

Factor de impacto sobre artículos: propuesta de indicadores que obtengan el aporte de los artículos individuales en la comunidad científica

Jesús Vázquez-Ruiz¹, Beatriz Alejandra Olivares-Zepahua²,
María Antonieta Abud-Figueroa³

División de Estudios de Posgrado e Investigación, Instituto Tecnológico de Orizaba
jvazquezruiz@acm.org; {bolivares; mabud}@ito-depi.edu.mx

Resumen. En la actualidad hay un gran número de indicadores que miden el impacto de las revistas científicas, como el Factor de Impacto, o indicadores para medir el impacto de los investigadores como el Índice H, pero dichos indicadores realizan una medida generalizada de la revista o del investigador, sin evaluar el aporte de un artículo de forma individual; para solventar tal problemática, se presenta la propuesta de arquitectura de una aplicación para medir el impacto de los artículos de los investigadores del Instituto Tecnológico de Orizaba; la arquitectura está definida en capas: la primera para la extracción de citas de artículos e información referente a estos en fuentes como *ScienceDirect*, *IEEE Xplore*, *ACM DL* y en redes sociales como *Facebook* y *Twitter*; la segunda capa realiza el cálculo de impacto de los artículos mediante indicadores propios y la tercera capa es una aplicación Web para la visualización de los resultados.

Palabras clave: Revistas científicas, factor de impacto, indicadores, índice H.

1. Introducción

Desde mediados del siglo pasado, se estableció una medida para evaluar el factor de impacto que tiene una publicación, es decir, qué tan relevante es una revista considerada como científica; anualmente este nivel se mide y se publica por el ISI (*Institute for Scientific Information*, Instituto para la Información Científica); sin embargo, esta medida no refleja completamente la relevancia que un artículo en particular tiene en la comunidad científica o, en el caso de programas de orientación profesional, en el área empresarial correspondiente. Existen varias herramientas, la mayoría con costo, que permiten conocer y analizar el factor de impacto de las publicaciones indizadas por el ISI. También existen buscadores especializados dentro de la Web para consultar las citas que existen sobre un artículo, sin embargo, todos estos casos presentan el mismo problema: un usuario hace las búsquedas de forma individual, lo que no resulta óptimo dado el alto nivel de cambio que tiene la información.

Por lo anterior, se propone el desarrollo de una aplicación Web que, partiendo de la productividad específica de los investigadores de la Maestría en Sistemas Computacionales del Instituto Tecnológico de Orizaba en Veracruz, México, periódicamente recolecte en la Web la información de citas en los artículos incluidos en dicha productividad, para esta recolección de citas se necesita la creación de *Crawlers*, la recolección aplica Minería de Contenido sobre sitios de fuentes fidedignas en busca de citas sobre los artículos, analizar la cita en cuestión y contabilizarla.

2. Trabajos relacionados

Para el planteamiento de indicadores de impacto a nivel de artículo se revisaron indicadores existentes a nivel de revista y de ahí se hizo el análisis para su adaptación o la creación de nuevos en su caso, algunos de los indicadores provienen de [1], en este artículo se discute que tanto el Factor de Impacto como el Factor de Prestigio de los artículos están limitados, ya que no miden el impacto de las revistas donde se citan, para ello se proponen indicadores para medir el valor de las citas en función del impacto y/o prestigio de las revistas donde se producen dichas citas; otros indicadores analizados son de [2], ya que es importante saber cómo identificar la calidad de una revista, el investigador más prestigiado y qué artículo fue el más importante en un determinado año entre otros aspectos.

Cabe señalar que existen herramientas que obtuvieron el impacto de revistas y crearon sus propios indicadores, como es el caso de [3], proyecto que se presenta como una alternativa para medir el impacto de las publicaciones o de los investigadores; el OAS (*Open Access Statistics*, Estadísticas de Acceso Abierto) y también en [4] obtuvieron el impacto de los investigadores por medio de los SECES (*search engine count estimates*, es decir el recuento de páginas o *hits*) al utilizar motores de búsqueda como son *Google*, *Yahoo* y *MSN*. También se comprobó que una red social puede medir el impacto de una revista científica como en [5], donde trabajaron con *Twitter* contabilizando los *tweets* publicados por los usuarios sobre artículos científicos, es decir publicaciones con referencias que apuntan a artículos de *JMIR (Journal of Medical Internet Research)* y con ello saber la cantidad de atención que prestan los usuarios a un tema. En [6] se comprobó que un sitio colaborativo es capaz de medir el impacto de las revistas científicas, en dicho trabajo se busca medir el impacto de los artículos de *PubMed* (motor de búsqueda de libre acceso a la base de datos *MEDLINE* de citas y resúmenes de artículos de investigación biomédica) que son citados en *WIKIPEDIA*, por medio de sus PMID (*PubMed Unique Identifier*, un número único asignado a cada cita de un artículo de revistas biomédicas y biológicas que recolecta *PubMed*). Las herramientas mencionadas son algunas que se realizaron para obtener el impacto de las revistas científicas pero también hay propuestas para realizar una clasificación de los artículos ya publicados, como en [7] que hace hincapié en el desconocimiento del tipo de artículo publicado en una revista, es decir, si un artículo fue solamente polémico al momento de publicarlo, si tiene una gran contribución, o si sólo cubrió una necesidad puntual. Con relación al tema de extracción de contenido, en [8] se habla de las novedades en la recolección de información en la Web al ocupar

