

Universidad de las Ciencias Informáticas

Facultad 1



Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

Título: Estado de la producción científica en Web Semántica en el período 2005 – 2009 según la disponibilidad de *Computers & Applied Sciences Complete*.

Autor: Yaniel Rodríguez Báez

Tutores: Lic. Sergio Carbonell de la Fé
MSc. Delly Lien González Hernández

La Habana, julio de 2011

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo. Para que así conste firmamos la presente a los ____ días del mes de _____ del año 2011.

Firma del Autor

Firma del Tutor

Lic. Sergio Carbonell de la Fé

Firma del Tutor

MSc. Delly Lien González Hernández

AGRADECIMIENTOS

En primer lugar quiero agradecer especialmente a Dios que es tan grande y poderoso, que ha sabido guiarme en mi camino, a mis padres que a pesar de estar divorciados siempre los dos han estado cerca de mi apoyándome y dándome todo su amor y confianza, a mis abuelos y demás familiares por ayudarme siempre a salir adelante cada día. Agradezco a mi novia por darme todo el amor del mundo y estar cada día a mi lado, en las buenas y en las malas, siendo un pilar fundamental en mi vida. Reconocer el apoyo de todos mis amigos, que siempre estuvieron ahí, para brindarme el consejo oportuno en las diferentes situaciones del destino y a mis tutores que fueron un apoyo fundamental en el desarrollo de esta investigación, dando paso a otra etapa de mi vida.

DEDICATORIA

Dedico la presente investigación especialmente a mis padres y demás familiares que siempre confiaron en mí y dieron todo de sí, para que pudiera ser lo que hoy soy. También a todas las personas especiales que encontré en mi vida, que me ayudaron y brindaron todo su apoyo incondicionalmente.

RESUMEN

En el presente trabajo se realiza un estudio de la producción científica en Web Semántica en el período 2005-2009 a partir de la base de datos bibliográfica *Computers & Applied Sciences Complete* distribuida por EBSCO. Se realiza un estudio del estado del arte en Web Semántica según conceptos y tecnologías utilizadas en el desarrollo de soluciones y se consideran los antecedentes de estudios métricos aplicados a la Ciencia de la Computación. Se utilizan como herramientas: EndNote X4 para la gestión y normalización de los registros; Bibexcel v.2011 como sistema de generación de matrices y Ucinet 6 para la realización de metáforas visuales de redes sociales. Por medio del uso de métodos métricos de información e indicadores bibliométricos se determina la zona núcleo de revistas para la producción científica en Web Semántica, autores más productivos y relaciones de colaboración, líneas temáticas trabajadas y la representatividad y colaboración a nivel institucional y de país.

Palabras clave: *Computers & Applied Sciences Complete*, estudios métricos de la información, producción científica, Web Semántica.

