1	Universidad Nacional Autónoma de México	375	192
2	Centro de Investigación Científica de Yucatán	253	81
3	Centro de Investigación y de Estudios Avanzados- Unidad Mérida	226	113
4	Centro de Investigación y de Estudios Avanzados- Ciudad de México	208	94
5	Instituto Politécnico Nacional	175	89

Tabla 5.7 Países con mayor colaboración con la UADY

<b>Posición</b>	<b>País</b>	<b>Colaboraciones</b>	<b>Publicaciones</b>
1	Estados Unidos	1144	350
2	España	306	147
3	Francia	213	61
4	Reino Unido	206	94
5	Alemania	160	49

Tabla 5.8 Instituciones extranjeras con mayor colaboración con la UADY

<b>Posición</b>	<b>Institución</b>	<b>País</b>	<b>Colaboraciones</b>	<b>Publicaciones</b>
1	Colorado State University	Estados Unidos	136	33
2	Iowa State University	Estados Unidos	69	29
3	Université de Toulouse	Francia	56	8
4	Baylor College of Medicine	Estados Unidos	54	4
5	Michigan Technological University	Estados Unidos	45	15

A partir de estos resultados, se procede a realizar dos estudios:

- 1) Análisis de correlación de los indicadores propuestos en diversos escenarios en una facultad de la UADY
- 2) Uso de técnicas de minería de datos para el análisis del estado actual de la investigación en la UADY

En los estudios de correlación, se enumeran los indicadores que tienen una correlación fuerte y se analiza el motivo de acuerdo al contexto de la problemática. Se estudia el estado actual de la investigación en la Facultad de Matemáticas (FMAT), perteneciente al Campus de Ciencias Exactas e Ingenierías de la UADY. Así mismo se efectuaron análisis de correlación mediante el coeficiente de Spearman (Conover, 1999), para determinar si las métricas de la producción científica se relacionan directa o inversamente. Para ello se utilizó el paquete estadístico STATGRAPHICS Centurion 18 Versión 18.1.06 (Statgraphics, 2017), y las correlaciones fueron consideradas significativas cuando  $P < 0.05$ . En la Tabla 5.9 se presenta la notación de los indicadores utilizados para los estudios de correlación.

Tabla 5.9 Notación de los indicadores utilizados para los estudios de correlación

<b>Métrica</b>	<b>Descripción</b>
RIC	Relevancia en cuerpo académico
RICE	Relevancia en otras colaboraciones
RPI	Relevancia individual
PNI	Relevancia en publicaciones no indizadas
RDI	Relevancia del investigador
RCA	Relevancia del cuerpo académico
PCA	Producción en el cuerpo académico
PCE	Producción colaboración externa
PI	Producción individual
PNind	Producción no indizada
PTCA	Producción total del cuerpo académico
PTinv	Producción total investigador

En lo referente a la aplicación de técnicas de minería de datos se dan a conocer los resultados del uso de las técnicas de agrupamiento a partir de la discretización de los datos y de las reglas de asociación generadas, con la finalidad de caracterizar a los cuerpos académicos y a los investigadores con base en su relevancia científica.

## 5.1 CORRELACIÓN DE LOS INDICADORES PARA LOS CUERPOS ACADÉMICOS DE LA FMAT

Considerando los valores de las cuatro métricas sobre la producción científica de los cuerpos académicos de la FMAT, todas las correlaciones resultaron positivas y significativas (Tabla 5.10), a mayor valor de una métrica se incrementó el valor de la otra.

Tabla 5.10 Coeficientes de correlación de Spearman entre las métricas de la producción científica de los cuerpos académicos de la FMAT (n=8)

	PT	GCC	PTS
GCC	$r_s=0.9271$ $P=0.0142$		
PTS	$r_s=0.7807$ $P=0.0389$	$r_s=0.8810$ $P=0.0198$	
RCA	$r_s=0.8783$ $P=0.0201$	$r_s=0.9286$ $P=0.0140$	$r_s=0.9524$ $P=0.0117$

A partir de la Tabla 5.10, se observa que a mayor relevancia del cuerpo académico (*RCA*), mayor es la producción total del cuerpo académico, mayor el grado de colaboración (*GCC*) y mayor es la producción total indizada por Scopus (*PTS*).

## 5.2 CORRELACIÓN DE LOS INDICADORES PARA LA FMAT POR ÁREAS

Se determinó si las métricas de la producción científica de los investigadores del área de Computación (Tabla 5.11) y del área de Matemáticas (Tabla 5.12) de la FMAT se relacionan directa o inversamente, sin considerar si pertenecen o no a un cuerpo académico.

Tabla 5.11 Coeficientes de correlación de Spearman entre las métricas de la producción científica de los investigadores del área de Computación de la FMAT (n=42)

	RIC	RICE	RPI	PNI	RDI	RCA	PCA	PCE	PI	PNind	PTCA
RICE	0.2701										
	P=0.0838										
RPI	0.2214	0.1812									
	P=0.1562	P=0.2459									
PNI	-0.2786	0.3716	0.1997								
	P=0.0744	P=0.0173	P=0.2011								
RDI	0.5485	0.8866	0.1616	0.3098							
	P=0.0004	P<0.0001	P=0.3007	P=0.0473							
RCA	0.9128	0.3090	0.1917	-0.3127	0.5520						
	P<0.0001	P=0.0479	P=0.2196	P=0.0453	P=0.0004						
PCA	0.9023	0.3287	0.2862	-0.1901	0.5312	0.8255					
	P<0.0001	P=0.0353	P=0.0668	P=0.2236	P=0.0007	P<0.0001					
PCE	0.2491	0.9559	0.2287	0.4224	0.8471	0.2809	0.3490				
	P=0.1107	P<0.0001	P=0.1431	P=0.0068	P<0.0001	P=0.0721	P=0.0254				
PI	0.2214	0.1812	1.0000	0.1997	0.1616	0.1917	0.2862	0.2287			
	P=0.1562	P=0.2459	P<0.0001	P=0.2011	P=0.3007	P=0.2196	P=0.0668	P=0.1431			
PNind	-0.2832	0.3816	0.1461	0.9931	0.3146	-0.3156	-0.1943	0.4277	0.1461		
	P=0.0698	P=0.0145	P=0.3495	P<0.0001	P=0.0440	P=0.0433	P=0.2134	P=0.0062	P=0.3495		
PTCA	0.9148	0.2749	0.2669	-0.3110	0.5075	0.9369	0.8615	0.2589	0.2669	-0.3145	
	P<0.0001	P=0.0783	P=0.0874	P=0.0464	P=0.0012	P<0.0001	P<0.0001	P=0.0973	P=0.0874	P=0.0441	
PTinv	0.3864	0.8557	0.2414	0.5383	0.8716	0.3856	0.5170	0.9075	0.2414	0.5418	0.3631
	P=0.0134	P<0.0001	P=0.1222	P=0.0006	P<0.0001	P=0.0135	P=0.0009	P<0.0001	P=0.1222	P=0.0005	P=0.0201

Tabla 5.12 Coeficientes de correlación de Spearman entre las métricas de producción científica de investigadores del área de Matemáticas de la FMAT (n=53)

	RIC	RICE	RPI	PNI	RDI	RCA	PCA	PCE	PI	PNind	PTCA
RICE	0.2920										
	P=0.0352										
RPI	0.3246	0.2951									
	P=0.0192	P=0.0333									
PNI	-0.1467	0.1094	-0.1635								
	P=0.2901	P=0.4304	P=0.2385								
RDI	0.5486	0.8776	0.3329	0.1818							
	P=0.0001	P<0.0001	P=0.0164	P=0.1900							
RCA	0.9006	0.3061	0.4358	-0.1272	0.5426						
	P<0.0001	P=0.0273	P=0.0017	P=0.3589	P=0.0001						
PCA	0.9973	0.2942	0.3391	-0.1567	0.5458	0.9009					
	P<0.0001	P=0.0339	P=0.0145	P=0.2586	P=0.0001	P<0.0001					
PCE	0.2870	0.9675	0.3414	0.1561	0.8451	0.3030	0.2923				
	P=0.0385	P<0.0001	P=0.0138	P=0.2602	P<0.0001	P=0.0289	P=0.0351				
PI	0.3185	0.2860	0.9992	-0.1636	0.3272	0.4328	0.3334	0.3321			
	P=0.0216	P=0.0392	P<0.0001	P=0.2381	P=0.0183	P=0.0018	P=0.0162	P=0.0166			
PNind	-0.1358	0.1219	-0.1636	0.9980	0.1915	-0.1213	-0.1460	0.1704	-0.1637		
	P=0.3273	P=0.3792	P=0.2381	P<0.0001	P=0.1672	P=0.3818	P=0.2924	P=0.2193	P=0.2377		
PTCA	0.9006	0.3061	0.4358	-0.1272	0.5426	1.0000	0.9009	0.3030	0.4328	-0.1213	
	P<0.0001	P=0.0273	P=0.0017	P=0.3589	P=0.0001	P<0.0001	P<0.0001	P=0.0289	P=0.0018	P=0.3818	
PTinv	0.5838	0.7897	0.3805	0.3445	0.9372	0.5925	0.5849	0.8180	0.3753	0.3557	0.5925
	P<0.0001	P<0.0001	P=0.0061	P=0.0130	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P=0.0068	P=0.0103	P<0.0001

A partir de los resultados de las Tablas 5.11 y 5.12, se realiza una comparación entre las correlaciones importantes de los indicadores presentados a continuación.

#### *Relevancia del investigador en su CA (RIC)*

Tanto para el área de Computación como para el de Matemáticas,

- A mayor relevancia del investigador en su CA (*RIC*)
  - Mayor es la relevancia del investigador (*RDI*)
  - Mayor es la relevancia de su CA (*RCA*)
  - Mayor es su producción en el CA (*PCA*)
  - Mayor es la producción total de su CA (*PTCA*)
  - Mayor es su producción total (*PTinv*)

Además de lo expuesto previamente, en el área de Matemáticas, el indicador *RIC* tiene correlación importante con los indicadores que se presentan a continuación:

- A mayor relevancia del investigador en su CA (*RIC*)
  - Mayor es la relevancia del investigador en colaboración externa al CA (*RICE*)
  - Mayor es la relevancia de la producción individual (*RPI*)
  - Mayor es su producción en colaboración externa al CA (*PCE*)
  - Mayor es su producción individual (*PI*)

#### *Relevancia del investigador en colaboración externa al CA (RICE)*

Tanto para el área de Computación como para el de Matemáticas,

- A mayor relevancia del investigador en colaboración externa al CA (*RICE*)
  - Mayor es la relevancia del investigador (*RDI*)
  - Mayor es la relevancia de su CA (*RCA*)
  - Mayor es su producción en su CA (*PCA*)
  - Mayor es su producción en colaboración externa al CA (*PCE*)
  - Mayor es su producción total (*PTinv*)

Además, otros indicadores tienen correlación importante con el indicador *RICE* de acuerdo con el área de estudio.

Para el área de Computación

- A mayor relevancia del investigador en colaboración externa al CA (*RICE*)
  - Mayor es el peso de su producción no indizada (*PTPNI*)
  - Mayor es su producción no indizada (*PNind*)

Es decir, en el área de Computación, la relevancia del investigador en colaboraciones externas al cuerpo académico tiene una correlación importante con la producción y peso de las publicaciones no indizadas.

Para el área de Matemáticas

- A mayor relevancia del investigador en colaboración externa al CA (*RICE*)
  - Mayor es la relevancia del investigador dentro de su CA (*RIC*)

- Mayor es la relevancia de su producción individual (*RPI*)
- Mayor es su producción individual (*PI*)
- Mayor es la producción total de su cuerpo académico (*PTCA*)

En otras palabras, en el área de Matemáticas, la relevancia del investigador en colaboraciones externas al cuerpo académico tiene correlación importante con su relevancia dentro de su cuerpo académico y de sus publicaciones individuales, al igual que con su producción individual y la de su cuerpo académico.

#### *Relevancia de la producción individual del investigador (RPI)*

Tanto para el área de Computación como para el de Matemáticas, a mayor relevancia de la producción individual del investigador, mayor es su producción individual. Pero en el área de Matemáticas, dicho indicador tiene correlaciones importantes con ocho métricas más, quedando de la siguiente manera:

Para el área de Matemáticas

- A mayor relevancia de la producción individual del investigador (*RPI*)
  - Mayor es la relevancia del investigador dentro de su CA (*RIC*)
  - Mayor es su relevancia en colaboración externa al CA (*RICE*)
  - Mayor es su relevancia del investigador (*RDI*)
  - Mayor es la relevancia de su cuerpo académico (*RCA*)
  - Mayor es su producción dentro del CA (*PCA*)
  - Mayor es su producción en colaboración externa al cuerpo académico (*PCE*)
  - Mayor es la producción total de su cuerpo académico (*PTCA*)
  - Mayor es su producción total (*PTinv*)

#### *Relevancia del investigador (RDI)*

Tanto para el área de Computación como para el de Matemáticas

- A mayor relevancia del investigador (*RDI*)
  - Mayor es la relevancia del investigador dentro de su cuerpo académico (*RIC*)
  - Mayor es su relevancia en colaboración externa al CA (*RICE*)
  - Mayor es la relevancia de su cuerpo académico (*RCA*)

- Mayor es su producción dentro del CA (*PCA*)
- Mayor es su producción en colaboración externa al cuerpo académico (*PCE*)
- Mayor es la producción total de su cuerpo académico (*PTCA*)
- Mayor es su producción total (*PTinv*)

Además, otros indicadores tienen correlación importante con el indicador *RDI* de acuerdo con el área de estudio.

Para el área de Computación

- A mayor relevancia del investigador (*RDI*)
  - Mayor es el peso de su producción no indizada (*PNI*)
  - Mayor es su producción no indizada (*PNind*)

En otras palabras, en el área de Computación, la relevancia del investigador tiene correlación importante con el peso y la producción de las publicaciones no indizadas.

Para el área de Matemáticas

- A mayor relevancia del investigador (*RDI*)
  - Mayor es la relevancia de su producción individual (*RPI*)
  - Mayor es su producción individual (*PI*)

Es decir, en el área de Matemáticas, la relevancia del investigador tiene correlación importante con la relevancia y la producción de las publicaciones individuales.

*Relevancia del cuerpo académico de un investigador (RCA)*

Tanto para el área de Computación como para el de Matemáticas,

- A mayor relevancia del cuerpo académico de un investigador (*RCA*)
  - Mayor es la relevancia del investigador dentro de su cuerpo académico (*RIC*)
  - Mayor es su relevancia en colaboración externa al cuerpo académico (*RICE*)
  - Mayor es su relevancia del investigador (*RDI*)
  - Mayor es su producción dentro del cuerpo académico (*PCA*)
  - Mayor es la producción total de su cuerpo académico (*PTCA*)
  - Mayor es su producción total (*PTinv*)

Además, otros indicadores tienen correlación importante con el indicador *RCA* de acuerdo con el área de estudio.

Para el área de Computación

- A mayor relevancia del cuerpo académico de un investigador
  - Mayor es el peso de su producción no indizada
  - Mayor es su producción no indizada

Es decir, en el área de Computación, la relevancia del cuerpo académico de un investigador tiene correlación importante con el peso y la producción de las publicaciones no indizadas.

Para el área de Matemáticas

- A mayor relevancia del cuerpo académico de un investigador
  - Mayor es la relevancia de su producción individual
  - Mayor es su producción individual
  - Mayor es su producción en colaboración externa al CA

En otras palabras, en el área de Matemáticas, la relevancia del cuerpo académico de un investigador tiene correlación importante con la relevancia y producción de sus publicaciones individuales, y también con la producción en colaboraciones externas al cuerpo académico.

### 5.3 CORRELACIÓN DE LOS INDICADORES PARA LA FMAT PARA INVESTIGADORES CON CA

A continuación, se presentan los estudios de correlación entre indicadores de los investigadores que pertenecen a un cuerpo académico, tanto para el área de Computación (matriz en la Tabla 5.13) como de Matemáticas (matriz en la Tabla 5.14). Se estudian qué indicadores tienen mayor influencia sobre otros.

