



UNIVERSIDAD DE CUENCA

Facultad de Ingeniería

Carrera de Ingeniería de Sistemas

Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos.

Trabajo de titulación previo a la obtención del título de Ingeniera de Sistemas

Modalidad: Proyecto Técnico

Autores:

Tania Maricela Landívar Ordóñez

CI: 0105905913

Correo electrónico: tania72landi@gmail.com

Ronaldo Sebastián Rendón Loja

CI: 0106633449

Correo electrónico: ronnyrendon28@gmail.com

Directora:

Ing. Lorena Catalina Sigüenza Guzmán

CI: 0102659687

Cuenca - Ecuador

02-marzo-2022



Resumen

La gestión de colecciones en bibliotecas universitarias es una actividad fundamental, ya que la misma es la encargada de proporcionar un conjunto de fuentes bibliográficas que satisfagan las necesidades de información. Sin embargo, en algunos casos el acceso a determinadas fuentes puede tener un alto valor económico. Partiendo de esta premisa, surge la necesidad de contar con estrategias/métodos que ayuden a la toma de decisiones con respecto a la gestión de colecciones. En la Universidad de Cuenca, este proceso se maneja en base a ciertas directrices como bases de datos con mayor trascendencia mundial o estadísticas de uso generadas por parte de los proveedores de las bases de datos, las mismas que reflejan parcialmente la actividad científica y académica, así como las necesidades de información. Por este motivo, el propósito de este trabajo de titulación es el de implementar un prototipo de sistema web que permita generar una base de conocimiento en función del descubrimiento de patrones de publicación y citación que de soporte a este tipo de decisiones. Para este proceso, se tomó en consideración los artículos científicos de investigadores con filiación a la Universidad de Cuenca indexados en Scopus y Latindex en el periodo (2016 - 2020); se aplicó una metodología para el análisis y recolección de datos, también ciertos conceptos de análisis bibliométrico como la Ley de Bradford y técnicas de minería de datos para el procesamiento de los mismos. Adicionalmente, para el desarrollo del sistema informático se tomó como referencia la metodología de desarrollo tradicional de software. Entre los principales resultados se pudo obtener una lista de fuentes bibliográficas (revistas científicas) donde más se publican, citan, consultan y que mayor factor de impacto tienen. Estas fuentes se verifican si están disponibles en las bases de datos bibliográficas contratadas por la Universidad de Cuenca, información de gran relevancia que apoya al proceso de toma de decisiones.

Palabras Clave: Base de datos bibliográfica. Métodos e indicadores bibliométricos. Ley de Bradford. Bibliominería. Minería de datos. Metodología de desarrollo tradicional.



Abstract

Collections management in university libraries is a fundamental activity, since it is in charge of providing a set of bibliographic sources that satisfy the information needs. However, in some cases access to certain sources may have a high economic value. Based on this premise, there is a need for strategies/methods to help make decisions regarding collection management. At the University of Cuenca, this process is managed based on certain guidelines such as databases with greater global significance or usage statistics generated by the providers, which partially reflect the scientific and academic activity, as well as the information needs. For this reason, the purpose of this degree work is to implement a web system that allows the generation of a knowledge base based on the discovery of publication and citation patterns that support this type of decisions. For this process, scientific articles of researchers with affiliation to the University of Cuenca indexed in Scopus and Latindex in the period (2016 - 2020) are taken into consideration, a methodology for data analysis and collection was applied, as well as certain concepts of bibliometric analysis such as Bradford's law and data mining techniques for data processing. In addition, the traditional development methodology was used as a reference for the development of the system. Among the main results, it was possible to obtain a list of bibliographic sources (scientific journals) where most is published, cited, consulted and have the highest impact factor. These sources are verified if it is available in the bibliographic databases contracted by the University of Cuenca, information of great relevance that supports the decision making process.

Keywords: Bibliographic database. Bibliometric methods and indicators. Bradford's Law. Bibliomining. Data mining. Traditional development methodology.



Índice

Resumen	2
Abstract.....	3
Índice	4
Índice de figuras.....	8
Listado de abreviaturas	16
Agradecimiento	17
Agradecimiento	18
Dedicatoria.....	19
Dedicatoria.....	20
Capítulo 1: Introducción	21
1.1. Identificación del problema y justificación.....	21
1.2. Alcance	23
1.3. Objetivos.....	24
1.3.1. Objetivo General.....	24
1.3.2. Objetivos Específicos	24
1.4. Estructura de la tesis.....	25
Capítulo 2: Marco Teórico	26
2.1. Conceptos Generales	26
2.1.1. Bases de datos bibliográficas	26
2.1.2. Bibliometría	26
2.1.3. Métodos e indicadores bibliométricos	27
2.1.4. Minería de datos.....	27
2.1.5. Bibliominería.....	28
2.1.6. Ley De Bradford	28
2.1.7. Sistema Web	28
2.1.8. Arquitectura Cliente – Servidor	29
Capítulo 3: Estado del Arte	30
3.1. Metodología	30



3.2. Criterios de suscripción de bases bibliográficas	33
3.3. Métodos e indicadores bibliométricos utilizados en el análisis de la producción científica	34
3.4. Técnicas de minería de datos aplicada a patrones de publicación y citación.	38
3.5. Herramientas de recopilación y visualización de la producción científica.	39
3.6. Recapitulación	42
Capítulo 4: Metodología	43
4.1. Aplicación de la Ley de Bradford	43
4.2. Metodología de desarrollo tradicional	44
4.3. Análisis de requerimientos	46
4.3.1. Descripción de la situación actual	47
4.3.2. Identificación y evaluación de las oportunidades de mejora	48
4.3.3. Sistema de Información Propuesto	50
4.3.3.1. Descripción de los requerimientos	50
4.3.3.2. Restricciones	55
4.3.3.3. Actores relacionados	55
4.4. Diseño del Prototipo	56
4.4.1. Descripción de la arquitectura del sistema	56
4.4.1.1. Arquitectura Lógica	56
4.4.1.2. Arquitectura Física	60
4.4.2. Diseño del Modelo de Clases del Sistema	60
4.4.2.1. Diseño de la vista	63
4.4.2.2. Diseño del Controlador	63
4.4.2.3. Diseño de la Persistencia	64
4.4.2.4. Diseño de la Base de Datos	65
4.4.3. Definición de la Interfaz del Usuario	68
4.4.3.1. Especificación de la Interfaz de Pantalla y Navegación	68
4.5. Desarrollo e implementación del prototipo	72
4.5.1. Prototipo implementado	72
4.5.1.1. Sección Publicaciones	73



4.5.1.2.	Sección Autores	75
4.5.1.3.	Sección Referencias	78
4.5.1.4.	Sección Estadísticas Proveedores.....	81
4.5.1.5.	Sección Análisis Datos	84
4.5.1.6.	Sección Ley de Bradford	86
4.6.	Pruebas de funcionamiento y Despliegue	90
Capítulo 5:	Resultados y Discusión	91
Capítulo 6:	Conclusiones y Recomendaciones	105
6.1	Conclusiones finales.....	105
6.2	Líneas de trabajos futuros	108
Bibliografía	109
Anexos.....	117
Anexo 1: Diseño Del Prototipo.....	117
Dashboard	117
Pantalla para el ingreso de autores y su relación con sus publicaciones.	117
Pantalla para el ingreso de los medios de publicación.	118
Pantalla para el ingreso de las áreas Frascati y Unesco.	119
Pantalla para el ingreso de las bases de datos bibliográficas que provee la Universidad de Cuenca para su uso.	120
Anexo 2: Manual de Usuario	122
Índice	122
Índice de figuras.....	123
Presentación	125
Desarrollo del Manual de Usuario	125
Dashboard.....	125
Publicaciones	126
Autores.....	128
Referencias	130
Medios de publicación	132
Áreas.....	135
Bases de datos bibliográficas	136
Estadísticas de los proveedores	137



Configuración	139
Ley de Bradford	140
Análisis de datos	141
Anexo 3: Manual Técnico	142
Índice	142
Índice de figuras.....	143
Presentación	145
Objetivo.....	145
Requisitos del sistema	145
Herramientas utilizadas para el desarrollo	145
Instalación de aplicaciones	147
Modelo de clases	148
Diagrama Entidad – Relación.....	150
Diccionario de datos	152
Servicios web implementados en el desarrollo del sistema	165
Anexo 4: Artículo Presentado en el ICAT 2021	180



Índice de figuras

Figura 1. Cantidad de información recuperada sobre criterios de adquisición y suscripción por país.	33
Figura 2. Número de publicaciones sobre métodos e indicadores bibliométricos por país (Scopus y Scielo).....	35
Figura 3. Áreas de investigación de Scopus y Scielo sobre métodos e indicadores bibliométricos.	36
Figura 4. Cantidad de información en Scopus y Scielo sobre herramientas de recopilación y visualización.	40
Figura 5. Etapas establecidas en la Metodología Tradicional para el desarrollo el Prototipo	45
Figura 6. Modelo para la Evaluación de la Importancia de la Colección	46
Figura 7. Modelo conceptual actualizado de la metodología.....	48
Figura 8. Modelo conceptual actual con visión futura.	49
Figura 9. Arquitectura Lógica del Sistema Web.	56
Figura 10. Arquitectura Física Cliente - Servidor.	60
Figura 11. Entidades y operaciones que interactúa en el prototipo de sistema.	62
Figura 12. Modelo Conceptual de la Base de Datos del Prototipo.	66
Figura 13. Diagrama entidad – relación (Modelo lógico del Prototipo).	67
Figura 14. Opciones del menú de navegación de la Pantalla Principal.	68
Figura 15. Pantalla del ingreso del detalle de las publicaciones y sus respectivas referencias.	69
Figura 16. Pantalla para la extracción del detalle de las referencias por cada publicación.	70
Figura 17. Pantalla para el ingreso de estadísticas de uso proporcionadas por los proveedores.	71
Figura 18. Pantalla para el análisis de datos a través de un análisis estadístico y técnicas de minería de datos.	72
Figura 19. Pantalla para el mantenimiento y visualización de las publicaciones.	73
Figura 20. Visualización de las referencias en texto por cada publicación.	74
.....	74



Figura 21. Visualización del detalle de las referencias separada en campos por cada publicación.	74
Figura 22. Pantalla para el mantenimiento de los autores y sus publicaciones. 76	
Figura 23. Publicaciones por autor.	76
Figura 24. Gráfico de burbuja en función del número de publicaciones por autor. 77	
Figura 25. Gráfico de burbuja de acuerdo con el número de publicaciones según el orden del autor en la publicación.	77
Figura 26. Pantalla para la extracción de las referencias y el detalle de las referencias por publicación.	79
Figura 27. Pantalla para el ingreso de forma manual y automática de las referencias y el detalle de mismas.	79
Figura 28. Pantalla para la extracción del detalle de la referencia a través de la API de Google Académico.	80
Figura 29. Pantalla para el ingreso de las estadísticas de uso de las bases de datos bibliográficas y sus revistas o fuentes bibliográficas.	82
Figura 30. Gráfico de barras de la estadística de búsqueda de las bases de datos bibliográficas.	82
Figura 31. Gráfico de barras de la estadística de búsqueda de las fuentes bibliográficas indexadas en las bases de datos.	83
Figura 32. Sección análisis de datos	85
Figura 33. Sección para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto. 86	
Figura 34. Tabla de resultados para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto. 87	
Figura 35. Número de publicaciones según el orden de autor	93
Figura 36. Número de publicaciones según el primer autor	93
Figura 37. Número de publicaciones según el último autor.	94
Figura 38. Autores con el mayor número de publicaciones.	94
Figura 39. Número de publicaciones por Área Unesco.	95
Figura 40. Número de publicaciones por Área Frascati.	95
Figura 41. Número de Publicaciones por Medios de Publicación.	96



Figura 42. Número de publicaciones por Cuartil	97
Figura 43. Predicción del número de búsqueda de Ebsco con Science Direct en los próximos seis meses.	98
Figura 44. Ley de Bradford: Evaluación de los medios de publicación en función del número de publicaciones, número de citas, número de búsquedas y el factor de impacto.....	99
Figura 45. Medios de publicación más citados en el área de Ingenierías Eléctrica, Electrónica e Información.	100
Figura 46. Medios de publicación con más publicaciones en el área de Ingenierías Eléctrica, Electrónica e Información.....	101
Figura 47. Medios de Publicación con la mayor calificación, resultado de la aplicación de Ley de Bradford en el área de Ingenierías Eléctrica, Electrónica e Información	102
Figura 48. Número de medios de publicación núcleo proporcionados por las bases de datos bibliográficas.	103



Índice de tablas

Tabla 1. Cadenas utilizadas y resultados obtenidos durante la búsqueda en las bases de datos bibliográficas.....	31
Tabla 2. Métodos Bibliométricos más utilizados para el análisis de la producción científica	37
Tabla 3. Indicadores Bibliométricos más utilizados en el análisis de la producción científica (Scielo y Scopus).....	38
Tabla 4. Herramienta de Recopilación y Visualización Scopus.....	41
Tabla 5. Herramienta de Recopilación y Visualización Scielo.....	42
Tabla 6. Requisitos funcionales del Prototipo.	55
Tabla 7. Servicios Web para la Pantalla Publicaciones.....	75
Tabla 8. Servicios Web para el mantenimiento de los autores.....	78
Tabla 9. Servicios Web para la extracción y almacenamiento de las referencias y detalle de las referencias.	81
Tabla 10. Servicios web para el ingreso y mantenimiento de las estadísticas de uso de las bases de datos bibliográficas y sus fuentes.	84
Tabla 11. Servicios web para la sección análisis de datos	86
Tabla 12. Servicios web para la evaluación de las revistas a través de la ley de Bradford.....	89
Tabla 13. Personal del CDRJBV de la Universidad de Cuenca que revisaron el prototipo de sistema	90



Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Yo, Tania Maricela Landívar Ordóñez en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulación “Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos”, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 02 de marzo de 2022


Tania Maricela Landívar Ordóñez

C.I: 0105905913



Cláusula de licencia y autorización para publicación en el Repositorio
Institucional

Yo, Ronaldo Sebastián Rendón Loja en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulación "Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 02 de marzo de 2022

Ronaldo Sebastián Rendón Loja

C.I: 0106633449



Cláusula de Propiedad Intelectual

Yo, Tania Maricela Landívar Ordóñez, autor/a del trabajo de titulación "Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 02 de marzo de 2022

Tania Maricela Landívar Ordóñez

C.I: 0105905913



Cláusula de Propiedad Intelectual

Yo, Ronaldo Sebastián Rendón Loja, autor/a del trabajo de titulación "Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 02 de marzo de 2022

Ronaldo Sebastián Rendón Loja

C.I: 0106633449



Listado de abreviaturas

- **Back-end:** parte de la aplicación de software en la cual se desarrolla toda la parte lógica como la persistencia con la base de datos, consultas, generación de servicios web, entre otros.
- **MVC (Modelo, Vista, Controlador):** Arquitectura de diseño de software.
- **Front-end:** es la parte visible de la aplicación de software que interactúa con los usuarios.
- **JavaScript:** Lenguaje interpretado que se ejecuta en el lado del cliente, en el navegador.
- **HTTP:** Protocolo utilizado en cada transacción de la World Wide Web.
- **MySQL:** sistema de gestión de bases de datos relacional, multihilo y multiusuario.
- **Web Services:** servicios web, es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones.
- **RESTful:** Los servicios web RESTful permite realizar una serie de operaciones como: GET, POST, UPDATE, PUT, DELETE y PATCH a través de una URL y un método HTTP.
- **GET, POST:** Método de request soportado por el protocolo HTTP.
- **API:** La interfaz de programación de aplicaciones, conocida también por la sigla API, en inglés, application programming interface, es un conjunto de subrutinas, funciones y procedimientos que ofrece cierta librería para ser utilizada por otro software como una capa de abstracción.



Agradecimiento

Primeramente, a Dios, ya que tengo la bendición de tener unos maravillosos padres Patricio y Yolanda, sin duda alguna este camino ha sido duro, pero ellos con su apoyo y confianza han hecho que pueda cumplir esta meta.

A mi hermana y mi sobrina, que se convirtieron en una motivación para llegar a culminar este objetivo, espero ser un ejemplo que puedan seguir.

A todos los docentes, quienes compartieron su conocimiento y me motivaron a ser una buena profesional, de manera especial a la tutora de tesis, Ing. Lorena Sigüenza Guzmán, por guiarnos en este trabajo de titulación, haciendo que reflejemos los conocimientos y habilidades obtenidas durante nuestra formación.

A la Lic. Rocío Campoverde Carpio y al Ing. Juan José Sáenz que fueron quienes nos ayudaron en el desarrollo de este trabajo de titulación, gracias por su tiempo y colaboración.

A todos mis amigos y compañeros de carrera, fue increíble haber compartido esta etapa con ustedes, en especial a mi compañero de tesis Ronaldo Rendón. “Lo logramos amigo”.

Finalmente quiero agradecer a mi enamorado Juan Carlos Morocho, que en el transcurso de esta etapa estuvo en mis mejores momentos, pero más en aquellos días cuando necesitaba ánimos.

Tania Landívar



Agradecimiento

A Dios, por todas las bendiciones recibidas y por guiarme siempre durante mi existencia. A mis padres Kleber y Diana por su apoyo total, por su amor incondicional, por siempre confiar y creer en mí, gracias por sus consejos, valores y principios que me han sabido inculcar a lo largo de mi vida.

A mi hermano, Oscar por su cariño y apoyo incondicional durante este largo trayecto, por creer en mí y permitirle ayudarlo con sus estudios, ya que esto fue una motivación para mí, para estar siempre en constante aprendizaje.

A todos los docentes, quienes me brindaron su conocimiento y su tiempo, lo que me ha permitido crecer como un buen profesional, de manera especial a mi tutora de tesis, Ing. Lorena Sigüenza Guzmán, por su tiempo, todo su apoyo, y por hacer todo lo posible para que culminemos con éxito este trabajo de titulación.

A la Lic. Rocío Campoverde Carpio y al Ing. Juan José Sáenz que fueron quienes nos ayudaron en el desarrollo de este trabajo de titulación, gracias por su tiempo y colaboración.

A mi compañera de tesis Tania Landivar, por este largo trayecto de aprendizaje, por los buenos y malos momentos que pudimos superarlos para llegar a cumplir nuestro objetivo, la obtención de la tesis. “Si se pudo”.

Finalmente agradezco a mi enamorada Gabriela Verdugo, por apoyarme siempre a cumplir todas mis metas, agradezco sus consejos y palabras de aliento que hicieron de mí una mejor persona, por brindarme su amor y apoyo incondicional que me motivaron a alcanzar esta meta y me ayudaron a sobrellevar los momentos difíciles.

Ronaldó Rendón



Dedicatoria

A mis padres Patricio y Yolanda, quienes tuvieron el sueño de que su hija se convierta en una profesional. Esto es para ustedes y por ustedes.!!

Tania Landívar



Dedicatoria

A mis padres Kleber y Diana, por su amor, esfuerzo, y sacrificio en todos estos años que me han permitido llegar a cumplir un sueño más.

A mi abuelito Julio que siempre estuvo pendiente de este objetivo y me motivaba para que nunca me rinda. “Oso lo logré”

Ronaldo Rendón



Capítulo 1: Introducción

El actual proyecto técnico resalta el uso de indicadores bibliométricos y técnicas de minería de datos como recurso para la toma de decisiones respecto a la adquisición o contratación de base de datos bibliográficos. En este capítulo se presentan los puntos relevantes de la problemática actual acerca de la gestión de colecciones en las bibliotecas universitarias, el porqué de mejorar los procesos para dicha gestión, el estado del arte en la cual se realiza una compilación de otras investigaciones acerca del descubrimiento de patrones de publicación y citación. Finalmente, se detallan los objetivos y la estructura de desarrollo de este trabajo de titulación.

1.1. Identificación del problema y justificación

La bibliometría se puede definir como la aplicación de métodos estadísticos y matemáticos para analizar el uso de los recursos de información de una biblioteca. A nivel mundial, esta metodología ha sido utilizada durante décadas para medir la actividad científica y su repercusión (Glänzel, 2003; González Alcaide & Gorraiz, 2018; Abramo, 2017; Beaudry & La rivièrre, 2016). En Latinoamérica, existen también importantes estudios. Por ejemplo, Guerrero-Sosas et al. (2019) presentaron en México, un sistema que permite medir la productividad y el impacto de las publicaciones mediante el uso de indicadores bibliométricos, metadatos y similitud textual semántica. En Ecuador, Zhimnay Valverde et al. (2019) analizaron la importancia de la medición de la ciencia para evaluar la excelencia académica de universidades. Por otro lado, Gureev y Mazov (2015) en Rusia, demuestran que la bibliometría puede ser también utilizada para mejorar la gestión de colecciones (i.e., construcción de un acervo bibliográfico en concordancia con las necesidades del usuario). Sin embargo, la bibliometría también ha sido criticada por no proporcionar un análisis completo del uso o importancia de la colección (Duy & Vaughan, 2006). Adicionalmente, las bases de datos de citas más importantes son universales y no acopladas a las necesidades locales (Duy & Vaughan, 2006). Para superar estas limitaciones, varios estudios coinciden que la combinación de diferentes métodos permite obtener una indicación más sólida del uso de la colección y de las necesidades de los usuarios (Beile et al., 2004; Duy & Vaughan, 2006; Enger, 2009) Una propuesta es la de Siguenza-Guzman et al. (2013), quienes evalúan la importancia de la colección combinando cuatro tipos de análisis de patrones: 1) de publicación, las revistas donde publican los docentes/investigadores de la institución y sus colaboraciones; 2) de citación, las referencias citadas en dichas publicaciones; 3) de descarga, estadísticas de revistas consultadas y descargadas suministradas por los proveedores; y 4), de factor de impacto, la importancia de una revista mediante el establecimiento de rankings por especialidad.



La gestión de las colecciones es una tarea constante. Una parte se enfoca a que las universidades canalicen sus recursos financieros para obtener acceso a bases de datos de publicaciones científicas. No obstante, debido a su alto costo y al presupuesto siempre limitado (Bordons & Zulueta, 1999), este acceso aún es reducido. La Universidad de Cuenca, por ejemplo, ha incrementado últimamente su presupuesto para suscripciones; sin embargo, no ha sido suficiente para cubrir los mínimos niveles de acceso. Por este motivo, las bibliotecas ecuatorianas han recurrido al financiamiento compartido a través de consorcios específicos (Chowdhury, 2014), como la red [CEDIA¹ \(Red Nacional de Investigación y Educación Ecuatoriana\)](#). Esta red vio conveniente la conformación de un consorcio que trabaje conjuntamente para gestionar la suscripción a bases de datos bibliográficas. Contardi (2004) destaca que los principales beneficios de un consorcio son la simplificación de los procesos de suscripción, el aumento del acceso a recursos actualizados y la disminución de costos al momento de su adquisición. Empero, Thornton (2000) indica que se requiere tener especial cuidado con la gestión de la colección ya que, se corre el riesgo de obtener material innecesario y así gastar injustificadamente el presupuesto. Por tal motivo, es vital contar con información fiable que permita tomar decisiones sobre las suscripciones a revistas científicas, a través de un modelo ajustado a las necesidades actuales (locales e internacionales) y futuras.

Desafortunadamente, en el Ecuador, este tipo de información no se encuentra disponible ni a nivel de universidades ni de consorcios. Las decisiones de selección han sido manejadas “ad hoc”, sin mayor sustento cuantitativo y escasa evaluación del impacto obtenido. En el mejor de los casos, la selección se realiza en función de las estadísticas de consulta generadas por parte de los proveedores en un periodo académico y directrices establecidas por la institución, como bases de datos con mayor trascendencia a nivel mundial, o multidisciplinarias que abarquen todas las áreas del conocimiento. Las mismas que reflejan parcialmente la actividad científica y académica, así como las necesidades de información.

Por lo tanto, tomando en consideración el modelo propuesto por Siguenza-Guzman et al. (2013) para la evaluar la importancia de la colección, se propone formar una base de conocimiento en función de los cuatro ejes principales de estudio de patrones mencionados anteriormente: publicación, citación, descarga y factor de impacto; con el objetivo de proporcionar otra fuente de información fiable que de soporte a la toma de decisiones para la construcción de la colección.

¹ <https://www.cedia.edu.ec/>



1.2. Alcance

Este proyecto de titulación de modalidad proyecto técnico propone desarrollar un prototipo de Sistema Informático Web para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca, a través de indicadores bibliométricos y técnicas de minería de datos. El universo de datos está constituido por artículos científicos de investigadores con filiación a la Universidad de Cuenca, indexados en Scopus y Latindex en un periodo de cinco años (2016-2020). En función de los artículos publicados y citados se busca obtener un listado de fuentes bibliográficas (revistas científicas) en las que más se publican, citan, consultan y que mayor factor de impacto tiene. Analizando de esta manera la productividad y consumo de las fuentes bibliográficas y si las mismas están siendo proporcionadas por la Universidad de Cuenca. Además, se busca obtener autores, medios de publicación, áreas de investigación más prolíficos de las publicaciones pertenecientes a los autores con filiación a la Universidad de Cuenca. Resultados que apoyen el proceso de toma de decisiones para adquirir o suscribirse a una fuente o base de datos bibliográfica.

Por lo anterior, este trabajo de titulación busca responder a las siguientes preguntas de investigación:

P.I.1: ¿Qué indicadores y métodos bibliométricos son los más sobresalientes para descubrir patrones de publicación y citación de las publicaciones (artículos científicos) de los autores con filiación a la Universidad de Cuenca?

P.I.2: ¿Cuáles son las fuentes bibliográficas (revistas) más atractivas donde publican los autores con filiación a la Universidad de Cuenca?

P.I.3: ¿Qué fuentes bibliográficas son las más consumidas por los autores con filiación a la Universidad de Cuenca para la elaboración y publicación de sus artículos?

P.I.4: ¿Pueden los patrones de publicación y citación apoyar al proceso de toma de decisiones para la adquisición o suscripción de base de datos bibliográficas?



1.3. Objetivos

1.3.1. Objetivo General

Desarrollar un prototipo de sistema informático para el descubrimiento de patrones de publicación y citación en la producción científica de la Universidad de Cuenca, a través de indicadores bibliométricos y técnicas de minería de datos, con el fin de proporcionar información que ayude al soporte de toma de decisiones y oriente la adquisición y suscripción de bases bibliográficas conforme a las necesidades de los usuarios.

1.3.2. Objetivos Específicos

1. Analizar los requerimientos mediante una estructura de recolección de datos a través de la fusión de análisis de citas y publicación, estadísticas de uso suministradas por los proveedores y los valores del factor de impacto de las fuentes bibliográficas (revistas), con el fin de obtener información que oriente a la adquisición y suscripción de bases de datos bibliográficas.
2. Diseñar el prototipo describiendo la funcionalidad de las interfaces, la arquitectura cliente-servidor del sistema y el diseño de la base de datos definida por un modelo conceptual y un modelo lógico.
3. Desarrollar e implementar el prototipo con la herramienta de desarrollo web React y el lenguaje de programación Python, en tres niveles: cliente, servidor y base de datos para la recolección e integración de los datos de las publicaciones, ejecución de técnicas de minería de datos y la visualización de los resultados.
4. Realizar pruebas de funcionamiento y comparar las estadísticas que utiliza la Universidad de Cuenca para la adquisición y/o suscripción de bases de datos bibliográficas con los resultados obtenidos en este trabajo de titulación.



1.4. Estructura de la tesis

A continuación, se describen los seis capítulos que componen el presente trabajo de titulación:

- **Capítulo 1: Introducción**

Describe la identificación y justificación del problema, el alcance, los objetivos tanto el general como específicos.

- **Capítulo 2: Marco teórico**

Se conceptualizan los temas relevantes de la investigación como, bases de datos bibliográficas, bibliometría, métodos e indicadores bibliométricos, minería de datos, bibliominería, ley de Bradford, sistema web y arquitectura cliente - servidor.

- **Capítulo 3: Estado del Arte**

Presenta el estado actual del tema general de investigación. Se lo realiza por medio de una revisión sistemática de literatura de los diferentes temas en los que se centra el trabajo. Esto permite tener una vista objetiva de las soluciones existentes al problema planteado y determina el estado de la investigación.

- **Capítulo 4: Metodología**

Detalla en qué consiste la metodología de desarrollo de software tradicional y las cinco etapas: análisis de requerimientos, diseño del prototipo, desarrollo e implementación, pruebas y despliegue. También, se menciona el modelo propuesto por Siguenza-Guzman et al. (2013) que es el punto de partida para el análisis del presente trabajo y la aplicación de la ley de Bradford en el prototipo.

- **Capítulo 5: Resultados y Discusión**

Presenta los resultados obtenidos del estudio como: patrones de publicación, patrones de citación y fuentes bibliográficas más consumidas para la producción científica universitaria, análisis e interpretación de los resultados obtenidos.

- **Capítulo 6: Conclusiones y Recomendaciones**

Se presentan las conclusiones obtenidas de los resultados del Capítulo 5 y algunas recomendaciones para posibles trabajos futuros que puedan surgir del presente estudio.



Capítulo 2: Marco Teórico

En este capítulo se dan a conocer los conceptos principales de los términos, métodos y herramientas utilizados tanto para el análisis de la producción científica como para el descubrimiento de patrones de publicación y citación.

2.1. Conceptos Generales

Los principales términos utilizados en este trabajo de titulación son: base de datos bibliográficas, bibliometría, métodos e indicadores bibliométricos, minería de datos, bibliominería, ley de Bradford, sistema web y la arquitectura cliente-servidor.

2.1.1. Bases de datos bibliográficas

Las bases de datos o repositorios bibliográficos representan una fuente de información digital especializada y organizada. Son sistemas informáticos dedicados a gestionar trabajos científicos y académicos de diferentes instituciones (Mariscal Orozco & Girarte, 2017). Existen base de datos especializadas de acceso abierto, es decir, se tiene acceso inmediato a todo tipo de resultados de investigación, pero una de las desventajas es que falta aún en ellos una evaluación tanto cualitativa como cuantitativa que refleje completamente la calidad de un trabajo (Perakakis et al., 2016). Por otro lado, las bases de datos privadas tienen acceso restringido, ya sea por el costo, infraestructura y servicios de bibliotecas universitarias. Estas bases proporcionan trabajos de calidad debido a que se realiza una evaluación completa a los mismos para que puedan ser colocadas en el repositorio (Mariscal Orozco & Girarte, 2017).

2.1.2. Bibliometría

La bibliometría se convirtió en un campo de investigación interdisciplinaria que puede extenderse a casi todos los campos científicos (Glänzel, 2003). Es un término utilizado para indicar la calidad, consumo y producción de un artículo a través de métodos estadísticos y matemáticos (Ma, 2014). El análisis bibliométrico es un método documental cuyos objetivos son tanto el análisis del tamaño, crecimiento y distribución de la producción científica, como el análisis del comportamiento de los productores y consumidores de estos documentos, a través de indicadores bibliométricos (Dios et al., 1997).



2.1.3. Métodos e indicadores bibliométricos

Los métodos bibliométricos permiten realizar un diagnóstico objetivo sobre un conjunto de publicaciones científicas, a través de indicadores, técnicas y modelos de análisis basados tanto en las características del conjunto de estudio, patrones de producción y consumo de información (Boeris, 2010). Según Glänzel y Moed (2013), un indicador bibliométrico caracteriza y evalúa unidades de análisis mediante métodos cuantitativos. Dios et al. (1997) mencionan que estos indicadores pueden ser divididos dentro de tres aspectos: la calidad, importancia e impacto científico.

- Los *indicadores de calidad* científica son de tipo subjetivo; es decir, se miden a través de la percepción de un grupo de personas expertas en el área.
- Los *indicadores de importancia* científica se obtienen tomando en cuenta criterios como la cantidad de publicaciones, su distribución, productividad y colaboración de los autores.
- Los *indicadores de impacto* analizan el consumo de material científico en una región de estudio. Se obtienen a partir del análisis de citas y miden la cantidad de citas recibidas de un artículo por publicaciones posteriores.

2.1.4. Minería de datos

La minería de datos es el estudio de recopilar, limpiar, procesar, analizar y obtener información útil de los datos (Aggarwal, 2015). Implica el descubrimiento de relaciones significativas, patrones y tendencias que son observadas al examinar grandes cantidades de datos (Hernández Orallo et al., 2004). Según Han et al. (2011), la minería de datos surge gracias a la evolución de las tecnologías de la información pues, debido al volumen de datos, se vio necesario el análisis de los mismos buscando información útil para la toma de decisiones (Linoff & Berry, 2011).

Se podría decir que la minería de datos es un proceso que integra datos de diferentes fuentes, para posteriormente extraer un importante conocimiento, es decir, identificar información trascendente, valiosa y útil, de la cual las instituciones van a poder tomar alguna significativa decisión (Daza, 2016).

La minería de datos se convierte en una tarea esencial en una gran cantidad de dominios como la banca, comercio minorista, bioinformática, etc., que permite extraer patrones útiles desconocidos (Gupta & Chandra, 2020). Este proceso emplea una combinación de métodos para manejar varios tipos de datos, tareas de minería de datos y áreas de aplicación (Fu, 1997). La minería de datos ayuda a las empresas en la toma de decisiones mediante el análisis de



patrones y tendencias ocultas, para obtener conocimiento o aumentar su rentabilidad mediante la mejora de procesos y operaciones (Gupta & Chandra, 2020).

2.1.5. Bibliominería

Bibliominería (en inglés, Bibliomining) fue introducido por Nicholson y Stanton (2003) para hacer referencia a la combinación de minería de datos, bibliometría y data warehousing que es un almacén de datos que comprende un conjunto de conceptos y herramientas que apoyan con material de información para la toma de decisiones (Gatzui, 1999). Esta disciplina tiene el propósito de analizar los servicios de una biblioteca. Bibliomining tiene una estrecha relación con la bibliometría y la documentación (Mancini, 1996), en el sentido de que aporta a la identificación de patrones que realmente representan el conocimiento basado en mediciones y el análisis de grandes cantidades de datos por medio de métodos estadísticos de especialidades métricas, ya conocidas como la bibliometría y la infometría (Aenta, 2011).

Debido al auge tecnológico y el acelerado crecimiento de la información, la gestión de grandes volúmenes de datos y su forma de utilización para que aporten conocimiento a la toma de decisiones es uno de los mayores problemas para las instituciones ya sean con o sin fines de lucro (Peñalver, 2014). Obtener dicho conocimiento es difícil, dada la heterogeneidad de la fuente de datos. Sin embargo, mediante procesos de descubrimiento de conocimiento esta tarea es cada vez más factible (Haro Valle et al., 2014).

2.1.6. Ley De Bradford

La ley de Bradford fue propuesta en 1934, se conoce también como la ley de dispersión que establece: “Si las revistas científicas se ordenan en secuencia decreciente de productividad de artículos sobre un tema dado, éstas pueden dividirse en un núcleo de revistas dedicadas más en particular al tema y varios grupos o zonas, conteniendo el mismo número de artículos que el núcleo, donde el número de revistas en el núcleo y las zonas sucesivas estará en la relación de $1 : n : n^2$.” (Spinak, 1996). Permitiendo identificar revistas científicas no sólo las más productivas sino también más relevantes para cubrir una determinada área del conocimiento (Urbizagástegui Alvarado, 2016). Se trata de demostrar que la distribución de los artículos es altamente desigual, donde gran parte de los artículos están concentrados en una pequeña población de revistas, mientras que una mínima cantidad de artículos se distribuye sobre un alto número de revistas (Urbizagástegui, 1996).

2.1.7. Sistema Web

Sistema web son aquellas aplicaciones donde los usuarios pueden acceder mediante un servidor web a través del internet o intranet. También se lo conoce



como aplicación (Software) que no necesariamente se encuentra instalado en una computadora si no que está en otro equipo que funciona como servidor en la red. Los sistemas web son atractivos debido a lo práctico del navegador web como cliente ligero, a la independencia del sistema operativo, así como a la facilidad para actualizar y mantener aplicaciones web sin distribuir e instalar software a miles de usuarios potenciales (Guerrero, 2016).

2.1.8. Arquitectura Cliente – Servidor

Es aquel en donde las tareas se reparten entre proveedores de recursos llamados servidores y los consumidores llamados clientes, los mismos que acceden a los servicios a través de peticiones (Marini, 2012).

- **Cliente:** Conocido como Front-end, permite al usuario formular los requerimientos y pasarlos al servidor para ejecutar determinadas funciones, por ejemplo, procesar la lógica de la aplicación, generar requerimientos de la base de datos, recibir resultados del servidor, etc. (Marquez, 1980).
- **Servidor:** A este proceso se le conoce con el término de Back-end, que es el encargado de atender a las peticiones de los clientes. Entre las funciones que lleva a cabo son: aceptar y procesar requerimientos de base de datos por parte de los clientes, procesar la lógica de la aplicación, entre otros (Marquez, 1980).



Capítulo 3: Estado del Arte

El análisis de la producción científica ha tomado fuerza en los últimos años. Existen varios métodos, procesos y herramientas que ayudan a la generación y evaluación de dicha producción científica; por tal motivo, fue necesario realizar una revisión del estado actual del tema general de investigación, el mismo que se detalla en este capítulo.

3.1. Metodología

En esta revisión de literatura se aplicó la metodología propuesta por Fink (2014). El procedimiento de Fink permite una revisión sistemática, estructurada y enfocada con fines de investigación. Esta metodología se conforma por siete pasos: (1) selección de las preguntas de investigación, (2) selección de las bases de datos bibliográficas, (3) selección de los términos de búsqueda, (4) aplicación de criterios prácticos de detección, (5) aplicación de criterios metodológicos de detección, (6) desarrollo de la revisión y (7) resumen de los resultados.

El paso inicial es la selección de preguntas de investigación empleando la estructura PICO para diseñar las preguntas y desarrollar estrategias de búsqueda que permitan delimitar el objetivo del estudio. (Kitchenham & Charters, 2007; Shokraneh, 2016). El nombre PICO proviene de *Población*, donde se identifica el problema, que para el presente estudio analiza si parte del proceso de decisión sobre la adquisición de bases de datos bibliográficas en bibliotecas universitarias, se basa en un análisis metodológico de la producción generada por la institución a través del descubrimiento de patrones de publicación y citación. *Intervención*, donde se establecen los criterios de intervención como métodos/estrategias de adquisición y suscripción, uso de métodos e indicadores bibliométricos, técnicas de minería de datos y herramientas para la recolección y visualización de la producción científica. *Comparación*, entre los términos mencionados anteriormente. Y, *Resultado* (del inglés, Outcome), del análisis donde se esperan obtener los criterios más relevantes en cuanto a la adquisición de bases bibliográficas, métodos e indicadores bibliométricos, técnicas de minería de datos, herramientas más utilizadas para el análisis de la producción científica con el descubrimiento de patrones de publicación y citación; y si estos patrones se han utilizado como fuente para el proceso de toma de decisiones en relación a la adquisición y suscripción de bases de datos bibliográficas. En consecuencia, se han considerado importantes las siguientes preguntas: (a) ¿Cuáles son los criterios, métodos o estrategias para la adquisición de bases de datos bibliográficas en las bibliotecas universitarias? (b) ¿Qué métodos bibliométricos son utilizados



para el descubrimiento de patrones de publicación y citación? (c) ¿Cuáles son los indicadores bibliométricos más utilizados para el análisis de patrones de publicación y citación? (d) ¿Cuáles son las técnicas de minería de datos más utilizadas para el análisis de patrones de publicación y citación? (e) ¿Qué herramientas de recopilación y visualización de la producción científica son utilizadas para un análisis de patrones de publicación y citación?

El segundo paso corresponde a la selección de bases de datos bibliográficas a consultar: Scopus de gran relevancia internacional (Zhu & Liu, 2020) y Scielo que se destaca en la región hispana (Canales et al., 2020), permitiendo realizar un estudio completo a nivel geográfico. Posteriormente, como tercer paso, fue necesario determinar las palabras clave, las cuales, para esta investigación, se derivaron principalmente de los siguientes términos básicos: “Adquisición y Suscripción”, “Métodos e Indicadores Bibliométricos”, “Técnicas de Minería de Datos” y “Herramientas de Recolección y Visualización”. El detalle de las bases de datos bibliográficas, términos de búsqueda y número de resultados obtenidos se muestra en la Tabla 1.

Cadena de búsqueda	Resultados Scopus	Resultados Scielo
("Acquisition" OR "Subscription") AND "e-journals" AND "University Libraries"	9	0
"Bibliometric Methods" AND ("Publication Patterns" OR "Citation Patterns")	22	1
"Bibliometric Indicators" AND ("Publication Patterns" OR "Citation Patterns")	36	25
"Analysis" AND "Scientific Publications" AND "Data Mining" AND "Techniques"	11	0
("Collection Tool" OR "Visualization Tool") AND ("Publication Patterns" OR "Citation Patterns")	47	16
Total	125	42

Tabla 1. Cadenas utilizadas y resultados obtenidos durante la búsqueda en las bases de datos bibliográficas.

El cuarto paso fue la aplicación de los criterios prácticos de cribado ya que en las búsquedas preliminares se obtuvo un gran número de artículos. Se



seleccionaron artículos publicados en el periodo 2016-2020 en inglés y español. El idioma español permitió encontrar principios, pautas e indicadores aplicados en las instituciones regionales. En cuanto al inglés, al ser el idioma científico universal (Gutiérrez Ramírez & Landeros Falcón, 2010), se considera fundamental en todas las investigaciones. Al aplicar los criterios prácticos, se define que solo se revisen los documentos de tipo artículo de revista. El motivo principal de esta decisión fue que estos artículos representan actualmente el más alto nivel de investigación con información actual y de alta calidad, a diferencia de los artículos de conferencias que generalmente son trabajos en progreso que no han sido concluidos (Siguenza-Guzman et al., 2015). El resultado fue 167 documentos analizados, de los cuales el 5.4% (9 de 167) corresponden a la búsqueda de criterios de suscripción a bases de datos bibliográficas, el 6.6% (11 de 167) a técnicas de minería de datos, el 13.8% (23 de 167) a métodos bibliométricos, el 36.5% (61 de 167) a indicadores bibliométricos, y el 37.7% (63 de 167) a herramientas de recopilación y visualización de la producción científica.

Una vez seleccionada la información, el quinto paso fue la aplicación de los criterios de selección metodológica, en la que se establece la revisión de la sección de metodología, resultados y conclusión por artículo. El objetivo fue encontrar criterios de adquisición o suscripción a bases de datos bibliográficas, métodos usados para el análisis e indicadores bibliométricos, técnicas de minería de datos y herramientas de software utilizadas para la extracción y visualización de la producción científica. Se requirió el uso de una “Matriz de Hallazgos”, que tiene la función de listar los datos importantes luego de la implementación del filtro práctico y metodológico. Esta matriz contiene información estratégica como el año de publicación, el nombre del autor, el título del documento, la información del contenido, área de investigación, resumen y conclusiones. Se utiliza para realizar un análisis rápido del contenido lo que ayuda a determinar si los artículos contienen información relevante para la investigación.

El sexto paso corresponde al estudio de la literatura seleccionada. Para ello, se leyeron todos los trabajos recuperados con el fin de identificar la información que permitiera dar respuesta a las preguntas formuladas en el primer paso.

Se utilizó el análisis de metadatos y el análisis de contenido. Dado que el análisis de metadatos es un enfoque para mejorar la precisión del descubrimiento de recursos, permite que la información se organice de forma estructurada (Chen et al., n.d.), por lo que puede ser necesario una serie de criterios y procedimientos para diferentes tipos de metadatos. El análisis de contenido contribuye a desarrollar un diálogo sobre la evaluación de la calidad de los metadatos (Moen et al., 1997) y si cumple con el propósito y los objetivos del estudio seleccionado.

Finalmente, el último paso formula los resultados. Por lo tanto, los datos se exportaron a una hoja de cálculo para realizar análisis de contenido y metadatos. Para el análisis de contenido, se establecieron varias secciones dentro de la matriz según los criterios de intervención desarrollados en la metodología PICO. El objetivo fue encontrar criterios y estrategias para la adquisición de base de datos bibliográficas, métodos e indicadores bibliométricos con su respectiva descripción y principales características.

Por otra parte, para el análisis de metadatos, se realizó un estudio estadístico descriptivo a través de las frecuencias de la información, como el año de publicación, productividad por país, área de investigación, métodos e indicadores bibliométricos y herramientas de recolección y visualización.

3.2. Criterios de suscripción de bases bibliográficas

Como punto de partida se inicia con los *criterios de adquisición y suscripción para las bases de datos bibliográficas*, donde se tienen nueve estudios realizados, los mismos que reflejan que este tema no ha tenido gran impacto en la investigación. En el año 2020, se tiene una productividad del 44.4% (4 de 9), 33.3% (3 de 9) del año 2019, 11.1% (1 de 9) del año 2018 y el año 2017, respectivamente. El país con más publicaciones es India con un 33.3% (3 de 9) seguido de Nigeria y Estados Unidos con un 22.2% (2 de 9) cada uno, como se puede visualizar en la Figura 1.

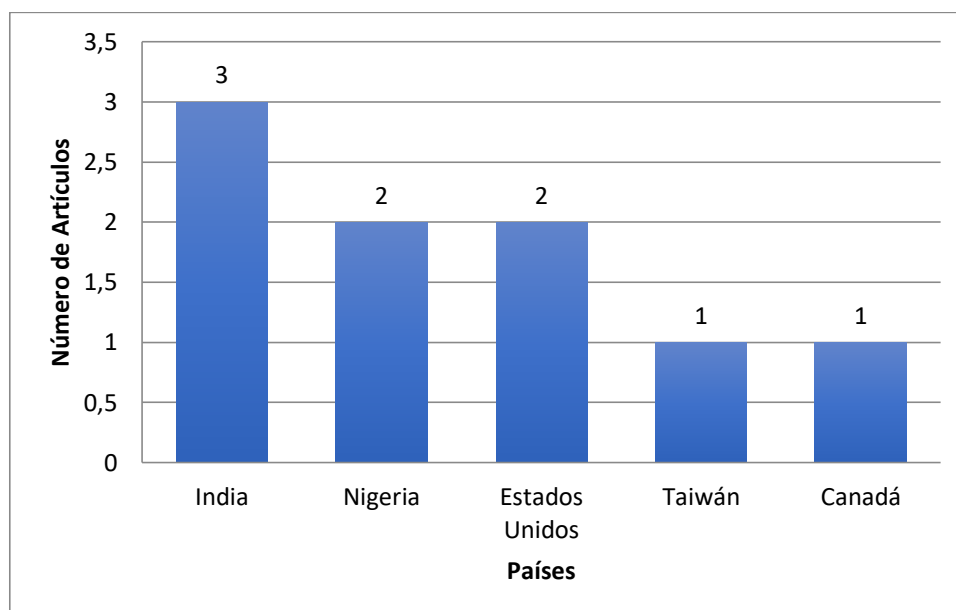


Figura 1. Cantidad de información recuperada sobre criterios de adquisición y suscripción por país.



Los estudios se han centrado en evaluar la disponibilidad y uso de los recursos de información en las bibliotecas universitarias (Tukur & Kannan, 2020), perspectivas de la formación de consorcios (Zhao, Zhao, & MacGillivray, 2017), los inventarios de suscripciones a revistas electrónicas para mejorar la documentación interna, además de analizar el grado de satisfacción de los usuarios con los recursos bibliotecarios e instalaciones (Urban, 2019).

En cuanto a la pregunta: 1) *¿Cuáles son los criterios, métodos o estrategias para la adquisición de bases de datos bibliográficas en las bibliotecas universitarias?* Anteriormente se ubicaron los artículos que describen a las entidades o instituciones que realizan la suscripción o adquisición a un repositorio bibliográfico. Los artículos se enfocan en la evaluación de la disponibilidad de recursos digitales y servicios bibliotecarios (Tukur & Kannan, 2020), mencionan la gestión del conocimiento para el desarrollo de colecciones (Proctor, 2019) y hacen énfasis en proporcionar y mantener acceso a las publicaciones periódicas electrónicas (Zhao et al., 2017).

Como estrategias más relevantes para la adquisición o suscripción de bases de datos bibliográficas se tienen. Primero, formalizar y fortalecer los vínculos entre estudiantes de pregrado y personal de la biblioteca para mejorar el acceso a la información (Tukur & Kannan, 2020). Segundo, creación de redes, catalogación, clasificación e intercambio de materiales de información (Babatunde et al., 2020). Tercero, uso de selectores, que es personal de la biblioteca que posee un gran cantidad de conocimiento táctico y áreas temáticas, además saber qué libros o revistas son de interés y qué formatos se prefieren (Proctor, 2019). Y, finalmente, formar un consorcio entre bibliotecas en donde cada una tenga un representante que participa en la toma de decisiones para la colección del consorcio (Zhao et al., 2017). Cabe mencionar que, si no se tiene conocimiento de compras, concesión de las licencias, negociaciones con los proveedores, planificación presupuestaria y la creación de políticas, puede llegar a ser un desafío para quienes no han estado directamente involucrados (Proctor, 2019).

3.3. Métodos e indicadores bibliométricos utilizados en el análisis de la producción científica

Respecto a *métodos e indicadores bibliométricos*, las publicaciones de Scopus y Scielo se encuentran distribuidas de la siguiente manera: con un 23.9% (20 de 84) del año 2020, 22.6% (19 de 84) del año 2019, 20.2% (17 de 84) del año 2018, 15.4% (13 de 84) del año 2017 y 17.9% (15 de 84) del año 2016. Estos resultados reflejan un incremento en el número de investigaciones publicadas año tras año. En la Figura 2 se muestra los países más prolíficos de acuerdo

con Scopus, China con 13.8% (8 de 58) e India con 10.3% (6 de 58). En Scielo, sobresale Cuba con un 26.9% (7 de 26).

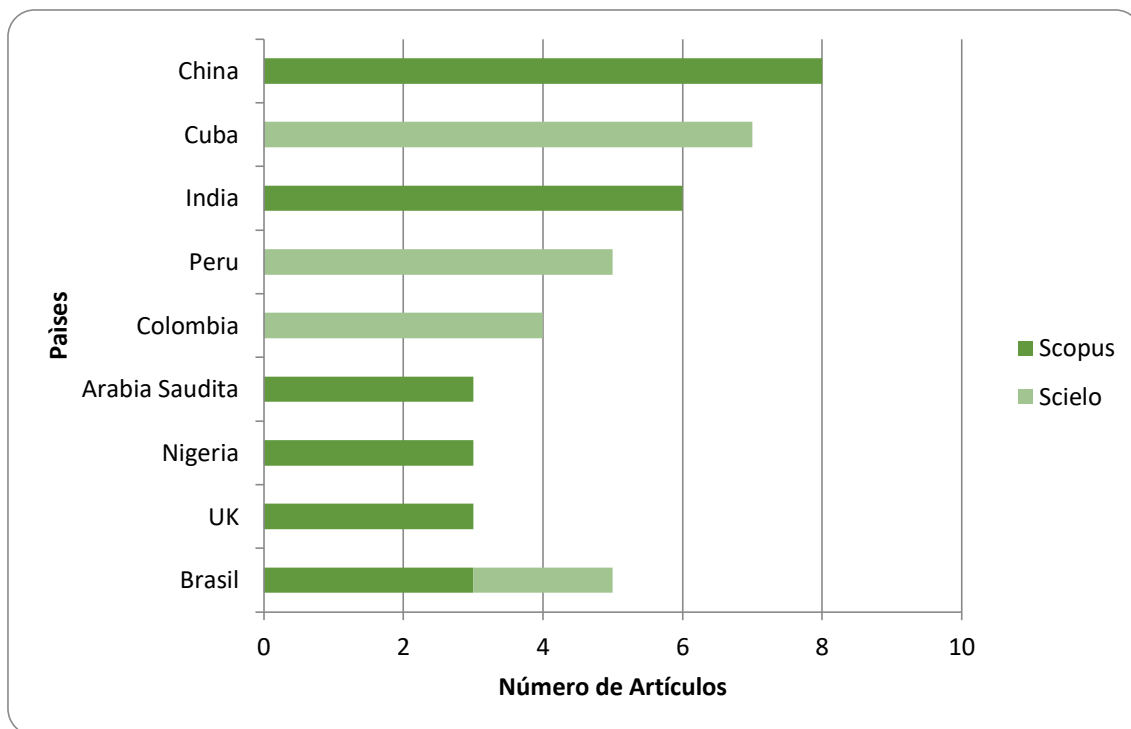
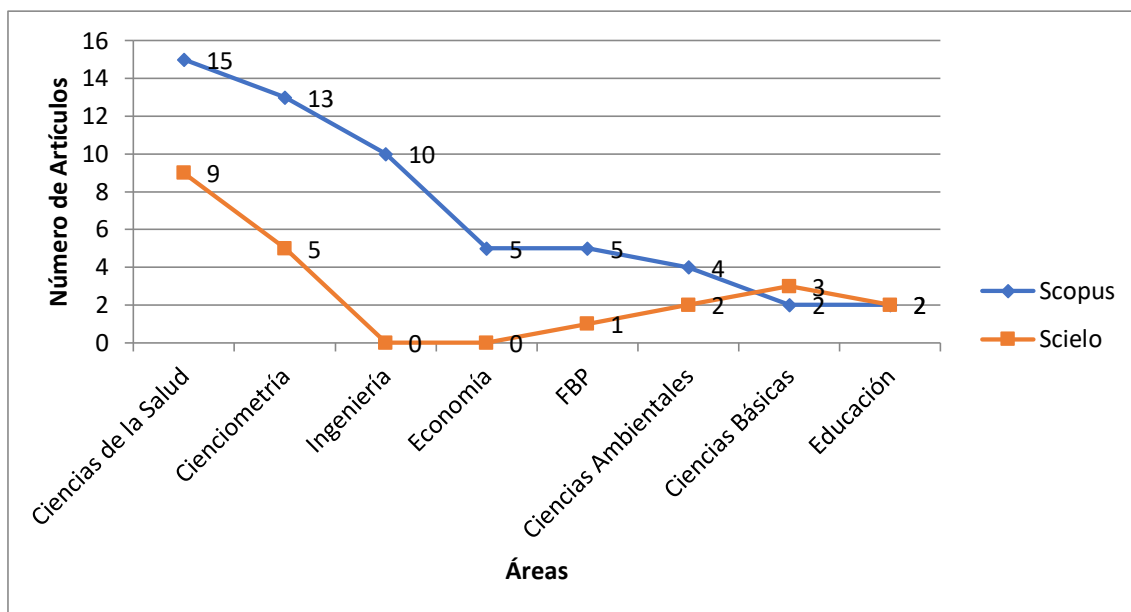


Figura 2. *Número de publicaciones sobre métodos e indicadores bibliométricos por país (Scopus y Scielo)*

Los métodos e indicadores bibliométricos han permitido encontrar años de mayor producción, autores destacados, procedencia de los autores, campo específico de estudio (Danell, 2020; Simao et al., 2020), visibilidad e impacto de la investigación, ráfagas de citas y mapa de la línea de tiempo (Guo et al., 2020), por nombrar algunos ejemplos. Esto hace posible un mapeo confiable del desarrollo científico (Ellegaard, 2018).