Minería Web (minería de contenido, uso y estructura) y define la minería de contenido como “la extracción de información o texto en las páginas Web”. Para realizar dicha extracción existen herramientas como la descrita en [9], la cual elimina el ruido excesivo como es la publicidad, enlaces a otros sitios y contenido que no tiene relación con la información principal; otra herramienta se presenta en [10] y sirve para crear directorios de servicios/proveedores en la Web, o el caso de [11], la herramienta EVA (*Evolutionary Virtual Agent*, Agente Virtual Evolutivo) en su cuarta versión, trata de un agente que simula una conversación real, por medio de un cliente (*chat*) al que se le solicita información y éste le responde por texto o voz. Esta herramienta tiene 2 componentes, el primero es la interfaz con un rostro en 3D el cual, al responder, lo hace con gestos, el segundo es el motor de comportamiento (cerebro) el cual extrae pertinentemente información de la Web y la usa en el flujo de la conversación.

La Minería Web tiene, entre otras ventajas, identificar por ejemplo enlaces inválidos, en [12] se muestra la problemática derivada del gran número de URLs de artículos no válidos (eliminadas, caducadas, cambiadas de servidor) y presenta recomendaciones para disminuirla. Cabe mencionar que los artículos a los que se les calculará el impacto se encuentran en una ontología de productividad propia del programa, con información previamente verificada, para ello se investigó sobre ontologías en el contexto de artículos, en [13] se utilizan métodos para clasificar documentos automáticamente y obtener la relación entre ellos por medio de ontologías y metadatos *SÁHKE*; y finalmente en [14] se mencionan las características que conforman las ontologías y la Web Semántica.

3. Conceptos

A continuación se describen los conceptos más relevantes para el mejor entendimiento del presente documento.

- Indicadores de impacto

Los indicadores de impacto son métodos para medir la relevancia que tiene una publicación, una revista o un investigador entre otros. Estas métricas o indicadores dan una idea del éxito o fracaso del elemento a medir [2].

- Ontología

Una ontología define los términos a utilizar para describir y representar un área de conocimiento. Las ontologías las utilizan las personas, las bases de datos, y las aplicaciones que necesitan compartir un dominio de información (un dominio es simplemente un área de temática específica o un área de conocimiento, tales como la medicina, fabricación de herramientas, bienes inmuebles o productividad entre otros). Las ontologías incluyen definiciones de conceptos básicos del dominio y las relaciones entre ellos que son útiles para las computadoras, en este sentido, hacen el conocimiento reutilizable [15].

- *Crawler*

Para la extracción de información de la Web se utilizan programas llamados “arañas” (también conocidos como *Spiders*, *Crawlers* o *Bots*), dichos programas se encargan de llevar a cabo un tipo concreto de tareas, en particular es responsable de recorrer las páginas Web de Internet, descargarlas o copiarlas a la computadora local y procesarlas para la obtención de información [16].

- *Selenium*

Selenium es un marco de trabajo que permite descargar el código fuente de las páginas Web, también cuenta con funciones para extraer elementos del DOM (Modelo de Objetos del Documento) para realizar pruebas o procesamiento de información. Actualmente está disponible para *Java*, *Ruby*, *Python*, *Perl*, *.Net* y *PHP* [17].

- *Jena*

Jena es un marco de trabajo en Java para crear aplicaciones de Web Semántica, Jena proporciona un conjunto de herramientas y bibliotecas de *Java* para ayudar a desarrollar la Web semántica y enlazar datos de aplicaciones, herramientas y servidores [18].

- Protocolo OAuth

OAuth (*Open Authorization*, Autorización Abierta) es un protocolo abierto que permite la autorización segura y sencilla para acceder a datos o aplicaciones que requieren autenticación sin la necesidad de estar registrados en ellas, ya sea en aplicaciones Web, móviles o de escritorio [19].

- Scribe OAuth

Es un marco de trabajo para trabajar con el protocolo OAuth en el lenguaje *Java*, *Scribe* tiene soporte para muchos servicios populares que utilizan OAuth, tales como *Facebook*, *Twitter* y *Google* entre otros [20].