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO I: ENFOQUE DEL DESARROLLO EVIDENCIADO EN EL CAMPO DE LA WEB SEMÁNTICA	7
1.1. WEB SEMÁNTICA, ORÍGENES Y DEFINICIÓN	7
1.2. LA WEB TRADICIONAL	9
1.3. INFRAESTRUCTURA Y TECNOLOGÍAS DE LA WEB SEMÁNTICA.....	10
1.3.2- TECNOLOGÍAS Y LENGUAJES RECIENTES EN DESARROLLO.....	15
1.3.3- HERRAMIENTAS EN DESARROLLO	18
1.4. INVESTIGACIONES REALIZADAS SOBRE WEB SEMÁNTICA.....	21
1.4.2- A NIVEL NACIONAL	22
1.4.3- INVESTIGACIONES EN LA UCI	23
1.5. PAPEL DE LA PRODUCCIÓN CIENTÍFICA.....	24
1.6. APORTES INVESTIGATIVOS.....	28
CAPÍTULO II: MATERIALES, MÉTODOS Y HERRAMIENTAS	29
2.1 MÉTODOS Y PROCEDIMIENTOS	29
2.2 INDICADORES BIBLIOMÉTRICOS	30
2.3 FUENTES.....	32
2.4 ESTRATEGIA DE BÚSQUEDA	35
2.5 HERRAMIENTAS	36
2.5.1 GESTIÓN Y MANEJO DE LOS REGISTROS BIBLIOGRÁFICOS	37
2.5.2 BIBEXCEL (v.2011)	37
2.5.3 UCINET 6.....	37
CAPÍTULO III: ANÁLISIS DE LOS RESULTADOS	38
3.1 PRODUCCIÓN DE ARTÍCULOS POR AÑO	38
3.2 SOLAPAMIENTO Y SUPERPOSICIÓN DE LA COMPUTERS & APPLIED SCIENCES COMPLETE CON SCOPUS Y EL SCIENCE CITATION INDEX EXPANDEX (SCIE)	38
3.3 PRODUCCIÓN Y DISTRIBUCIÓN POR PUBLICACIONES	39
3.4 PRODUCCIÓN Y DISTRIBUCIÓN POR AUTORES	41
3.5 PRODUCCIÓN Y DISTRIBUCIÓN POR INSTITUCIÓN.....	45
3.6 PRODUCCIÓN Y DISTRIBUCIÓN POR PAÍSES	49
3.7 PRODUCCIÓN TEMÁTICA A TRAVÉS DEL ANÁLISIS DE PALABRAS CLAVE	53
CONCLUSIONES	57
RECOMENDACIONES	59
BIBLIOGRAFÍA	60

INTRODUCCIÓN

El lanzamiento del primer satélite artificial, el *Sputnik*, en octubre de 1957 es señalado por muchos autores como el primer antecedente de Internet. El acontecimiento repercutió de diferentes maneras en los Estados Unidos, principal oponente de la entonces Unión Soviética. Estas reacciones fueron desde la asignación de subsidios a proyectos científicos de investigación asociados a la esfera militar hasta la creación de nuevas instituciones, como la *Advanced Research Project Agency* (ARPA) en 1957 (ALBARELLO, 2005).

Al año siguiente, en 1958, el Pentágono, crea una nueva agencia de coordinación de los contratos federales de investigación: *Defense Advances Research Projects Agency* (DARPA), para investigaciones en computadoras avanzadas. Con el fin de facilitar los intercambios entre los distintos equipos contratantes, crea diez años más tarde la red ARPANET, que conectaba a los centros de computación de la *System Development Corporation* en Santa Mónica, California, al Proyecto Genie en la Universidad de California Berkeley y al Proyecto Multics del Instituto Tecnológico de Massachusetts (WIKIPEDIA, 2009). Es considerado por los historiadores que el núcleo fundacional de Internet es la red conocida como ARPANET. La red seguiría extendiéndose y en 1973 realiza su primera conexión internacional con Inglaterra y Noruega (MARTÍNEZ, 2009).

Paralelamente al crecimiento de las redes se irían desarrollando tecnologías como el *Transmission Control Protocol/Internet Protocol* (TCP/IP), el *HyperText Markup Language* (HTML), el *Uniform Resource Locator* (URL), entre otras, y cuya razón de ser fueron optimizar la comunicación y localización de los contenidos que se compartían (MARTÍNEZ, 2009). El desarrollo de las técnicas de telecomunicación propició la creación de nuevas formas y servicios de comunicación, como el correo electrónico, en 1971 por Ray Tomlinson (CAVENDER, 1998), *gopher* –predecesor de cierto modo de la Web, consistente en el acceso a la información a través de menús-, intercambio de fichero por medio de FTP, que modificaron la forma en que las personas consumían y producían información.

En 1989, el *European Organization for Nuclear Research* (CERN) poseía el nodo de Internet más grande de Europa y uno de sus investigadores, llamado Timothy John Berners-Lee vio la oportunidad de unir Internet y el hipertexto (HTTP y HTML). Un año

más tarde, en 1990, crea las herramientas necesarias para el funcionamiento de la web: un servidor web –llamado NeXTcube-, el programa de navegación –más tarde conocido como navegador o *browser*- y las primeras páginas web que al mismo tiempo describían el proyecto (CASTELLS, 2003). La *World Wide Web* es un sistema de distribución de información basado en hipertexto o hipermedias enlazados y accesibles a través de Internet (SEGAL, 1995). Se basa principalmente en documentos escritos en HTML. El lenguaje sirve, sobre todo, para crear hipertexto en Internet, pero ofrece pocas posibilidades para categorizar los elementos que configuran el texto más allá de las típicas funciones estructurales (W3C, 2007).

