

<https://doi.org/10.23913/ride.v12i24.1144>

Artículos científicos

Uso de la minería de datos para la caracterización de investigadores y cuerpos académicos

Use of Data Mining to Characterize Researchers and Academic Groups

Uso de mineração de dados para caracterizar pesquisadores e órgãos acadêmicos

Víctor H. Menéndez Domínguez

Universidad Autónoma de Yucatán, México

mdoming@correo.uady.mx

<https://orcid.org/0000-0003-3587-1263>

Jared D.T. Guerrero Sosa

Universidad Autónoma de Yucatán, México

jared.guerrero@correo.uady.mx

<https://orcid.org/0000-0001-7999-9870>

María Enriqueta Castellanos Bolaños

Universidad Autónoma de Yucatán, México

enriqueta.c@correo.uady.mx

<https://orcid.org/0000-0001-6294-5948>

José William Cervera Pérez

Universidad Autónoma de Yucatán, México

w.cervera.p@gmail.com

<https://orcid.org/0000-0002-1378-3070>

Resumen

En este trabajo se presentan los comportamientos y tendencias en la producción académica de profesores asociados a una universidad pública del sureste de México. Se empleó una aplicación propia que recopila y procesa la producción de los académicos publicada en Scopus. Para el análisis de la información se utilizaron técnicas de minería de datos para clasificar la producción de cuerpos académicos, la relación de académicos activos con base en el número de publicaciones recientes y su grado de certificación, los grupos definidos según los campus de la universidad, así como reglas para identificar las tendencias de los grupos de investigadores. Uno de los descubrimientos fue la correlación del número de integrantes de un grupo con su producción, y la correlación entre la producción indizada y no indizada de los académicos. Si bien el estudio fue hecho para una universidad en particular, la metodología puede ser reproducida para situaciones similares.

Palabras clave: cuerpos académicos, procesamiento de datos, producción científica, universidades.

Abstract

In this work, the behaviors, and trends in the production of researchers associated with an important public university in south-eastern Mexico is presented. A proprietary application was used that collects and processes the production of the academics published in Scopus. For the analysis of the information, data mining techniques were used to classify the production of academic bodies, the ratio of active academics based on the number of recent publications and their certification degree, the groups defined according to the university campuses, as well as rules to identify trends in researchers' groups. One of the discoveries was the correlation between the number of group members with their production, and the correlation between the indexed and non-indexed researchers' production. Although the study was done for a particular university, the methodology can be reproduced for similar situations.

Keywords: academic groups, data processing, scientific production, universities.

Resumo

Este artículo presenta los comportamientos y tendencias en la producción académica de profesores asociados a una universidad pública en el sureste de México. Se utilizó un aplicativo propio que recolecta y procesa la producción académica publicada en Scopus. Para el análisis de la información, se utilizaron técnicas de minería de datos para clasificar la producción de los cuerpos académicos, la proporción de académicos activos con base en el número de publicaciones recientes y su grado de certificación, los grupos definidos de acuerdo con los campos universitarios, así como reglas para identificar tendencias en grupos de investigación. Una de las descubiertas fue la correlación del número de miembros de un grupo con su producción, y la correlación entre la producción indexada y no indexada de los académicos. Aunque el estudio se realizó para una determinada universidad, la metodología puede ser reproducida para situaciones similares.

Palabras-clave: órganos académicos, procesamiento de datos, producción científica, universidades.

Fecha Recepción: Agosto 2021

Fecha Aceptación: Febrero 2022

Introducción

Hoy en día, un factor relevante para determinar la relevancia de una universidad radica en la producción científica que esta genera (Leahey, 2016). Este atributo comúnmente se mide con base en el número de publicaciones que producen los investigadores de forma individual o grupal, pero también pueden intervenir otros atributos como dónde se publica, el número de autores y el número de citas, entre otros (Menéndez, Guerrero, Castellanos y Zurita, 2020). Comúnmente, este proceso está asociado a uno o más repositorios digitales: depósitos de archivos digitales que tienen diferentes clasificaciones y pueden ser accedidos, difundidos y preservados (Texier, De Giusti, Oviedo, Villarreal y Lira, 2012).

Normalmente, el análisis de las publicaciones de los investigadores se realiza de manera manual. Se consideran los resultados obtenidos del mecanismo de búsqueda que ofrece el repositorio, el cual conserva almacenados los metadatos para cada publicación conforme a un formato estándar de descripción (Chuttur, 2014). Esto conlleva la posibilidad de cometer errores en la captura de información para la búsqueda, en la selección de los productos científicos o al interpretar los resultados (Cechinel, Sánchez y Sicilia, 2009). Dicho margen de error puede influir en la caracterización de la producción de los investigadores.

En este sentido, las técnicas de extracción de conocimiento, específicamente la minería de datos, pueden resultar relevantes para identificar los comportamientos de los investigadores asociados a una institución. Es así como el objetivo de esta investigación fue analizar la producción científica de los académicos de una institución utilizando algoritmos de minería de datos con la finalidad de identificar los patrones y tendencias de sus actividades de investigación.

Como caso de estudio se utilizó la producción científica de los académicos de la Universidad Autónoma de Yucatán (UADY), una importante institución pública del sureste de México. En suma, se emprendió una búsqueda de nueva información que normalmente no podría ser obtenida en un análisis realizado de forma manual.

Estado del conocimiento

Producción científica y repositorios

Algo que caracteriza a la ciencia contemporánea es la constante colaboración entre científicos en proyectos multidisciplinarios y transdisciplinarios (González y Gómez, 2014). Sin duda se trata de una ventaja para analizar conceptos y concebir otros. Huelga decir que este grupo de personas conforma lo que se conoce como *comunidad científica*. De igual modo, existen diferentes grupos de investigadores en universidades que comparten intereses y realizan actividades colaborativas y a los cuales se les conoce como *cuerpos académicos* o *grupos de investigación*. Dichas comunidades y cuerpos (al igual que los investigadores individualmente) generan diversos tipos de productos científicos: artículos de revista, libros, ponencias, entre otros, y pueden ser analizados desde distintas perspectivas (Guerrero, Menéndez, Castellanos y Curi, 2019).

Esta producción se suele encontrar en uno o varios repositorios digitales propiedad de la institución donde el investigador se encuentra adscrito (Guerrero, Menéndez y Castellanos, 2018). En México existe el Repositorio Nacional (<https://www.repositorionacionalcti.mx/>), el cual se define como una plataforma digital encargada de proporcionar acceso abierto a una amplia variedad de recursos de información de tipo académica, científica y tecnológica generada en México (Consejo Nacional de Ciencia y Tecnología [Conacyt], 2017). Este repositorio integra los repositorios institucionales (plataformas digitales de instituciones pertenecientes a los sectores social, privado y gubernamental) y sus respectivos autores.

A nivel internacional, Scopus es una de las mayores bases de datos de citas y resúmenes de literatura científica revisada por pares: revistas científicas, libros y actas de conferencias (Elsevier, 2020). Cuenta con más de 70 millones de recursos, 70 000 perfiles de instituciones y 16 millones de perfiles de autores (Elsevier, 2019). Ofrece un exhaustivo resumen de los resultados de la investigación mundial en diversos campos de la ciencia, por lo que numerosas instituciones y organizaciones lo utilizan para conocer a través de sus indicadores la productividad de sus integrantes.

Scopus ha logrado reconocimiento debido a que integra en sus índices de revistas un importante número de títulos correspondientes a países en vías de desarrollo, combinando tanto las revistas de alcance internacional como regional y local (Luna, Luna y Luna, 2018).

La calidad del investigador en México

El Sistema Nacional de Investigadores (SNI) fue creado para reconocer los diversos trabajos hechos por personas dedicadas al ámbito científico y tecnológico en México (Conacyt, 2019). El SNI presenta tres distinciones que un miembro puede obtener (Conacyt, 27 de enero de 2017).

