

Adopción de software de Business Intelligence: Una revisión sistemática de literatura aplicando minería de texto

Adoption of Business Intelligence software: A systematic literature review applying text mining

Cristian Hugo MORALES Alarcón [1](#); Ciro Diego RADICELLI García [2](#); María Fernanda JARAMILLO Pinos [3](#); Elba María BODERO Poveda [4](#)

Recibido: 29/01/2018 • Aprobado: 01/03/2018

Contenido

[1. Introducción](#)

[2. Metodología](#)

[3. Resultados](#)

[4. Conclusiones](#)

[Referencias bibliográficas](#)

RESUMEN:

Este estudio realizó una revisión sistemática de literatura (SLR) para la adopción de software de Business Intelligence, se analizaron 367 trabajos publicados entre 2010 y 2015 indizados en cuatro bases datos científicas. Este trabajo es de tipo documental, utilizó el protocolo para una SLR y aplica minería de texto en varias actividades de este procedimiento. Obteniendo como resultado el ranking de: plataformas de inteligencia de negocios más utilizadas, populares y sectores que aplican con mayor frecuencia software de BI.

Palabras-Clave: Inteligencia de negocios, Minería de Texto, Revisión Sistemática de Literatura, WordStat

ABSTRACT:

This study carried out a systematic literature review (SLR) for the adoption of Business Intelligence software, analyzing 367 papers published between 2010 and 2015 indexed in four scientific data bases. This work is documentary type, used the protocol for an SLR and applies text mining in several activities of this procedure. Obtaining as a result a ranking of: most used, popular business intelligence platforms and sectors that most frequently apply BI software.

Keywords: Business Intelligence, Text Mining, Systematic Literature Review, WordStat

1. Introducción

La elección de plataformas de inteligencia de negocios en las organizaciones antes de adquirirlas e implementarlas es de vital importancia, estos sistemas deben ser valorados con la finalidad de que puedan apoyar de forma concreta a las decisiones de los gerentes en las organizaciones (Rouhani, Ghazanfari, & Jafari, 2012). El objetivo central de esta investigación es, realizar una revisión sistemática de literatura (SLR) aplicando minería de

texto en la adopción de software de Inteligencia de Negocios.

El término de Business Intelligence (BI), ha existido desde 1958 cuando Hans Peter Luhn lo utilizó para describir un sistema automático para la difusión de información mediante la utilización del computador para el tratamiento de datos, y así, abstraer, codificar y archivar la información de una organización (Luhn, 1958). BI se define ampliamente como el proceso de transformar datos en información útil para la generación de ideas estratégicas y operativas más eficaces y de esta manera facilitar la toma de decisiones brindando beneficios reales a la empresa (Duan & Da Xu, 2012). Así, mediante la inteligencia de negocios se obtiene conocimiento nuevo que puede contribuir a tomar mejores decisiones empresariales (Washington, DC: U.S Patente nº 9,183,529, 2015).

Un sistema de BI se considera una herramienta que permite analizar y procesar los datos existentes de la empresa y convertir los datos en conocimiento (Lv, Xie, Wang, & Cheng, 2012), además los sistemas de inteligencia empresarial brindan apoyo y mejoran los procesos de toma de decisiones en las organizaciones (Arnott & Pervan, 2005), es por esta razón que analizar la preferencia de software de Business Intelligence en los diferentes sectores de la industria es de vital importancia, con la finalidad de maximizar la probabilidad de éxito en la implementación de un sistema de estas características.

Business Intelligence incluye: Lógica de consulta asociativa, cuadro de mando integral (scorecarding), gestión del rendimiento empresarial, planificación empresarial, reingeniería de procesos de negocio, análisis competitivo, gestión de relaciones con clientes (CRM), minería de datos (DM), cultivo de datos y almacenes de datos, sistemas de apoyo a las decisiones (DSS), almacenes y gestión de documentos, sistemas de gestión empresarial, sistemas de información ejecutivos (EIS), finanzas y presupuestación, gestión de recursos humanos, gestión del conocimiento, sistemas de información de gestión (SIG), sistemas de información geográfica (SIG), procesamiento analítico en línea (OLAP) y el análisis multidimensional, inteligencia de negocios en tiempo real, estadísticas y datos técnicos de análisis, gestión de la cadena de suministro, gestión de la cadena de la demanda, sistemas de inteligencia, análisis de tendencia, web mining, consulta de usuario final y presentación de informes (Ranjan, 2009).

Por su parte, una Revisión Sistemática de Literatura (SLR) se define como un mecanismo para analizar e interpretar una investigación, responde a una o varias preguntas de investigación, fenómeno o área de interés. La SLR tiene la finalidad de realizar una evaluación justa, mediante la utilización de una metodología fiable, rigurosa y auditable (Kitchenham & Charters, 2007). Una Revisión Sistemática de Literatura identifica vacíos disponibles para efectuar una investigación, resume la evidencia existente de una temática definida previamente y provee un marco de antecedentes para investigaciones futuras (Kitchenham, 2004), además, una SLR se ha convertido en una metodología ampliamente aplicada en el ámbito de la ingeniería de software (Marshall & Brereton, 2013).

El proceso de revisión de los casos de estudio a nivel científico debe ser riguroso y cuantificable, para evitar resultados erróneos y un posible sesgo de la investigación, para ello es necesario la aplicación de una SLR, que permitirá responder a una o varias preguntas de investigación. Debido a la naturaleza amplia y rigurosa de los trabajos de investigación, el realizar una SLR en una organización es difícil y requiere mucho tiempo, porque las actividades se llevan a cabo de forma manual, razón por lo cual es necesario disponer de herramientas o técnicas que apoyen la realización de una SLR (Felizardo et al., 2011).

La minería de texto se define como la extracción automática de información a partir de texto, la cual se encuentra previamente desconocida y que es potencialmente útil (Sahadevan, Hofmann-Apitius, & Schellander, 2012). Es así que una revisión sistemática de literatura (SLR) puede apoyarse en la utilización de técnicas de text mining y así contribuir a responder una serie de preguntas de investigación, como las planteadas en este estudio las cuales se encuentran relacionadas a la adopción de software de Business Intelligence.

1.1. Trabajos relacionados

En esta sección se detallan en un principio los estudios generales, hasta llegar a los

particulares y más íntimamente relacionados con este tema de investigación, permitiendo enfocar este estudio en el contexto correcto.

"Selection and deployment of a Business Intelligence system (BI) at a hospital's Clinical Engineering Department" (Pérez-Martín et al., 2014), el objetivo de este estudio fue proveer directrices para seleccionar y aplicar una plataforma de Business Intelligence comercial en el Departamento de Ingeniería Clínica de un hospital, se examinó varias opciones, las dos plataformas preseleccionadas para su posterior análisis, fueron Business Objects de SAP y Pentaho, debido a razones de índole económica este último fue finalmente seleccionado.