Con el lanzamiento de la red de redes en 1991 la web dejaría de ser un espacio reservado para científicos e instituciones de investigación para convertirse en una herramienta de trabajo utilizada por cualquier tipo de personas y para cualquier fin. Al crecimiento de la red y de sus usos lo seguiría el desarrollo de diferentes tecnologías enfocadas a la gestión y producción de contenidos que si bien han permitido el desarrollo de una web más dinámica, potente y segura no han demostrado estar preparadas para el increíble crecimiento de los contenidos. Tal es el caso de la generación dinámica de páginas, el acoplamiento con bases de datos, la diversidad de formatos, la acumulación de datos a partir de una mayor interactividad de los usuarios (CASTELLS, 2003) y el tardío descubrimiento de la necesidad de representación y organización de los contenidos que han incidido en hacer de la web actual un caos en lo que respecta a la búsqueda y recuperación de la información (RAMÍREZ, 2004).

Desde la misma concepción de la web, Berners-Lee tuvo presente la necesidad de ordenar sus contenidos mediante la introducción de descripciones explícitas del significado, ese enfoque es identificado por él como Web Semántica en 2001 (BERNERS-LEE *et al.*, 2001). La idea sería retomada por el *World Wide Web Consortium* (W3C), organismo internacional de estandarización de tecnologías para internet, fundado en 1994, lanzando un proyecto a corto, medio y largo plazo cuyo objetivo es, mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos usando la Web Semántica o como también es llamada “web de los datos”, para que la información sea compartida y se pueda reutilizar a través de aplicaciones, empresas y fronteras comunitarias (CODINA y ROVIRA, 2006).

En la Universidad de las Ciencias Informáticas (UCI) de la Habana, Cuba, dentro del departamento de Universidad Digital del Centro de Informatización Universitaria (CENIA) de la Facultad 1, se desea crear una línea de desarrollo y producción de soluciones enfocada a la “Web Semántica” que permita el dominio de las nuevas tecnologías y herramientas de Internet al mismo nivel que se trabajan en el resto del mundo. El punto de partida sería el análisis de información científica actualizada que se produce sobre el tema, en el mundo y en Cuba.

La producción científica tiene su más fiel reflejo en los canales de comunicación de revistas científicas que son los encargados de diseminar los resultados obtenidos por las comunidades de investigadores. El acceso a estos recursos de información es difícil para las instituciones del tercer mundo que no cuentan generalmente, con el respaldo financiero necesario para costear su adquisición, otras veces son subutilizadas y no se les saca su máximo provecho.

La UCI posee una suscripción mediante pago a la base de datos EBSCO *Computers & Applied Sciences Complete*, donde existen publicaciones científicas de gran importancia que podrían brindar un reflejo del estado en la temática de Web Semántica y aportar elementos para el diseño de la nueva línea de trabajo del centro CENIA.

A partir de la necesidad social antes expresada, se plantea el siguiente problema científico de la investigación: ¿Cómo evaluar el estado del arte en Web Semántica en la base de datos EBSCO *Computers & Applied Sciences Complete* en el período 2005-2009?

El diseño de la investigación permite al autor declarar como objeto de estudio: la producción científica en Web Semántica y como campo de acción: el estado del arte sobre la producción científica en Web Semántica en el período 2005-2009.

En el contexto investigativo está enmarcado como objetivo general: Describir el estado de la producción científica en Web Semántica en el período 2005 – 2009 a partir de la información disponible en EBSCO *Computers & Applied Sciences Complete*.

La idea a defender será que el análisis de la producción científica en Web Semántica

permitirá un mejor diseño del proyecto de desarrollo de Web Semántica del departamento de Universidad Digital del centro CENIA de la Facultad 1.