- *Candidato a Investigador Nacional*. Tiene productos y publicaciones en los ámbitos científicos o tecnológicos.
- *Investigador Nacional*. Se divide en tres niveles:
 - Nivel 1. Tiene productos científicos o tecnológicos de calidad, dirige tesis de licenciatura o posgrado, o imparte asignaturas, y participa en otras actividades docentes.
 - Nivel 2. Además de lo necesario para pertenecer al nivel I, colabora con otros investigadores en productos originales y de calidad para evidenciar alguna línea de investigación, así como dirige tesis de posgrado y forma recursos humanos.
 - Nivel 3. Además de lo necesario para pertenecer al nivel II, tiene investigación que cause impacto en la actualidad, realiza actividades de liderazgo nacional en ciencia y tecnología, y tiene 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, por lo menos 15 años adscrito al

SNI y tres evaluaciones ininterrumpidas obteniendo la distinción Investigador Nacional nivel 3.

Cada aspirante debe cumplir con los lineamientos establecidos por el sistema, además de contar con el grado de doctor.

Por otra parte, el Programa para el Desarrollo Profesional Docente para el Tipo Superior (Prodep) pretende profesionalizar a los profesores de tiempo completo (PTC) en México para que formen cuerpos académicos y realicen actividades de docencia, investigación y desarrollo de la tecnología e innovación haciendo uso de la responsabilidad social (Dirección General de Educación Superior Universitaria e Intercultural [DGESUI], 2014). El perfil Prodep se otorga a aquellos académicos con posgrado que realizan investigación además de docencia y tutoría.

El Prodep también considera la conformación de grupos de investigación, denominados *cuerpos académicos*, y los clasifica en tres grupos (Programa de Mejoramiento del Profesorado [Promep], 2020). 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 en formación (CAEF).** Son cuerpos académicos que nacen a partir de una o más líneas de investigación y se encuentran en una etapa temprana. Sus características son: 1) los integrantes están identificados, 2) al menos la mitad de sus integrantes cuentan con el perfil 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.

- **Cuerpo académico en consolidación (CAEC).** Es el nivel intermedio en el que se puede clasificar un cuerpo académico. Se caracteriza por: 1) más de la mitad de sus integrantes cuentan con doctorado, 2) cuentan con productos académicos de calidad derivados de las líneas de investigación consolidadas, 3) al menos la tercera parte de sus integrantes cuentan con el perfil Prodep, 4) participan de forma conjunta en líneas de investigación o aplicación del conocimiento, 5) amplia experiencia en docencia y en formación de recursos humanos y 6) colaboran con otros cuerpos académicos.

- **Cuerpo académico consolidado (CAC).** Es el nivel máximo que puede alcanzar un cuerpo académico. Sus características son: 1) la mayoría de sus integrantes cuenta doctorado, 2) amplia experiencia docente, 3) la mayoría de sus integrantes cuentan

perfil Prodep, 4) colaboración y producción científica y académica, 5) participan en congresos, seminarios, mesas, talleres, etcétera, de forma regular y frecuente y 6) intensa participación en redes de intercambio académico.

En resumen, las instancias mexicanas expuestas, el SNI y el Prodep, son opciones que los investigadores mexicanos 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 académico 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.

Minería de datos

La minería de datos es un campo de la estadística y las ciencias de la computación. A través de diversas técnicas, se extrae información de una base de datos para generar conocimiento, el cual puede ser expresado a través de conceptos, reglas, leyes, patrones, entre otros (Romero y Ventura, 2020).

La minería de datos es un tema que implica el aprendizaje práctico, no tanto teórico (Witten, Frank y Hall, 2011); busca técnicas para encontrar y describir patrones estructurales en los datos. Es una herramienta para ayudar a explicar dichos datos y hacer predicciones de ellos. De acuerdo con Romero y Ventura (2006), existen tres técnicas básicas para descubrir patrones y conocimiento:

1) **Clasificación.** Consiste en determinar nuevos patrones con base en un conjunto de datos previamente identificados. Algunos de los algoritmos más usados para esta técnica son: ID3, J48, C4.5, Naive Bayes, algoritmos evolutivos, entre otros.

2) **Agrupamiento.** Su objetivo principal es concentrar datos que tienen características similares. Para ello, se analizan los datos almacenados en la base de datos, y de acuerdo con reglas de clasificación, se genera una colección de recursos agrupados en clases. Algunos de los algoritmos más representativos son: Single-link, Complete-link, SimpleKMeans, Kmedia, entre otros.

3) **Asociación.** Su objetivo principal es establecer reglas que asocien los valores de atributos diferentes de una misma base de datos. La asociación y la correlación son usados para la búsqueda de un ítem frecuente de entre una gran cantidad de información. Algunos de los algoritmos más representativos son: Apriori, Predictive A priori, entre otros.

Metodología

La metodología empleada fue el descubrimiento de conocimiento en bases de datos (KDD, por sus siglas en inglés) (Guarascio, Manco y Ritacco, 2019). Consta de cinco fases (Camacho, Zapata, Menéndez y Canto 2018), las cuales se describen a continuación (figura 1).

Figura 1. Fases de la propuesta metodológica



Fuente: Elaboración propia

1) *Selección*. Se compone de dos subfases. La primera consiste en el aprendizaje del dominio del conocimiento, sobre todo aquel que resulte relevante y las metas de la aplicación. La segunda consiste en seleccionar las bases de datos objetivo. En este caso, aquellas que almacenan la producción científica y las que contienen información de los investigadores.

2) *Preprocesamiento*. Consiste en utilizar operaciones básicas que permitan depurar los datos que no se requieren, seleccionar los necesarios y los que podrían ser útiles. En este caso, los datos elementales de cada investigador y los datos descriptivos de cada publicación (título, autores, palabras clave, identificadores, entre otros).

3) *Transformación*. En esta etapa, diversos campos de la base de datos de tipo numérico se transforman en etiquetas lingüísticas para una mejor caracterización. Algunas técnicas útiles son el uso de percentiles para asociar rangos a claves.

4) *Minería de datos*. Se utilizan diversas técnicas de minería de datos, de acuerdo con las necesidades de la problemática planteada. Por cada técnica, se estudian algoritmos

relacionados y se selecciona el adecuado en cada caso. Entre las técnicas utilizadas se encuentran: clasificación, agrupamiento y reglas de asociación.

5) *Análisis de los resultados.* Se examinan los resultados generados, lo cual concluye en la generación de nuevo conocimiento respecto a cuerpos académicos e investigadores. A partir de esto, es posible llevar a cabo la toma de decisiones.

Es importante señalar que no se emplean instrumentos de recolección de datos, dado el origen de la información que será analizada. Los datos son recolectados de forma automática mediante una aplicación web que accede a la información almacenada en los repositorios, para luego ser procesados según la metodología descrita.

Caso de estudio

La UADY es una de las instituciones a nivel superior más importantes en el sureste de México. Hasta febrero del 2019, en la UADY se registraron 824 profesores de tiempo completo y 78 cuerpos académicos, distribuidos en 15 facultades y dos centros de investigación que se agrupan en seis campus (UADY, 2020). En la tabla 1 se presentan algunas estadísticas que los profesores de la UADY han reportado a febrero de 2019. Para propósito de este trabajo denominaremos *investigador* al profesor de tiempo completo que cuenta con producción.