Las áreas de investigación a las que se han dirigido los estudios de métodos e indicadores bibliométricos referentes a Scopus y Scielo se pueden visualizar en la Figura 3. En Scopus se destaca el área de Ciencias de la Salud con un 25.9% de publicaciones (15 de 58), seguidos del área de la Cienciometría con un 22.4% (13 de 58) e Ingeniería con un 17.2% (10 de 58). Por otro lado, respecto a Scielo, los resultados ratifican a las áreas de Ciencias de la Salud y la Cienciometría como las más relevantes con un 34.6% (9 de 26) y 19.2% (5 de 26), respectivamente.



* FBP = Filosofía y Práctica Bibliotecaria

Figura 3. Áreas de investigación de Scopus y Scielo sobre métodos e indicadores bibliométricos.

Dentro de las Ciencias de la Salud se tratan temas como la Oncología (Danell, 2020), Medicina Complementaria (Danell et al., 2020) y Ortopedia (Shon et al., 2019), sobre las cuales se realizan estudios bibliométricos que buscan obtener años de mayor producción, autores destacados, procedencia de los autores, entre otros (Shi et al., 2018). Le sigue la Cienciometría que desarrolla actividades como: tendencias de publicación (Ullah, 2019) y análisis de la cinética de autocitas de una revista (Heneberg, 2016), colaboración científica internacional de investigadores y mapeo de la comunicación académica en publicaciones (B S & Rajgoli, 2017), preprints como acelerador de la comunicación académica (Wang et al., 2020), análisis de texto (Ranaei et al., 2020), medición y visualización de la colaboración y la productividad de la investigación (Leidolf et al., 2018), e impacto de la investigación y productividad de los países (Barrot, 2016). Finalmente, como una de las tres áreas relevantes está la Ingeniería, incluyendo temas relacionados con análisis comparativo de patrones de citas en proyectos de investigación de ingeniería mecánica y civil (Osueke et al., 2018).

Respecto a las preguntas de investigación, 2) *¿Qué métodos bibliométricos son utilizados para el descubrimiento de patrones de publicación y citación?*, y 3) *¿Cuáles son los indicadores bibliométricos más utilizados para el análisis de patrones de publicación y citación?* Para su determinación se analizó la frecuencia de los diferentes métodos e indicadores bibliométricos que más se han utilizado en otros trabajos. Para los estudios bibliométricos, como fuente de datos más empleada, se tiene como principal a Web of Science con un 51.7%



(30 de 58), seguido de Scopus con un 29.3% (17 de 58), PubMed y Medline con un 5.2% (3 de 58) cada una.

En cuanto a *métodos bibliométricos*, en Scopus se pueden observar en la Tabla 2 que sobresale el uso de análisis de citas con 27.6% (16 de 58), que permite determinar enlaces y conexiones en los trabajos de investigación (Cañedo Andalia, 1999); el análisis co-citas con un 10.3% (6 de 58), que reconoce la afinidad entre los autores co-citados (Miguel et al., 2007); el análisis de co-palabras con un 10.3% (6 de 58), que mapea la estructura conceptual y temática de un dominio específico (Galvez, 2018). Adicionalmente, tanto la Ley de Lotka, la Ley de Bradford y la Ley de Zipf representan igualmente el 6.9% (4 de 58) de métodos bibliométricos utilizados en la literatura. La ley de Lotka describe la relación cuantitativa entre los autores y los artículos producidos en un campo dado y un periodo de tiempo (Alvarado, 1999). La ley de Bradford permite seleccionar las publicaciones no solo las más productivas sino también las más relevantes para cubrir una determinada área del conocimiento (Alvarado, 2016). Finalmente, la ley de Zipf proporciona mecanismos para identificar y seleccionar las palabras clave de un texto (Alvarado & Arango, 2011).

Método Bibliométrico	# de Artículos
Análisis de citación	16
Análisis de co-cita	6
Análisis de co-ocurrencia de palabras	6
Análisis de co-autores	5
Ley de Lotka	4
Ley de Bradford	4
Ley de Zipf	4

Tabla 2. Métodos Bibliométricos más utilizados para el análisis de la producción científica

Como *indicador bibliométrico* más utilizado en Scopus, se puede observar en la Tabla 3, que destaca el número de citas con un 37.9% (22 de 58), seguido del número de artículos con un 25.9% (15 de 58), número de autores con un 20.7% (12 de 58), el índice h con un 18.9% (11 de 58), el factor de impacto con 13.7%



(8 de 58) y CiteScore con 5.2% (3 de 58). Asimismo, para el análisis bibliométrico se extraen metadatos de las publicaciones como autor, tema, país, institución, revista, año de publicación, categoría, idioma, afiliación y título. También, se presentan los resultados que corresponden a Scielo, en el cual se destaca el número de artículos con un 80.8% (21 de 26), seguido del número de citas con un 65.4% (17 de 26) y el número de autores con un 53.9% (14 de 26), resultados que ratifican a los expuestos en Scopus que determinan el número de artículos y el número de citas como los más relevantes.

Indicador Bibliométrico	Scopus	Scielo
Número de artículos	15	21
Número de citas	22	17
Número de autores	12	14
Índice h	11	7
Factor de impacto	8	5
CiteScore	3	3

Tabla 3. Indicadores Bibliométricos más utilizados en el análisis de la producción científica (Scielo y Scopus).

3.4. Técnicas de minería de datos aplicada a patrones de publicación y citación.

Existe un número reducido de publicaciones que mencionan explícitamente el uso de técnicas de minería de datos como agrupación, clasificación o reglas de asociación. El número de publicaciones por año en Scopus se describe de la siguiente manera, con un 27.3% (3 de 11) en el año 2020, 18.1% (2 de 11) en el año 2019, 27.3% (3 de 11) en el año 2018, 9.1% (1 de 11) en el año 2017 y 18.2% (2 de 11) en el año 2016. Estos resultados convierten a la minería de datos como un campo poco productivo para los estudios de investigación relacionados con patrones de publicación y citación. Por otro lado, en la productividad por país, sobresale España con un 18.2% (2 de 11).



En cuanto a la pregunta, 4) *¿Cuáles son las técnicas de minería de datos más utilizadas para el análisis de patrones de publicación y citación?* Las técnicas de minería de datos se utilizan normalmente para encontrar información oculta en grandes cantidades de datos (Alvarez-Jareño et al., 2018). Se puede aplicar a diferentes áreas. Por ejemplo, Torres-Berru et al., (2020) presentan un estudio cuyo propósito es encontrar algoritmos que identifiquen características específicas de fraude o corrupción. Por otro lado, Marcos-Pablos y García-Peñalvo (2020) describen una metodología con el objetivo de ayudar a adaptar, filtrar y sincronizar la búsqueda y selección de artículos para revisiones bibliográficas.

Se obtiene una lista de las técnicas de minería de datos entre el segundo, tercer y cuarto término de búsqueda respecto a Scopus, debido a que en Scielo no se encontraron resultados referentes al tema de discusión. Entre los resultados se tiene agrupamiento con el 11.6% (8 de 69), clasificación con el 10.1% (7 de 69) y regresión múltiple con el 5.8% (4 de 69), siendo estos porcentajes relativamente bajos. Sin embargo, la siguiente pregunta muestra el uso de herramientas dedicadas exclusivamente a estudios bibliométricos que tienen incluidos funcionalidades de minería de datos.

3.5. Herramientas de recopilación y visualización de la producción científica.

Por último, el uso de herramientas de recopilación y visualización de la producción científica en estudios bibliométricos han sido notables. Scopus refleja en el año 2020 un número de publicaciones del 31.9% (15 de 47), 23.4% en el año 2019 (11 de 47), 19.1% en el año 2018 (9 de 47), 17% en el año 2017 (8 de 47) y 8.5% en el año 2016 (4 de 47). Estos resultados tienen una estrecha relación con los resultados presentados respecto a métodos e indicadores bibliométricos, ya que, para realizar dichos estudios, en la mayoría de las publicaciones utilizan por lo menos una herramienta de recopilación y visualización. En la Figura 4 se presentan los países más prolíficos de Scopus y Scielo, dentro de los cuales en Scopus se enmarca a Estados Unidos con un 25.5% (12 de 47), China con 12.8% (6 de 47) e India con un 10.6% (5 de 47). Por otro lado, los resultados relacionados con Scielo precisan una productividad del 6.3% (1 de 16) en el año 2020, 31.3% (5 de 16) en el año 2019, 25% (4 de 16) en el año 2018, 31.3% (5 de 16) en el año 2017 y 6.3% (1 de 16) en el año 2016. Como se puede apreciar en la Figura 4, el país con más publicaciones es Brasil con un 81.2% (13 de 16), muy por encima de Argentina, México y Singapur que presentan productividades del 6.3% (1 de 16), respectivamente. Debido a que Scielo es una biblioteca electrónica más regional a nivel de Latinoamérica, existe una diferencia considerable respecto a la productividad con Scopus que es una base de datos mundial, llegando a la

deducción de que en Latinoamérica se está trabajando en el uso de este tipo de herramientas, sin embargo, aún existe una brecha con otros países en temas de investigación.

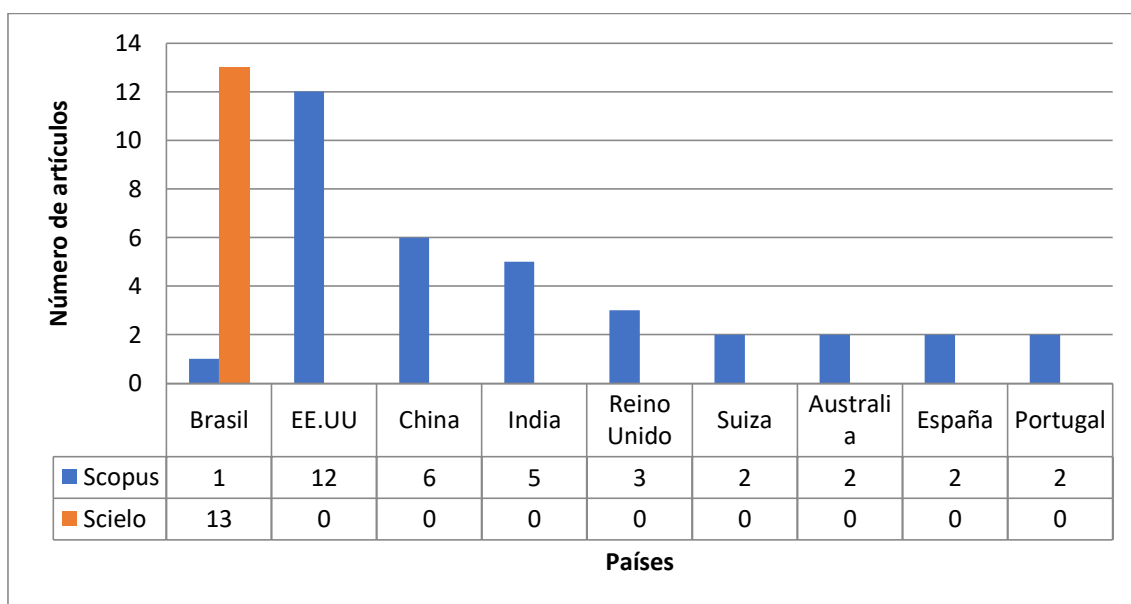


Figura 4. Cantidad de información en Scopus y Scielo sobre herramientas de recopilación y visualización.

Sobre la pregunta de investigación: 5) *¿Qué herramientas de recopilación y visualización de la producción científica, son utilizadas para un análisis de patrones de publicación y citación?* Para la identificación de las herramientas de recopilación y visualización se toma el total de documentos obtenidos de Scopus entre los términos de búsqueda de métodos e indicadores bibliométricos, conjuntamente con el término de herramientas de recopilación y visualización dando un total de 105 documentos. Las herramientas que sobresalen en la recopilación de datos son HistCite con un 4.8% (5 de 105), CiteSpace con un 2.9% (3 de 105), ScientoPy con un 1.9% (2 de 105) y Endnote con un 1.9%. Respecto a las herramientas de visualización se tiene Excel con un 20.9% (22 de 105), Tableau, DAVID, SPSS, Sci2, Ucinet y Pajek, todas con un porcentaje de 1.9% (2 de 105). Por último, la herramienta más utilizada tanto para recopilación y visualización con un 28% (29 de 105) es VOSviewer ya que, es una herramienta de software para construir y visualizar redes bibliométricas. Estas redes pueden incluir, por ejemplo, revistas, investigadores o publicaciones individuales, y pueden construirse sobre la base de la citación, el acoplamiento bibliográfico, la co-citación o las relaciones de coautoría. VOSviewer también ofrece una funcionalidad de minería de textos que puede utilizarse para construir y visualizar redes de co-ocurrencia de términos importantes extraídos de un cuerpo de literatura científica (Van Eck & Waltman, 2018). En la Tabla 4 se observan las herramientas de recopilación y



visualización encontradas en Scopus que más sobresalen. En cuanto a Scielo, las herramientas que destacan en recopilación y visualización de datos son Excel con un 45.2% (19 de 42), VOSviewer con un 14.3% (6 de 42), EndNote con un 9.5% (4 de 42) y Cochrane con un 7.1% (3 de 42), mismas que se visualizan en la Tabla 5.

Herramienta	# de Artículos
VOSviewer	29
Excel	22
HistCite	5
CiteSpace	3
Pajek	2
ScientoPy	2
EndNote	2

Tabla 4. Herramienta de Recopilación y Visualización Scopus.

Herramienta	# de Artículos
Excel	19
VOSviewer	6
EndNote	4
Cochrane	3
SPSS	3
Ucinet	3
NewDraw	2



Tabla 5. Herramienta de Recopilación y Visualización Scielo.

3.6. Recapitulación

A través del desarrollo de esta revisión, fue posible determinar que no se ha explotado la capacidad del uso de patrones de publicación y citación para la gestión de colecciones. Patrones que orienten a la adquisición o suscripción de revistas científicas, base de datos o repositorios bibliográficos, decisión que las bibliotecas universitarias deben tomar constantemente. Como se puede apreciar, dichas decisiones se plantean bajo el criterio de personal con alto grado de conocimiento en concesión de licencias, planificación presupuestaria y demás. Sin embargo, se considera que si existiera otra fuente fiable sobre la cual apoyar tales decisiones haría más objetivo el análisis y, por consiguiente, mejoraría el mantenimiento de las colecciones.

Como siguiente punto, se tienen los métodos bibliométricos, en los cuales se destaca el análisis de citas que permite determinar enlaces y conexiones en los trabajos de investigación. Por otro lado, el análisis de co-citas que reconoce la afinidad entre los autores co-citados. Y, por último, el análisis de co-palabras que mapea la estructura conceptual y temática de un dominio específico. Estos métodos son valiosos ya que después de un determinado proceso dan como resultado mapas de co-autoría, colaboración entre países, conexiones entre temas de investigación, en fin, una gran cantidad de información para analizar. Sin embargo, para aplicar estos métodos es necesaria la unidad de medida denominada “indicador bibliométrico” y en el presente estudio los más destacados son: número de citas, número de artículos y el índice h, los mismos que permiten evaluar rasgos diversos de la actividad científica, vinculados tanto a la producción como al consumo de la información.

Los estudios bibliométricos han aumentado de manera considerable, por lo que hoy en día existe un sin número de herramientas que facilitan dicha actividad. Por mencionar, VOSviewer que es la más utilizada no solo para la recolección y visualización de la información, sino que también proporciona funcionalidades de minería de datos. Además, se debe tomar en cuenta a Excel como una herramienta clave que puede considerarse especial, ya que según el tema con el cual se trabaje se adapta a las necesidades de los usuarios, que va desde el manejo de los datos, realizar cálculos con la información, hasta su visualización mediante gráficas de resumen de resultados.



Capítulo 4: Metodología

En este capítulo, mediante la Ley de Bradford, se evalúan las revistas que más se destacan en número de publicaciones, número de citas, número de búsquedas y factor de impacto. Por otra parte, se explica en qué consiste la metodología de desarrollo tradicional y se menciona el modelo propuesto por Siguenza-Guzman et al. (2013). Las fases de la metodología son: Análisis de requerimientos, Diseño del prototipo, Desarrollo e implementación del prototipo, y Pruebas de funcionamiento y despliegue. Mediante las etapas que abarca esta metodología se determinan las características y funcionalidades de cada sección que posee el prototipo del sistema.

4.1. Aplicación de la Ley de Bradford

Dentro del prototipo se utilizó la Ley de Bradford para encontrar el núcleo de medios de publicación y revistas, haciendo uso de la fórmula propuesta dentro del trabajo de titulación (1).

$$\begin{aligned} JC = & (P1 * Nro.Publicaciones) + (P2 * Nro.Citaciones) \\ & + (P3 * Nro.Búsquedas) + (P4 * SJR^2) \\ & + (P5 * Indexado) \end{aligned} \quad (1)$$

En donde:

- JC (Journal Calification): Calificación total del recurso bibliográfico (revista)
- P1: Peso de publicación
- P2: Peso de citación
- P3: Peso de búsqueda
- P4: Peso del factor de impacto
- P5: Peso indexado

La calificación total del recurso bibliográfico (revista) se calcula en función de cuatro indicadores bibliométricos: número de publicaciones, número de citas, número de búsquedas y factor de impacto (SJR). Para cada uno de los indicadores se le asigna un peso que determina la importancia del mismo, ya que por ejemplo si se le asigna un peso de cero a P1, en la fórmula el número de publicaciones no va a ser considerado.

² SJR: [Scimago Journal & Country Rank](#)



Adicionalmente, se tiene el valor de *Indexado* que permite determinar si la fuente bibliográfica esta indexada y en qué bases de datos consta. La finalidad de la fórmula es evaluar la revista bajo diferentes directrices de consumo y producción que permita obtener una lista de las fuentes bibliográficas de mayor estimación.

Los filtros que intervienen para hacer uso de la fórmula descrita anteriormente son: Periodo de tiempo en años, Área Unesco y Frascati que son áreas del conocimiento utilizadas en el Centro de Documentación Regional “Juan Bautista Vázquez” (CDRJBV) de la Universidad de Cuenca y estas a su vez se dividen en tres categorías: amplio, específico y detallado. Finalmente, luego de aplicar la Ley de Bradford se obtiene el total y porcentaje acumulado de acuerdo con el área Unesco o Frascati en un periodo de tiempo determinado. Con estos valores se conocen qué revistas o publicaciones tienen mayor relevancia y productividad dentro del CDRJBV.

4.2. Metodología de desarrollo tradicional

Al inicio, el desarrollo del software era artesanal en su totalidad. La fuerte necesidad de mejorar el proceso y llevar los proyectos a la meta deseada, obligó a importarse la concepción y fundamentos de metodologías existentes en otras áreas y adaptarlas al desarrollo de software. Esta nueva etapa de adaptación contenía el desarrollo dividido en etapas de manera secuencial que de algo mejoraba la necesidad latente en el campo del software. Entre las principales metodologías tradicionales se tiene la metodología o modelo en cascada (Figueroa-Diaz et al., 2007).

El modelo en cascada propiciado por Winston Royce en 1970 sugiere un enfoque sistemático y secuencial, disciplinado y basado en análisis, diseño, pruebas y mantenimiento. Al final de cada etapa se reúnen y revisan los documentos para garantizar que se cumplen los requerimientos antes de avanzar a la fase siguiente (Garcés & Egas, 2013). Pionero en guiar el proceso de desarrollo de software dirigido por un plan, introduciendo una planificación de cada fase antes de empezar a trabajar en ella.

Para el desarrollo del prototipo se utilizó la metodología tradicional o en cascada, de cinco etapas: análisis de requerimientos, diseño del prototipo, desarrollo e implementación, pruebas y despliegue. A continuación, en la Figura 5 se puede visualizar las etapas de esta metodología con los componentes principales en relación con el actual caso de estudio.

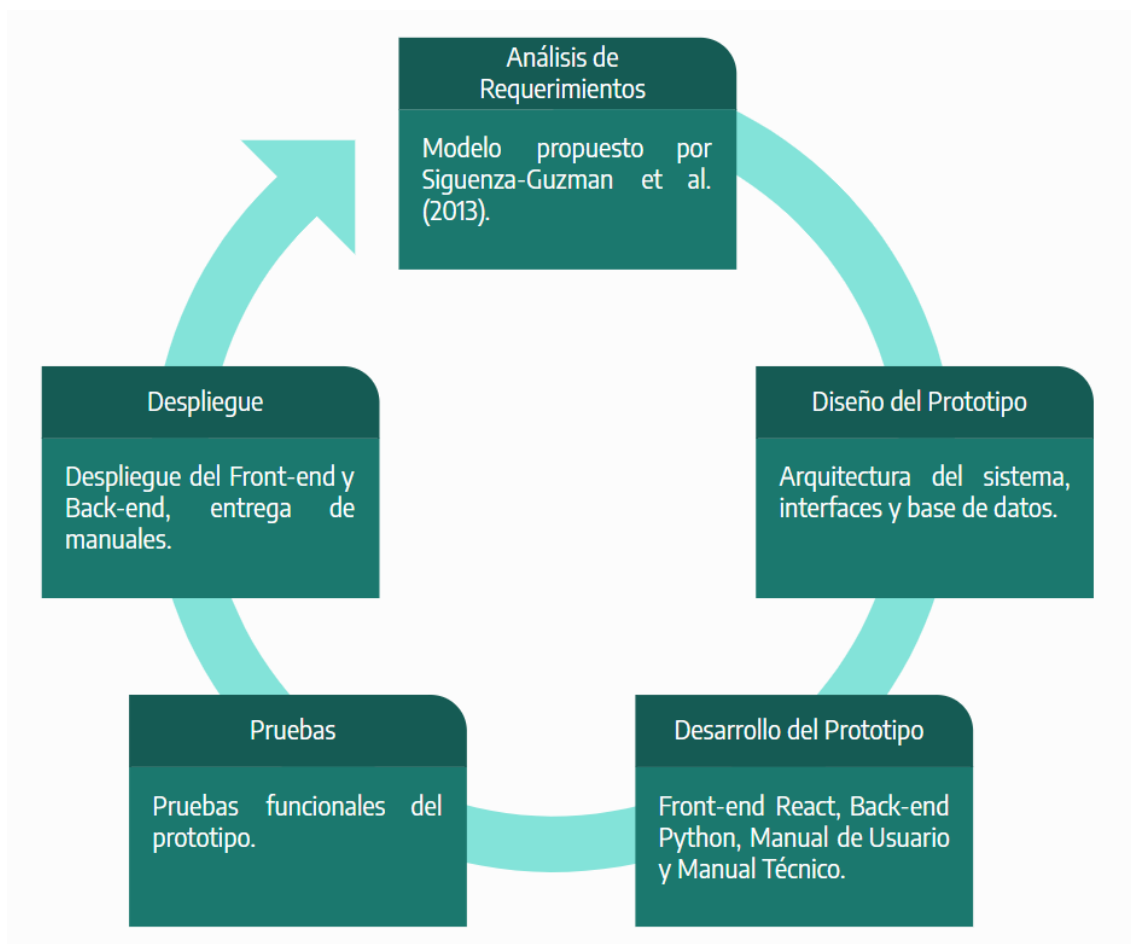


Figura 5. *Etapas establecidas en la Metodología Tradicional para el desarrollo el Prototipo*

Como punto de partida se tiene el análisis, para el cual se tomó el modelo propuesto por Siguenza-Guzman et al. (2013) que se plasma en la Figura 6. El diseño incluyó la descripción de la funcionalidad de las interfaces, la arquitectura cliente-servidor del sistema y el diseño de la base de datos definida por un modelo conceptual y un modelo lógico. Para la implementación se utilizaron herramientas de desarrollo web. Es decir, React, para el lado del cliente, que se maneja a nivel de componentes, los mismos que pueden ser reutilizables haciendo al sistema escalable y fácil de mantener. Y, Python, en el lado del servidor, por ser uno de los lenguajes más utilizados para técnicas de minería de datos. Se realizaron pruebas funcionales del prototipo y, finalmente, el despliegue.



Figura 6. *Modelo para la Evaluación de la Importancia de la Colección*

Respecto al análisis y toma de decisiones en relación a la adquisición/suscripción de bases digitales se utilizó la Ley de Bradford, que permite encontrar el núcleo de medios publicación/revistas, tomando en consideración el número de citas, número de publicaciones, factor de impacto y estadísticas de los proveedores, seleccionando no sólo las más productivas sino también las más relevantes para cubrir una determinada área del conocimiento (Alvarado, 2016).

4.3. Análisis de requerimientos

Para el análisis de requerimientos se tomó como base el análisis de la situación actual de cómo se maneja el procesamiento para la adquisición o suscripción de bases de datos bibliográficas, se plantearon las oportunidades de mejora y la propuesta.



4.3.1. Descripción de la situación actual

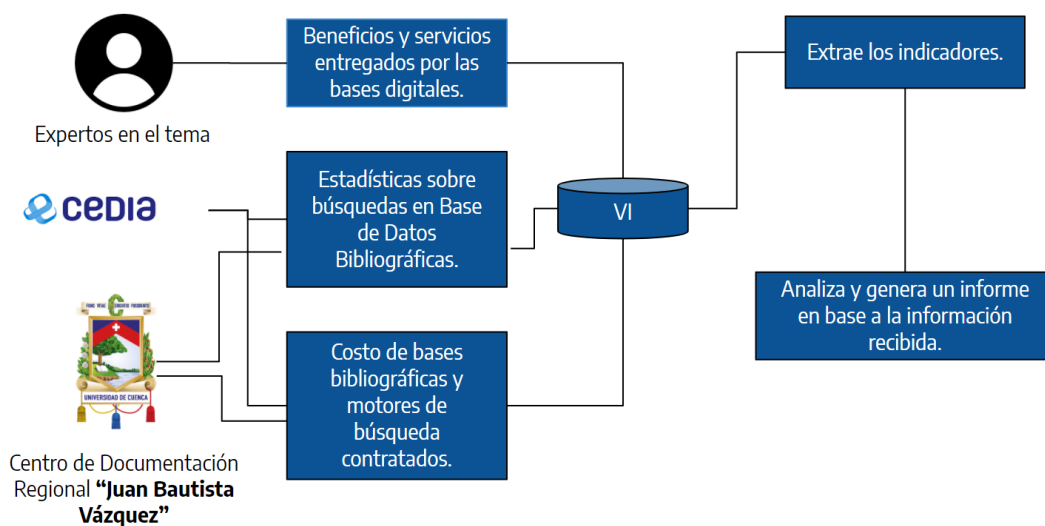
En la Universidad de Cuenca, el Vicerrectorado de Investigación es el encargado del análisis para la toma de decisiones acerca de la renovación o adquisición de bases de datos bibliográficas, el mismo que cuenta con información obtenida desde las siguientes fuentes:

- Estadísticas sobre búsquedas realizadas en las distintas bases de datos bibliográficas, información entregada por los proveedores de bases de datos al CDRJBV y por la red CEDIA.
- Información sobre los beneficios y servicios entregados por las distintas bases de datos bibliográficas consultadas a expertos en el tema, investigaciones y otros usuarios.
- Información acerca del costo de las bases bibliográficas y motores de búsqueda contratados y otras opciones disponibles en el mercado académico. El precio para cada base digital fue tomado de las distintas plataformas entregadas por el CDRJBV y por la red CEDIA. Se considera para todo un valor anual de la suscripción.

Los indicadores que utilizan para el análisis son:

1. Número de búsquedas mensuales en las distintas bases digitales, tomando en cuenta los 12 últimos meses.
2. El precio y el total de búsquedas para cada base digital.
3. Promedio de búsquedas por base digital y su desviación estándar.

Una vez obtenidos los resultados se analizan cada una de las bases de datos bibliográficas, planteando diferentes escenarios tanto para renovación o nuevas adquisiciones. Luego, se elabora una tabla detallada con el repositorio bibliográfico, proveedor, costo actual, costo aproximado en función de los escenarios y la suscripción con su respectiva descripción y recomendaciones. También es importante mencionar que se posee las áreas y servicios que proporciona cada repositorio bibliográfico; como resultado obtienen un informe que sirve como fundamento para la toma de decisiones en relación con las bases de datos bibliográficas (Cabrera y Saenz, 2021). A continuación, en la Figura 7 se presenta el modelo conceptual de la metodología descrita.



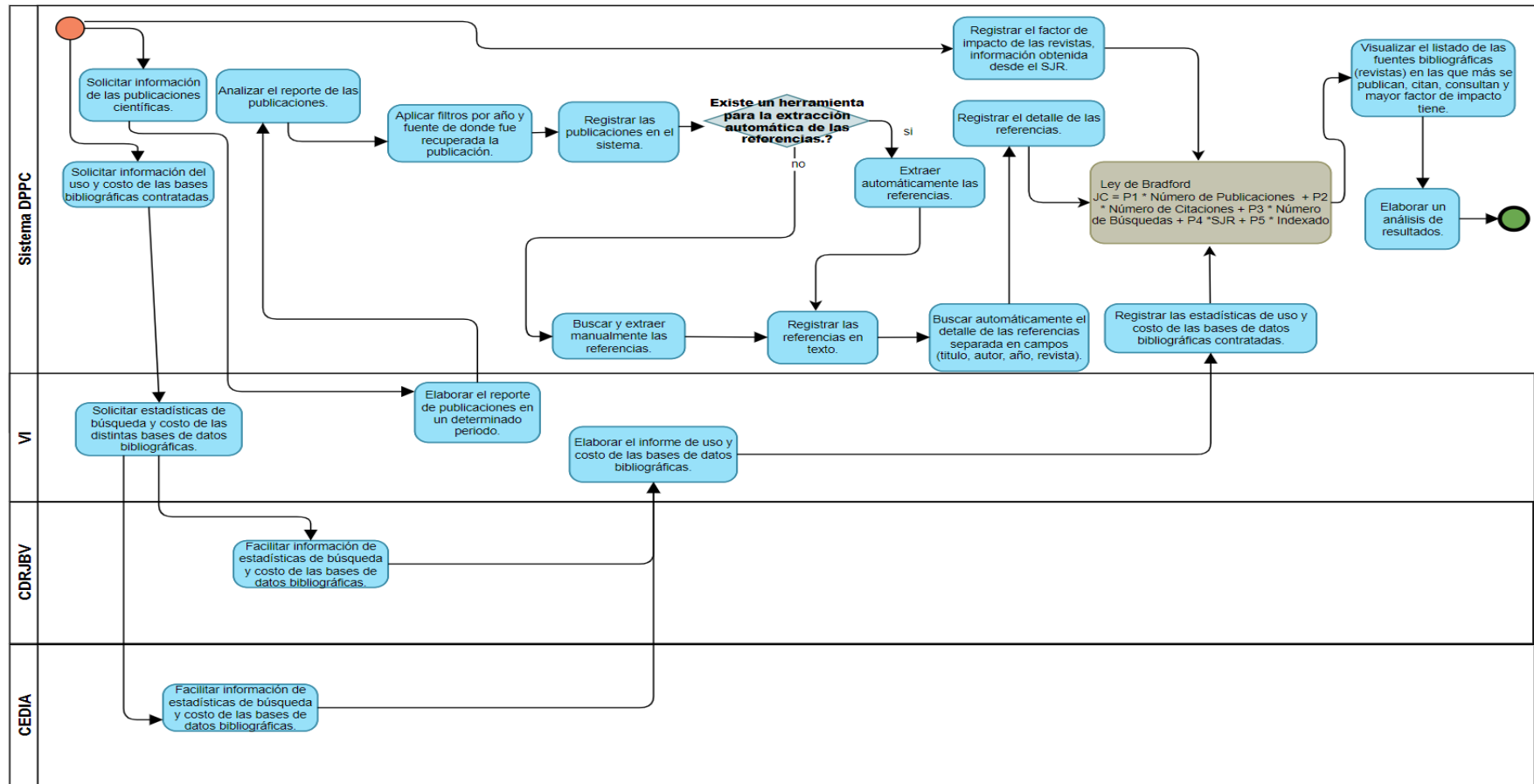
* VI = Vicerrectorado de Investigación

Figura 7.

Modelo conceptual actualizado de la metodología.

4.3.2. Identificación y evaluación de las oportunidades de mejora

La toma de decisiones tiene como base fundamental tres fuentes de información, como: estadísticas de búsqueda, beneficios y servicios entregados por las distintas bases de datos bibliográficas y la información de costo de bases bibliográficas y motores de búsqueda. Se propone agregar una nueva fuente de información, la misma que se basa en la evaluación de la actividad científica de los investigadores con filiación a la Universidad de Cuenca. El proceso consiste en obtener los artículos científicos de los investigadores, extraer los indicadores bibliométricos y las referencias de cada documento, con el objetivo de analizar y obtener conocimiento del número de artículos y qué tipo y calidad literatura se consumió (ej. revistas); en función del número de publicaciones, número de citas, estadísticas de los proveedores y factor de impacto. En la Figura 8 se muestra el mapa conceptual con visión futura.



- * DPCC = Descubrimiento de Patrones de Publicación y Citación
- * CDRJBV = Centro de Documentación Regional Juan Bautista Vázquez
- * VI = Vicerrectorado de Investigación

Figura 8. Modelo conceptual actual con visión futura.



4.3.3. Sistema de Información Propuesto

Se plantea desarrollar un prototipo de sistema que registre publicaciones (artículos científicos) y sus correspondientes referencias, obtención de metadatos de cada una de las referencias de manera automática, aplicación de métodos e indicadores bibliométricos, técnicas de minería de datos y un análisis estadístico descriptivo. Este prototipo tiene como propósito obtener patrones de publicación y citación que proporcionen información que apoye a la toma de decisiones acerca de la renovación o adquisición de bases de datos bibliográficas.

Este nuevo punto de vista para analizar el uso de las bases de datos bibliográficas será de gran utilidad para el personal encargado de generar informes para la toma de decisiones.

4.3.3.1. Descripción de los requerimientos

Se requiere crear el prototipo de sistema para gestionar la información de los artículos y que permita integrar metodologías e indicadores de análisis bibliométrico con operaciones de minería de datos y un análisis estadístico descriptivo, con el objetivo de descubrir información útil acerca del uso de bases de datos bibliográficas. Para el análisis de requerimientos se toma como base el modelo propuesto por Siguenza-Guzman et al. (2013) que evalúa la importancia de la colección combinando cuatro tipos de análisis de patrones: publicación, citación, descarga y factor de impacto.

Patrones de publicación: Se toma en consideración los metadatos de los artículos publicados por los autores con filiación a la Universidad de Cuenca, entre los cuales se tiene: área de investigación, medio de publicación, cuartil y orden del autor.

Patrones de citación: Se analizan las referencias citadas en cada uno de los artículos de los autores con filiación a la Universidad de Cuenca, extrayendo sus metadatos, tomando como principal el medio de publicación (revista).

Patrones de descarga: Se analizan las estadísticas proporcionadas por los proveedores de bases de datos e información recolectada por el Vicerrectorado de Investigación, por ejemplo, número de búsqueda mensual de cada base de datos bibliográfica, descripción de la suscripción y el detalle de las áreas y servicios que suministran.



Factor de impacto: Se analiza el factor de impacto extraído directamente desde el SJR (SCImago Journal Rank). Se analizan los medios de publicación de las referencias de acuerdo con el número de citas que las mismas han tenido por los investigadores de la Universidad de Cuenca.

Los indicadores bibliométricos seleccionados fueron: Número de publicaciones, número de citas, factor de impacto y cuartil. Respecto a las técnicas de minería de datos se seleccionó el algoritmo de predicción Arima, con el fin de predecir el número de búsquedas realizadas en las revistas de Ebsco conjuntamente con Science Direct para los próximos seis meses, en función del número de búsquedas desde octubre del 2020 a octubre del 2021. Adicionalmente, se realizó un análisis estadístico descriptivo respecto al número de publicaciones por área, medio de publicación y cuartil.

A continuación, en la Tabla 6, se presentan los requerimientos funcionales que abarca el prototipo:

Número	Requisito Funcional	Descripción
1	Gestionar metadatos de las publicaciones de los autores y coautores con filiación a la Universidad de Cuenca	Fuente: Hoja de cálculo. Datos: Base de datos digital, URL de la publicación, título, título alternativo, palabras clave, abstract, resumen, área Frascati, área Unesco, tipo de publicación, año de publicación, enlace de la revista, DOI, estado de la publicación, enlace del documento, factor de impacto, cuartil, autor identificación, orden autor, nombres, nombre filiación, medio de publicación. Operaciones: <ul style="list-style-type: none">• Ingresar, listar y eliminar datos de las publicaciones.• Listar las referencias de las publicaciones.• Listar detalles de las referencias de las publicaciones• Buscar una publicación por autor, título, año, tipo de publicación, medio de publicación, área, fuente, cuartil o DOI.
2	Gestionar datos de los autores con filiación en la Universidad de Cuenca y sus respectivas	Fuente: Hoja de cálculo. Datos: Identificación del autor, título,



	publicaciones	<p>enlace del documento.</p> <p>Operaciones:</p> <ul style="list-style-type: none"> • Ingresar, listar y eliminar datos de los autores. • Listar las publicaciones asociadas a un determinado autor. • Eliminar las publicaciones individuales de cada autor • Buscar autor por número de identificación o por nombre • Buscar publicación por nombre dentro de ver publicaciones de los autores
3	Extraer referencias de las publicaciones de manera automática. En el caso de que no sea posible, el registro se realizará de manera manual	<p>Fuente: Fuente externa (ej. Scopus) permite descargar las referencias de los artículos. Caso contrario, a través del URL de la publicación facilita el ingreso manual de las referencias.</p> <p>Datos: Referencia</p> <p>Operaciones:</p> <ul style="list-style-type: none"> • Listar datos de las referencias de la publicación • Ingreso de las referencias de manera manual a través de la publicación a la que pertenece • Ingreso de las referencias de manera automática haciendo uso de la API de Scopus • Ingreso de las referencias detalladas de manera automática haciendo uso de la API de Google Académico • Listar las referencias extraídas por las APIs • Buscar las referencias extraídas por su nombre • Buscar las referencias de las publicaciones por autor, título, año, tipo de publicación, medio de publicación, área, fuente o DOI.
4	Gestionar los medios de publicación donde se encuentran publicados los	<p>Datos: Nombre del medio de publicación.</p>



	documentos de los investigadores con filiación a la Universidad de Cuenca	<p>Operaciones:</p> <ul style="list-style-type: none"> • Ingresar, listar, editar y eliminar los medios de publicación • Buscar un medio de publicación por el nombre
5	Registro del ranking de medios de publicación, citación y número de búsquedas	<p>Datos: Nombre del medio de publicación.</p> <p>Operaciones:</p> <ul style="list-style-type: none"> • Listar y buscar los rankings de medios de publicación por nombre o número de publicaciones • Listar y buscar los rankings de medios de citación por nombre o número de citas • Listar y buscar los rankings de números de búsquedas por nombre o número de búsquedas
6	Registro del índice de factor de impacto proporcionado por SJR (SCImago Journal Rank)	<p>Fuente: Hoja de cálculo.</p> <p>Datos: Identificación del SRJ, título, tipo, ISSN, índice SJR (factor de impacto), cuartil</p> <p>Operaciones:</p> <ul style="list-style-type: none"> • Ingresar y listar datos acerca de la influencia científica de las revistas académicas según el número de citas en otros medios y periódicos o revistas de importancia proporcionados por SJR • Buscar por ranking, id del recurso, título, tipo, ISSN, SJR, cuartil
7	Análisis de la Ley de Bradford	<p>Permitirá ejecutar los cálculos de la Ley de Bradford para determinar cuáles son los medios de publicación de uso principales.</p> <p>Operaciones:</p> <ul style="list-style-type: none"> • Seleccionar año inicial • Seleccionar año final • Seleccionar área Unesco o Frascati



		<ul style="list-style-type: none">Listar los resultados de aplicar la Ley de Bradford: total y porcentaje acumulados
8	Gestionar áreas a las cuales pertenecen los documentos publicados	Datos: nombre del área. Operaciones: <ul style="list-style-type: none">Ingresar, listar, editar y eliminar las áreas a las cuales pertenece una publicaciónBuscar un área por su id o por el nombre
9	Gestionar área SJR (SCImago Journal Rank)	Datos: nombre del área. Operaciones: <ul style="list-style-type: none">Ingresar, listar, editar y eliminar un área SJRAsignar una categoría a un área existenteBuscar un área por su nombreListar y buscar las categorías de un área por su nombre
10	Registrar base de datos bibliográfica	Datos: Nombre de la base de datos bibliográfica, proveedor, costo actual, suscripción/descripción, área/servicio. Operaciones: <ul style="list-style-type: none">Crear, listar, editar y eliminar una base de datos bibliográficaBuscar una base de datos bibliográfica por id, nombre, proveedor, costo actual, suscripción/descripción o área/servicio
11	Gestionar las estadísticas de uso de base de datos bibliográficas contratadas	Datos: Base de datos bibliográfica, revista (medio de publicación), año, mes, número de búsquedas. Operaciones: <ul style="list-style-type: none">Crear, listar, editar y eliminar una estadística de un proveedor



		<ul style="list-style-type: none">• Visualizar una gráfica de las estadísticas de los proveedores seleccionando el nombre de la base de datos bibliográfica, medio de publicación, año o mes
12	Análisis estadístico y visualización de resultados	Permite visualizar gráficas de barras y de pastel con información del número de artículos por área, medio de publicación, cuartil y factor de impacto.
13	Técnicas de minería de datos: Predicción del número de búsquedas	Visualización del total del número de búsquedas que tiene una revista académica alojada en un repositorio bibliográfico.
14	Gestionar los parámetros que son necesarios para la aplicación de la Ley de Bradford	<p>Datos: código del parámetro, nombre, valor.</p> <p>Operaciones:</p> <ul style="list-style-type: none">• Crear, listar, editar y eliminar un parámetro de configuración• Buscar un parámetro de acuerdo con su código, nombre o valor

Tabla 6. Requisitos funcionales del Prototipo.

4.3.3.2. Restricciones

- Esta fase comprende únicamente artículos indexados en Scopus y Latindex en el periodo 2016-2020.
- Tipo de Documento: Artículo de Revista y Artículo de Conferencia
- Se cuenta con información de las bases de datos bibliográficas: Scopus, Science Direct y Ebsco.
- Para hacer uso del API de Google Académico se debe estar conectado a la intranet de la Universidad de Cuenca.

4.3.3.3. Actores relacionados

- Vicerrectorado de Investigación
- Centro de Documentación Regional “JUAN BAUTISTA VÁZQUEZ”
- Red CEDIA

4.4. Diseño del Prototipo

4.4.1. Descripción de la arquitectura del sistema

La siguiente sección describe la arquitectura del sistema tanto en su forma lógica como física.

4.4.1.1. Arquitectura Lógica

El sistema consta de un solo módulo cuyo objetivo es descubrir patrones de publicación y citación para determinar el uso de bases de datos bibliográficas. Se divide en dos partes fundamentales, Front-end y Back-end (ver la Figura 9). A continuación, se detallan las funcionalidades que realiza cada una de estas partes.

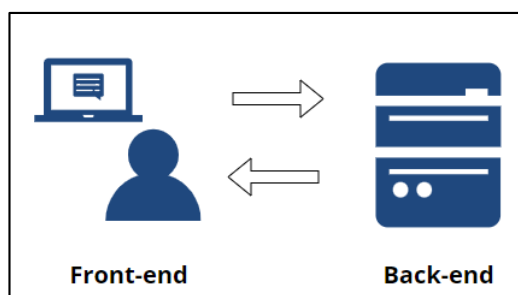


Figura 9. *Arquitectura Lógica del Sistema Web.*

Front-end

El Front-end está dividido en once secciones con las siguientes funcionalidades:

Dashboard

1. Consulta total de publicaciones ingresadas.
2. Consulta total de publicaciones sin referencias.
3. Consulta total de referencias ingresadas.
4. Consulta total detalle referencias extraídas.

Publicaciones

1. Ingreso de publicaciones.
2. Consulta de publicaciones.
3. Consulta referencias en texto por publicación.
4. Consulta referencias detallada en campos por publicación.



5. Editar detalle de la publicación.
6. Eliminar publicación.

Autores

1. Ingreso autores.
2. Consulta autores.
3. Consulta publicaciones por autor.
4. Editar detalle del autor.
5. Visualización gráfica de los autores por orden de autor.
6. Visualización gráfica de los autores por número de publicaciones.

Referencias

1. Consulta publicaciones sin referencias.
2. Ingreso referencias manuales.
3. Ingreso referencias automático a través de la API de Scopus.
4. Consulta publicaciones sin completar referencias.
5. Busca detalle de la referencia a través de la API de Google Académico.

Áreas

1. Ingreso áreas Unesco.
2. Ingreso áreas Frascati.
3. Ingreso áreas SJR.
4. Consultar áreas Unesco.
5. Consultar áreas Frascati.
6. Consultar áreas SJR.
7. Editar área Unesco.
8. Editar área Frascati.
9. Editar área SJR.
10. Eliminar área Unesco.
11. Eliminar área Frascati.
12. Eliminar área SJR.
13. Asignar categorías a un área SJR.
14. Consultar categorías por área.
15. Eliminar categoría asignada a un área.

Base de datos bibliográfica

1. Ingresar base de datos bibliográfica.
2. Consultar base de datos bibliográfica.
3. Editar base de datos bibliográfica.
4. Eliminar base de datos bibliográfica.



5. Ingresar revistas o fuentes bibliográficas pertenecientes a cada base de datos bibliográfica.
6. Consultar revistas o sub-base de datos por base de datos bibliográfica.

Estadísticas Proveedores

1. Ingresar estadísticas de búsqueda por base de datos bibliográfica.
2. Consultar estadísticas de búsqueda por base de datos bibliográfica.
3. Eliminar estadísticas de búsqueda por base de datos bibliográfica.
4. Ingresar estadísticas de búsqueda por revista o subbase de datos bibliográfica.
5. Consultar estadísticas de búsqueda por revista o subbase de datos bibliográfica.
6. Eliminar estadísticas de búsqueda por revista o subbase de datos bibliográfica.
7. Visualización gráfica de las estadísticas de búsqueda por base de datos bibliográfica en función del número de búsquedas por mes y año.
8. Visualización gráfica de las estadísticas de búsqueda por revistas o subbase de datos bibliográfica en función del número de búsquedas por mes y año.

Parámetro

1. Ingresar parámetro.
2. Consultar parámetro.
3. Editar parámetro.
4. Eliminar parámetro.
5. Asignar equivalencia entre área Unesco y área SJR.
6. Consultar equivalencias área SJR por área Unesco.
7. Eliminar equivalencia entre área Unesco y área SJR.
8. Asignar equivalencia entre área Unesco y área Frascati.
9. Consultar equivalencias área Frascati por área Unesco.
10. Eliminar equivalencia entre área Unesco y área Frascati.

Medios de Publicación

1. Ingresar medio de publicación.
2. Consultar medio de publicación.
3. Editar medio de publicación.
4. Eliminar medio de publicación.
5. Lista ranking de medios de publicación.
6. Lista ranking de medios de citación.
7. Lista ranking de número de búsquedas por revista o subbase de datos bibliográfica.



8. Lista revistas con su factor de impacto del SJR.

Ley de Bradford

1. Calcula el ranking de medios de publicación, citación, número de búsquedas con filtro por año y área.
2. Ejecuta la ley de Bradford basada en cuatro indicadores bibliométricos y con filtro por año y área.

Análisis de Datos

1. Consulta y visualización de estadísticas de datos en gráficas de barra y de pastel: Número de publicaciones por área, medio de publicación, cuartil.
2. Ejecución y visualización de los resultados de las técnicas de minería de datos: predicción de los próximos seis meses del número de búsquedas de un medio de publicación.

Back-end

El Back-end procesa las peticiones que recibe del Front-end, las mismas que se encuentran distribuidas de la siguiente manera.

- Gestión de publicaciones
- Gestión de referencias
- Gestión del detalle de la referencia
- Gestión de las áreas Unesco, Frascati y SJR.
- Gestión de las categorías SJR y su conexión con las áreas SJR
- Gestión de la equivalencia de áreas Unesco con las áreas SJR
- Gestión de la equivalencia de áreas Unesco con las áreas Frascati
- Gestión de los medios de publicación
- Gestión de medios de publicación en función de las estadísticas de uso
- Gestión de medios de publicación en función del número de citas
- Gestión de medios de publicación en función del número de publicaciones
- Gestión de las bases de datos bibliográficas
- Gestión de las revistas indexadas en las bases de datos bibliográficas
- Gestión de las revistas registradas en el SJR y su conexión con las categorías SJR
- Gestión de autores
- Gestión de las publicaciones correspondientes a cada autor
- Gestión de estadísticas de uso de las bases de datos bibliográficas

- Gestión de estadísticas de uso de las revistas correspondientes a las bases de datos bibliográficas
- Gestión de la aplicación de técnicas de minería de datos
- Gestión del análisis estadístico descriptivo, a través de las frecuencias de la información.
- Gestión de la Ley de Bradford
- Gestión de los parámetros generales del sistema

Como se mencionó anteriormente, una de las partes fundamentales es el Back-end que se diseñó en base a la arquitectura MVC (Modelo, Vista, Controlador) y el diseño de Servicios RESTful. Esto permite manejar una estructura interna para la generación de servicios web, dichos servicios facilitan la notificación de una petición realizada en el Front-end al Back-end a través de una URI (Identificador de Recursos Uniforme).

4.4.1.2. Arquitectura Física

La arquitectura física del prototipo del sistema fue a través de un enfoque cliente-servidor que se puede visualizar en la Figura 10. El cual, permita una fácil comunicación entre los **usuarios** que interactúan directamente con el Front-end y los **servidores** del sistema donde se encuentran alojados los mismos para su acceso a través del Internet.

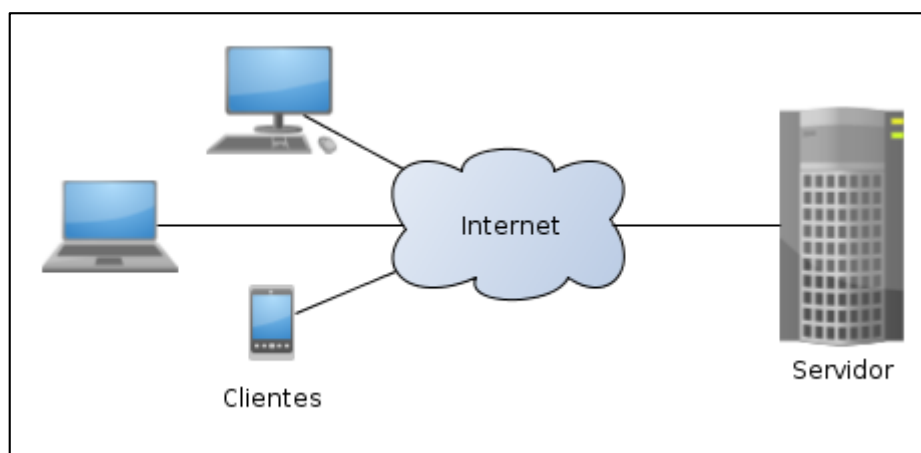


Figura 10. *Arquitectura Física Cliente - Servidor.*

4.4.2. Diseño del Modelo de Clases del Sistema

En la Figura 11 se puede observar el diagrama que representa cada una de las entidades que interactúan en el sistema con sus correspondientes operaciones. La entidad principal es el artículo que gestiona las operaciones para el ingreso de las publicaciones de los autores con filiación a la Universidad de Cuenca,



además de sus conexiones con las áreas, categorías, revistas, medios de publicación y base de datos bibliográficas. Estas permiten tener a detalle las características de cada una de las publicaciones y fuentes bibliográficas (revistas).