4. Propuesta

En la figura 1 se muestran las capas de la arquitectura propuesta y posteriormente se define cada una de ellas.

4.1 Extracción

La capa de extracción es responsable de identificar los artículos a los que se les calcularán los indicadores de impacto, acceder a fuentes de información para extraer

información de citas siguiendo la definición existente en una ontología de indicadores.

A continuación se muestran los elementos que conforman esta capa.

4.1.1 Fuentes de Información

Las citas de los artículos se buscan en fuentes de información confiables, como son las bibliotecas virtuales de organizaciones reconocidas; por otro lado, debido al crecimiento de las redes sociales, también se considera la revisión de dichas fuentes. A continuación se describen dichas fuentes:



Figura 1. Capas de la herramienta de impacto

ACM. La ACM es la principal organización de Ciencias Computacionales que brinda a sus miembros, y a toda la comunidad de estudiantes la oportunidad de tener a su alcance los últimos avances en tecnología. Fomenta la interacción entre estudiantes, universidades y profesionistas, que es la base para un buen desarrollo académico. La ACM tiene una biblioteca virtual (ACM Digital Library) que da acceso a artículos y libros por medio de una suscripción [21].

IEEE. La IEEE es el Instituto de Ingenieros Eléctricos y Electrónicos, una asociación técnico-profesional mundial dedicada a la generación de estándares en las áreas de Ingenierías Eléctrica y Electrónica entre otras cosas.

Su objetivo es promover la creatividad, el desarrollo y la integración, compartir y aplicar los avances en las tecnologías de la información, electrónica y ciencias en general para beneficio de la humanidad y de los mismos profesionales. Ésta cuenta con una biblioteca virtual para la distribución de libros electrónicos y artículos tanto de acceso restringido por suscripción como de acceso abierto. Dicha biblioteca virtual es IEEE Xplore [22].

ScienceDirect. ScienceDirect es una parte de *Elsevier* (*Elsevier* forma parte del *Reed Elsevier PLC Group*, una importante editorial mundial y proveedor de información), es una base de datos que ofrece artículos de más de 2.500 revistas y capítulos de más de 11.000, así como también acceso a bases de datos secundarias como *Pergamon* y *NorthHolland*. ScienceDirect facilita la colaboración, premia la innovación y acelera el propio proceso de investigación. La dirección de esta plataforma es [23].

Facebook. Esta es una red social que cuenta con 600 millones de miembros que se comunican en más de 70 idiomas [24], la visibilidad del contenido de un usuario depende de la configuración, si lo restringe o lo hace libre en la Web.

En esta red social se puede compartir enlaces (URLs) de todo tipo, ya sea audio, video o documentos; para el caso del Análisis de Sentimientos planteado para este proyecto se consideran las calificaciones aprobatorias “likes” que tiene el enlace compartido del artículo, cuántos enlaces compartidos y los comentarios tanto positivos como negativos.

Twitter. Esta es una red social que cuenta con 500 millones de miembros, dicha red permite enviar mensajes de texto plano de corta longitud, con un máximo de 140 caracteres, llamados *tweets*, que se muestran en la página principal del usuario [25].

Para la aplicación del Análisis de Sentimientos planteado se considera el número de *tweets* que tiene el artículo y sus comentarios tanto positivos como negativos.

4.1.2. Ontología de Productividad. La ontología de productividad contiene información relevante de los investigadores del Instituto Tecnológico de Orizaba, así como también todos los artículos que han elaborado cada uno de ellos, esta ontología fue desarrollada en *Protégé* [26]; cabe señalar que la herramienta de extracción partirá de esta ontología para buscar las citas de cada uno de los artículos ahí contenidos.

4.1.3. Ontología de indicadores. La ontología de indicadores, contiene toda la información acerca de los indicadores propuestos, sus nombres, fórmulas, información de las fuentes e información de las citas.

4.1.4. Crawlers o herramientas de extracción. Para elaborar la herramienta de extracción se propone trabajar con el marco de trabajo *Selenium*, ya que a pesar de hacer pruebas con *Crawlers* ya existentes como son: *JSpider Crawler*, *Jobo* y *Web Harvest* [27], las fuentes seleccionadas detectaban los procesos automáticos de estas herramientas, impidiendo la extracción de información que se necesita. Para cada una de las fuentes se cuenta con un elemento que busca el artículo y otro que extrae la información pertinente.