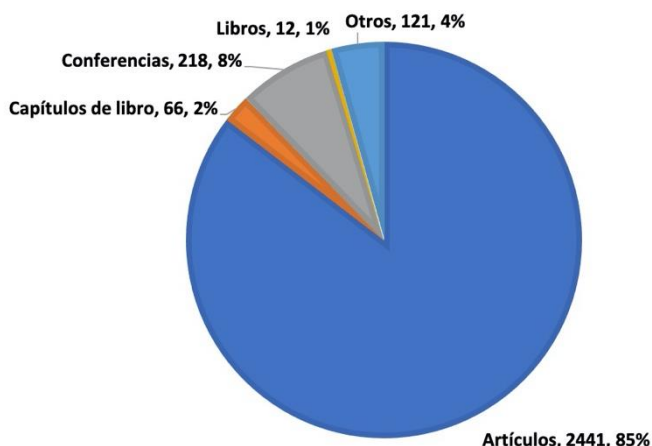
Tabla 1. Algunas estadísticas de los profesores de la UADY

Profesores con producción	438
Doctores	280
Profesores en el SIN	193
Doctores con SIN	64.5 %
Producción promedio de los doctores	15.5 publicaciones
Producción promedio de doctores SNI	19.73 publicaciones
Doctores en cuerpos académico	7.85 %
Doctores SNI pertenecientes a un cuerpo académico	78.75 %

Fuente: Elaboración propia

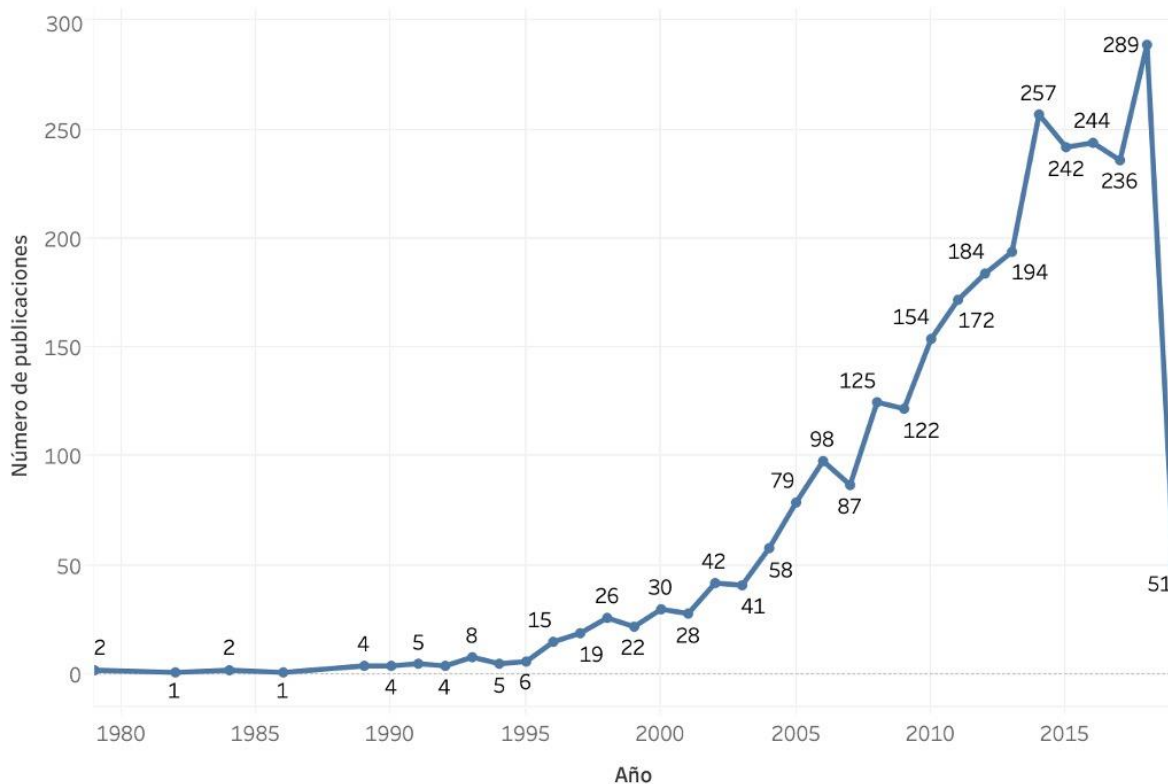
En las figura 2 y 3 se presenta la distribución de la producción de la UADY almacenada en Scopus y el repositorio nacional en el período comprendido desde la primera publicación que resultó estar en Scopus (1979) a febrero de 2019.

Figura 2. Distribución de la producción científica de la UADY por tipo



Fuente: Elaboración propia

Figura 3. Producción científica de la UADY por año



Fuente: Elaboración propia

Selección

La información se obtuvo mediante una aplicación web propia, basada en el lenguaje de programación Python (McKinney, 2017), que recupera la producción científica de la UADY a través de una interfaz de consulta de Scopus y del Repositorio Nacional. Utiliza una base de datos en MongoDB (MongoDB, 2019) para almacenar también la información relevante de los 78 cuerpos académicos. En la tabla 2 se describen los atributos utilizados para cuerpos académicos.

Tabla 2. Atributos para cuerpos académicos

Atributo	Descripción
Nombre	Nombre del cuerpo académico
número_de_integrantes	Número de integrantes
Facultad	Facultad a la que pertenece
Campus	Campus al que pertenece
Tipo	Tipo de cuerpo académico: CAEF: cuerpo académico en formación CAEC: cuerpo académico en consolidación CAC: cuerpo académico consolidado
artículos_indizados	Número de artículos indizados por Scopus
capítulos_indizados	Número de capítulos indizados por Scopus
libros_indizados	Número de libros indizados por Scopus
otros_indizados	Número de productos de otra índole indizados por Scopus
producción_repositorio	Número de productos almacenados en el Repositorio Nacional
total_de_publicaciones	El total de producción indizada y no indizada de los cuerpos académicos

Fuente: Elaboración propia

La tabla 3 presenta la distribución de cuerpos académicos por campus y el promedio de producción indizada por Scopus y el Repositorio Nacional.

Tabla 3. Distribución de los cuerpos académicos en los campus de la UADY y la producción promedio

Campus	CAC	Producción promedio	CAEC	Producción promedio	CAEF	Producción promedio
Campus de Arquitectura, Hábitat y Diseño	1	0	2	0	0	N/A
Campus de Ciencias Biológicas y Agropecuarias	8	10.3	2	12.5	0	N/A
Campus de Ciencias de la Salud	4	8.75	8	3.62	1	0
Campus de Ciencias Exactas e Ingenierías	10	11.88	9	7.22	2	1
Campus de Ciencias Sociales, Económico Administrativas y Humanidades	9	3.44	9	0	3	0
Centro de Investigaciones Regionales Dr. Hideyo Noguchi	6	19.66	2	17	2	1.5

Fuente: Elaboración propia

Se recuperaron los atributos listados en la tabla 4 para los 438 profesores que cuentan con producción científica.

Tabla 4. Atributos para profesores

Atributo	Descripción
Prodep	Verifica si el profesor cuenta con perfil Prodep (sí o no)
último_grado	Último grado de estudios (licenciatura, maestría o doctorado)
Género	Género (masculino o femenino)
Sin	Nivel de SNI con el que cuenta el profesor (candidato, nivel 1, nivel 2, nivel 3 o no)
cuerpo_académico	Cuerpo académico al que pertenece
Facultad	Facultad a la que pertenece
Campus	Campus al que pertenece
Cuartil	Cuartil de citas al que pertenece
Activo	Verifica si el profesor se encuentra activo. Para ello, deberá tener al menos tres publicaciones en los tres años anteriores.
total_scopus	Número total de publicaciones indizadas por Scopus
total_repositorio	Número total de publicaciones indizadas por el Repositorio Nacional

Fuente: Elaboración propia

En la tabla 5 se puede apreciar los grupos de investigadores según su grado académico, así como su distribución en los campus.