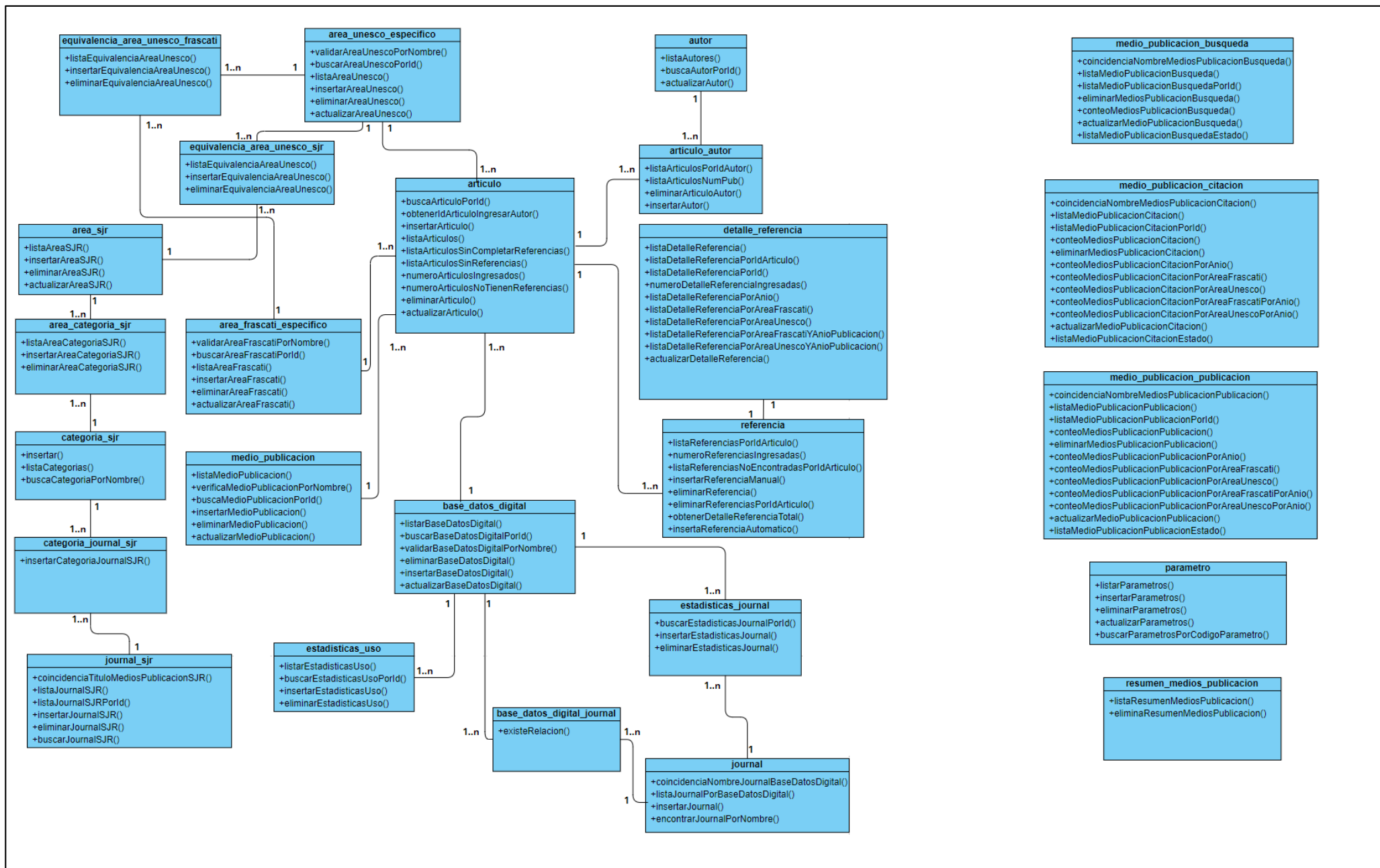


Figura 11. Entidades y operaciones que interactúa en el prototipo de sistema.

Tania Maricela Landívar Ordóñez
 Ronaldo Sebastián Rendón Loja



4.4.2.1. Diseño de la vista

En este caso, se define la manera en la que van a estar expuestos los servicios web para el consumo desde el Front-end, de la siguiente manera:

Rutas

- servicio_articulo
- servicio_referencia
- servicio_detalle_referencia
- servicio_autor
- servicio_articulo_autor
- servicio_area_frascati
- servicio_area_unesco
- servicio_equivalencia_area_unesco_frascati
- servicio_area_sjr
- servicio_equivalencia_area_unesco_sjr
- servicio_categorias_sjr
- servicio_area_categoria_sjr
- servicio_journal_sjr
- servicio_base_datos_bibliografica
- servicio_estadisticas_uso_base_datos_bibliografica
- servicio_base_datos_bibliografica_journal
- servicio_estadisticas_journal
- servicio_medio_publicacion
- servicio_medio_publicacion_busqueda
- servicio_medio_publicacion_citacion
- servicio_medio_publicacion_publicacion
- servicio_resumen_medios_publicacion
- servicio_analisis_estadistico
- servicio_ley_bradford
- servicio_mineria_datos
- servicio_parametro

4.4.2.2. Diseño del Controlador

El controlador es el encargado de realizar todas las operaciones y está definido por las siguientes clases:

Controladores

Tania Maricela Landívar Ordóñez
Ronaldo Sebastián Rendón Loja



- ArticuloController
- ReferenciaController
- DetalleReferenciaController
- AutorController
- ArticuloAutorController
- AreaFrascatiController
- AreaUnescoController
- EquivalenciaAreaUnescoFrascatiController
- AreaSJRCController
- EquivalenciaAreaUnescoSJRCController
- CategoriasSJRCController
- AreaCategoriaSJRCController
- JournalSJRCController
- CategoriasJournalSJRCController
- JournalController
- BaseDatosBibliograficaController
- EstadisticasUsoBaseDatosBibliograficaController
- BaseDatosBibliograficaJournalController
- EstadisticasJournalController
- MedioPublicacionController
- MedioPublicacionBusquedaController
- MedioPublicacionCitacionController
- MedioPublicacionPublicacionController
- ResumenMedioPublicacionController
- DatosAnalisisEstadisticoController
- LeyBradfordController
- MineríaDatosController
- ParametroController

4.4.2.3. Diseño de la Persistencia

En el caso de la persistencia se detallan los modelos que van a interactuar directamente con la base de datos, siendo los mismos:

Modelos

- Articulo
- Referencia
- DetalleReferencia
- Autor
- ArticuloAutor
- AreaFrascati



- AreaUnesco
- AreaSJR
- EquivalenciaAreaUnescoSJR
- EquivalenciaAreaUnescoFrascati
- CategoriasSJR
- AreaCategoriaSJR
- JournalSJR
- CategoriaJournalSJR
- Journal
- BaseDatosBibliografica
- EstadisticasUsoBaseDatosBibliografica
- BaseDatosBibliograficaJournal
- EstadisticasJournal
- MedioPublicacion
- MedioPublicacionBusqueda
- MedioPublicacionPublicacion
- MedioPublicacionCitacion
- ResumenMediosPublicacion
- Parametro

4.4.2.4. Diseño de la Base de Datos

A continuación, la Figura 12 presenta un modelo conceptual de las entidades consideradas para el diseño de la base de datos, que es un esquema inicial de las entidades a utilizar y sus relaciones partiendo de la entidad principal “artículo”; una vez definido el modelo conceptual se detallan los atributos que caracterizan a cada una de las entidades, así como también se establece las relaciones de cardinalidad entre las mismas, dando como resultado el modelo lógico que contiene los atributos y sus relaciones que se visualiza en la Figura 13.

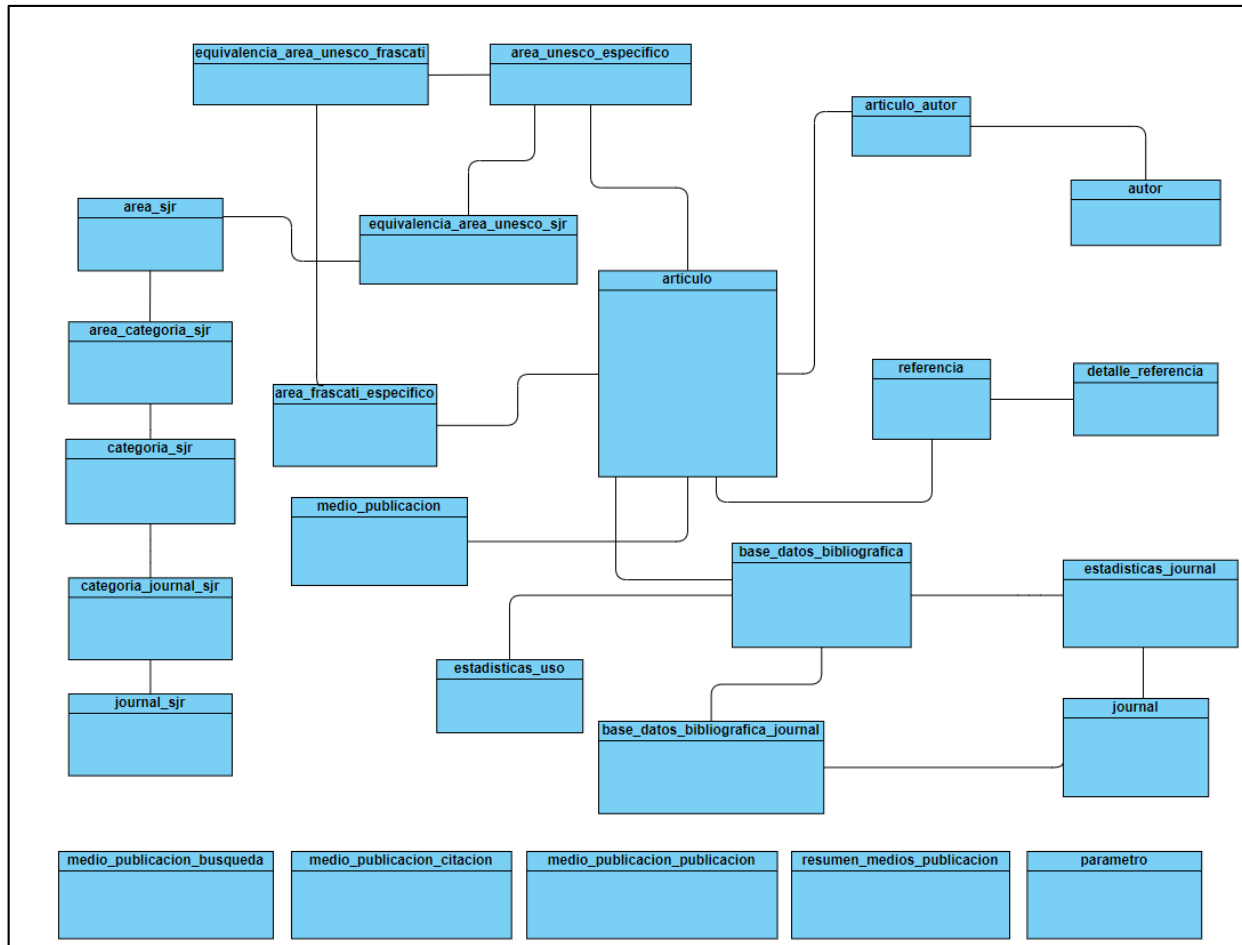


Figura 12. Modelo Conceptual de la Base de Datos del Prototipo.

Tania Maricela Landívar Ordóñez
 Ronaldo Sebastián Rendón Loja

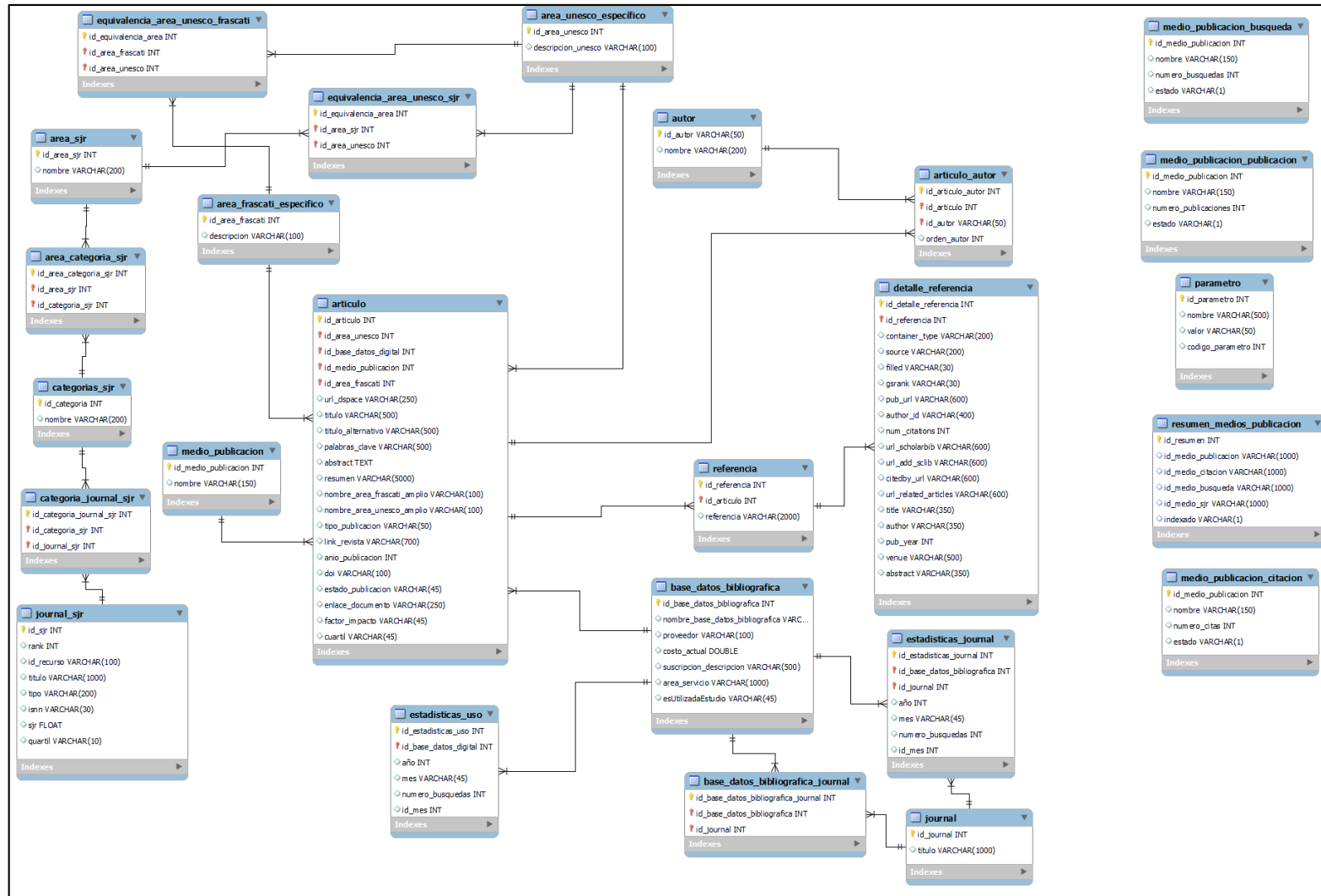


Figura 13. Diagrama entidad – relación (Modelo lógico del Prototipo).

Tania Maricela Landívar Ordóñez
 Ronaldo Sebastián Rendón Loja

4.4.3. Definición de la Interfaz del Usuario

En esta sección se presenta el diseño de las interfaces gráficas con las que el usuario interactúa en el prototipo, las mismas que fueron desarrolladas con la herramienta de diseño Balsamiq. Esta herramienta permite generar bocetos digitales de la idea o concepto para una aplicación web, facilitando la comprensión antes de escribir cualquier código (Balsamiq, 2008). Se especifica el formato individual de cada pantalla con sus respectivos elementos.

4.4.3.1. Especificación de la Interfaz de Pantalla y Navegación

La pantalla principal se compone de un menú de navegación situada en la zona izquierda de la pantalla. La misma se puede visualizar en la Figura 14.

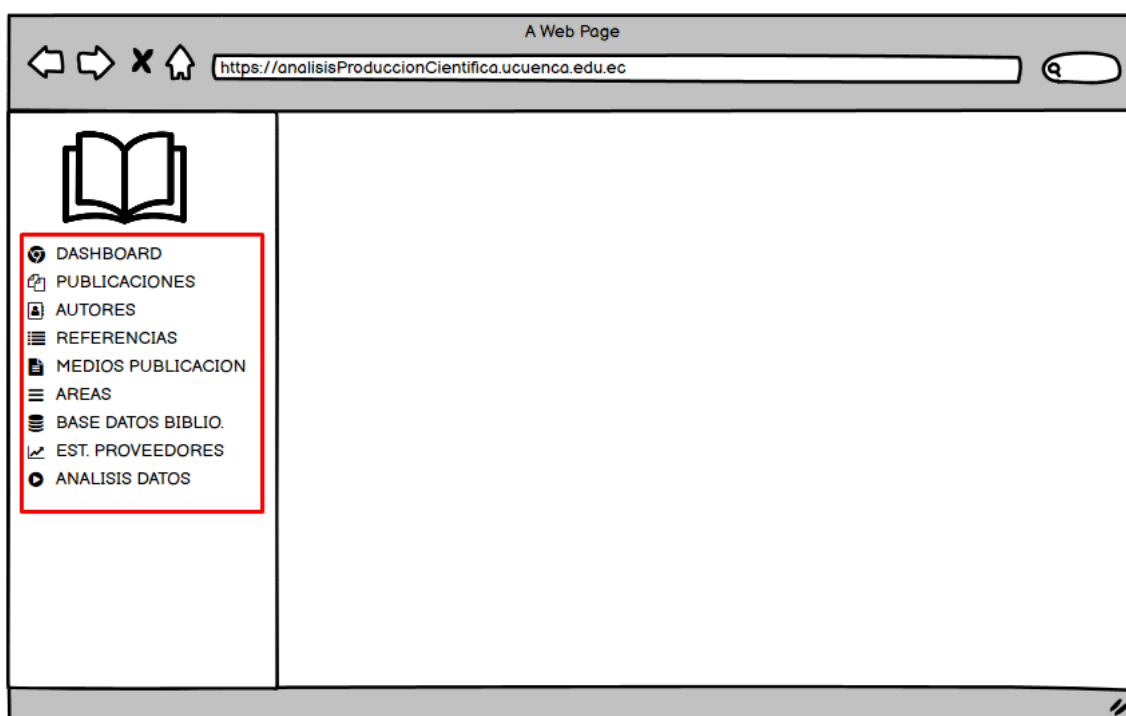
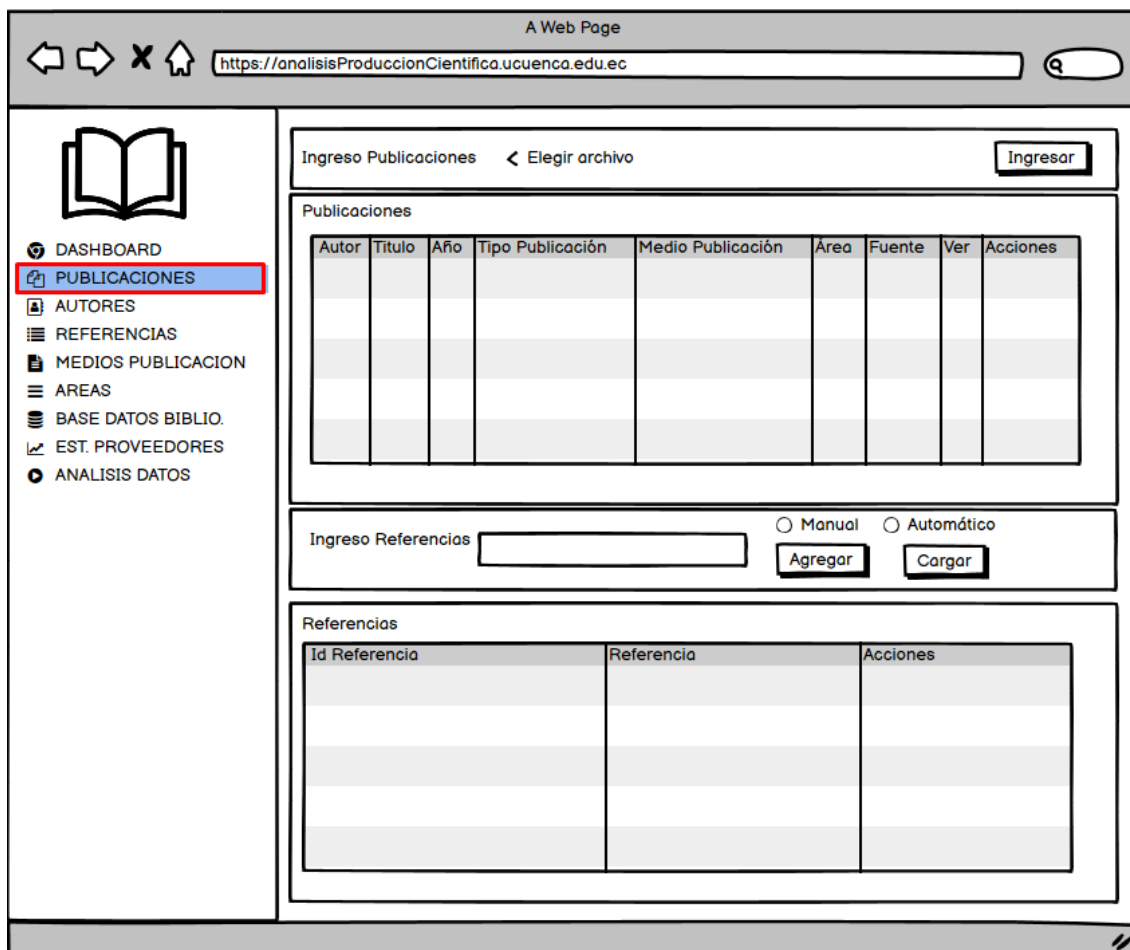


Figura 14. Opciones del menú de navegación de la Pantalla Principal.

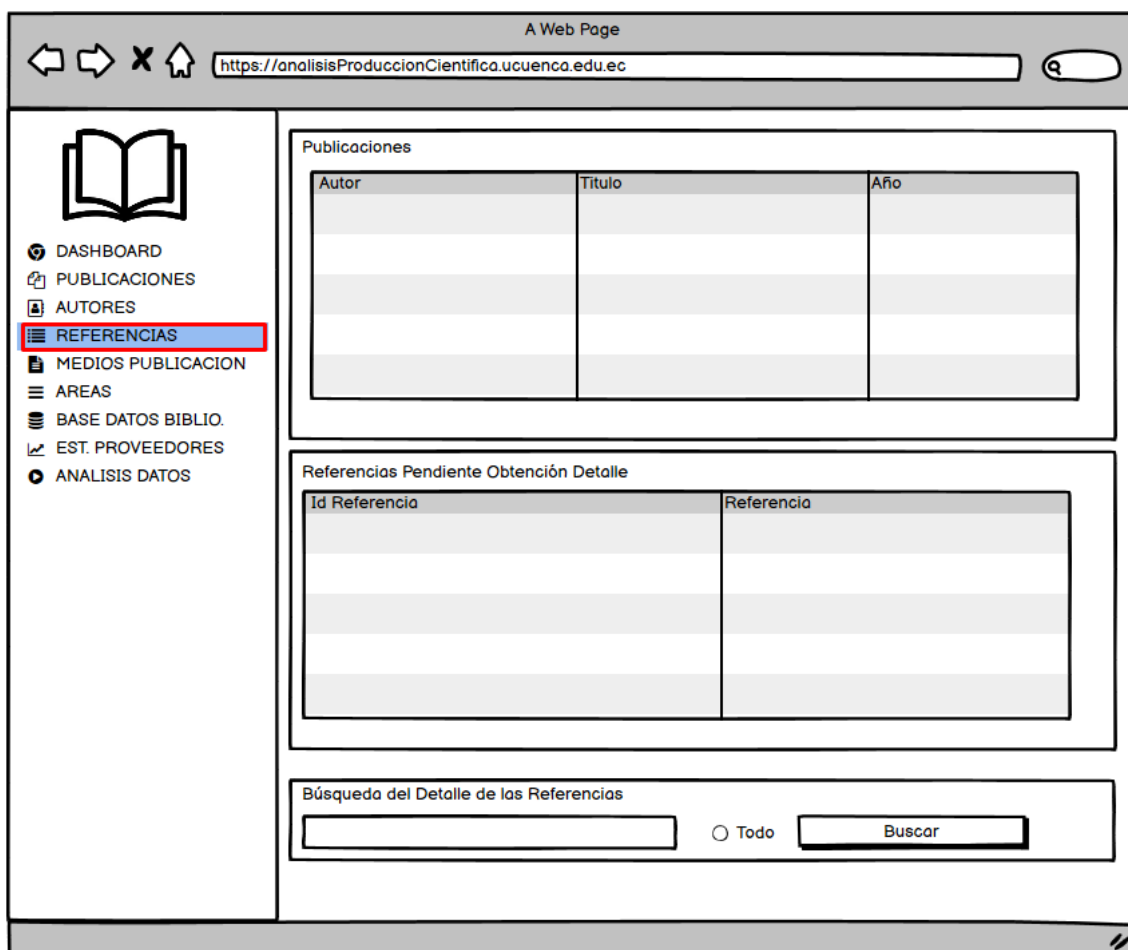
En la Figura 15, la pantalla para la sección de publicaciones presenta un formulario que permite su ingreso a través de un archivo de Excel. Contiene una tabla en donde se listan las publicaciones, permite el ingreso individual de las referencias y lista el detalle de las mismas en una tabla.



The screenshot shows a web browser window with the URL `https:// analisisProduccionCientifica.ucuenca.edu.ec`. The page title is "A Web Page". On the left is a navigation menu with the following items: DASHBOARD, PUBLICACIONES (highlighted in red), AUTORES, REFERENCIAS, MEDIOS PUBLICACION, AREAS, BASE DATOS BIBLIO., EST. PROVEEDORES, and ANALISIS DATOS. The main content area is divided into two sections. The top section is titled "Ingreso Publicaciones" and includes a "Elegir archivo" button and an "Ingresar" button. Below this is a table with the following columns: Autor, Titulo, Año, Tipo Publicación, Medio Publicación, Área, Fuente, Ver, and Acciones. The bottom section is titled "Ingreso Referencias" and includes a text input field, radio buttons for "Manual" and "Automático", and "Agregar" and "Cargar" buttons. Below this is a table with the following columns: Id Referencia, Referencia, and Acciones.

Figura 15. Pantalla del ingreso del detalle de las publicaciones y sus respectivas referencias.

Se puede visualizar en la Figura 16, la sección de Referencias, en la cual se listan las publicaciones pendientes de extraer el detalle de las referencias a través de la API de Scopus y Google Académico. El objetivo es extraer metadatos como el nombre de la revista o medio de publicación donde se encuentra la publicación referenciada.



A Web Page

https:// analisisProduccionCientifica.ucuenca.edu.ec

REFERENCIAS

DASHBOARD
PUBLICACIONES
AUTORES
REFERENCIAS
MEDIOS PUBLICACION
AREAS
BASE DATOS BIBLIO.
EST. PROVEEDORES
ANALISIS DATOS

Publicaciones

Autor	Titulo	Año

Referencias Pendiente Obtención Detalle

Id Referencia	Referencia

Búsqueda del Detalle de las Referencias

Todo

Figura 16. Pantalla para la extracción del detalle de las referencias por cada publicación.

La pantalla para el manejo de las estadísticas de uso de las bases de datos bibliográficas proporcionadas por los proveedores se puede visualizar en la Figura 17. Esta pantalla permite el ingreso del número de búsquedas por año y mes, lista las estadísticas de búsqueda por bases de datos en una tabla y adicionalmente presenta un gráfico estadístico con la información.

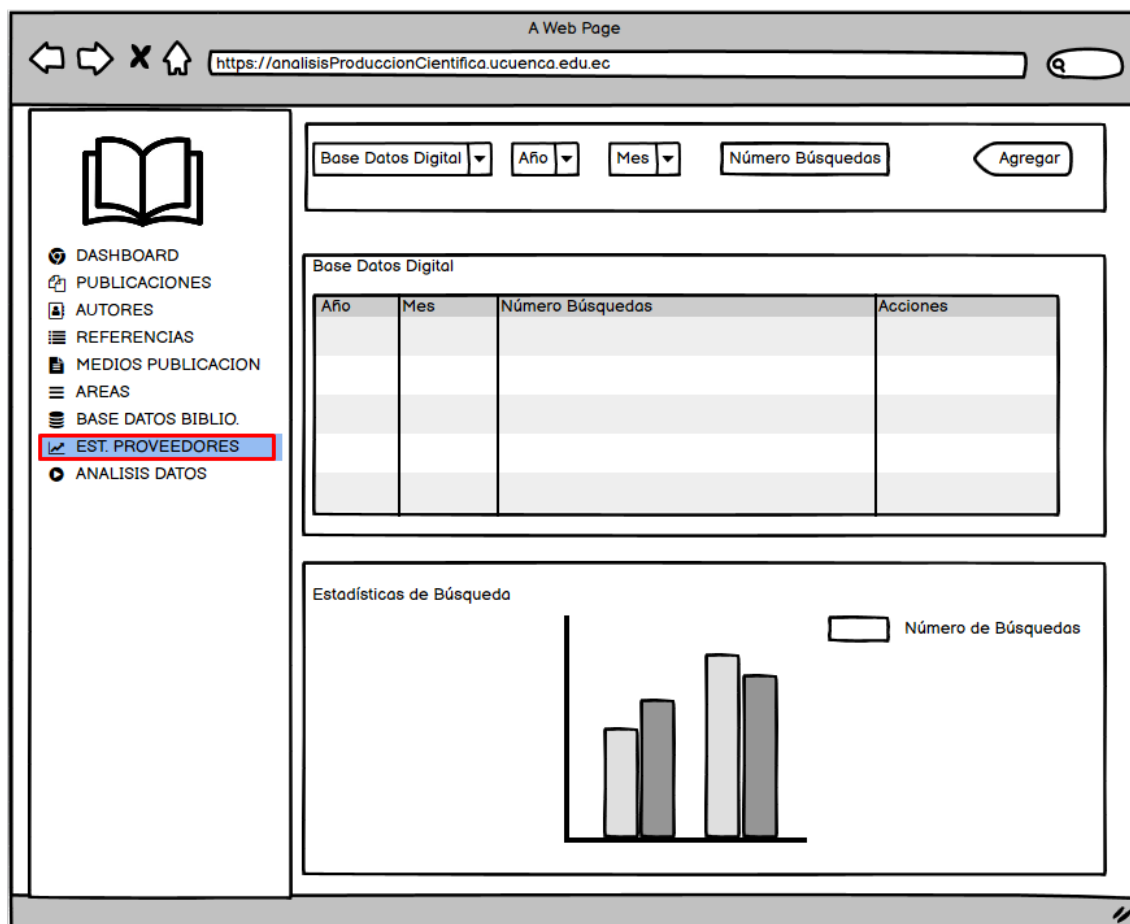


Figura 17. Pantalla para el ingreso de estadísticas de uso proporcionadas por los proveedores.

Finalmente, la pantalla para analizar los datos puede visualizarse en la Figura 18, en la cual se pueden ejecutar operaciones de minería de datos y análisis estadístico, por ejemplo, visualizar las áreas con el mayor número de publicaciones.

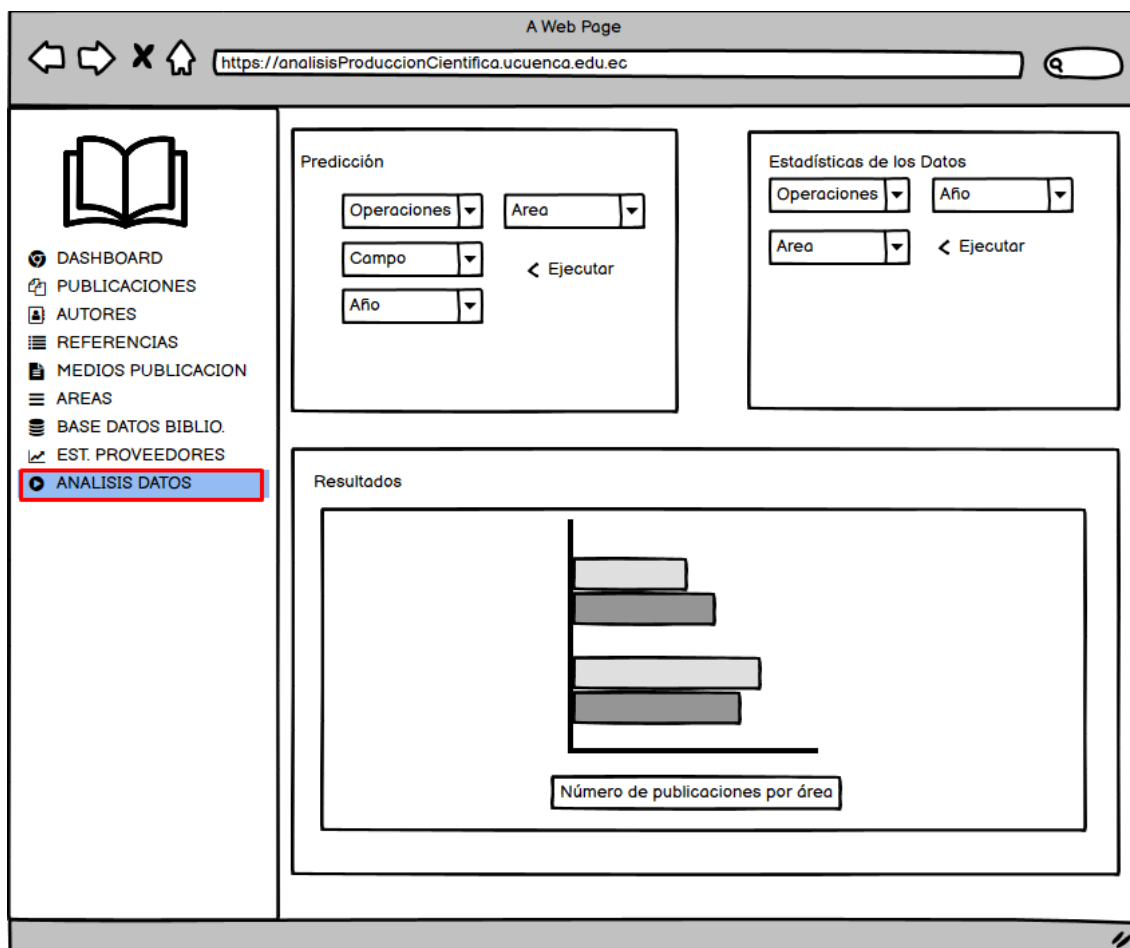


Figura 18. Pantalla para el análisis de datos a través de un análisis estadístico y técnicas de minería de datos.

Los diseños de las demás pantallas se pueden visualizar en Anexos sección Diseño del Prototipo.

4.5. Desarrollo e implementación del prototipo

Para la implementación del prototipo se utilizó la arquitectura cliente-servidor (Front-end y Back-end), Modelo Vista Controlador (MVC), servicios RESTful y algunas tecnologías para desarrollo web, las mismas que están detalladas en el Manual Técnico sección de Anexos.

4.5.1. Prototipo implementado

El prototipo tiene un menú con once opciones desde el ingreso y mantenimiento de las publicaciones y referencias hasta el análisis de datos y configuración general del prototipo de sistema.

4.5.1.1. Sección Publicaciones

En la Figura 19 se puede visualizar la pantalla para el mantenimiento y visualización de las publicaciones. En esta sección se ingresan los artículos de los investigadores con filiación a la Universidad de Cuenca indexadas en Scopus y Latindex en el periodo de 2016-2020. El ingreso se hace a través de un archivo de Excel con los siguientes campos: URL de la publicación, título, título alternativo, palabras clave, abstract, resumen, nombre área Frascati amplio, nombre área Frascati específico, nombre área Unesco amplio, nombre área Unesco específico, tipo publicación, año publicación, nombre medio publicación, link de la revista, DOI, nombre de la base de datos bibliográfica, estado de la publicación, enlace del documento, factor de impacto, cuartil, identificación del autor, orden del autor, nombre del autor y nombre de la filiación.

Figura 19. Pantalla para el mantenimiento y visualización de las publicaciones.

Una vez ingresada la publicación se puede editar el detalle de misma, visualizar las referencias, el detalle de las referencias y eliminar la publicación. Opciones que se pueden visualizar en las Figuras 20 y 21.

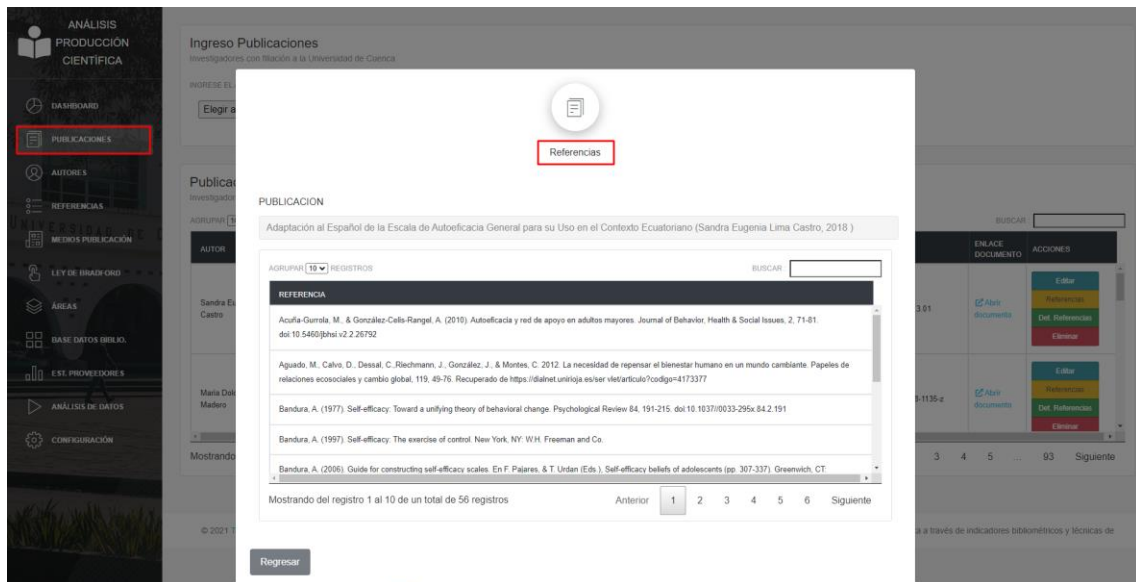


Figura 20. Visualización de las referencias en texto por cada publicación.

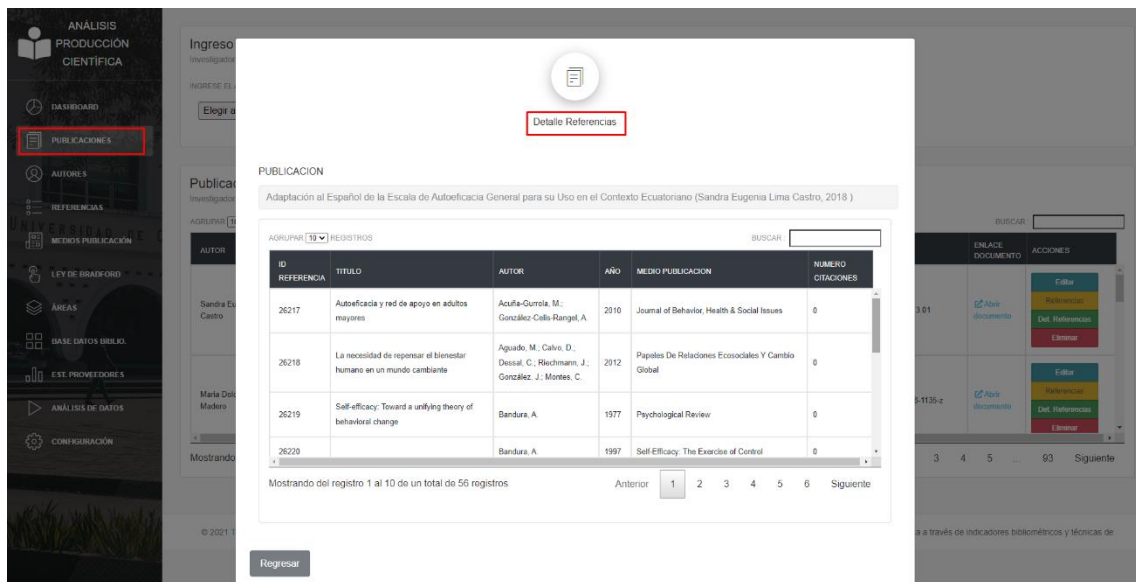


Figura 21. Visualización del detalle de las referencias separada en campos por cada publicación.

Los servicios web utilizados para la funcionalidad de esta pantalla se describen en la Tabla 7.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/articulo/insertar	Ingresar Publicaciones	POST	Publicaciones: Archivo Excel	Notificación del proceso



http://localhost:5000/articulo/listar	Lista Publicaciones	GET	Sin Parámetros	Lista <Publicaciones>
http://localhost:5000/articulo/eliminar	Elimina una publicación	GET	Id del Artículo	Notificación del proceso
http://localhost:5000/articulo/actualizar	Actualiza campos de una publicación	POST	Autor, título, año publicación, tipo publicación, cuartil.	Notificación del proceso
http://localhost:5000/referencia/listarReferenciasPorIdArticulo	Lista Referencias por Artículo	GET	Id del Artículo	Lista <Referencias>
http://localhost:5000/detalleReferencia/listaDetalleReferenciaPorIdArticulo	Lista Detalle Referencia Por Artículo	GET	Id del Artículo	Lista <Detalle Referencia>

Tabla 7. Servicios Web para la Pantalla Publicaciones.

4.5.1.2. Sección Autores

El detalle de la pantalla de Autores se puede ver en la Figura 22, en donde se observa dos opciones, “Datos” y “Gráfico de Burbuja”. En la primera se ingresan los autores por artículo a través de un archivo de Excel con los siguientes campos: nombre de la base de datos bibliográfica, título, año de publicación, identificación del autor, orden del autor y nombre del autor.

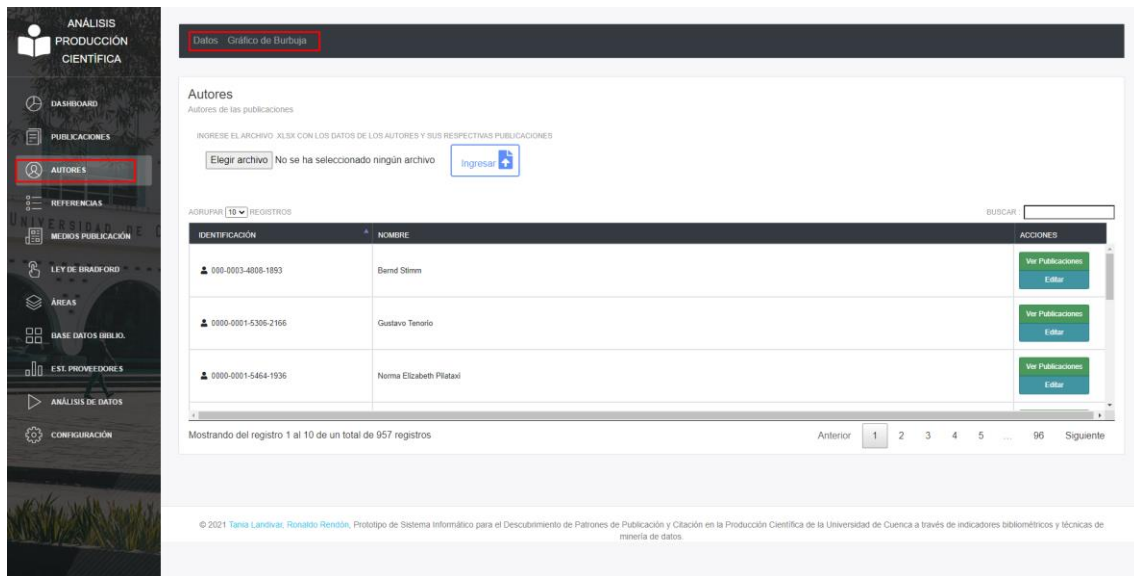


Figura 22. Pantalla para el mantenimiento de los autores y sus publicaciones.

Continuando en esta pantalla, se pueden ver las publicaciones en las que ha participado cada autor perteneciente a la Universidad de Cuenca (Ver la Figura 23).

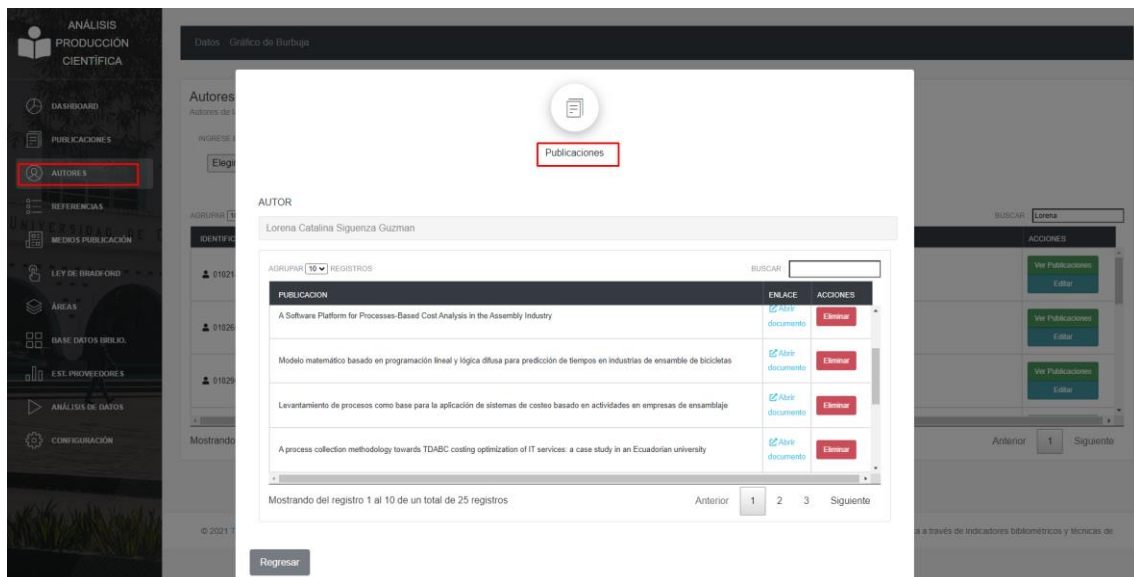


Figura 23. Publicaciones por autor.

En la segunda opción como se observa en las Figuras 24 y 25 se puede visualizar la información de los autores mediante un gráfico. Por ejemplo, el número de publicaciones por autor y el número de publicaciones según el orden del autor en la publicación.

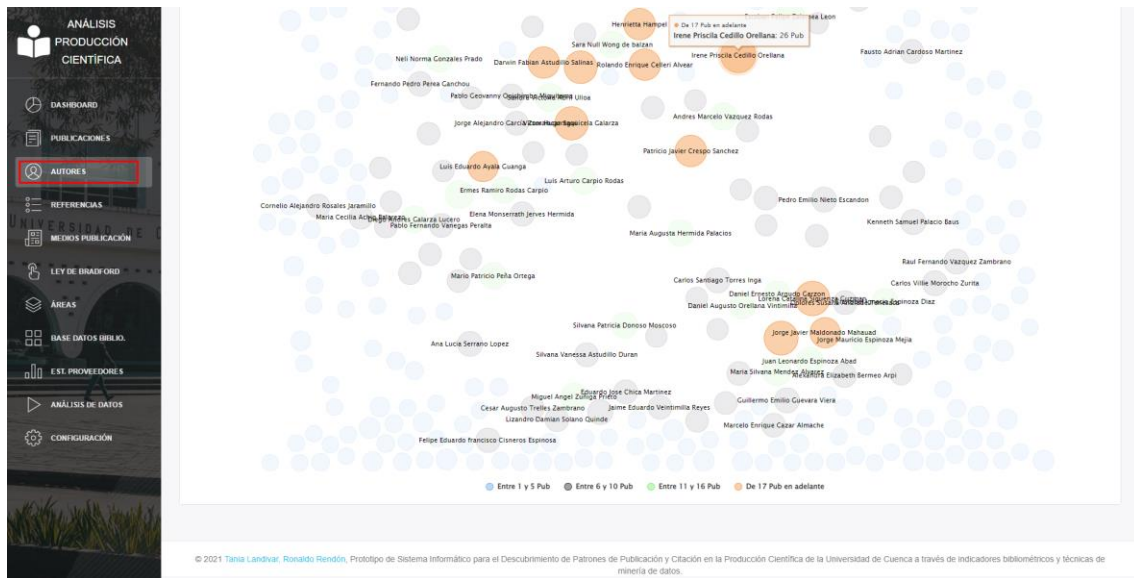


Figura 24. Gráfico de burbuja en función del número de publicaciones por autor.

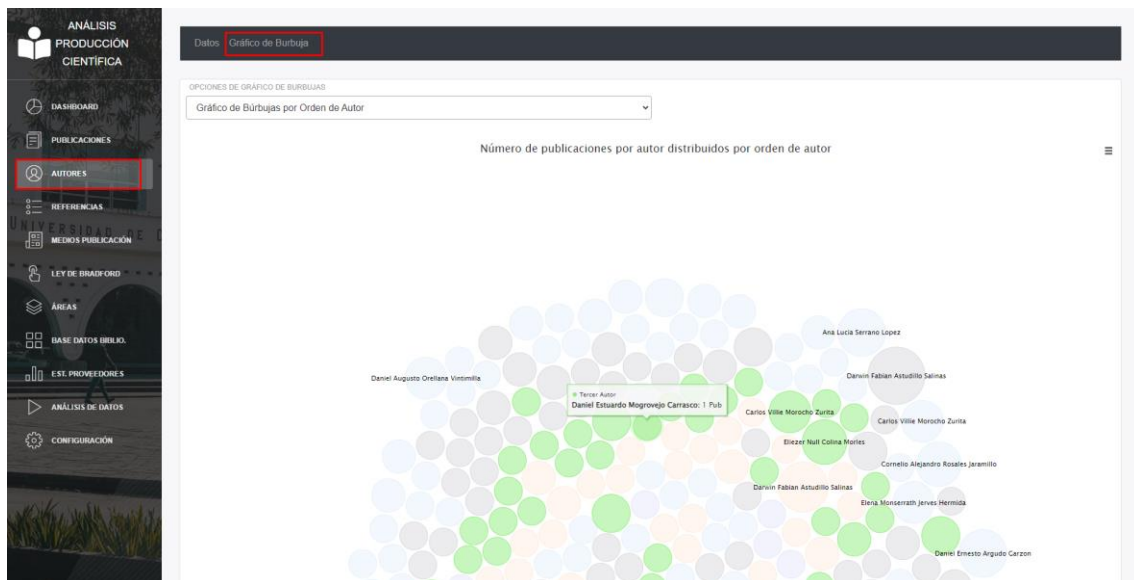


Figura 25. Gráfico de burbuja de acuerdo con el número de publicaciones según el orden del autor en la publicación.

En la Tabla 8 se encuentran los servicios web utilizados para la funcionalidad de la pantalla de Autores:

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/articuloAutor/insertar	Ingresar Autores y la relación con las publicaciones	POST	Autores Publicación: Archivo Excel	Por Notificación del proceso

http://localhost:5000/autor/listar	Listar los Autores	GET	Sin Parámetros	Lista <Autores>
http://localhost:5000/autor/actualizar	Actualiza la información del autor	POST	Id del Autor, Nombre	Notificación del Proceso
http://localhost:5000/articuloAutor/eliminar	Eliminar la relación entre el autor y la publicación.	GET	Id del Artículo Autor	Notificación del proceso
http://localhost:5000/articuloAutor/listarPorIdAutor	Listar las publicaciones por autor.	GET	Id del Autor	Lista<Publicaciones>
http://localhost:5000/autor/ejecutarDatosHighChart	Listar los autores según su orden de autor	GET	Sin Parámetros	Lista <autores por orden de autor>

Tabla 8. Servicios Web para el mantenimiento de los autores.

4.5.1.3. Sección Referencias

La sección de referencias es uno de los enfoques principales para el desarrollo de este prototipo, ya que a través de estas se buscan obtener los medios de publicación (revistas) que más consume la comunidad universitaria en los trabajos de investigación. La pantalla para el mantenimiento de las referencias se puede visualizar en la Figura 26.

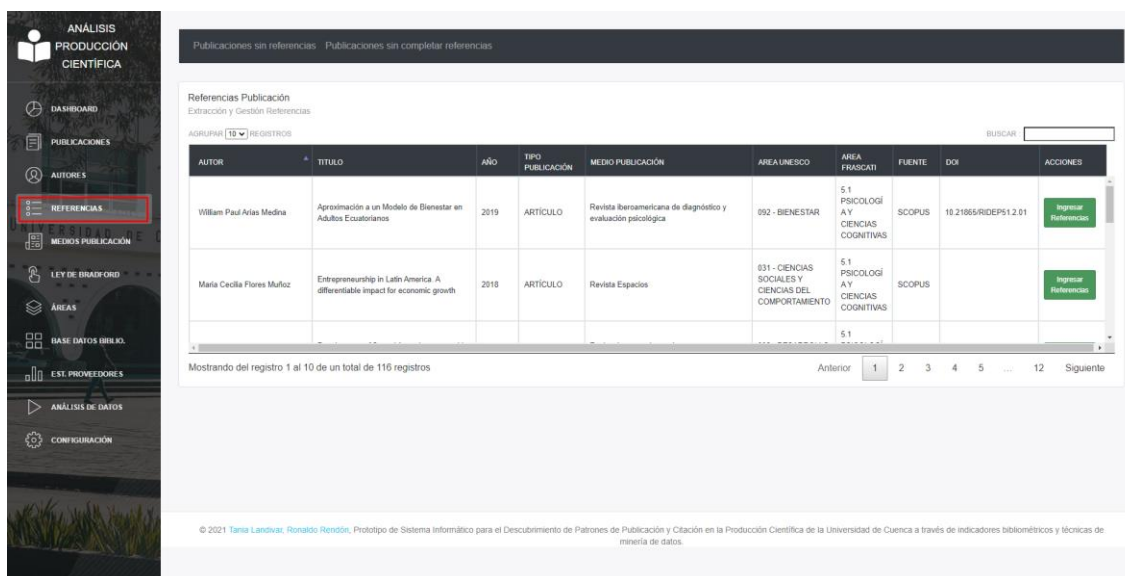


Figura 26. Pantalla para la extracción de las referencias y el detalle de las referencias por publicación.

La primera opción del menú superior visualiza las publicaciones sin referencias. Para las publicaciones indexadas en Scopus de los investigadores con filiación a la Universidad de Cuenca, se utiliza el DOI para obtener las referencias en texto y el detalle de la referencia separada en campos. Para el caso de las publicaciones indexadas en Latindex, se extrae una a una la referencia de forma manual (ver la Figura 27).