"Directions in Business Intelligence: An analysis of applications" (Hayen, 2008), este trabajo estudia los frameworks de Sistemas de Soporte a la Decisión (DSS), con la finalidad de definir un conjunto de características útiles en aplicaciones basadas en de casos de BI.

"Open source alternatives for Business Intelligence: Critical success factors for adoption" (Zhao, Navarrete, & Iriberry, 2012, la finalidad de este trabajo investigación fue identificar cuales son los factores críticos que afectan a la adopción de software de código abierto de inteligencia de negocios y además comparar las diferencias con herramientas propietarias.

"The usage of open-source Business Intelligence in the Czech Republic" (Nemec & Menclova, 2011). Esta investigación presenta un estudio relacionado a la adopción de software de Business Intelligence en varias industrias, concuerda con el objetivo de este trabajo de investigación.

En los estudios presentados no se ha aplicado una SLR y no se ha hecho uso de técnicas de minería de texto, además, utilizan una fuente de información distinta a este trabajo de investigación, sin embargo, direccionan la adopción de software de Business Intelligence y en parte los sectores de aplicación de BI.

Relacionado a la aplicación de revisiones sistemáticas de literatura y selección de herramientas de inteligencia de negocios se puede destacar varios estudios como:

"Business Intelligence: An analysis of the literature" (Jourdan, Rainer, & Marshall, 2008). Ese estudio recopila 167 trabajo sobre una variedad de temas relacionados con la inteligencia de negocios publicada en 1997 y 2006 en diez publicaciones líderes de sistemas de información (SI).

"Business Intelligence in construction: A review" (Shi, Peng, & Xu, 2012). Presenta el desarrollo de la inteligencia empresarial en los últimos años. Arquitectura, tecnologías, evaluación del rendimiento y aplicaciones, se discuten cuatro aspectos para ilustrar la tecnología utilizada de BI y la tendencia futura.

Estos estudios se diferencian con este trabajo de investigación, que no se apoyan en técnicas de text mining, el periodo de análisis, las fuentes de información y las preguntas de investigación son distintas en este estudio.

Existen varios trabajos que se relacionan de forma más específica a los objetivos de esta investigación que es la revisión sistemática de literatura (SLR) con la aplicación de técnicas de minería de texto, estos proporcionan información útil para el desarrollo de este estudio. Se presentan a continuación trabajos que contemplan investigaciones en el campo de la SLR, minería de texto y de forma conjunta en estas temáticas, estos son los siguientes:

"A Systematic Mapping on the use of Visual Data Mining to Support the Conduct of Systematic Literature Reviews" (Felizardo et al, 2012), este estudio tuvo por objetivo recoger y evaluar las pruebas sobre el uso de la técnica de minería de datos visual, para apoyar el proceso de la SLR.

"Text mining Business Intelligence: a small sample of what words can say" (Ishikiriya, Miro, & Gomes, 2015). Este documento aplica una metodología dividida en dos pasos: buscar documentos de literatura y al uso de una técnica de minería de texto, analiza 35 documentos relacionados a la inteligencia de negocios mediante el software R-project.

"Business Intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation" (Moro, Cortez, & Rita, 2015). Este trabajo analiza 219 documentos de las revistas más relevantes en el ámbito de BI para la industria bancaria, publicados entre 2002 y 2013.

“Using Visual text mining to Support the Study Selection Activity in Systematic Literature Reviews” (Felizardo, Salleh, Martins, Mendes, & MacDonell, 2011). Se propone un enfoque denominado Revisión Sistemática de Literatura basada en minería de datos visual (SLR-VTM), para apoyar a la actividad primaria de selección de estudios utilizando técnicas de minería de texto visual (VTM). Para realizar este estudio se apoya en cuatro estudiantes de doctorado en la selección principal de estudios de forma manual y utilizando el método SLR-VTM.

Es necesario diferenciar varios aspectos relevantes entre los trabajos presentados y este estudio: estos estudios no se han enfocado en la adopción de software de inteligencia de negocios con una revisión sistemática de literatura aplicando técnicas de minería de texto, el protocolo que utilizan para la revisión sistemática de literatura difiere al de este estudio, utilizan herramientas de software diferentes para realizar la minería de texto, hacen uso de técnicas de minería de texto principalmente en la selección de los estudios primarios y en esta investigación, además, se aplica en la creación de cadenas de búsqueda y en la depuración de estudios candidatos.

El resto del artículo se encuentra organizado de la siguiente manera: En la Sección 2, se explica la metodología de investigación, y se describe el proceso que se lleva a cabo en la revisión sistemática de literatura aplicando técnicas de text mining. La sección 3 presenta los resultados de la revisión y analiza los datos que se ha recogido en la revisión. Y finalmente, la Sección 4, concluye el documento respondiendo a las preguntas de investigación planteadas.

2. Metodología

Este trabajo de investigación es de tipo documental, debido a que hace uso de una gran cantidad de documentos analizados a través de una revisión sistemática de literatura (SLR), los cuales permiten profundizar y ampliar el conocimiento relacionado a la adopción de software de inteligencia de negocios, estos documentos (artículos, libros y capítulos) son obtenidos de la exploración de fuentes bibliográficas de las bases de datos científicas: SCOPUS, Science Direct, IEEE Xplore y EBSCO. En esta investigación se utilizó además un método analítico, con la finalidad examinar cada uno de los elementos del fenómeno de estudio ordenadamente por separado.

Este estudio presenta un protocolo de SLR basada en la metodología desarrollada por Kitchenham and Charter´s (Kitchenham & Charters, 2007) y complementada con el estudio “Conceptualización e Infraestructura para la Investigación Experimental en Ingeniería del Software” (Fonseca Carrera, 2014) de esta forma se cumplió la rigurosidad científica que conlleva responder a las preguntas de investigación.

En esta investigación se aplica el método inductivo con la finalidad de analizar un caso particular (la adopción de software de Business Intelligence) a partir de procedimientos y técnicas probadas a nivel general (protocolo de revisión y técnicas de minería de texto). Los instrumentos y técnicas que se utilizar en este estudio se encuentran validadas por una gran cantidad de investigadores en varios trabajos de investigación, algunos ejemplos pueden observarse en la sección de trabajos relacionados. Se han utilizado varias matrices de datos en la SLR, las mismas cumplen con el protocolo de revisión establecido.

Como apoyo en el desarrollo de la revisión sistemática de literatura de este estudio intervinieron cuatro investigadores voluntarios, los cuales fueron capacitados en la cátedra de Gestión de Conocimiento, dictada en la Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios de la Universidad de las Fuerzas Armadas-ESPE, a los cuales se agradece su participación, estos viabilizaron el desarrollo de varias actividades de la SLR y permitieron realizar un proceso de arbitraje usando una validación cruzada en las actividades cruciales de la SLR.