Tabla 5.13 Coeficientes de correlación de Spearman entre las métricas de producción científica de investigadores con cuerpo académico del área de Computación de la FMAT (n=19)

	RIC	RICE	RPI	PNI	RDI	RCA	PCA	PCE	PI	PNind	PTCA
RICE	0.5024										
	P=0.0330										
RPI	0.1723	0.2207									
	P=0.4648	P=0.3492									
PNI	0.3435	0.4093	0.6097								
	P=0.1451	P=0.0825	P=0.0097								
RDI	0.7738	0.8874	0.1722	0.4172							
	P=0.0010	P=0.0002	P=0.4650	P=0.0767							
RCA	0.6477	0.6054	0.0928	0.2964	0.7326						
	P=0.0060	P=0.0102	P=0.6937	P=0.2086	P=0.0019						
PCA	0.7523	0.5953	0.3296	0.5028	0.6677	0.4599					
	P=0.0014	P=0.0115	P=0.1620	P=0.0329	P=0.0046	P=0.0510					
PCE	0.4458	0.9136	0.3314	0.4545	0.8021	0.5363	0.6352				
	P=0.0586	P=0.0001	P=0.1597	P=0.0538	P=0.0007	P=0.0229	P=0.0070				
PI	0.1723	0.2207	1.0000	0.6097	0.1722	0.0928	0.3296	0.3314			
	P=0.4648	P=0.3492	P=0.0000	P=0.0097	P=0.4650	P=0.6937	P=0.1620	P=0.1597			
PNind	0.3425	0.4125	0.5086	0.9924	0.4274	0.3042	0.4956	0.4429	0.5086		
	P=0.1463	P=0.0801	P=0.0309	P=0.0000	P=0.0698	P=0.1969	P=0.0355	P=0.0602	P=0.0309		
PTCA	0.3931	0.5371	0.3944	0.4294	0.4585	0.2354	0.7577	0.5378	0.3944	0.4127	
	P=0.0953	P=0.0227	P=0.0943	P=0.0685	P=0.0518	P=0.3180	P=0.0013	P=0.0225	P=0.0943	P=0.0800	
PTinv	0.6280	0.8470	0.3478	0.5266	0.8351	0.6471	0.8439	0.9164	0.3478	0.5195	0.6503
	P=0.0077	P=0.0003	P=0.1401	P=0.0255	P=0.0004	P=0.0060	P=0.0003	P=0.0001	P=0.1401	P=0.0275	P=0.0058

Tabla 5.14 Coeficientes de correlación de Spearman entre las métricas de producción científica de investigadores con cuerpo académico del área de Matemáticas de la FMAT (n=19)

ICAAM	RIC	RICE	RPI	PNI	RDI	RCA	PCA	PCE	PI	PNind	PTCA
RICE	0.3777										
	P=0.1091										
RPI	-0.0941	0.4086									
	P=0.6896	P=0.0830									
PNI	-0.0443	0.0624	-0.2539								
	P=0.8510	P=0.7912	P=0.2815								
RDI	0.7095	0.8613	0.3418	-0.0414							
	P=0.0026	P=0.0003	P=0.1471	P=0.8605							
RCA	0.7734	0.5313	0.1041	-0.0247	0.7833						
	P=0.0010	P=0.0242	P=0.6589	P=0.9166	P=0.0009						
PCA	0.9605	0.3865	-0.0514	-0.1268	0.6896	0.7721					
	P=0.0000	P=0.1011	P=0.8275	P=0.5906	P=0.0034	P=0.0011					
PCE	0.2937	0.9039	0.4844	0.0901	0.7518	0.4136	0.3363				
	P=0.2128	P=0.0001	P=0.0399	P=0.7023	P=0.0014	P=0.0793	P=0.1536				
PI	-0.1497	0.3636	0.9927	-0.2557	0.2987	0.0629	-0.1023	0.4374			
	P=0.5254	P=0.1229	P=0.0000	P=0.2779	P=0.2051	P=0.7896	P=0.6642	P=0.0635			
PNind	0.0000	0.0958	-0.2550	0.9956	0.0000	-0.0135	-0.0816	0.1238	-0.2569		
	P=1.0000	P=0.6843	P=0.2794	P=0.0000	P=1.0000	P=0.9544	P=0.7293	P=0.5994	P=0.2758		
PTCA	0.7734	0.5313	0.1041	-0.0247	0.7833	1.0000	0.7721	0.4136	0.0629	-0.0135	
	P=0.0010	P=0.0242	P=0.6589	P=0.9166	P=0.0009	P=0.0000	P=0.0011	P=0.0793	P=0.7896	P=0.9544	
PTinv	0.6476	0.7841	0.4291	0.0125	0.9062	0.7331	0.6827	0.8429	0.3862	0.0532	0.7331
	P=0.0060	P=0.0009	P=0.0687	P=0.9575	P=0.0001	P=0.0019	P=0.0038	P=0.0003	P=0.1013	P=0.8214	P=0.0019

A partir de los resultados de las Tablas 5.13 y 5.14, se realiza una comparación entre las correlaciones importantes entre los indicadores.

#### *Relevancia del investigador en su CA (RIC)*

Tanto para el área de Computación como para el de Matemáticas,

- A mayor relevancia del investigador en su CA (*RIC*)
  - Mayor es la relevancia del investigador (*RDI*)
  - Mayor es la relevancia de su CA (*RCA*)
  - Mayor es su producción en el CA (*PCA*)
  - Mayor es su producción total (*PTinv*)

En ambas áreas existen indicadores que tienen correlación importante con la relevancia del investigador en su cuerpo académico, quedando de la siguiente manera.

Para el área de Computación, a mayor relevancia del investigador en su cuerpo académico (*RIC*), mayor es la relevancia de su producción en colaboración externa a su cuerpo académico (*RICE*).

Para el área de Matemáticas, a mayor relevancia del investigador en su cuerpo académico (*RIC*), mayor es la producción total de su cuerpo académico (*PTCA*).

*Relevancia del investigador en colaboración externa al CA (RICE)*

Tanto para el área de Computación como para el de Matemáticas

- A mayor relevancia del investigador en colaboración externa al cuerpo académico (*RICE*)
  - Mayor es la relevancia del investigador (*RDI*)
  - Mayor es la relevancia de su cuerpo académico (*RCA*)
  - Mayor es la producción total de su cuerpo académico (*PTCA*)
  - Mayor es su producción total (*PTinv*)

Además, otros indicadores tienen correlación importante con el indicador *RICE* de acuerdo con el área de estudio.

En el área de Computación

- A mayor relevancia del investigador en colaboración externa al cuerpo académico (*RICE*)
  - Mayor es la relevancia del investigador dentro de su cuerpo académico (*RIC*)
  - Mayor es su producción dentro del cuerpo académico (*PCA*)

Es decir, en el área de Computación, la relevancia del investigador en colaboraciones externas a su cuerpo académico tiene correlación importante con su relevancia y producción dentro del cuerpo académico.

En el área de Matemáticas, a mayor relevancia del investigador en colaboración externa al cuerpo académico (*RICE*), mayor es su producción en colaboraciones externas al cuerpo académico (*PCE*).

### *Relevancia de la producción individual del investigador (RPI)*

En ambas áreas, a mayor relevancia de la producción individual del investigador (*RPI*), mayor es su producción individual (*PI*). Además, se debe tomar en cuenta lo siguiente:

En el área de Computación

- A mayor relevancia de la producción individual del investigador (*RPI*)
  - Mayor es el peso de su producción no indizada (*PNI*)
  - Mayor es la producción no indizada (*PNind*)

Es decir, en el área de Computación, la relevancia de la producción individual del investigador tiene correlación importante con el peso y la producción de publicaciones no indizadas.

En el área de Matemáticas, a mayor relevancia de la producción individual del investigador (*RPI*), mayor es su producción en colaboración externa al cuerpo académico (*PCE*).

### *Peso de la producción no indizada (PNI)*

En ambas áreas, a mayor peso de la producción no indizada (*PNI*), mayor producción no indizada (*PNind*). Esta afirmación se puede decir que es natural, ya que es de esta producción que se calcula el indicador.

Sin embargo, en el área de Computación otros indicadores tienen correlación con el peso de la producción no indizada, y se presentan a continuación.

Para el área de Computación

- A mayor peso de la producción no indizada (*PNI*)
  - Mayor es la relevancia de su producción individual (*RPI*)
  - Mayor es su producción dentro de su cuerpo académico (*PCA*)
  - Mayor es su producción individual (*PI*)
  - Mayor es su producción total (*PTinv*)

En otras palabras, en el área de Computación, el peso de la producción no indizada tiene correlación importante con la relevancia de la producción individual, las publicaciones que el investigador ha realizado dentro de su cuerpo académico y de forma individual, así como con su producción total.

### *Relevancia del investigador (RDI)*

Tanto para el área de Computación como para el de Matemáticas

- A mayor relevancia del investigador (*RDI*)
  - Mayor es la relevancia del investigador dentro de su cuerpo académico (*RIC*)
  - Mayor es su relevancia en colaboración externa al cuerpo académico (*RICE*)
  - Mayor es la relevancia de su cuerpo académico (*RCA*)
  - Mayor es su producción dentro del cuerpo académico (*PCA*)
  - Mayor es su producción en colaboración externa al cuerpo académico (*PCE*)
  - Mayor es su producción total (*PTinv*)

Además, por parte del área de Matemáticas, a mayor relevancia del investigador (*RDI*), mayor es la producción total de su cuerpo académico (*PTCA*).

### *Relevancia del cuerpo académico de un investigador (RCA)*

Tanto para el área de Computación como para el de Matemáticas,

- A mayor relevancia del cuerpo académico de un investigador (*RCA*)
  - Mayor es la relevancia del investigador dentro de su cuerpo académico (*RIC*)
  - Mayor es su relevancia en colaboración externa al cuerpo académico (*RICE*)
  - Mayor es su relevancia del investigador (*RDI*)
  - Mayor es su producción total (*PTinv*)

En ambas áreas existen otros indicadores que se correlacionan de modo importante con el indicador de la relevancia del cuerpo académico, quedando de la siguiente manera:

Para el área de Computación, a mayor relevancia del cuerpo académico de un investigador (*RCA*), mayor es su producción en colaboraciones externas al cuerpo académico (*PCE*). Y para el área de Matemáticas

- A mayor relevancia del cuerpo académico de un investigador (*RCA*)
  - Mayor es su producción dentro del cuerpo académico (*PCA*)
  - Mayor es la producción total de su cuerpo académico (*PTCA*)

Es decir, en el área de Matemáticas, el indicador de la relevancia del cuerpo académico de un investigador tiene correlación importante con la producción del investigador dentro de su cuerpo académico y con la producción total del cuerpo académico.

#### 5.4 CORRELACIÓN DE LOS INDICADORES PARA LA FMAT PARA INVESTIGADORES SIN CA

Para estos análisis estadísticos, se omitieron las métricas relacionadas con la producción y la relevancia de los cuerpos académicos. Las matrices de correlación se presentan en la Tabla 5.15 y Tabla 5.16, correspondientes a las áreas de Computación y Matemáticas, respectivamente.

Tabla 5.15 Coeficientes de correlación de Spearman entre las métricas de producción científica de investigadores sin cuerpo académico del área de Computación de la FMAT (n=23)

IsinCAAC	RICE	PNI	RDI	PCE	PNind
PNI	0.5840				
	P=0.0062				
RDI	0.9632	0.6987			
	P<0.0001	P=0.0010			
PCE	0.9689	0.6227	0.9375		
	P<0.0001	P=0.0035	P<0.0001		
PNind	0.5991	0.9902	0.7005	0.6400	
	P=0.0050	P<0.0001	P=0.0010	P=0.0027	
PTinv	0.8672	0.8633	0.9340	0.9018	0.8741
	P<0.0001	P=0.0001	P<0.0001	P<0.0001	P<0.0001

Tabla 5.16 Coeficientes de correlación de Spearman entre las métricas de producción científica de investigadores sin cuerpo académico del área de Matemáticas de la FMAT (n=34)

<b>IsinCAAM</b>	RICE	PNI	RDI	PCE	PNind
PNI	0.1801				
	P=0.3009				
RDI	0.9253	0.4415			
	P<0.0001	P=0.0112			
PCE	0.9825	0.2267	0.9091		
	P<0.0001	P=0.1929	P<0.0001		
PNind	0.1943	0.9960	0.4466	0.2441	
	P=0.2644	P<0.0001	P=0.0103	P=0.1608	
PTinv	0.8165	0.6329	0.9508	0.8411	0.6429
	P<0.0001	P=0.0003	P<0.0001	P<0.0001	P=0.0002

*Relevancia de la producción en colaboración (RICE)*

Tanto para el área de Computación como para el de Matemáticas

- A mayor relevancia de la producción en colaboración de un investigador que no pertenece a un cuerpo académico (*RICE*)
  - Mayor es su relevancia del investigador (*RDI*)
  - Mayor es su producción en colaboración (*PCE*)
  - Mayor es su producción total (*PTinv*)

En el área de Computación también existe correlación significativa con otros dos indicadores, quedando de la siguiente manera:

Para el área de Computación

- A mayor relevancia de la producción en colaboración de un investigador que no pertenece a un cuerpo académico (*RICE*)
  - Mayor es el peso de su producción no indizada (*PNI*)
  - Mayor es su producción no indizada (*PNind*)

Es decir, en el área de Computación, la relevancia de la producción en colaboración de un investigador que no pertenece a un cuerpo académico tiene correlación importante con el peso y la producción de sus publicaciones no indizadas.

#### *Peso de la producción no indizada (PNI)*

Tanto para el área de Computación como para el área de Matemáticas

- A mayor peso de la producción no indizada de un investigador que no pertenece a un cuerpo académico (*PNI*)
  - Mayor es su producción en colaboración (*PCE*)
  - Mayor es su producción no indizada (*PNind*)
  - Mayor es su producción total (*PTinv*)

Al igual que el caso anterior, en el área de Computación, el indicador se correlaciona con dos métricas más.

Para el área de Computación

- A mayor peso de la producción no indizada de un investigador que no pertenece a un cuerpo académico
  - Mayor es la relevancia de su producción en colaboración
  - Mayor es su relevancia del investigador

Entonces, en el área de Computación, el peso de la producción no indizada de un investigador que no pertenece a un cuerpo académico tiene correlación importante con su relevancia en publicaciones en colaboración y con su relevancia del investigador.

#### *Relevancia del investigador (RDI)*

Tanto para el área de Computación y de Matemáticas

- A mayor relevancia de un investigador que no pertenece a un cuerpo académico (*RDI*)
  - Mayor es la relevancia de su producción en colaboración (*RICE*)
  - Mayor es el peso de su producción no indizada (*PNI*)
  - Mayor es su producción en colaboración (*PCE*)
  - Mayor es su producción no indizada (*PNind*)
  - Mayor es su producción total (*PTinv*)

## 5.5 CORRELACIÓN DE LOS INDICADORES EN CADA CA DE COMPUTACIÓN

En la Facultad de Matemáticas de la UADY existen tres cuerpos académicos pertenecientes al área de Computación. Cada uno de ellos se analizó para localizar la correlación de los indicadores considerando a cada uno de sus integrantes.

### ***Cuerpo Académico de Computación 1 (CAC1)***

Para investigadores del CAC1 las métricas *RPI* y *PI* tienen valor cero en todos los registros ( $n = 5$ ), además *RCA* es igual a 828.2174 en todos los registros y *PTCA* es igual a 12 (Figura 5.3).

Por lo que sólo se consideraron las otras ocho en los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *PNI* (peso de la producción no indizada del investigador)
- 4) *RDI* (relevancia del investigador)
- 5) *PCA* (producción del investigador en su cuerpo académico)
- 6) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 7) *PNind* (producción no indizada del investigador)
- 8) *PTinv* (producción total del investigador)

Sin embargo, ningún par resultó significativo ( $P > 0.05$ ).

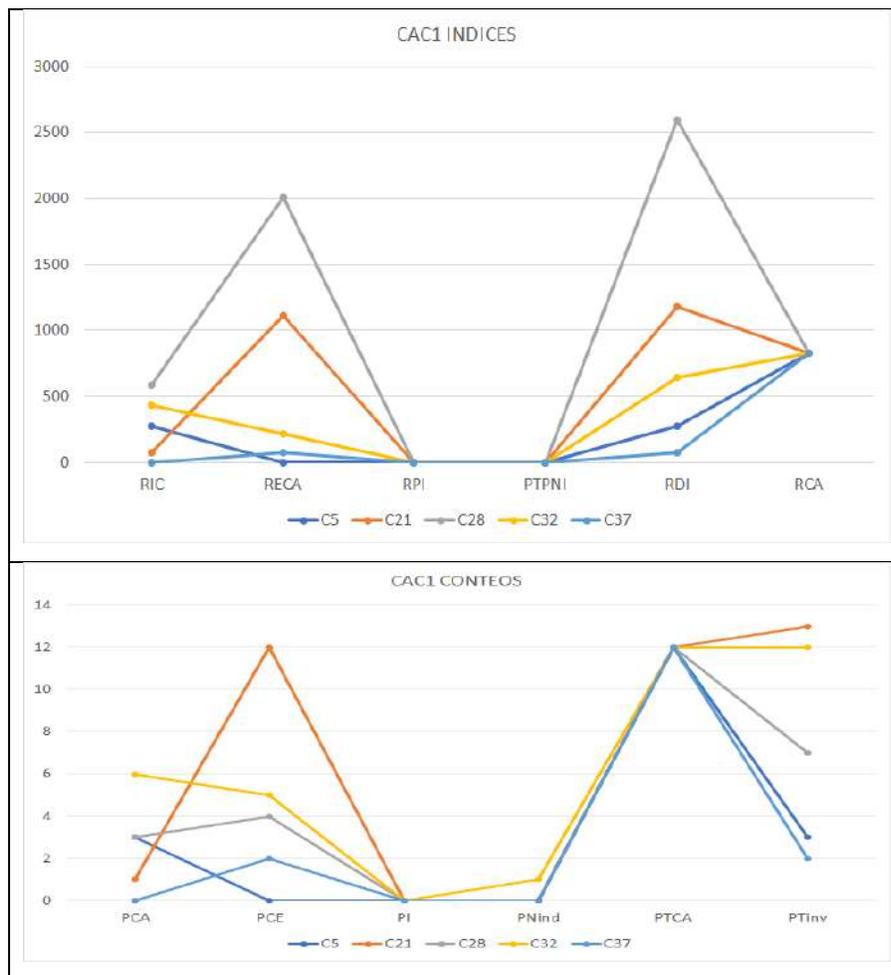


Figura 5.3 Métricas de la producción científica de investigadores del CAC1 de la FMAT

### ***Cuerpo Académico de Computación 2 (CAC2)***

Para investigadores del CAC2, *RCA* es igual a 288.4142 en todos los registros ( $n = 5$ ) y *PTCA* es igual a 20 (Figura 5.4). Por lo que sólo se consideraron los otros diez indicadores en los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *RPI* (relevancia de la producción individual del investigador)
- 4) *PNI* (peso de la producción no indizada del investigador)
- 5) *RDI* (relevancia del investigador)

- 6) *PCA* (producción del investigador en su cuerpo académico)
- 7) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 8) *PI* (producción individual del investigador)
- 9) *PNind* (producción no indizada del investigador)
- 10) *PTinv* (producción total del investigador)