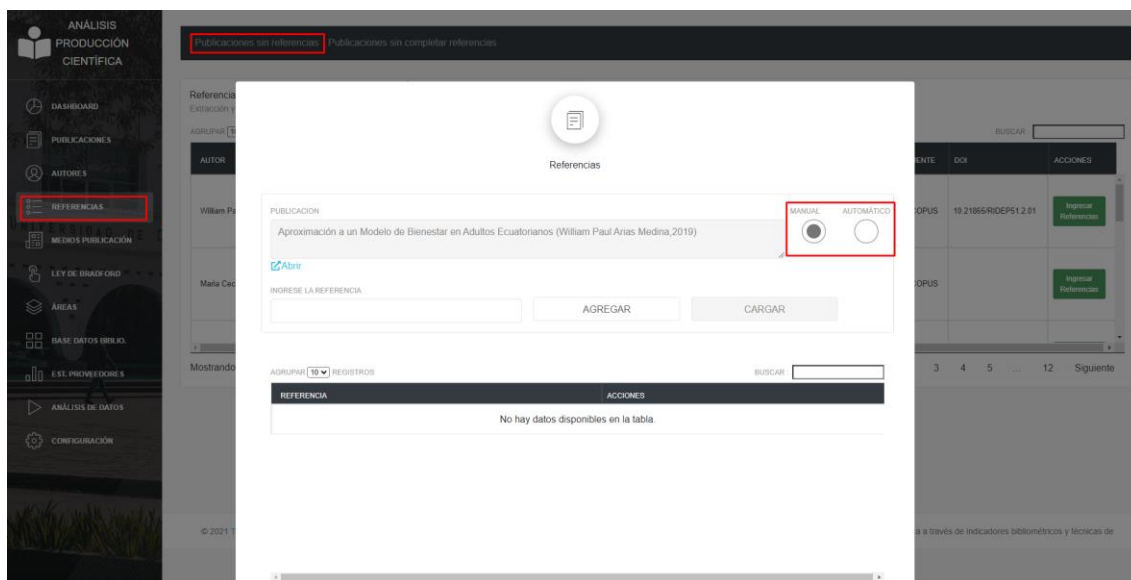


Figura 27. Pantalla para el ingreso de forma manual y automática de las referencias y el detalle de mismas.

En la Figura 28 se presenta el proceso para complementar la extracción del detalle de las referencias, como en el caso de Latindex que no cuenta con una herramienta que permita extraer esta información. Primero, se ingresan las referencias una a una en la sección de publicaciones sin referencias; luego, en la sección de publicaciones sin completar referencias, se busca el detalle de las mismas con la API de Google Académico.

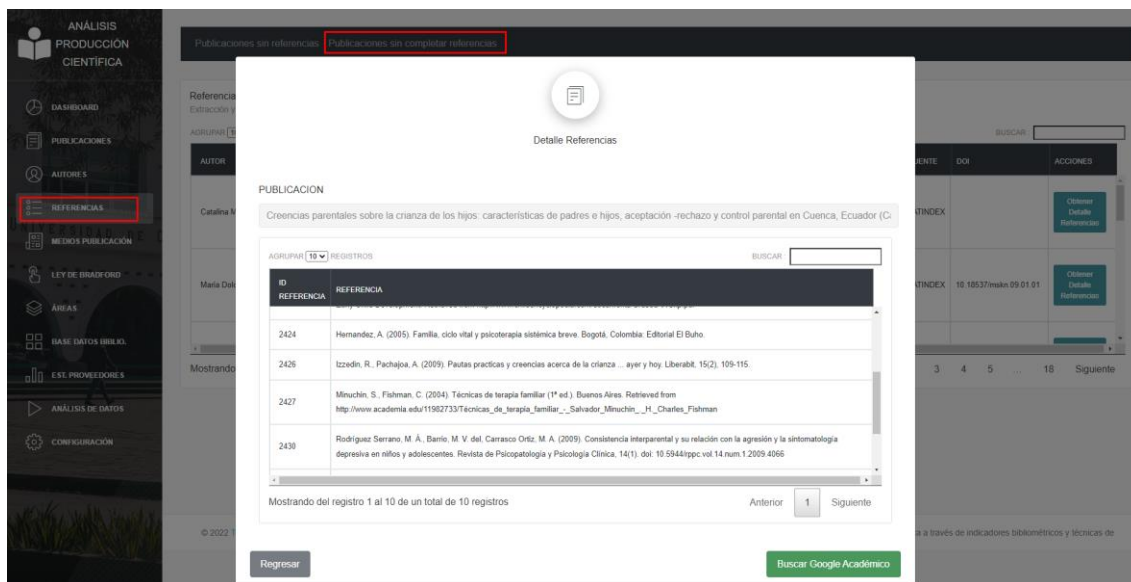


Figura 28. Pantalla para la extracción del detalle de la referencia a través de la API de Google Académico.

A continuación, en la Tabla 9 se describen los servicios web para la operatividad de la pantalla de Referencias.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/articulo/listarArticulosSinReferencias	Lista Publicaciones Sin Referencias	GET	Sin Parámetros	Lista<Publicaciones>
http://localhost:5000/articulo/listarArticulosSinCompletarReferencias	Lista Publicaciones Sin Completar Referencias	GET	Sin parámetros	Lista<Publicaciones>
http://localhost:5000/referencia/insertarManual	Inserta Referencia de Forma Manual	POST	Id del artículo, Referencia	Notificación del Proceso
http://localhost:5000/referencia/listarReferenciasPorIdArticulo	Lista Referencias por Artículo	GET	Id del Artículo	Lista <Referencias>
http://localhost:5000/referencia/insertarAutomaticoScopus	Inserta Referencia y Detalle Referencia Automático con el uso de la API de Scopus	POST	Id del Artículo, Id de la Base de Datos	Notificación del Proceso



			Bibliográfica	
http://localhost:5000/referencia/listarReferenciasNoEncontradasPorIdArticulo	Lista las Referencias Pendientes por Id del Artículo	GET	Id del Artículo	Lista <Referencias>
http://localhost:5000/referencia/obtenerDetalleReferenciaTotal	Obtiene el Detalle de la Referencia utilizando la API Google Académico	POST	Id del Artículo	Notificación del proceso
Servicio Web Externo				
Scopus.py	Busca las referencias y el detalle de las referencias de la publicación indexada en Scopus		DOI de la publicación	references:{ fulltext,title, authors, sourcetitle, publicationyear, citedbycount}
Scholar.py (API)	Busca el Detalle de la Publicación por la Referencia en Texto		Referencia	{author_id, bib: {abstract, author, pub_year, title, objects,venue}, citedby_url, eprint_url, filled, gsrnk, num_citations, pub_url, source,url_add_sclib, url_scholarbib}

Tabla 9. Servicios Web para la extracción y almacenamiento de las referencias y detalle de las referencias.

4.5.1.4. Sección Estadísticas Proveedores

En esta sección se tiene como referencia las estadísticas proporcionadas por los proveedores del CDRJBV, que permite ingresar las estadísticas de uso de las bases de datos bibliográficas, o en su defecto de una revista o medio de publicación que corresponde a estas bases de datos. Además, en la opción de

gráfico de la sección se puede visualizar una gráfica de las estadísticas de búsqueda por mes y año de una determinada base de datos bibliográfica o revista, como se observa en las Figuras 29, 30 y 31.

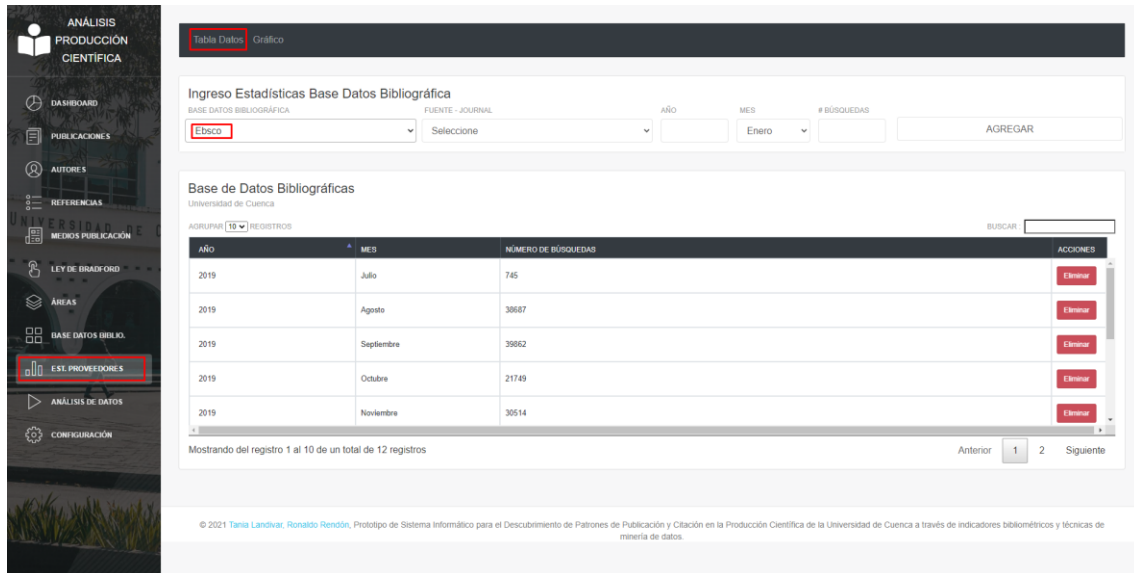


Figura 29. Pantalla para el ingreso de las estadísticas de uso de las bases de datos bibliográficas y sus revistas o fuentes bibliográficas.

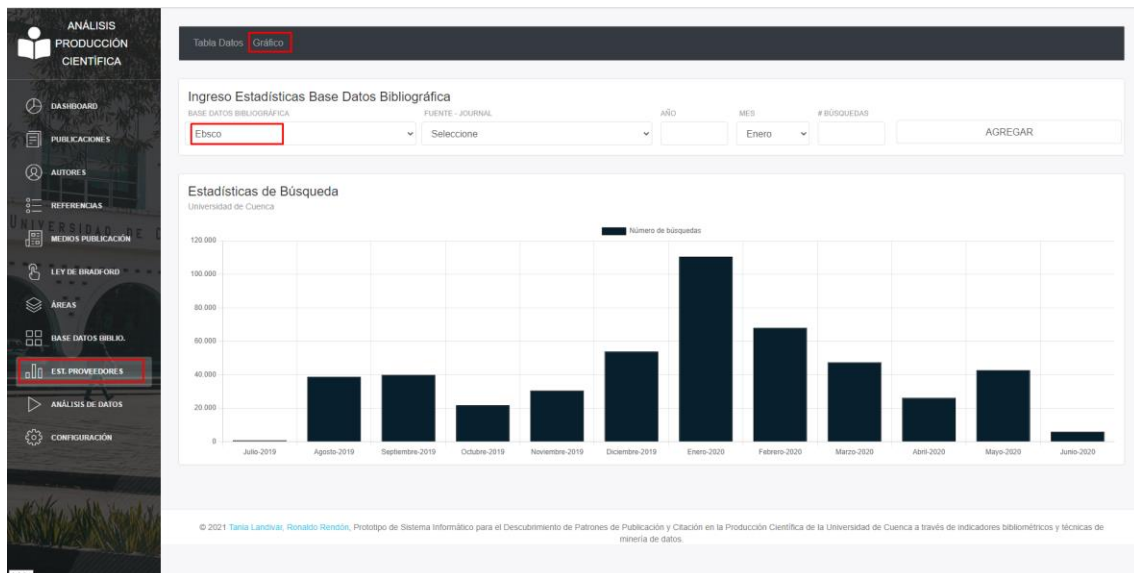


Figura 30. Gráfico de barras de la estadística de búsqueda de las bases de datos bibliográficas.

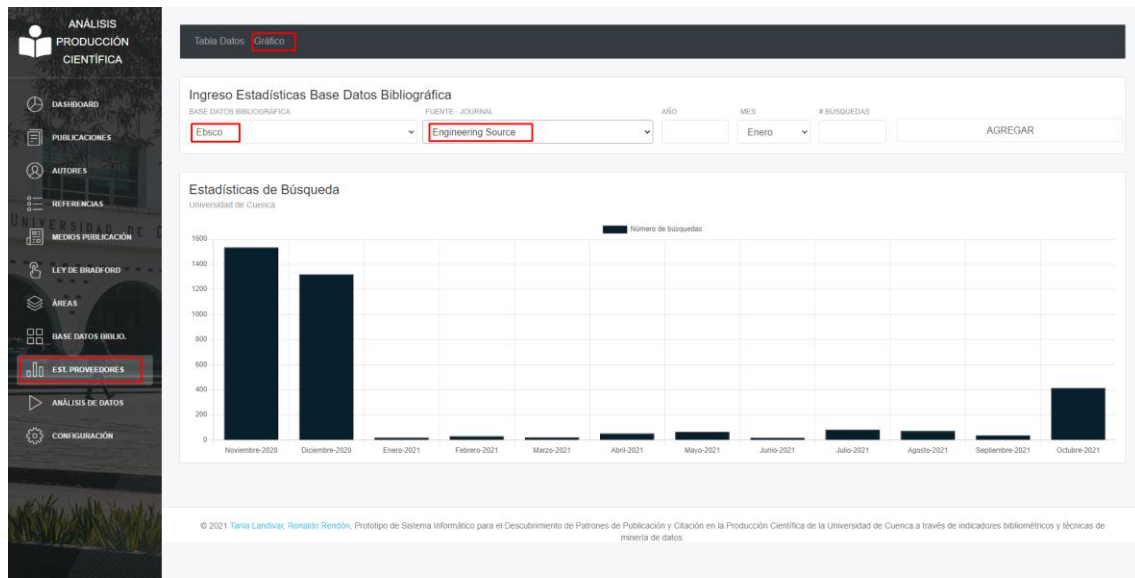


Figura 31. Gráfico de barras de la estadística de búsqueda de las fuentes bibliográficas indexadas en las bases de datos.

En la Tabla 10 se presentan los servicios web empleados para el mantenimiento de las estadísticas de uso de las bases de datos bibliográficas y sus revistas o fuentes bibliográficas.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/baseDatosBibliografica/listar	Lista base de datos bibliográfica	GET	Sin Parámetros	Lista <Bases de datos bibliográfica >
http://localhost:5000/journa/listarPorBaseDatosBibliografica	Lista las revistas por base de datos bibliográfica	GET	Id de la Base de Datos Bibliográfica	Lista<Revistas>
http://localhost:5000/estadisticasUso/insertar	Inserta estadísticas de uso por base de datos bibliográfica	POST	Id Base de datos bibliográfica, año, id mes, número de búsquedas	Notificación del Proceso
http://localhost:5000/estadisticasJournal/insertar	Insertar estadísticas de búsqueda a las revistas correspondientes a cada base de datos bibliográfica	POST	Id de la base de datos bibliográfica, id de la revista, año, id mes,	Notificación del Proceso



			número de búsquedas	
http://localhost:5000/estadisticasUso/buscarEstadisticasUsoPorId	Busca las estadísticas de uso por base de datos bibliográfica	POST	Id de la base de datos bibliográfica	Lista <Estadísticas de Uso por Base de Datos Bibliográfica >
http://localhost:5000/estadisticasJournal/buscarEstadisticasJournalPorId	Lista las estadísticas de uso de la revista	POST	Id de la base de datos bibliográfica, Id de la revista	Lista<Estadísticas de Uso de la Revista>
http://localhost:5000/estadisticasUso/eliminar	Elimina la estadística de uso de la base de datos bibliográfica	GET	Id de la estadística de uso	Notificación del proceso
http://localhost:5000/estadisticasJournal/eliminar	Elimina la estadística de uso de las revistas de las bases de datos bibliográficas	GET	Id de la estadística de uso	Notificación del proceso

Tabla 10. Servicios web para el ingreso y mantenimiento de las estadísticas de uso de las bases de datos bibliográficas y sus fuentes.

4.5.1.5. Sección Análisis Datos

Dentro de esta sección se hace referencia a toda la información que contiene cada una de las publicaciones, con el fin de poder visualizar el tratamiento de los datos y análisis estadístico. Es por lo que, en esta sección, se encuentran dos opciones, Estadísticas y Predicción de los datos las cuales se pueden ver en la Figura 32.

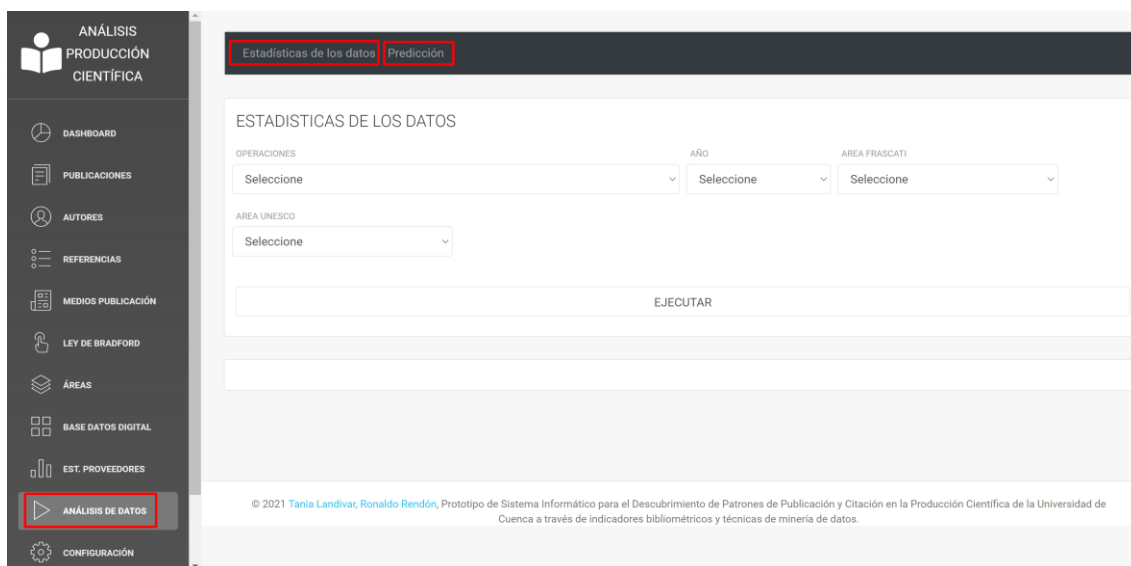


Figura 32. Sección análisis de datos

En la Tabla 11 se presentan los servicios web empleados para el mantenimiento de la sección análisis de datos.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/ analisisEstadistico/n umeroPublicaciones PorAreaUnesco	Lista número de publicaciones por área Unesco	GET	Sin Parámetros	Lista <número de publicaciones por área Unesco >
http://localhost:5000/ analisisEstadistico/ numeroPublicacione sPorAreaFrascati	Lista número de publicaciones por área Frascati	GET	Sin Parámetros	Lista <número de publicaciones por área Frascati >
http://localhost:5000/ areaFrascati/listar	Lista las áreas pertenecientes a Frascati	GET	Sin Parámetros	Lista <área Frascati >
http://localhost:5000/ areaUnesco/listar	Lista las áreas pertenecientes a Unesco	GET	Sin Parámetros	Lista <área Unesco>
http://localhost:5000/ prediccion/ejecutarP rediccion	Lista de valores de la predicción a 6 meses	GET	Sin Parámetros	Lista <valor de predicción del número de búsquedas>

Tabla 11. Servicios web para la sección análisis de datos

4.5.1.6. Sección Ley de Bradford

En esta sección se realiza un análisis de las revistas o fuentes bibliográficas, de acuerdo con los siguientes indicadores bibliométricos, la más publicada, la más citada, con el mayor número de búsquedas y factor de impacto. Se evalúa a la revista en base a los indicadores antes mencionados y se le asigna un peso a cada uno de ellos, para finalmente calificar desde diferentes puntos de vista a la revista. Adicionalmente, se establece si la revista que tiene mayor calificación se encuentra indexada en los repositorios bibliográficos ya contratados por la Universidad de Cuenca. Dejando a consideración si los recursos contratados están cubriendo las necesidades de información de los usuarios.

La sección de la Ley de Bradford permite realizar el análisis de acuerdo con el área Unesco o área Frascati en un periodo de tiempo, ver la Figura 33. Se presenta en una tabla los indicadores bibliométricos con sus pesos y resultados finales, evaluado en su totalidad a la revista (ver la Figura 34).

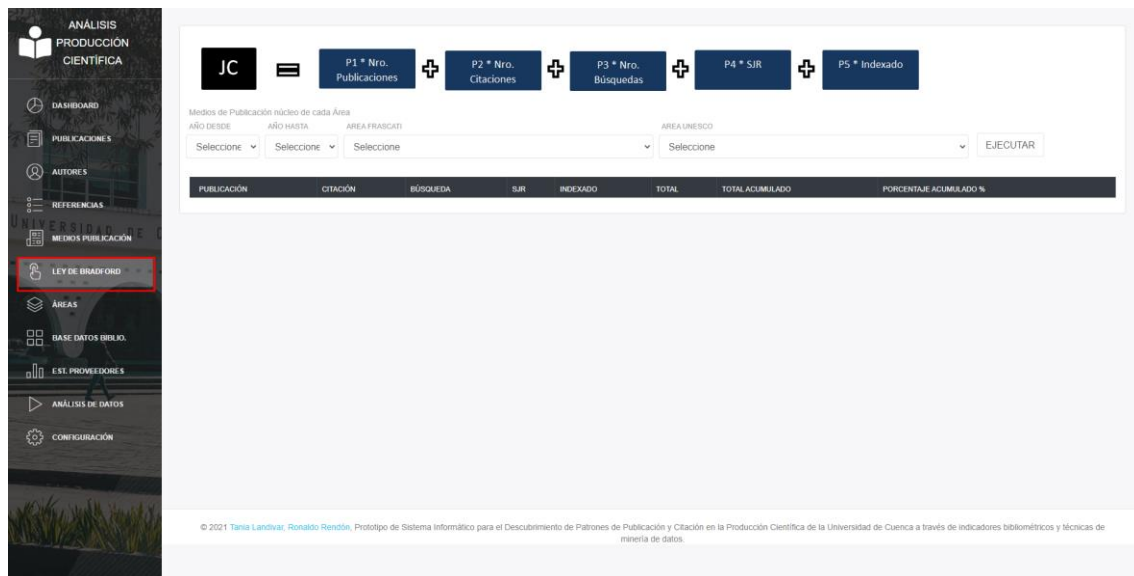


Figura 33. Sección para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto.

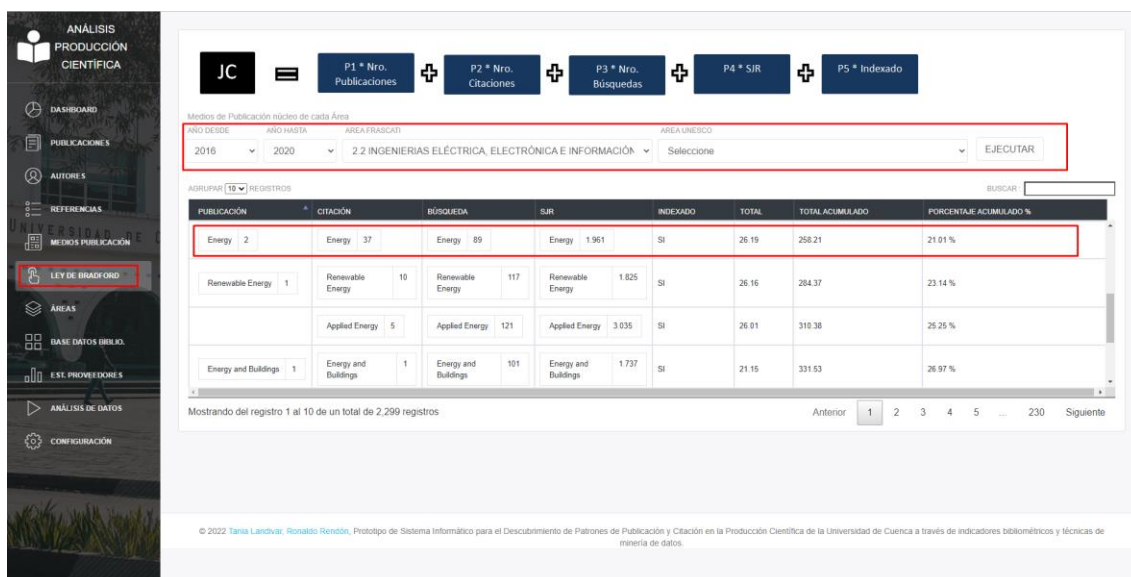


Figura 34. *Tabla de resultados para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto.*

En la Tabla 12 se presentan los servicios web empleados para la evaluación de las revistas en la sección de la ley de Bradford.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/medioPublicacionBúsqueda/conteoMediosPublicacionBúsqueda	Cuenta el número de búsquedas realizadas por revista y las ordena descendientemente	GET	Sin Parámetros	Notificación del Proceso
http://localhost:5000/medioPublicacionCita/conteoMediosPublicacionCita	Cuenta el número de citas realizadas a las revistas y las ordena descendientemente	GET	Sin parámetros	Notificación del proceso
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacion	Cuenta el número de publicaciones realizadas en las revistas y las ordena descendientemente	GET	Sin parámetros	Notificación del proceso
http://localhost:5000/medioPublicacionCita/conteoMediosPublicacionCita	Cuenta el número de citas realizadas a las revistas en un periodo de tiempo y las ordena	GET	Año desde, año hasta	Notificación del Proceso



PorAnio	descendentemente			
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAnio	Cuenta el número de publicaciones realizadas en las revistas en un periodo de tiempo y las ordena descendentemente	GET	Año desde, año hasta	Notificación del Proceso
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaFrascati	Cuenta el número de citas realizadas a las revistas por área Frascati y las ordena descendentemente	GET	Id área Frascati	Notificación del proceso
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaFrascati	Cuenta el número de publicaciones realizadas en las revistas por área Frascati y las ordena descendentemente	GET	Id área Frascati	Notificación del proceso
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaUnesco	Cuenta el número de citas realizadas a las revistas por área Unesco y las ordena descendentemente	GET	Id área Unesco	Notificación del proceso
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaUnesco	Cuenta el número de publicaciones realizadas en las revistas por área Unesco y las ordena descendentemente	GET	Id área Unesco	Notificación del proceso
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaFrascatiPorAnio	Cuenta el número de citas realizadas a las revistas por área Frascati en un periodo determinado y las ordena descendentemente	GET	Id área Frascati, año desde, año hasta	Notificación del proceso
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaFrascati	Cuenta el número de publicaciones realizadas en las revistas por área Frascati en un periodo	GET	Id área Frascati, año desde, año	Notificación del proceso



acionPorAreaFrascatiPorAnio	determinado y las ordena descendentemente		hasta	
http://localhost:5000/medioPublicacionCita cion/conteoMediosPublicacionCita cionPorAreaUnescoPorA nio	Cuenta el número de citas realizadas a las revistas por área Unesco en un periodo determinado y las ordena descendentemente	GET	Id área Unesco, año desde, año hasta	Notificación del proceso
http://localhost:5000/medioPublicacionPu blicacion/conteoMedi osPublicacionPublic acionPorAreaUnesc oPorAnio	Cuenta el número de publicaciones realizadas en las revistas por área Unesco en un periodo determinado y las ordena descendentemente	GET	Id área Unesco, año desde, año hasta	Notificación del proceso
http://localhost:5000/ leyBradford/coincide ncia	Obtiene y almacena el número de publicaciones, el número de citas y el número de búsquedas en función de la coincidencia del nombre de la revista	GET	Sin Parámetros	Notificación del proceso
http://localhost:5000/ resumenMediosPubli cacion/listar	Lista el resumen de los medios de publicación, que contiene el número de publicaciones, número de citas, número de búsquedas, factor de impacto y si esta indexada o no en los repositorios bibliográficos.	GET	Sin Parámetros	Lista<Calificaciones de la Revista>

Tabla 12. Servicios web para la evaluación de las revistas a través de la ley de Bradford.

El detalle y funcionalidad de las demás pantallas del prototipo pueden encontrarlas en el manual técnico y el manual de usuario en la Sección de Anexos.



4.6. Pruebas de funcionamiento y Despliegue

Para la validación del prototipo del sistema se contó con el apoyo del personal del CDRJBV, que se mencionan en la Tabla 13:

Cargo	Persona encargada
Coordinadora General CDRJBV	Lic. Rocío Campoverde Carpio
Bibliotecaria del CDRJBV	Lic. Lourdes Naula
Ingeniero de Sistemas Vicerrectorado de Investigación	Ing. Juan José Sáenz

Tabla 13. Personal del CDRJBV de la Universidad de Cuenca que revisaron el prototipo de sistema

El prototipo se despliega de manera local y uno de los requerimientos para su funcionamiento es estar conectado a la red de la Universidad para el consumo de las APIs. Por lo que, para desplegar en un dominio propio de la Universidad de Cuenca, el prototipo debe pasar una serie de procesos y directrices de desarrollo.



Capítulo 5: Resultados y Discusión

Para responder tanto a los objetivos específicos como a las preguntas de investigación se analizaron los datos obtenidos a través de las diferentes secciones antes mencionadas.

Respecto al primer objetivo se realizó el análisis de requerimientos donde inicialmente se evaluó la situación actual de cómo se maneja el proceso de adquisición y suscripción de las bases de datos bibliográficas, luego se planteó las oportunidades de mejora considerando el modelo propuesto por Siguenza-Guzman et al. (2013). Lo que permitió definir los requisitos funcionales del prototipo de sistema y proponer una fórmula (2) que evalúa las revistas que más se destacan en número de publicaciones, número de citas, número de búsquedas y factor de impacto (SJR). Dando como resultado un listado de las fuentes bibliográficas que más estimación tienen en función de los indicadores bibliométricos antes mencionados.

$$JC = (P1 * Nro. Publicaciones) + (P2 * Nro. Citaciones) + (P3 * Nro. Búsquedas) + (P4 * SJR^3) + (P5 * Indexado) \quad (2)$$

En donde:

- JC (Journal Calification): Calificación total del recurso bibliográfico (revista)
- P1: Peso de publicación
- P2: Peso de citación
- P3: Peso de búsqueda
- P4: Peso del factor de impacto
- P5: Peso indexado

Una vez obtenida la lista que puede apreciar las fuentes bibliográficas en las que más se publican, citan, consultan y mayor factor de impacto tienen, a su vez se verifica si la fuente bibliográfica se encuentra indexada en las bases de datos bibliográficas ya contratadas por la Universidad de Cuenca, dejando a la vista si cubre las necesidades de información o qué fuentes pueden adquirirse o suscribirse.

³ SJR: [Scimago Journal & Country Rank](#)



Respecto al segundo objetivo se diseñó el prototipo definiendo el modo de interacción del sistema, seleccionando la arquitectura cliente-servidor en la cual el cliente presenta las opciones con las que contará el usuario y el servidor representa la parte operacional del mismo, es decir ejecuta las peticiones solicitadas por el cliente mediante servicios RESTful. Luego se definió la base de datos que almacena la información capturada y extraída por el sistema, se inició con un modelo conceptual que es una visión inicial de las entidades que interactuarán en el sistema y se concluyó con un modelo lógico que especifica los atributos y relaciones. Finalmente se describió la funcionalidad de las interfaces a través de bocetos realizados en Balsamiq, que representaron las pantallas con las que el usuario trabajará.

El tercer objetivo es el desarrollo e implementación del prototipo, se creó la base de datos en MySQL, luego se procedió a generar los servicios web en Python encargados de dar mantenimiento a las entidades que se definieron en el modelo de la base de datos, además de otras operaciones como la fórmula de la Ley de Bradford, predicción, etc. Las pantallas de interacción para el usuario se desarrollaron en React, mismas que son las encargadas de capturar los datos y armar la estructura para consumir los servicios web y presentar en pantalla los resultados.

A continuación se presentan los resultados obtenidos en las pruebas de funcionamiento, mismas que se encuentran en tres secciones diferentes: Autores, Análisis de datos y Ley de Bradford.

Iniciando con la Sección de Autores, se puede visualizar mediante un gráfico de burbuja de la Figura 35, el número de publicaciones distribuidas según su orden de autor. El primero, último, segundo y tercero son los autores principales del documento, ya que los autores declaran sus contribuciones de forma explícita indicando que el que más contribuye es el primer autor seguido del último autor, el segundo y luego los demás coautores según su colaboración dentro del artículo (Sauermann & Haeussler, 2017). El gráfico de burbuja donde se visualizan los primeros autores se puede ver en la Figura 36. Cabe recalcar que el último autor es importante (ver Figura 37) ya que en ocasiones es considerado el científico más consolidado del grupo, ejerciendo funciones como un director de proyecto (Repiso, 2020).

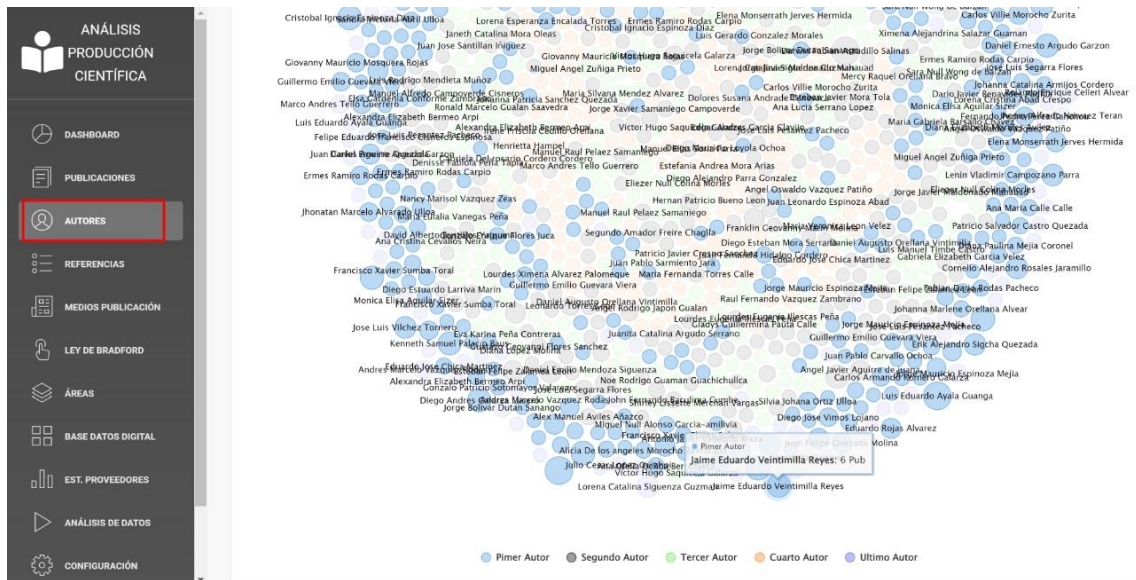


Figura 35. *Número de publicaciones según el orden de autor*

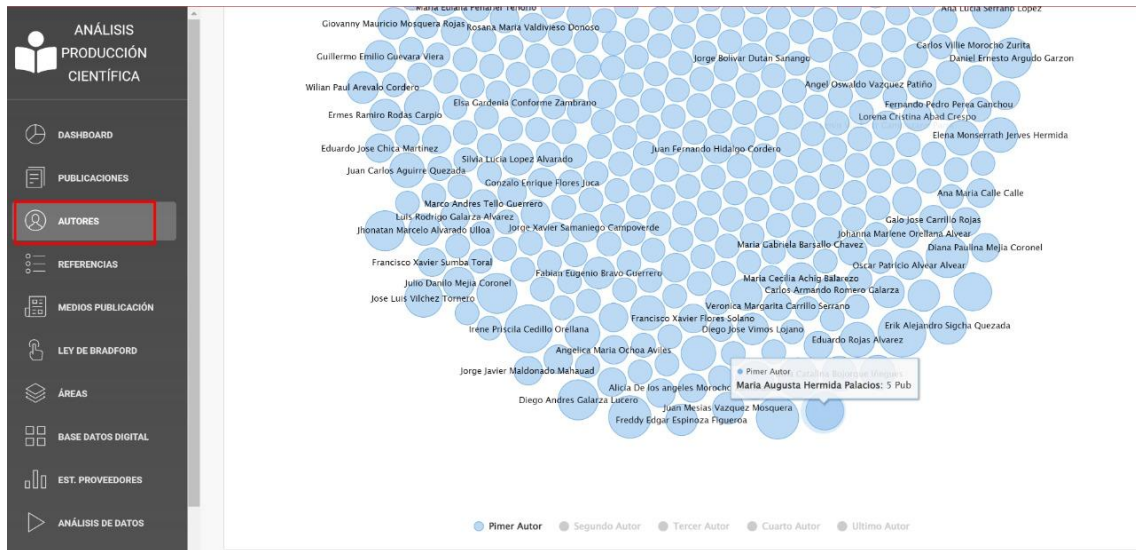


Figura 36. *Número de publicaciones según el primer autor*

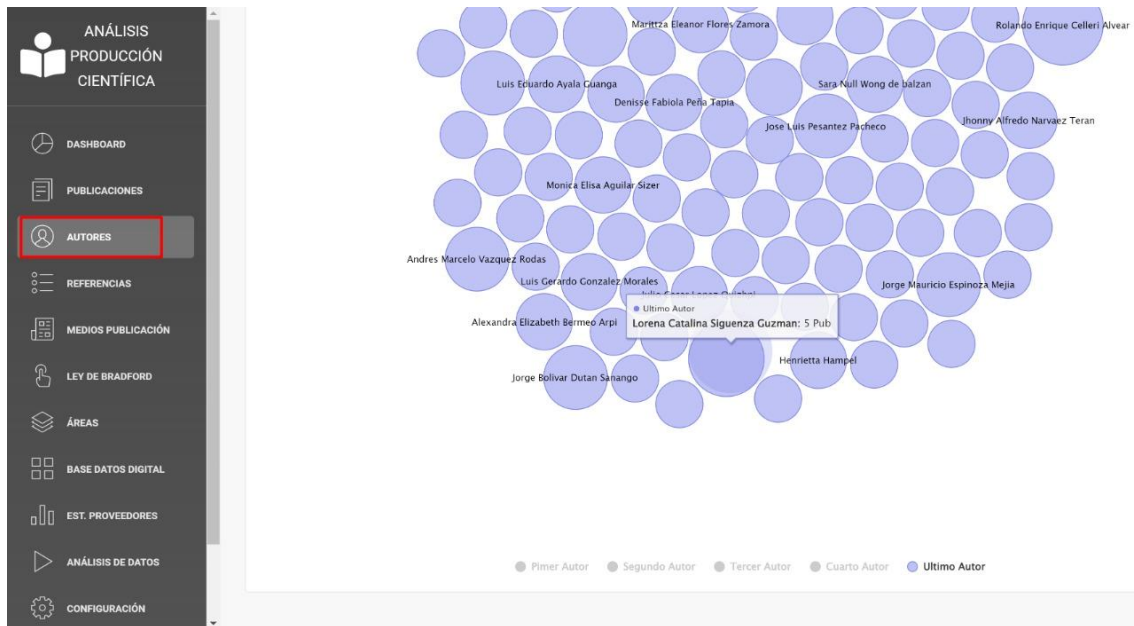


Figura 37. *Número de publicaciones según el último autor*

Continuando con el análisis de autores, se tiene otro gráfico de burbuja que resalta los autores con mayor número de publicaciones, Figura 38. Para el mismo, se definió diferentes rangos; de uno a cinco, de seis a diez, de once a dieciséis y de diecisiete en adelante.

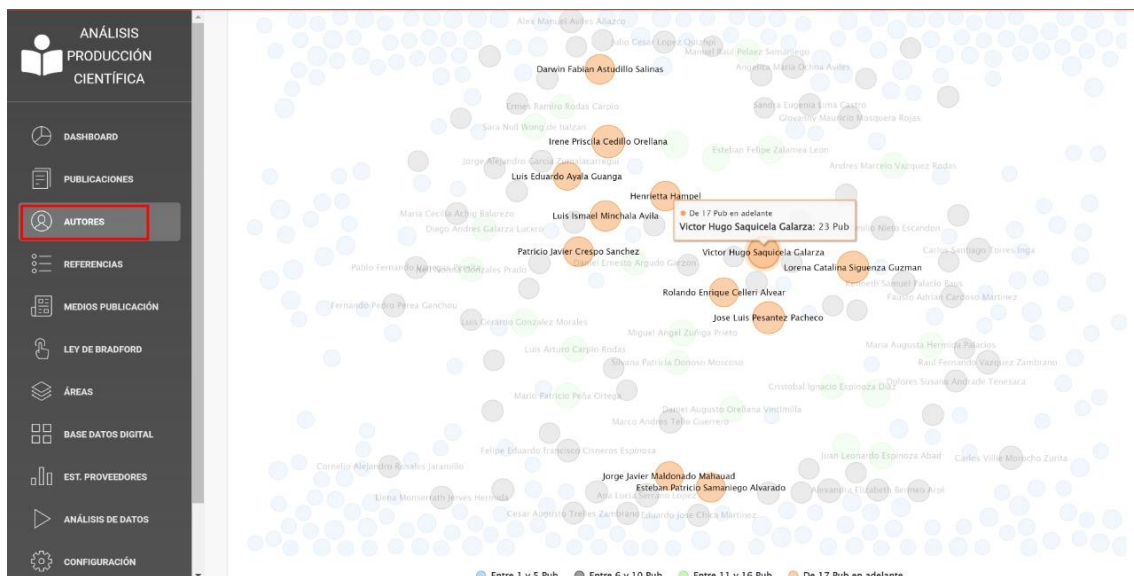


Figura 38. *Autores con el mayor número de publicaciones*

En la Sección Análisis de Datos, se tiene hasta la fecha 926 publicaciones ingresadas, que corresponden a los autores con filiación de la Universidad de Cuenca. En la opción Estadísticas de los datos, se realiza un análisis estadístico descriptivo tomando como indicador bibliométrico el número de

publicaciones que permitió hacer las siguientes evaluaciones: número de publicaciones por Área Unesco, Área Frascati, Medio de Publicación y Cuartil. En la Figura 39 se puede visualizar la productividad por Área Unesco en la cual sobresale Medio Ambiente con 14.25% (132 de 926) e Información y Comunicación (TIC) con un 14.15% (131 de 926).

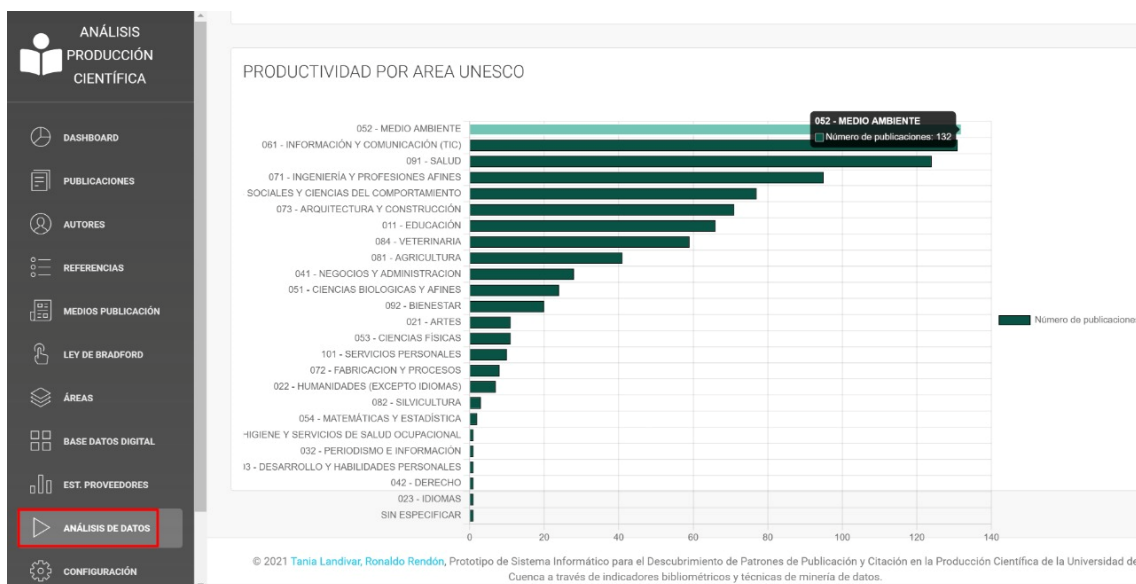


Figura 39. Número de publicaciones por Área Unesco.

Respecto a la productividad por Área Frascati, en la Figura 40 se puede visualizar que el área de Ingeniería Eléctrica, Electrónica e Información sobresale con un 13.50% (125 de 926).

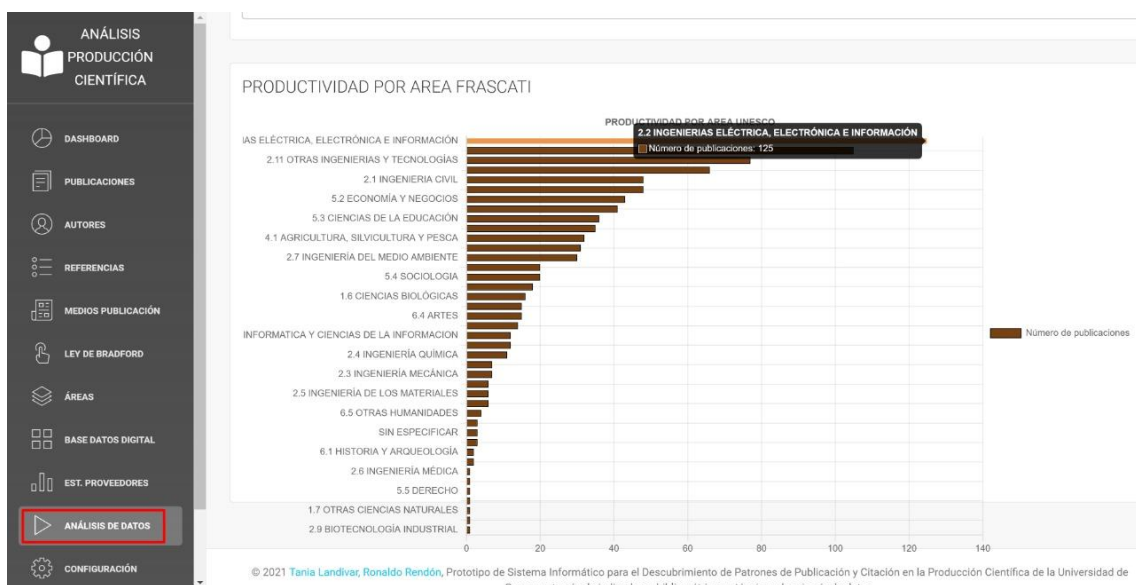


Figura 40. Número de publicaciones por Área Frascati.

Los medios de publicación a los que se orientan los investigadores de la Universidad de Cuenca a publicar se pueden ver en la Figura 41, siendo los más importantes Maskana con un 3.24% (30 de 926), Archivos Venezolanos de Farmacología y Terapéutica con un 2.27% (21 de 926) y Advances in Intelligent Systems and Computing con un 1.62% (15 de 926).

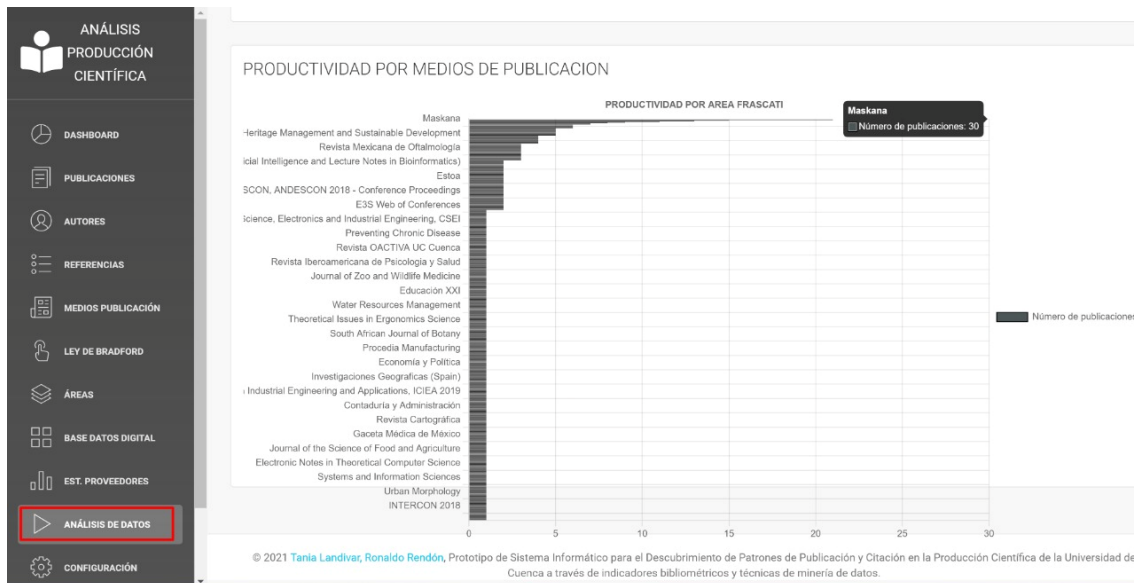


Figura 41. *Número de Publicaciones por Medios de Publicación.*

El cuartil es un indicador bibliométrico que permite evaluar la calidad de una publicación que se encuentra indexada en Scopus, de acuerdo con el valor del factor de impacto se ubica la publicación en un cuartil, siendo el más alto Q1 y el más bajo Q4. En la Figura 42 se puede ver la distribución de las publicaciones respecto a este indicador. El 42.87% (397 de 926) de publicaciones no están asignadas a ningún cuartil, el cuartil (Q1) tiene un 22.57% (209 de 926), el cuartil (Q2) tiene un 12.96% (120 de 926), el cuartil (Q3) tiene un 12.63% (117 de 926) y el cuartil (Q4) tiene un 8.96% (83 de 926). Algo muy importante de resaltar es que gran cantidad de las publicaciones no tienen un valor de factor de impacto que pueda ser considerado para ubicarse dentro de un cuartil.

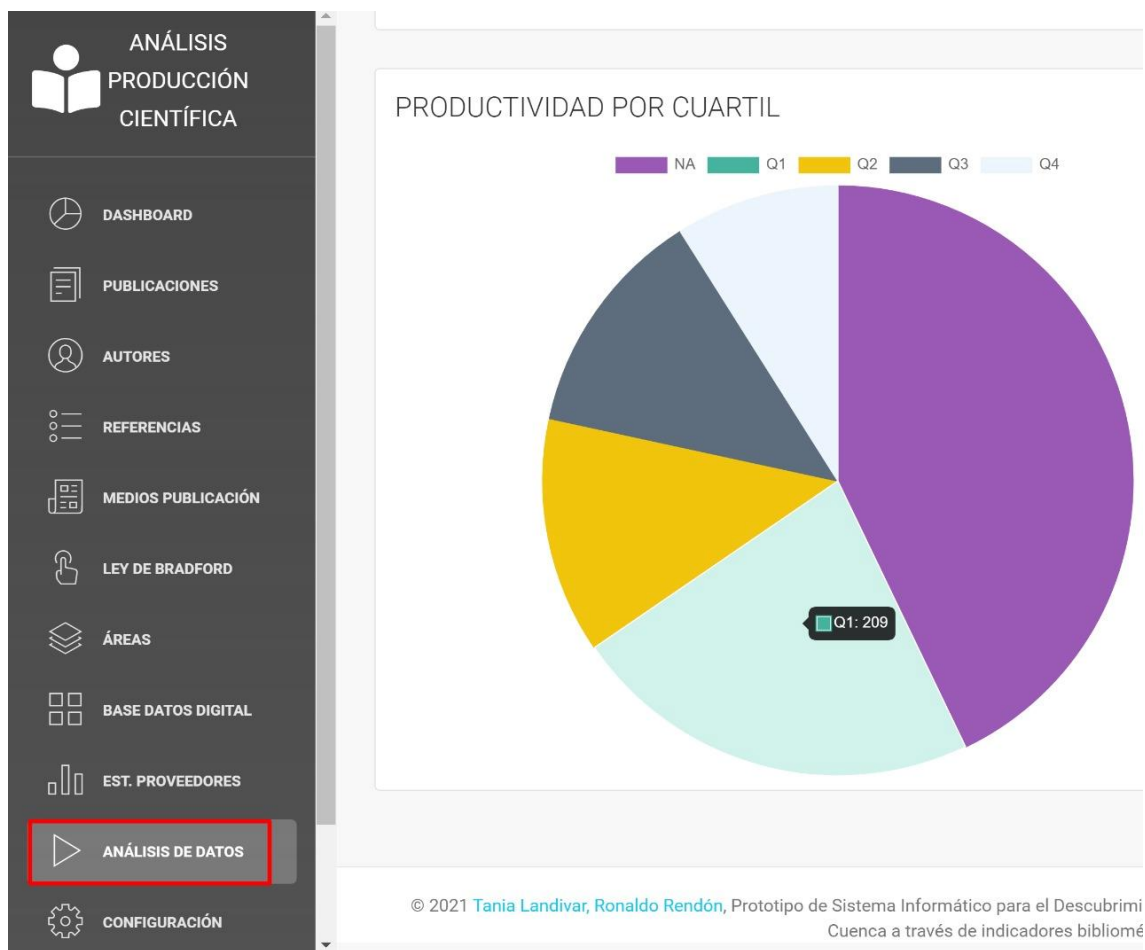


Figura 42. *Número de publicaciones por Cuartil*

Continuando con el análisis de datos en la opción de predicción que se visualiza en la Figura 43, se presenta una gráfica que informa la predicción del número de búsquedas realizadas en las revistas de Ebsco conjuntamente con Science Direct para los próximos seis meses, en función del número de búsquedas desde octubre del 2020 a octubre del 2021.



Figura 43. *Predicción del número de búsqueda de Ebsco con Science Direct en los próximos seis meses.*

Se puede observar que el número de búsquedas tiende a bajar en los meses de enero a febrero lo cual se puede relacionar al receso académico que existe entre los semestres, y tiende a subir desde marzo hasta abril por inicio de un nuevo semestre.

Antes de continuar es importante responder la pregunta de investigación número uno, en donde, se definieron como sobresalientes los indicadores bibliométricos: número de publicaciones, número de búsquedas, número de citas y factor de impacto SJR, para descubrir patrones de publicación y citación de los autores con filiación a la Universidad de Cuenca, ya que permite evaluar la productividad científica bajo diferentes líneas de acción. Por otro lado, la Ley de Bradford se destacó como el método bibliométrico que no sólo identifica las fuentes bibliográficas más productivas, sino también las más relevantes para cubrir un área del conocimiento determinada. Por lo que se integró este método con los indicadores bibliométricos antes mencionados, para generar información que apoye el proceso de toma de decisiones para la adquisición o suscripción de bases de datos bibliográficas. En la siguiente sección se presenta más a detalle esta integración y su aplicación.