Tabla 5. Distribución de los profesores con producción en los campus de la UADY

	Campus de Ciencias Biológicas y Agropecuarias	Campus de Ciencias de la Salud	Campus de Ciencias Exactas e Ingenierías	Campus de Ciencias Sociales, Económico Administrativo y Humanidades	Centro de Investigaciones Regionales Dr. Hideyo Noguchi
Profesores con producción	76	86	159	54	63
Doctores	59	34	98	41	47
Profesores con perfil Prodep	75	84	157	54	61
Profesores no adscritos al SIN	39	68	94	17	27
Candidato al SIN	0	6	10	10	2
SNI 1	25	9	46	19	25
SNI 2	6	3	7	7	9
SNI 3	6	0	2	1	0
Profesores SNI con cuerpo académico	33	13	54	23	29

Fuente: Elaboración propia

Preproceso

Para llevar a cabo el proceso de minería de datos, el sistema desarrollado genera archivos .csv y .arff a partir de la base de datos. Los archivos recopilan la información relevante para el propósito de la investigación: para los cuerpos académicos lista el nombre, número de integrantes, campus, tipo, producción indizada por los dos repositorios y total de publicaciones; para profesores se lista el último grado, perfil Prodep, género, nivel de SNI, cuerpo académico, campus, total de publicaciones, producción indizada por los dos repositorios, cuartil de citas y actividad de cada uno.

Transformación

Se categorizaron algunos atributos numéricos para cuerpos académicos y profesores. Esto con el fin de clasificar los datos para su uso en la siguiente fase del proceso. A continuación, se describen los principales.

Para cuerpos académicos:

- *Campus*. Este es un nuevo atributo. Para obtenerlo, se tuvo como base el atributo “facultad”, donde, dependiendo de la facultad del cuerpo académico, se asignó el campus correspondiente.

- *Total de publicaciones*. Se definieron categorías dependiendo del número de publicaciones totales de un cuerpo académico. Las etiquetas son:

- Muy poco: de 0 a 20 publicaciones.
- Poco: de 21 a 41 publicaciones.
- Regular: de 42 a 62 publicaciones.
- Mucho: de 63 a 83 publicaciones.
- Demasiado: más de 84 publicaciones.

Para profesores:

- *Cuartil*. Asigna una letra del alfabeto dependiendo del número de citas que un académico tenga:

- De cero a dos citas.
- De tres a cinco citas.
- De seis a nueve citas.
- De 10 a 19 citas.
- De 20 a 29 citas.

- De 30 a 39 citas.
- De 40 a 79 citas.
- De 80 a 932 citas.
- Más de 933 citas.
- *Total Scopus*. Se tiene una categorización con el fin de asignar un valor dependiendo del número de publicaciones indizadas por Scopus. Dichos valores son:
 - Muy poco: de 0 a 30 publicaciones.
 - Poco: de 31 a 60 publicaciones.
 - Regular: de 61 a 90 publicaciones.
 - Mucho: de 91 a 120 publicaciones.
 - Demasiado: más de 121 publicaciones.
- *Total repositorio*. Para tener una mejor representación de datos, se manejó una escala y así se pudo asignar una etiqueta dependiendo del número de publicaciones de un investigador indizadas por el Repositorio Nacional. Dichas etiquetas son:
 - Muy poco: de cero a una publicación.
 - Poco: de dos a tres publicaciones.
 - Regular: de cuatro a cinco publicaciones.
 - Mucho: de seis a siete publicaciones.
 - Demasiado: de ocho a nueve publicaciones.

Es importante mencionar que el promedio de producción total de los cuerpos académicos analizados está en el rango de 0 a 20 (etiqueta “Muy poco”), y lo mismo ocurre con los investigadores (0 a 30, etiqueta “Muy poco”).

Minería de datos

Para el análisis de los archivos, se utilizó el *software* WEKA (Hall *et al.*, 2009). En dicha herramienta se aplicaron los algoritmos J48 (clasificación), SimpleKMeans (agrupamiento) y A priori (asociación) (Witten *et al.*, 2011) y se obtuvo una colección de árboles de clasificación, patrones y reglas, que se expondrán más adelante.

Análisis de los resultados de las técnicas de minería de datos

En esta fase se analizaron los resultados generados con las tres técnicas de minería de datos presentadas previamente.

Resultados

Esta sección contiene los resultados de la implementación de la minería de datos (con ayuda de los algoritmos de clasificación, agrupamiento y asociación) con el *software* WEKA a partir de la información generada de la producción de cuerpos académicos e investigadores. En primera instancia, se aplicaron los algoritmos a los cuerpos académicos de la UADY, y en segunda instancia, se realizó lo mismo para sus académicos. El orden que se siguió fue aplicar algoritmos de clasificación para generar un árbol de clasificación, luego algoritmos de agrupamiento para dividir por grupos tanto a cuerpos académicos como investigadores, y finalmente extraer reglas con ayuda del algoritmo de asociación. A continuación, se encuentran los resultados obtenidos.

Algoritmo de clasificación

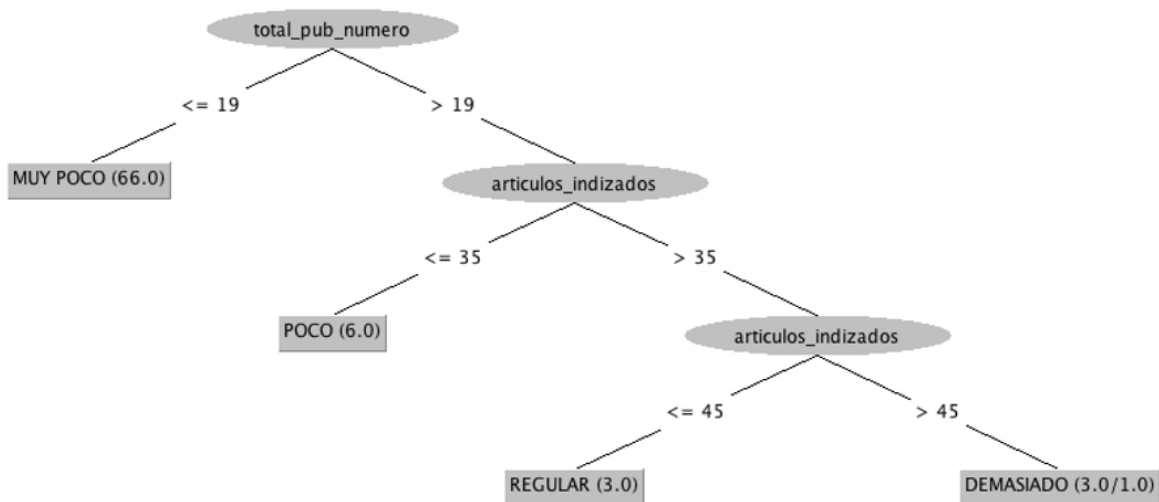
La problemática de clasificación de individuos o entidades ha sido de gran interés para investigaciones (Romero y Ventura, 2010, 2020). En este trabajo se experimentó con el algoritmo J48.

Primero, se validó con los cuerpos académicos de la universidad. Para ello, se utilizó como atributo principal el total de publicaciones, las cuales se encontraban en un intervalo de 0 a 103 publicaciones. El algoritmo J48 para este experimento tiene un grado de corrección de 91.0256 %. Una vez aplicado, se obtuvieron los siguientes resultados (representados en la figura 4):

- Si el total de publicaciones es menor o igual a 19, entonces son muy pocas.
- Si el total de publicaciones es mayor a 19 y menor o igual a 35, siendo artículos indizados, entonces son pocas.
- Si el total de publicaciones es mayor a 35 y menor o igual a 45, siendo artículos indizados, entonces es regular.
- Si el total de publicaciones es mayor a 45, siendo artículos indizados, entonces son demasiadas.