Como objetivos específicos:

- Identificar la producción científica correspondiente a Web Semántica dentro de *Computers & Applied Sciences Complete*.
- Elaborar una base de datos para la normalización y sistematización de la información identificada.
- Describir la producción científica en Web Semántica utilizando indicadores cuantitativos clásicos.
- Identificar las principales relaciones entre autores, instituciones, países y temáticas.

Para guiar el proceso investigativo se determinan las siguientes tareas científicas:

- Búsqueda bibliográfica.
- Elaboración del estado del arte en la temática de Web Semántica.
- Selección y descripción de las fuentes bibliográficas seleccionadas.
- Elaboración de una base de datos con la producción científica en Web Semántica en el período 2005-2009.
- Normalización de la base datos.
- Discusión y análisis de los resultados.

Los métodos de investigación utilizados fueron:

Métodos teóricos:

- Analítico-sintético: se utilizó en la investigación de información relacionada con el desarrollo alcanzado en la temática de "Web Semántica", sus aspectos y relaciones, así como se analizaron documentos que permitieron obtener datos necesarios para entender la esencia de los elementos que se describen en este trabajo.

- Análisis Histórico Lógico: este método permitió estudiar de forma analítica la trayectoria histórica de las estrategias usadas anteriormente a diferentes niveles, para conocer y resolver problemáticas relacionadas con el creciente desarrollo de tecnologías y aplicaciones en la temática de “Web Semántica”.
- Análisis documental: se utilizó para el análisis y obtención de los conocimientos teóricos que sustentan la investigación y en el proceso de revisión bibliográfica para la elaboración del estado del arte.

El presente documento está estructurado, además de la presente Introducción, en los siguientes capítulos:

- Capítulo I: Enfoque del desarrollo evidenciado en el campo de la Web Semántica.
Se realiza un estado del arte de la temática “Web Semántica” a través de investigaciones internacionales y cubanas; se analizan los principales conceptos manejados por la comunidad científica; se describen los principales problemas de la web y mejoras que introduce el enfoque semántico y experiencias anteriores en la evaluación de la Ciencia de la Computación por medio de la producción científica.
- Capítulo II: Materiales, métodos y herramientas.
Se analizan las fuentes de información utilizadas tanto para la búsqueda y recuperación de los registros como para la validación de los datos. Se describe la estrategia de búsqueda y recuperación seleccionada y las herramientas empleadas.
- Capítulo III: Análisis de los resultados:
Se presentan y analizan los resultados de la investigación.
- Conclusiones
- Recomendaciones
- Bibliografía

Para la confección de la bibliografía se emplearon las normas ISO 690 e ISO 690-2.

Limitaciones del estudio:

- El estudio se limita cronológicamente a la producción científica generada en el período comprendido entre 2005 y 2009.
- Solo se considerarán como fuentes los títulos de publicaciones periódicas recopiladas por EBSCO *Computers & Applied Sciences Complete* (CASC).
- No se consideraron fuentes de acceso el *Directory of Open Journal Systems* pues no permitía la exportación normalizada de los registros bibliográficos, por la misma razón se desechó la utilización del servicio *CS Digital Library IEEE Computer Society*, a lo que se le añade que esta fuente solo da cobertura a las publicaciones editadas por la IEEE.
- No se pudo tener acceso a las bases de datos compiladas por el *Web of Science* y *Scopus*.

Resultados esperados:

- Identificación de las líneas de investigación más trabajadas dentro de la temática de Web Semántica en el período estudiado.
- Identificación de las instituciones con mayor producción científica.
- Identificación de los autores más productivos.
- Identificación de las relaciones de colaboración a niveles nacionales, institucionales y personales.
- Identificación de las publicaciones especializadas en la temática.