De todos los pares posibles, dos resultaron significativos:

- 1) A mayor valor de *RICE*, mayor es el valor de *RDI* ( $r_s=1$ ,  $P < 0.0001$ )
- 2) A mayor valor de *PCE*, mayor es el valor de *PTinv* ( $r_s=1$ ,  $P < 0.0001$ )

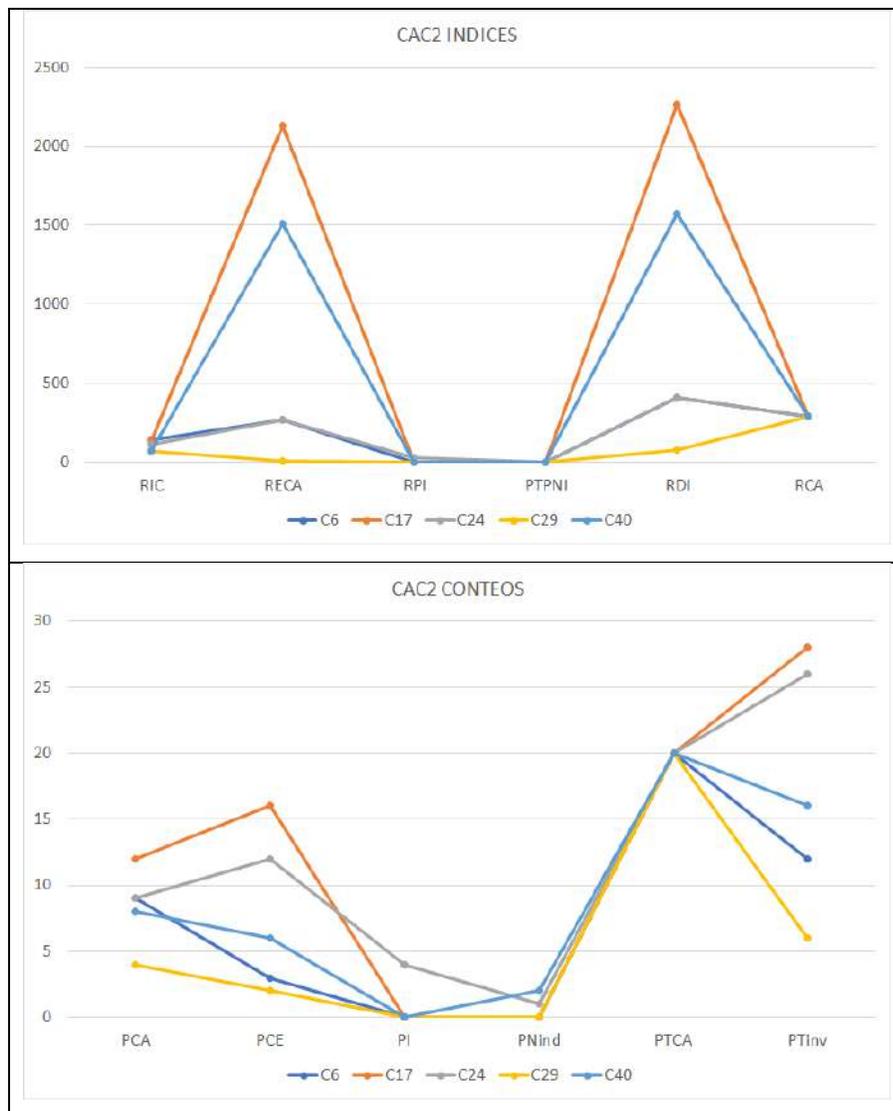


Figura 5.4 Métricas de la producción científica de investigadores del CAC2 de la FMAT

### ***Cuerpo Académico de Computación 3 (CAC3)***

Para investigadores del CAC3 las métricas *RPI*, *PNI*, *PI* y *PNind* tienen valor cero en todos los registros ( $n = 9$ ), además *RCA* es igual a 50.5710 en todos los registros y *PTCA* es igual a 12 (Figura 5.5). Por lo que sólo se consideraron los otros seis indicadores en los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *RDI* (relevancia del investigador)
- 4) *PCA* (producción del investigador en su cuerpo académico)
- 5) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 6) *PTinv* (producción total del investigador)

De todos los pares posibles, tres resultaron significativos:

- 1) A mayor valor de *RICE*, mayor es el valor de *PCE* ( $r_s=0.9586$ ,  $P = 0.0067$ )
- 2) A mayor valor de *PCA*, mayor es el valor de *PTinv* ( $r_s = 0.8206$ ,  $P = 0.0203$ )
- 3) A mayor valor de *PCE*, mayor es el valor de *PTinv* ( $r_s = 0.7457$ ,  $P = 0.0349$ )

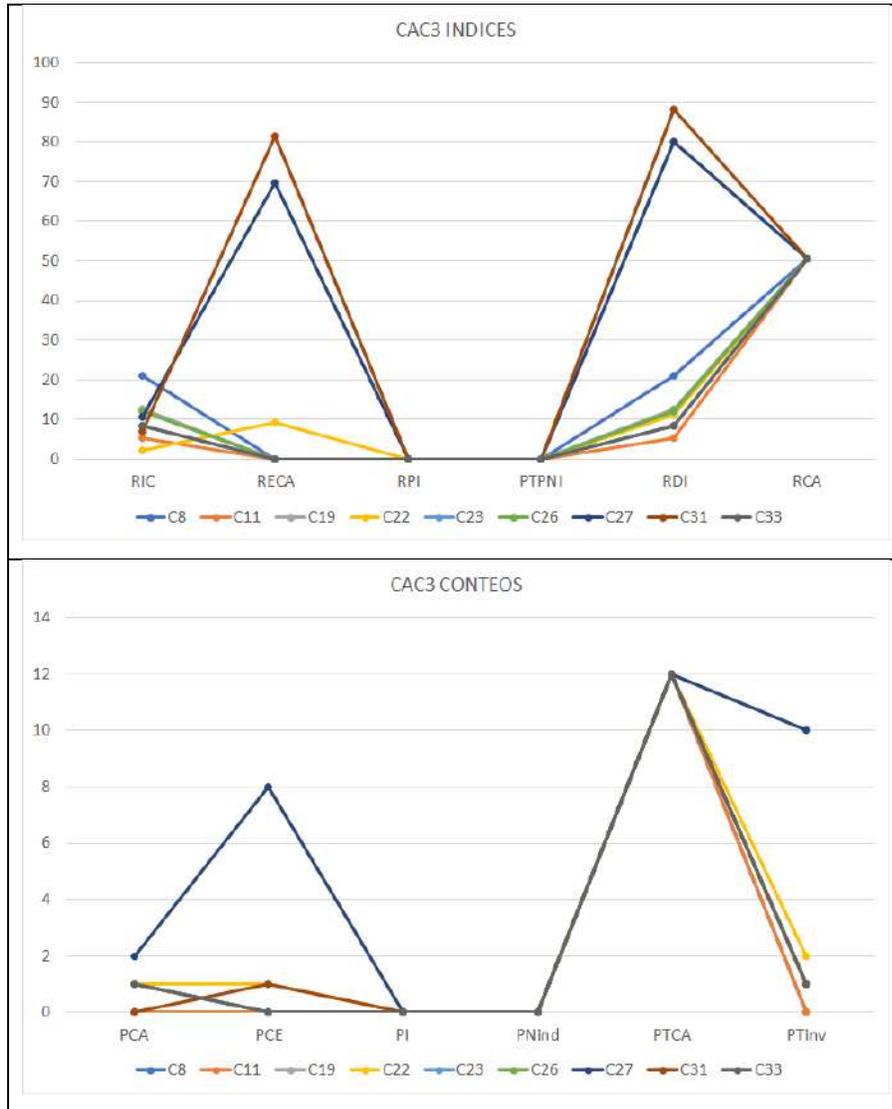


Figura 5.5 Métricas de la producción científica de investigadores del CAC3 de la FMAT

A continuación se presenta la distribución de las métricas *RCA* y *PTCA*, cuyos valores son iguales para cada uno de los integrantes del cuerpo académico correspondiente. Es de resaltar que el CAC1 presentó mayor relevancia, sin embargo, su producción total fue igual al del CAC3, y el que presentó la mayor producción total fue CAC2 (Figura 5.6).

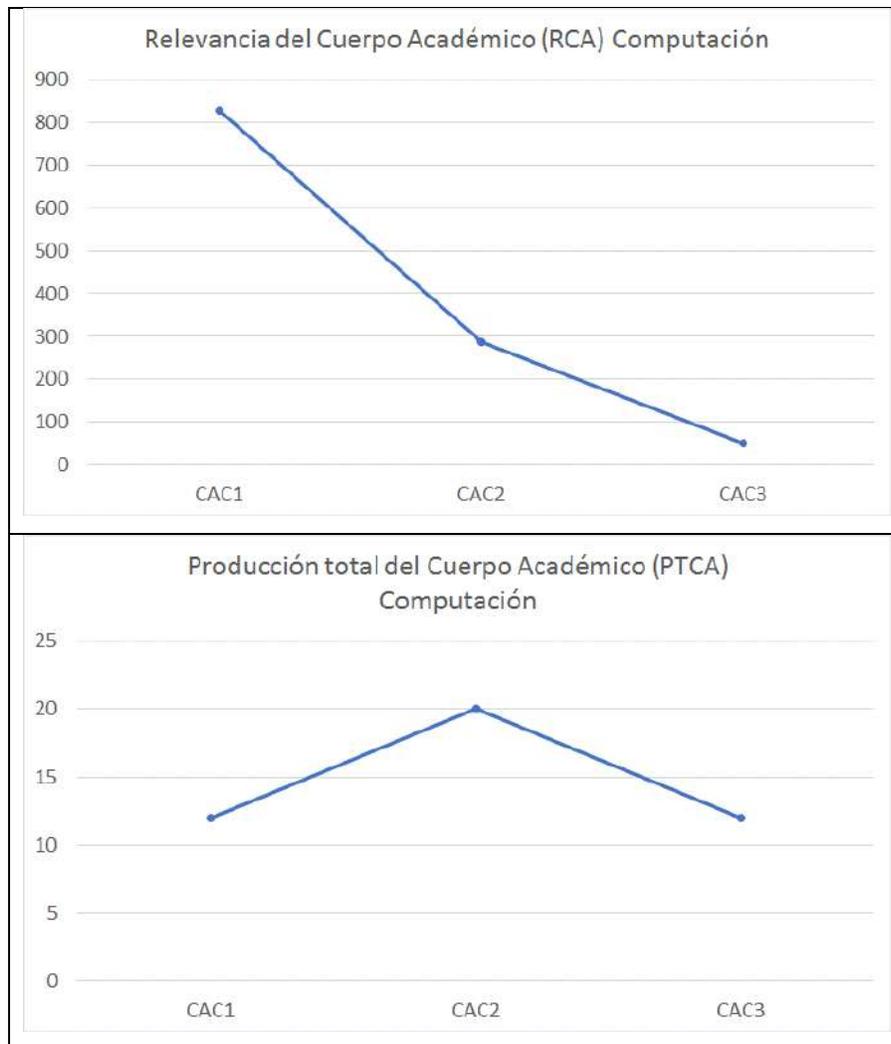


Figura 5.6 Relevancia y producción total de los cuerpos académicos del área de Computación de la FMAT

## 5.6 CORRELACIÓN DE LOS INDICADORES EN CADA CA DE MATEMÁTICAS

En la Facultad de Matemáticas de la UADY existen cinco cuerpos académicos pertenecientes al área de Matemáticas. Cada uno de ellos se analizó para localizar la correlación de los indicadores considerando a cada uno de sus integrantes. Debido al tamaño de muestra pequeño se utilizaron los valores críticos (Zar, 2010) para la significancia de las correlaciones. Como el mínimo tamaño de muestra para los valores críticos es cuatro, las métricas de los CAM1 y CAM2 sólo fueron descritas.

### ***Cuerpo Académico de Matemáticas 1 (CAM1)***

Para investigadores del CAM1 las métricas *RICE*, *RPI*, *PNI*, *PCE*, *PI* y *PNind* tienen valor cero en todos los registros ( $n = 3$ ), además *RCA* es igual a 7.8550 en todos los registros y *PTCA* es igual a 1 (Figura 5.7), por lo que sólo se consideraron los otros cuatro indicadores en los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RDI* (relevancia del investigador)
- 3) *PCA* (producción del investigador en su cuerpo académico)
- 4) *PTinv* (producción total del investigador)

De todos los pares posibles, dos resultaron significativos:

- 1) A mayor valor de *RIC*, mayor es el valor de *RDI*
- 2) A mayor valor de *PCA*, mayor es el valor de *PTind*

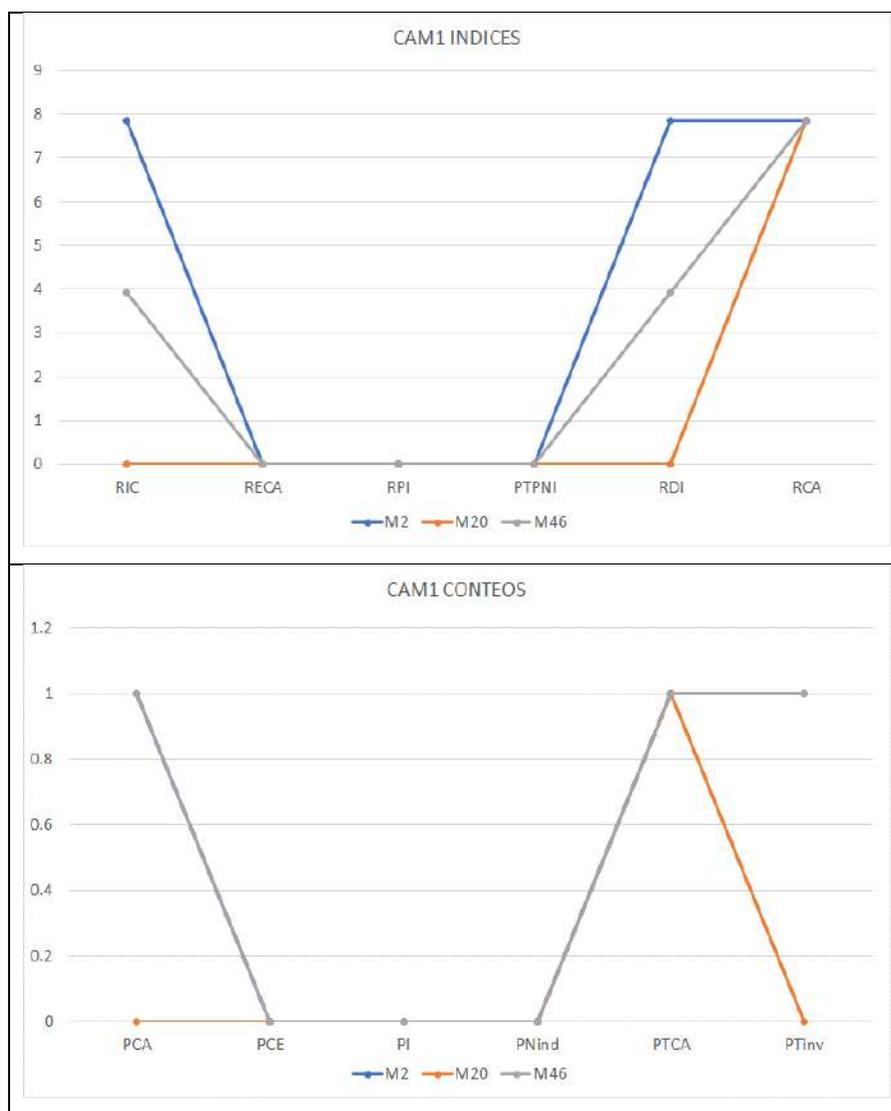


Figura 5.7 Métricas de la producción científica de investigadores del CAM1 de la FMAT

### ***Cuerpo Académico de Matemáticas 2 (CAM2)***

Para investigadores del CAM2 las métricas *RPI* y *PI* tienen valor cero en todos los registros ( $n = 3$ ), además *RCA* es igual a 86.5920 en todos los registros y *PTCA* es igual a 10 (Figura 5.8). Se consideraron los siguientes indicadores para los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *PNI* (peso de la producción no indizada del investigador)
- 4) *RDI* (relevancia del investigador)

- 5) *PCA* (producción del investigador en su cuerpo académico)
- 6) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 7) *PNind* (producción no indizada del investigador)
- 8) *PTinv* (producción total del investigador)

A continuación se observan las relaciones obtenidas.

- 1) A mayor valor de *RIC*, mayores son los valores de: i) *PNI* y ii) *PNind*.
- 2) A mayor valor de *RICE*, mayores son los valores de: i) *RDI*, ii) *PCA*, iii) *PCE* y iv) *PTinv*.
- 3) A mayor valor de *RDI*, mayores son los valores de: i) *PCA*, ii) *PCE* y iii) *PTinv*.
- 4) A mayor valor de *PCA*, mayores son los valores de: i) *PCE* y ii) *PTinv*.
- 5) A mayor valor de *PCE*, mayor es el valor de *PTinv*.