La sección de La Ley de Bradford permite conocer el núcleo de los medios de publicación en donde se encuentran publicadas las referencias utilizadas por los autores de la Universidad de Cuenca, de acuerdo con el área de investigación y un periodo de tiempo.

En la Figura 44 se tiene el ejemplo del área “INGENIERÍAS ELÉCTRICA, ELECTRÓNICA E INFORMACIÓN” correspondiente al área Frascati desde el año 2016 – 2020.

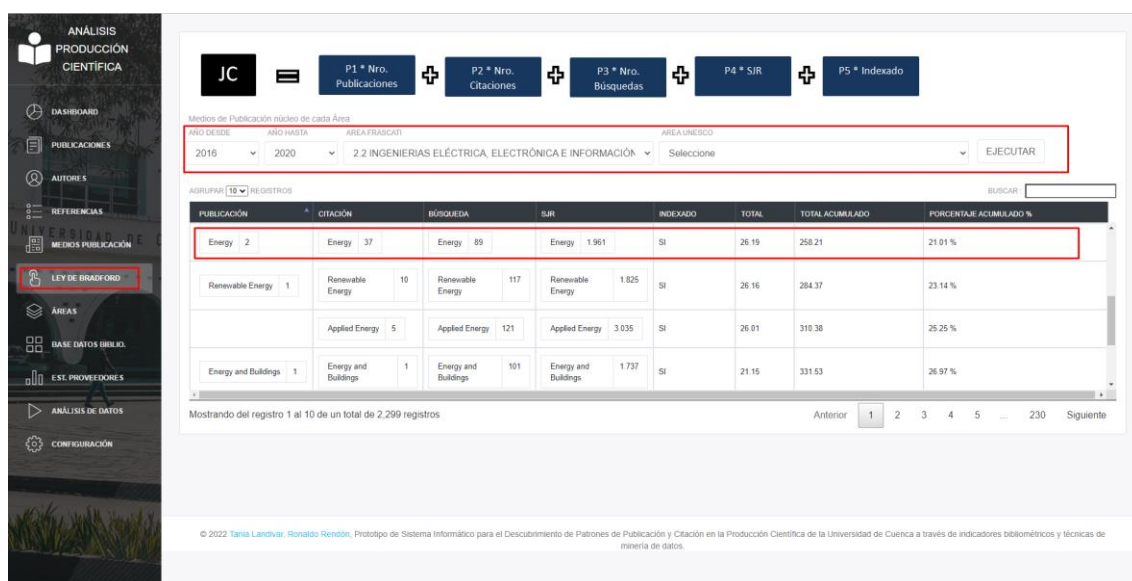


Figura 44. Ley de Bradford: Evaluación de los medios de publicación en función del número de publicaciones, número de citas, número de búsquedas y el factor de impacto

De la aplicación de la Ley de Bradford se listan los medios de publicación con su respectiva evaluación ordenadas de mayor a menor. Se consideran los pesos (P1, P2, P3, P4, P5) y valores correspondientes a cada uno de los indicadores bibliométricos: número de publicaciones, número de citas, número de búsquedas y el SJR. Los pesos permiten darle más importancia a uno o varios indicadores bibliométricos.

El 50% del porcentaje acumulado que se observa en la Figura 44 representa los medios de publicación núcleo, siendo estos los más notables y con mayor calificación, evaluados de acuerdo a los indicadores antes mencionados.

Para llegar a la evaluación final del medio de publicación; primero, se obtienen los medios de publicación con el mayor número de citas (ver la Figura 45), que representa el consumo de estas fuentes bibliográficas por parte de los autores de la Universidad de Cuenca. En el área seleccionada para el estudio se tiene un total de 2292 citas, de las cuales el 29.9% (687 de 2292) de los medios de publicación tienen más de tres citas.

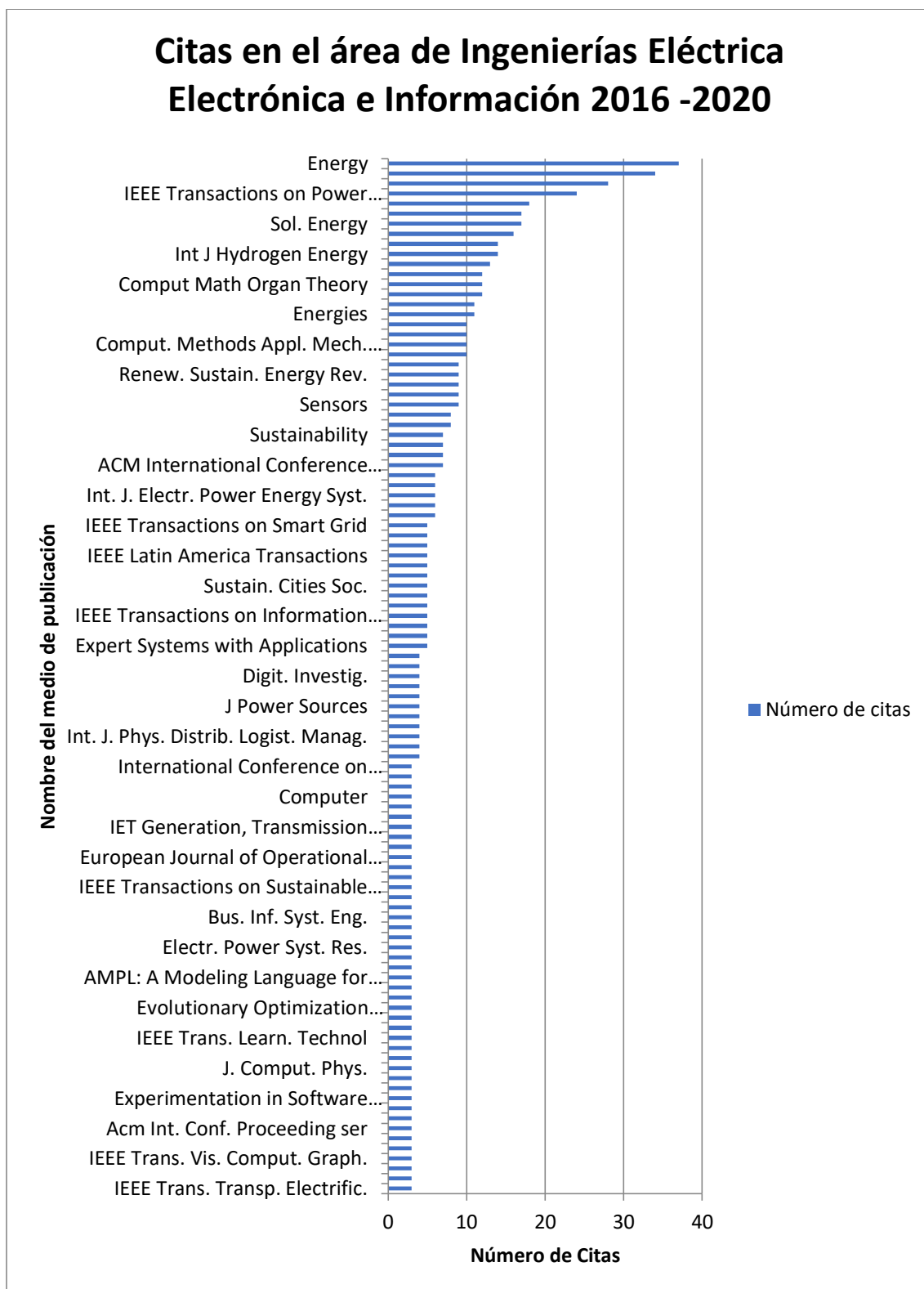


Figura 45. Medios de publicación más citados en el área de Ingenierías Eléctrica, Electrónica e Información.

Para responder a la pregunta de investigación sobre que fuentes bibliográficas son las más consumidas por los autores con filiación a la Universidad de Cuenca, en “INGENIERÍAS ELÉCTRICA, ELECTRÓNICA E INFORMACIÓN” se analiza la Figura 45 en donde se visualizan las más sobresalientes como “Energy”.

También se obtienen los medios de publicación con el mayor número de artículos (ver la Figura 46), que representan las fuentes bibliográficas más atractivas donde los investigadores publican sus artículos científicos. El total de publicaciones en esta área fue de 125, de las cuales el 44.8% (56 de 125) de los medios de publicación tienen más de dos publicaciones.

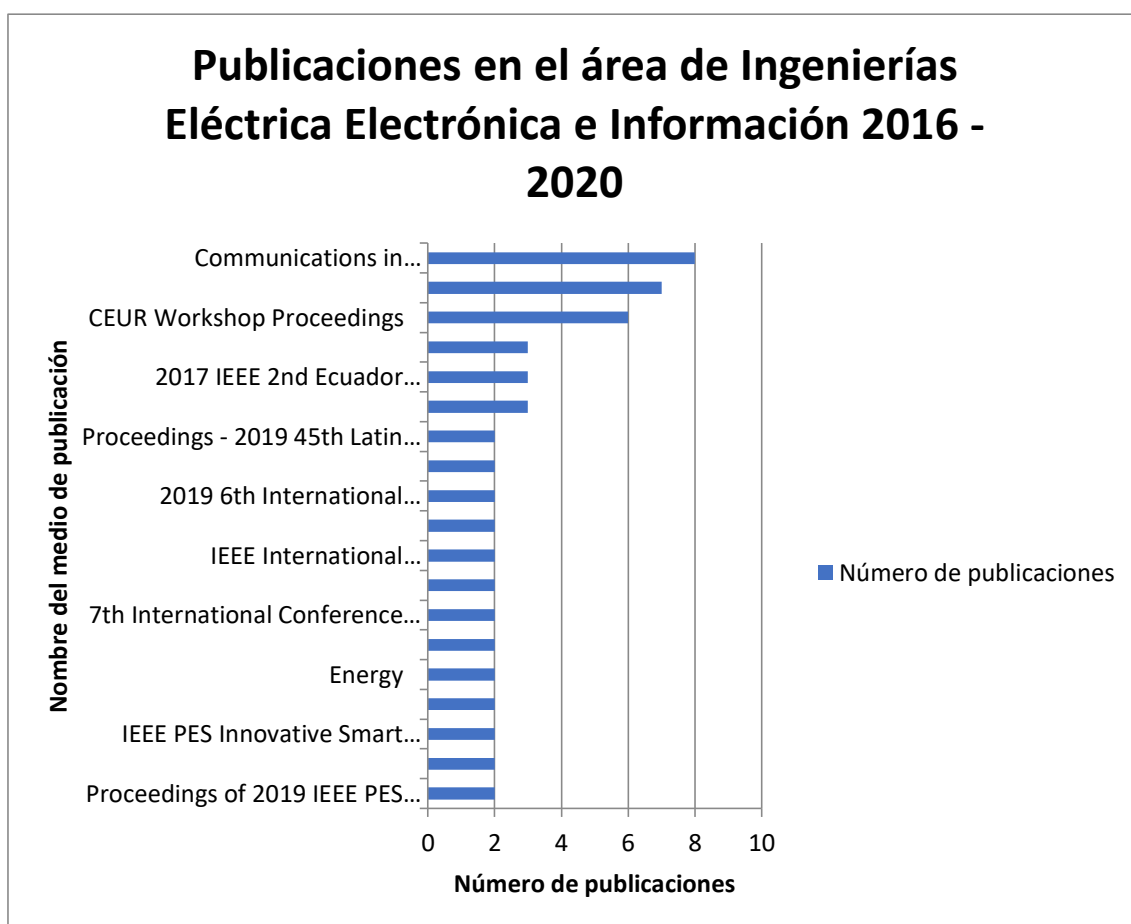


Figura 46. Medios de publicación con más publicaciones en el área de Ingenierías Eléctrica, Electrónica e Información

Continuado con las preguntas de investigación, las fuentes bibliográficas más atractivas donde publican los autores con filiación a la Universidad de Cuenca respecto para el área “INGENIERÍAS ELÉCTRICA, ELECTRÓNICA E INFORMACIÓN” se puede ver en la Figura 46, teniendo como la principal “Communications in Computer and Information Science”.

Cabe mencionar que en el prototipo se puede ejecutar estos análisis para diferentes áreas.

Una vez obtenidos los medios de publicación con el mayor número de publicaciones y citas, se calcula la Ley de Bradford, para lo cual se considera también el número de búsquedas, el factor de impacto (SJR) y si el medio de publicación evaluado está indexado en las bases de datos bibliográficas (Ebsco, Science Direct y Scopus) ya contratadas por la Universidad de Cuenca. En la Figura 47 se pueden observar los medios de publicación que mayor calificación tienen de acuerdo con los valores de sus pesos e indicadores bibliométricos. A los pesos (P1, P2, P3, P4, P5) para cada indicador bibliométrico se les asignó el valor del 20%, dando el total del 100% en los cinco. Este valor puede ser calibrado en función de las necesidades del CDRJBV.

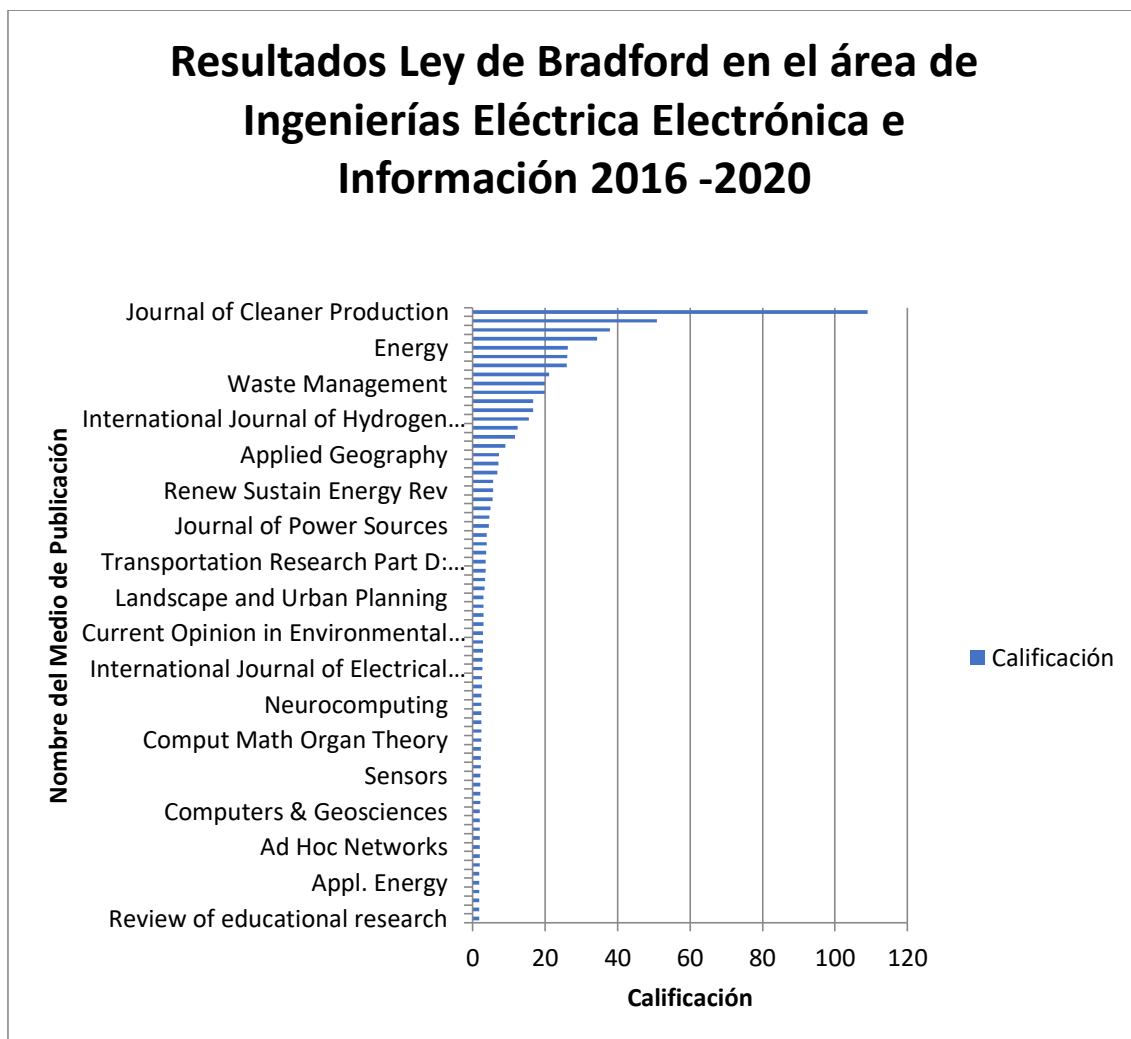


Figura 47. Medios de Publicación con la mayor calificación, resultado de la aplicación de Ley de Bradford en el área de Ingenierías Eléctrica, Electrónica e Información

Finalmente, se toma el núcleo de los medios de publicación que está representado entre el 0% y 50% del porcentaje acumulado presentado en la tabla en la sección Ley de Bradford (Figura 44), dando un total de 69 medios de publicación núcleo. Posteriormente, se determina si los mismos están siendo proporcionados por las bases de datos bibliográficas, dando como resultado que el 71% (49 de 69) de los medios de publicación utilizados tanto para publicar, citar, buscar y que tienen factor de impacto, están siendo suministrados por las bases de datos bibliográficas. Y, el 29% (20 de 69) de medios de publicación no se cuenta para su uso.

Del 71% (49 de 69) de los medios de publicación que están siendo suministrados por las bases de datos bibliográficas, el 73.5% (36 de 49) están proporcionadas tanto por Scopus y Science Direct, el 22.4% (11 de 49) por Scopus, y solamente el 4% (2 de 49) por Science Direct (ver Figura 48).

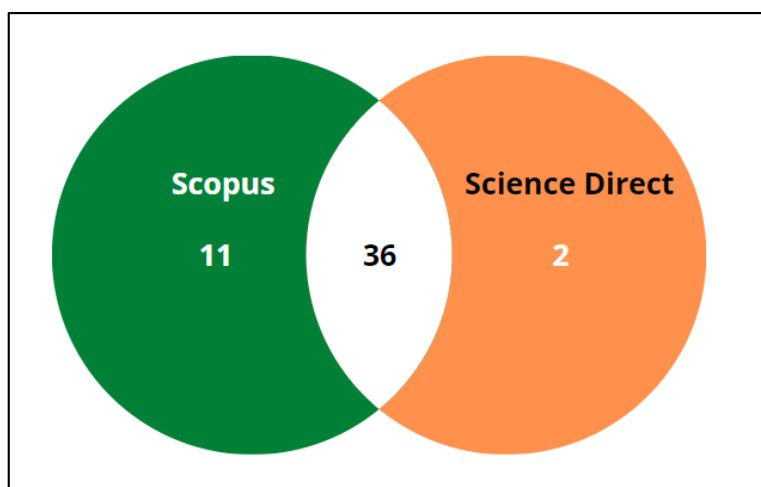


Figura 48. *Número de medios de publicación núcleo proporcionados por las bases de datos bibliográficas.*

Como último objetivo se tiene la comparación de las estadísticas que utiliza la Universidad de Cuenca con los resultados obtenidos en este trabajo de titulación. La universidad basa sus decisiones de adquisición y suscripción de bases de datos bibliográficas en función del número de búsquedas por mes proporcionadas por los proveedores (CEDIA, Publicencia y Ebsco) y algunas directrices establecidas por la institución como por ejemplo contratar bases de datos multidisciplinarias, así como también realizan una comparación entre el precio y el total de búsquedas por base de datos bibliográfica (Cabrera y Saenz, 2021). El presente prototipo considera además del número de búsquedas, el número de publicaciones, número de citas y el factor de impacto, permitiendo determinar las fuentes bibliográficas (revistas) en donde más se publican, citan, consultan y su calidad evaluada por el factor de impacto. Además, como se puede ver en la Figura 48, la distribución de las



revistas por base de datos bibliográfica, nos indica que la misma fuente está siendo proporcionada por dos bases de datos, situación que aporta al momento de tomar una decisión sobre adquisición o suscripción.

Y finalizando con la última pregunta de investigación, la información obtenida como patrones de publicación y citación en el sistema permite realizar diferentes análisis, que proveen una visión de cómo están siendo usadas las bases de datos y fuentes bibliográficas mismas que apoyan a la toma de decisiones de suscripción o adquisición. Del porcentaje de medios de publicación que no está siendo proporcionado por las bases de datos bibliográficas actualmente suscritas, se puede analizar y buscar la mejor opción que satisfaga esta demanda de información.



Capítulo 6: Conclusiones y Recomendaciones

Dentro de este capítulo se presentan las conclusiones obtenidas en la ejecución de este trabajo de titulación. Además, se plantean ciertas líneas de investigación y trabajos futuros que pueden surgir.

6.1 Conclusiones finales

La decisión para la contratación o adquisición de las bases de datos bibliográficas es proceso que constantemente se realiza en las bibliotecas universitarias, ya que se tiene que estar actualizando para intentar cubrir las necesidades de información de los estudiantes y docentes. En la actualidad, las bibliotecas, entre ellas la de la Universidad de Cuenca, basan sus decisiones en función de selectores quienes poseen un gran conocimiento en áreas temáticas, libros y revistas de interés. Una estrategia que ha dado buenos resultados es la formación de consorcios entre bibliotecas, ya que cada una tiene un representante que participa en la toma de decisiones para la colección del consorcio. Es importante que el personal encargado deba tener un conocimiento en compras, negociaciones con los proveedores y planificación presupuestaria. Sin embargo, existe un déficit en el análisis del consumo de los recursos como las revistas que proveen las bases de datos bibliográficas, ya que este es otro enfoque notable que apoya a la toma de decisiones.

Es por lo mencionado que, en este trabajo de titulación, se ha presentado y desarrollado el “Prototipo de sistema para el descubrimiento de patrones de publicación y citación de la producción científica, a través de indicadores bibliométricos y técnicas de minería de datos”. El prototipo presenta y visualiza información acerca de las revistas en las que más se publican, citan, consultan y mayor factor de impacto tienen. La misma que orienta a la adquisición y suscripción de bases de datos bibliográficas debido a que refleja tanto el consumo como la producción de publicaciones en determinadas fuentes bibliográficas, verificando si las mismas están indexadas en las bases de datos bibliográficas contratadas por la Universidad de Cuenca.

El análisis de requerimientos y la recolección de la información se realizó en base al modelo propuesto por Siguenza-Guzman et al. (2013): 1) Patrones de publicación, se consideró la información de los artículos publicados por los autores con filiación a la Universidad de Cuenca, de los cuales el medio de publicación permitió analizar cuál es la revista más atractiva donde publican dichos autores. 2) Patrones de citación, se tomó las fuentes bibliográficas (revistas) de las referencias citadas en los artículos de los autores de la



Universidad de Cuenca, permitiendo analizar el consumo que se ha tenido para el desarrollo de estos trabajos. 3) Patrones de descarga, el análisis se realizó en función de las estadísticas suministradas por los proveedores, obteniendo el número de búsquedas realizadas a las bases de datos y fuentes bibliográficas. Y, 4) para el factor de impacto, los datos fueron obtenidos desde el SJR que proporciona un ranking de las revistas con mayor valoración, ya sea a nivel general, por área o por años; esta valoración determina el impacto de las revistas en la comunidad científica internacional. Se obtuvo una tabla con los requisitos funcionales del sistema, así como una fórmula propuesta para la evaluación de las fuentes bibliográficas en función de los indicadores bibliométricos y la Ley de Bradford.

El diseño del prototipo es un paso de la metodología de desarrollo que describe la arquitectura del sistema, el modelo de la base de datos y la funcionalidad de las interfaces, se definió el modo de interacción del sistema (cliente - servidor), la estructura de la base de datos y sus relaciones. Además, se diseñó a nivel visual como estarían distribuidos los elementos en pantalla y su funcionalidad. Esta documentación sirvió como base para el próximo paso que es la implementación.

Al desarrollar el prototipo en una arquitectura cliente - servidor se pudo separar las funcionalidades, ya que por un lado el diseño del Front-end se desarrollaba sin contratiempos hasta que los servicios RESTful en el lado del Back-end estén listos, para posteriormente coordinar la interacción entre el cliente y el servidor. Obteniendo como resultado, un prototipo que permite desde la recolección, tratamiento, interpretación y carga de la información, hasta el análisis estadístico descriptivo, minería de datos y visualización de los resultados.

Se suministra diferentes líneas de evaluación que apoyen la toma de decisiones para seleccionar las fuentes bibliográficas que se necesitan adquirir y también si las bases bibliográficas ya contratadas están satisfaciendo la demanda de información. Se puede realizar el análisis a nivel general o como se presentó en este documento respecto a un área del conocimiento específica.

Los indicadores bibliométricos más sobresalientes para descubrir patrones de publicación y citación en el desarrollo de este prototipo son: número de publicaciones, número de citas, número de búsquedas y factor de impacto (SJR), que están relacionados a los índices productividad, consumo y calidad.

Para la evaluación de las fuentes bibliográficas se aplicó la Ley de Bradford ya que es un método bibliométrico que identifica las fuentes bibliográficas más relevantes para cubrir un área del conocimiento determinada.



En relación a las fuentes bibliográficas (revistas) más atractivas donde publican los autores con filiación a la Universidad de Cuenca respecto al área de “INGENIERÍAS ELÉCTRICA, ELECTRÓNICA E INFORMACIÓN” son: Communications in Computer and Information Science, Advances in Intelligent Systems and Computing y CEUR Workshop Proceedings. También las fuentes más consumidas por esta área son: Energy, IEEE Trans. Power Syst. y Renew Sustain Energy Rev.

Por lo antes mencionado, se puede concluir que al integrar los indicadores y el método bibliométrico “Ley de Bradford” se construyó una herramienta que proporciona gran cantidad de información (patrones de publicación y citación), la misma que complementa las decisiones de adquisición o suscripción de bases de datos bibliográficas. Y en comparación con las estadísticas que maneja la Universidad de Cuenca, el prototipo considera otros indicadores que analizan las necesidades de información desde diferentes perspectivas.

Finalmente, algo que destacar es que conforme pasa el tiempo, el número de publicaciones aumenta a nivel local y global, por lo que, es de vital importancia que la Universidad de Cuenca proporcione fuentes bibliográficas que satisfagan la demanda de información para el desarrollo de publicaciones. Ya que, por ejemplo, se espera que, si los investigadores cuentan con publicaciones indexadas en revistas con alto factor de impacto, la calidad de la publicación también se elevaría.



6.2 Líneas de trabajos futuros

La evaluación de la producción científica es un tema de interés en cualquier campo tanto para académicos como para profesionales, porque la investigación es un esfuerzo mundial y aplicable a la comunidad diversa.

En este trabajo se resalta principalmente la evaluación de una fuente bibliográfica (revistas) en función de la productividad, consumo y calidad; sin embargo, existen muchas aplicaciones que se pueden realizar con esta información.

El prototipo está diseñado para ser utilizado en cualquier institución educativa superior, por lo que no se descarta la implementación de mismo en otras instituciones, debido a que proporciona una base del conocimiento que apoya el proceso de toma de decisiones. Es por ello que, como líneas de trabajos futuros, se sugiere evaluar a los proveedores haciendo un análisis de costos y los recursos bibliográficos que facilitan.

Otra mejora que se podría implementar sobre el sistema en general es la integración directa y despliegue sobre los servidores de la Universidad de Cuenca, ya que, al tenerlos separados, se tienen que pedir los datos mediante una hoja de cálculo como se mencionó anteriormente y luego usarlos en el sistema. Mientras que, si se tuviera el sistema unificado, todos los procesos de carga de datos serían de manera automática y se evitaría y ahorraría tiempo para tratar toda la información que se maneja dentro del CDRJBV.

Finalmente, sería importante trabajar en conjunto con CEDIA para agilizar el proceso de búsqueda y extracción de información de calidad, ya que ellos poseen herramientas adecuadas para este tipo de análisis.



Bibliografía

- Abramo, G. (2017). Bibliometric evaluation of research performance: Where do we stand? *Voprosy Obrazovaniya*, 2017(1), 112–127. <https://doi.org/10.17323/1814-9545-2017-1-112-127>
- Aenta. (2011). *Sistema de gestión para evaluar y monitorear publicaciones científicas en la AENTA*.
- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-14142-8>
- Alvarado, R. U. (1999). La ley de Lotka y la literatura de bibliometría. *Investigación Bibliotecológica*, 13(27), 125–141. <https://dialnet.unirioja.es/servlet/articulo?codigo=962863>
- Alvarado, R. U. (2016). El crecimiento de la literatura sobre la ley de Bradford. *Investigación Bibliotecológica*, 30(68), 51–72. <https://dialnet.unirioja.es/servlet/articulo?codigo=5820273>
- Alvarado, R. U., & Arango, C. R. (2011). La ley de Zipf y el punto de transición de Goffman en la indización automática. *Investigación Bibliotecológica*, 25(54), 71–92. <https://dialnet.unirioja.es/servlet/articulo?codigo=5817230>
- Alvarez-Jareño, J. A., Badal Valero, E., & Pavía, J. M. (2018). Aplicación de métodos estadísticos, económicos y de aprendizaje automático para la detección de la corrupción. *Revista Internacional de Transparencia e Integridad*, 9, 7. <https://dialnet.unirioja.es/servlet/articulo?codigo=6977094>
- B S, M., & Rajgoli, I. (2017). Mapping of Scholarly Communication in Publications of the Astronomical Society of Australia, Publications of the Astronomical Society of Japan , and Publications of the Astronomical Society of the Pacific : A Bibliometric Approach. *Science & Technology Libraries*, 36, 351. <https://doi.org/10.1080/0194262X.2017.1368427>
- Babatunde, T., Alhassan, J., & Dr, B. (2020). *Effective Resource Sharing Services In University Libraries In North Central Nigeria*.
- Barrot, J. (2016). Research impact and productivity of Southeast Asian countries in language and linguistics. *Scientometrics*, 110. <https://doi.org/10.1007/s11192-016-2163-3>
- Beaudry, C., & Larivière, V. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy*, 45(9), 1790–1817. <https://doi.org/10.1016/J.RESPOL.2016.05.009>



- Beile, P., Boote, D., & Killingsworth, E. (2004). A Microscope or a Mirror?: A Question of Study Validity Regarding the Use of Dissertation Citation Analysis for Evaluating Research Collections (in Education). *Journal of Academic Librarianship*, 30(5), 347–353.
- Boeris, C. E. (2010). *Aplicación de métodos bibliométricos a la evaluación de colecciones: el caso de la Biblioteca del Instituto Argentino de Radioastronomía* [Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación]. <http://eprints.rclis.org/15824/>
- Bordons, M., & Zulueta, M. (1999). [Evaluation of the scientific activity through bibliometric indices]. *Revista Espanola de Cardiologia*, 52(10), 790–800. [https://doi.org/10.1016/S0300-8932\(99\)75008-6](https://doi.org/10.1016/S0300-8932(99)75008-6)
- Cabrera, Y., & Saenz, J. J. (2021). *Informe de uso de bases digitales - 2021*.
- Canales, C. B., Sanz-Valero, J., & Bojo-Canales, C. (2020). Indicadores de impacto y de prestigio de las revistas de ciencias de la salud indizadas en la red SciELO: estudio comparativo. *Rev Esp Salud Pública*, 94, 9–10. <https://repisalud.isciii.es/handle/20.500.12105/10992>
- Cañedo Andalia, R. (1999). Los análisis de citas en la evaluación de los trabajos científicos y las publicaciones seriadas. In *ACIMED*. <http://eprints.rclis.org/5281/>
- Chen, Y., Chen, S., Chang, Y., Lin, S. C., Chen, Y., Chen, S., & Lin, S. C. (n.d.). *A Metadata Lifecycle for Content Analysis in Digital Libraries A Metadata Lifecycle for Content Analysis in Digital Libraries*. 1–16.
- Chowdhury, G. (2014). Sustainability of digital libraries: A conceptual model and a research framework. *International Journal on Digital Libraries*, 14(3–4), 181–195. <https://doi.org/10.1007/S00799-014-0116-0>
- Contardi, S. (2004). Adquisición de publicaciones electrónicas en consorcios de bibliotecas. *Información, Cultura y Sociedad*, 0(11), 117–128. <https://doi.org/10.34096/ICS.111.916>
- Danell, J.-A. B. (2020). Integrative oncology from a bibliometric point of view. *Complementary Therapies in Medicine*, 52, 102477. <https://doi.org/10.1016/j.ctim.2020.102477>
- Danell, J.-A. B., Danell, R., & Vuolanto, P. (2020). Scandinavian research on complementary and alternative medicine: A bibliometric study. *Scandinavian Journal of Public Health*, 48(6), 609–616. <https://doi.org/10.1177/1403494819834099>
- Daza, A. (2016). *DATA MINING: MINERIA DE DATOS*.



- desarrolloweb.com. (n.d.). *Qué es React. Por qué usar React*. Retrieved December 22, 2021, from <https://desarrolloweb.com/articulos/que-es-react-motivos-uso.html>
- Dios, F. J. G. de, Benavent, M. M., & Hernández, M. A. M. (1997). Indicadores bibliométricos: características y limitaciones en el análisis de la actividad científica. *Anales Españoles de Pediatría: Publicación Oficial de La Asociación Española de Pediatría (AEP)*, 47(3 (SEPTIEMBRE)), 235–244. <https://dialnet.unirioja.es/servlet/articulo?codigo=4412628>
- Duy, J., & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination1. *Journal of Academic Librarianship*, 32(5), 512–517. <https://doi.org/10.1016/J.ACALIB.2006.05.005>
- Ellegaard, O. (2018). The application of bibliometric analysis: disciplinary and user aspects. *Scientometrics*, 116(1), 181–202. <https://doi.org/10.1007/s11192-018-2765-z>
- Enger, K. B. (2009). Using citation analysis to develop core book collections in academic libraries. *Library and Information Science Research*, 31(2), 107–112. <https://doi.org/10.1016/J.LISR.2008.12.003>
- Figueroa-Díaz, R., Sólis, C., & Cabrera, A. (2007). *METODOLOGÍAS TRADICIONALES VS. METODOLOGÍAS ÁGILES*. <https://doi.org/10.13140/RG.2.1.2897.3206>
- Fink, A. (2014). *Evaluation Fundamentals: Insights into Program Effectiveness, Quality, and Value*. SAGE Publications.
- Galvez, C. (2018). Análisis de co-palabras aplicado a los artículos muy citados en Biblioteconomía y Ciencias de la Información (2007-2017). *Transinformação*, 30(3), 277–286. <https://www.redalyc.org/jatsRepo/3843/384357985001/index.html>
- Garcés, L., & Egas, L. M. (2013). Evolución de las Metodologías de desarrollo de la Ingeniería de software en el proceso la Ingeniería de Sistemas Software. *Revista Científica y Tecnológica UPSE*, 1(3). <https://doi.org/10.26423/rctu.v1i3.29>
- Gatzui, S. (1999). Data Warehousing: Concepts and Mechanisms. *Wirtschaftsinformatik Als Mittler Zwischen Technik, Ökonomie Und Gesellschaft*, 61–69. https://doi.org/10.1007/978-3-322-94873-1_6
- Glänzel, W. (2003). Bibliometrics as a research field: A course on Theory and Application of Bibliometric Indicators. *Researchgate*, May, 115. https://www.researchgate.net/publication/242406991_Bibliometrics_as_a_r



research_field_A_course_on_theory_and_application_of_bibliometric_indicators

- Glänzel, W., & Moed, H. F. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, 96(1), 381–394. <https://doi.org/10.1007/s11192-012-0898-z>
- González Alcaide, G., & Gorraiz, J. I. (2018). Assessment of Researchers Through Bibliometric Indicators: The Area of Information and Library Science in Spain as a Case Study (2001–2015). *Frontiers in Research Metrics and Analytics*, 3, 15. <https://doi.org/10.3389/FRMA.2018.00015>
- Guerrero-Sosas, J. D. T., Chicharro, F. P. R., Serrano-Guerrero, J., Menendez-Dominguez, V., & Castellanos-Bolaños, M. E. (2019). A proposal for a recommender system of scientific relevance. *Procedia Computer Science*, 162, 199–206. <https://doi.org/10.1016/J.PROCS.2019.11.276>
- Guo, Y., Hao, Z., Zhao, S., Gong, J., & Yang, F. (2020). Artificial Intelligence in Health Care: Bibliometric Analysis. *Journal of Medical Internet Research*, 22(7). <https://doi.org/10.2196/18228>
- Gureev, V. N., & Mazov, N. A. (2015). Assessment of the relevance of journals in research libraries using bibliometrics (a review). *Scientific and Technical Information Processing 2015* 42:1, 42(1), 30–40. <https://doi.org/10.3103/S0147688215010050>
- Gutiérrez Ramírez, M., & Landeros Falcón, I. (2010). Importancia del lenguaje en el contexto de la aldea global. *Horizontes Educativos*, 15(1), 95–108.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.
- Haro Valle, V. A., Pérez Rocano, W. R., Sigüenza Guzmán, L., Cattrysse, D., Saquicela Galarza, V. H., Cuenca, U. de, Cuenca, D. de I. de la U. de, & DIUC. (2014). *Diseño e implementación de un sistema de soporte de decisiones para el Centro de Documentación Regional “Juan Bautista Vázquez.”* <http://dspace.ucuenca.edu.ec/handle/123456789/21411>
- Heneberg, P. (2016). From Excessive Journal Self-Cites to Citation Stacking: Analysis of Journal Self-Citation Kinetics in Search for Journals, Which Boost Their Scientometric Indicators. *PloS One*, 11(4), e0153730. <https://doi.org/10.1371/journal.pone.0153730>
- Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2004). *Introducción a la minería de datos*.



- Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*.
- Leidolf, A., Baker, M., Porter, A., & Garner, J. (2018). Measuring and Visualizing Research Collaboration and Productivity. *Journal of Data and Information Science*, 3, 54–81. <https://doi.org/10.2478/jdis-2018-0004>
- LEMONCODE. (2016). *React y D3.js, trabajando juntos I - Introducción — Lemoncode formacion*. <https://lemoncode.net/lemoncode-blog/2018/6/5/react-y-d3js-trabajando-juntos>
- Linoff, G. S., & Berry, M. J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons.
- Ma, J. (2014). Bibliometric indicators for evaluating the quality of scientific publications. *The Journal of Contemporary Dental Practice*, 15(2), 258–262. <https://doi.org/10.5005/jp-journals-10024-1525>
- Mancini, D. (1996). Mining your automated system for systemwide decision making. *Library Administration Management*.
- Marcos-Pablos, S., & García-Peñalvo, F. J. (2020). Information retrieval methodology for aiding scientific database search. *Soft Computing*, 24(8), 5551–5560. <https://doi.org/10.1007/s00500-018-3568-0>
- Marini, I. E. (2012). *El Modelo Cliente / Servidor*. 1–11.
- Mariscal Orozco, J. L., & Girarte, J. L. (2017). Digital repositories for the training and researching processes in Cultural Management. *Córima*, 2. <https://doi.org/10.32870/cor.a2n3.6591>
- Marquez, L. (1980). *Capítulo 5. Cliente-Servidor*. 5.1. 1–9. http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/marquez_a_bm/capitulo5.pdf
- Miguel, S., De-Moya-Anegón, F., & Herrero-Solana, V. (2007). El análisis de citas como método de investigación en Bibliotecología y Ciencia de la Información. In *Investigación Bibliotecológica*. <http://eprints.rclis.org/15311/>
- Moen, W. E., Stewart, E. L., & McClure, C. R. (1997). The role of content analysis in evaluating metadata for the us government information locator service (GILS): results from an exploratory study. *IEEE Computer Society Metadata Conference*, 1--14.
- Nicholson, S., & Stanton, J. M. (2003). Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. *Organizational Data Mining: Leveraging*



- Enterprise Data Resources for Optimal Performance*, 247-262.
<http://arizona.openrepository.com/arizona/handle/10150/106383>
- Osueke, C. O., Idiegbeyan-ose, J., Botu, T., Aregbesola, A., & Emmanuel, O. (2018). Analysis of mechanical engineering research activities using bibliometric method: A case study of undergraduate projects. *International Journal of Mechanical Engineering and Technology*, 9, 1014–1021.
- Peñalver, A. G. (2014). *Técnicas para el descubrimiento de patrones temporales= Techniques for the discovery of temporal patterns* [Universidad de Murcia].
<https://dialnet.unirioja.es/servlet/tesis?codigo=96909>
- Perakakis, P., Ponsati-Obiols, A., Bernal, I., Sierra, C., Osman, N., Mosquera-de-Arancibia, C., & Lorenzo, E. (2016). *Módulo de evaluación por pares en abierto para repositorios de acceso abierto*. <http://eprints.rclis.org/30132/>
- Proctor, J. (2019). Knowledge Management for Collection Development: Transforming Institutional Knowledge into Tools for Selectors. *The Serials Librarian*, 76(1–4), 118–122.
<https://doi.org/10.1080/0361526X.2019.1551668>
- Ranaei, S., Suominen, A., Porter, A., & Carley, S. (2020). Evaluating technological emergence using text analytics: two case technologies and three approaches. *Scientometrics*, 122(1), 215–247.
<https://doi.org/10.1007/s11192-019-03275-w>
- Repiso, R. (2020). *La autoría: ¿Cuántos firman, quiénes y en qué orden? – Comunicar. Escuela de Autores*.
<https://www.revistacomunicar.com/wp/escuela-de-autores/la-autoria-cuantos-firman-quienes-y-en-que-orden/>
- Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosures. *Science Advances*, 3(11).
<https://doi.org/10.1126/SCIADV.1700404>
- Shi, X., Cai, L., & Jia, J. (2018). The Evolution of International Scientific Collaboration in Fuel Cells during 1998–2017: A Social Network Perspective. *Sustainability*, 10(12), 4790.
<https://doi.org/10.3390/su10124790>
- Shokraneh, F. (2016). *PICO Framework: Two Decades of Variation and Application*.
https://www.researchgate.net/publication/328334814_PICO_Framework_Two_Decades_of_Variation_and_Application
- Shon, W. Y., Yoon, B.-H., Jung, E.-A., Kim, J. W., Ha, Y.-C., Han, S. H., & Kim,



- H.-S. (2019). Assessment of Korea's Orthopedic Research Activities in the Top 15 Orthopedic Journals, 2008–2017. *Clinics in Orthopedic Surgery*, 11(2), 237. <https://doi.org/10.4055/cios.2019.11.2.237>
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. *Proceedings of IATUL Conferences, May 2014*, 1–9. [https://lirias.kuleuven.be/bitstream/123456789/399278/1/IATUL Conference - Siguenza-Guzman et al.pdf](https://lirias.kuleuven.be/bitstream/123456789/399278/1/IATUL%20Conference%20-%20Siguenza-Guzman%20et%20al.pdf)
- Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature Review of Data Mining Applications in Academic Libraries. *The Journal of Academic Librarianship*, 41(4), 499–510. <https://doi.org/10.1016/J.ACALIB.2015.06.007>
- Simao, L. B., Carvalho, L. C., & Madeira, M. J. (2020). Intellectual structure of management innovation: bibliometric analysis. *Management Review Quarterly 2020 71:3*, 71(3), 651–677. <https://doi.org/10.1007/S11301-020-00196-4>
- Thornton, G. A. (2000). Impact of electronic resources on collection development, the roles of librarians, and library consortia. *Library Trends*, 48(4), 842–856.
- Torres-Berru, Y., Batista, V. F. L., & Torres-Carrión, P. (2020). Data mining to detect and prevent corruption in contracts: Systematic mapping review. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, 2020(E29), 13–26.
- Tukur, L., & Kannan, S. (2020). An Appraisal of Availability and Utilization of Information Resources and Library Services by Undergraduate Students in Three Agriculture University Libraries in Northern Nigeria (ABU , FUAM, FUTM). *Library Philosophy and Practice (e-Journal)*. <https://digitalcommons.unl.edu/libphilprac/4591>
- Ullah, S. (2019). Publication Trends Of Pakistan Heart Journal: A Bibliometric Study. *Library Philosophy and Practice (e-Journal)*. <https://digitalcommons.unl.edu/libphilprac/2237>
- Urban, S. G. (2019). Using E-Journal Subscription Inventories to Improve Internal Documentation and Patron Access. *https://doi.org/10.1080/07317131.2019.1621568*, 36(3), 269–280. <https://doi.org/10.1080/07317131.2019.1621568>
- Van Eck, N. J., & Waltman, L. (2018). VOSviewer Manual: Manual for VOSviewer version 1.6.7. *Univeriteit Leiden, February*, 51.



https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf

- Wang, Z., Chen, Y., & Glänzel, W. (2020). Preprints as accelerator of scholarly communication: An empirical analysis in Mathematics. *Journal of Informetrics*, 14(4), 101097. <https://doi.org/10.1016/j.joi.2020.101097>
- Zhao, W., Zhao, S., & MacGillivray, K.-S. (2017). Providing and Maintaining Access to Electronic Serials: Consortium and Member University Library's Perspectives. *The Serials Librarian*, 72(1–4), 144–151. <https://doi.org/10.1080/0361526X.2017.1309831>
- Zhimnay Valverde, C., Fernandez, J., Albarracín, J., Sádaba-Rodríguez, I., & Sucozhanay, D. (2019). Mapping of Scientific Production in Social Sciences in Ecuador. *INTED2019 Proceedings*, 1(March), 9501–9509. <https://doi.org/10.21125/inted.2019.2354>
- Zhu, J., & Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321–335. <https://doi.org/10.1007/s11192-020-03387-8>

Anexos

Anexo 1: Diseño Del Prototipo

Dashboard

En la Figura A1.1 se observa el cuadro de resumen de resultados más relevantes, a través de gráficos y diagramas para la visualización de datos.

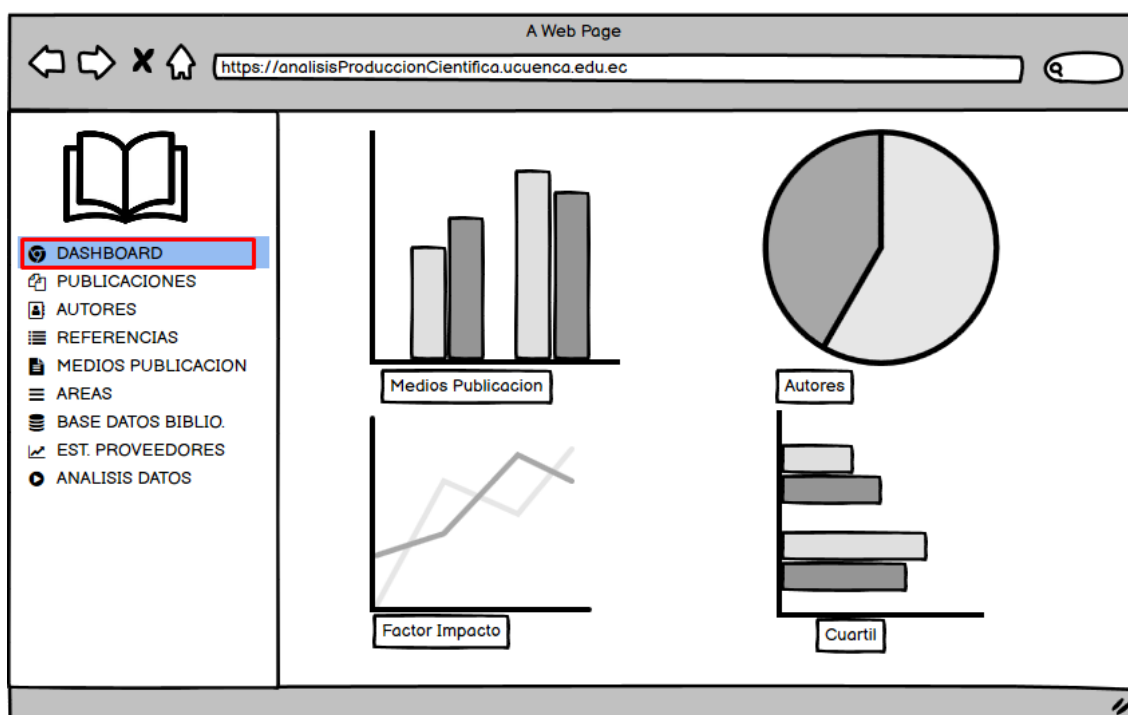


Figura A1.1 *Diseño conceptual para la pantalla Dashboard.*

Pantalla para el ingreso de autores y su relación con sus publicaciones.

Se puede visualizar la sección de Autores (ver Figura A1.2) que permite el ingreso de los autores que han intervenido en cada una de las publicaciones, el ingreso se realiza a través de un archivo de Excel.

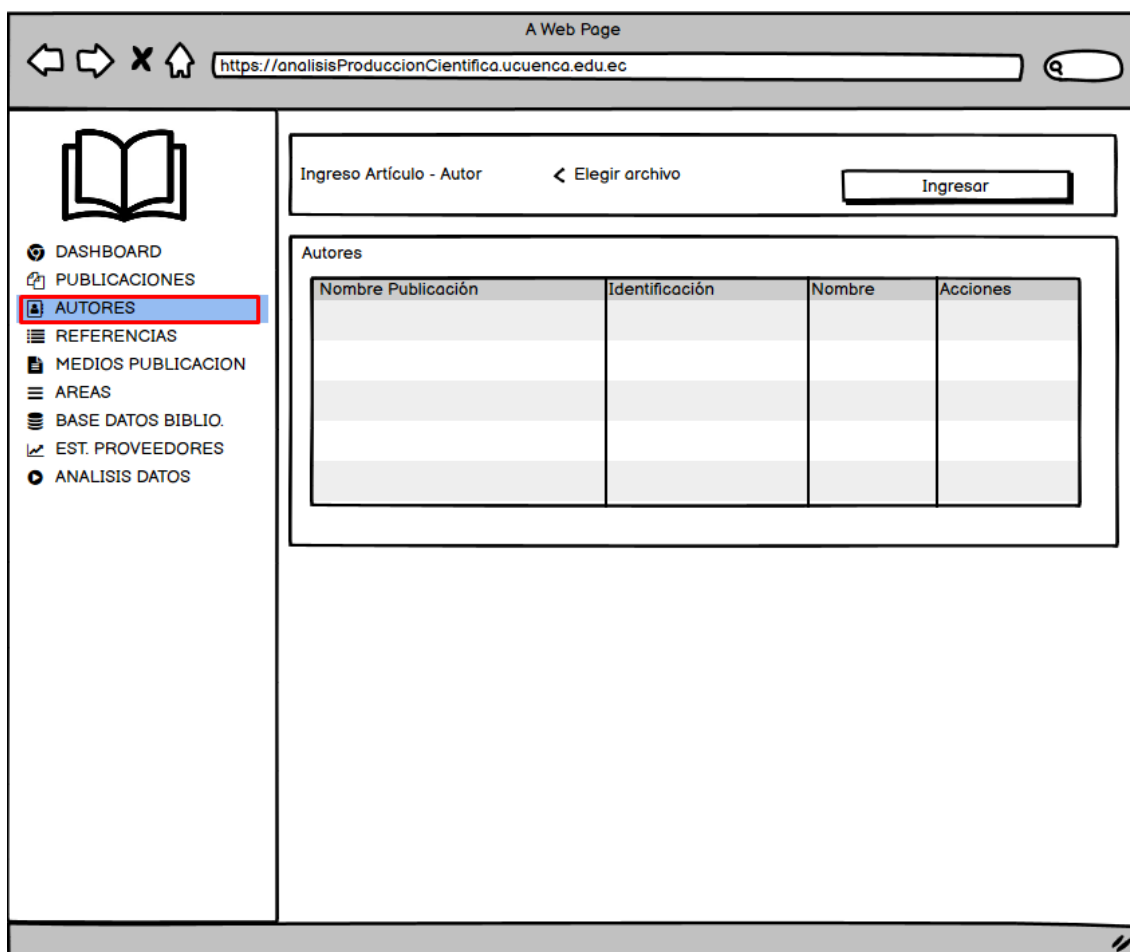


Figura A1. 2 *Diseño conceptual de la pantalla autores*

Pantalla para el ingreso de los medios de publicación.

Pantalla para el manejo de información de los Medios de Publicación donde se encuentran indexadas las publicaciones de los autores de la Universidad de Cuenca, se puede observar en la Figura A1.3.