El procedimiento para realizar la SLR, presenta varias actividades, estas fueron analizadas con la finalidad de determinar en cuales de estas es factible la aplicación de text mining, se pudo determinar que es viable la aplicación de minería de texto en: la construcción de la cadena de búsqueda, en la depuración de estudios candidatos y en la extracción de

características de los estudios primarios.

En esta investigación, se utilizó la herramienta WordStat 7.1, paquete complementario al software QDA Miner 4.0. Así el proceso para realizar text mining comprendió: La colección de documentos de texto, en esta fase se recopilaron los trabajos de investigación en relación al tema estudio, estos formatos fueron descargados con la extensión .pdf. Posteriormente se realizó el pre-procesamiento de texto, que consta de la selección, limpieza y un pre-procesamiento de la información de las bases de datos no estructuradas, transformando este contenido en representaciones estructuradas para facilitar su análisis. Posteriormente se explotó la estructura sintáctica y su semántica, utilizando diferentes representaciones como: caracteres, palabras, términos o conceptos de los documentos (Rai & Vijaya Murari, 2014). La siguiente fase fue la de procesamiento de texto, en la cual se identificó las relaciones conceptuales, proporcionando características significativas. Finalmente se realizó el análisis de texto, para la evaluación de la información obtenida en las fases anteriores, lo que permitió determinar si fue descubierto nuevo conocimiento y su importancia en el estudio.

2.1. Preguntas de investigación

La definición de las preguntas de investigación es una actividad que fue crucial para cumplir el objetivo de una revisión sistemática de literatura. En este contexto, esta revisión sistemática de literatura está dirigida a responder tres preguntas de investigación: **RQ1**. ¿Cuáles son las plataformas de Business Intelligence más utilizadas? **RQ2**. ¿Cuáles son las plataformas de Business Intelligence más populares? **RQ3**. ¿En qué tipo de industrias o sectores son más utilizados los Sistemas de Business Intelligence?

RQ1. y RQ2. buscan definir un ranking de las herramientas de inteligencia de negocios, la primera con el objetivo de encontrar las plataformas que más se aplican en los estudios recuperados y la segunda busca definir el software de inteligencia de negocios que se presenta con mayor frecuencia en los trabajos recopilados; para responder a estas preguntas fue creado un diccionario de datos con la mayoría de herramientas de Business Intelligence que se encuentran en el mercado, esto fue realizado a partir de la revisión de múltiples sitios web y trabajos de investigación.

RQ3. Esta pregunta se planteó para definir las industrias que más aplican BI, estos sectores han sido creados según la clasificación expuesta por varios artículos científicos que fueron mencionados en la Sección de trabajos relacionados de este documento; también, se ha considerado la clasificación de industrias que presentan varias empresas de BI como: Qlik Tech, Microsoft Corporation, Tableau, entre otras. Además, se ha considerado la clasificación de sectores de BI de la empresa Technology Evaluation Center. Es así que se ha planteado el siguiente listado de sectores de aplicación de BI que permite clasificar a todos los trabajos científicos analizados, estos fueron: 1) administración y gobierno; 2) construcción; 3) economía, finanzas y valores; 4) educación; 5) empresas de fabricación; 6) energía, minas y petróleo; 7) industria minorista; 8) productos de consumo; 9) salud; 10) seguros; 11) tecnologías de la información; 12) telecomunicaciones y 13) transportación.

2.2. Criterios de inclusión y exclusión

Los criterios de inclusión y exclusión, fueron planteados y discutidos por quienes condujeron esta investigación. Estos criterios permiten definir los requisitos que debieron cumplir los estudios para ser considerados válidos, tiene como finalidad analizar una muestra representativa de trabajos de investigación, cuyos resultados permitan responder de forma correcta las preguntas de investigación planteadas.

Se incluyeron trabajos: de cinco años atrás a la fecha de revisión en el periodo 2010 a 2015 (libros, capítulos de libros, artículos y ponencias), en el caso de este último, siempre y cuando los mismos hayan tenido una estructura de artículo científico y hayan estado disponibles en la web. Además, trabajos de investigación que especificaron: el sector de aplicación, la herramienta utilizada y los resultados que se obtuvieron al llevar a cabo la implementación. Para esto se utilizaron las siguientes bases de datos para la búsqueda de trabajos relacionados: SCOPUS, Science Direct, IEEE y EBSCO, las cuales son reconocidas a

nivel mundial por su calidad.

Se ha excluido en esta revisión: Trabajos de investigación que no poseen un título específico, como por ejemplo, estudios que especifican únicamente el nombre del congreso o seminario, en lugar del título del trabajo de investigación. Además, análisis comparativos, revisiones de literatura que presentan resultados similares a este estudio. También se excluyen trabajos que no presentaron herramientas de inteligencia de negocios aplicados a un estudio en particular. Finalmente, estudios cuyo contenido se encontraba en idiomas diferentes al inglés, español, o artículos que contaban con un aviso de retracción.

2.3. Conformación del grupo de control

Se realizó una revisión previa en la base de datos SCOPUS para constatar que existan estudios relacionados que puedan responder las preguntas de investigación. Se solicitó dos artículos por investigador que cumplan con los criterios de inclusión y exclusión. Posteriormente, se realizó una validación cruzada que permitió garantizar que los estudios cumplieran con los requerimientos necesarios, como resultado se realizó un listado de estudios del grupo de control con un total de ocho casos.

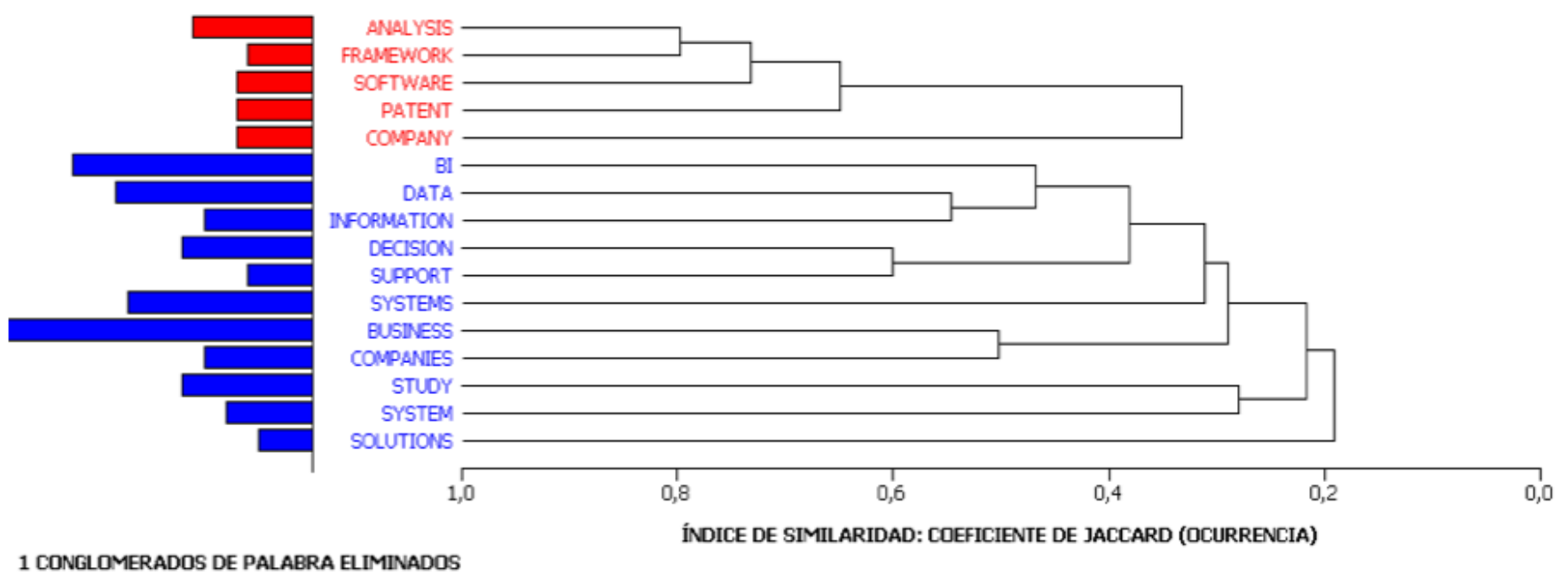
2.4. Construcción de la cadena de búsqueda