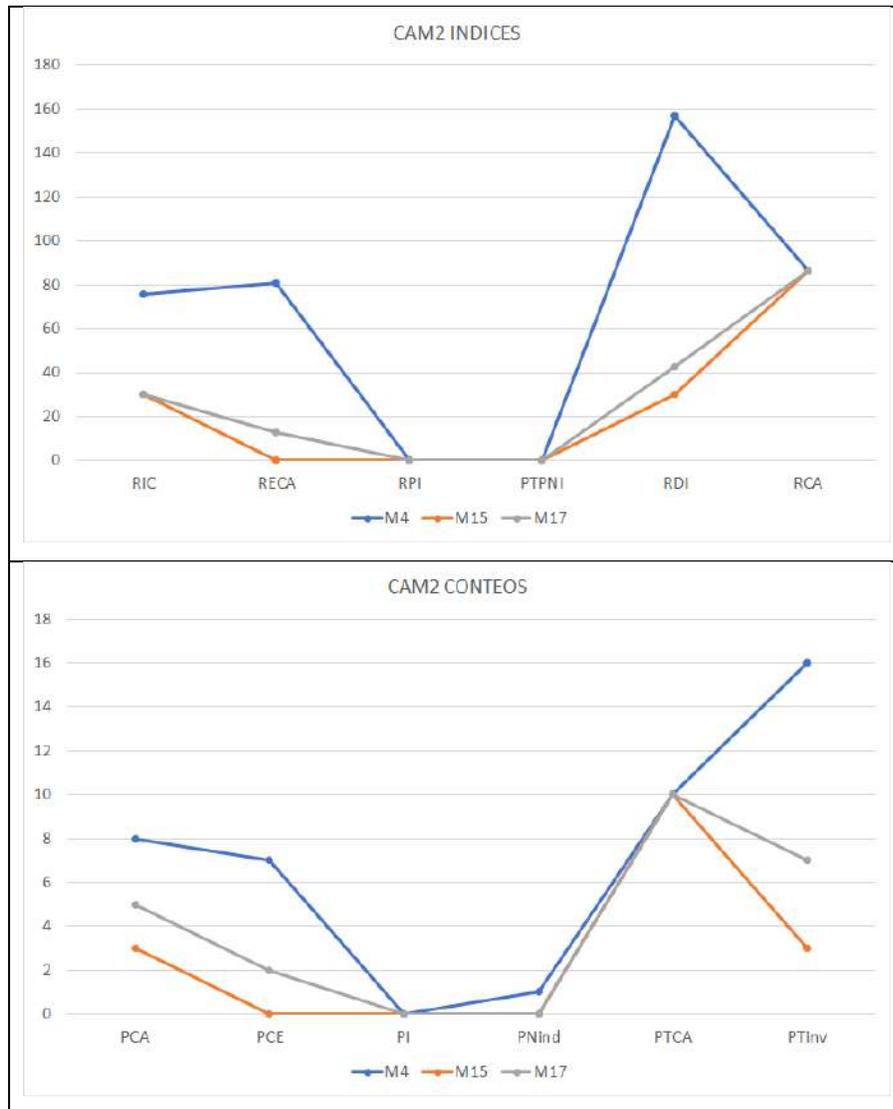


Figura 5.8 Métricas de la producción científica de investigadores del CAM2 de la FMAT

### ***Cuerpo Académico de Matemáticas 3 (CAM3)***

Para investigadores del CAM3 las métricas *PNI* y *PNind* tienen valor cero en todos los registros ( $n = 4$ ), además *RCA* es igual a 141.5260 en todos los registros y *PTCA* es igual a 12 (Figura 5.9). Por lo que sólo se consideraron los otros ocho indicadores en los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *RPI* (relevancia de la producción individual del investigador)
- 4) *RDI* (relevancia del investigador)

- 5) *PCA* (producción del investigador en su cuerpo académico)
- 6) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 7) *PI* (producción individual del investigador)
- 8) *PTinv* (producción total del investigador)

De todos los pares posibles, sólo uno resultó significativo: a mayor valor de *RICE*, mayor es el valor de *RDI* ( $r_s=1$ ,  $P = 0.05$ ).

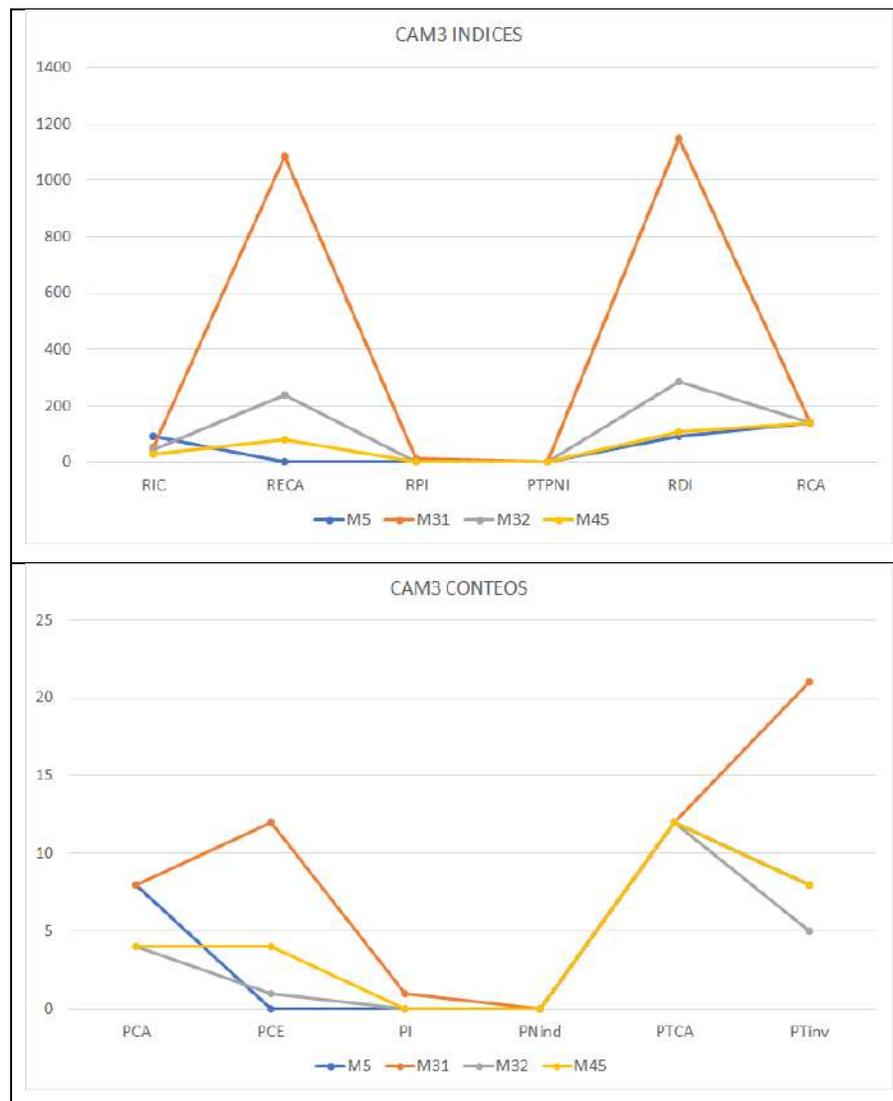


Figura 5.9 Métricas de la producción científica de investigadores del CAM3 de la FMAT

#### ***Cuerpo Académico de Matemáticas 4***

Para investigadores del CAM4 la métrica *PCA* tiene valor uno en todos los registros ( $n = 4$ ), además *RCA* es igual a 12.3840 en todos los registros y *PTCA* es igual a 2 (Figura 5.10). Por lo que sólo se consideraron los otras nueve indicadores en los análisis de correlación, y se enlistan a continuación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *RPI* (relevancia de la producción individual del investigador)
- 4) *PNI* (peso de la producción no indizada del investigador)
- 5) *RDI* (relevancia del investigador)
- 6) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 7) *PI* (producción individual del investigador)
- 8) *PNind* (producción no indizada del investigador)
- 9) *PTinv* (producción total del investigador)

De todos los pares posibles, dos resultaron significativos:

- 1) A mayor valor de *RICE*, mayor es el valor de *PCE* ( $r_s=1, P = 0.05$ )
- 2) A mayor valor de *RDI*, mayor es el valor de *PTinv* ( $r_s=1, P = 0.05$ )

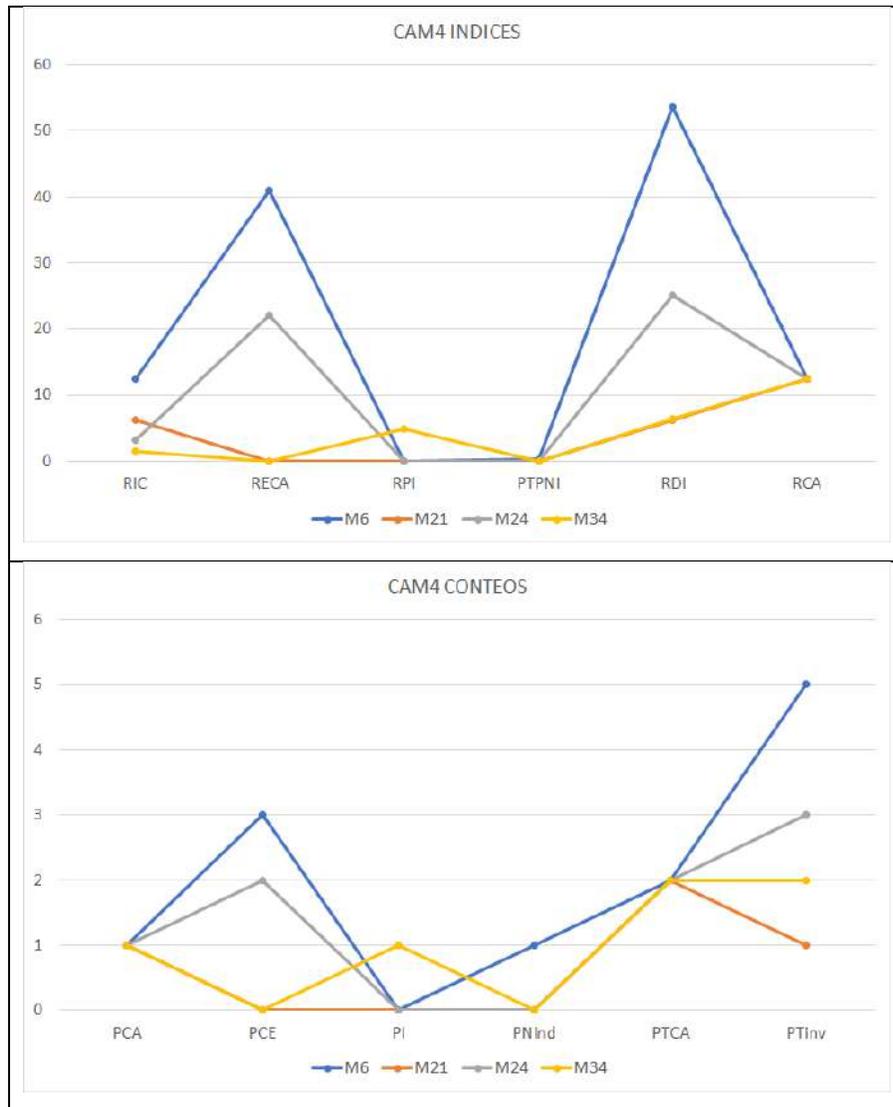


Figura 5.10 Métricas de la producción científica de investigadores del CAM4 de la FMAT

### ***Cuerpo Académico de Matemáticas 5 (CAM5)***

Para investigadores del CAM5, *RCA* es igual a 25.6460 en todos los registros ( $n = 5$ ) y *PTCA* es igual a 3 (Figura 5.11). Por lo que sólo se consideraron los otros diez indicadores en los análisis de correlación:

- 1) *RIC* (relevancia del investigador en su cuerpo académico)
- 2) *RICE* (relevancia del investigador en colaboraciones externas al cuerpo académico)
- 3) *RPI* (relevancia de la producción individual del investigador)
- 4) *PNI* (peso de la producción no indizada del investigador)

- 5) *RDI* (relevancia del investigador)
- 6) *PCA* (producción del investigador en su cuerpo académico)
- 7) *PCE* (producción del investigador en colaboraciones externas al cuerpo académico)
- 8) *PI* (producción individual del investigador)
- 9) *PNind* (producción no indizada del investigador)
- 10) *PTinv* (producción total del investigador)

De todos los pares posibles, sólo uno resultó significativo: a mayor valor de *RICE*, mayor es el valor de *RDI* ( $r_s = 1$ ,  $P = 0.01$ ).

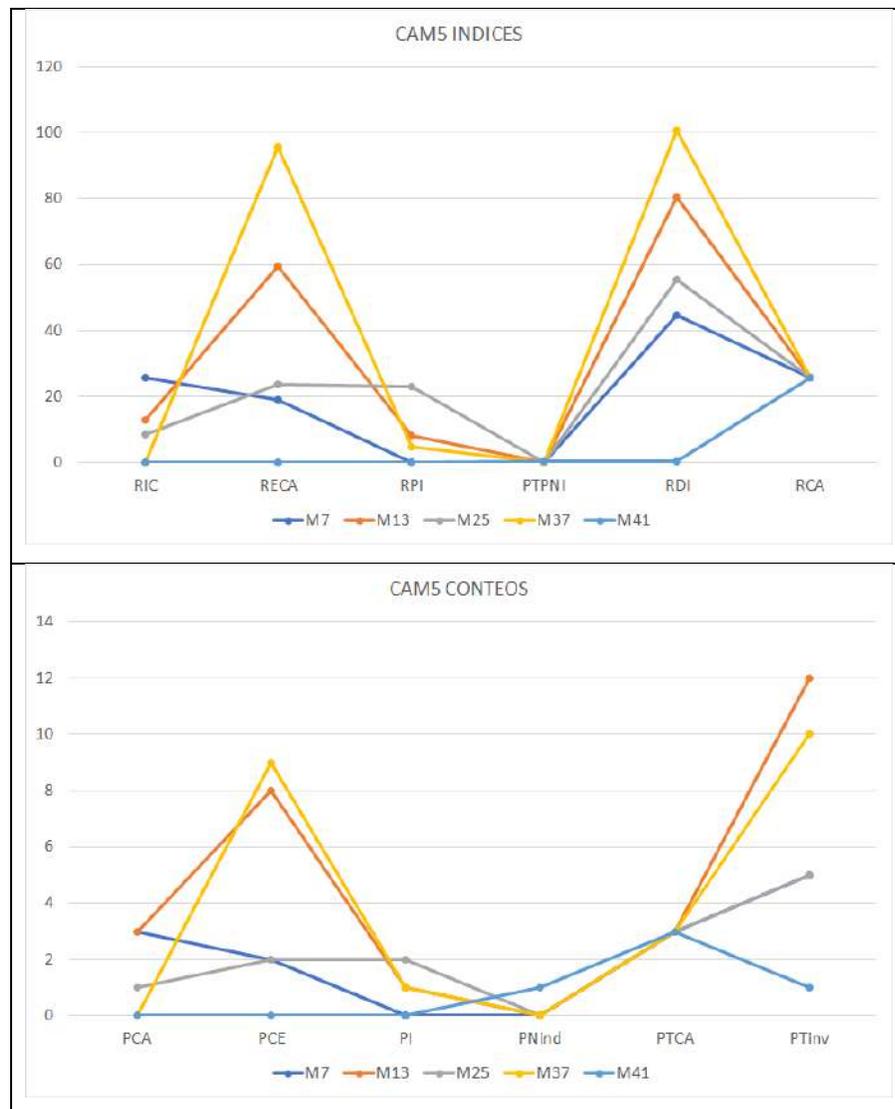


Figura 5.11 Métricas de la producción científica de investigadores del CAM5 de la FMAT

A continuación, se presenta la distribución de las métricas *RCA* y *PTCA* de los cuerpos académicos de Matemáticas, cuyos valores son iguales para cada uno de los integrantes del cuerpo académico correspondiente. Cabe destacar que ambas métricas presentan el mismo comportamiento  $CAM1 < CAM4 < CAM5 < CAM2 < CAM3$  (Figura 5.12).

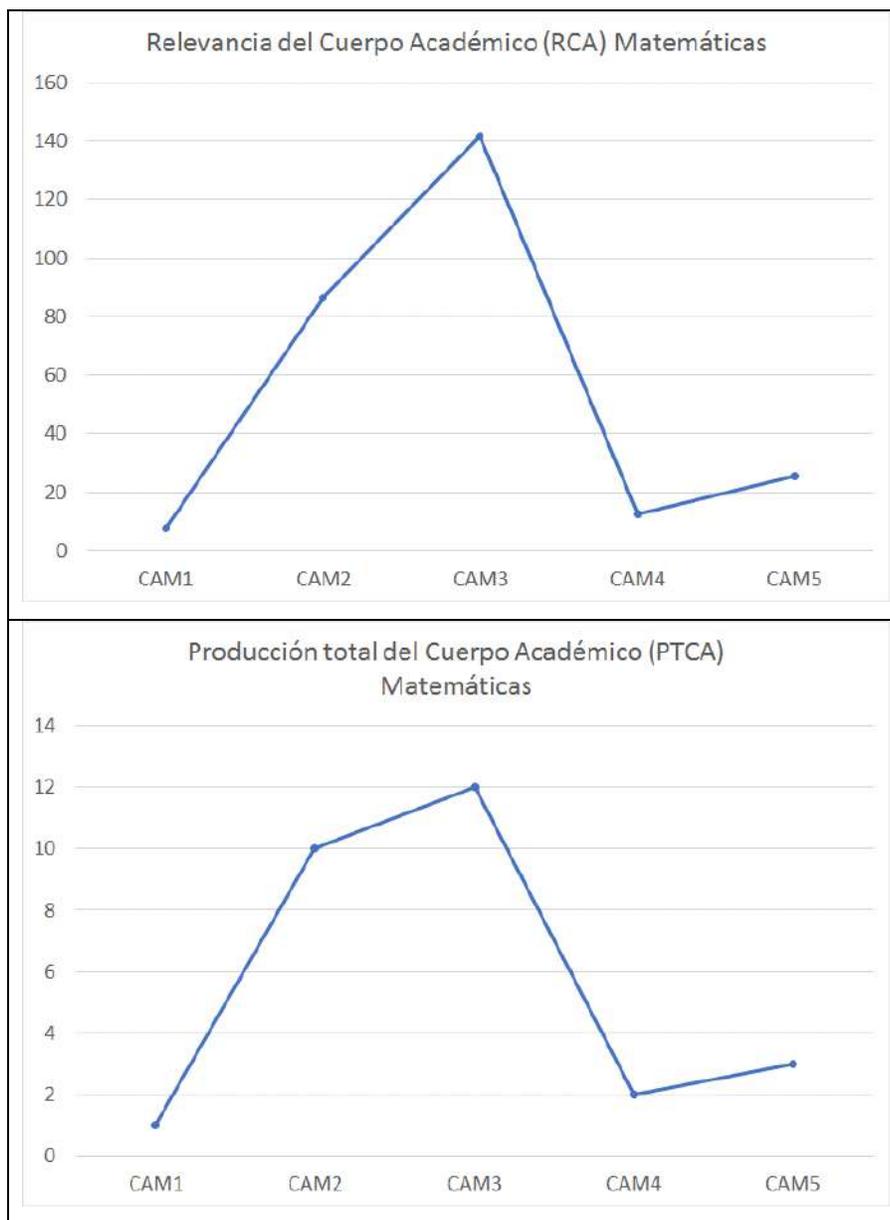


Figura 5.12 Relevancia y producción total de los cuerpos académicos del área de Matemáticas de la FMAT

## 5.7 ANÁLISIS DE LA RELEVANCIA CIENTÍFICA A PARTIR DE TÉCNICAS DE MINERÍA DE DATOS

### *Cuerpos académicos*

Los 78 cuerpos académicos de la UADY están consolidados de acuerdo con lo establecido por los lineamientos del PRODEP y los resultados de su última evaluación. El PRODEP considera la producción científica como la forma óptima de presentar el trabajo en conjunto de innovación (PRODEP, 2017).