Analizando estos resultados, se puede apreciar que el algoritmo ha desechado la clasificación “Mucho” y “Demasiado”, así como otros recursos indizados por Scopus o el Repositorio Nacional, esto debido a la mínima cantidad que estos representan con respecto al total de la producción analizada (figura 4).

Figura 4. Árbol de clasificación J48 para el total de publicaciones de los cuerpos académicos de la UADY



Fuente: Elaboración propia

En segunda instancia, se experimentó con los investigadores de la universidad. Para ello, se utilizó como atributo principal si el profesor se encuentra activo o no y la producción generada en el período de enero de 2016 hasta enero de 2019. Para efectos del experimento, se considera que un profesor es activo si cuenta con al menos tres publicaciones desde enero de 2016 hasta enero de 2019.

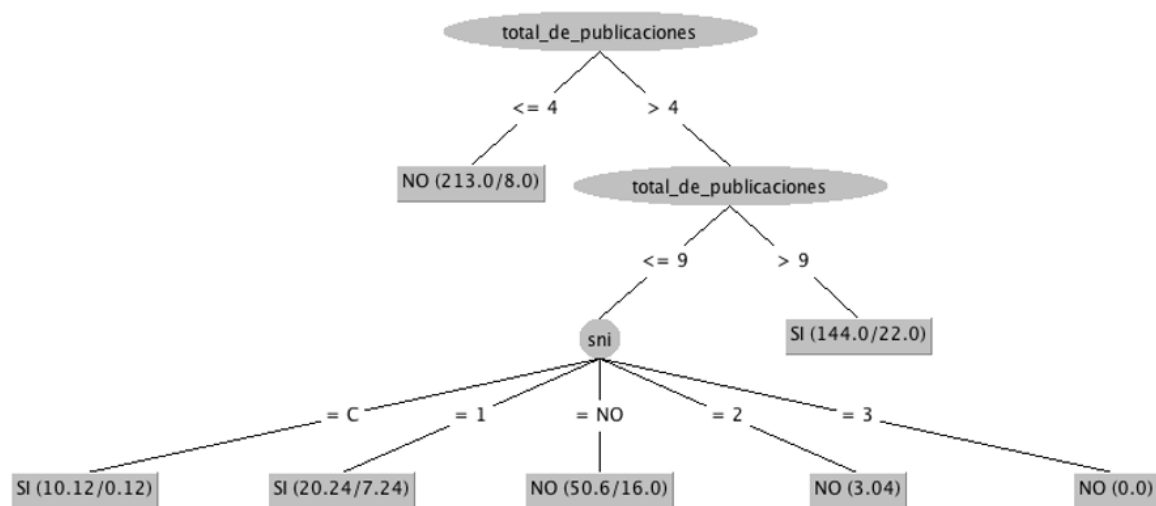
El algoritmo J48 para este experimento tiene un grado de corrección de 84.3537 %. Al ser aplicado, se obtuvieron los siguientes resultados (representados en la figura 5):

- Si el total de publicaciones de un profesor es menor o igual a cuatro, entonces no se encuentra activo.
- Si el total de publicaciones de un profesor es mayor a cuatro y menor o igual a nueve, y además es candidato a SNI, entonces se encuentra activo.
- Si el total de publicaciones de un profesor es mayor a cuatro y menor o igual a nueve, y además cuenta con el nivel 1 de SNI, entonces se encuentra activo.
- Si el total de publicaciones de un profesor es mayor a cuatro y menor o igual a nueve, y además cuenta con el nivel 2 de SNI, entonces no se encuentra activo.
- Si el total de publicaciones de un profesor es mayor a cuatro y menor o igual a nueve, y además cuenta con el nivel 3 de SNI, entonces no se encuentra activo.
- Si el total de publicaciones de un profesor es mayor a cuatro y menor o igual a nueve, y además no cuenta con SNI, entonces no se encuentra activo.

- Si el total de publicaciones de un profesor es mayor a nueve, entonces se encuentra activo.

Para los niveles 2 y 3 de SNI no se cuentan con casos suficientes, por lo que el algoritmo J48 considera a los profesores con dichos niveles como no activos (figura 5).

Figura 5. Árbol de clasificación J48 para el total de publicaciones de profesores de la UADY

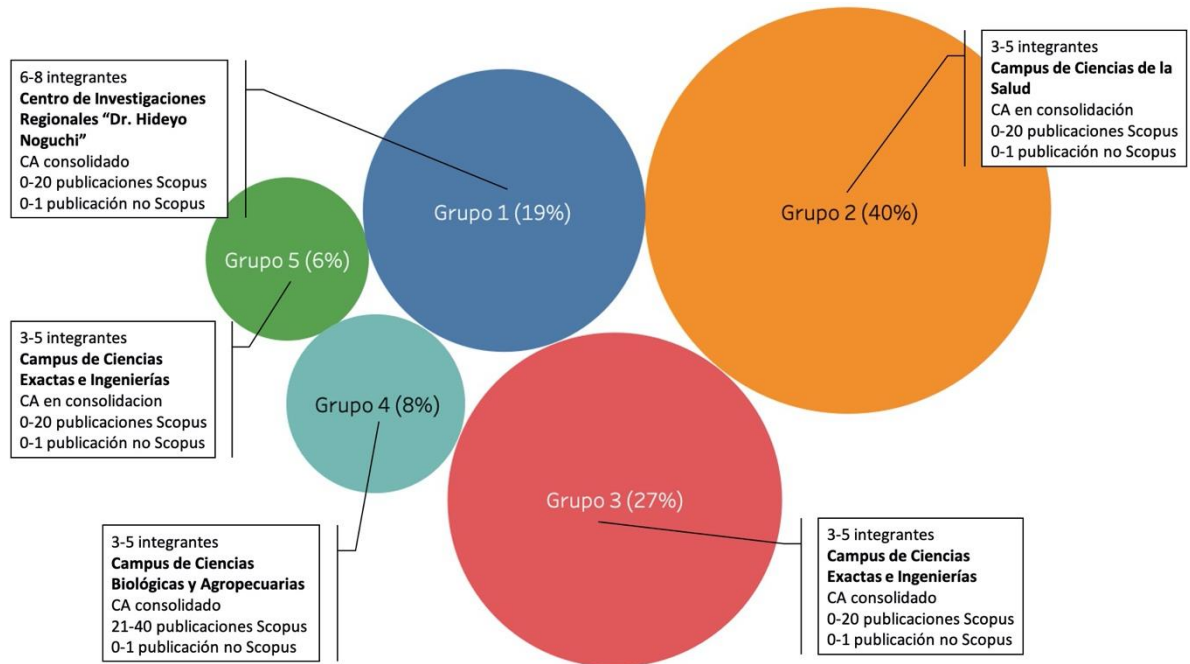


Fuente: Elaboración propia

Algoritmo de agrupamiento

Se experimentó con SimpleKMeans como algoritmo de agrupamiento. Para ello, se probó crear cinco grupos de cuerpos académicos, esperando que la separación de los grupos fuera con base en los campus con los que cuenta la universidad. Sin embargo, esto no fue así (figura 6), pues varios cuerpos académicos no fueron considerados relevantes para el algoritmo y fueron absorbidos por otros con mayor peso.