En esta actividad para generar la cadena de búsqueda aplicando text mining, se utiliza la frecuencia de documentos, para obtener un resumen estadístico de las palabras que aparecen un mayor número de veces, así como el número de casos en donde se encuentra contenida la palabra y el total de estas en los segmentos de texto que se encuentran asociados. Además, se utilizó una agrupación jerárquica (Hierarchical Clustering) a través de gráficos de árbol (dendrogramas) y se usó un gráfico de red para el análisis de los vínculos en las palabras.

Se realizó el análisis de los trabajos de investigación del grupo de control a nivel del título, resumen y palabras clave, con el objetivo de extraer términos generales, y términos que son comunes entre estudios, pero además que sean coherentes con el objetivo de la revisión sistemática de literatura, para ello se creó un listado de artículos del grupo de control. Con este listado se realizó un análisis de contenido con la herramienta WordStat 7.1, el cual es un paquete complementario del software QDA Miner 4.0, se aplica un análisis descriptivo, el cual fue filtrado con una frecuencia superior a cinco, valor que es recomendado por la herramienta. La Figura 1, muestra un dendrograma, en donde el eje vertical ubicado en la parte izquierda se compone de las palabras del grupo de control y el eje horizontal representa los grupos de palabras formados por el procedimiento de agrupamiento. Las palabras que tienden a aparecer juntas son combinadas en una etapa inicial, mientras que las que son independientes las unas de las otras tienden a ser combinadas al final del proceso de aglomeración.

Figura 1

Agrupación jerárquica de las palabras del grupo de control

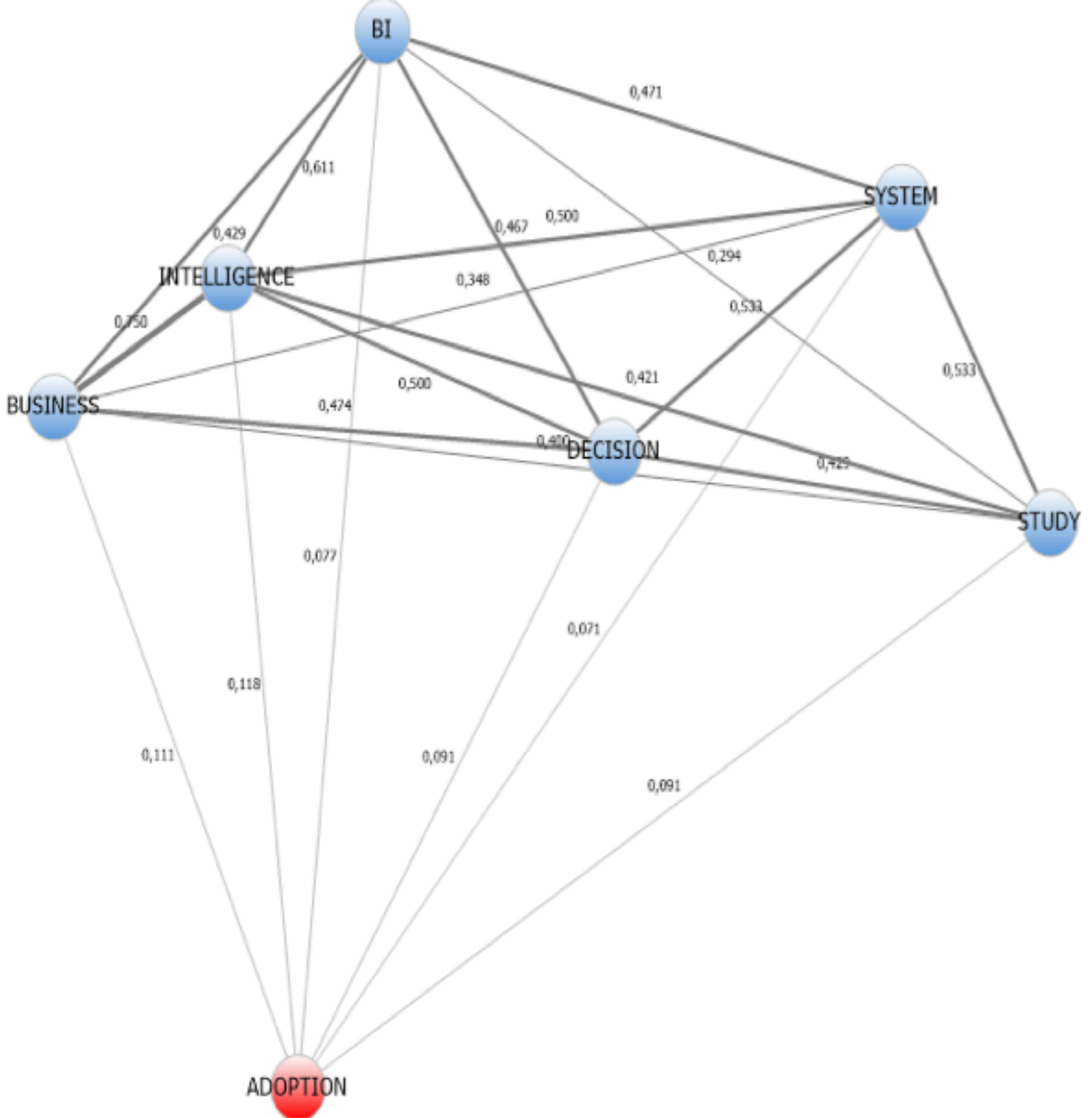


Fuente: Elaboración Propia

En la Figura 2, se puede observar un gráfico de red que permite la exploración de relaciones a través de las conexiones entre las palabras encontradas en el listado del grupo de control, este gráfico permitió detectar los patrones y estructuras de coocurrencias, que resultará en una vista de red de sus elementos, donde cada uno de ellos es representado como un nodo y su relación es presentada como una línea que conecta a los nodos y su distancia significa la fuerza de vínculo entre las palabras encontradas.