4.2 Controlador de cálculo

En esta segunda capa se muestra la forma en la que se calculará el impacto de los artículos, como se había mencionado existen indicadores de impacto como el “Factor de Impacto” para revistas científicas, o el “Índice de Hirsch” para investigadores,

estos indicadores se obtienen a través de las citas que tengan los artículos, pero las citas se contabilizan de forma general y surge la problemática de que el impacto no es equilibrado, por ejemplo si hubiera 2 revistas con el mismo número de citas en 2 años (el mismo impacto) ¿será lo mismo el impacto de la revista cuyas citas provienen de una revista de alto impacto que el impacto de otra cuyas citas provienen de una revista de bajo impacto?, para ello en [1] propusieron indicadores que pudieran solventar esa problemática y de cierta forma darle un valor más justo al impacto. Parecido al problema ya comentado aparece uno similar para el “Instituto Tecnológico de Orizaba”, en el caso donde un investigador tuviera 2 artículos, cada uno tiene 1 cita, la incógnita sería: ¿los 2 artículos tendrán el mismo aporte en la comunidad científica?, y la respuesta es que no, a pesar de tener el mismo número de citas se tiene que conocer el impacto de cada cita para contestar qué artículo tiene un mayor aporte o contribución, ya que no sería lo mismo que una de las citas tuviera a su vez 300 citas y la otra solo 2 citas, en este caso se reflejaría que el artículo de la cita que tiene 300 citas tiene mayor aporte que el otro, ya que está siendo parte de otra investigación la cual está provocando mayor interés y desarrollo en la comunidad científica.

También es importante comentar que las redes sociales son una gran oportunidad de dar a conocer información ya que es muy fácil que se pueda distribuir; con ello un investigador podría dar a conocer sus resultados y conocer el impacto que pueda tener gracias a la popularidad que pueda provocar. Por lo que se ha planteado sobre el impacto de artículos, se han adaptado indicadores que reflejen dicho impacto, los cuales a continuación se definen y se muestra cómo se obtienen.

4.2.1. Factor de impacto de Artículo (FIA)

Se refiere al impacto que tiene un artículo de forma general, es decir sin importar el impacto que tengan las citas de los artículos quienes lo citan y permitiendo la autocita. Se calcula dividiendo el número de citas totales entre una frecuencia de tiempo de 2 años. Su fórmula es representada de la siguiente forma:

$$\text{FIA} = \# \text{ citas totales} / \text{frecuencia (2 años)}$$

4.2.2. Indicador de impacto por artículo de donde proviene la cita (IIAPC).

Este indicador ayudará a identificar si el artículo quien cita al artículo principal tiene mayor impacto, con ello conocer si el artículo principal está contribuyendo a una investigación con mayor desarrollo. Se calcula sumando el impacto del artículo principal más el impacto del artículo quien cita entre 2 en un lapso de 2 años. Su fórmula es representada de la siguiente forma:

$$\text{IIAPC} = \text{FIA (artículo principal)} + \text{FIA (artículo que citó)} / 2$$

4.2.3. Indicador de impacto por artículo ponderado (IIAP).

Este indicador mostrará un impacto más justo de un artículo a diferencia del FIA, ya que es el promedio de la suma del impacto de cada artículo que cita al artículo principal. Este indicador se obtiene de la suma del impacto de cada artículo entre el

número total de citas en un periodo de 2 años. Su fórmula es representada de la siguiente forma:

$$\text{IIAP} = \text{FIA1} + \text{FIA2} + \text{FIAn} / \# \text{ Artículos}$$

4.2.4. Indicador de impacto por artículo en *Twitter* (IIAT).

Este indicador mostrará el impacto que tiene un artículo en la red social *Twitter*. Se calcula mediante el número de *tweets* sobre un periodo de tiempo de 6 meses. Su fórmula es representada de la siguiente forma:

$$\text{IIAT} = \# \text{ tweets} / \text{Frecuencia (6 meses)}$$

4.2.5. Indicador de impacto por artículo en *Facebook* (FIAF).

Este indicador mostrará el impacto que tienen los artículos en la red social *Facebook*. Se calcula mediante el número de *likes* que tiene un artículo en un periodo de tiempo de 6 meses. Su fórmula es representada de la siguiente forma:

$$\text{FIAF} = \# \text{ likes} / \text{Frecuencia (6 meses)}$$

Cabe señalar que en los indicadores para redes sociales se han elegido 6 meses por que el tiempo de vida de una publicación en ellas es más corto que en una fuente normal.

4.2.6. Indicador de impacto por artículo sin interacción mutua (IIASIM).

Este indicador mostrará el impacto que tiene un artículo pero omitiendo la autocita, ya que reflejará el impacto sobre otros artículos de autores diferentes, es decir, así el investigador verá cual es el interés que provoca su artículo en otros investigadores. Se calcula por el número de citas sin auto citas entre un lapso de tiempo de 2 años. Su fórmula es representada de la siguiente forma:

$$\text{IIASIM} = \# \text{citas totales} / \text{frecuencia (2 años)}$$

4.3 Presentación

Para la presentación de los datos extraídos se propone una aplicación Web en la que los investigadores consulten el impacto de sus artículos, como también qué artículo tiene el mayor impacto en un año específico, cuál es el artículo más citado, el artículo de mayor impacto en una red social, conocer el impacto que tienen los artículos quienes citan al artículo principal y conocer información pertinente sobre las citas como son: autores, nombre del artículo, fecha de publicación, dónde lo publicaron y lugar, entre otros.