Figura 6. Generación de cinco grupos de cuerpos académicos con el algoritmo SimpleKMeans



Fuente: Elaboración propia

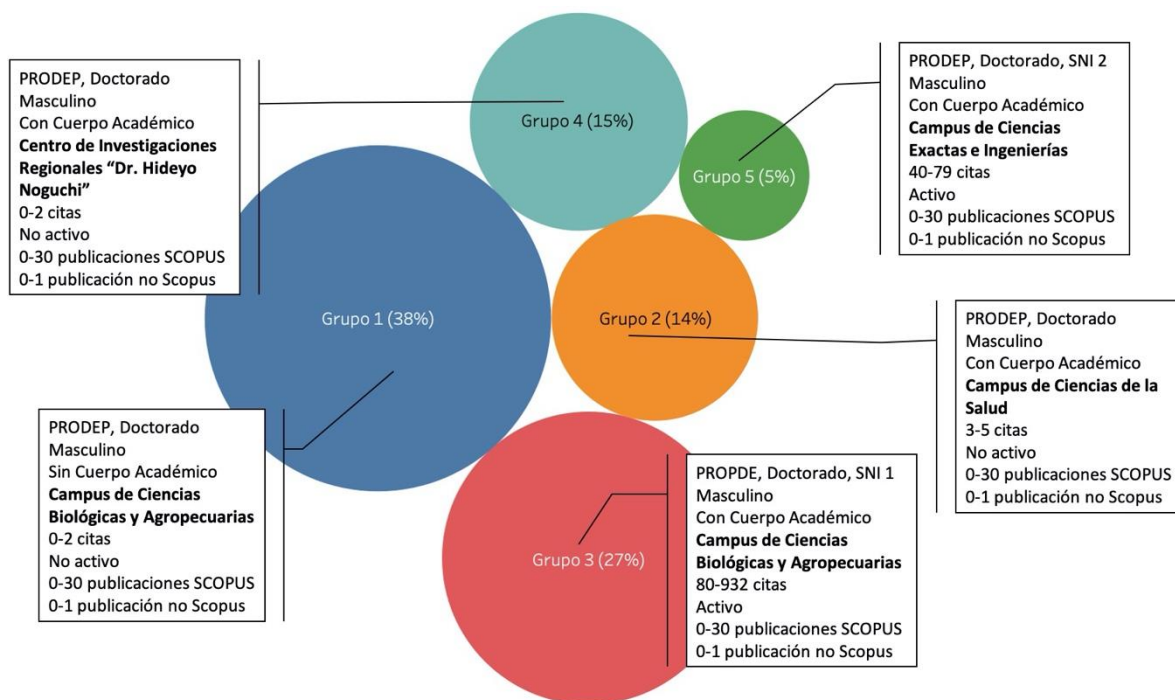
Los grupos representativos reportados en la figura 6 señalan una alta presencia de cuerpos académicos consolidados (54 %), seguida de cuerpos académicos en consolidación (46 %). El número de integrantes en su mayoría es pequeño (entre tres y cinco [81 %]) a excepción del grupo 1 (que tiene entre seis y ocho integrantes). Si bien casi todos los grupos tienen una producción entre 0 y 20 publicaciones indizadas en SCOPUS, el grupo 4 (8 %) tiene un intervalo de 21-40 publicaciones en Scopus, están consolidados y pertenecen al Campus de Ciencias Biológicas y Agropecuarias. El grupo más significativo de cuerpos académicos (32 [40 %]) está en proceso de consolidación con tres a cinco integrantes, pertenece al Campus de Ciencias de la Salud y tiene de 0 a 20 publicaciones Scopus.

En segunda instancia, se experimentó de igual modo con la creación de cinco grupos representativos de investigadores esperando que se agruparan por campus. Nuevamente no fue así, pues los académicos del Campus de Ciencias Biológicas y Agropecuarias poseen una mayor producción, por lo que prevalecen sobre académicos de otros campus y provocan que estos últimos no sean relevantes para el algoritmo.

En la figura 7 se puede observar que el grupo más significativo (38 %) agrupa a doctores que tienen perfil Prodep sin cuerpo académico, cuyo nivel de citación es muy bajo y un número de entre 0 a 30 publicaciones. En dos grupos los profesores publican cada año,

pues tienen SNI, entre 0 a 30 publicaciones, con un alto nivel de citas (80 a 932 citas para el grupo 3 [27 %]; 40 a 79 citas para el grupo 5 [5 %]). Todos los grupos representativos de investigadores tienen Prodep y doctorado como último grado de estudios, pero la mayoría no cuentan con SNI (tres grupos).

Figura 7. Generación de cinco grupos de profesores con el algoritmo SimpleKMeans



Fuente: Elaboración propia

Algoritmos de asociación

La generación de reglas de asociación es la última técnica de minería de datos utilizada en este caso de estudio. Se utilizó el algoritmo Apriori para observar el comportamiento de los cuerpos académicos y profesores de la UADY. Para ello, se generaron 25 reglas para profesores y 25 reglas para cuerpos académicos, de las cuales se presentan las 10 más confiables (tabla 6).

Las reglas tienen uno o más antecedentes que generan un consecuente. Por cada componente de una regla se establece el número de casos que fueron considerados para generarla, lo que da el índice de confianza de esta.

Muchas de las reglas confirman supuestos o la correlación existente entre los atributos. Por ejemplo, el número de integrantes de un cuerpo académico, el nivel de

certificación alcanzado y su productividad o entre investigadores, su grado, el perfil Prodep y su actividad.

Tabla 6. Reglas generadas con el algoritmo Apriori

Antecedente	Consecuente	Interpretación en lenguaje natural	Índice de confianza
no_integrantes = Pocos tipo = CAEC (28)	total_publicaciones = Muy poco (28)	Si el cuerpo académico está en consolidación y el número de integrantes es poco, entonces el grupo tiene muy poca producción.	1
último_grado = D (279)	prodep = Sí (278)	Si el último grado del investigador es doctorado, entonces cuenta con perfil Prodep.	1
activo = No (272)	total_scopus = Muy poco (271)	Si el investigador no se encuentra activo, entonces tiene muy pocas publicaciones en Scopus.	1
total_socups = Muy poco total_repositorio = Muy poco (356)	prodep = Sí (349)	Si un investigador cuenta con muy pocas publicaciones Scopus y Repositorio Nacional, entonces cuenta con perfil Prodep.	0.98
tipo = CAEC (32)	total_publicaciones = Muy poco (31)	Si el cuerpo académico está en consolidación, entonces tiene muy poca producción.	0.97

<p>producción_scopus = Muy poco (66)</p>	<p>producción_repositorio = Muy poco (62)</p>	<p>Si la producción en Scopus de un cuerpo académico es muy poca, entonces su producción en el Repositorio Nacional será muy poca.</p>	<p>0.94</p>
<p>no_integrantes = Pocos (61)</p>	<p>producción_scopus = Muy poco (56)</p>	<p>Si el número de integrantes del cuerpo académico es poco, entonces tiene muy poca producción en el Repositorio Nacional.</p>	<p>0.92</p>
<p>no_integrantes = Pocos produccion_repositorio = Muy poco (58)</p>	<p>producción_scopus = Muy poco (53)</p>	<p>Si el número de integrantes del cuerpo académico es poco y tiene muy poca producción en el Repositorio Nacional, entonces tiene muy poca producción en Scopus.</p>	<p>0.91</p>
<p>prodep = Sí total_repositorio = Muy poco (382)</p>	<p>total_scopus = Muy poco (349)</p>	<p>Si un investigador cuenta con perfil Prodep y cuenta con muy pocas publicaciones en el Repositorio Nacional, entonces cuenta con muy pocas publicaciones en Scopus.</p>	<p>0.91</p>

prodep = Sí (434)	total_scopus = Muy poco (395)	Si el investigador cuenta con perfil Prodep, entonces cuenta con muy pocas publicaciones en Scopus.	0.91
----------------------	-------------------------------------	--	------

Fuente: Elaboración propia

Discusión

Para el algoritmo de clasificación se pudo observar un recorte en la publicación total de los cuerpos académicos. Únicamente se centró en un intervalo de 0 a 45, con los cuales se obtuvieron las diferentes etiquetas relacionadas a la cantidad de publicaciones de un cuerpo académico. Este hecho se refuerza con los resultados obtenidos a través de un sistema de índices (Guerrero, Menéndez y Castellanos, 2021) para la evaluación de la producción de un cuerpo académico y de un investigador, pues el promedio de las publicaciones de un cuerpo académico es de 11.25, y la mayoría de estos grupos cuentan con hasta 19 publicaciones.