Capítulo I: Enfoque del desarrollo evidenciado en el campo de la Web Semántica

1.1. Web Semántica, orígenes y definición

A finales de los 90 se realizan los primeros abordajes a la temática de la Web Semántica (BERNERS-LEE *et al.*, 2001). Su precursor, Berners-Lee, intentó desde el principio incluir información semántica en su creación, la *World Wide Web*, pero por diferentes causas no fueron implementadas (CARVIN, 2005). Simultáneamente la W3C creó un equipo de desarrolladores dedicados a mejorar, extender y estandarizar el sistema. Muchos lenguajes, publicaciones y herramientas han sido ya desarrollados, pero a pesar de ello, las tecnologías de la Web Semántica aún están en estado muy temprano (CASTELLS, 2003).

El W3C es el organismo que regula aspectos esenciales de la Web encargándose de la elaboración de normas técnicas, la sistematización de tecnologías de desarrollo web, como pueden ser el HTML, el XML y las CSS y la recomendación de buenas prácticas para el desarrollo de aplicaciones en entorno web. Puede decirse que es el organismo de normalización más importante para Internet, sus recomendaciones y propuestas de normalización, a pesar de no ser de obligatorio cumplimiento, son acatadas y tomadas como referentes dentro de la misma comunidad de desarrolladores web y organismos de certificación. Su director fundador es Berners-Lee, por lo que no es de extrañar que muchas de sus ideas y concepciones se vean reflejadas en el trabajo de la organización. La Web Semántica es uno de los proyectos que lleva a cabo en estos momentos la institución (CODINA y ROVIRA, 2006).

El proyecto aspira a mejorar la interacción entre los sistemas informáticos y los usuarios, donde según Cueva (2008) los primeros se dotarían de una mayor inteligencia y autonomía, y a los segundos de un nuevo infouniverso donde será posible la conversión de la información en conocimiento. En la definición oficial dada por el W3C se afirma que:

“La Web Semántica proporciona un marco común que permite que los datos sean compartidos y reutilizados a través de aplicaciones, empresas y fronteras comunitarias. (...). Está basada en *Resource Description Framework* (RDF) e integra una variedad de aplicaciones utilizando *eXtended Markup Language* (XML) para la sintaxis y *Uniform Resource Identifier* (URI) para las denominaciones” (W3C, 2009).

La definición anterior se considera un poco críptica para el entendimiento y comprensión de lo que es la Web Semántica, ya que en realidad no se ha podido aún materializar del todo este proyecto.

Berners-Lee, caracteriza a la Web Semántica de la siguiente forma:

“Es colocar los datos en la Web de un modo en que las máquinas puedan entenderlos naturalmente o convertirlos a esa forma. Esto crea lo que yo llamo una Web Semántica: una red de datos que pueden ser procesados directamente o indirectamente por máquinas, dado que es una extensión de la actual Web en la cual la información se da mediante un significado bien definido, lo que facilita que los ordenadores y los usuarios trabajen en cooperación” (BERNERS-LEE *et al.*, 2001).

También Codina y Rovira (2006) hacen alusión a dos grandes visiones que definen la Web Semántica, separadas por iniciativas estructurales y funcionales en cuanto a inteligencia artificial y procesamiento robusto:

“Definición 1. La visión de la inteligencia artificial: La Web Semántica es un conjunto de iniciativas destinadas a promover una futura web cuyas páginas estén organizadas, estructuradas y codificadas de tal manera que los ordenadores sean capaces de efectuar inferencias y razonar a partir de sus contenidos” (CODINA y ROVIRA, 2006).

“Definición 2. La visión del procesamiento robusto: La Web Semántica es un conjunto de iniciativas destinadas a convertir la *World Wide Web* en una gran base de datos capaz de soportar un procesamiento sistemático y consistente de la información” (CODINA y ROVIRA, 2006).

Se han planteado conceptos diferentes pero todos coinciden al definir que bajo el término de Web Semántica se incluyen un conjunto de acciones que no tienen otro fin que el de dar significado a los contenidos disponibles en la web de forma tal que estos significados puedan ser procesados por las computadoras. A partir de los conceptos anteriores, este autor asume que la Web Semántica es una extensión de la Web tradicional, centrada en los significados de las palabras y expresiones que se encuentran en la Web, logrando un resultado más eficiente para los usuarios que navegan en internet, desde el punto de vista técnico es soportada por una serie de tecnologías y lenguajes (URI, XML, OWL, RULE, SPARQL y RDF), asociados entre ellos en un sistema de capas, donde todas deben realizar su función. Trabaja con mucha más cantidad de datos que la Web tradicional, para realizar un filtrado automático preciso de la información y el contenido que reside en la Web será más entendible por las propias máquinas, más allá de su simple estructura sintáctica.