5. Desarrollo de la investigación

En la figura 2 se muestra el proceso de una forma muy general para que la herramienta llegue a obtener el impacto de los artículos de forma automática desde la ontología de productividad hasta la aplicación Web para la visualización de resultados, así como también la tecnología que se usará para que cada proceso se lleve a cabo. Como primer caso se consulta, mediante el uso de Jena, el nombre de los artículos en la ontología de productividad del Instituto, para que posteriormente se envíen a la herramienta de extracción. Una vez que el extractor tiene la lista de artículos, buscará cada uno de ellos en las fuentes indicadas en la ontología de indicadores, al encontrarlo descargará tanto el número de citas con un periodo de 2 años, como también información relevante de cada una de las citas ya mencionado en la descripción de la capa de presentación.

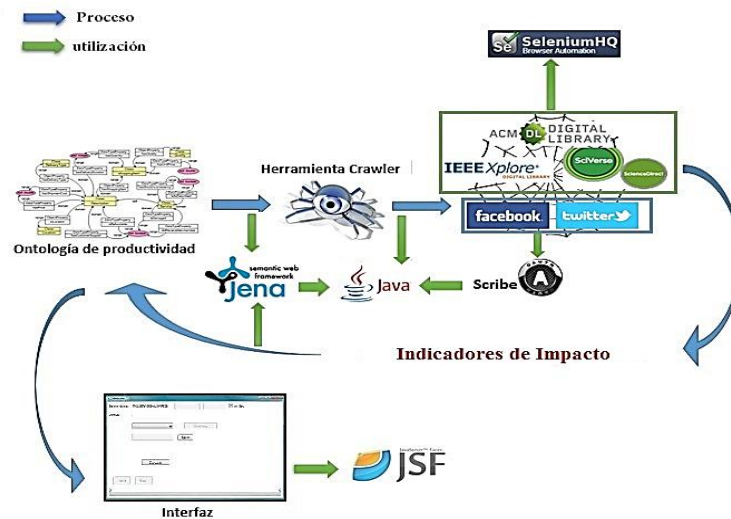


Figura 2. Ilustración del proceso de la herramienta para la obtención de impacto de los artículos de los investigadores del Instituto Tecnológico de Orizaba.

Como se mencionó anteriormente, para extraer los datos de las diferentes fuentes es necesario utilizar el marco de trabajo *Selenium*; para el caso de *Facebook* se extraerá el número de *likes* que tenga el artículo en el perfil del investigador o en la cuenta del programa de maestría en esta red; en el caso de *Twitter* también con las mismas cuentas, pero en dicha red, se extraerá el número de *tweets*. Una vez obtenido el número de citas e la información posteriormente se realiza el cálculo de cada uno de los indicadores mencionados en el apartado anterior; ya realizado el cálculo, con el auxilio de *Jena*, se guarda el resultado de los indicadores para cada artículo en la ontología de productividad y la información de las citas en la ontología de indicadores. Por

último, ya realizada la extracción y actualización de las ontologías, la aplicación Web visualizará el resultado de los indicadores para cada artículo como información a detalle de cada una de las citas que tengan.

6. Discusión

La limitación de las fuentes seleccionadas (ACM, IEEE, ScienceDirect) se debe a que proporcionan información detallada de las citas bajo acceso libre pese a ser bibliotecas virtuales de suscripción, también se les considera líderes en su tipo e incluyen a todas las revistas indizadas del área.

En el caso de redes sociales *Facebook* y *Twitter* se eligen por ser las que tienen el mayor impacto en la actualidad y se puede acceder a la información necesaria para el cálculo de indicadores mediante el protocolo OAUTH. La justificación de utilizar indicadores distintos a los usados actualmente en el medio es que, en programas de posgrado con orientación profesional, es menos frecuente la publicación en revistas indizadas, lo que da como resultado que se desconozca el impacto de muchos artículos, los cuales, es posible que sean de relevancia en el área académica aunque no se hayan publicado en una revista de alto impacto.

7. Resultados parciales

La primera versión del extractor de citas está formada por una serie de componentes que se muestran en la figura 3. El extractor ya obtiene la lista de artículos de la ontología de productividad y elige el Extractor específico a utilizar para extraer la información pertinente de cada fuente, pues dicha información no está representada de la misma manera en todas las páginas.