Mientras que, para los profesores, se pudo observar que el árbol J48 es mucho más específico en los resultados obtenidos, lo que da a entender que los autores con publicaciones menores o iguales a cuatro tienden a ser inactivos. Esto probablemente porque la producción con la que cuenta se realizó previo a los tres años anteriores. Mientras que si son mayores a nueve tienden a ser activos; en el intervalo entre estos dos valores se pueden observar tendencias a no tener SNI, lo cual, como se comentó previamente, se debe a la poca cantidad de datos de profesores de nivel 2 y 3.

El caso de los investigadores activos es complementado con los resultados obtenidos a través de un modelo ontológico para la representación del conocimiento en el dominio de la producción científica (Guerrero, Menéndez, Castellanos y Gómez, 2019). A través de un motor de consultas SPARQL, dicho estudio identificó que la mayoría de los profesores activos pertenecen al SNI en el nivel 1. Por lo que, con lo obtenido en el árbol de clasificación, se infiere que estos profesores tienen más de nueve publicaciones en los últimos tres años.

Cabe mencionar que el árbol de decisión de los cuerpos académicos considera todos los cuerpos académicos, incluso aquellos que a los que no fue posible localizarles producción. En cambio, el árbol de decisión de los profesores excluye todo aquel profesor que no tenga producción.

Para el algoritmo de agrupamiento, se pudo observar el gran impacto que tienen los campus de Ciencias de la Salud, Ciencias Biológicas y Agropecuarias y Ciencias Exactas e Ingenierías, ya que representan un porcentaje muy grande en los resultados obtenidos una vez clasificados en grupos. Estos resultados se presentaron incluso utilizando una mayor cantidad de agrupaciones o clústeres, lo que implica que los demás campus no tienen una gran aportación de artículos científicos para la universidad.

Incluso en otro trabajo donde se estudiaron las colaboraciones entre los profesores de la UADY (independientemente si son cuerpos académicos o no) (Guerrero, Menéndez, Castellanos y Curi, 2020) se confirmó la existencia de colaboraciones en las dependencias de la universidad, incluso entre ellas, generando conocimiento multidisciplinario. El modelo ontológico mencionado previamente infirió que existen publicaciones que abarcan hasta seis áreas de conocimiento de siete diferentes.

De igual modo, se puede ver una tendencia a tener grupos académicos consolidados y profesores con un grado de doctorado, esto refuerza los requisitos del Prodep para los cuerpos académicos consolidados, donde se solicita que la mayoría de los integrantes de un cuerpo académico debe contar con doctorado. Incluso corroboran los resultados de otro estudio sobre la producción y colaboración de los cuerpos académicos (Guerrero, Menéndez, Castellanos y Guerra 2020), el cual indica que, a mayor consolidación, mayor es el número de instituciones nacionales e internacionales colaboradoras. Esto a su vez confirma el grado de consolidación que se les ha asignado, pues la definición de lazos con instituciones externas para generar y difundir nuevo conocimiento forma parte de los criterios de evaluación.

Aunque, de igual modo, se puede ver que, tanto en los cuerpos académicos como en los investigadores, existe una tendencia a tener muy poca producción tanto indizada como no indizada. Se considera muy poca debido a que el valor máximo en esta variable es muy alto. Uno de los datos más interesantes es la tendencia en los grupos de profesores a contar casi en su mayoría con profesores del género masculino, lo que da una idea de la diferencia que existe en los profesores dentro de la universidad.

Finalmente, para el algoritmo de asociación, las reglas generadas para los cuerpos académicos reflejan tendencias proporcionales al número de integrantes de un cuerpo con su

producción. Mientras que para los profesores se obtuvieron reglas diferentes, tales como la relación entre el perfil Prodep con el último grado de estudios, género y número de publicaciones, así como reglas proporcionales con el número de publicaciones indizadas y las no indizadas, esto refuerza el cumplimiento en los requerimientos del perfil Prodep y del SNI.

La tendencia de los investigadores de contar con muy pocas publicaciones es reforzada por el uso de un sistema de índices para la evaluación de la producción de cuerpos académicos y de investigadores (Guerrero, Menéndez y Castellanos, 2021), ya que, en los resultados obtenidos, se indica que en la UADY el número promedio de publicaciones es alrededor de 11 (y muy pocos investigadores cuentan con más de 121 publicaciones en Scopus, lo que se considera una cifra muy alta de publicaciones en dicha base de datos bibliográfica).

Conclusiones

Los repositorios digitales de documentación científica han permitido que la información acerca de las publicaciones realizadas por investigadores esté al alcance de cualquier persona, 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. Su análisis genera información relevante para la toma de decisiones.

Diversas instancias como el SNI y el Prodep, ambos en México, se encargan de evaluar y otorgar reconocimientos a los investigadores por medio de su producción científica almacenada en diversos repositorios.

En este trabajo se ha presentado cómo las técnicas de minería de datos permiten caracterizar a los investigadores a nivel grupal o individual de una institución. Para ello, se definió un grupo de indicadores que considera el número de publicaciones, citas, el prestigio de las revistas y, en menor medida, la producción no indizada y de acceso abierto.

Los resultados obtenidos permitirían corroborar o desechar supuestos relacionados con la productividad científica de una institución, lo que facilita la toma de decisiones que incentiven o condicionen políticas institucionales.

Futuras líneas de investigación

Para investigaciones futuras se pretende corroborar los resultados obtenidos mediante la metodología de minería de datos con nuevos datos en situaciones particulares, como, por ejemplo, caracterizar a los profesores y cuerpos académicos a nivel campus o áreas de conocimientos, lo que permitiría la identificación de profesores y grupos con perfiles de producción afines para posibles colaboraciones.

Se pretende realizar un nuevo estudio que incorpore el impacto de la publicación de un cuerpo académico y de un investigador por cada cita recibida. Se valoraría cada cita de acuerdo con el prestigio de la revista o publicación de origen, incluyendo las autocitaciones. Esto permitiría la caracterización de la divulgación del conocimiento generado.

La propuesta del trabajo permite construir la base para nuevos estudios complementarios donde la bibliometría, en conjunto con las ciencias de la computación, permita evaluar el impacto de las publicaciones científicas en la sociedad en general, y no únicamente en la comunidad científica, ya que los recursos que se encuentran en múltiples repositorios son referenciados en otras plataformas con enfoque social y académico.

Además, para el caso de las universidades y centros de investigación que ofertan programas de posgrado, se espera caracterizar, por la línea de investigación, las temáticas más frecuentes y a través de indicadores estadísticos, identificar las de mayor relevancia; así facilitar diversas actividades relacionadas con la difusión y promoción de dichos programas.

Por otra parte, se pretende extender el panorama de una a varias instituciones, lo que permitirá la caracterización de las colaboraciones entre cuerpos académicos por zonas geográficas y temáticas.

Es importante destacar los valores de los indicadores propuestos y la importancia de su cálculo por medio de un proceso automático capaz de analizar el estado actual de la investigación de una institución abarcando aspectos que difícilmente se pueden obtener en un proceso manual. En este sentido, se está trabajando en una *software* que implemente la metodología y los algoritmos presentados. La aplicación empleará una arquitectura basada en servicios web y protocolos para la interoperabilidad, lo que facilitaría su extensibilidad para que pueda ser adaptada con facilidad de acuerdo con nuevas necesidades.