En la Tabla 5.17 se describe el estado de la producción y relevancia de los cuerpos académicos. Se observa que en todos los casos, los valores máximos se encuentran muy distantes del promedio.

Tabla 5.17 Estado de la producción y relevancia de los cuerpos académicos

Número mínimo de integrantes	3
Promedio de integrantes	4.74
Número máximo de integrantes	11
Número mínimo de producción indizada	0
Promedio de producción indizada	10.88
Número máximo de producción indizada	102
Número mínimo de producción no indizada	0
Promedio de producción no indizada	0.37
Número máximo de producción no indizada	9
Número mínimo de producción total	0
Promedio de producción total	11.25
Número máximo de producción total	103
Número mínimo de la producción total de Scopus ( <i>PTS</i> )	0
Promedio de la producción total de Scopus ( <i>PTS</i> )	180.84
Número máximo de la producción total de Scopus ( <i>PTS</i> )	2373
Número mínimo del grado de colaboración ( <i>GCC</i> )	0
Promedio del grado de colaboración ( <i>GCC</i> )	81.44
Número máximo del grado de colaboración ( <i>GCC</i> )	750
Número mínimo del peso de la producción no indizada ( <i>PTPNI</i> )	0
Promedio del peso de la producción no indizada ( <i>PTPNI</i> )	0.73
Número máximo del peso de la producción no indizada ( <i>PTPNI</i> )	18
Número mínimo de la relevancia del cuerpo académico	0
Promedio de la relevancia del cuerpo académico ( <i>RCA</i> )	263.03
Número máximo de la relevancia del cuerpo académico	3057

A partir de estos datos, se utilizó la herramienta web de Aprendizaje Automático BigML para realizar tareas de clustering y reglas de asociación. Para el primer caso, se utilizó el algoritmo G-Means, el cual realiza agrupamiento jerárquico utilizando K-means hasta que se dejan de cumplir criterios estadísticos especificados con un valor crítico que indica si se desean más o

menos grupos. Se analizaron los resultados de los clusters para caracterizar por medio de una etiqueta que los distinga del resto. Los elementos a considerar para el clustering fueron:

- Valor crítico: 2
- Clústers obtenidos: 3
- Variables:
  - Número de integrantes
  - Producción total indizada por Scopus (*PTS*)
  - Grado de colaboración (*GCC*)
  - Peso de la producción no indizada (*PTPNI*)
  - Producción total

Los grupos o clusters se describen a continuación:

- Cuerpos académicos menos relevantes (58 instancias)

Variable	Centroide
Integrantes	3.96
PTS	48.26
GCC	35.53
PTPNI	0.31
produccion_total	5.46

- Cuerpos académicos relevantes (12 instancias)

Variable	Centroide
Integrantes	7.18
PTS	109.17
GCC	63.40
PTPNI	0.36
produccion_total	8.71

- Cuerpos académicos más relevantes (8 instancias)

Variable	Centroide
Integrantes	6.36
PTS	1,159.11
GCC	444.77
PTPNI	0.75
produccion_total	55.38

Con la finalidad de representar las variables en grupos que compartan características similares, se llevó a cabo el proceso de discretización de cada variable considerada para el clustering, tanto para cuerpos académicos como para investigadores.

Para esto, se llevó a cabo un clustering por cada variable y se seleccionó aquel que mejor representa y distribuye los datos. Las clases para los indicadores quedan de la siguiente manera:

- Integrantes: Bajo, medio y alto
- PTS: Muy bajo, bajo, alto y muy alto
- GCC: Bajo, medio y alto
- PTPNI: Bajo y alto
- Producción total: Muy baja, baja, media, alta y muy alta

A diferencia de la lógica booleana, donde un elemento del universo puede pertenecer o no a un conjunto, en los conjuntos borrosos asociados a un valor lingüístico con una etiqueta lingüística, un elemento puede pertenecer a un conjunto con cierto grado de pertenencia. Las funciones de pertenencia se asignaron considerando los valores mínimos, máximos y los centroides del clustering de cada variable. En Figuras 5.13-5.17 se presentan los conjuntos de cada variable divididos en subconjuntos (las clases mencionadas previamente) con sus respectivas funciones de pertenencia.

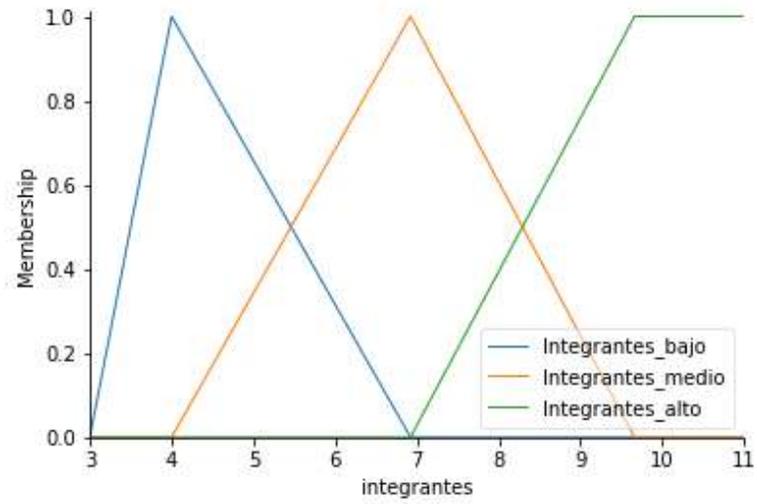


Figura 5.13 Conjuntos borrosos del número de integrantes de los cuerpos académicos

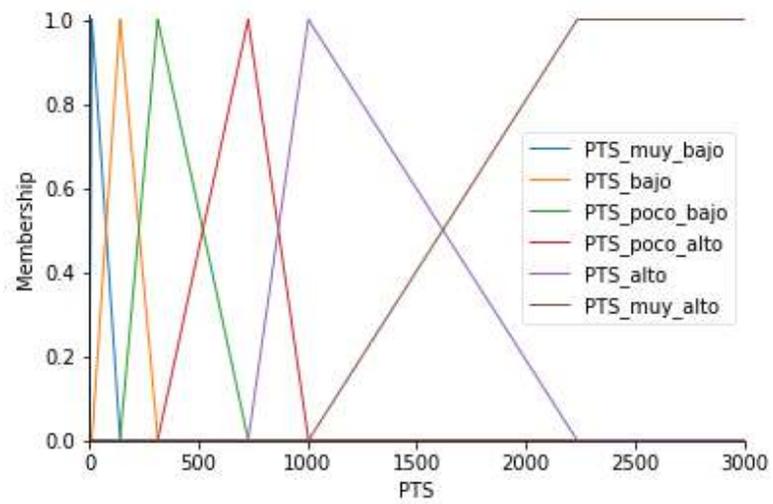


Figura 5.14 Conjuntos borrosos del PTS de los cuerpos académicos

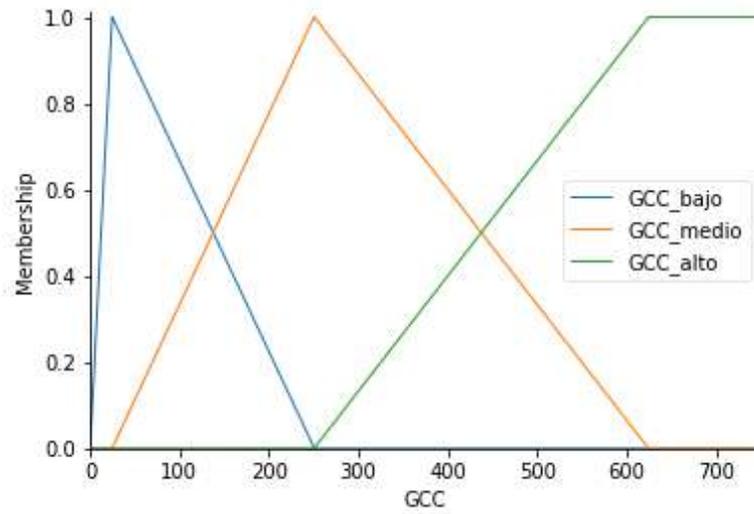


Figura 5.15 Conjuntos borrosos del GCC de los cuerpos académicos

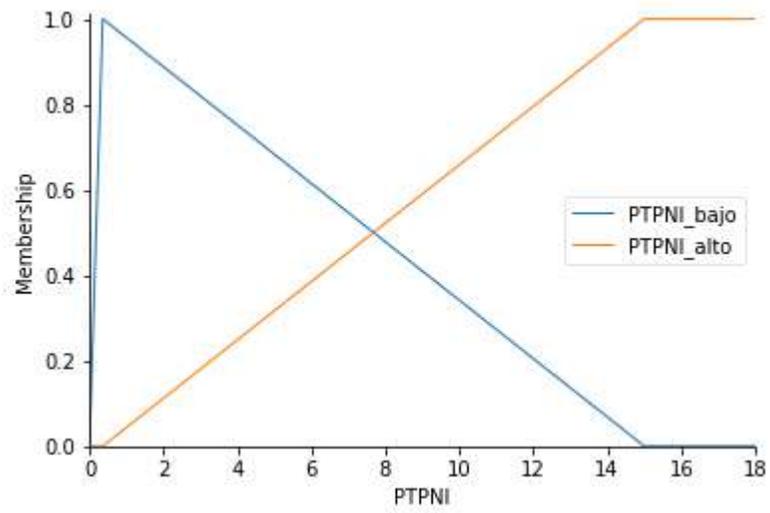


Figura 5.16 Conjuntos borrosos del PTPNI de los cuerpos académicos

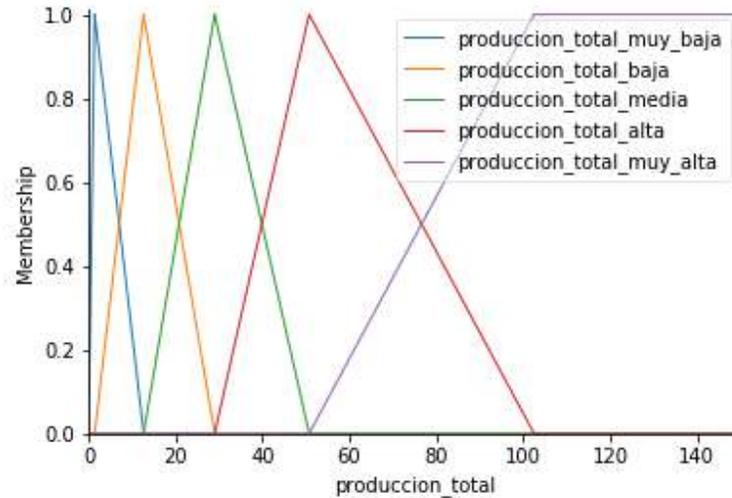


Figura 5.17 Conjuntos borrosos de la producción total de los cuerpos académicos

A partir de los valores discretizados se obtuvieron algunas reglas de asociación que representan la caracterización de los cuerpos académicos. Se consideran aquellas que tienen un nivel de confianza mayor a 0.7

- Si un cuerpo académico tiene un número bajo de integrantes, entonces es uno de los menos relevantes.
- Si un cuerpo académico tiene un número medio de integrantes, entonces es uno de los relevantes.
- Si un cuerpo académico tiene un grado de colaboración bajo (*GCC*), entonces es uno de los menos relevantes.
- Si un cuerpo académico tiene un grado de colaboración alta (*GCC*), entonces es uno de los más relevantes.
- Si un cuerpo académico tiene un número de integrantes que no sea bajo y tiene un grado de colaboración baja, entonces es uno de los relevantes.
- Si un cuerpo académico tiene un grado de colaboración (*GCC*) que no es bajo y su producción total en Scopus (*PTS*) no es poco baja, entonces es uno de los más relevantes.
- Si un cuerpo académico tiene un grado de colaboración (*GCC*) bajo y su número de integrantes es bajo, entonces es de los menos relevantes.
- Si la producción total en Scopus (*PTS*) de un cuerpo académico es muy baja, entonces es uno de los menos colaborativos.

- Si la producción total de un cuerpo académico es muy baja, entonces es uno de los menos colaborativos.
- Si la producción total de un cuerpo académico es muy baja, entonces es uno de los que tienen menor *PTS*.
- Si la producción total en Scopus (*PTS*) de un cuerpo académico es muy baja, entonces es uno de los menos productivos.
- Si un cuerpo académico tiene un grado de colaboración (*GCC*) bajo, entonces es uno de los que tienen menor *PTS*.
- Si un cuerpo académico tiene un grado de colaboración (*GCC*) bajo, entonces es uno de los menos productivos.

### *Investigadores*

Los profesores de tiempo completo pueden contar con el perfil PRODEP y pertenecer al SNI. En la UADY existen 824 profesores de tiempo completo, de los cuales 438 cuentan con producción científica.

En la Tabla 5.18 se describe el estado de la producción y relevancia de los investigadores. Se observa que en todos los casos, los valores máximos se encuentran muy distantes del promedio.

Tabla 5.18 Estado de la producción y relevancia de los investigadores

Número mínimo de publicaciones	1
Promedio de publicaciones	11.42
Número máximo de publicaciones	157
Número mínimo del indicador de relevancia en el cuerpo académico ( <i>RIC</i> )	0
Promedio del indicador de relevancia en el cuerpo académico ( <i>RIC</i> )	81.21
Número máximo del indicador de relevancia en el cuerpo académico ( <i>RIC</i> )	2937
Número mínimo del indicador de relevancia en colaboración externa al cuerpo académico ( <i>RICE</i> )	0
Promedio del indicador de relevancia en colaboración externa al cuerpo académico ( <i>RICE</i> )	208.17
Número máximo del indicador de relevancia en colaboración externa al cuerpo académico ( <i>RICE</i> )	3210
Número mínimo del indicador de relevancia de la producción individual en Scopus ( <i>RPI</i> )	0
Promedio del indicador de relevancia de la producción individual en Scopus ( <i>RPI</i> )	2.31
Número máximo del indicador de relevancia de la producción individual en Scopus ( <i>RPI</i> )	178
Número mínimo del indicador del peso de la producción no indizada ( <i>PNI</i> )	0
Promedio del indicador del peso de la producción no indizada ( <i>PNI</i> )	0.146
Número máximo del indicador del peso de la producción no indizada ( <i>PNI</i> )	6
Número mínimo de la relevancia del investigador ( <i>RDI</i> )	0.015
Promedio de la relevancia del investigador ( <i>RDI</i> )	292.5
Número máximo de la relevancia del investigador ( <i>RDI</i> )	4361

A partir de estos datos se utilizó la herramienta web de Aprendizaje Automático BigML para realizar tareas de clustering y reglas de asociación. En el primer caso, se utilizó el algoritmo G-Means, el cual realiza agrupamiento jerárquico utilizando K-means hasta que se dejan de cumplir los criterios estadísticos especificados con un valor crítico que indica si se desean más

o menos grupos (Hamerly y Elkan, 2003). Previamente se realizó el análisis de los componentes principales (*PCA*) para reducir la dimensionalidad de un conjunto de datos, incluyendo:

- RIC
- RICE
- RPI
- RDI

No se consideró *PNI* ya que el valor en gran parte de los casos es 0 y no resulta relevante. Dichos componentes principales en conjunto tienen una varianza acumulada del 82.13% de los datos. Como se observa en la Figura 5.18, el primer componente principal tiene correlación fuerte con la relevancia del investigador (*RDI*), mientras que el segundo componente tiene correlación con la relevancia de la producción individual (*RPI*). En los componentes principales, los valores menores representan mayores valores en los indicadores con los que tiene mayor correlación.