Figura 2

Escalamiento multidimensional de las palabras del grupo de control



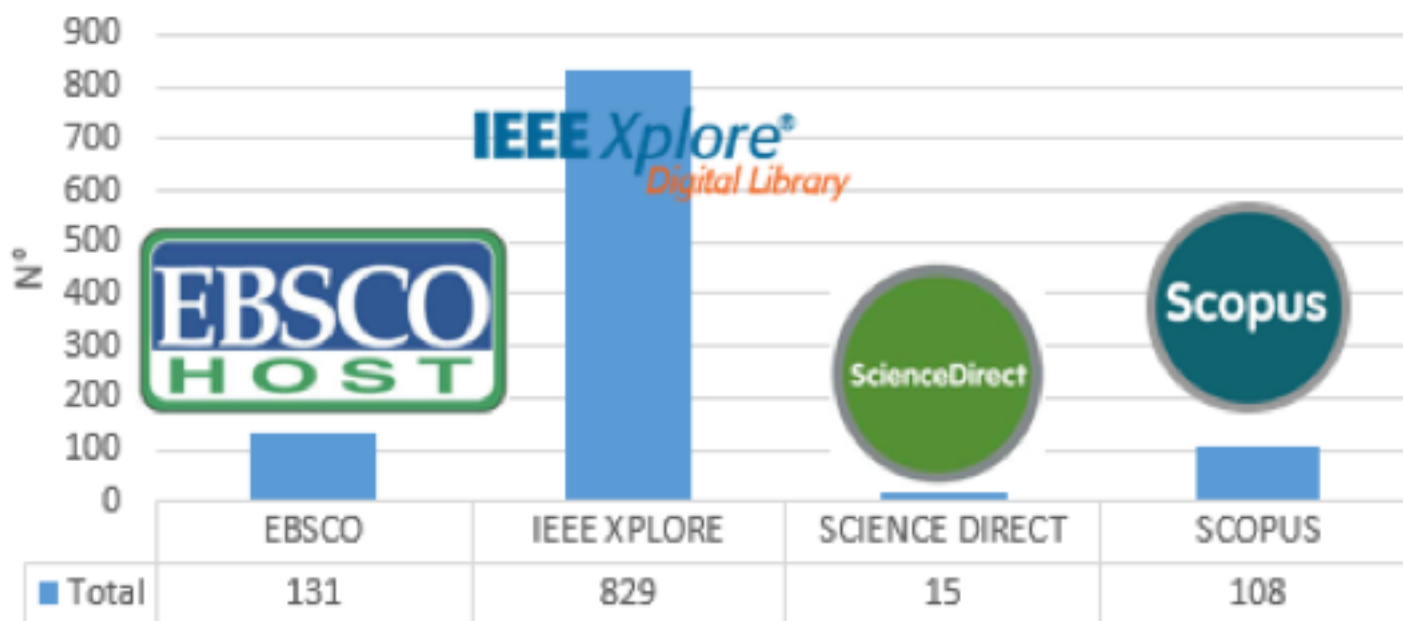
Fuente: Elaboración Propia

Una vez que se obtuvo el dendrograma y el gráfico de red, se dio la libertad a cada investigador para que plantee una cadena búsqueda. Estas cuatro cadenas fueron aplicadas en la base de datos de SCOPUS, y se analizó el número de trabajos del grupo de control devueltos por cada una de ellas, es así, que se eligió la cadena de búsqueda "(BUSINESS INTELLIGENCE OR BI) AND (SYSTEMS OR SYSTEM OR SOLUTIONS OR SOFTWARE) AND (COMPANY OR COMPANIES) AND (STUDY) AND (DECISION)", debido a que la misma devolvió un mayor número de artículos del grupo de control.

2.5. Búsqueda de estudios candidatos

Con la cadena de búsqueda definida, se configuró para que funcione en cada una de las bases de datos científicas, la Figura 3. indica el número de estudios devueltos, haciendo un total de 1083 estudios candidatos, en la cual se puede observar que el mayor número de trabajos obtenidos es de la base de datos IEEE Xplore y en contraste la que menos trabajos devolvió fue la de Science Direct.