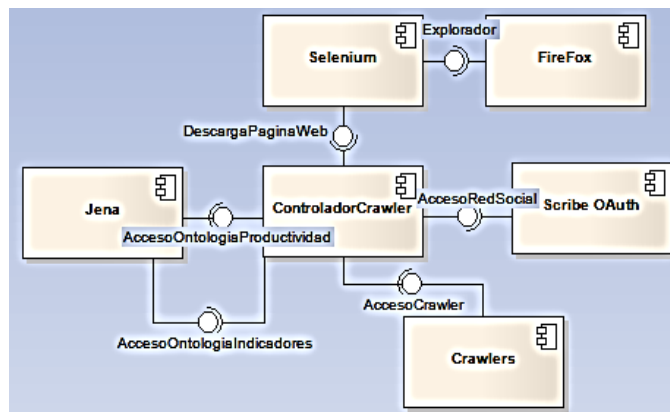


Figura 3. Diagrama de componentes del extractor de citas.

En la tabla 1 se muestra el resultado de una primera muestra de artículos de cuatro de los investigadores del Instituto Tecnológico de Orizaba: Giner Alor Hernández, Ulises Juárez Martínez, María Antonieta Abud Figueroa, Ana María Chávez Trejo, este resultado cuenta con el número de citas y el impacto de sólo 3 indicadores ya que falta trabajar con la extracción en redes sociales. También se muestra el resultado de un artículo externo “*Analysis of recommendation algorithms for e-commerce*” [28] para demostrar el resultado con un artículo que rebasa las 100 citas.

Tabla 1. Resultados parciales de los indicadores de impacto.

Investigadores	Artículo	Citas	FIA	IIAP	IIASIM
Giner Alor-Hernandez, Alberto Aguilar-Lasserre, Ulises Juárez-Martínez, Ruben Posada-Gomez, Guillermo Cortes-Robles, Mario Alberto Garcia-Martínez, Juan Miguel Gomez-Berbis, Alejandro Rodriguez-Gonzalez	HYDRA: A Middleware oriented Integrated Architecture for e-procurement in Supply Chains	1	0.0416	0.0416	0.0
Ulises Juárez-Martínez, Jose Oscar Olmedo-Aguirre	A Join Point Model for Fine Grained Aspects	1	0.0416	0.0	0.0
Giner Alor-Hernandez, Ruben Posada-Gomez, Ulises Juárez-Martínez, Gustavo Pelaez-Camarena, María Antonieta Abud-Figueroa	VEGITO: A Virtual Enterprise Generator	0	0.0	0.0	0.0
Giner Alor-Hernandez, Ulises Juárez-Martínez, Ruben Posada-Gomez, Ana María Chávez-Trejo, Jose Saul Rocha-Aragon	Defining an SOA for Stock Quote Management	0	0.0	0.0	0.0
Alor-Hernández G, Sánchez-Cervantes	ITOHealth: a multimodal	0	0.0	0.0	0.0

Investigadores	Artículo	Citas	FIA	IIAP	IIASIM
JL, Juárez-Martínez U , Posada-Gómez R, Cortes-Robles G, Aguilar-Lasserre A.	middleware-oriented integrated architecture for discovering medical entities				
Alejandro Rodríguez, Myriam Mencke, Giner Alor-Hernandez , Ruben Posada-Gomez, Juan Miguel Gomez, Alberto A. Aguilar-Lasserre	MEDBOLI: Medical Diagnosis Based on Ontologies and Logical Inference	2	0.0833	0.0208	0.0
Juan Miguel Gomez, Giner Alor-Hernandez , Ruben Posada-Gomez, Ma. Antonieta Abud-Figueroa , Angel Garcia-Crespo	SITIO: A Social Semantic Recommendation Platform	0	0.0	0.0	0.0
Giner Alor-Hernandez , Alberto Aguilar-Lasserre, Ulises Juárez-Martínez , Ruben Posada-Gomez, Guillermo Cortes-Robles, Mario Alberto Garcia, Martinez, Juan Miguel, Gomez, Myriam Mencke, Alejandro Rodriguez Gonzalez	A Hybrid Architecture for E Procurement	0	0.0	0.0	0.0
Ulises Juárez-Martínez , José Oscar Olmedo-Aguirre	Énfasis: a model for local variable crosscutting	1	0.0416	0.0416	0.0
Ulises Juárez-Martínez , Giner Alor-Hernandez	Advise Weaving in Énfasis	0	0.0	0.0	0.0

Investigadores	Artículo	Citas	FIA	IIAP	IIASIM
Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl	Analysis of recommenda- tion algo- rithms for e- commerce	147	6.125	1.547	0.0