Figura 5.18 Correlación de cada variable en los dos componentes principales considerados para el clustering de los investigadores

Se analizaron los resultados de los clusters para caracterizarlos por medio de etiquetas . Los elementos a considerar para el clustering fueron:

- Valor crítico: 1
- Clústers obtenidos: 6
- Variables:
  - PC1
  - PC2
  - Producción total

Los grupos o clusters se describen a continuación:

- Investigadores menos relevantes (294 instancias)

Variable	Centroide
PC1	0.64
PC2	0.05
produccion_total	3.43

- Investigadores muy poco relevantes (78 instancias)

Variable	Centroide
PC1	-0.20
PC2	0.14
produccion_total	14.97

- Investigadores poco relevantes (34 instancias)

Variable	Centroide
PC1	-1.02
PC2	0.18
produccion_total	35.23

- Investigadores relevantes (9 instancias)

Variable	Centroide
PC1	-1.49
PC2	-2.69
produccion_total	27.15

- Investigadores muy relevantes (15 instancias)

Variable	Centroide
PC1	-4.45
PC2	1.12
produccion_total	36.68

- Investigadores más relevantes (8 instancias)

Variable	Centroide
PC1	-6.46
PC2	0.35
produccion_total	110.22

Con la finalidad de representar las variables en grupos que comparten características similares, se llevó a cabo el proceso de discretización de cada variable considerada para el clustering.

Para esto, se llevó a cabo un clustering por cada variable y se seleccionó aquel que mejor representa y distribuye los datos. Las clases para los indicadores quedan de la siguiente manera:

- PC1: Muy bajo, bajo, alto y muy alto
- PC2: Muy bajo, bajo, medio, alto y muy alto
- Producción total: Muy baja, baja, media, alta y muy alta

Al igual que en los cuerpos académicos, se crearon conjuntos borrosos con sus respectivas funciones para la discretización de los datos, las cuales se presentan en las Figuras 5.19-5.21.

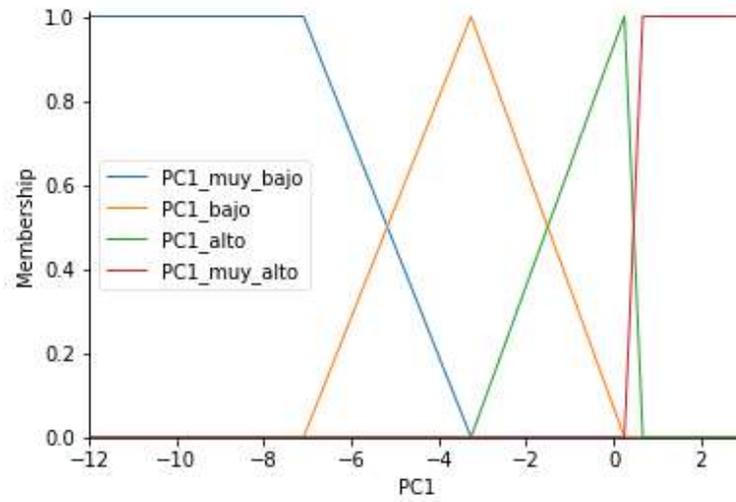


Figura 5.19 Conjuntos borrosos del PC1 de los investigadores

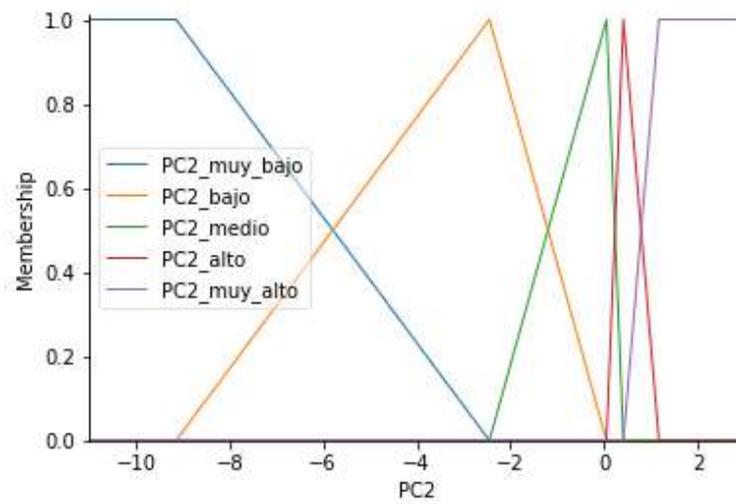


Figura 5.20 Conjuntos borrosos del PC2 de los investigadores

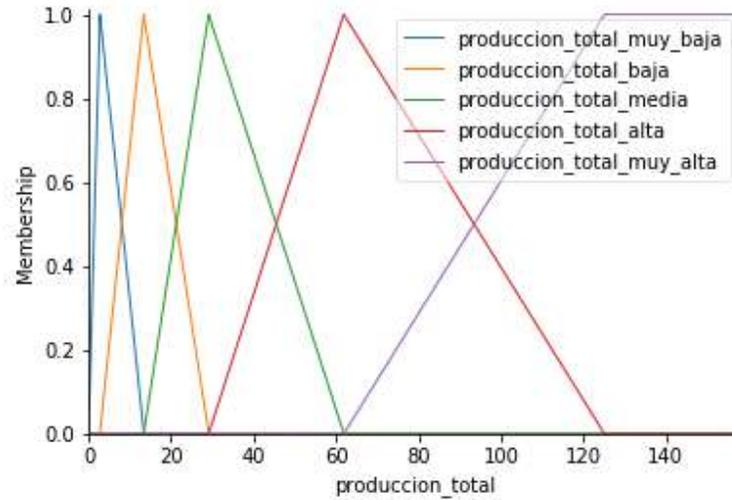


Figura 5.21 Conjuntos borrosos de la producción total de los investigadores

A partir de los valores discretizados se obtuvieron algunas reglas de asociación que representan la caracterización de los investigadores. Se consideran aquellas que tienen un nivel de confianza mayor a 0.7.

- Si un investigador tiene una muy alta producción total, entonces es uno de los más relevantes
- Si el valor de PC2 de un investigador es medio y su producción total es muy baja, entonces es uno de los menos relevantes
- Si un investigador tiene una producción total muy baja, entonces es de los menos relevantes
- Si el valor de PC1 de un investigador es alto, entonces es de los menos relevantes
- Si el valor de PC2 de un investigador es medio y su producción total es muy baja, entonces es de los menos relevantes
- Si la producción total de un investigador es medio, entonces es uno de los relevantes

## 5.8 DISCUSIÓN DE LOS RESULTADOS

En cuanto a la relevancia de los cuerpos académicos de la FMAT, existe correlación importante con el grado de colaboración, la relevancia de la producción indizada por Scopus y la producción

total, lo que permite observar que el comportamiento del indicador es el correcto, ya que se encarga de sumarizar los valores de los indicadores de la producción y del grado de colaboración.

Para los investigadores de FMAT, sin distinguir si pertenecen o no a un cuerpo académico, su relevancia tiene correlación con sus colaboraciones (dentro y fuera del cuerpo académico), con la relevancia de su cuerpo académico y con su producción total. Además, en el área de Computación también se encontró que la producción no indizada da lugar a la relevancia del investigador, mientras que en el área de Matemáticas, la producción individual es importante para la relevancia del investigador.

En el análisis donde se consideran únicamente a los investigadores que pertenecen a un cuerpo académico, la relevancia de los investigadores en sus cuerpos académicos involucra la relevancia y producción total del investigador y de su cuerpo académico. Para la relevancia del investigador, tanto en el área de Computación como en el área de Matemáticas, este indicador tiene correlación importante con la relevancia y producción del investigador en colaboración dentro y fuera de su cuerpo académico, así como la relevancia total del cuerpo académico. Sin embargo, en el área de Matemáticas la producción total del cuerpo académico es importante para la relevancia del investigador, mientras que en el área de Computación no.

Para los investigadores de FMAT que no pertenecen a algún cuerpo académico, la relevancia del investigador tiene correlación importante con la producción y relevancia en colaboraciones, el peso y el total de publicaciones no indizadas y la producción total. Se puede observar que para los investigadores pertenecientes a algún cuerpo académico, la producción no indizada no resulta importante para su relevancia, debido a que las instancias evaluadoras se enfocan en la producción de alto impacto, y para mantener o aumentar el nivel de consolidación, este tipo de publicaciones son prioritarias, por lo que se puede garantizar que los cuerpos académicos de la FMAT cumplen con lo solicitado.

Después del análisis efectuado sobre los tres cuerpos académicos del área de Computación, se encontró que, en el primero no resultó relevante la correlación de los indicadores. En el

segundo, la relevancia de las colaboraciones externas al cuerpo académico se incrementó a mayor relevancia del investigador, y la producción en colaboraciones externas al cuerpo académico se incrementó a mayor producción total del investigador. Y en el tercero, la relevancia de la producción en colaboraciones externas al cuerpo académico se incrementó a mayor producción de la misma, la producción del investigador en su cuerpo académico se incrementó a mayor producción total del investigador y la producción en colaboraciones externas al cuerpo académico se incrementó a mayor producción total del investigador.

Con respecto al análisis realizado sobre los cinco cuerpos académicos del área de Matemáticas, se encontró que, en el primero, la relevancia del investigador dentro de su cuerpo académico incrementó a mayor relevancia del investigador, además de que la producción del investigador dentro de su cuerpo académico incrementó a mayor producción individual. En el segundo, entre los indicadores importantes, la relevancia del investigador en su cuerpo académico se incrementó a mayor peso y producción de las publicaciones no indizadas, la relevancia del investigador incrementó a mayor producción del mismo en colaboración dentro y fuera de su cuerpo académico, así como la producción individual. En el tercero, la relevancia de las colaboraciones externas al cuerpo académico incrementó a mayor relevancia del investigador. En el cuarto, la relevancia de las colaboraciones externas al cuerpo académico incrementó a mayor producción del mismo tipo y la relevancia del investigador incrementó a mayor producción total del investigador. En el quinto, la relevancia de las colaboraciones externas al cuerpo académico incrementó a mayor relevancia del investigador.

En el análisis del estado de la relevancia utilizando minería de datos, el uso del algoritmo G-Means permitió obtener un mejor agrupamiento que utilizando K-Means, debido a que se desconoce el número óptimo de clústers. Los cuerpos académicos y los investigadores se agruparon de acuerdo con su relevancia. Las reglas de asociación obtenidas permiten reforzar el conocimiento previo que en teoría debería cumplirse. Además, con estos resultados es posible tomar decisiones relacionadas con investigadores y cuerpos académicos que realizan publicaciones relevantes para la institución y su área, así como aquellos que pueden mejorar su relevancia e incentivarlos a publicar en revistas de mayor calidad, entre otras sugerencias.

## 5.9 CONCLUSIONES

En este capítulo se presentaron casos de estudio de correlación entre los indicadores propuestos para localizar en diferentes escenarios los elementos que mayor importancia tienen en los valores de los indicadores de la relevancia de los cuerpos académicos y de los investigadores en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán. Por otra parte, se presentó el estudio del estado de la producción científica en la UADY por medio de técnicas de minería de datos.

## 6 CONCLUSIONES Y TRABAJO FUTURO

Este capítulo tiene como objetivo exponer los resultados de una valoración global de los productos obtenidos en esta investigación.

Se inicia analizando la consecución de los objetivos planteados en el trabajo, para proseguir con una revisión de las principales aportaciones generadas como resultado de la investigación.

También se presentan las distintas publicaciones realizadas en estos últimos años que confirman los resultados obtenidos. Finalmente, se establecen las líneas futuras de investigación.

### 6.1 INTRODUCCIÓN

Los repositorios digitales de documentación científica han permitido que la información acerca de las publicaciones realizadas por profesores e investigadores estén al alcance de cualquier persona. Incluso los repositorios de acceso abierto brindan contenidos en texto completo, ayudando a la divulgación y aprovechamiento de los avances científicos y tecnológicos.

Cada recurso almacenado en un repositorio digital puede ser descrito, localizado y referenciado por medio de sus metadatos. Sin embargo, los metadatos suelen adolecer de integridad, completitud, exactitud y consistencia.

Diversas instancias como el Sistema Nacional de Investigadores (SNI) y el Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP), ambas en México, se encargan de evaluar y otorgar reconocimientos a los investigadores por medio de su producción científica almacenada en diversos repositorios. Para esta tarea, existe una serie de indicadores basados en las citas y en el prestigio de las revistas.

Se ha propuesto un modelo de indicadores para medir la relevancia tanto de grupos de investigación (cuerpos académicos) así como de los investigadores, utilizando como criterios la importancia de la publicación indizada por Scopus de acuerdo con su tipo según las instancias

evaluadoras, el grado de colaboración interno y externo y en menor medida, la producción no indizada y de acceso abierto.

Considerando que los metadatos posiblemente contienen valores erróneos y con base en los indicadores propuestos, se ha implementado un sistema de índices para valorar la producción académica y la investigación, a partir de repositorios digitales y metadatos. Entre las tareas del sistema se encuentran la recuperación, depuración y almacenamiento de la producción indizada por Scopus y la producción de acceso abierto; el cálculo de los indicadores propuestos; la representación del conocimiento por medio de un modelo ontológico y la preparación del formato de los archivos para la presentación de los valores de los indicadores para la interfaz de usuario.

Es importante destacar los valores de los indicadores propuestos y su cálculo por medio de un proceso automático, permitiendo analizar el estado actual de la investigación de una institución abarcando aspectos que difícilmente se pueden obtener en un proceso manual, como el grado de colaboración y la relevancia de las publicaciones.

## 6.2 ANÁLISIS DE LA CONSECUCCIÓN DE LOS OBJETIVOS

El modelo de indicadores propuesto mide la relevancia científica de los cuerpos académicos y de los investigadores. La relevancia consiste en la importancia de las publicaciones de acuerdo con su tipo con base en lo solicitado por las instancias evaluadoras, el grado de colaboración y el impacto de las publicaciones en la comunidad científica.

En este trabajo se ha presentado la implementación del sistema que con base en los indicadores propuestos, mide la relevancia de los cuerpos académicos y de los investigadores, a partir de la producción almacenada en diversos repositorios, desde la consulta y recuperación de los datos hasta la presentación de los resultados en la interfaz de usuario. Al inicio de este documento se expusieron los objetivos parciales que se pretendían cumplir para lograr el objetivo general que enmarca esta tesis que es:

**Desarrollar una solución tecnológica que valore la relevancia de un investigador a través de la evaluación de los metadatos de su producción científica, los cuales se pueden consultar en uno o más repositorios digitales.**

A continuación se hace una valoración de la consecución de los objetivos parciales presentados en el primer capítulo:

**1. Realizar una revisión de las herramientas existentes que den a conocer los indicadores cuantitativos estadísticos de la producción científica de un investigador.**

Se ha revisado en la bibliografía cuáles son los indicadores que se utilizan actualmente para medir la relevancia de un investigador, así como algunas propuestas que consideran aspectos fuera de los indicadores tradicionales (*sección 2.3.2 Indicadores cuantitativos para la productividad científica*)

También se localizaron y se estudiaron los diversos repositorios que contienen la producción científica, tanto indizada como de acceso abierto (*sección 2.5 Repositorios institucionales de acceso abierto*).

**2. Establecer las herramientas y técnicas que permitan recuperar la producción científica almacenada en diversos repositorios.**

Al elegir los repositorios necesarios para el cálculo de los indicadores propuestos, se localizaron los recursos que dichos repositorios disponen para recuperar sus datos almacenados. En el caso de Scopus, se utilizó su API de autor como de búsqueda. (*sección 4.4.1 Consulta y almacenamiento de las publicaciones científicas*).

También se realizó la cosecha de metadatos a partir de repositorios de acceso abierto (*sección 4.4.1 Consulta y almacenamiento de las publicaciones científicas*) a partir del protocolo para la interoperabilidad OAI-PMH (*sección 2.1.4 Interoperabilidad*).

- 3. Utilizar técnicas de extracción de datos que permitan obtener los nombres de los investigadores y su producción desde los repositorios digitales, aún cuando sus registros incluyan inconsistencias.**

Los datos de las publicaciones recuperadas a partir de los repositorios de acceso abierto no presentaron anomalías para la recuperación, puesto que los metadatos son validados por las instituciones propietarias de dichos repositorios, coincidiendo los nombres de los autores en la base de datos como en los resultados de las consultas.

Sin embargo, para la producción indizada por Scopus, que es la base de la mayoría de los indicadores, se analizaron los resultados de las consultas para diseñar una serie de condiciones que validen cuándo un perfil pertenece a un investigador en específico (*sección 4.4.1 Consulta y almacenamiento de las publicaciones científicas*).

- 4. Definir métricas para la evaluación de los indicadores propuestos para la relevancia científica.**

Utilizando como referencia las evaluaciones de las instancias mexicanas evaluadoras de la producción académica y científica, se propuso una serie de indicadores que miden la relevancia de cuerpos académicos (aplicables para cualquier otro tipo de grupo de investigación) (*sección 3.1 Relevancia de un cuerpo académico*) y de los investigadores (*sección 3.2 Relevancia de un investigador*). La relevancia consiste en la importancia de las publicaciones de acuerdo con su tipo con base en lo solicitado por las instancias evaluadoras, el grado de colaboración y el impacto de las publicaciones en la comunidad científica.

- 5. Implementar una solución tecnológica que calcule por indicadores cuantitativos estadísticos la relevancia de la producción académica y científica de los investigadores.**

Después de definir los indicadores, se procedió a la implementación del sistema. Se diseñó una arquitectura en capas (*sección 4.1 Marco arquitectónico*) implementado por diversas herramientas tecnológicas (*sección 4.2 Especificaciones técnicas*), las diversas funcionalidades del sistema, como la consulta y almacenamiento de las publicaciones científicas y la generación de los indicadores (*sección 4.4 Funcionalidad del sistema*).