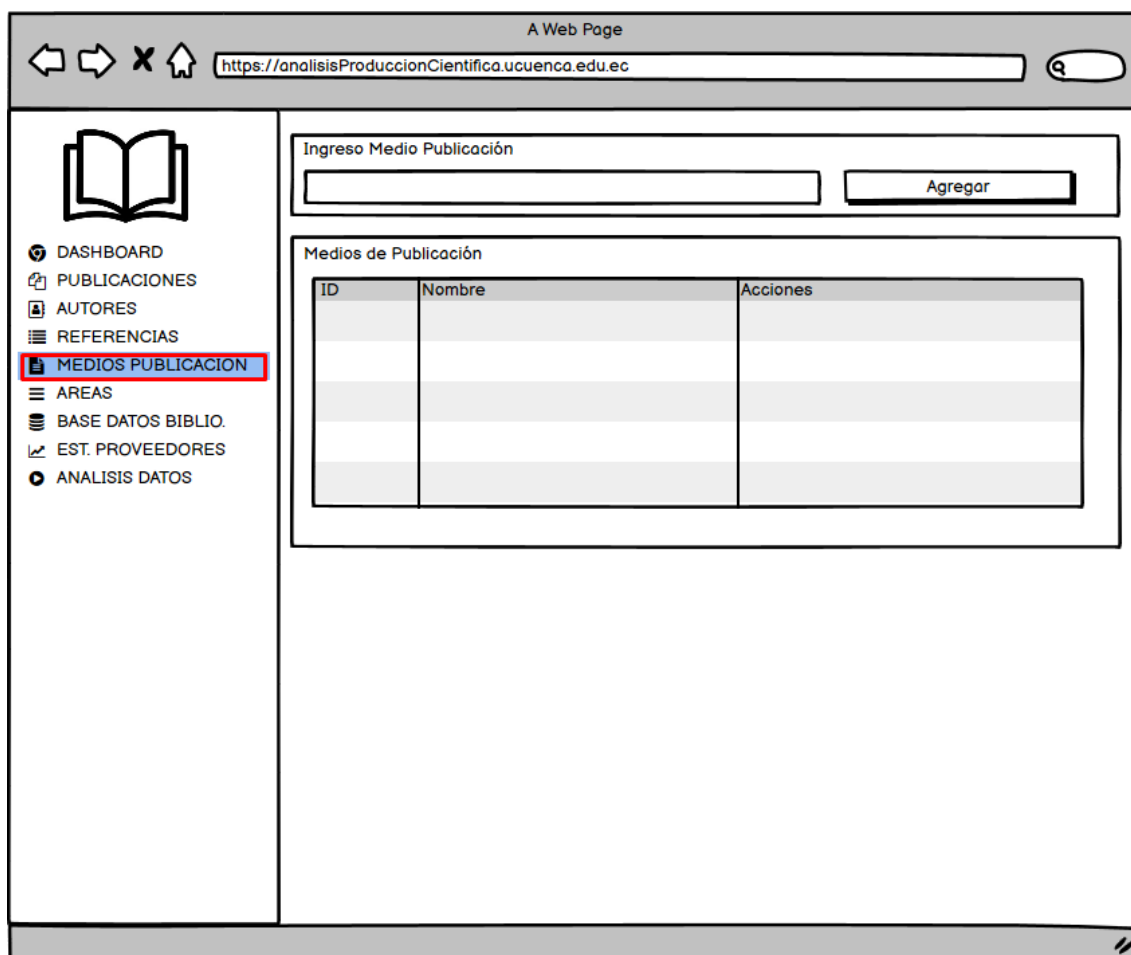
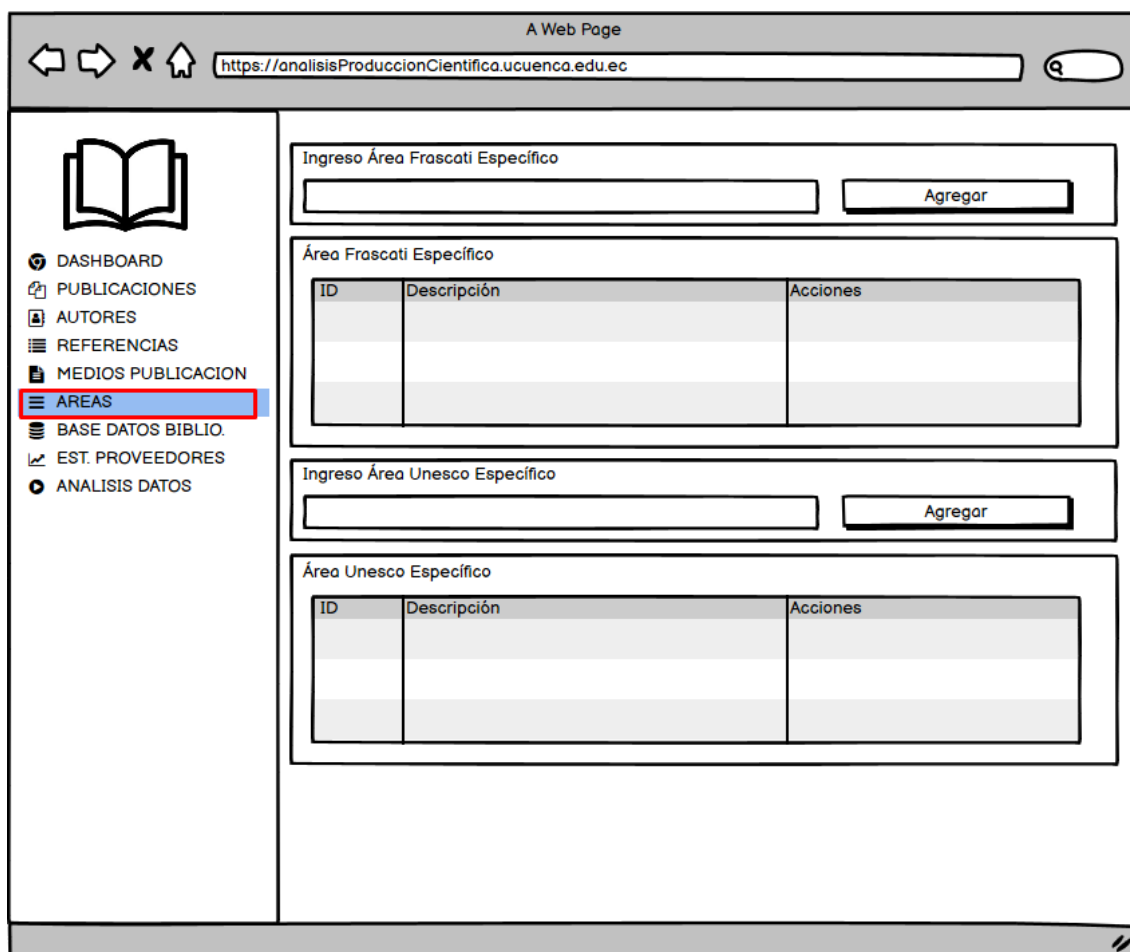


Figura A1. 3 *Diseño conceptual de la pantalla medios de publicación*

Pantalla para el ingreso de las áreas Frascati y Unesco.

Pantalla para el manejo de información de las diferentes áreas a las que corresponden las publicaciones de los autores con filiación a la Universidad de Cuenca (ver Figura A1.4), las mismas que se encuentran en dos categorías: Área Frascati y Área Unesco.

Figura A1.4 *Diseño conceptual de la pantalla áreas*

Pantalla para el ingreso de las bases de datos bibliográficas que provee la Universidad de Cuenca para su uso.

La pantalla para el ingreso de las bases de datos bibliográficas se puede visualizar en la Figura A1.5, cuyos campos requeridos son: Nombre, Proveedor, Costo, Área/Servicio, Suscripción/Descripción. Finalmente, se listan los que ya están ingresados en una tabla.

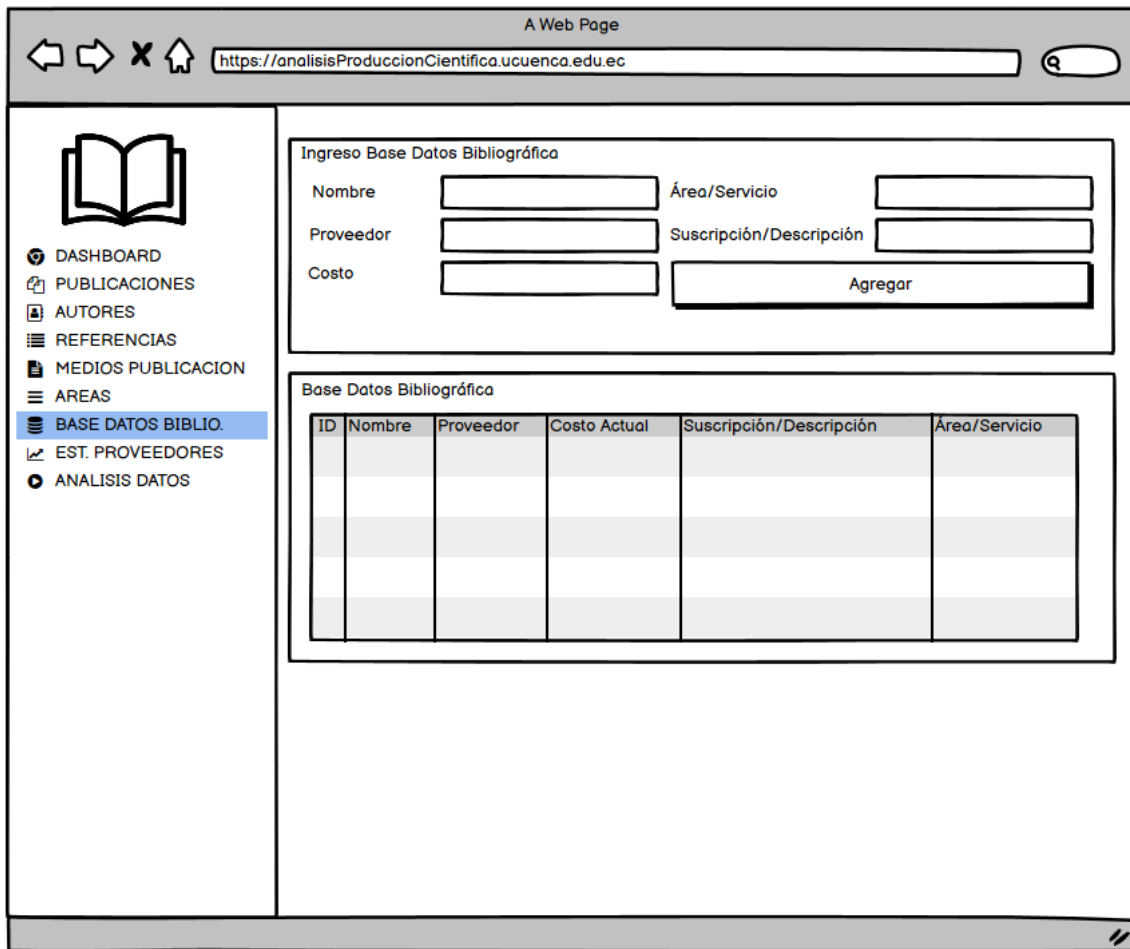


Figura A1. 5 Diseño conceptual de la pantalla base de datos bibliográficas



Anexo 2: Manual de Usuario

Índice

Índice	122
Índice de figuras	123
Presentación	125
Desarrollo del Manual de Usuario	125
Dashboard	125
Publicaciones	126
Autores	128
Referencias	130
Medios de publicación	132
Áreas	135
Bases de datos bibliográficas	136
Estadísticas de los proveedores	137
Configuración	139
Ley de Bradford	140
Análisis de datos	141



Índice de figuras

Figura A2. 1. Sección Dashboard, indicadores de almacenamiento.	126
Figura A2. 2. Pantalla para el mantenimiento y visualización de las publicaciones.	126
Figura A2. 3. Visualización de las referencias en texto por cada publicación.	127
Figura A2. 4. Visualización del detalle de las referencias separada en campos por cada publicación.	127
Figura A2. 5. Pantalla para el mantenimiento de los autores y sus publicaciones.	128
Figura A2. 6. Publicaciones por autor.	129
Figura A2. 7. Gráfico de burbuja en función del número de publicaciones por autor.	129
Figura A2. 8. Gráfico de burbuja de acuerdo con el número de publicaciones según el orden del autor en la publicación.	130
Figura A2. 9. Pantalla para la extracción de las referencias y el detalle de las referencias por publicación.	130
Figura A2. 10. Pantalla para el ingreso de forma manual y automática de las referencias y el detalle de mismas.	131
Figura A2. 11. Pantalla para la extracción del detalle de la referencia a través de la API de Google Académico.	132
Figura A2. 12. Pantalla Medios Publicación opción medios de publicación.	132
Figura A2. 13. Pantalla Medios Publicación opción ranking medios publicación.	133
Figura A2. 14. Pantalla Medios Publicación opción ranking medios citación.	134
Figura A2. 15. Pantalla Medios Publicación opción ranking búsquedas journal-database.	134
Figura A2. 16. Pantalla Medios Publicación opción factor de impacto SJR.	135
Figura A2. 17. Pantalla para visualizar la sección de áreas	136



Figura A2. 18. Pantalla para la sección de mantenimiento de las bases de datos bibliográficas	136
Figura A2. 19. Pantalla para el ingreso de medios de publicación como revistas que proporcionan las bases de datos bibliográficas	137
Figura A2. 20. Pantalla para el ingreso de las estadísticas de uso de las bases de datos bibliográficas y sus revistas o fuentes bibliográficas.	138
Figura A2. 21. Gráfico de barras de la estadística de búsqueda de las bases de datos bibliográficas.	138
Figura A2. 22. Gráfico de barras de la estadística de búsqueda de las fuentes bibliográficas indexadas en las bases de datos.....	139
Figura A2. 23. Sección	configuración. 139
Figura A2. 24. Sección para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto.	140
Figura A2. 25. Tabla de resultados para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto.	141
Figura A2. 26. Sección	análisis de datos. 141



Presentación

El siguiente manual contiene información acerca de cómo manipular el prototipo de sistema informático para el descubrimiento de patrones de publicación y citación en la producción científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos.

Desarrollo del Manual de Usuario

Las secciones o tareas que forman parte del prototipo de sistema son las siguientes:

- Dashboard
- Publicaciones
- Autores
- Referencias
- Medios de publicación
- Ley de Bradford
- Áreas
- Bases de datos bibliográficas
- Estadísticas de los proveedores
- Análisis de datos
- Configuración

Dashboard

Para acceder a esta sección del sistema, se debe dar clic sobre la opción “Dashboard” que se encuentra en la parte superior izquierda de la pantalla como se observa en la Figura A2. 1.

Una vez que se accede a la opción de Dashboard, se presenta la página principal del sistema, en donde se visualiza el número total de publicaciones almacenadas pertenecientes a los autores con filiación de la Universidad de Cuenca, el total de publicaciones que no tienen referencias, el total de referencias, y el número total del detalle de las referencias extraídas.



Figura A2. 1 Sección Dashboard, indicadores de almacenamiento.

Publicaciones

Para ingresar a la sección “Publicaciones” se debe presionar el botón con el mismo nombre que se encuentra en la parte izquierda de la pantalla como se observa en la Figura A2. 2. Una vez presionado la opción “Publicaciones”, se puede visualizar la pantalla para el mantenimiento y visualización de las publicaciones. En esta sección se ingresan los artículos de los investigadores con filiación a la Universidad de Cuenca indexadas en Scopus y Latindex en el periodo de 2016-2020. El ingreso se hace a través de un archivo de Excel con los siguientes campos: URL de la publicación, título, título alternativo, palabras clave, abstract, resumen, nombre área Frascati amplio, nombre área Frascati específico, nombre área Unesco amplio, nombre área Unesco específico, tipo publicación, año publicación, nombre medio publicación, link de la revista, DOI, nombre de la base de datos bibliográfica, estado de la publicación, enlace del documento, factor de impacto, cuartil, identificación del autor, orden del autor, nombre del autor y nombre de la filiación.

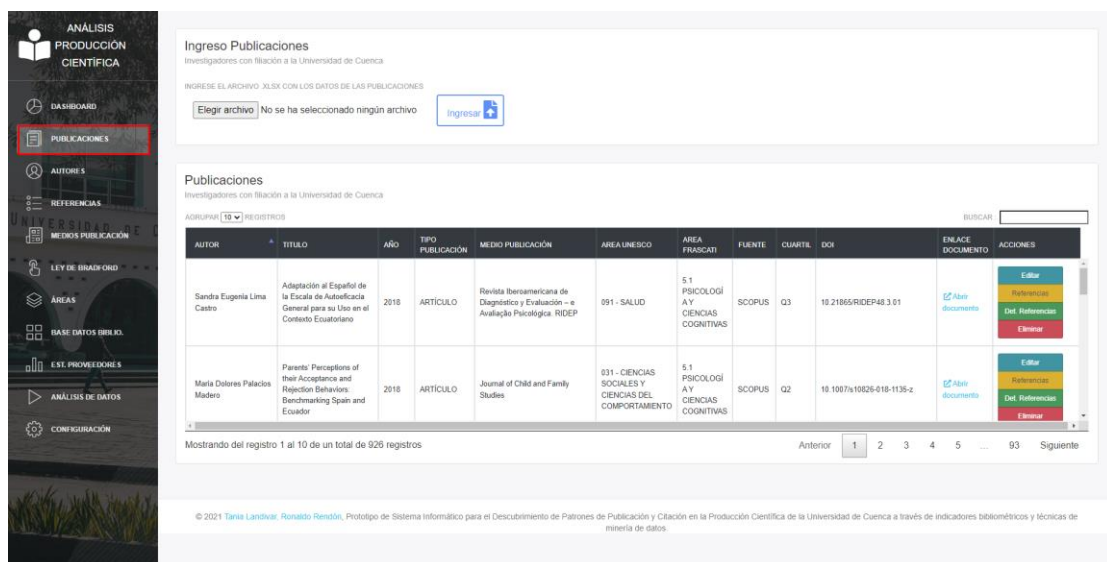


Figura A2. 2 Pantalla para el mantenimiento y visualización de las publicaciones.

Una vez ingresada la publicación se puede editar el detalle de misma, visualizar las referencias, el detalle de las referencias y eliminar la publicación. Opciones que se pueden visualizar en las Figuras A2. 3 y A2. 4.

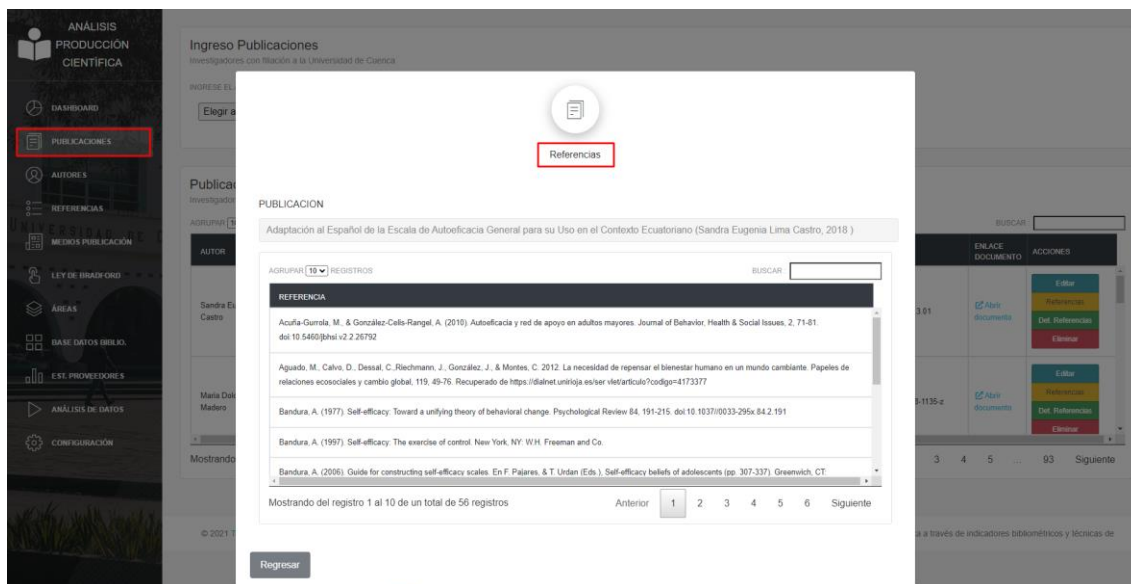


Figura A2. 3 Visualización de las referencias en texto por cada publicación.

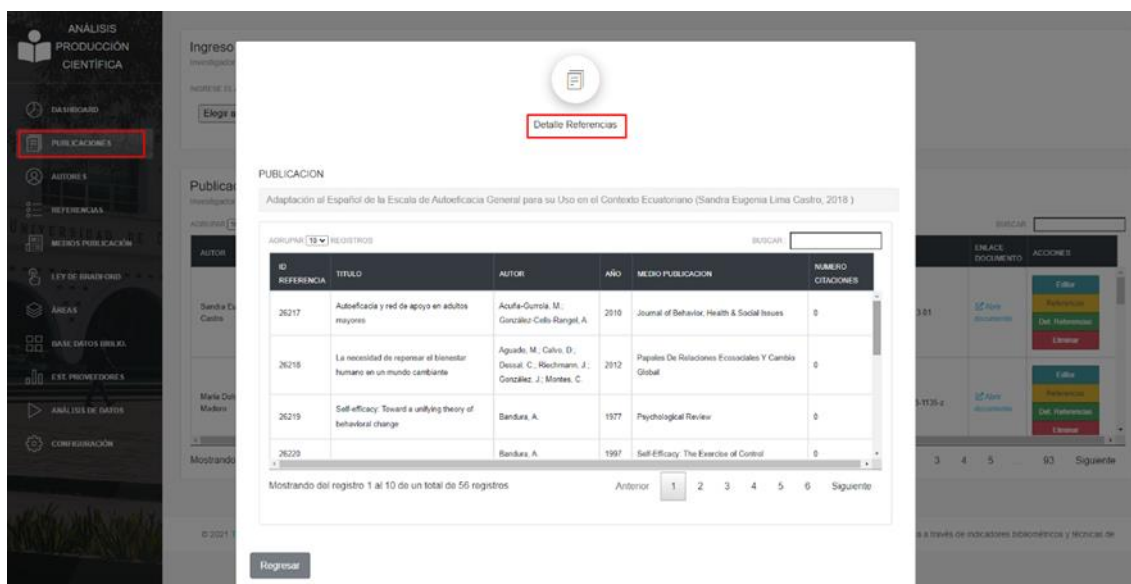
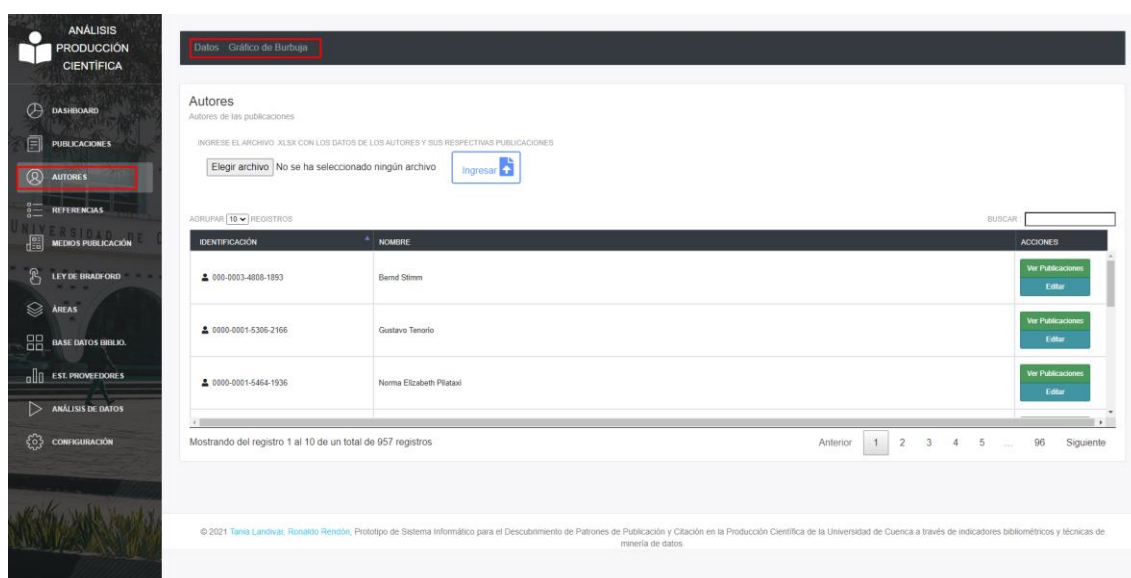


Figura A2. 4 Visualización del detalle de las referencias separada en campos por cada publicación.

Autores

La sección “Autores” se encuentra en la parte izquierda de la pantalla por debajo de la sección “Publicaciones”. El detalle de la pantalla de Autores se puede ver en la Figura A2. 5, en donde se observan dos opciones, “Datos” y “Gráfico de Burbuja”. En la primera se ingresan los autores por artículo a través de un archivo de Excel con los siguientes campos: nombre de la base de datos bibliográfica, título, año de publicación, identificación del autor, orden del autor y nombre del autor.



IDENTIFICACIÓN	NOMBRE	ACCIONES
000-0003-4006-1893	Bened Stemm	Ver Publicaciones Editar
0000-0001-5306-2166	Gustavo Tenorio	Ver Publicaciones Editar
0000-0001-5464-1936	Norma Elizabeth Plataxi	Ver Publicaciones Editar

Figura A2. 5 Pantalla para el mantenimiento de los autores y sus publicaciones

Continuando en esta pantalla, se pueden ver las publicaciones en las que ha participado cada autor perteneciente a la Universidad de Cuenca (Ver la Figura A2. 6).

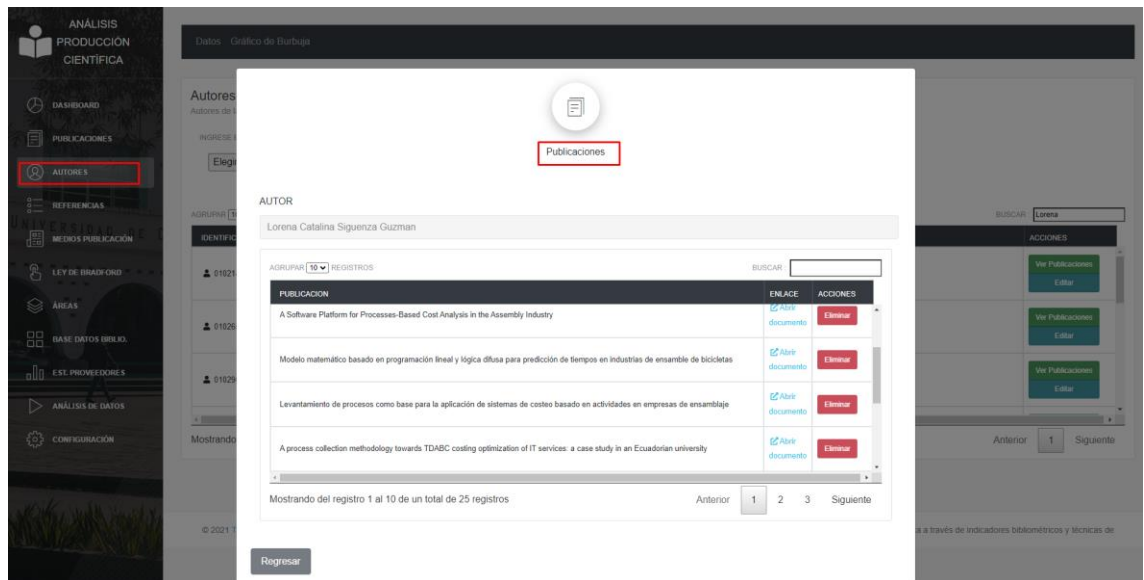


Figura A2. 6 Publicaciones por autor.

En la segunda opción, como se observa en las Figuras A2. 7 y A2. 8, se puede visualizar la información de los autores mediante un gráfico. Por ejemplo, el número de publicaciones por autor y el número de publicaciones según el orden del autor en la publicación.

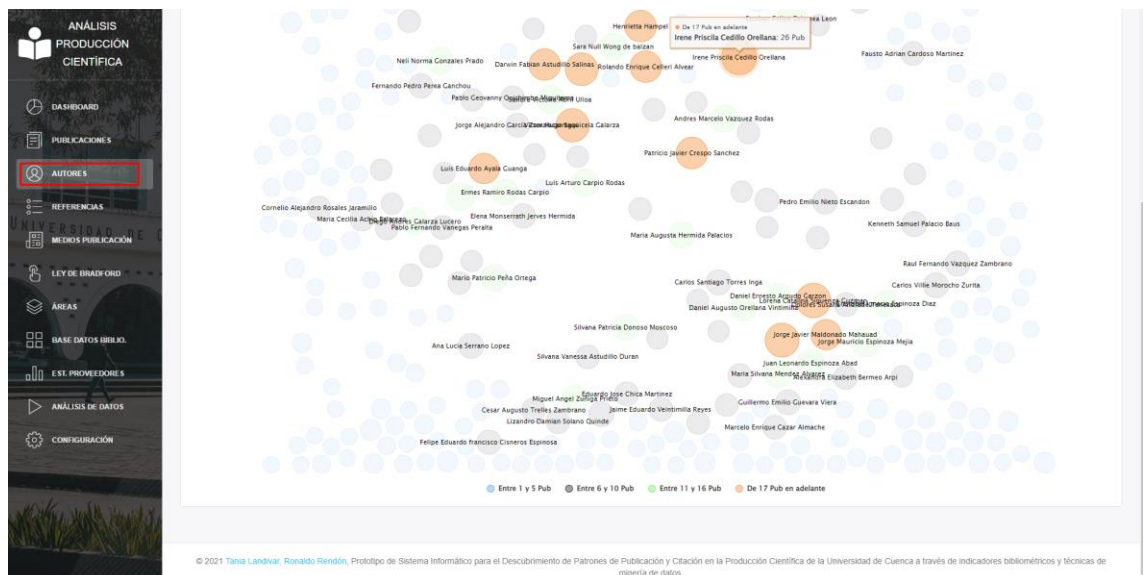


Figura A2. 7 Gráfico de burbuja en función del número de publicaciones por autor.

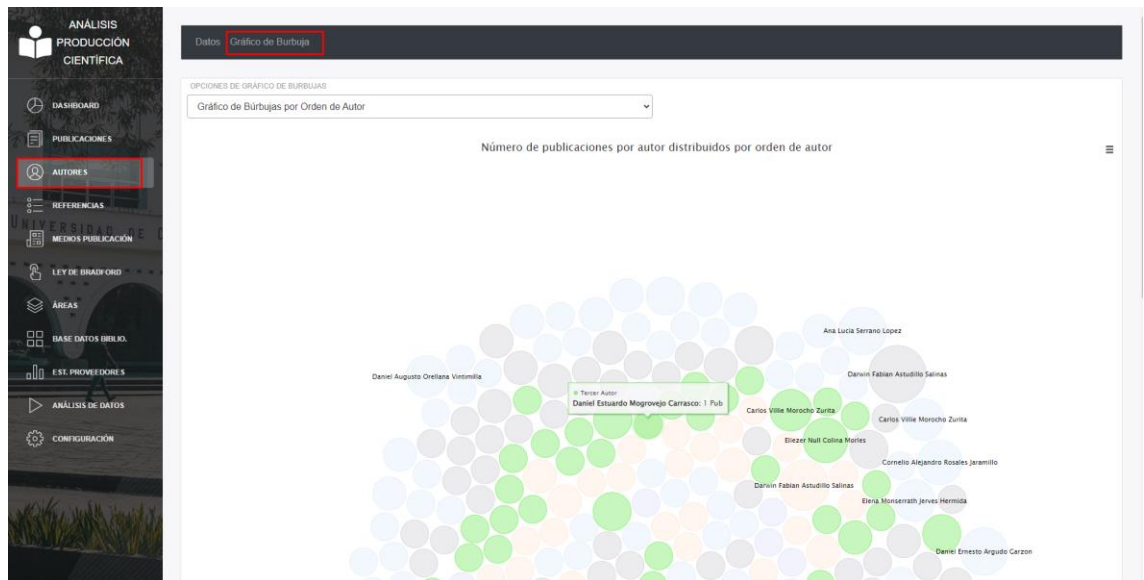


Figura A2. 8 Gráfico de burbuja de acuerdo con el número de publicaciones según el orden del autor en la publicación.

Referencias

La sección “Referencias” se encuentra debajo de la sección “Autores”. Referencias es uno de los enfoques principales para el desarrollo de este prototipo ya que, a través de estas, se buscan obtener los medios de publicación (revistas) que más consume la comunidad universitaria en los trabajos de investigación. La pantalla para el mantenimiento de las referencias se puede visualizar en la Figura A2. 9.

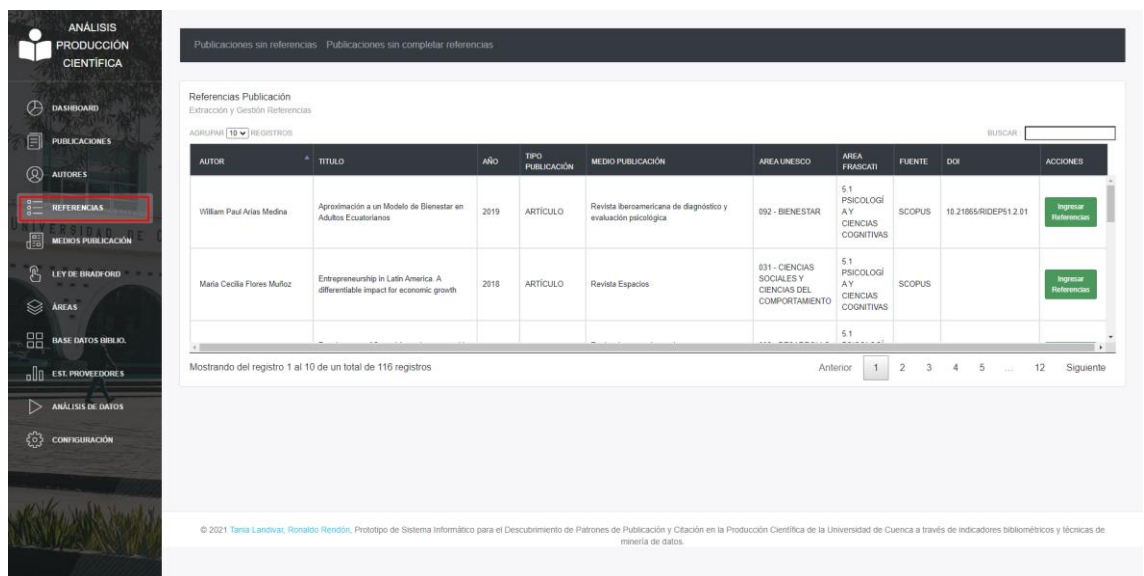


Figura A2. 9 Pantalla para la extracción de las referencias y el detalle de las referencias por publicación.

La primera opción del menú superior visualiza las publicaciones sin referencias. Para las publicaciones indexadas en Scopus de los investigadores con filiación a la Universidad de Cuenca, se utiliza el DOI para obtener las referencias en texto y el detalle de la referencia separada en campos. Para el caso de las publicaciones indexadas en Latindex, se extrae una a una la referencia de forma manual (ver la Figura A2. 10).

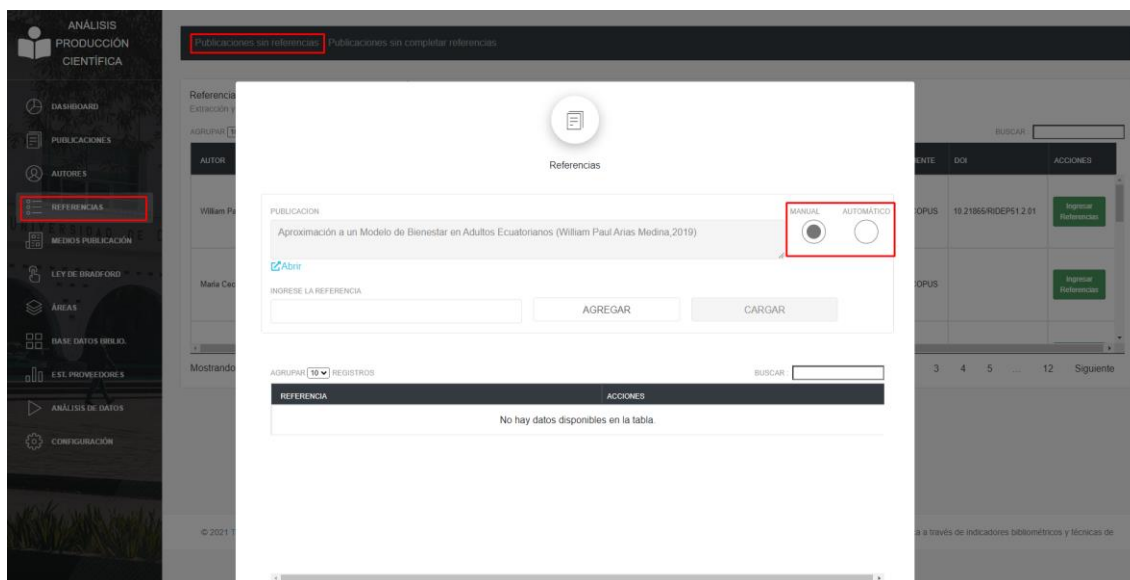


Figura A2. 10 Pantalla para el ingreso de forma manual y automática de las referencias y el detalle de mismas.

En la Figura A2. 11 se presenta el proceso para complementar la extracción del detalle de las referencias, como en el caso de Latindex que no cuenta con una herramienta que permita extraer esta información. Primero se ingresan las referencias una a una en la sección de publicaciones sin referencias, luego en la sección de publicaciones sin completar referencias se busca el detalle de las mismas con la API de Google Académico.

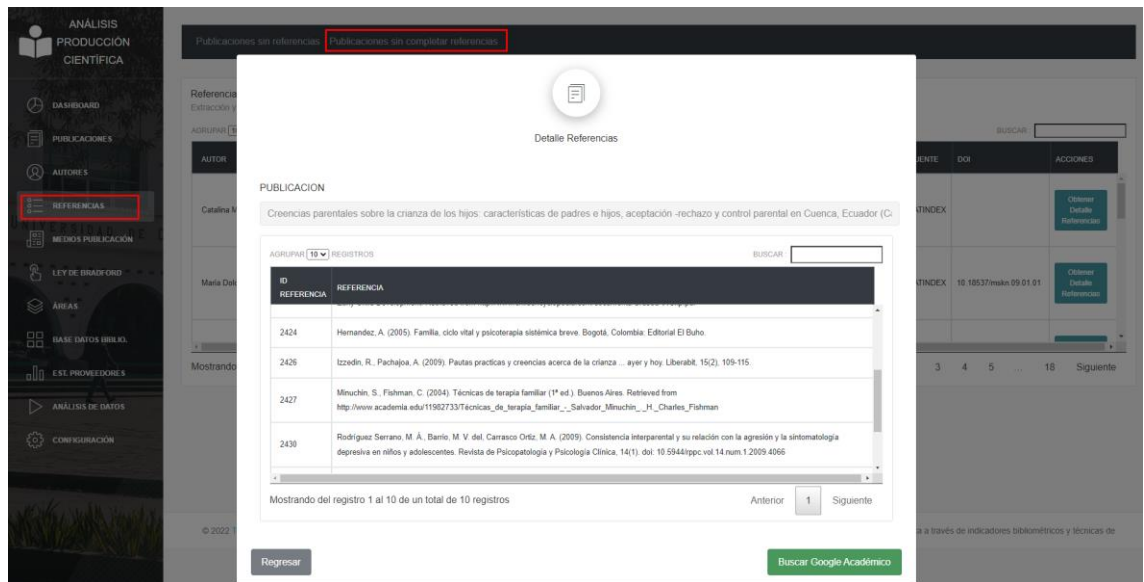


Figura A2. 11 Pantalla para la extracción del detalle de la referencia a través de la API de Google Académico.

Medios de publicación

La opción “Medios de Publicación” que se encuentra debajo de la sección “Referencias”, permite el mantenimiento de los medios enlazados a las publicaciones propias de los investigadores con filiación a la Universidad de Cuenca. En la Figura A2. 12 se observa la pantalla principal de esta sección.

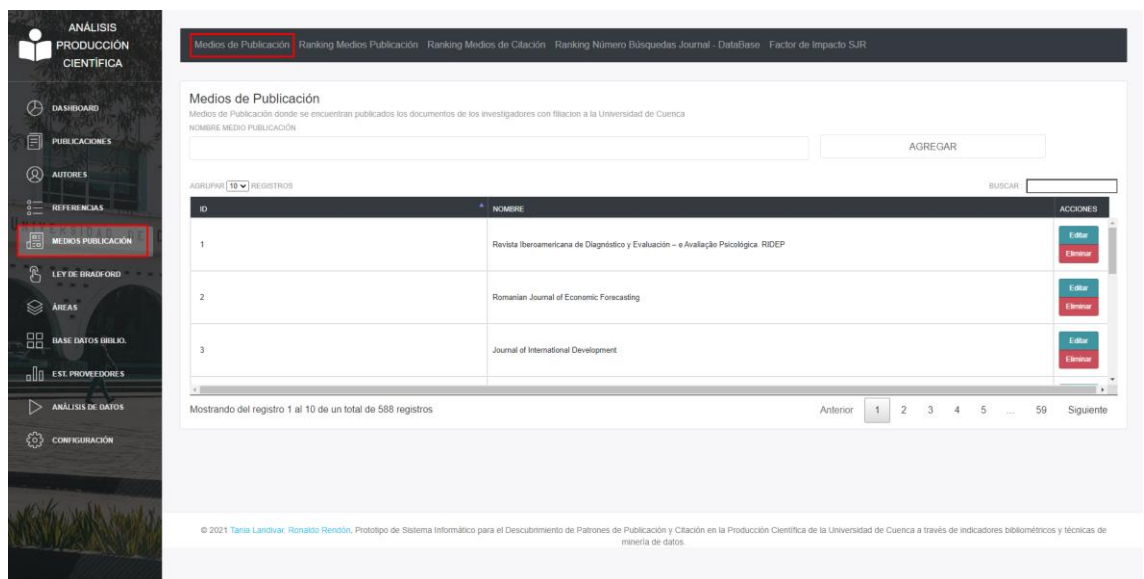
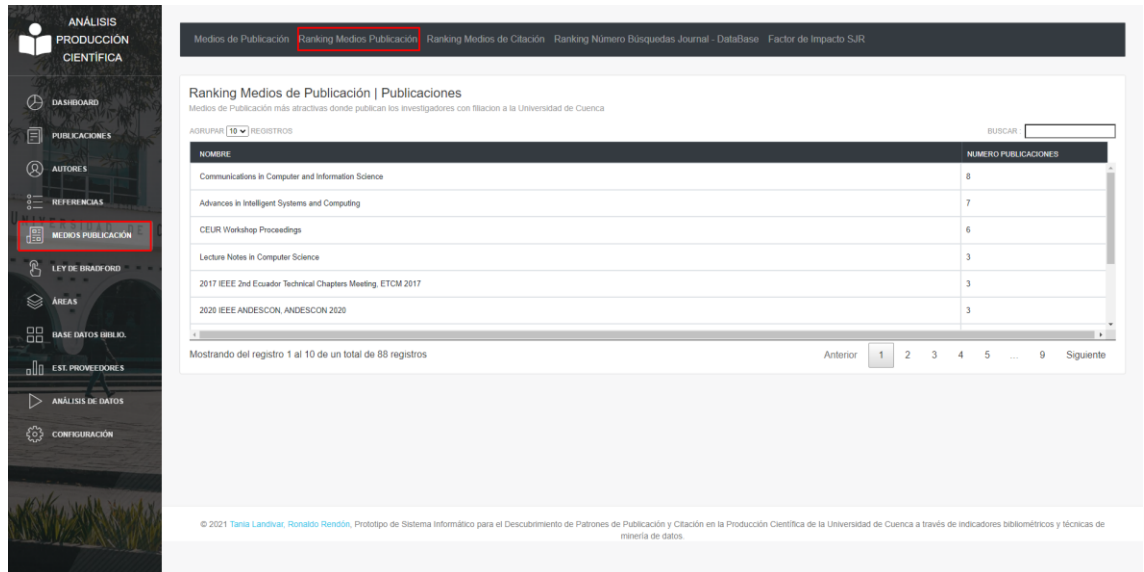


Figura A2. 12 Pantalla Medios Publicación opción medios de publicación.

La opción “Ranking Medios Publicación” como se observa en la Figura A2. 13, permite visualizar los medios de publicación más destacados en donde

publican los investigadores con filiación a la Universidad de Cuenca. Este apartado posibilita el filtrado de cualquier medio de publicación que se desee conocer mediante un cuadro de búsqueda, así como también una tabla con el nombre de los medios de publicación ordenados de mayor a menor de acuerdo con el número de publicaciones.



NOMBRE	NUMERO PUBLICACIONES
Communications in Computer and Information Science	8
Advances in Intelligent Systems and Computing	7
CEUR Workshop Proceedings	6
Lecture Notes in Computer Science	3
2017 IEEE 2nd Ecuador Technical Chapters Meeting, ETCM 2017	3
2020 IEEE ANDESCON, ANDESCON 2020	3

Figura A2. 13 Pantalla Medios Publicación opción ranking medios publicación.

La Figura A2. 14 presenta la siguiente opción “Ranking Medios Citación” en donde se puede ver los medios de publicación más relevantes de donde se basan los investigadores para sus citas. Al igual que la opción anterior, esta permite realizar una búsqueda y muestra la información en orden descendente de acuerdo con el número de citas por cada medio de publicación.

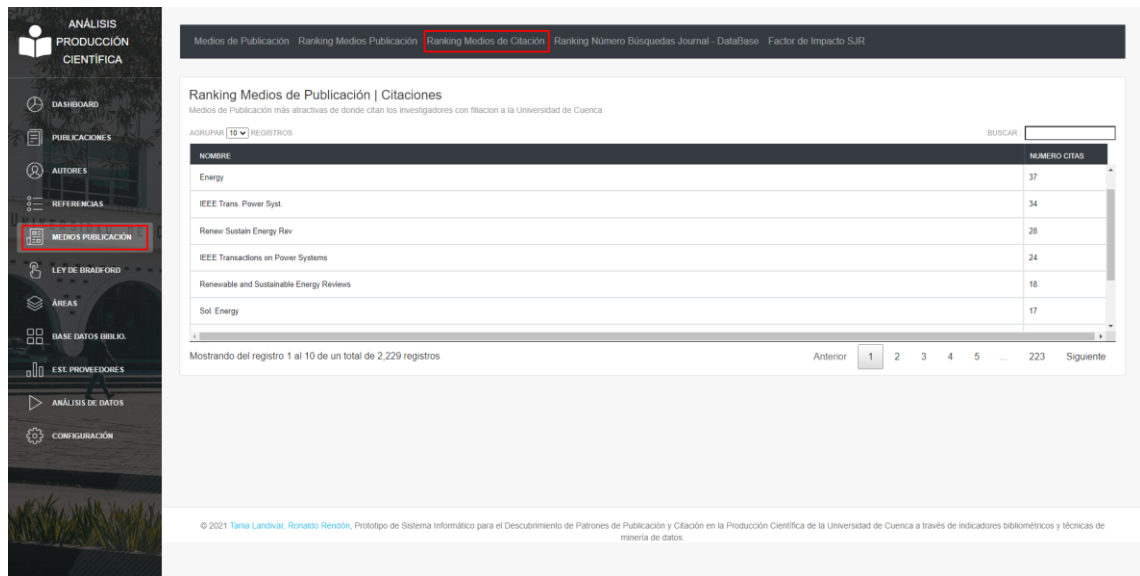


Figura A2. 14 Pantalla Medios Publicación opción ranking medios citación.

La siguiente opción “Ranking Número Búsquedas Journal - DataBase”, al dar clic sobre esta, se muestra la información sobre el número de búsquedas que tiene una revista académica en particular que se encuentra alojada en un repositorio bibliográfico. La tabla está organizada de mayor a menor número de búsquedas como se observa en la Figura A2. 15.

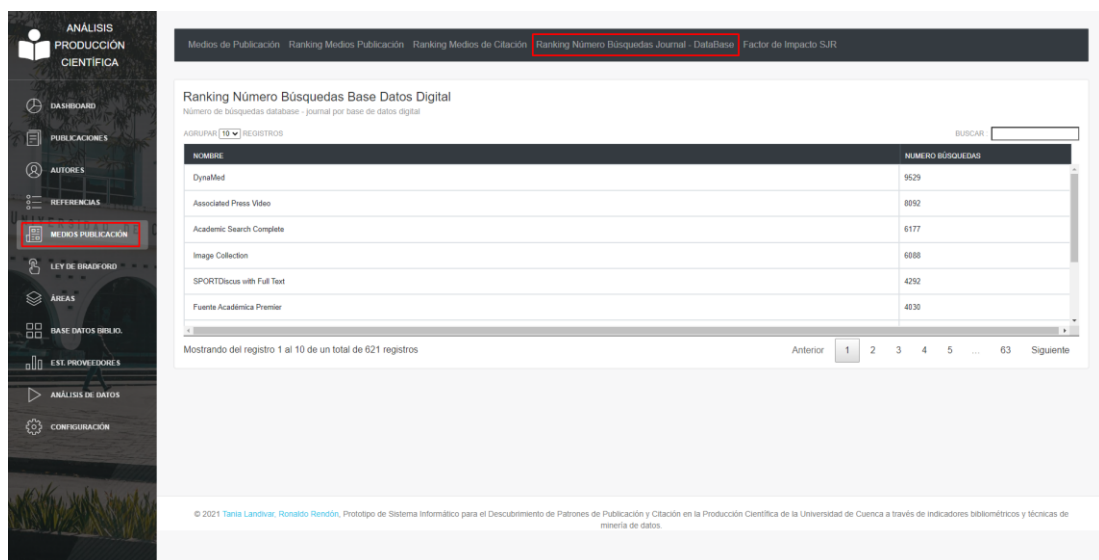


Figura A2. 15 Pantalla Medios Publicación opción ranking búsquedas journal-database.

La última opción de la sección “Medios de Publicación” denominada “Factor de Impacto SJR” corresponde a los datos que proporciona SCImago Journal Rank acerca de la influencia científica de las revistas académicas según el número de citas en otros medios y periódicos o revistas de importancia. Estos datos se

pueden cargar a través de un archivo de Excel mediante el botón “Seleccionar archivo” como se muestra en la Figura A2. 16.

The screenshot shows a web application interface for 'Medios de Publicación'. The left sidebar contains navigation options: ANÁLISIS PRODUCCIÓN CIENTÍFICA, DASHBOARD, PUBLICACIONES, AUTORES, REFERENCIAS, MEDIOS PUBLICACIÓN (highlighted), LEY DE BRADFORD, ÁREAS, BASE DATOS BIBLIÓ., EST. PROVEEDORES, ANÁLISIS DE DATOS, and CONFIGURACIÓN. The main content area is titled 'Ranking | Scimago Journal And Country Rank' and includes a 'Factor de Impacto SJR' tab. Below the title, there is a section for uploading an Excel file with a 'Elegir archivo' button and an 'Ingresar' button. A table displays the following data:

RANK	ID RECURSO	TÍTULO	TIPO	ISSN	SJR	SJR BEST QUARTILE
1	26773	Ca-A Cancer Journal for Clinicians	Journal	15424863, 00079235	62.937	Q1
2	19434	MMWR Recommendations and Reports	Journal	10575967, 15458601	40.949	Q1
3	20315	Nature Reviews Molecular Cell Biology	Journal	14710072, 14710080	37.461	Q1
4	29431	Quarterly Journal of Economics	Journal	00335533, 15314650	34.573	Q1
5	21100812243	Nature Reviews Materials	Journal	20568437	32.011	Q1
6	58530	National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System	Journal	15518922, 15518930	28.083	Q1

Below the table, it indicates 'Mostrando del registro 1 al 10 de un total de 32,952 registros' and includes pagination controls for 'Anterior', '1', '2', '3', '4', '5', '...', '3296', and 'Siguiente'. At the bottom, a copyright notice reads: '© 2021 Tania Landívar, Ronaldo Rendón, Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos.'

Figura A2. 16 Pantalla Medios Publicación opción factor de impacto SJR.

Áreas

Dar clic en la opción “Áreas” que se encuentra debajo de la sección “Ley de Bradford”. En esta sección se realiza el mantenimiento de las áreas Unesco, Frascati y SJR, que son utilizadas por el CDRJBV (Centro de Documentación Regional “Juan Bautista Vázquez”) de la Universidad de Cuenca. El propósito de esta pantalla es listar, crear y eliminar cada una de las áreas pertenecientes a Unesco, Frascati y SJR. Esto se puede visualizar en la Figura A2. 17.

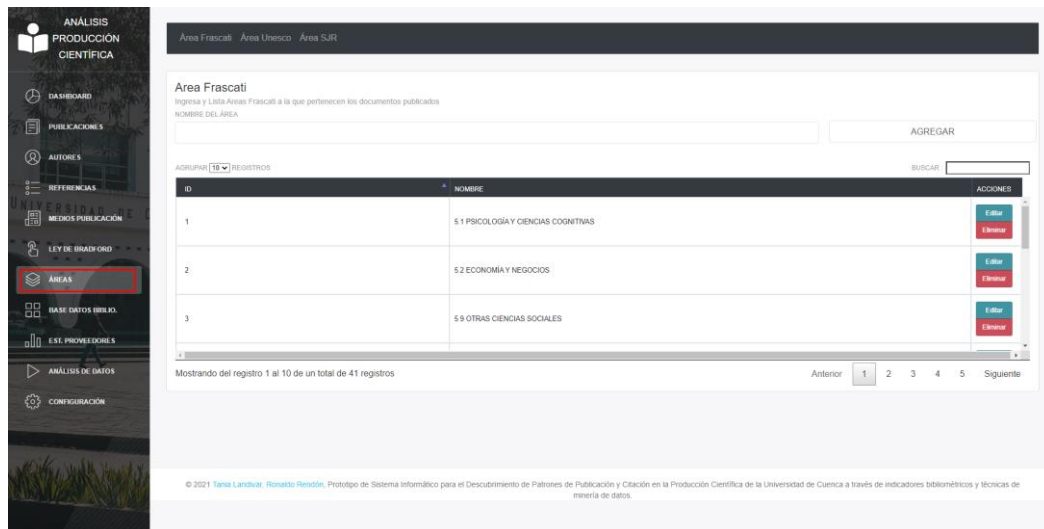


Figura A2. 17 Pantalla para visualizar la sección de áreas.

Bases de datos bibliográficas

En la Figura A2. 18 se presenta la sección donde realiza el mantenimiento de las bases de datos bibliográficas a las que se suscribe el CDRJBV de la Universidad de Cuenca, con el propósito de listar, crear, editar y eliminar cada una de las bases bibliográficas. Para acceder a la misma, se debe presionar sobre la opción “Base Datos Biblio” que se encuentra en la parte inferior izquierda de la pantalla, debajo de la sección “Areas”.