Para mostrar la información extraída de las citas desde el extractor, en la figura 4 se muestra el resultado de la búsqueda individual del artículo “*HYDRA: a middleware-oriented integrated architecture for e-procurement in supply chains*” [29] del investigador Giner Alor Hernández, tiene 1 cita que corresponde al artículo “*Query expansion methods and performance evaluation for reusing linking open data of the European public procurement notices*” [30] que tiene como autor principal a José María Álvarez. También se muestra la bitácora del Controlador donde se observa que obtuvo el número de citas totales, el número de citas que están dentro de la diferencia de 2 años respecto al año actual, el número de auto citas y también los datos relevantes de cada cita como son autores, fecha de publicación, año, descripción, nombre del artículo, la URL de donde está localizado y cuántas citas tiene a su vez. Cabe señalar que la contabilidad de citas es realizada cuando el artículo es encontrado en alguna de las fuentes, si el artículo a buscar no es encontrado o no existe, simplemente el extractor no contabiliza las citas ni la información de ellas.

```

Artículo a buscar: HYDRA: a middleware-oriented integrated architecture for e-procurement in supply chains
Artículo encontrado en ACM link: http://dl.acm.org/citation.cfm?id=1880469.1880469&coll=DL&dl=GUI&CFID=2073655324CPTOKEN=42560873
Autor principal: Giner Alor-Hernández
numero de citas totales: 1
numero de citas con diferencia de 2 años al año actual: 1
numero de autocitas: 0
*****
Cita: Jose María Álvarez , José Emilio Labra , Ramón Calmeau , Ángel Marín , José Luis Marín, Query expansion methods and performance evaluation for reusing
ruta: http://dl.acm.org/citation.cfm?id=2075618&amp;CFID=207365548&amp;CPTOKEN=40929362
Autores:
Jose Maria Álvarez
José Emilio Labra
Ramón Calmeau
Ángel Marín
José Luis Marín
nombre de artículo: Query expansion methods and performance evaluation for reusing linking open data of the European public procurement notices
description: The aim of this paper is to present some methods to expand user queries and a performance evaluation to retrieve public procurement notices in t
fuente: http://dl.acm.org/
revista publicada: Proceedings of the 14th international conference on Advances in artificial intelligence: spanish association for artificial intelligence
fecha: November 07-11
año: 2011
lugar: La Laguna, Spain
Número de citas que tiene este artículo: 1
Nombre de los artículos de estas citas de este artículo:
Towards a theoretical foundation for the harmonization of linked data

Artículo No encontrado en ScienceDirect
Artículo No encontrado en IEEE

```

Figura 4. Resultado de extracción para el artículo “*HYDRA: a middleware-oriented integrated architecture for e-procurement in supply chains*”, para este resultado el artículo no se encontró en ScienceDirect ni IEEE

8. Conclusiones

La combinación de diferentes tecnologías y protocolos emergentes como la Minería de Contenido, las ontologías, las redes sociales y OAuth entre otros, con técnicas de Ingeniería de Software sofisticadas como la arquitectura en capas y la creación de componentes, permite el desarrollo de aplicaciones complejas bajo aproximaciones iterativas. Estas aplicaciones permiten obtener de manera más fácil información relevante para los usuarios con miras a procesarlas para la adquisición de conocimiento específico, en este caso el impacto del trabajo de los investigadores. Si bien es cierto que ya existen en el mercado herramientas que obtienen datos parecidos, la presente propuesta cuenta con las ventajas de realizar el trabajo de forma automática, es decir, no es necesario indicar cada uno de los artículos sino que los toma de las fuentes propias de la Institución, la búsqueda de citas se realiza en fuentes confiables previamente definidas y, además, considera indicadores especiales, también parte de la propuesta, que toman en cuenta la situación particular del programa de posgrado y que se refleja a nivel de artículo y no de publicación.

9. Trabajo a futuro

Las siguientes etapas incluyen la creación de la capa responsable del cálculo de los indicadores así como de la aplicación Web que permita la consulta específica de los datos obtenidos de las citas.

También se tiene considerado un Análisis de Sentimientos, para definir si los comentarios publicados ahí, con respecto a los artículos, se pueden considerar como favorables o desfavorables a estos.

Agradecimientos

Este trabajo cuenta con apoyo por parte del Consejo Nacional de Ciencia y Tecnología (CONACyT).

Aclaración

Las referencias a ACM, IEEE, ScienceDirect no implican de manera alguna una vinculación directa de su parte con el presente trabajo.