Los resultados se presentan de forma estética a través de mapas de árbol y grafos de colaboración por medio de la interfaz de usuario (*sección 4.3 Interfaz web*). Además, se diseñó un modelo ontológico para la representación del conocimiento (*sección 4.5 Representación del conocimiento*).

## **6. Analizar experimentalmente el funcionamiento de la solución tecnológica con el Repositorio Digital de la Universidad Autónoma de Yucatán.**

Se recuperó la producción no indizada por Scopus a partir del repositorio nacional del Consejo Nacional de Ciencia y Tecnología (CONACYT) y uno de los repositorios que lo componen es el Repositorio Digital de la UADY (REDI-UADY) (*sección 4.4.1 Consulta y almacenamiento de las publicaciones científicas*). Se calcularon los indicadores de la producción no indizada (*sección 4.4.2 Generación del indicador de la productividad de cuerpos académicos, sección 4.4.3 Generación del indicador de la productividad de los investigadores*).

También con los resultados de los indicadores se realizaron estudios de correlación entre los indicadores propuestos en diferentes escenarios en la Facultad de Matemáticas de la UADY, así como un estudio del estado actual de la investigación utilizando técnicas de minería de datos (*Sección 5 Resultados*).

Estos objetivos parciales han sido generados a partir del objetivo general, que ha sido:

*Desarrollar una solución tecnológica que valore la relevancia de un investigador a través de la evaluación de los metadatos de su producción científica, los cuales se pueden consultar en uno o más repositorios digitales.*

Considerando que se ha logrado la consecución de los objetivos parciales, puede decirse que el objetivo general de esta investigación se ha cumplido, puesto que se ha definido, desarrollado y evaluado un sistema de índices para valorar la relevancia de la producción académica y la investigación a partir de repositorios digitales y metadatos.

La veracidad de la hipótesis presentada en el Capítulo 1 *Introducción* se basaba en el cumplimiento del objetivo general. De manera que a partir de los resultados obtenidos, puede concluirse que la hipótesis ha quedado demostrada y por lo tanto puede afirmarse que:

**La relevancia de un investigador puede medirse mediante una herramienta tecnológica que considere la producción almacenada en uno o más repositorios digitales; aún cuando los metadatos asociados a dicha producción posiblemente adolecen de integridad, completitud, exactitud y consistencia.**

### 6.3 ANÁLISIS DE LAS APORTACIONES

En este trabajo se ha especificado un sistema de índices para valorar la relevancia de la producción académica y la investigación a partir de repositorios digitales y metadatos. Se han presentado los diversos elementos que conforman la propuesta y cuya realización permite lograr el objetivo planteado.

El concepto de relevancia del investigador de acuerdo con las instancias evaluadoras ha sido la base para la propuesta. Para dicho concepto, se considera la importancia de las publicaciones de acuerdo con sus tipos, el grado de colaboración y el impacto de la producción en la comunidad científica.

Se han propuesto tres indicadores que en conjunto definen la relevancia de un cuerpo académico y cuatro indicadores que conforman la relevancia de un investigador. De estos siete indicadores, cuatro corresponden a la producción indizada por Scopus, dos a la no indizada por Scopus y disponible en repositorios de acceso abierto y uno considera aspectos de ambos tipos de publicaciones, asignando de esta manera un mayor peso a la producción indizada, que es lo que representa la relevancia de un investigador.

Este modelo de indicadores se ha implementado en un sistema independiente de cualquier herramienta tecnológica, utilizando servicios web y protocolos para la interoperabilidad, permitiendo que el sistema puede ser adaptado con facilidad de acuerdo a nuevas necesidades.

Las características a evaluar por los indicadores, así como su cálculo de manera automática, permiten un análisis más detallado y conciso que una evaluación realizada de manera manual, observando incluso el estado de la colaboración interna y externa, y su fácil representación, tal como se ha presentado por medio de la Teoría de grafos.

Además, los mapas de árbol presentan de manera estética y de fácil evaluación, los resultados de los indicadores de los cuerpos académicos y de los investigadores, localizando los elementos fuertes y débiles por medio de un interfaz de usuario.

Los resultados de los estudios de correlación de los indicadores en los diversos escenarios en la FMAT de la UADY permiten identificar que en la relevancia de los investigadores influyen diferentes aspectos, dependiendo del área de estudio o si el investigador es miembro de un cuerpo académico.

También el estudio a través de la minería de datos permite localizar diversos grupos de acuerdo al nivel de relevancia de los cuerpos académicos y los investigadores, además de validar que en la UADY se los lineamientos de las instancias evaluadoras por medio de las reglas de asociación.

## 6.4 CONTRASTE DE RESULTADOS

Con la finalidad de presentar los resultados de las diversas fases de la investigación, se han realizado publicaciones en revistas (2) y ponencias en congresos internacionales (3). A continuación, se enlistan agrupadas de manera correspondiente y ordenadas cronológicamente de forma descendiente.

### 6.4.1 ARTÍCULOS EN REVISTAS

1. Guerrero Sosa, J. D. T., Menéndez Domínguez, V. H. y Castellanos Bolaños, M. E. (2018). Indicadores de calidad en investigaciones científicas: Antecedentes. *Abstraction and Application*, 19, 6–24.
2. Guerrero-Sosa, J. D. T., Menendez-Domínguez, V. y Uc-Cetina, V. Redes neuronales recurrentes y su uso para la conversión de cadenas. *Komputer Sapiens*. Aceptado.
3. Guerrero-Sosa, J., Menéndez-Domínguez, V., Castellanos-Bolaños, M. y Gómez-Montalvo, J. Use of an Ontological Model to Assess the Relevance of Scientific Production. *IEEE Latin America Transactions*. Aceptado.

### 6.4.2 PONENCIAS EN CONGRESOS INTERNACIONALES

1. Guerrero Sosa, J. D. T., Menéndez Domínguez, V. H. y Castellanos Bolaños, M. E. (2018). Sistema de índices para valorar la calidad de la producción académica y la investigación, a partir de repositorios digitales y metadatos. En M. E. Prieto-Méndez, S. J. Pech-Campos y A. Francesa-Alfaro (Eds.), *X Conferencia Conjunta Internacional sobre Tecnologías y Aprendizaje* (pp. 45–52). Cartago: CIATA.org-UCLM.
2. Guerrero, J., Sánchez, D., Menéndez, V., Castellanos, M. E. y Gómez, J. (2019). Tools for interoperability between repositories of digital resources. En L. Gómez Chova, A. López Martínez e I. Candel Torres (Eds.), *Proceedings of INTED2019* (pp. 6292–6300). Valencia: IATED.

3. Guerrero-Sosa, J. D. T., Menendez-Domínguez, V., Castellanos-Bolaños, M.-E. y Curi-Quintal, L. F. (2019). Use of graph theory for the representation of scientific collaboration. En N. T. Nguyen (Ed.), *11th International Conference on Computational Collective Intelligence*. Hendaye: Springer.
4. Guerrero-Sosa, J. D. T., Romero, F. P., Serrano-Guerrero, J., Menendez-Dominguez, V. y Castellanos-Bolaños, M. E. (2019). A proporsal for a recommender system of scientific relevance. En *International Conference on Information Technology and Quantitative Management*. Granada.

## 6.5 TRABAJO FUTURO

Anteriormente se presentó un análisis del cumplimiento de los objetivos iniciales planteados, así como una revisión de las aportaciones conseguidas. A lo largo del desarrollo de la presente investigación se han presentado diferentes líneas de trabajo que permiten el desarrollo, ampliación y mejora de los resultados obtenidos a este momento.

### 6.5.1 MEJORAS EN LAS TÉCNICAS UTILIZADAS

- **Definición de un indicador que mida la utilidad de las citas de un cuerpo académico y de un investigador.** Scopus cuenta con una API que permite recuperar las citas que ha recibido cada publicación por medio de alguno de sus identificadores (DOI y Scopus-ID).

Considerando los valores de estos datos, es posible definir un indicador que se encargue de evaluar el impacto de la investigación y el impacto de la publicación de cada cita recibida. De este modo, se considera el impacto que tiene cada cita de acuerdo al prestigio de la revista o publicación de origen.

También es posible localizar las autocitaciones, eliminarlas y considerar como parte de la utilidad de las citas únicamente las que no pertenecen al mismo autor o grupo de investigación.

O bien, es posible considerar las autocitaciones si la publicación en cuestión proviene de otro trabajo de investigación del mismo autor o grupo de investigación, siendo esto posible a través de un estudio semántico para identificar la similitud entre las publicaciones, considerando el título, los resúmenes y las palabras clave.

- **Ajuste de la definición del modelo de indicadores a valores de 0 a 100.** Si bien, los indicadores permiten identificar por medio de los mapas de árbol en el interfaz de usuario qué elementos son más relevantes que otros, es necesario ajustar los indicadores para que sus valores se encuentren en un rango de 0 a 100, ya que esta escala permite una fácil evaluación de los valores.

Al realizar este ajuste, las tomas de decisiones y evaluaciones que requieran el uso de los indicadores propuestos, pueden ser más transparentes y de fácil comprensión.

- **Definición de técnicas de extracción de datos de la producción científica a partir de otros repositorios como Google Scholar y Web of Science.** Los indicadores propuestos utilizan las publicaciones indizadas por Scopus. Sin embargo, Web of Science sigue siendo una base de datos confiable para medir la productividad y relevancia de los investigadores, según las universidades y las instancias evaluadoras.

Se propone analizar el formato y los valores de los metadatos de las consultas realizadas a Web of Science para definir las técnicas de extracción de datos, al igual que con la producción de Google Scholar. En este último caso, sería necesario identificar cuáles recursos pertenecen a publicaciones científicas, puesto que este repositorio indexa otro tipo de documentos.

#### 6.5.2 MEJORAS EN EL SISTEMA

- **Identificar la producción científica de un profesor o un grupo de profesores (Cuerpo académico, núcleo base, Facultad, Campus, UADY) durante un intervalo de tiempo.** A pesar de que los valores de los indicadores permiten ver la relevancia de los cuerpos académicos y los investigadores considerando toda su producción, se

considera importante conocer el estado de la investigación en un intervalo de tiempo, considerando los miembros en dicha ventana de tiempo de los cuerpos académicos, de los núcleos base, de una facultad, de un campus y de la UADY.

A través de esta mejora, es posible obtener la evolución de la producción en los diversos escenarios, facilitando la toma de decisiones y evaluar, con base en los indicadores, qué ha ocasionado una mayor o menor relevancia en un momento determinado.

- **Caracterizar el impacto de los programas de becas en la producción científica generada en un programa de posgrado.** Los indicadores propuestos evalúan únicamente a los profesores. Sin embargo, es posible utilizar una base de datos de estudiantes de posgrado de la UADY para evaluar igualmente la producción que ha generado durante el programa al que se encuentra inscrito.

De esta manera, es posible evaluar cuál es el impacto y aprovechamiento de las becas de posgrado, identificando los proyectos que resultan relevantes y quiénes son los profesores que incentivan a sus estudiantes a realizar producción científica.

- **Establecer el impacto de los convenios de colaboración en la producción científica de la UADY.** Es posible adaptar el sistema para identificar y evaluar el impacto de los convenios que la UADY tiene con otras instituciones nacionales e internacionales, a través de las publicaciones científicas.

De esta manera sería posible demostrar el aprovechamiento de los convenios.

- **Identificar profesores con perfiles de producción afines para posibles colaboraciones.** Por medio de un sistema que emita recomendaciones con base en estudios de similitud semántica entre las temáticas de las diversas publicaciones científicas almacenadas en la base de datos, se pretende localizar investigadores que trabajen temas similares para apoyar la colaboración interna en la UADY.

Para la evaluación de la similitud semántica se pretende utilizar los títulos, los resúmenes y las palabras clave de las publicaciones. Además, estas recomendaciones apoyarían fuertemente la colaboración multidisciplinaria, ya que las comparaciones se realizarían entre todos los profesores de la UADY.

- **Identificación de grupos de profesores que pueden conformar cuerpos académicos a partir de temáticas de investigación similares.** Se pretende realizar un sistema de recomendaciones para la creación de cuerpos académicos con base en la similitud semántica de las temáticas de los investigadores que no son miembros de un cuerpo académico.

Además, como los cuerpos académicos, de acuerdo con el PRODEP, pueden ser mutidisciplinarios, esta herramienta apoyaría este aspecto. La formación de nuevos cuerpos académicos permite que los investigadores sean más relevantes, puesto que uno de los indicadores para la relevancia de los investigadores está asociado con la producción del investigador en colaboración con el cuerpo académico, considerando el grado de participación y el peso de la publicación de acuerdo a su importancia según las instancias evaluadoras.

## 7 BIBLIOGRAFÍA

- AEBOE. (2019). Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Madrid, España. Recuperado de <https://www.boe.es/buscar/pdf/2011/BOE-A-2011-9617-consolidado.pdf>
- Agudelo, M. (2009). Los metadatos. *Gestión Contenidos Educ. Virtual Calid.* Antioquía, Colombia: Ministerio de Educación Nacional de Colombia. Recuperado de [http://aprendeenlinea.udea.edu.co/lms/men/docsoac3/0301\\_metadatos.pdf](http://aprendeenlinea.udea.edu.co/lms/men/docsoac3/0301_metadatos.pdf)
- Al-Khalifa, H. S. y Davis, H. C. (2006). The evolution of metadata from standards to semantics in E-learning applications. En *Proceedings of the seventeenth conference on Hypertext and hypermedia - HYPERTEXT '06* (p. 69). <https://doi.org/10.1145/1149941.1149956>
- ANECA. (2017). Criterios de evaluación [noviembre-2017] - Aneca. Recuperado el 1 de diciembre de 2017, de <http://www.aneca.es/Programas-de-evaluacion/ACADEMIA/Criterios-de-evaluacion-noviembre-2017>
- Avison, D., Lau, F., Myers, M. y Nielsen, P. A. (1999). Action research. *Communications of the ACM*, 42(1), 94–97. <https://doi.org/10.1145/291469.291479>
- Azcorra, A., Bernardos, C., Gallego, Ó. y Soto, I. (2001). *Informe sobre el estado de la Teleeducación en España.* Madrid, España. Recuperado de <http://bibliotecadigital.tamaulipas.gob.mx/archivos/descargas/53f96c4bc2f9bdd3046cc141e1da668b52fcb643.pdf>
- Bakkalbasi, N., Bauer, K., Glover, J. y Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3(7). <https://doi.org/10.1186/1742-5581-3-7>
- Barker, P. (2005). What is IEEE Learning Object Metadata/IMS Learning Resource Metadata? *Cetis Standards Briefings Series*, 1, 1–4. Recuperado de <http://publications.cetis.ac.uk/wp-content/uploads/2011/02/WhatIsIEEELOM.pdf>
- Barton, J., Currier, S. y Hey, J. M. N. (2003). Building Quality Assurance into Metadata Creation: an Analysis based on the Learning Objects and e-Prints Communities of Practice. *International conference on Dublin Core and metadata applications supporting communities of discourse and practicemetadata research applications*, 1–10. Recuperado de [http://eprints.soton.ac.uk/20/1/BartonCurrierHey\\_DC2003.doc](http://eprints.soton.ac.uk/20/1/BartonCurrierHey_DC2003.doc)

- Berners-Lee, T. y Hendler, J. (2001). Publishing on the semantic web. *Nature*. Nature Publishing Group. <https://doi.org/10.1038/35074206>
- Bornmann, L. y Marx, W. (2011). The h index as a research performance indicator. *European Science Editing*, 37(3), 77–80.
- Brasher, A. y McAndrew, P. (2004). *Human-generated learning object metadata*. *OTM Workshops 2004, Lecture Notes in Computer Science 3292* (Vol. 3292).
- Buela-Casal, G. (2003). Evaluación de la calidad de los artículos y de las revistas científicas: Propuesta del factor de impacto ponderado y de un índice de calidad. *Psicothema*, 15(1), 23–35. <https://doi.org/10.1016/j.acalib.2011.02.006>
- Castagné, M. (2013). *Institutional repository software comparison: DSpace, EPrints, Digital Commons, Islandora and Hydra*. Vancouver, Canadá. <https://doi.org/10.14288/1.0075768>
- COAR. (2011). El caso de Interoperabilidad para Repositorios de Acceso Abierto. Recuperado de <http://coar-repositories.org>
- CONACYT. (2014). Lineamientos Técnicos para el Repositorio Nacional y los Repositorios Institucionales. Recuperado el 17 de enero de 2018, de <http://www.siiicyt.gob.mx/index.php/normatividad/2-conacyt/4-conacyt/1499-lineamientos-tecnicos-para-el-repositorio-nacional-y-los-repositorios-institucionales/file>
- CONACYT. (2016). Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología.
- CONACYT. (2017a). Área I: Físico-Matemáticas y Ciencias de la Tierra Criterios Internos de Evaluación Área I: Físico-Matemáticas y Ciencias de la Tierra Criterios Internos de Evaluación, 1–4. Recuperado de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13714-criterios-especificos-ai/file>
- CONACYT. (2017b). Área II: Biología, Química y Ciencias de la Vida. Criterios Específicos de Evaluación. Recuperado el 27 de noviembre de 2017, de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13715-criterios-especificos-aii/file>
- CONACYT. (2017c). Área III: Medicina y Ciencias de la Salud. Criterios Específicos de Evaluación. Recuperado el 27 de noviembre de 2017, de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13716-criterios-especificos-aiii/file>