Referencias

- Camacho, P. E., Zapata, A., Menéndez, V. H. y Canto, P. J. (2018). Análisis del desempeño del profesorado universitario en el uso de MOODLE a través de técnicas de minería de datos: propuestas de necesidades formativas. *RED. Revista de Educación a Distancia*, (58). Recuperado de <https://doi.org/10.6018/red/58/10>.
- Cechinel, C., Sánchez, S. y Sicilia, M. Á. (2009). Empirical Analysis of Errors on Human-Generated Learning Objects Metadata. Paper presented at the Research Conference on Metadata and Semantic Research. Milan, October 1-2, 2009. Retrieved from https://doi.org/10.1007/978-3-642-04590-5_6.
- Chuttur, M. Y. (2014). Investigating the effect of definitions and best practice guidelines on errors in Dublin Core metadata records. *Journal of Information Science*, 40(1), 28-37. Retrieved from <https://journals.sagepub.com/doi/10.1177/0165551513507405>.
- Consejo Nacional de Ciencia y Tecnología [Conacyt]. (27 de enero de 2017). Acuerdo por el que se emite el nuevo reglamento del Sistema Nacional de Investigadores. *Diario Oficial de la Federación*. Recuperado de http://www.dof.gob.mx/nota_detalle.php?codigo=5470107&fecha=27/01/2017.
- Consejo Nacional de Ciencia y Tecnología [Conacyt]. (2017). Lineamientos jurídicos de ciencia abierta. Recuperado de <https://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>.
- Consejo Nacional de Ciencia y Tecnología [Conacyt]. (2019). Sistema Nacional de Investigadores. Recuperado de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores>.
- Dirección General de Educación Superior Universitaria e Intercultural [DGESUI]. (2014). Programa para el Desarrollo Profesional Docente, para el Tipo Superior (Prodep). Recuperado de <https://www.dgesui.ses.sep.gob.mx/PRODEP.htm>.
- Elsevier. (2019). Scopus. Expertly curated abstract & citation database Retrieved from <https://www.elsevier.com/solutions/scopus>.
- Elsevier. (2020). *Scopus. Content Coverage Guide*. Elsevier. Retrieved from <https://www.elsevier.com/solutions/scopus/how-scopus-works/content>.

- González, G. y Gómez, J. (2014). La colaboración científica: principales líneas de investigación y retos de futuro. *Revista Española de Documentación Científica*, 37(4). Recuperado de <https://doi.org/10.3989/redc.2014.4.1186>.
- Guarascio, M., Manco, G. y Ritacco, E. (2019). Knowledge Discovery in Databases. In Ranganathan, S., Gribskov, M., Nakai, K. and Schönbach, C. (eds.), *Encyclopedia of Bioinformatics and Computational Biology* (vol. 1) (pp. 336-341). United States: Elsevier. Retrieved from <https://doi.org/10.1016/B978-0-12-809633-8.20456-1>.
- Guerrero, J. D. T., Menéndez, V. H. y Castellanos, M. E. (2018). Indicadores de calidad en investigaciones científicas: antecedentes. *Abstraction and Application*, 19(9), 6-24. Recuperado de <https://intranet.matematicas.uady.mx/journal/descargar.php?id=134>.
- Guerrero, J. D. T., Menéndez, V. H. y Castellanos, M. E. (2021). An indexing system for the relevance of academic production and research from digital repositories and metadata. *The Electronic Library*, 39(1), 33-58. Retrieved from <https://doi.org/10.1108/EL-06-2020-0160>.
- Guerrero, J. D. T., Menéndez, V., Castellanos, M. E. y Curi, L. F. (2019). Use of Graph Theory for the Representation of Scientific Collaboration. Paper presented at the International Conference on Computational Collective Intelligence, Hendaye, September 4-6, 2019. Retrieved from https://doi.org/10.1007/978-3-030-28374-2_47.
- Guerrero, J. D. T., Menéndez, V. H., Castellanos, M. E. y Curi, L. F. (2020). Analysis of Internal and External Academic Collaboration in an Institution Through Graph Theory. *Vietnam Journal of Computer Science*, 7(4), 391-415. Retrieved from <https://doi.org/10.1142/S2196888820500220>.
- Guerrero, J. D. T., Menéndez, V. H., Castellanos, M. E. y Gómez, J. R. (2019). Use of an Ontological Model to Assess the Relevance of Scientific Production. *IEEE Latin America Transactions*, 17(9), 1424-1431. Retrieved from <https://doi.org/10.1109/TLA.2019.8931135>.
- Guerrero, J. D. T., Menéndez, V. H., Castellanos, M. E. y Guerra, C. A. (2020). Metodología para la generación de grafos de colaboración científica de una institución académica. *Abstraction and Application*, 29, 88-101. Recuperado de <https://intranet.matematicas.uady.mx/journal/descargar.php?id=208>.

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, 11(1), 10-18. Retrieved from <https://doi.org/10.1145/1656274.1656278>.
- Leahey, E. (2016). From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration. *Annual Review of Sociology*, 42(1), 81-100. Retrieved from <https://doi.org/10.1146/annurev-soc-081715-074219>.
- Luna, M., Luna, E. y Luna, S. (2018). La UADY en la literatura científica registrada en Web of Science y Scopus: 1900-2016. *Revista Educación y Ciencia*, 7(50), 17-29. Recuperado de <http://www.educacionyciencia.org/index.php/educacionyciencia/article/view/470>.
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. United States: O'Reilly Media.
- Menéndez, V. H., Guerrero, J. D. T., Castellanos, M. E. y Zurita, E. (2020). Análisis de la producción de cuerpos académicos basado en teoría de grafos. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 10(20). Recuperado de <https://doi.org/10.23913/ride.v10i20.603>.
- MongoDB. (2019). The Database for Modern Applications. Retrieved from <https://www.mongodb.com/>.
- Programa de Mejoramiento del Profesorado [Promep]. (2020). Cuerpos académicos. Conceptos básicos. Recuperado de <http://promep.sep.gob.mx/ca1/Conceptos2.html>.
- Romero, C. y Ventura, S. (eds.) (2006). *Data Mining in E-Learning*. Southampton, United Kingdom: WIT Press.
- Romero, C. y Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618. Retrieved from <https://doi.org/10.1109/TSMCC.2010.2053532>.
- Romero, C. y Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3). Retrieved from <https://doi.org/10.1002/widm.1355>.
- Texier, J., De Giusti, M. R., Oviedo, N., Villarreal, G. L. y Lira, A. J. (2012). El uso de repositorios y su importancia para la educación en Ingeniería. Ponencia presentada en el World Engineering Education Forum (WEEF). Buenos Aires, 2012. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/22943>.

Universidad Autónoma de Yucatán [UADY]. (2020). Nuestra universidad. Quiénes somos.
Recuperado de <https://www.uady.mx/nuestra-universidad/>.

Witten, I. H., Frank, E. and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Burlington, United States: Elsevier. Retrieved from <https://doi.org/10.1016/C2009-0-19715-5>.

Rol de Contribución	Autor (es)
Conceptualización	Víctor Hugo Menéndez Domínguez
Metodología	Jared David Tadeo Sosa
Software	José William Cervera Pérez
Validación	María Enriqueta Castellanos Bolaños
Análisis Formal	María Enriqueta Castellanos Bolaños
Investigación	Víctor Hugo Menéndez Domínguez
Recursos	Jared David Tadeo Sosa
Curación de datos	José William Cervera Pérez
Escritura - Preparación del borrador original	Jared David Tadeo Sosa
Escritura - Revisión y edición	María Enriqueta Castellanos Bolaños
Visualización	María Enriqueta Castellanos Bolaños
Supervisión	Víctor Hugo Menéndez Domínguez
Administración de Proyectos	Víctor Hugo Menéndez Domínguez
Adquisición de fondos	Víctor Hugo Menéndez Domínguez