Figura 3
Estudios devueltos

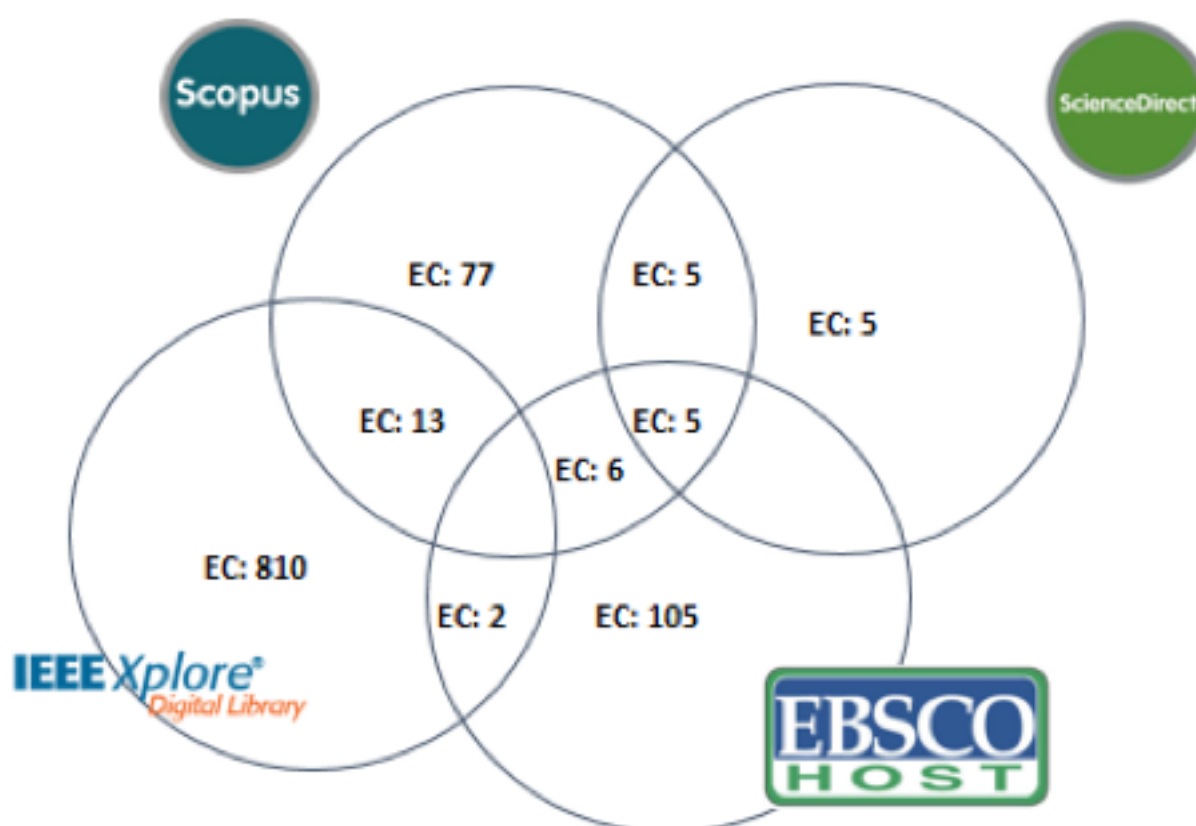


Fuente: Elaboración Propia

2.6. Depuración de estudios candidatos

En el proceso de depuración de las bases de datos se utilizó la herramienta WordStat, se hizo uso de la ficha de extracción, para posteriormente utilizar la funcionalidad de "Palabra clave en el contexto". La configuración se realizó con un mínimo de cinco, que se consideró como el número más corto de palabras que puede contener el título de un trabajo científico, se asigna además un número máximo de nueve palabras, dato que posterior a realizar varios intentos, fue el valor que más se acercó al resultado deseado, finalmente se estipula una frecuencia mínima de dos, debido a que se busca eliminar los estudios duplicados. Posteriormente se realizó la unificación de las bases de datos utilizando el mismo procedimiento con la herramienta WordStat. Así los estudios que tuvieron menos de cinco palabras se analizaron por separado, de forma manual. La Figura 4 muestra los estudios candidatos en donde se pueden observar los trabajos que se replican en las bases de datos científicas.

Figura 4
Estudios candidatos

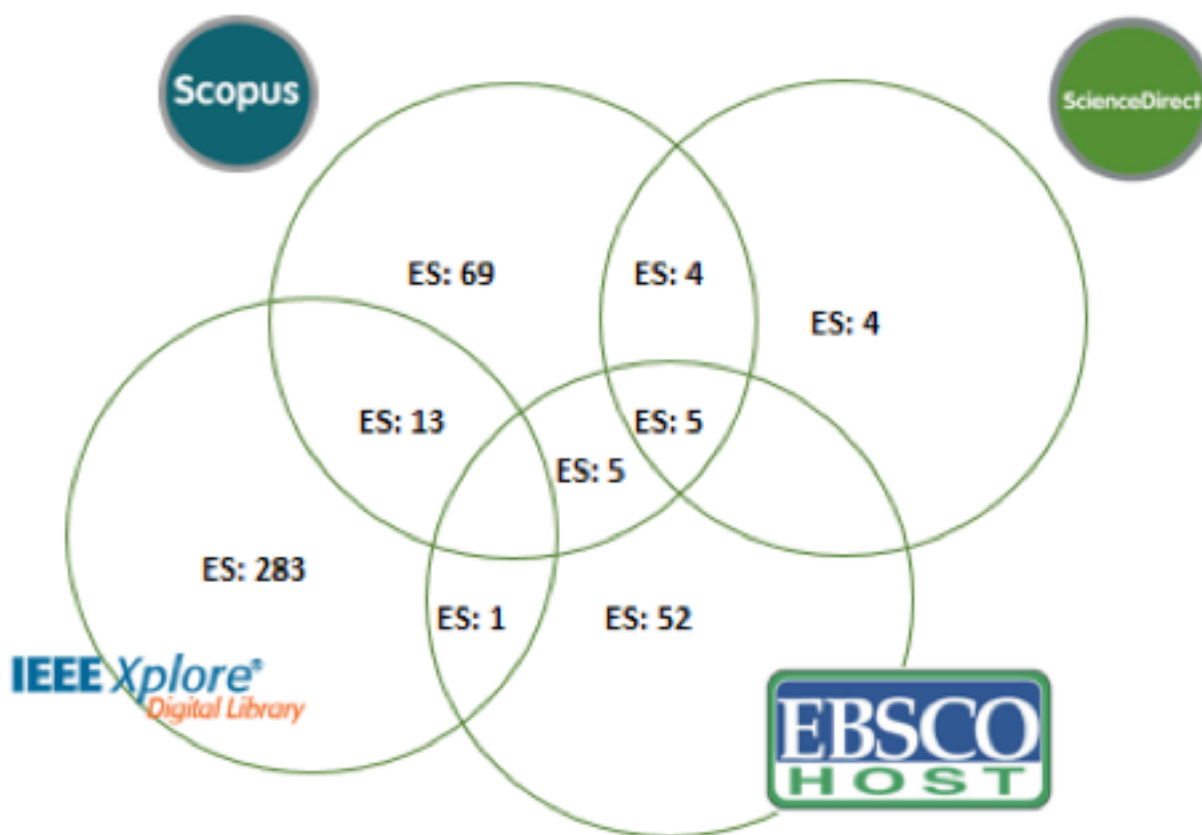


Fuente: Elaboración Propia

2.7. Selección de estudios

En esta subfase no se aplicó text mining, debido a que en los casos de estudio en su mayoría las herramientas de BI utilizadas no son mencionadas únicamente en el título, resumen o palabras clave; sino probablemente en la descripción completa del documento y al no tener recuperados la totalidad de los trabajos, aplicar técnicas de text mining para esta subfase, podía conllevar a obtener resultados erróneos. Tomando en consideración este precedente, se subdividió 257 artículos por cada investigador, para realizar la actividad de selección de estudios, los cuales revisaron el título, resumen y palabras clave, de los estudios candidatos depurados. Posteriormente y con la finalidad de evitar errores, se realizó una validación cruzada entre dos investigadores, los cuales validaron la opinión de su pareja investigador, en el caso de discrepancia, quienes condujeron esta investigación tomaron la decisión de incluir o no un trabajo. La Figura 5, indica los estudios seleccionados por base de datos científica, se han eliminado 436 trabajos duplicados.

Figura 5
Estudios seleccionados



Fuente: Elaboración Propia

2.8. Selección de estudios primarios

Se unificó el listado de estudios seleccionados entregados por los investigadores, de las cuales, representa un mayor número de estudios IEEE Xplore, seguido de Scopus, posteriormente de EBSCO y finalmente Science Direct. La obtención de los trabajos se realizó en primera instancia de todos los artículos de libre descarga, posteriormente se acudió a varias universidades con las cuales han adquirido acceso estas bases de datos científicas: La Universidad Nacional de Chimborazo, la Escuela Superior Politécnica de Chimborazo, la Universidad de las Fuerzas Armadas-ESPE y la Universidad Técnica de Ambato; además, los estudios fueron recuperados a través del contacto con investigadores que poseen membresías de las bases de datos científicas, obteniendo un total de 407 trabajos de 436 estudios totales es decir un 93% de estudios fueron recuperados, lo cual es una cifra aceptable para continuar con la revisión sistemática de literatura, únicamente 29 trabajos que corresponde al 7%, no pudieron ser recuperados, debido a que la base de datos pertenecía a una organización privada o en su defecto el enlace no se encontraba disponible. Una vez descargados los trabajos científicos de las diferentes bases de datos, se realizó una revisión de estos, principalmente para excluir varios de los que no cumplían los

criterios de idioma, la estructura general de un artículo científico o que poseían un aviso de retracción, estos hacen un total de 40 artículos, finalmente 367 trabajos científicos fueron los estudios primarios.