- CONACYT. (2017d). Área IV: Humanidades y Ciencias de la Conducta. Criterios Específicos de Evaluación. Recuperado el 27 de noviembre de 2017, de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13717-criterios-especificos-aiv/file>
- CONACYT. (2017e). Área V: Ciencias Sociales Criterios Específicos de Evaluación. Recuperado el 27 de noviembre de 2017, de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13718-criterios-especificos-av/file>
- CONACYT. (2017f). Área VI: Biotecnología y Ciencias Agropecuarias Criterios Internos de Evaluación Área VI: Biotecnología y Ciencias Agropecuarias Criterios Internos de Evaluación, 1–4. Recuperado de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13719-criterios-especificos-avi/file>
- CONACYT. (2017g). Área VII: Ingenierías Criterios Específicos de Evaluación, 1–8. Recuperado de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal/criterios-sni/13720-criterios-especificos-avii/file>
- CONACYT. (2017h). Declaración de protección de los Derechos de Autor y otros Derechos de Propiedad Intelectual. Recuperado el 16 de enero de 2018, de <https://www.repositorionacionalcti.mx/proteccion>
- CONACYT. (2017i). Inicio|Siicyt. Recuperado el 27 de noviembre de 2017, de <http://www.siicyt.gob.mx/>
- CONACYT. (2017j). Lineamientos Jurídicos de Ciencia Abierta. Recuperado el 30 de noviembre de 2017, de <http://www.siicyt.gob.mx/index.php/normatividad/conacyt-normatividad/programas-vigentes-normatividad/lineamientos/lineamientos-juridicos-de-ciencia-abierta/3828-lineamientos-juridicos-de-ciencia-abierta/file>
- CONACYT. (2017k). Programas del CONACYT - Google Public Data Explorer. Recuperado el 27 de noviembre de 2017, de [https://www.google.com/publicdata/explore?ds=cnkd64ja0fa0u\\_](https://www.google.com/publicdata/explore?ds=cnkd64ja0fa0u_)
- CONACYT. (2017l). Repositorio Nacional/Sección Informativa. Recuperado el 30 de noviembre de 2017, de <https://www.repositorionacionalcti.mx/>
- CONACYT. (2017m). Repositorio Nacional. Recuperado el 25 de octubre de 2017, de

- <https://www.repositorionacionalcti.mx/>
- CONACYT. (2017n). Sistema Nacional de Investigadores. Recuperado el 20 de noviembre de 2017, de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores>
- Conover, W. J. (1999). *Practical nonparametric statistics*.
- Corryn, J. Fair Access to Science and Technology Research Act of 2017 (2017).
- Crossan, M. M., Lane, H. W. y White, R. E. (1999). An organizational learning framework: From intuition to institution. *Academy of Management Review*, 24(3), 522–537. <https://doi.org/10.5465/AMR.1999.2202135>
- DCMI. (2005). Using Dublin Core.
- DDI. (2017). Document, Discover and Interoperate. Recuperado el 14 de noviembre de 2017, de <http://www.ddialliance.org/>
- De Castro, P., Shearer, K. y Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. En J. Keith, A. Clements, P. De Castro y D. Luzi (Eds.), *Procedia Computer Science* (Vol. 33, pp. 39–46). Rome, Italy: Elsevier. <https://doi.org/10.1016/j.procs.2014.06.007>
- DGESU. (2017). Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP). Recuperado el 25 de octubre de 2017, de <http://www.dgesu.ses.sep.gob.mx/PRODEP.htm>
- DGESU. (2018). Dirección General Educación Superior Universitaria | Inicio. Recuperado el 22 de agosto de 2018, de <http://www.dgesu.ses.sep.gob.mx/PRODEP.htm>
- DOF. (2017). Acuerdo por el que se emite el nuevo reglamento del Sistema Nacional de Investigadores. Recuperado el 20 de noviembre de 2017, de [http://www.dof.gob.mx/nota\\_detalle.php?codigo=5470107&fecha=27/01/2017](http://www.dof.gob.mx/nota_detalle.php?codigo=5470107&fecha=27/01/2017)
- DuraSpace. (2018). DSpace - A Turnkey Institutional Repository Application. Recuperado el 14 de marzo de 2019, de <https://duraspace.org/dspace/>
- Editeur. (2009). ONIX for Books. Recuperado el 14 de noviembre de 2017, de <http://www.editeur.org/83/Overview/>
- Elsevier. (2017). Elsevier Scopus APIs. Recuperado el 18 de enero de 2018, de [https://dev.elsevier.com/sc\\_apis.html](https://dev.elsevier.com/sc_apis.html)
- Elsevier. (2018a). Content- How Scopus Works.

- Elsevier. (2018b). Scopus. Recuperado el 13 de diciembre de 2017, de <https://www.scopus.com/freelookup/form/author.uri>
- Elsevier. (2018c). Who uses Scopus.
- Elsevier. (2019). The largest database of peer-reviewed literature- Scopus. Recuperado el 17 de marzo de 2019, de <https://www.elsevier.com/es-mx/solutions/scopus>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. y Pappas, G. (2007). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*. <https://doi.org/10.1096/fj.07-9492lsf>
- Gómez Dueñas, L. (2010). *Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales: caso biblioteca digital colombiana*. Recuperado de [http://eprints.rclis.org/14878/1/MODELOS\\_DE\\_interoperabilidad\\_BDCOL.pdf](http://eprints.rclis.org/14878/1/MODELOS_DE_interoperabilidad_BDCOL.pdf)
- González-Pereira, B., Guerrero-Bote, V. P. y Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics*, 4, 379–391. <https://doi.org/10.1016/j.joi.2010.03.002>
- González Alcaide, G. y Gómez Ferri, J. (2014). La colaboración científica: principales líneas de investigación y retos de futuro. *Revista española de Documentación Científica*, 37(4), 1–15. <https://doi.org/10.3989/redc.2014.4.1186>
- Google. (2017). Google Académico. Recuperado el 12 de diciembre de 2017, de <https://scholar.google.com.mx/>
- Gottschling, M. (2017). 7 things you should know about...Institutional Repositories, CRIS Systems, and their Interoperability. Recuperado de <https://www.coar-repositories.org/community/events/archive/repository-observatory-third-edition/7-things-you-should-know-about-institutional-repositories-cris-systems-and-their-interoperability/#anchor-what-are-the-main-differences-between-cris-and-irs>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5, 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Guerrero-Sosa, J. D. T., Menendez-Domínguez, V., Castellanos-Bolaños, M.-E. y Curi-Quintal, L. F. (2019). Use of graph theory for the representation of scientific collaboration. En N. T. Nguyen (Ed.), *11th International Conference on Computational Collective Intelligence*. Hedaye: Springer.
- Guerrero-Sosa, J. D. T., Menendez-Domínguez, V. y Uc-Cetina, V. (s/f). Redes neuronales

- recurrentes y su uso para la conversión de cadenas. *Komputer Sapiens*.
- Guerrero-Sosa, J. D. T., Romero, F. P., Serrano-Guerrero, J., Menendez-Dominguez, V. y Castellanos-Bolaños, M. E. (2019). A proposal for a recommender system of scientific relevance. En *International Conference on Information Technology and Quantitative Management*. Granada.
- Guerrero-Sosa, J., Menéndez-Domínguez, V., Castellanos-Bolaños, M. y Gómez-Montalvo, J. (s/f). Use of an Ontological Model to Assess the Relevance of Scientific Production. *IEEE Latin America Transactions*.
- Guerrero, J., Sánchez, D., Menéndez, V., Castellanos, M. E. y Gómez, J. (2019). Tools for interoperability between repositories of digital resources. En L. Gómez Chova, A. López Martínez y I. Candel Torres (Eds.), *Proceedings of INTED2019* (pp. 6292–6300). Valencia: IATED.
- Guerrero Sosa, J. D. T., Menéndez Domínguez, V. H. y Castellanos Bolaños, M. E. (2018a). Indicadores de calidad en investigaciones científicas: Antecedentes. *Abstraction and Application*, 19, 6–24. Recuperado de <https://intranet.matematicas.uady.mx/journal/descargar.php?id=134>
- Guerrero Sosa, J. D. T., Menéndez Domínguez, V. H. y Castellanos Bolaños, M. E. (2018b). Sistema de índices para valorar la calidad de la producción académica y la investigación, a partir de repositorios digitales y metadatos. En M. E. Prieto-Méndez, S. J. Pech-Campos y A. Francesa-Alfaro (Eds.), *X Conferencia Conjunta Internacional sobre Tecnologías y Aprendizaje* (pp. 45–52). Cartago: CIATA.org-UCLM.
- Hamerly, G. y Elkan, C. (2003). Learning the K in K-Means. En S. Thrun, L. K. Saul y B. Schölkopf (Eds.), *NIPS'03 Proceedings of the 16th International Conference on Neural Information Processing Systems* (pp. 281–288). Whistler, Canadá: MIT Press Cambridge.
- Harzing, A.-W. y Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- IOSR. (2019). Main features of CERIF. Recuperado el 15 de octubre de 2019, de <https://www.eurocris.org/cerif/main-features-cerif>
- Ip, A., Morrison, I. y Currie, M. (2001). What is a learning object, technically? En W. Fowler y J. Hasebrook (Eds.), *Proceedings of WebNet 2001- World Conference on the WWW and*

- Internet* (pp. 580–586). Association for the Advancement of Computing in Education (AACE). Recuperado de <https://files.eric.ed.gov/fulltext/ED466945.pdf>
- ISO. (2012). ISO/IEC 19506:2012 - Information technology -- Object Management Group Architecture-Driven Modernization (ADM) -- Knowledge Discovery Meta-Model (KDM). Recuperado el 14 de noviembre de 2017, de <https://www.iso.org/standard/32625.html>
- Jacso, P. (2005). As we may search - Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537–1547.
- Keefer, A. (2008). Los repositorios digitales universitarios y los autores. *Anales de Documentación*, 10(0), 205–214. <https://doi.org/10.6018/1151>
- KNB. (2017). Ecological Metadata Language (EML). Recuperado el 14 de noviembre de 2017, de <https://knb.ecoinformatics.org/#tools/eml>
- Lamarca Lapuente, M. J. (2013). *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. Universidad Complutense de Madrid. Recuperado de <http://www.hipertexto.info/documentos/metadatos.htm>
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- López G., M. del S., Cabrales G., F. y Schmal S., R. (2005). Gestión del conocimiento: una revisión teórica y su asociación con la Universidad. *Panorama Socioeconómico*, 30. Recuperado de <http://www.redalyc.org/articulo.oa?id=39903004>
- López Guzmán, C. (2005). 6. Interoperabilidad de los Repositorios de Objetos de Aprendizaje a través de IMS. Recuperado el 20 de mayo de 2018, de <http://www.biblioweb.tic.unam.mx/libros/repositorios/interoperabilidad.htm>
- Méndez Rodríguez, E. (2006). Dublin Core, metadatos y vocabularios. *El Profesional de la Información*, 15(2), 84–86. <https://doi.org/10.3145/epi.2006.mar.01>
- Méndez Rodríguez, E. M. (1999). RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *Les biblioteques i els centres de documentació*, 487–498. <https://doi.org/10.3145/epi.2006.mar.05>
- Myrseth, P., Stang, J. y Dalberg, V. (2011). A data quality framework applied to e-government metadata: A prerequisite to establish governance of interoperable e-services. En *2011 International Conference on E-Business and E-Government, ICEE2011 - Proceedings* (pp.

- 9–12). <https://doi.org/10.1109/ICEBEG.2011.5881298>
- Nelson, M. y Warner, S. (2015, agosto 1). The Open Archives Initiative Protocol for Metadata Harvesting. Recuperado el 1 de abril de 2018, de <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- Nonaka, I. y Takeuchi, H. (1995). *The Knowledge-creating company. How Japanese companies create the dynamics of innovation*. Oxford University Press. Oxford University Press.
- Noy, N. F. y McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University Knowledge Systems Laboratory Technical Report KSL-01-05. Stanford, California. Recuperado de [https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf)
- NSF. (2017). Picture Yourself at NSF | NSF - National Science Foundation. Recuperado el 1 de diciembre de 2017, de <https://www.nsf.gov/careers/>
- NSF. (2018). Open Archives Initiative Protocol for Metadata Harvesting. Recuperado el 8 de marzo de 2018, de <https://www.openarchives.org/pmh/>
- Núñez Jover, J., Félix Montalvo, L. y Pérez Ones, I. (2006). La gestión del conocimiento, la ciencia, la tecnología y la innovación en la Nueva Universidad: una aproximación conceptual. *Revista Pedagogía Universitaria*, XI(2), 31–43.
- Ochoa, X. y Duval, E. (2009). Quantitative analysis of learning object repositories. *IEEE Transactions on Learning Technologies*, 2(3), 226–238. <https://doi.org/10.1109/TLT.2009.28>
- ORCID. (2017). ORCID. Nuestra Misión. Recuperado el 21 de noviembre de 2017, de <https://orcid.org/about/what-is-orcid/mission>
- Palavitsinis, N. (2013). *Metadata Quality Issues in Learning Repositories*. World Conference on Educational Multimedia, Hypermedia and Telecommunications. Universidad de Alcalá. Recuperado de <http://dspace.uah.es/dspace/bitstream/handle/10017/20664/ThesisPalavitsinis.pdf?sequence=1&isAllowed=y>
- Park, J.-R. (2009). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly*, 47(3–4), 213–228. <https://doi.org/10.1080/01639370902737240>
- Peh, W. C. G. y Ng, K. H. (2008). Basic structure and types of scientific papers. *Singapore Medical Journal*, 49(7), 522–525.

- Prieto Méndez, M. E., Menéndez-Domínguez, V. H. y Vidal Castro, C. L. (2014). Metadata and Ontologies in e-Learning. En Miguel A. Sicilia (Ed.), *Handbook of Metadata, Semantics and Ontologies*. World Scientific Publishing Company.
- PRODEP. (2017). Reconocimiento a Perfil Deseable (VIGENTE). Recuperado el 26 de noviembre de 2017, de [http://www.dgesu.ses.sep.gob.mx/Documentos/ultimosbenef/Perfiles\\_2017.pdf](http://www.dgesu.ses.sep.gob.mx/Documentos/ultimosbenef/Perfiles_2017.pdf)
- RECOLECTA. (2018). OpenAIRE. Recuperado el 8 de marzo de 2018, de <https://recolecta.fecyt.es/open-aire>
- REMEDI. (2017a). Portal REMEDI- Servicios.
- REMEDI. (2017b). Portal REMEDI - Información. Recuperado el 25 de octubre de 2017, de <http://www.remedi.org.mx/portal/acerca.html>
- SciELO. (2018). SciELO- Scientific Electronic Library Online.
- Senso, J. A. y Rosa Piñero, A. de la. (2003). El concepto de metadato: algo más que descripción de recursos electrónicos. *Ciência da Informação*, 32(2), 95–106. <https://doi.org/10.1590/S0100-19652003000200011>
- Sharma, O. P. (2012). Quality Indicators of Scientific Research. *Indian Journal of Microbiology*, 52(2), 305–306. <https://doi.org/10.1007/s12088-012-0246-2>
- Sicilia, M.A., Garcia, E., Pages, C., Martinez, J. J. y Gutierrez, J. M. (2005). Complete metadata records in learning object repositories: some evidence and requirements. *International Journal of Learning Technology*, 1(4), 411–424. <https://doi.org/10.1504/IJLT.2005.007152>
- Smith, E. A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of Knowledge Management*, 5(4), 311–321. <https://doi.org/10.1108/13673270110411733>
- Smith, J. R. y Schirling, P. (2006). Metadata standards roundup. *IEEE Multimedia*, 13(2), 84–88. <https://doi.org/10.1109/MMUL.2006.34>
- Statgraphics. (2017). STATGRAPHICS | Data Analysis Solutions.
- Suber, P. (2006). Open Access in the United States. En N. Jacobs (Ed.), *Open access: Key Strategic, Technical and Economic Aspects* (pp. 149–160). Elsevier.
- Texier, J., De Giusti, M., Oviedo, N., Villarreal, G. y Lira, A. (2012). El uso de repositorios y su importancia para la educación en Ingeniería. En H. J. Hoyer y U. R. Cukierman (Eds.), *World Engineering Education Forum (WEEF) (Buenos Aires, 2012)* (pp. 1–10). Buenos

- Aires, Argentina: edUTecNe. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/22943>
- UADY. (2017a). Repositorio Digital Institucional de la UADY.
- UADY. (2017b). SIGA - Quiénes Somos. Recuperado el 16 de enero de 2018, de <http://www.recursoSIGA.uady.mx/quienes.php>
- UAEM. (2017). Acerca de REDALYC.
- UAEM. (2019). Sistema de Información Científica Redalyc Red de Revistas Científicas de América Latina y el Caribe, España y Portugal.
- UNAM. (2018). Latindex.
- Vergara González, E. P. (2008). Capital intelectual y gestión del conocimiento. En J. Laviña Orueta y L. Mengual Pavón (Eds.), *Libro Blanco de la Universidad Digital 2010* (1a ed., pp. 249–278). Barcelona: Fundación Colección Telefónica.
- Whalen, M. (2008). Rights Metadata Made Simple. *Introduction to Metadata*, 1–8.
- Wolf Iszaevich, G. E., Uriarte Santillán, E., Galina Russell, I. y Miranda Quevedo, P. (2014). Red de Acervos Digitales de la UNAM (RAD-UNAM): Construyendo una red de contenidos universitarios. *Revista Digital Universitaria*, 15(10), 1–15. <https://doi.org/http://ru.tic.unam.mx/handle/123456789/2260>
- Yen, N. Y., Shih, T. K., Chao, L. R. y Jin, Q. (2010). Ranking metrics and search guidance for learning object repository. *IEEE Transactions on Learning Technologies*, 3(3), 250–264. <https://doi.org/10.1109/TLT.2010.15>
- Zar, J. H. (2010). *Biostatistical Analysis* (5a ed.). New Jersey, EUA: Prentice-Hall.