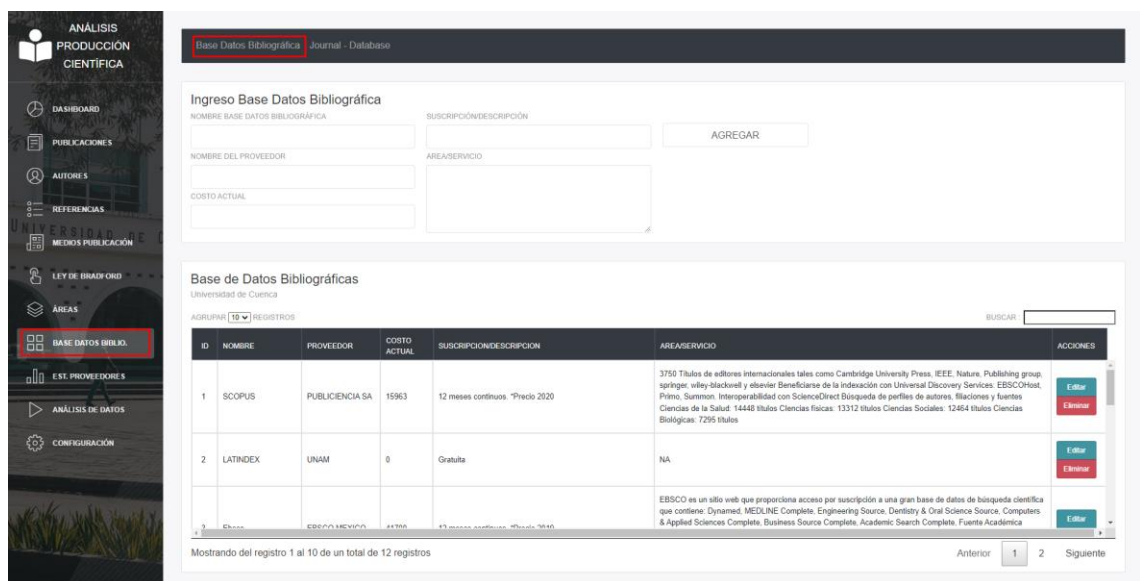


Figura A2. 18 Pantalla para la sección de mantenimiento de las bases de datos bibliográficas.

También, permite registrar los medios de publicación. Por ejemplo, las revistas que están habilitadas para su uso por cada base de datos bibliográfica. Los proveedores proporcionan una lista de las revistas por base de datos y es ingresada a través de un archivo de Excel, ver la Figura A2. 19.

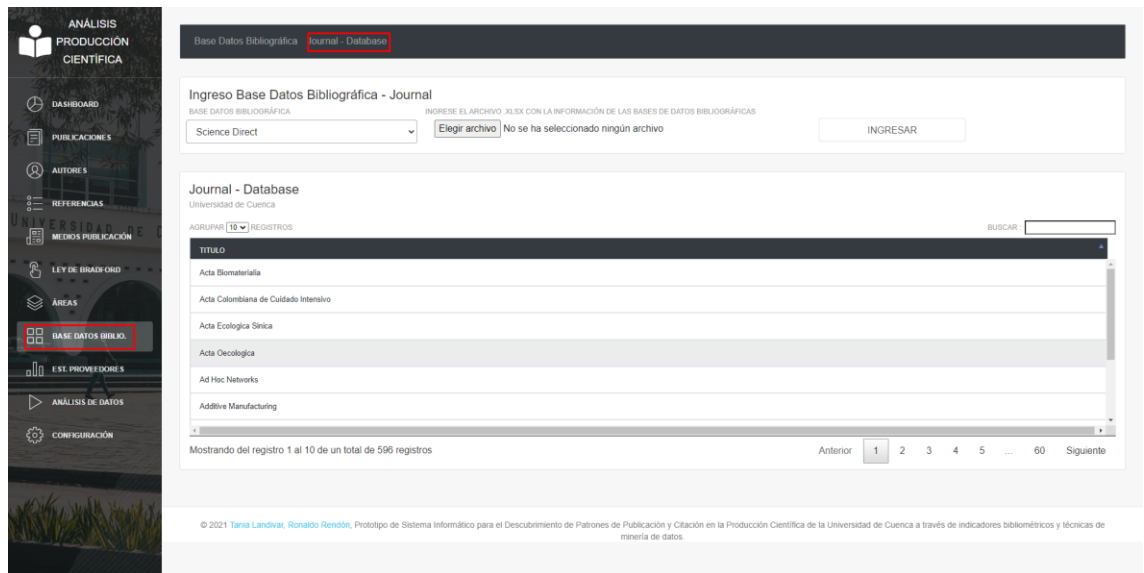


Figura A2. 19 Pantalla para el ingreso de medios de publicación como revistas que proporcionan las bases de datos bibliográficas

Estadísticas de los proveedores

La sección “Estadística de los proveedores” se encuentra debajo de la sección “Bases de datos bibliográficas” y se accede al dar clic sobre la misma. En esta sección, se tiene como referencia las estadísticas proporcionadas por los proveedores del CDRJBV, que permite ingresar las estadísticas de uso de las bases de datos bibliográficas, o en su defecto de una revista o medio de publicación que corresponde a estas bases de datos. Además, en la opción de gráfico de la sección se puede visualizar una gráfica de las estadísticas de búsqueda por mes y año de una determinada base de datos bibliográfica o revista, como se observa en las Figuras A2. 20, A2. 21 y A2. 22.

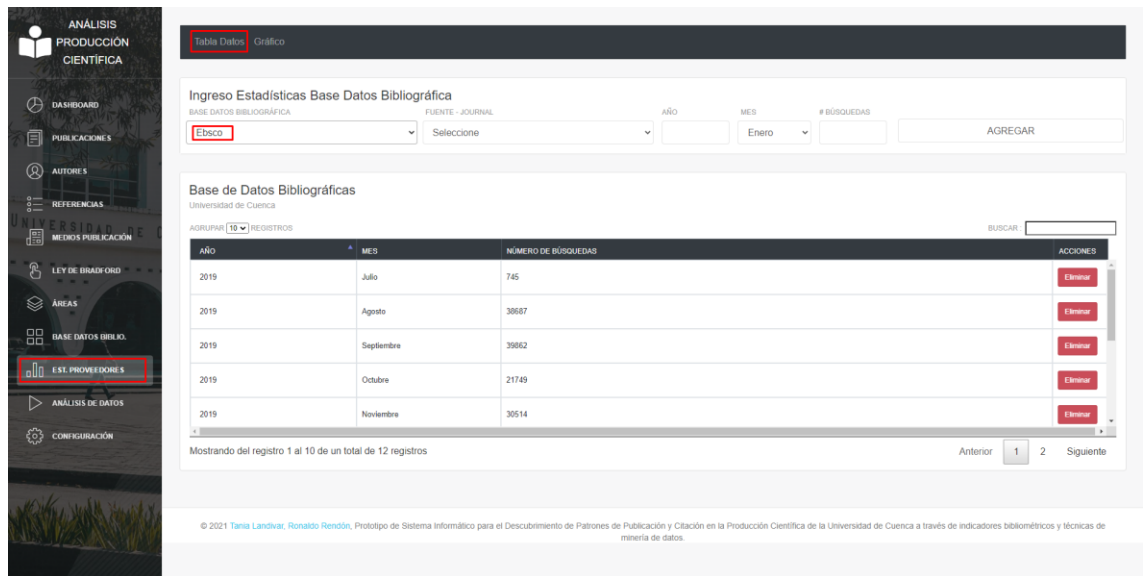


Figura A2. 20 Pantalla para el ingreso de las estadísticas de uso de las bases de datos bibliográficas y sus revistas o fuentes bibliográficas.

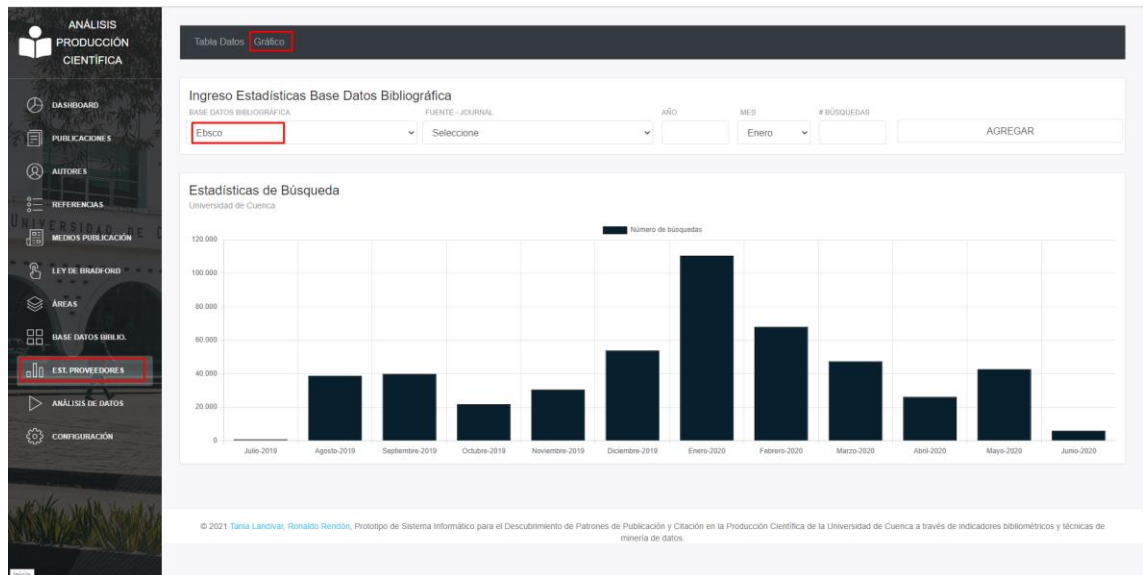


Figura A2. 21 Gráfico de barras de la estadística de búsqueda de las bases de datos bibliográficas.

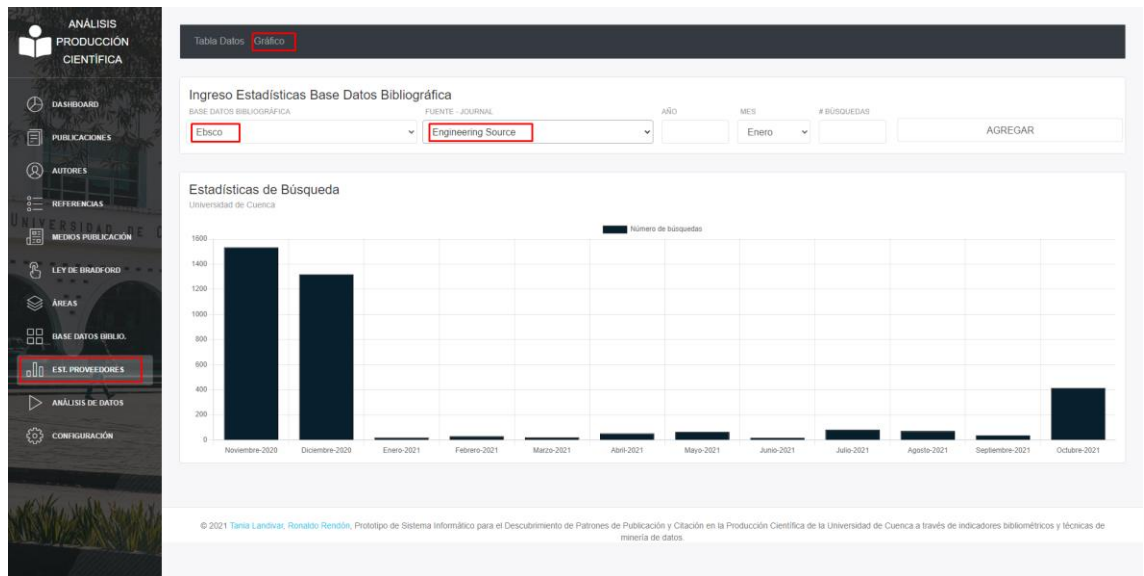


Figura A2. 22 Gráfico de barras de la estadística de búsqueda de las fuentes bibliográficas indexadas en las bases de datos.

Configuración

La última sección denominada “Configuración” se encuentra en la parte inferior izquierda de la pantalla. Esta sección muestra características específicas del sistema necesarias para el cálculo en el apartado de la Ley de Bradford, en donde se permite ingresar un nombre y un valor de parámetro o la edición o eliminación de los mismos. También se visualizan las opciones de Equivalencia donde se muestra la relación entre áreas Unesco, Frascati y SJR, como se observa en la Figura A2. 23.

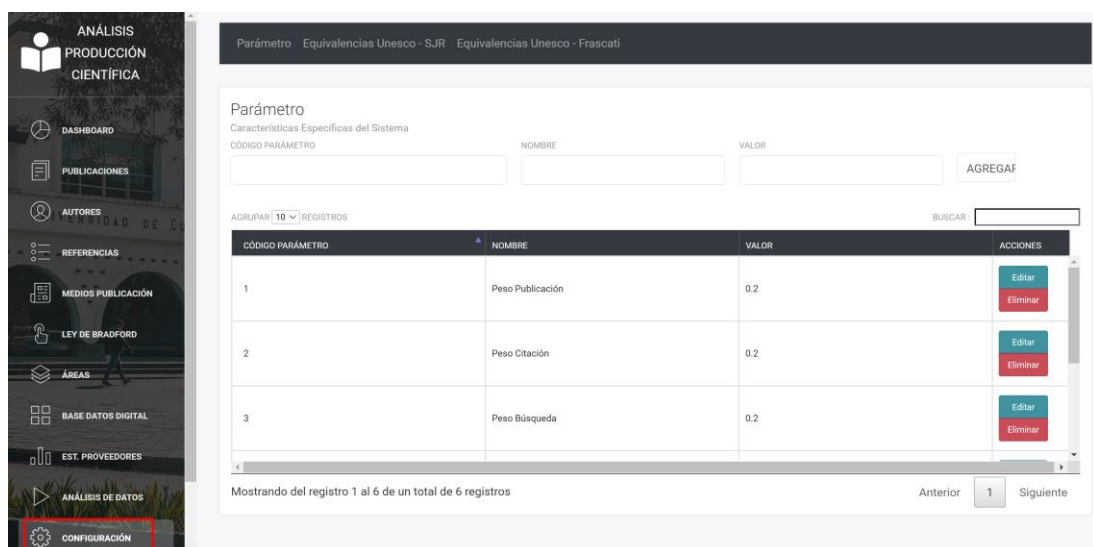
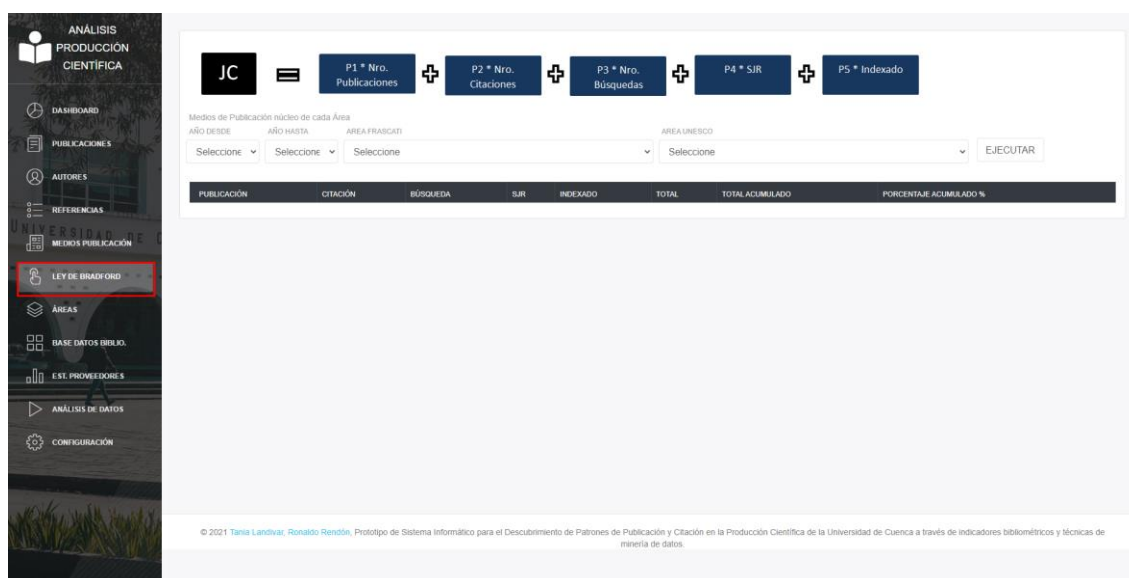


Figura A2. 23 Sección configuración.

Ley de Bradford

Para acceder a esta sección se debe dar clic en el apartado “Ley de Bradford” que se encuentra en la parte centro-izquierda de la pantalla. En esta sección se realiza un análisis de las revistas o fuentes bibliográficas, de acuerdo con los siguientes indicadores bibliométricos, la más publicada, la más citada, con el mayor número de búsquedas y factor de impacto. Se evalúa a la revista en base a los indicadores antes mencionados y se le asigna un peso a cada uno de ellos, para finalmente calificar desde diferentes puntos de vista a la revista. Adicionalmente, se establece si la revista que tiene mayor calificación se encuentra indexada en los repositorios bibliométricos ya contratados por la Universidad de Cuenca. Dejando a consideración si los recursos contratados están cubriendo las necesidades de información de los usuarios.

La sección de la Ley de Bradford permite realizar el análisis de acuerdo con el área Unesco o área Frascati en un periodo de tiempo, ver la Figura A2. 24. Se presenta en una tabla los indicadores bibliométricos con sus pesos y resultados finales, siendo las revistas que tienen mayor puntuación las más importantes (ver la Figura A2. 25).



ANÁLISIS PRODUCCIÓN CIENTÍFICA

DASHBOARD

PUBLICACIONES

AUTORES

REFERENCIAS

MEJORES PUBLICACIONES

LEY DE BRADFORD

ÁREAS

BASE DATOS BIBLIJ.

EST. PROVEEDORES

ANÁLISIS DE DATOS

CONFIGURACIÓN

JC = P1 * Nro. Publicaciones + P2 * Nro. Citaciones + P3 * Nro. Búsquedas + P4 * SIR + P5 * Indexado

Medios de Publicación núcleo de cada Área

AÑO DESDE: Seleccione AÑO HASTA: Seleccione ÁREA FRASCATI: Seleccione ÁREA UNESCO: Seleccione EJECUTAR

PUBLICACIÓN	CITACIÓN	BÚSQUEDA	SIR	INDEXADO	TOTAL	TOTAL ACUMULADO	PORCENTAJE ACUMULADO %
-------------	----------	----------	-----	----------	-------	-----------------	------------------------

© 2021 Tania Landívar, Ronaldo Rendón, Prototipo de Sistema Informático para el Descubrimiento de Patrones de Publicación y Citación en la Producción Científica de la Universidad de Cuenca a través de indicadores bibliométricos y técnicas de minería de datos.

Figura A2. 24 Sección para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto.

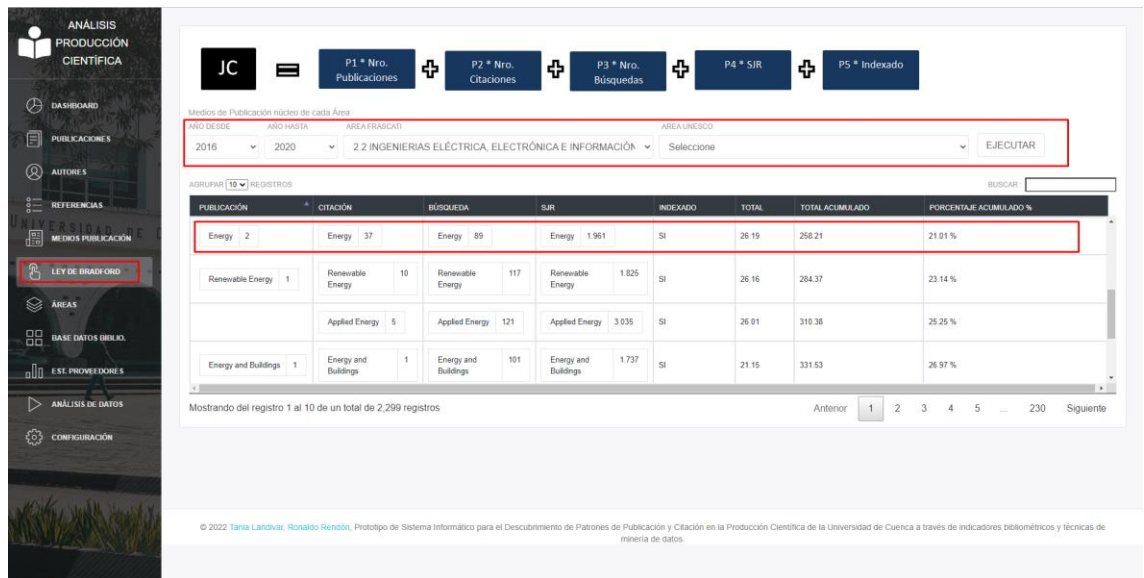


Figura A2. 25 Tabla de resultados para la evaluación de la revista en función del número de publicaciones, número de citas, número de búsquedas y factor de impacto.

Análisis de datos

La penúltima sección “Análisis de datos” se encuentra por debajo de la sección “Estadísticas de los proveedores”. Dentro de esta sección se hace referencia a toda la información que contiene cada una de las publicaciones, con el fin de poder visualizar el tratamiento de los datos y análisis estadístico. En esta sección se encuentran dos opciones, Estadísticas y Predicción de los datos, las cuales se pueden ver en la Figura A2. 26.

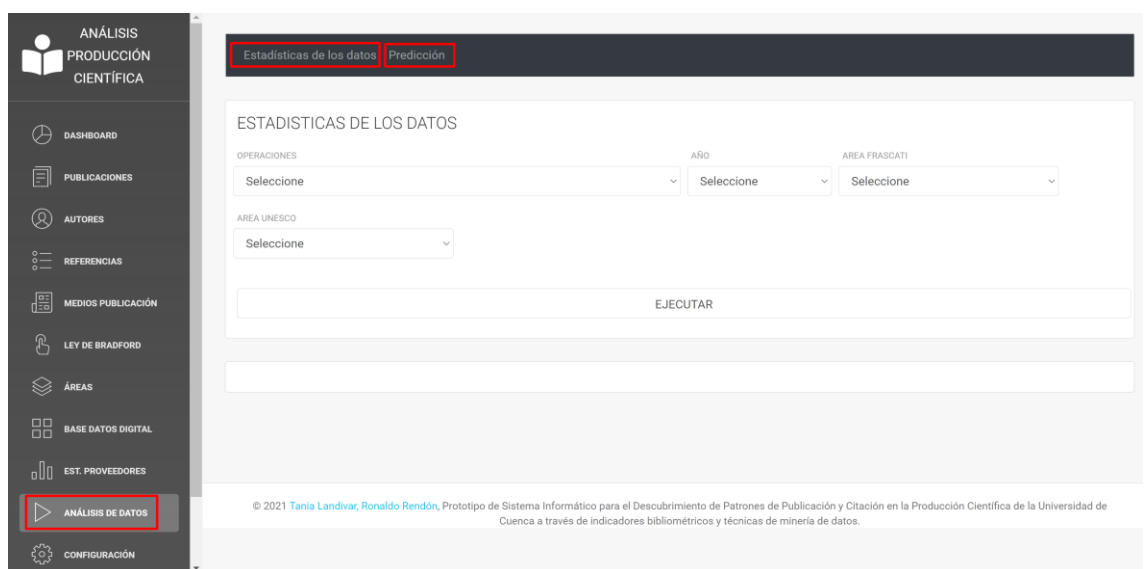


Figura A2. 26 Sección análisis de datos.



Anexo 3: Manual Técnico

Índice

Índice	142
Índice de figuras	143
Presentación	145
Objetivo.....	145
Requisitos del sistema	145
Herramientas utilizadas para el desarrollo	145
Instalación de aplicaciones	147
Modelo de clases	148
Diagrama Entidad – Relación.....	150
Diccionario de datos	152
Servicios web implementados en el desarrollo del sistema	165
Bibliografía	109



Índice de figuras

Figura A3. 1. Entidades y operaciones que interactúan en el prototipo de sistema 149

Figura A3. 2. Diagrama entidad – relación (Modelo lógico del Prototipo). 151



Índice de tablas

Tabla A3. 1.Servicios Web para la Sección Publicaciones.	166
Tabla A3. 2.Servicios Web para el mantenimiento de los autores.	167
Tabla A3. 3.Servicios Web para la extracción y almacenamiento de las referencias y detalle de las referencias.	168
Tabla A3. 4.Servicios web para el ingreso y mantenimiento de las estadísticas de uso de las bases de datos bibliográficas y sus fuentes.	169
Tabla A3. 5.Servicios web para la sección análisis de datos	170
Tabla A3. 6.Servicios web para la evaluación de las revistas a través de la ley de Bradford.	172
Tabla A3. 7.Servicios Web para el mantenimiento de los Medios Publicación.	173
Tabla A3. 8.Servicios Web Ranking Medios de Publicación.	174
Tabla A3. 9.Servicios Web Ranking Medios de Citación.	175
Tabla A3. 10.Servicios Web Número Búsquedas Base Datos Bibliográfico.	176
Tabla A3. 11.Servicios web para el mantenimiento de las áreas	177
Tabla A3. 12.Servicios web para la sección del mantenimiento de las bases de datos bibliográficas y el ingreso de sus revistas.....	178
Tabla A3. 13.Servicios web para la sección configuración	179



Presentación

En este manual se guiará a los usuarios que deseen realizar mantenimiento del sistema, se detallan los requerimientos y la estructura del aplicativo, también se especifican las herramientas necesarias para la construcción y la funcionalidad del mismo.

Objetivo

Especificar los principales componentes de la estructura y el desarrollo del sistema, con el fin de que se pueda realizar soporte, modificaciones y actualizaciones que sean necesarias.

Requisitos del sistema

- Requerimientos de hardware
 - Equipo, teclado, mouse, monitor.
 - Memoria RAM 4GB
 - Tarjeta de red LAN y/O Wireless
 - Procesador 1.4 GHz.
- Requerimientos de software
 - Sistema operativo (Windows 8 en adelante)
 - Acceso a Internet
 - Navegador Chrome versión 96+
 - Navegador Mozilla Firefox versión 96+
- Requerimientos de red
 - Estar conectado a la red universitaria para hacer uso de la API de Google Académico y Scopus.

Herramientas utilizadas para el desarrollo

Visual Studio Code (v1.64.2) es el editor de código fuente utilizado para la implementación del prototipo, con los siguientes lenguajes de programación: JavaScript en el Front-end y Python (v3.9.2) en el Back-end.

Tecnologías y Librerías Front-End:



- React(v17.0.1): Biblioteca de JavaScript que permite crear interfaces de forma sencilla, se basa en componentes que son reutilizables y fáciles de mantener (desarrolloweb.com, n.d.).
- Node.js (v12.18.3): Entorno de ejecución para JavaScript.
- D3 (v7.0.0): Biblioteca de Java Script para la creación y manipulación de interfaces para la visualización interactiva de datos con transiciones, animaciones, etc. (LEMONCODE, 2016).
- Highcharts (v9.3.3): librería de software para gráficos escrita en JavaScript puro.

Tecnologías y Librerías Back-End:

- flask (v2.0.x): Es un marco web de Python que permite crear rápidamente aplicaciones con un mínimo número de líneas de código.
- flask_sqlalchemy (v2.0.x): Maneja el modelo de datos de la aplicación.
- Pymysql (v1.0.2): Permite la interacción con base de datos mysql.
- marshmallow_sqlalchemy (v2.0.x): Biblioteca de serialización y deserialización de objetos.
- Scholarly (v1.5.1): Módulo que permite recuperar información sobre el autor y las publicaciones de Google Académico.
- pybliometrics.scopus: Librería que permite recuperar información de los artículos indexados en Scopus a través del identificador de objeto digital (DOI).
- scraperapi-sdk (v1.0.6): Protege el rastreador web contra solicitudes bloqueadas, fallas proxy, fugas IP, fallas de navegador y CAPTCHA.
- flask_cors (v3.0.10): Habilita los cors que permiten el acceso a los recursos a través de los servicios RESTful, es decir, se trata de una medida de seguridad entre el consumo de los recursos desde el Front-end al Back-end.
- Numpy (v1.22.2): Crea vectores y matrices grandes multidimensionales junto con una gran colección de funciones matemáticas de alto nivel.
- Pandas (v1.4.1): Herramienta de manipulación y análisis de datos.
- hermetrics levenshtein (v0.1.13): Librería para métricas de distancia y similitud.
- Pmdarima (v1.8.4): Librería estadística diseñada para llenar el vacío de las capacidades de análisis de series temporales de Python.

Tecnologías Base de Datos:



- MySQL (v5.7): Gestor de base de datos para el almacenamiento de las publicaciones, referencias, bases de datos bibliográficas y estadísticas de uso de las fuentes bibliográficas proporcionadas por los proveedores con sus respectivos atributos y relaciones.

API (Interfaz de programación de aplicaciones):

- Google Académico: El API de Google Académico denominado Google Scholar permite extraer metadatos mediante un token. La información que se extrae con el api es la siguiente: título, autor, año de publicación, número de citas, revista, URL de la publicación, medio de publicación y abstract.
- Scopus API: El API de Elsevier denominado Scopus Search API permite extraer metadatos como el nombre de la revista o medio de publicación donde se encuentra la publicación referenciada.

Instalación de aplicaciones

1. Descargar e instalar Python y Node.js en el equipo.
2. Descargar e instalar Visual Studio Code, luego abrir el IDE y agregar en extensiones Python.
3. Abrir el proyecto del Back-end en Visual Studio Code. Luego instalar las librerías mencionadas en “Tecnologías y Librerías Back-End” haciendo uso del siguiente comando:

```
pip install {nombre de la librería}
```

Ejemplo:

```
pip install hermetrics levenshtein==0.1.13
```

4. Abrir el proyecto del Front-end en Visual Studio Code y ejecutar el siguiente comando:

```
npm install
```

Este proyecto cuenta con un archivo llamado package.json que se encarga de descargar automáticamente todas las librerías necesarias para que se ejecute el Front-end.

5. Instalar MySQL y luego ejecutar el script de la base de datos sobre su administrador, con el fin de que se importe los datos de las entidades y sus relaciones. La base importada tiene el nombre de “bd_bibliometria”.



6. Una vez importada la base de datos, en el Front-end se debe ejecutar el siguiente comando:

npm start

Y, para el Back-end:

flask run

Finalizado estos pasos puede hacer uso del sistema, no olvide que debe estar conectada a la red de la Universidad para que pueda hacer uso de las API de Google Académico y Scopus.

Modelo de clases

En la Figura A3. 1 se puede observar el diagrama que representa cada una de las entidades que interactúan en el sistema con sus correspondientes operaciones. La entidad principal es el artículo que gestiona las operaciones para el ingreso de las publicaciones de los autores con filiación a la Universidad de Cuenca, además de sus conexiones con las áreas, categorías, revistas, medios de publicación y base de datos bibliográficas. Estas permiten tener a detalle las características de cada una de las publicaciones y fuentes bibliográficas (revistas).

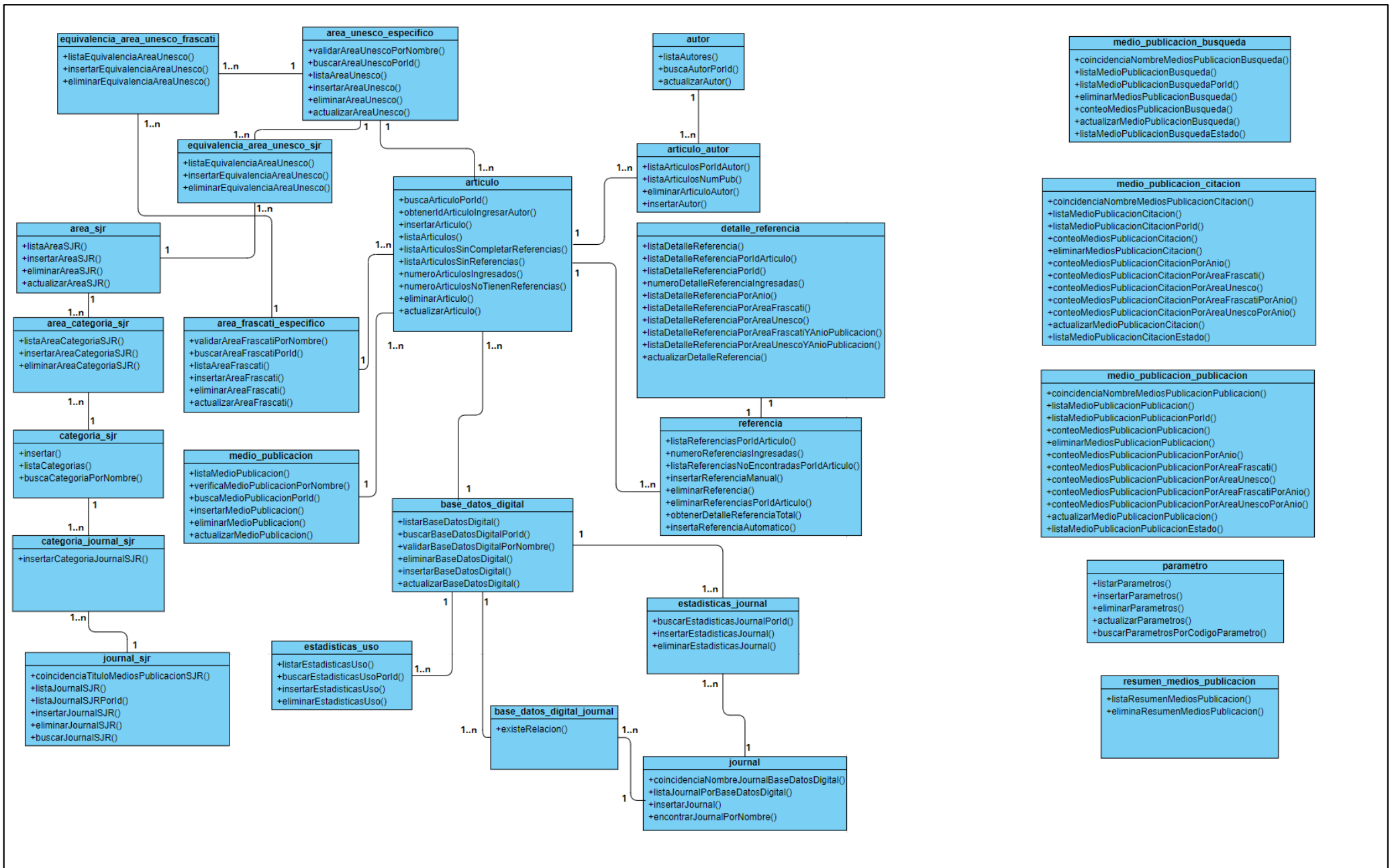


Figura A3. 1 Entidades y operaciones que interactúan en el prototipo de sistema

Tania Maricela Landívar Ordóñez
 Ronaldo Sebastián Rendón Loja



Diagrama Entidad – Relación

A continuación, la Figura A3. 2 presenta las entidades consideradas para el diseño de la base de datos, partiendo desde la entidad principal “artículo” se detallan los atributos que caracterizan de cada una de las entidades, así como también se establecen las relaciones de cardinalidad entre las mismas.

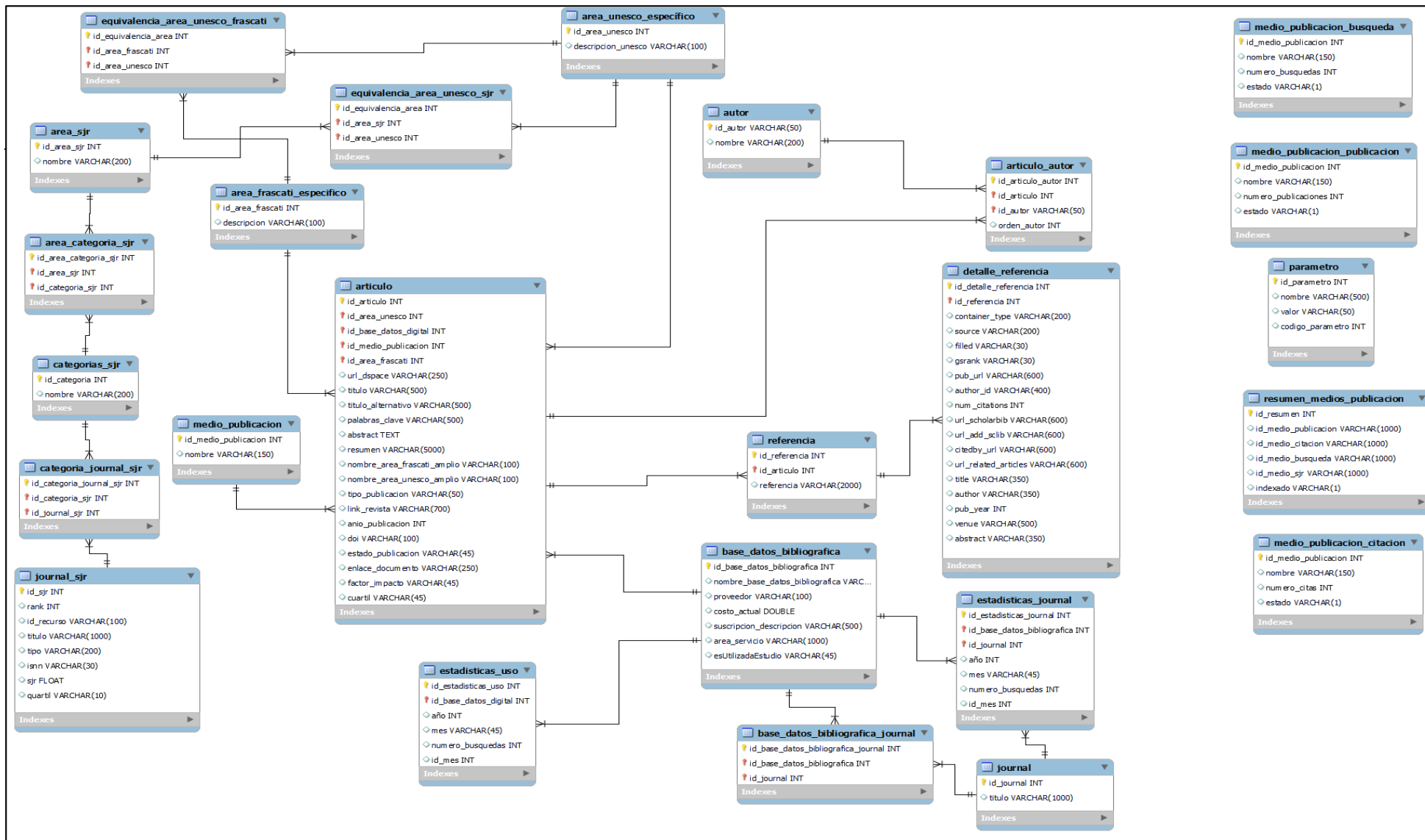


Figura A3. 2 Diagrama entidad – relación (Modelo lógico del Prototipo).

Tania Maricela Landívar Ordóñez
 Ronaldo Sebastián Rendón Loja



Diccionario de datos

Para el almacenamiento de datos del software, se definen los campos necesarios para cada una de las entidades relacionadas con el prototipo de sistema.

Diccionario de Datos Modelo Artículo

Nombre:	articulo	Versión:	1.0	
Descripción:	Tabla de artículos			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_articulo	int	11	Id del artículo consecutivo, autoincremental	Clave
id_base_datos_digital	int	11	Id de la base de datos digital	Clave Foránea
url_dspace	varchar	250	Url donde se encuentra el artículo en el dspace	
titulo	varchar	500	Título del artículo	
titulo_alternativo	varchar	500	Título alternativo del artículo	
palabras_clave	varchar	500	Palabras clave del artículo	
abstract	text		Abstract del artículo	
resumen	varchar	5000	Resumen del artículo	



nombre_area_frascati_amplio	varchar	100	Nombre del área frascati amplio	
nombre_area_unesco_amplio	varchar	100	Nombre del área unesco amplio	
tipo_publicacion	varchar	50	Tipo de publicación	
anio_publicacion	int	11	Año de publicación	
link_revista	varchar	700	Link de la revista donde se encuentra publicada el artículo	
doi	varchar	100	Doi del artículo	
estado_publicacion	varchar	45	Estado de publicación	
enlace_documento	varchar	250	Enlace del documento original	
factor_impacto	varchar	45	Factor de impacto del artículo	
cuartil	varchar	45	Cuartil del artículo	
autor_identificación	varchar	45	Identificación del autor que registró el artículo	
orden_autor	int	11	Orden del autor en la publicación	
nombres	varchar	100	Nombres del autor	
nombre_afiliacion	varchar	200	Nombre de la afiliación	
nombre_area_frascati_especif	varchar	100	Nombre área frascati	



ico			específico	
nombre_area_unesco_especifico	varchar	100	Nombre área unesco específico	
nombre_medio_publicacion	varchar	150	Nombre del medio de publicación	
id_area_unesco	int	11	Id área unesco	Clave Foránea
id_area_frascati	int	11	Id área frascati	Clave Foránea
id_medio_publicacion	int	11	Id Medio de Publicación	Clave Foránea

Diccionario de Datos Modelo Área Frascati Específico

Nombre:	area_frascati_especifico	Versión:	1.0	
Descripción:	Tabla de área frascati específico			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_area_frascati	int	11	Id del área frascati consecutivo, autoincremental	Clave
descripcion	varchar	100	Descripción del área frascati	



Diccionario de Datos Modelo Área Unesco Específico

Nombre:	area_unesco_especifico	Versión:	1.0	
Descripción:	Tabla de área unesco específico			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_area_unesco	int	11	Id del área unesco consecutivo, autoincremental	Clave
descripcion_unesco	varchar	100	Descripción del área unesco	

Diccionario de Datos Modelo Artículo Autor

Nombre:	articulo_autor	Versión:	1.0	
Descripción:	Tabla de los autores que pertenecen a cada uno de los artículos			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_autor	int	11	Id del autor consecutivo, autoincremental	Clave
id_articulo	int	11	Id del artículo	Clave Foránea
identificacion	varchar	15	Número de identificación del autor	
orden_autor	int	11	Orden del autor en el artículo	
nombre	varchar	200	Nombre completo del autor	



Diccionario de Datos Modelo Base Datos Bibliográfica

Nombre:	base_datos_digital	Versión:	1.0	
Descripción:	Tabla con detalle de las base de datos bibliográficas			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_base_datos_digital	int	11	Id de la base de datos digital consecutivo, autoincremental	Clave
nombre_base_datos_digital	varchar	200	Nombre de la base de datos digital	
proveedor	varchar	100	Nombre del proveedor de la base de datos digital	
costo_actual	double		Costo actual pagado de la base de datos digital	
suscripcion_descripcion	varchar	500	Detalle de la suscripción	
area_servicio	varchar	1000	Detalle de las áreas y servicios que ofrecen cada una de las bases de datos digital	
esUtilizadaEstudio	varchar	2	Controla si la base es utilizada para el estudio y análisis	

Diccionario de Datos Modelo Estadísticas de Uso

Nombre:	estadisticas_uso	Versión:	1.0	
----------------	------------------	-----------------	-----	--



Descripción:	Tabla con las estadísticas de uso de las base de datos bibliográficas			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_estadisticas_uso	int	11	Id de las estadísticas de uso consecutivo, autoincremental	Clave
id_base_datos_digital	int	11	Id de la base de datos digital	Clave Foránea
año	int	11	Año	
mes	varchar	45	Mes	
numero_búsquedas	int	11	Número de búsquedas por base de datos digital	

Diccionario de Datos Modelo Medios de Publicación

Nombre:	medio_publicacion	Versión:	1.0	
Descripción:	Tabla con los medios de publicación correspondiente a cada uno de los artículos			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_medio_publicacion	int	11	Id del medio de publicación consecutivo, autoincremental	Clave
nombre	varchar	150	Nombre del medio de publicación	



Diccionario de Datos Modelo Medios de Publicación Búsqueda

Nombre:	medio_publicacion_búsqueda	Versión:	1.0	
Descripción:	Tabla con los medios de publicación de búsquedas			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_medio_publicacion	int	11	Id del medio de publicación consecutivo, autoincremental	Clave
nombre	varchar	150	Nombre del medio de publicación	Clave Foránea
numero_búsquedas	int	11	Número de búsquedas de una publicación	
estado	varchar	1	Estado de la publicación: indexado o no indexado	

Diccionario de Datos Modelo Medios de Publicación Citación

Nombre:	medio_publicacion_citacion	Versión:	1.0	
Descripción:	Tabla con los medios de publicación de citación			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_medio_publicacion	int	11	Id del medio de publicación consecutivo,	Clave



			autoincremental	
nombre	varchar	150	Nombre del medio de publicación	Clave Foránea
numero_citas	int	11	Número de citas de una publicación	
estado	varchar	1	Estado de la publicación: indexado o no indexado	

Diccionario de Datos Modelo Medios de Publicación

Nombre:	medio_publicacion_publicacion	Versión:	1.0	
Descripción:	Tabla con los medios de publicación de número de publicaciones			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_medio_publicacion	int	11	Id del medio de publicación consecutivo, autoincremental	Clave
nombre	varchar	150	Nombre del medio de publicación	Clave Foránea
numero_publicaciones	int	11	Número de publicaciones	
estado	varchar	1	Estado de la publicación: indexado o no indexado	



Diccionario de Datos Modelo Referencia

Nombre:	referencia	Versión:	1.0	
Descripción:	Tabla con las referencias de cada uno de los artículos			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_referencia	int	11	Id de la referencia consecutivo, autoincremental	Clave
id_articulo	int	11	Id del artículo	Clave Foránea
referencia	varchar	2000	Referencia en texto plano	

Diccionario de Datos Modelo Detalle de Referencia

Nombre:	detalle_referencia	Versión:	1.0	
Descripción:	Tabla con el detalle de las referencias en donde el principal campo es el medio de publicación(venue)			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_detalle_referencia	int	11	Id del detalle de la referencia consecutivo, autoincremental	Clave
id_referencia	int	11	Id de la tabla referencia	Clave Foránea
container_type	varchar	200	Tipo de contenedor	



source	varchar	200	Fuente	
filled	varchar	30	Lleno	
gsrank	varchar	30	Rango	
pub_url	varchar	600	Url de la publicación	
author_id	varchar	400	Id de los autores	
num_citations	int	11	Número de citas	
url_scholarbib	varchar	600	Url del scholarbib	
url_add_sclib	varchar	600	Url del sclib	
citedby_url	varchar	600	Url de las citas	
url_related_articles	varchar	600	Url de los artículos relacionados	
title	varchar	350	Titulo del articulo	
author	varchar	350	Autores	
pub_year	int	11	Año de publicación	
venue	varchar	500	Medio de publicación	
abstract	varchar	350	Abstract del artículo	



Diccionario de Datos Modelo Base de Datos Bibliográfica - Revistas

Nombre:	base_datos_digital_journal	Versión:	1.0	
Descripción:	Tabla con el detalle de las bases de datos bibliográficas y sus respectivas revistas o fuentes bibliográficas.			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_base_datos_digital_journal	int	11	Id de la base de datos bibliográfica del journal	Clave
id_base_datos_digital	int	11	Id de la base de datos bibliográfica	Clave Foránea
id_journal	int	11	Id del journal	

Diccionario de Datos Modelo Categorías SJR

Nombre:	categorias_sjr	Versión:	1.0	
Descripción:	Tabla con el nombre de las categorías correspondientes del SJR			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_categoria	int	11	Id de la base de datos bibliográfica del journal	Clave
nombre	varchar	200	Nombre del SJR	



Diccionario de Datos Modelo Journal SJR

Nombre:	journal_sjr	Versión:	1.0	
Descripción:	Tabla con el detalle de los datos del factor de impacto proporcionados por el SJR			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_sjr	int	11	Id del SJR	Clave
titulo	varchar	1000	Título correspondiente al journal	Clave Foránea
rank	int	11	Número del ranking	
tipo	varchar	200	Tipo de journal	
issn	varchar	30	Identificador del journal	
sjr	float		Factor de impacto	
cuartil	varchar	10	División del cuartil	

Diccionario de Datos Modelo Equivalencia Área Unesco Frascati

Nombre:	equivalencia_area_unesco_frascati	Versión:	1.0	
----------------	-----------------------------------	-----------------	-----	--



Descripción:	Tabla para identificar la correcta correspondencia entre áreas Unesco y Frascati			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_equivalencia	int	11	Id de la equivalencia del área	Clave
id_area_unesco	int	11	Id del area Unesco	Clave Foránea
id_area_frascati	int	11	Id del area Frascati	Clave Foránea

Diccionario de Datos Modelo Equivalencia Área Unesco SJR

Nombre:	equivalencia_area_unesco_sjr	Versión:	1.0	
Descripción:	Tabla para identificar la correcta correspondencia entre áreas Unesco y SJR			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_equivalencia_area	int	11	Id de la equivalencia del área	Clave
id_area_unesco	int	11	Id del área Unesco	Clave Foránea
id_area_sjr	int	11	Id del área SJR	Clave Foránea



Diccionario de Datos Modelo Parámetro

Nombre:	parametro	Versión:	1.0	
Descripción:	Tabla realizar las configuraciones de los parámetros utilizados en el sistema			
Campo	Tipo de dato	Tamaño	Descripción	Llave
id_parametro	int	11	Id de la equivalencia del área	Clave
nombre	varchar	500	Nombre del parámetro o peso	
valor	int	11	Valor del peso	
codigo_parametro	int	11	Código correspondiente al peso del parámetro	

Servicios web implementados en el desarrollo del sistema

A continuación, se detallan los servicios que se implementaron en cada una de las secciones del prototipo de sistema.

Los servicios web utilizados para la funcionalidad de la sección “Publicaciones” se describen en la Tabla A3. 1.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/articulo/insertar	Ingresar Publicaciones	POST	Publicaciones: Archivo Excel	Notificación del proceso
http://localhost:5000/articulo/listar	Lista Publicaciones	GET	Sin Parámetros	Lista <Publicaciones>



http://localhost:5000/articulo/eliminar	Elimina una publicación	GET	Id del Artículo	Notificación del proceso
http://localhost:5000/articulo/actualizar	Actualiza campos de una publicación	POST	Autor, título, año publicación, tipo publicación, cuartil.	Notificación del proceso
http://localhost:5000/referencia/listarReferenciasPorIdArticulo	Lista Referencias por Artículo	GET	Id del Artículo	Lista <Referencias>
http://localhost:5000/detalleReferencia/listaDetalleReferenciaPorIdArticulo	Lista Detalle Referencia Por Artículo	GET	Id del Artículo	Lista <Detalle Referencia>

Tabla A3. 1 Servicios Web para la Sección Publicaciones.

En la Tabla A3. 2 se encuentran los servicios web utilizados para la funcionalidad de la sección de “Autores”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/articuloAutor/insertar	Ingresar Autores y la relación con las publicaciones	POST	Autores Por Publicación: Archivo o Excel	Notificación del proceso
http://localhost:5000/autor/listar	Listar los Autores	GET	Sin Parámetros	Lista <Autores>
http://localhost:5000/autor/actualizar	Actualiza la información del autor	POST	Id del Autor, Nombre	Notificación del Proceso



http://localhost:5000/articuloAutor/eliminar	Eliminar la relación entre el autor y la publicación.	GET	Id del Artículo o Autor	Notificación del proceso
http://localhost:5000/articuloAutor/listarPorIdAutor	Lista las publicaciones por autor.	GET	Id del Autor	Lista<Publicaciones>
http://localhost:5000/autor/ejecutarDatosHighChart	Lista los autores según su orden de autor	GET	Sin Parámetros	Lista <autores por orden de autor>

Tabla A3. 2 Servicios Web para el mantenimiento de los autores.

En la Tabla A3. 3 se describen los servicios web para la operatividad de la sección de “Referencias”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/articulo/listarArticulosSinReferencias	Lista Publicaciones Sin Referencias	GET	Sin Parámetros	Lista<Publicaciones>
http://localhost:5000/articulo/listarArticulosSinCompletarReferencias	Lista Publicaciones Sin Completar Referencias	GET	Sin parámetros	Lista<Publicaciones>
http://localhost:5000/referencia/insertarManual	Inserta Referencia de Forma Manual	POST	Id del artículo, Referencia	Notificación del Proceso
http://localhost:5000/referencia/listarReferenciasPorIdArticulo	Lista Referencias por Artículo	GET	Id del Artículo	Lista <Referencias>
http://localhost:5000/referencia/insertarAutomaticoScopus	Inserta Referencia y Detalle Referencia Automático con el uso de la API de Scopus	POST	Id del Artículo, Id de la Base de Datos Bibliográfica	Notificación del Proceso



http://localhost:5000/referencia/listarReferenciasNoEncontradasPorIdArticulo	Lista las Referencias Pendientes por Id del Artículo	GET	Id del Artículo	Lista <Referencias>
http://localhost:5000/referencia/obtenerDetalleReferenciaTotal	Obtiene el Detalle de la Referencia utilizando la API Google Académico	POST	Id del Artículo	Notificación del proceso
Servicio Web Externo				
Scopus.py	Busca las referencias y el detalle de las referencias de la publicación indexada en Scopus		DOI de la publicación	references:{ fulltext,title, authors, sourcetitle, publicationyear, citedbycount}
Scholar.py (API)	Busca el Detalle de la Publicación por la Referencia en Texto		Referencia	{author_id, bib: {abstract, author, pub_year, title, objects,venue},citedby_url, eprint_url, filled, gsrnk, num_citations, pub_url, source,url_add_sclib, url_scholarbib}

Tabla A3. 3 Servicios Web para la extracción y almacenamiento de las referencias y detalle de las referencias.

En la Tabla A3. 4 se presentan los servicios web empleados para la sección “Estadísticas de los proveedores”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/baseDatosBibliografica/listar	Lista base de datos bibliográfica	GET	Sin Parámetros	Lista <Bases de datos bibliográfica >
http://localhost:5000/journa/listarPorBaseDatosBibliografica	Lista las revistas por base de datos bibliográfica	GET	Id de la Base de Datos Bibliográfica	Lista<Revistas>



http://localhost:5000/estadisticasUso/insertar	Inserta estadísticas de uso por base de datos bibliográfica	POST	Id Base de datos bibliográfica, año, id mes, número de búsquedas	Notificación del Proceso
http://localhost:5000/estadisticasJournal/insertar	Insertar estadísticas de búsqueda a las revistas correspondientes a cada base de datos bibliográfica	POST	Id de la base de datos bibliográfica, id de la revista, año, id mes, número de búsquedas	Notificación del Proceso
http://localhost:5000/estadisticasUso/buscarEstadisticasUsoPorId	Busca las estadísticas de uso por base de datos bibliográfica	POST	Id de la base de datos bibliográfica	Lista <Estadísticas de Uso por Base de Datos Bibliográfica >
http://localhost:5000/estadisticasJournal/buscarEstadisticasJournalPorId	Lista las estadísticas de uso de la revista	POST	Id de la base de datos bibliográfica, Id de la revista	Lista<Estadísticas de Uso de la Revista>
http://localhost:5000/estadisticasUso/eliminar	Elimina la estadística de uso de la base de datos bibliográfica	GET	Id de la estadística de uso	Notificación del proceso
http://localhost:5000/estadisticasJournal/eliminar	Elimina la estadística de uso de las revistas de las bases de datos bibliográficas	GET	Id de la estadística de uso	Notificación del proceso

Tabla A3. 4 Servicios web para el ingreso y mantenimiento de las estadísticas de uso de las bases de datos bibliográficas y sus fuentes.