2.9. Extracción de características

Consistió en la realización de un análisis detallado de los trabajos obtenidos enfocados al objetivo de la investigación. Para lo cual se creó un listado de herramientas de BI, mismas que hacen un total de 170. Para el estudio de detallado de los estudios, se utilizó el software WordStat de QDA Miner, se agregó el diccionario con las 170 herramientas de BI, y se procedió a la recuperación de palabras clave. Una vez aplicada la recuperación de palabras y el método de las palabras claves en el contexto, se eliminaron las entradas correspondientes a las referencias bibliográficas, cuya descripción no representaba contenido válido de análisis. Posteriormente, se procedió a identificar que los casos devueltos correspondan a nombres propios, de esta manera se pudo identificar el nombre de una herramienta y no otro tipo de palabra, así por ejemplo, se eliminaron los registros que se encontraban con spectrum en lugar de Spectrum o SPECTRUM, pilot en lugar de Pilot o PILOT, etc. Se consideró también la herramienta que aparece con mayor frecuencia, con la finalidad de determinar con mayor precisión qué herramienta fue utilizada. Una vez finalizado este procedimiento se realizó el conteo de casos y se obtuvo un ranking de herramientas de BI más utilizadas, que contestó a la primera pregunta de investigación RQ1. se realizó además la comprobación manual revisando el corpus del texto devuelto por la herramienta de minería de texto. Para contestar a la segunda pregunta RQ2. las herramientas de BI más populares, bastó con ordenar por frecuencia el número de casos devueltos por la herramienta WordStat.

Para los sectores de aplicación de las herramientas de inteligencia de negocios RQ3. se creó un diccionario con 265 reglas las cuales se compusieron de un nombre, una palabra objetivo que en este caso estuvo relacionada con los sectores de aplicación de BI, un operador que en este caso fue "CERCA", que se utilizó con el objetivo de devolver el mayor número de coincidencias, finalmente se utilizó la palabra de enlace ("sector", "industry", "enterprise", "company" y "firm") que se ha establecido a una distancia máxima de cinco, que es el valor recomendado por la herramienta. Cabe destacar que en este proceso se combinó la aplicación de técnicas de text mining y una revisión por parte de los investigadores del corpus del texto devuelto por la herramienta WordStat, lo cual además permitió su comprobación.

3. Resultados

Una vez realizada la revisión sistemática de literatura, en la cual se ha aplicó técnicas de minería de texto en varias actividades de este procedimiento, en la Tabla 1., se muestra un resumen de la depuración de los estudios en las principales actividades de la SLR, en la cual se pudo observar que el número de estudios primarios analizados correspondió a 367 trabajos de investigación.

Tabla 1
Depuración de estudios de la SLR

ESTUDIOS	Número de trabajos científicos
CANDIDATOS	1083
DEPURADOS	1028
SELECCIONADOS	436
RECUPERADOS	407

Fuente: Elaboración Propia

3.1. Plataforma BI más utilizada

En la Tabla 2. se puede observar las 20 primeras posiciones de las plataformas de inteligencia de negocios, en donde prevalece la plataforma de Microsoft como la más utilizada de los estudios analizados, respondiendo a la pregunta **RQ1.** de esta revisión sistemática de literatura.

Tabla 2
Plataforma BI más utilizada

Posición	Plataforma para BI	Frecuencia de trabajos encontrados
1	MICROSOFT	42
2	IBM	39
3	MATHWORKS	16
4	SAP	12
5	ORACLE	8
6	ADAPTIVE	6
7	ARENA	5
8	ATLAS	5
9	ECLIPSE	5
10	PYTHON	4
11	QSR	4
12	APACHE	3
13	GAMS	3
14	SPECTRUM	3
15	INFOR	2
16	MINITAB	2
17	NEURO	2
18	QLIK TECH	2
19	SENTIWORDNET	2

Fuente: Elaboración Propia

3.2. Plataforma BI más popular

En la Tabla 3. Se muestran las 20 primeras posiciones de las plataformas de inteligencia de negocios, en donde prevalece también Microsoft como la plataforma más popular, dando respuesta a **RQ2.** de esta revisión sistemática de literatura.

Tabla 3
Plataforma BI más popular

Posición	Plataforma para BI	Frecuencia de trabajos encontrados
1	MICROSOFT	77
2	IBM	74
3	ORACLE	29
4	SAP	28
5	MATHWORKS	21
6	ADAPTIVE	12
7	INFOR	10
8	ARENA	8
9	ATLAS	7
10	ECLIPSE	7
11	GAMS	7
12	MICROSTRATEGY	7
13	QSR	7
14	PYTHON	6
15	QLIK TECH	5
16	SPECTRUM	5
17	TABLEAU	5
18	TERADATA	5
19	APACHE	4

Fuente: Elaboración Propia

3.3. Sectores de aplicación de BI

En la Tabla 4. se puede observar con exactitud la posición de todos los sectores en donde se aplicó con mayor frecuencia la inteligencia de negocios. Según la investigación realizada el sector de Tecnologías de Información, es la industria en donde se aplica el software de BI en mayor número, seguido de empresas de fabricación y organizaciones que se dedican al sector de energía, minas y petróleo. En contraste los sectores de educación y productos de consumo, son las industrias donde menos se aplica software de inteligencia de negocios.

Tabla 4
Sectores en los cuales se aplica BI con mayor frecuencia

Posición	Sector de aplicación de BI	Frecuencia de trabajos encontrados
1	TECNOLOGÍAS DE LA INFORMACIÓN	49
2	EMPRESAS DE FABRICACIÓN	47
3	ENERGÍA, MINAS Y PETRÓLEO	38
4	ADMINISTRACIÓN Y GOBIERNO	15
5	TELECOMUNICACIONES	14
6	ECONOMÍA, FINANZAS Y VALORES	13
7	TRANSPORTE	9
8	SALUD	8
9	CONSTRUCCIÓN	7
10	INDUSTRIA MINORISTA	7
11	SEGUROS	5
12	EDUCACIÓN	3
13	PRODUCTOS DE CONSUMO	3

Fuente: Elaboración Propia

4. Conclusiones

La SLR presenta varias actividades en las cuales se aplicaron minería de texto, para esto se hizo uso de la herramienta WordStat de QDA Miner, la cual fue crucial para el desarrollo de la revisión sistemática de literatura aplicando minería de texto, este software permitió: análisis de contenido, la revisión de texto no estructurado, escalamiento multidimensional de palabras, revisión de calidad y clasificación de información, en este proceso fue necesario disponer de un diccionario de herramientas de Business Intelligence y crear las reglas para

el análisis de contenido de los sectores de aplicación de BI.