Referencias

1. G. Buela Casal. Evaluación de la calidad de los artículos y de las revistas científicas: Propuesta del factor de impacto ponderado y de un índice de calidad. *Psicothema*, pp. 23-35, 2003.
2. N.P.K Nigam A. Citation Index and Impact factor. *Indian J Dermatol Venereol Leprol*, p. 511, 2012. with Graphs (MLG '10), ACM, pp. 86-93, 2010.
3. U. Herb. OpenAccess Statistics: Alternative Impact Measures for Open Access Documents? An Examination How to Generate Interoperable Usage Information From Distributed Open Access Services. *IFLA*, pp. 165-178, 2012.
4. M. and Tumasjan Spörrle. Using Search Engine Count Estimates as Indicators of Academic Impact: A Web-based Replication of Haggbloom et al. *The Open Psychology Journal*, pp. 12-18, 2011.
5. G. Eysenbach. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. In *Journal of Medical Internet Research*, 2011.
6. M. Krauthammer Perry Evans. Exploring the use of social media to measure journal article impact. *Proceedings of the AMIA Annual Symposium*, pp. 374-381, 2011.
7. H. Piwowar. LSE - London School of Economics and Political Science. <http://blogs.lse.ac.uk/impactofsocialsciences/2012/04/04/31-flavours-research-impact/>, abril de 2012.
8. R. Baeza Yates. Tendencias en minería de datos de la Web. *El profesional de la información*, pp. 5-10, 2009.
9. P.M. Joshi and S. Liu. Web document text and images extraction using DOM analysis and natural language processing. *Proceedings of the 9th ACM symposium on Document engineering (DocEng '09)*, ACM, pp. 218-221, 2009.
10. H. Davulcu, A. Joglekar, A. More, S. Mukherjee, S. Patil, and I. V. Ramakrishnan. P. Choudhari. YellowPager: a tool for ontology-based mining of service directories from web sources. *ACM*, pp. 458-458, 2002.
11. P. Millet and J. Claude Heudin. Web Mining in the EVA Intelligent Agent Architecture. *IEEE Computer Society*, pp. 368-371, 2007.
12. F. Coetzee, E. Glover, G. Flake, D. Pennock, B. Krovetz, F. Nielsen, A. Kruger, and L. Giles Steve Lawrence. Persistence of information on the web: analyzing citations contained in research articles. *Proceedings of the ninth international conference on Information and knowledge management (CIKM '00)*, pp. 235-242, 2000.
13. T. Raiko, T. Tiinanen, and E. Hyv Katariina Nyberg. Document classification utilising ontologies and relations between documents. *Proceedings of the Eighth Workshop on Mining and Learning*.
14. R. Pedraza-Jiménez, L. Codina, and C. Rovira. Web semántica y ontologías en el procesamiento de la información documental. *El profesional de la información*, pp. 569-578., 2007.
15. Grupo de trabajo W3C. OWL Web Ontology Language Use Cases and Requirements. <http://www.w3.org/2007/09/OWL-Overview-es.html>.
16. Heaton, J. HTTP Programming Recipes for Java Bots. Heaton Research Inc. 2007.
17. Selenium. <http://docs.seleniumhq.org/>, abril de 2013.
18. Apache Jena (2013,Abril) Apache Jena [Online]. <http://jena.apache.org/>.
19. OAUTH (2013,Abril) OAUTH [Online]. <http://oauth.net/>.
20. Scribe Java. (2013,Abril) GitHub [Online]. <https://github.com/cdryden/scribe-java>.
21. ACM. (2013, Feb.) ACM. [Online]. <http://dl.acm.org/>.

22. IEEE Explore. <http://ieeexplore.ieee.org/Xplore/guesthome.jsp>. Febrero de 2013.
23. Science Direct. (2013, Feb.) Science Direct. [Online]. <http://www.sciencedirect.com/>.
24. Facebook. (2013, Feb.) Facebook. [Online]. <http://www.facebook.com/>.
25. Twitter. (2013, Feb) Twitter. [Online]. <https://twitter.com/>.
26. Source.net/open-source/crawlers.
27. Protege Project. (2013, Feb) Protege Project. [Online]. <http://protege.stanford.edu/>.
28. Java source. (2013, Feb) Java source. [Online]. <http://java-source.net/open-source/crawlers>.
29. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2000. Analysis of recommendation algorithms for e-commerce. In Proceedings of the 2nd ACM conference on Electronic commerce (EC '00). ACM, New York, NY, USA, 158-167.
30. G. Alor-Hernandez., A. Aguilar-Lasserre, U. Juarez-Martinez, R. Posada-Gomez, G. Cortes-Robles, M. A. Garcia-Martinez & A. Rodriguez-Gonzalez. HYDRA: a middleware-oriented integrated architecture for e-procurement in supply chains. Transactions on computational collective intelligence I (pp. 1-20). 2010.
31. J. María Álvarez , J. Emilio Labra , R Calmeau , Á.l Marín , J. L. Marín. Query expansion methods and performance evaluation for reusing linking open data of the European public procurement notices. Proceedings of the 14th international conference on Advances in artificial intelligence: spanish association for artificial intelligence, November 07-11, 2011, La Laguna, Spain.