En la Tabla A3. 5 se presentan los servicios web empleados para el mantenimiento de la sección “Análisis de Datos”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
---------------	-------------	------	---------	--------



http://localhost:5000/ analisisEstadistico/n umeroPublicaciones PorAreaUnesco	Lista número de publicaciones por área Unesco	GET	Sin Parámetros	Lista <número de publicaciones por área Unesco >
http://localhost:5000/ analisisEstadistico/ numeroPublicacione sPorAreaFrascati	Lista número de publicaciones por área Frascati	GET	Sin Parámetros	Lista <número de publicaciones por área Frascati >
http://localhost:5000/ areaFrascati/listar	Lista las áreas pertenecientes a Frascati	GET	Sin Parámetros	Lista <área Frascati >
http://localhost:5000/ areaUnesco/listar	Lista las áreas pertenecientes a Unesco	GET	Sin Parámetros	Lista <área Unesco>
http://localhost:500/ prediccion/ejecutarP rediccion	Lista de valores de la predicción a 6 meses	GET	Sin Parámetros	Lista <valor de predicción del número de búsquedas>

Tabla A3. 5 Servicios web para la sección análisis de datos

En la Tabla A3. 6 se presentan los servicios web empleados para la evaluación de las revistas en la sección “Ley de Bradford”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/ medioPublicacionBúsqueda/conteoMediosPublicacionBusqueda	Cuenta el número de búsquedas realizadas por revista y las ordena descendentemente	GET	Sin Parámetros	Notificación del Proceso
http://localhost:5000/ medioPublicacionCita cion/conteoMedios PublicacionCita cion	Cuenta el número de citas realizadas a las revistas y las ordena descendentemente	GET	Sin parámetros	Notificación del proceso
http://localhost:5000/ medioPublicacionPu blicacion/conteoMedi osPublicacionPublic	Cuenta el número de publicaciones realizadas en las revistas y las ordena	GET	Sin parámetros	Notificación del proceso



acion	descendentemente			
http://localhost:5000/medioPublicacionCita cion/conteoMedios PublicacionCita cionPorAnio	Cuenta el número de citas realizadas a las revistas en un periodo de tiempo y las ordena descendentemente	GET	Año desde, año hasta	Notificación del Proceso
http://localhost:5000/medioPublicacionPu blicacion/conteoMedi osPublicacionPublic acionPorAnio	Cuenta el número de publicaciones realizadas en las revistas en un periodo de tiempo y las ordena descendentemente	GET	Año desde, año hasta	Notificación del Proceso
http://localhost:5000/medioPublicacionCita cion/conteoMedios PublicacionCita cionPorAreaFrascati	Cuenta el número de citas realizadas a las revistas por área Frascati y las ordena descendentemente	GET	Id área Frascati	Notificación del proceso
http://localhost:5000/medioPublicacionPu blicacion/conteoMedi osPublicacionPublic acionPorAreaFrasca ti	Cuenta el número de publicaciones realizadas en las revistas por área Frascati y las ordena descendentemente	GET	Id área Frascati	Notificación del proceso
http://localhost:5000/medioPublicacionCita cion/conteoMedios PublicacionCita cionPorAreaUnesco	Cuenta el número de citas realizadas a las revistas por área Unesco y las ordena descendentemente	GET	Id área Unesco	Notificación del proceso
http://localhost:5000/medioPublicacionPu blicacion/conteoMedi osPublicacionPublic acionPorAreaUnesc o	Cuenta el número de publicaciones realizadas en las revistas por área Unesco y las ordena descendentemente	GET	Id área Unesco	Notificación del proceso
http://localhost:5000/medioPublicacionCita cion/conteoMedios PublicacionCita cionPorAreaFrascatiPor	Cuenta el número de citas realizadas a las revistas por área Frascati en un periodo determinado y las ordena	GET	Id área Frascati, año desde, año hasta	Notificación del proceso



Anio	descendentemente			
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaFrascatiPorAnio	Cuenta el número de publicaciones realizadas en las revistas por área Frascati en un periodo determinado y las ordena descendentemente	GET	Id área Frascati, año desde, año hasta	Notificación del proceso
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaUnescoPorAnio	Cuenta el número de citas realizadas a las revistas por área Unesco en un periodo determinado y las ordena descendentemente	GET	Id área Unesco, año desde, año hasta	Notificación del proceso
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaUnescoPorAnio	Cuenta el número de publicaciones realizadas en las revistas por área Unesco en un periodo determinado y las ordena descendentemente	GET	Id área Unesco, año desde, año hasta	Notificación del proceso
http://localhost:5000/leyBradford/coincidencia	Obtiene y almacena el número de publicaciones, el número de citas y el número de búsquedas en función de la coincidencia del nombre de la revista	GET	Sin Parámetros	Notificación del proceso
http://localhost:5000/resumenMediosPublicacion/listar	Lista el resumen de los medios de publicación, que contiene el número de publicaciones, número de citas, número de búsquedas, factor de impacto y si esta indexada o no en los repositorios bibliográficos.	GET	Sin Parámetros	Lista<Calificaciones de la Revista>

Tabla A3. 6 Servicios web para la evaluación de las revistas a través de la ley de Bradford.



En la Tabla A3. 7 se presentan los servicios web empleados para el mantenimiento de la sección “Medios de publicación”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/medioPublicacion/listar	Lista los Medios de Publicación	GET	Sin Parámetros	Lista <Medios de Publicación>
http://localhost:5000/medioPublicacion/insertar	Inserta un Medio de Publicación	POST	Nombre Medio Publicación	Notificación del Proceso
http://localhost:5000/medioPublicacion/eliminar	Elimina un Medio de Publicación	GET	Id del Medio Publicación	Notificación del proceso
http://localhost:5000/medioPublicacion/verificaMedioPublicacionPorNombre	Verifica un Medio de Publicación	GET	Nombre	Notificación del Proceso
http://localhost:5000/medioPublicacion/actualizar	Actualiza un Medio de Publicación	POST	Nombre Medio Publicación	Notificación del Proceso
http://localhost:5000/medioPublicacion/buscar	Busca un Medio de Publicación	GET	Id del Medio Publicación	Lista <Medios de Publicación>

Tabla A3. 7 Servicios Web para el mantenimiento de los Medios Publicación.

En la Tabla A3. 8 se presentan los servicios web empleados para el mantenimiento de Ranking Medios de Publicación.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/medioPublicacionPublicacion/listar	Lista los Ranking Medios de Publicación	GET	Sin Parámetros	Lista <Ranking Medios de Publicación>
http://localhost:5000/medioPublicacionPublicacion/conteoMedios	Conteo del ranking de medios de publicación	GET	Sin Parámetros	Lista <Ranking Medios de Publicación>



osPublicacionPublicacion	sin filtros			Publicación>
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAnio	Conteo del ranking de medios de publicación por año	GET	Año desde Año hasta	Lista <Ranking Medios de Publicación>
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaFrascati	Conteo del ranking de medios de publicación por área Frascati	GET	Id del área	Lista <Ranking Medios de Publicación>
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaUnesco	Conteo del ranking de medios de publicación por área Unesco	GET	Id del área	Lista <Ranking Medios de Publicación>
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaFrascatiPorAnio	Conteo del ranking de medios de publicación por área Frascati y por año	GET	Id del área Año desde Año hasta	Lista <Ranking Medios de Publicación>
http://localhost:5000/medioPublicacionPublicacion/conteoMediosPublicacionPublicacionPorAreaUnescoPorAnio	Conteo del ranking de medios de publicación por área Unesco y por año	GET	Id del área Año desde Año hasta	Lista <Ranking Medios de Publicación>

Tabla A3. 8 Servicios Web Ranking Medios de Publicación.

En la Tabla A3. 9 se presentan los servicios web empleados para el mantenimiento de Ranking Medios de Citación.

Nombre del WS	Descripción	Tipo	Entrada	Salida
---------------	-------------	------	---------	--------



http://localhost:5000/medioPublicacionCitacion/listar	Lista los Ranking Medios de Citación	GET	Sin Parámetros	Lista <Ranking Medios de Citación>
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacion	Conteo del ranking de medios de citación sin filtros	GET	Sin Parámetros	Lista <Ranking Medios de Citación>
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAnio	Conteo del ranking de medios de citación por año	GET	Año desde Año hasta	Lista <Ranking Medios de Citación>
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaFrascati	Conteo del ranking de medios de citación por área Frascati	GET	Id del área	Lista <Ranking Medios de Citación>
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaUnesco	Conteo del ranking de medios de citación por área Unesco	GET	Id del área	Lista <Ranking Medios de Citación>
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaFrascatiPorAnio	Conteo del ranking de medios de citación por área Frascati y por año	GET	Id del área Año desde Año hasta	Lista <Ranking Medios de Citación>
http://localhost:5000/medioPublicacionCitacion/conteoMediosPublicacionCitacionPorAreaUnescoPorAnio	Conteo del ranking de medios de citación por área Unesco y por año	GET	Id del área Año desde Año hasta	Lista <Ranking Medios de Citación>

Tabla A3. 9 Servicios Web Ranking Medios de Citación.

En la Tabla A3. 10 se presentan los servicios web empleados para el mantenimiento del número búsquedas por revistas o fuentes bibliográficas pertenecientes a las bases datos bibliográficas utilizadas para el análisis.



Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/medioPublicacionBusqueda/listar	Lista los Ranking Número de Búsquedas	GET	Sin Parámetros	Lista <Ranking Número de Búsquedas>
http://localhost:5000/medioPublicacionBusqueda/conteoMediosPublicacionBusqueda	Conteo del ranking de número de búsquedas sin filtros	GET	Sin Parámetros	Lista <Ranking Número de Búsquedas>
http://localhost:5000/medioPublicacionBusqueda/conteoMediosPublicacionBusquedaPorAnio	Conteo del ranking de número de búsquedas por año	GET	Año desde Año hasta	Lista <Ranking Número de Búsquedas>

Tabla A3. 10 Servicios Web Número Búsquedas Base Datos Bibliográfico.

En la Tabla A3. 11 se presentan los servicios web empleados para el mantenimiento de la sección “Áreas”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/areaUnesco/listaAreaUnesco	Lista áreas de Unesco	GET	Sin Parámetros	Lista <Áreas de Unesco>
http://localhost:5000/areaUnesco/insertar	Inserta un Área Unesco	POST	Nombre Área Unesco	Notificación del Proceso
http://localhost:5000/areaUnesco/eliminar	Elimina un Área de Unesco	GET	Id del Área Unesco	Notificación del proceso
http://localhost:5000/areaFrascati/listaAreaFrascati	Lista áreas de Frascati	GET	Sin Parámetros	Lista <Áreas de Frascati>
http://localhost:5000/areaFrascati/insertar	Inserta un Área Frascati	POST	Nombre Área Frascati	Notificación del Proceso



http://localhost:5000/areaFrascati/eliminar	Elimina un Área de Frascati	GET	Id del Área Frascati	Notificación del proceso
http://localhost:5000/areaSJR/listaAreaSJR	Lista áreas de SJR	GET	Sin Parámetros	Lista <Áreas de SJR>
http://localhost:5000/areaSJR/insertar	Inserta un SJR	POST	Nombre Área SJR	Notificación del Proceso
http://localhost:5000/areaSJR/eliminar	Elimina un Área SJR	GET	Id del Área SJR	Notificación del proceso

Tabla A3. 11 Servicios web para el mantenimiento de las áreas

En la Tabla A3. 12 se presentan los servicios web empleados para el mantenimiento de la sección de “Bases de datos bibliográficas”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/baseDatosBibliografica/listar	Lista bases de datos bibliográficas	GET	Sin Parámetros	Lista <Bases de datos bibliográficas>
http://localhost:5000/baseDatosBibliografica/insertar	Inserta una base de datos bibliográfica	POST	Nombre Base de datos bibliográfica, nombre del proveedor, costo actual, suscripción/descripción y área/servicio	Notificación del Proceso
http://localhost:5000/baseDatosBibliografica/eliminar	Elimina una base de datos bibliográfica	GET	Id de la base de datos bibliográfica	Notificación del proceso
http://localhost:5000/baseDatosBibliografica/actualizar	Actualiza una base de datos bibliográfica	POST	Id de la base de datos bibliográfica, campos a actualizar	Notificación del proceso



http://localhost:5000/baseDatosDigital/buscarBaseDatosBibliograficaPorId	Busca una base de datos bibliográfica	POST	Id de la base de datos bibliográfica	Lista <Bases de datos digitales >
http://localhost:5000/journal/listarPorBaseDatosBibliografica	Listar revistas por base de datos bibliográfica	GET	Id de la Base de Datos Bibliográfica	Lista <Revistas>
http://localhost:5000/journal/insertarScienceDirect	Ingresa las revistas que corresponden a Science Direct	POST	Id de la Base de Datos Bibliográfica, Nombre de la revista	Notificación del Proceso
http://localhost:5000/journal/insertarEbsco	Ingresa las revistas que corresponden a Ebsco	POST	Id de la Base de Datos Bibliográfica, Nombre de la revista	Notificación del Proceso
http://localhost:5000/journal/insertarScopus	Ingresa las revistas que corresponden a Scopus	POST	Id de la Base de Datos Bibliográfica, Nombre de la revista	Notificación del Proceso

Tabla A3. 12 Servicios web para la sección del mantenimiento de las bases de datos bibliográficas y el ingreso de sus revistas.

En la Tabla A3. 13 se presentan los servicios web empleados para el mantenimiento de la sección “Configuración”.

Nombre del WS	Descripción	Tipo	Entrada	Salida
http://localhost:5000/parametro/listar	Listar nombres de los pesos del parámetro	GET	Sin Parámetros	Lista <nombres de los pesos del parámetro>
http://localhost:5000/parametro/crear	Creación de parámetros con nombre y valor del peso	GET	Sin Parámetros	Notificación del proceso



http://localhost:5000/ parametro/eliminar	Elimina un parámetro	GET	Sin Parámetros	Notificación del proceso
http://localhost:5000/ parametro/actualizar	Edita un nombre o valor del parámetro	POST	Id del parámetro	Notificación del proceso

Tabla A3. 13 Servicios web para la sección configuración.



Anexo 4: Artículo Presentado en el ICAT 2021

International Conference on Applied Technologies

Decision-making of the university libraries' digital collection through the publication and citation patterns analysis. A literature review

Tania Landívar¹, Ronaldo Rendon¹ and Lorena Siguenza-Guzman^{1,2*}[0000-0003-1367-5288]

¹ Department of Computer Sciences. Faculty of Engineering. Universidad de Cuenca, Cuenca, Ecuador

² Research Centre Accountancy, Faculty of Economics and Business. KU Leuven, Leuven, Belgium

lorena.siguenza@ucuenca.edu.ec

Abstract. In recent years, different bibliometric methods and indicators, data mining techniques, and collection and visualization tools have been presented to analyze publication patterns and citations of scientific production. Moreover, some criteria and strategies for acquiring bibliographic databases allow forming a knowledge base for decision-making. Unfortunately, these data, methods, and tools are scattered in the literature. In this context, the objective of this article is to determine if the discovery of publication and citation patterns is being used to analyze scientific production in university libraries that helps decision-making on the subscription/acquisition of bibliographic databases. Fink's methodology was applied in conjunction with the PICO structure for the literature review. This systematic review made it possible to compile relevant information from 167 articles published during the 2016-2020 period into a findings matrix and perform a metadata and content analysis. Among the study's primary results, it stands out that the analysis of citations, the number of citations, and the VOSviewer collection and visualization tool have been the most outstanding for the study of scientific production. However, it has not had a significant impact as an aid to the acquisition/subscription of bibliographic databases.

Keywords: Publication patterns, Citation patterns, Acquisition criteria, Subscription criteria, Bibliometric, Indicators, Data Mining Tools, Techniques, Library

1 Introduction

Bibliometric studies have increased significantly in the last five years, evaluating scientific production through statistical and mathematical methods [1]. As a result of these studies, publication and citation patterns are obtained, which can be used in decision-making. Around the world, these bibliometric studies have been used for decades to measure scientific activity and its impact, and Latin America is not the exception [2, 3, 4, 5]. For example, in Mexico, Guerrero-Sosa et al. presented a system to measure the productivity and impact of publications using bibliometric indica-



tors, metadata, and semantic textual similarity [6]. In Ecuador, Zhimnay et al. analyzed the importance of measuring science to evaluate the academic excellence of universities [7]. Bibliometrics has also been used to improve collection management. For example, Gureev and Mazov built a bibliographic collection in Russia according to user needs [8]. In Belgium, Siguenza-Guzman et al. evaluated the importance of the library collection by combining four types of pattern analysis: *publications*, the journals where the institution's professors/researchers publish and their collaborations; *citations*, the references cited in those publications; *downloads*, statistics of journals consulted and downloaded provided by database providers; and, *impact factor*, the importance of a journal when establishing rankings by specialty [9]. Additionally, to increase the value of bibliometric studies, data mining techniques have been introduced through tools that allow the collection, analysis, and visualization of scientific production. They are based on bibliographic mapping and clustering techniques to develop network maps, such as citation-based author networks, co-citations, and bibliographic coupling [10].

Therefore, the use of publication and citation patterns for decision-making has become a fundamental element. Some journals carry out this type of study to find publication trends by studying authorship and citation patterns, analyzing links and connections between institutions, countries/territories, and research areas [11, 12, 13]. This is to know how the journal develops and where it should go. Furthermore, there are many studies to determine the productivity of research in a particular field, finding open or restricted access journals, years, journals and institutions published the most, collaboration networks between authors, etc. Finally, the use of publication patterns and citations for collection management aims to determine which resources (e.g., journals, bibliographic databases, and repositories) are the most used by researchers, thus providing information for collection maintenance.

The objective of this study was to investigate and establish which methods, bibliometric indicators, data mining techniques, and collection and visualization tools have been the most used for the discovery of publication and citation patterns in the analysis of scientific production. And if they are being used to support the decision-making process about the subscription/acquisition of bibliographic databases in university libraries. This article is structured as follows. First, the second section, the methodology used for the systematic literature review is presented. Then, the third section presents the results obtained and their respective discussion, in which the meta data analysis and descriptive analysis are detailed. Finally, the last section presents the conclusions and findings that will contribute to science and future research related to the subject.

2 Methodology

In this literature review, the methodology proposed by Arlene Fink [14] was applied. This procedure allowed for a structured and focused systematic review for research purposes. Fink's method consists of seven steps: (1) selection of research questions, (2) selection of bibliographic databases, (3) selection of search terms, (4) application

of practical detection criteria, (5) application of methodological detection criteria, (6) development of the review, and (7) summary of the results.

The initial step was the selection of research questions for which the PICO structure was used to design the questions and develop search strategies that allow delimiting the study's objective [15, 16]. The term PICO stands for *Population* where the problem is identified. For the present study, this element analyzed whether deciding to acquire bibliographic databases in university libraries was based on a methodological analysis by discovering publication and citation patterns. *Intervention* where the intervention criteria were established in which the methods/strategies/criteria of acquisition and subscription, bibliometric methods and indicators, data mining techniques, and tools for collecting and visualizing scientific production were defined. *Comparison* between the terms mentioned above and *Outcomes* the analysis hopes to obtain the most relevant criteria for acquiring bibliographic databases, bibliometric methods, and indicators, data mining techniques, tools most used for the study of scientific production with the discovery of publication and citation patterns. It also analyzes whether these patterns have been used as a source for decision-making to acquire and subscribe. Consequently, the following questions were posed: (a) What are the criteria, methods, or strategies for acquiring bibliographic databases in university libraries? (b) What bibliometric methods are used for the discovery of publication patterns and citations? (c) What are the most used bibliometric indicators for the analysis of publication patterns and citations? (d) What are the most commonly used data mining techniques for analyzing publication and citation patterns? (e) What tools for the collection and visualization of scientific production are used to analyze publication patterns and citations?

The second step corresponded to the selection of bibliographic databases to consult. This step was carried out through a bibliographic search, and resources such as books, scientific journals, websites, and bibliographic databases were considered. This study focused on two digital databases: Scopus, of great international relevance [17], and Scielo, which stands out in the Hispanic region [18], allowing a complete investigation. Subsequently, as a third step, it was necessary to determine the keywords, which, for this research, were derived mainly from the following basic terms: "Acquisition and Subscription", "Bibliometric Methods and Indicators", "Data Mining Techniques and Visualization Tools". The detail of the bibliographic databases, search terms, and the number of results obtained is shown in Table 1.

Table 1. Search terms utilized for the literature review.

Query	Scopus	Scielo
("Acquisition" OR "Subscription") AND "e-journals" AND "University Libraries"	9	0
"Bibliometric Methods" AND ("Publication Patterns" OR "Citation Patterns")	22	1
"Bibliometric Indicators" AND ("Publication Patterns" OR "Citation Patterns")	36	25
"Analysis" AND "Scientific Publications" AND "Data Mining" AND	11	0



4

"Techniques" ("Collection Tool" OR "Visualization Tool") AND ("Publication Pat- terns" OR "Citation Patterns"	47	16
TOTAL	125	42

The fourth step was applying practical screening criteria since many articles were obtained in the preliminary searches. Articles published between 2016-2020, in English and Spanish, were selected. The Spanish language made it possible to find principles, guidelines, and indicators applied in local institutions. English is the universal scientific language [19]; therefore, it is essential in all investigations. Using the practical criteria, it was defined that only item-type documents would be reviewed. The main reason for this decision was that journal articles currently represent the highest level of research with current and high-quality information, unlike congresses that are working in progress and have not concluded; moreover, the information is more mature in the books. Therefore, the places where practical applications can be found are journal articles [20].

The result was 167 documents, 5.4% (9 out of 167) corresponded to the search criteria of subscription to bibliographic databases, 6.6% (11 out of 167) to data mining techniques, 13.8% (23 out of 167) to bibliometric methods, 36.5% (61 out of 167) to bibliometric indicators, and 37.7% (63 out of 167) to tools for collecting and visualizing scientific production.

Once the information has been selected, the fifth step was the application of the methodological selection criteria. This step established the review of the article's methodology, results, and conclusion sections to find criteria for acquisition or subscription to bibliographic databases, methods used for bibliometric analysis and bibliometric indicators, data mining techniques, and software tools used for the extraction and visualization of scientific production. To compile the results, a "Findings Matrix" was required, which has the function of listing essential data after implementing the practical and methodological filter. This matrix contains strategic information such as the publication year, the author's name, the document's title, the content information, research area, abstract, and conclusions. It was used to perform a quick content analysis that helped determine if the articles contained relevant information to the research.

The sixth step corresponded to the study of the selected literature. For this, each document was reviewed to identify the information that allowed answering the questions posed in the first step.

And finally, the results were formulated in the seventh step. The data was exported to a spreadsheet for content and metadata analysis. Metadata analysis is an approach to improve the accuracy of resource discovery, and it allows information to be organized in a structured manner [21]. A variety of criteria and procedures may be required for different types of metadata. In addition, it supports content analysis that contributes to developing a dialogue about the evaluation of the quality of metadata [22] and whether it fulfills the purpose and objectives of the selected study.

For the content analysis, several sections were established within the matrix according to the intervention criteria in the PICO methodology and those already set above to find criteria and strategies for the acquisition of bibliographic bases, meth-

ods, and bibliometric indicators with their respective description and main characteristics. Also, a descriptive statistical study was carried out for the latter for the meta data analysis through the information frequencies, such as year of publication, productivity by country, research area, bibliometric methods and bibliometric indicators, and collection and visualization tools.

3 Results and Discussion

In this section, the main findings obtained from the systematic literature review are specified, whose results allowed to know the better state of the art on the use of methods, indicators, and tools to discover publication and citation patterns.

3.1 Metadata Analysis

The meta data of the reviewed articles provided statistics such as productivity per year, by country, and research area. The results are detailed below.

Concerning the first metadata analysis, this began with the criteria of *purchase or subscription to bibliographic databases*, where it is observed that few studies have been carried out. There is a productivity of 44.4% (4 out of 9) in 2020, 33.3% (3 out of 9) in 2019, 11.1% (1 out of 9) in 2018 and 2017, respectively. The country with the most publications is India with 33.3% (3 out of 9), followed by Nigeria and the United States with 22.2% (2 out of 9), Taiwan and Canada with 11.1% (1 out of 9). Figure 1 shows the productivity by country mentioned above.

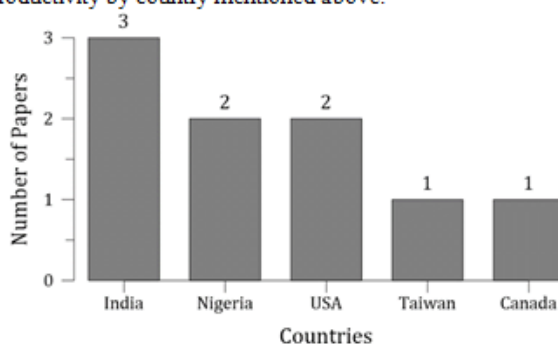


Fig. 1. Amount of information retrieved on productivity per country.

Studies have focused on evaluating the availability and use of information resources in university libraries [23], the prospects for the formation of consortia [24], inventories of electronic journal subscriptions to improve internal documentation, in addition to analyzing the degree of user satisfaction with library resources and facilities [25].

Regarding *bibliometric methods and indicators*, the publications in Scopus and Scielo are distributed as follows. With 23.9% (20 out of 84) of the year 2020, 22.6% (19 out of 84) of the year 2019, 20.2% (17 out of 84) in 2018, 15.4% (13 out of 84) in

2017 and 17.9% (15 out of 84) in 2016. These results reflect an increase in the number of studies published year after year. Figure 2 shows the most prolific countries according to Scopus, China 13.8% (8 out of 58) and India 10.3% (6 out of 58). In Scielo, Cuba stands out with 26.9% (7 out of 26).

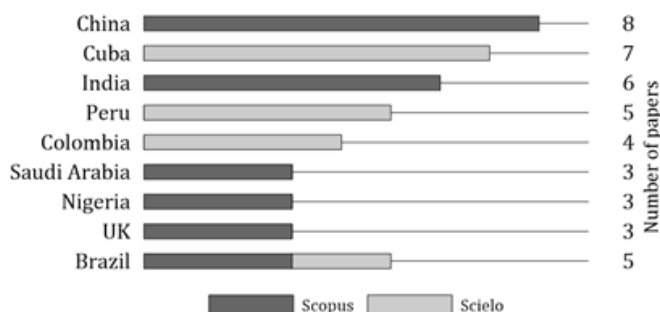
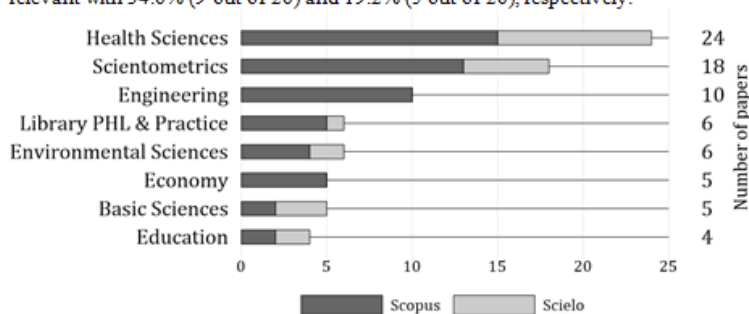


Fig. 2. Number of publications about bibliometric methods and indicators by country

Bibliometric methods and indicators have allowed finding years of higher production, prominent authors, the origin of the authors, specific field of study [26, 27], visibility and impact of research [28], the growth rate of publications, characteristics of research activities, citation bursts and chronological map [29], to name a few examples. This makes possible a reliable mapping of scientific development [30].

The research areas to which the studies of bibliometric methods and indicators have been directed both in Scopus and Scielo can be seen in Figure 3. In Scopus, the area of Health Sciences stands out with 25.9% (15 out of 58), followed by Scientometrics with 22.4% (13 out of 58) and Engineering with 17.2% (10 out of 58). Regarding Scielo, the results confirm Health Sciences and Scientometrics as the most relevant with 34.6% (9 out of 26) and 19.2% (5 out of 26), respectively.



* PHL = Philosophy

Fig. 3. Scopus and Scielo research areas on bibliometric methods and indicators.



Within the Health Sciences, topics such as Oncology [26], Complementary Medicine [31], and Orthopedics [32] are treated. Bibliometric studies are carried out on these topics seeking to obtain years of greater production, outstanding authors, the origin of authors, among others [33]. It is followed by Scientometrics, which develops activities such as publication trends [11] and analysis of the kinetics of self-citations of a journal [34], the international scientific collaboration of researchers [35], mapping of scholarly communication in publications [35], preprints such as accelerator of scholarly communication [36], text analysis [37], measurement and visualization of collaboration and research productivity [38], research impact and country productivity [39]. Finally, one of the three relevant areas is Engineering, including topics related to the comparative analysis of citation patterns in mechanical and civil engineering research projects [40].

A few publications explicitly mention the use of data mining techniques such as clustering, classification, or association rules. The productivity of publications per year in Scopus is described below, with 27.3% (3 out of 11) in 2020, 18.1% (2 out of 11) in 2019, 27.3% (3 out of 11) in 2018, 9.1% (1 out of 11) in 2017 and 18.2% (2 out of 11) in 2016, becoming a still little explored field for research studies related to publication and citation patterns. Scielo's results are null. Regarding productivity by country, Spain stands out with 18.2% (2 out of 11).

Finally, the use of tools to compile and visualize scientific production in bibliometric studies has been remarkable. Scopus reflects in 2020 a productivity of 31.9% (15 out of 47), 23.4% (11 out of 47) in 2019, 19.1% (9 out of 47) in 2018, 17% (8 out of 47) in 2017 and 8.5% (4 out of 47) in 2016. These results are closely related to those presented at the beginning of this section since they use at least one collection and visualization tool to carry out these studies. Figure 4 presents the most prolific Scopus and Scielo countries. In Scopus, the United States stands out with 25.5% (12 out of 47), China with 12.8% (6 out of 47), and India with 10.6% (5 out of 47). The results related to Scielo have a productivity of 6.3% (1 out of 16) in 2020, 31.3% (5 out of 16) in 2019, 25% (4 out of 16) in 2018, 31.3% (5 out of 16) in 2017 and 6.3% (1 out of 16) in 2016. Likewise, as shown in Figure 4, the country with the most publications is Brazil with 81.2% (13 of 16), high above Argentina, Mexico, and Singapore with a productivity of 6.3% (1 of 16), respectively. Scielo is a more local bibliographic library in Latin America; therefore, there is a considerable difference concerning productivity with Scopus, a global database. This allows concluding that there is still a gap in research with other countries despite the work done in Latin America.

8

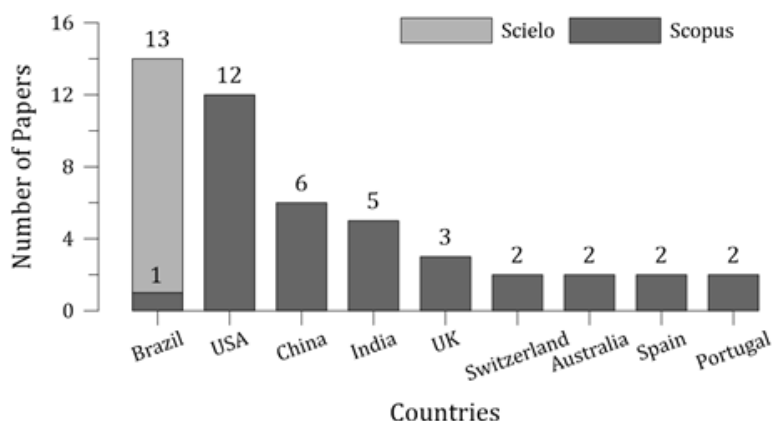


Fig. 4. Amount of information about collection and visualization tools (Scopus and Scielo).

3.2 Content Analysis

In this part of the paper, research questions established in Section 2 were answered. Regarding the question: *1) What are the criteria, methods, or strategies for acquiring bibliographic databases in university libraries?* The articles describing the entities or institutions that subscribe or acquire a digital repository were identified. The articles focus on the assessment of the availability of digital resources and library services [41], mention knowledge management for the development of collections [42], and emphasize providing and maintaining access to electronic journals [43].

The most relevant strategies for acquiring or subscribing to bibliographic databases are formalizing and strengthening the links between undergraduate students and library staff to improve access to information [41], networking, cataloging, classification, and exchanging information materials [44]. Another strategy consists in hiring selectors, i.e., library personnel who possess a great deal of tactical knowledge and subject areas, in addition to knowing which books or journals are of interest and which formats are preferred [42]. Likewise, an additional strategy is the formation of consortia between libraries where each has a representative participating in decision-making for the consortium's collection [43]. Finally, it is important to mention that all this can be a challenge for those who have not been directly involved if they do not know about purchasing, licensing, negotiations with suppliers, budget planning, and policy creation [42].

Regarding the research questions, *2) What bibliometric methods are used to discover publication and citation patterns, and 3) What are the most used bibliometric indicators for analyzing publication and citation patterns?* For their evaluation, the frequency of the different bibliometric methods and indicators was analyzed. Regarding bibliometric studies, Web of Science was the primary data source used with

51.7% (30 out of 58), followed by Scopus with 29.3% (17 out of 58), PubMed, and Medline with 5.2% (3 out of 58), respectively.

Regarding the *bibliometric methods* in Scopus, Table 2 shows that citation analysis stands out with 27.6% (16 out of 58), which allows determining links and connections in research works [45], co-citation analysis with 10.3% (6 out of 58) recognizing the affinity between co-cited authors [46], co-words analysis with 10.3% (6 out of 58) mapping the conceptual structure and thematic of a specific domain [47]. Additionally, either Lotka's Law, Bradford's Law, and Zipf's Law equally represent 6.9% (4 out of 58) of the bibliometric methods used in the literature. Lotka's Law describes the quantitative relationship between authors and articles produced in a given field and period [48]. Bradford's Law allows the selection of publications and is the most productive and relevant to cover a given area of knowledge [49]. Finally, Zipf's Law provides mechanisms to identify and select keywords in a text [50].

Table 2. Bibliometric methods utilized to evaluate the literature.

Bibliometric Method	# of articles
Citation analysis	16
Co-citation analysis	6
Co-words analysis	6
co-authorship analysis	5
Lotka's Law	4
Bradford's Law	4
Zipf's Law	4

As it can be seen in the Scopus results in Table 3, the number of citations stands out with 37.9% (22 out of 58) as the most used *bibliometric indicator*, followed by the number of articles with 25.9% (15 out of 58), the number of authors with 20.7% (12 out of 58), the h-index with 18.9% (11 out of 58), the impact factor with 13.7% (8 out of 58) and the citation score with 5.2% (3 out of 58). Likewise, the publications metadata are extracted for the bibliometric analysis, such as author, subject, country, institution, journal, year of publication, category, language, affiliation, and title. Finally, the results corresponding to Scielo are also presented, in which the number of articles with 80.8% (21 out of 26) stood out, followed by the number of citations with 65.4% (17 out of 26) and the number of authors with 53.9% (14 out of 26). These results ratify those presented in Scopus, which determines the number of articles and citations as the most relevant.

Table 3. Bibliometric Indicators Scielo y Scopus

Bibliometric indicator	Scopus	Scielo
Number of articles	15	21
Number of citations	22	17
Number of authors	12	14
h-index	11	7
Impac factor	8	5

CiteScore	3	3
-----------	---	---

Regarding question 4) *What are the most used data mining techniques for analyzing publication patterns and citations?* Data mining techniques are commonly used to find hidden information in large amounts of data [51]. It can be applied to different areas. For example, Torres Berrú et al. present a study whose purpose is to find algorithms that identify specific characteristics of fraud or corruption [52]. On the other hand, Marcos-Pablos et al. describe a methodology to help adapt, filter, and synchronize the search and selection of articles for bibliographic reviews [53].

A list of data mining techniques is obtained between the second, third, and fourth search terms for Scopus since no results on this topic were found in Scielo. The findings show that clustering stands out with 11.6% (8 out of 69), classification with 10.1% (7 out of 69), and multiple regressions with 5.8% (4 out of 69), these percentages being relatively low. However, the next question shows tools dedicated exclusively to bibliometric studies that include data mining functionalities.

Finally, on the research question, 5) *What tools for collecting and visualizing scientific production are used to analyze publication and citation patterns?* To identify the collection and visualization tools, the total number of documents obtained from Scopus was taken from the search terms of bibliometric methods and indicators, together with the collection and visualization tools, giving 105 papers. The tools that stand out in data collection are HistCite with 4.8% (5 of 105), CiteSpace with 2.9% (3 of 105), ScientoPy with 1.9% (2 of 105), and Endnote with 1.9%. For the visualization tools, Spreadsheet was found with 20.9% (22 of 105), Tableau, DAVID, SPSS, Sci2, Ucinet, and Pajek, all with a percentage of 1.9% (2 of 105), respectively. Finally, the most used tool for collecting and visualizing with 28% (29 out of 105) is VOSviewer since it is a software tool to build and visualize bibliometric networks. These networks can include, for example, individual journals, researchers, or publications and can be created from citations, bibliographic links, co-citations, or co-authorship relationships [54, 55]. In addition, VOSviewer also offers data mining functionality to build and visualize co-occurrence networks of important terms drawn from a body of scientific literature [56]. Table 4 shows the collection and visualization tools found in Scopus that stood out the most. Regarding Scielo, the tools that stand out in data collection and visualization are Spreadsheet with 45.2% (19 of 42), VOSviewer with 14.3% (6 of 42), EndNote with 9.5% (4 of 42), and Cochrane with 7.1% (3 of 42), shown in Table 5.

Table 4. Collection and Visualization tools reported in Scopus.

Tool	Number of articles
VOSviewer	29
Spreadsheet	22
HistCite	5
CiteSpace	3
Pajek	2
ScientoPy	2



EndNote	2
---------	---

Table 5. Collection and Visualization tools reported in Scielo.

Tool	Number of articles
Spreadsheet	19
VOSviewer	6
EndNote	4
Cochrane	3
SPSS	3
Ucinet	3
NetDraw	2

4 Conclusions

This article describes a systematic review to establish the most widely used bibliometric methods and indicators, data mining techniques, and collection and visualization tools to discover publication patterns and citations. In addition, they determine if they are being used in the decision-making process in university libraries regarding the subscription/acquisition of bibliographic databases. In this work, the PICO structure and Fink's model of the research literature review were applied based on a definition of research questions that allowed identifying resources related to the research topic [7]. In the first part of the literature review, two bibliographic databases (Scopus and Scielo) developed a search for information. Later, practical and methodological selection criteria were applied, establishing a sample of 167 articles with scientific content for their analysis.

The development of this review determined that the ability to use publication and citation patterns for collection management has not been exploited, specifically to decide which journal, database, or repository to acquire or subscribe, a decision that university libraries must constantly occupy. As can be seen, these decisions are made by staff with a high degree of knowledge in licensing, budget planning, etc. However, if there were another manner to support such decisions, it would make the analysis more objective and thus improve the maintenance of the collections.

The next point corresponds to the bibliometric methods, which stand out, the citation analysis allowing to determine links and connections in the research work; the co-citation analysis that recognizes the affinity between cited co-authors; and, finally, the study of common words mapping the conceptual and thematic structure of a specific domain. These methods are valuable because, after a certain process, they result in co-authorship maps, cross-country comparisons, connections between research topics, i.e., a lot of information to analyze. However, to apply these methods, the unit of measurement called "bibliometric indicator" is necessary. Also, the most important aspect in this study is the number of citations, number of articles, and H-index, which



allow evaluating different characteristics of science activity linked to both the production and the consumption of information.

Bibliometric studies have increased considerably; thus, there are a series of tools that facilitate this activity. VOSviewer is the most used not only for the collection and visualization of information but also for data mining capabilities. In addition, a Spreadsheet should be considered a key tool that can be regarded as unique because depending on the topic this works with, it adapts to the needs of users, ranging from data management, calculations with information to visualization of summary of results by graphs.

This article reviews a topic of interest in any field for academics and professionals because research is a worldwide effort applicable to the diverse scientific community and society in general. Furthermore, the work can be seen as an initial conceptual framework for developing research projects and constructing new tools for the analysis of scientific production. Therefore, for future work, it is recommended to create a software or tool that allows uniting these methods, indicators, and tool functionalities whose objective is to provide information for decision-making in university libraries for collection maintenance.

Appendix A – Full list of references retrieved in the literature review from Scopus and Scielo

The list of sources used in the literature review can be found online at <https://n9.cl/ctdhg>.

Acknowledgments

This research was funded by the Vice-Rector's Office for Research of the University of Cuenca.

References

1. Solis, B.P., Argüello, J.C.C., Barba, L.G., Gurrola, M.P., Zarhri, Z., TrejoArroyo, D.L.: Bibliometric Analysis of the Mass Transport in a Gas Diffusion Layer in PEM Fuel Cells. *Sustainability*. 11, 1–18 (2019)
2. Glänzel, W.: Bibliometrics as a research field: A course on theory and application of bibliometric indicators. *Course Handouts*. (2003)
3. González Alcaide, G., Gorraiz, J.I.: Assessment of Researchers Through Bibliometric Indicators: The Area of Information and Library Science in Spain as a Case Study (2001–2015). *Front. Res. Metr. Anal.* 3, (2018). <https://doi.org/10.3389/frma.2018.00015>
4. Abramo, G.: Bibliometric evaluation of research performance: where do we stand? *VO*. 112–127 (2017). <https://doi.org/10.17323/1814-9545-2017-1-112-127>
5. Beaudry, C., Larivière, V.: Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy*. 45, 1790–1817 (2016). <https://doi.org/10.1016/j.respol.2016.05.009>



6. Guerrero-Sosas, J.D.T., Chicharro, F.P.R., Serrano-Guerrero, J., Menendez-Dominguez, V., Castellanos-Bolaños, M.E.: A proposal for a recommender system of scientific relevance. *Procedia Computer Science*. 162, 199–206 (2019). <https://doi.org/10.1016/j.procs.2019.11.276>
7. Zhimnay, C., Fernández Landivar, J., Albarracín, J., Sadaba, I., Sucozhanay, D.: MAPPING OF SCIENTIFIC PRODUCTION IN SOCIAL SCIENCES IN ECUADOR. Presented at the March 1 (2019)
8. Gureyev, V., Mazov, N.: Assessment of the Relevance of Journals in Research Libraries Using Bibliometrics (a Review). *Scientific and Technical Information Processing*. 42, 30–40 (2015). <https://doi.org/10.3103/S0147688215010050>
9. Siguenza-Guzman, L., Holans, L., Abbeele, A.V.D., Vandewalle, J., Verhaeren, H., Cattrysse, D.: TOWARDS A HOLISTIC ANALYSIS TOOL TO SUPPORT DECISION-MAKING IN LIBRARIES. undefined. (2013)
10. Tuppal, C.P., Gallardo-Nimobla, M.M., Arquiza, G.S., Vega, P.D.: A bibliometric analysis of the Philippine Journal of Nursing for 1966 - 2017., <https://www.herdm.ph/index.php/herdm-home?view=research&cid=71335>
11. Ullah, S.: Publication Trends Of Pakistan Heart Journal: A Bibliometric Study. *Library Philosophy and Practice (e-journal)*. (2019)
12. Wang, W., Laengle Scarlazetta, S., Merigó Lindahl, J., Yu, D., Herrera Viedma, E., Cobo, M.J., Bouchon Meunier, B.: A Bibliometric analysis of the first twenty-five years of the international journal of uncertainty, fuzziness and knowledge-based systems. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*. (2018). <https://doi.org/10.1142/S0218488518500095>
13. Peter, M., Samuel, S.A., Idhris, M., Subbarayalu, A.: Saudi Arabian Top Four Medical journals Bibliometric study. *Library Philosophy and Practice (e-journal)*. (2020)
14. Fink, A.: *Evaluation Fundamentals: Insights into Program Effectiveness, Quality, and Value*. SAGE Publications (2014)
15. BA, K., Charters, S.: *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2, (2007)
16. Shokraneh, F.: PICO Framework: Two Decades of Variation and Application. (2016)
17. Zhu, J., & Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *ArXiv:2002.02608 [Cs]*. <http://arxiv.org/abs/2002.02608>
18. Bojo Canales, C., & Sanz-Valero, J. (2020). Indicadores de impacto y de prestigio de las revistas de ciencias de la salud indizadas en la red SciELO: Estudio comparativo. <https://repisalud.isciii.es/handle/20.500.12105/10992>
19. Ramírez, M.G., Falcón, I.A.L.: Importancia Del Lenguaje En El Contexto De La Aldea Global. *Horizontes Educativos*. 15, 95–107 (2010)
20. Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., Cattrysse, D.: Literature Review of Data Mining Applications in Academic Libraries. *The Journal of Academic Librarianship*. 41, 499–510 (2015). <https://doi.org/10.1016/j.acalib.2015.06.007>
21. Chen, Y., Chen, S., Chang, Y., Lin, S.C., Ya-ning, V., Shu-jiun, C., Yi-ting, C., Simon, C., Lin, C., Chen, Y., Chen, S., Chang, Y., Lin, S.C.: A Metadata Lifecycle for Content Analysis in Digital Libraries
22. Moen, W.E., Stewart, E.L., McClure, C.R.: The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS).
23. Tukur, L., Kannan, S.: An Appraisal of Availability and Utilization of Information Resources and Library Services by Undergraduate Students in Three Agriculture University Libraries in Northern Nigeria (ABU, FUAM, FUTM). *Library Philosophy and Practice (e-journal)*. (2020)



24. Zhao, W., Zhao, S., MacGillivray, K.-S.: Providing and Maintaining Access to Electronic Serials: Consortium and Member University Library's Perspectives. *The Serials Librarian*. 72, 144–151 <https://doi.org/10.1080/0361526X.2017.1309831> (2017).
25. Urban, S.G.: Using E-Journal Subscription Inventories to Improve Internal Documentation and Patron Access. *Technical Services Quarterly*. 36, 269–280 (2019). <https://doi.org/10.1080/07317131.2019.1621568>
26. Danell, J.-A.B.: Integrative oncology from a bibliometric point of view. *Complementary Therapies in Medicine*. 52, 102477 (2020). <https://doi.org/10.1016/j.ctim.2020.102477>
27. Simao, L.B., Carvalho, L.C., Madeira, M.J.: Intellectual structure of management innovation: bibliometric analysis. *Manag Rev Q*. 71, 651–677 (2021). <https://doi.org/10.1007/s11301-020-00196-4>
28. Sahoo, S., Pandey, S.: Bibliometric Analysis and Visualization of Global Ocean Acidification Research. *Science & Technology Libraries*. 39, 414–431 (2020). <https://doi.org/10.1080/0194262X.2020.1776194>
29. Guo, Y., Hao, Z., Zhao, S., Gong, J., Yang, F.: Artificial Intelligence in Health Care: Bibliometric Analysis. *J Med Internet Res*. 22, e18228 (2020). <https://doi.org/10.2196/18228>
30. Ellegaard, O.: The application of bibliometric analysis: disciplinary and user aspects. *Scientometrics*. 116, 181–202 (2018). <https://doi.org/10.1007/s11192-018-2765-z>
31. Danell, J.-A.B., Danell, R., Vuolanto, P.: Scandinavian research on complementary and alternative medicine: A bibliometric study. *Scand J Public Health*. 48, 609–616 (2020). <https://doi.org/10.1177/1403494819834099>
32. Shon, W.Y., Yoon, B.-H., Jung, E.-A., Kim, J.W., Ha, Y.-C., Han, S.H., Kim, H.-S.: Assessment of Korea's Orthopedic Research Activities in the Top 15 Orthopedic Journals, 2008–2017. *Clin Orthop Surg*. 11, 237–243 (2019). <https://doi.org/10.4055/cios.2019.11.2.237>
33. Shi, X., Cai, L., Jia, J.: The Evolution of International Scientific Collaboration in Fuel Cells during 1998–2017: A Social Network Perspective. *Sustainability*. 10, 4790 (2018). <https://doi.org/10.3390/su10124790>
34. Heneberg, P.: From Excessive Journal Self-Cites to Citation Stacking: Analysis of Journal Self-Citation Kinetics in Search for Journals, Which Boost Their Scientometric Indicators. *PLoS One*. 11, e0153730 (2016). <https://doi.org/10.1371/journal.pone.0153730>
35. B S, M., Rajgoli, I.: Mapping of Scholarly Communication in Publications of the Astronomical Society of Australia, Publications of the Astronomical Society of Japan, and Publications of the Astronomical Society of the Pacific: A Bibliometric Approach. *Science & Technology Libraries*. 36, 351 (2017). <https://doi.org/10.1080/0194262X.2017.1368427>
36. Wang, Z., Chen, Y., Glänzel, W.: Preprints as accelerator of scholarly communication: An empirical analysis in Mathematics. *Journal of Informetrics*. 14, 101097 (2020). <https://doi.org/10.1016/j.joi.2020.101097>
37. Ranaei, S., Suominen, A., Porter, A., Carley, S.: Evaluating technological emergence using text analytics: two case technologies and three approaches. *Scientometrics*. 122, 215–247 (2020). <https://doi.org/10.1007/s11192-019-03275-w>
38. Leidolf, A., Baker, M., Porter, A., Garner, J.: Measuring and Visualizing Research Collaboration and Productivity. *Journal of Data and Information Science*. 3, 54–81 (2018). <https://doi.org/10.2478/jdis-2018-0004>
39. Barrot, J.: Research impact and productivity of Southeast Asian countries in language and linguistics. *Scientometrics*. 110, (2016). <https://doi.org/10.1007/s11192-016-2163-3>
40. Osueke, C.O., Idiegbeyan-ose, J., Botu, T., Aregbesola, A., Emmanuel, O.: Analysis of mechanical engineering research activities using bibliometric method: A case study of un-

- dergraduate projects. *International Journal of Mechanical Engineering and Technology*. 9, 1014–1021 (2018)
41. Tukur, L.M., Kannan, S.: An Appraisal of Availability and Utilization of Information Resources and Library Services by Undergraduate Students in Three Agriculture University Libraries in Northern Nigeria (ABU, FUAM, FUTM). 27
 42. J Proctor, J. Knowledge management for collection development: Transforming institutional knowledge into tools for selectors. *Serials Librarian*, 76(1–4), 118–122. <https://doi.org/10.1080/0361526X.2019.1551668>. (2019).
 43. Zhao, W., Zhao, S., MacGillivray, K.-S.: Providing and Maintaining Access to Electronic Serials: Consortium and Member University Library's Perspectives. *The Serials Librarian*. 72, 144–151 <https://doi.org/10.1080/0361526X.2017.1309831> (2017).
 44. Babatunde, T., Alhassan, J., Dr, B.: Effective Resource Sharing Services In University Libraries In North Central Nigeria. (2020)
 45. Andalia, R.: Los análisis de citas en la evaluación de los trabajos científicos y las publicaciones seriadas. 7, (1999)
 46. Miguel, S., Moya-Anegón, F., Herrero-Solana, V.: El análisis de co-citas como método de investigación en Bibliotecología y Ciencia de la Información. *Investigación bibliotecológica*. 21, 139–155 (2007)
 47. Galvez, C.: Análisis de co-palabras aplicado a los artículos muy citados en *Biblioteconomía y Ciencias de la Información* (2007-2017). *Transinformação*. 30, 277–286 (2018). <https://doi.org/10.1590/2318-08892018000300001>
 48. Alvarado, R.U.: La ley de Lotka y la literatura de bibliometría. *Investigación bibliotecológica*. 13, 125–141 (1999)
 49. Urbizagástegui Alvarado, R., Urbizagástegui Alvarado, R.: El crecimiento de la literatura sobre la ley de Bradford. *Investigación bibliotecológica*. 30, 51–72 <https://doi.org/10.1016/j.ibbai.2016.02.003> (2016).
 50. Urbizagástegui, R., Restrepo Arango, C.: Zipf's law and Goffman's transition point in the automatic indexing. *Investigación Bibliotecológica*. 25, 71–92 (2011)
 51. Alvarez-Jareño, J., Badal-Valero, E., Pavia, J.: Aplicación de métodos estadísticos, económicos y de aprendizaje automático para la detección de la corrupción. 9, (2019)
 52. Torres Berrú, Y., Batista, V., Torres-Carrion, P.: Data Mining to detect and prevent corruption in contracts: Systematic Mapping Review. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*. 13–25 (2020)
 53. Marcos-Pablos, S., García-Peñalvo, F.J.: Information retrieval methodology for aiding scientific database search. *Soft Comput.* 24, 5551–5560 (2020). <https://doi.org/10.1007/s00500-018-3568-0>
 54. Pandey, S., Sahoo, S.: Research Collaboration and Authorship Pattern in the field of Semantic Digital Libraries. *DESIDOC Journal of Library & Information Technology*. 40, 375
 55. Sahoo, S., Pandey, S.: Evaluating research performance of Coronavirus and Covid-19 pandemic using scientometric indicators. *Online Information Review*. 44, 1443–1461 (2020). <https://doi.org/10.1108/OIR-06-2020-0252>
 56. Yi, W., Wang, Y., Tang, J., Xiong, X., Zhang, Y., Yan, S.: [Visualization analysis on treatment of coronavirus based on knowledge graph]. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue*. 32, 279–286 (2020). <https://doi.org/10.3760/cma.j.cn121430-20200225-00200>