Esta revisión sistemática de literatura analizó 1083 estudios, de los cuales 367 estudios primarios fueron examinados para responder a las preguntas de investigación planteadas. Sólo el 7% de los trabajos de investigación no pudieron recuperarse a texto completo, debido a que, la base de datos corresponde a una organización privada, no se tiene disponible el acceso gratuito a la base de datos, o el enlace no se encuentra disponible.

Una vez realizada la revisión sistemática de literatura se puede mencionar que las actividades en las cuales se pudo aplicar text mining fueron: i) construcción de la cadena de búsqueda, en la cual se creó una cadena de búsqueda acorde, para la recuperación de trabajos de investigación vinculados a este estudio. ii) en la depuración de estudios candidatos, para la eliminación de artículos duplicados o que presentaron un contenido similar. Fue necesario realizar una revisión manual de los estudios en los cuales existieron un número menor a cinco palabras en el contexto de la frase. Además, en la extracción de características se pudo utilizar los diccionarios y reglas creadas, la aplicación de técnicas de text mining en esta fase se considera crucial debido que es la actividad que mayor tiempo consume.

La plataforma de inteligencia de negocios de Microsoft fue la más utilizada y la más popular de los trabajos científicos revisados. El sector de Tecnologías de Información, seguido de Empresas de Fabricación fueron las industrias que aplican con mayor frecuencia software de Business Intelligence.

Finalmente, se considera que esta investigación puede ampliarse con la finalidad de realizar un análisis más profundo, en el cual se puede incorporar una clasificación por el tamaño de empresa, el tipo de organización (pública, privada), ubicación geográfica, etc. Estas categorías en un futuro pueden permitir realizar un análisis más detallado de la adopción de plataformas de inteligencia de negocios.

Referencias bibliográficas

- Arnott, D., & Pervan, G. (2005). A Critical Analysis of Decision Support Systems Research. *Journal of Information Technology*, 67–87.
- Duan, L., & Da Xu, L. (2012). Business intelligence for enterprise systems: a survey. *Industrial Informatics. IEEE Transactions on Industrial Informatics*, 679-687.
- Felizardo, K. R., MacDonell, S. G., Mendes, E., & Maldonado, J. C. (2012). A systematic mapping on the use of visual data mining to support the conduct of systematic literature reviews. *Journal of Software*, 450-461.
- Felizardo, K. R., Salleh, N., Martins, R. M., Mendes, E., & MacDonell, S. G. (2011). Using visual text mining to support the study selection activity in systematic literature reviews. *In 2011 International Symposium on Empirical Software Engineering and Measurement*, 77-86.
- Fonseca Carrera, E. R. (2014). *Conceptualización e Infraestructura para la Investigación Experimental en Ingeniería del Software (Doctoral dissertation, ETSI_Informatica)*. Madrid: Universidad Politécnica de Madrid.
- Hayen, R. (2008). Directions in business intelligence: An analysis of applications. *Americas Conference on Information Systems, AMCIS* , 656-665.
- Ishikiriyama, C. S., Miro, D., & Gomes, C. F. (2015). Text Mining Business Intelligence: a small sample of what words can say. *Procedia Computer Science*, 261-267.
- Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). Business intelligence: an analysis of the literature. *Information Systems Management*, 121-131.
- Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele, UK: Keele University.
- Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Keele: Software Engineering Group, School of Computer Science and Mathematics, Keele University.
- Luhn, H. (1958). A business intelligence system. *IBM Journal of Research and Development*

- Lv, H., Xie, Y., Wang, F., & Cheng, Y. (2012). Research on intelligent power consumption business intelligence system based on cloud computing. *In Computer Science & Service System (CSSS), 2012 International Conference on*, 1541-1546.
- Marshall, C., & Brereton, P. (2013). Tools to Support Systematic Literature Reviews in Software Engineering: A Mapping Study. *In 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 296-299.
- Moro, S., Cortez, P., & Rita, P. (2015). A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, 1314-1324.
- Nemec, R., & Menclova, E. (2011). The usage of open-source business intelligence in the Czech Republic. *Information Technology Interfaces (ITI)*.
- Peregrine, V. G. (2015). *Washington, DC: U.S Patente n° 9,183,529*.
- Pérez-Martín, C., Fernández-Aldecoa, J. C., Hernández-Armas, J., & Cánovas-Paradell, R. (2014). Selection and deployment of a business intelligence system (BI) at a hospital's Clinical Engineering Department. *In XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, 1100-1103.
- Predictive Analytics Today. (30 de 11 de 2015). <http://www.predictiveanalyticstoday.com/>.
Obtenido de <http://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/>
- Rai, P., & Vijaya Murari, T. (2014). Survey on Existing Text Mining Frameworks and A Proposed Idealistic Framework for Text Mining by Integrating IE and KDD. *International Journal of Computational Engineering Research*, 2250-3005.
- Ranjan, J. (2009). Business intelligence: Concepts, components, techniques and benefits. *Journal of Theoretical and Applied Information Technology*, 60-70.
- Rouhani, S., Ghazanfari, M., & Jafari, M. (2012). Evaluation model of business intelligence for enterprise systems using fuzzy TOPSIS. *Expert Systems with Applications*, 3764-3771.
- Sahadevan, S., Hofmann-Apitius, M., & Schellander, K. (2012). Text mining in livestock animal science: introducing the potential of text mining to animal sciences. *Journal of animal science*, 3666.
- Shi, H., Peng, C., & Xu, M. Z. (2012). Business intelligence in construction: A review. *In Advanced Materials Research*, 3049-3057.
- Zhao, Z., Navarrete, C., & Iriberry, A. (2012). Open source alternatives for business intelligence: Critical success factors for adoption. *Americas Conference on Information Systems 2012, AMCIS*, 18-32.

-
1. Mgs en Gestión de Sistemas de Información e Inteligencia de Negocios, EMBA. Analista de sistemas informáticos de la Universidad Nacional de Chimborazo (UNACH). cmorales@unach.edu.ec
 2. PhD en Telecomunicación. Docente investigador Grupo de Aprendizaje Ubicuo de la Universidad Nacional de Chimborazo (UNACH). cradicelli@unach.edu.ec
 3. Mgs en Administración de Empresas. Gerente de Proyectos de Noux CA. Docente de la Universidad de las Fuerzas Armadas-ESPE. maferja@yahoo.com
 4. Mgs en Tecnología de la Información y Multimedia Educativa. Docente de la Universidad Nacional de Chimborazo. ebodero@unach.edu.ec
-