Teniendo la definición anteriormente planteada, se hace necesario, para entender y comprender mejor lo que significa este gran proceso-proyecto, considerar cómo es la Web tradicional, a la que se puede denominar, por oposición a la Web Semántica, la Web Sintáctica.

1.2. La Web Tradicional

La mayoría de la Web está basada principalmente en documentos escritos en HTML (CUEVA, 2008), que es válido para adecuar el aspecto visual de un documento e incluir objetos multimedia en el texto. Está regida por el protocolo HTTP, que es el que rige las normativas y regula todo lo referido a las transferencias y accesibilidad de los hipertextos y por el sistema de localización de objetos en la web URL, dándole este, dirección o localización a los documentos escritos en el lenguaje antes mencionado.

Aunque es sumamente difícil medir el tamaño de la Web, se estimaba en 2001 en unos diez elevado a la nueve usuarios utilizan la Web, y que ésta contiene del orden de cuatro por diez elevado a la nueve documentos, un volumen de información equivalente a entre catorce y veinte y ocho millones de libros. (BERGMAN, 2001).

Años después un estudio realizado en 2005 usando distintos motores de búsqueda como: Google, MSN, Yahoo! y Ask Jeeves, estimó que existían 11.500 millones de páginas web (SIGNORINI, 2005). Otro estudio hecho en 2008 estimó que la cantidad había ascendido a 63.000 millones de páginas web (KUNDER, 2008), este crecimiento desmedido de la Web genera grandes volúmenes de información electrónica, creando algunas deficiencias por la aglomeración y desorganización del ciberespacio.

Debido a la gran capacidad y a la acumulación de contenidos en la Web, aparecen dificultades y una doble problemática basada en su codificación, porque no solamente carece de cualquier interpretación semántica y de un nulo valor semántico, sino que, además, sus etiquetas son susceptibles de uso inadecuado y se pueden utilizar de forma contraria a su función (ANTONIOU y VAN HARMELEN, 2008).

Otra deficiencia está en la utilización de los motores de búsqueda cuando se envía una pregunta, lo que hacen es comparar las palabras de la pregunta con las que están en su índice, creados estos con las palabras que aparecen en las páginas web, es decir, que si encuentra un documento que contenga las palabras de la pregunta, lo devolverá como respuesta. Esto es casi todo lo que pueden hacer los ordenadores en relación con procesamiento de información en páginas web (MELO y BOTIA, 2008).

Todos estos ejemplos son el síntoma de una causa común: la falta de capacidad de las representaciones en que se basa la web tradicional para expresar significados, la aglomeración de información no clasificada, por lo que ya está en desarrollo una nueva infraestructura y diversas tecnologías asociadas para solucionar todas las deficiencias de la que por su carácter opuesto a la Web Semántica es llamada Web Sintáctica.

1.3. Infraestructura y tecnologías de la Web Semántica

La Web Semántica propone superar las limitaciones de la web actual mediante la introducción de descripciones explícitas del significado, la estructura interna y la estructura global de los contenidos y servicios disponibles en la Web. Para poder lograr todos los objetivos que se han planteado en este prometedor proyecto deben contar con diferentes medios que serán descritos a grandes rasgos:

- 1- Mediante una codificación de páginas en la cual las etiquetas tengan, precisamente, carga semántica. Este apartado corresponde al estándar denominado XML.
- 2- Aportando metadatos en las páginas y sitios web, manteniendo el formato compatible de la estructura general de la Web y con diversas categorías de páginas e interoperabilidad entre los distintos sistemas informáticos, de esto se ocupa la norma RDF.
- 3- Mediante ontologías que permitan especificar conceptos, mediante el uso de un lenguaje fuertemente basado en lógica simbólica y que debe ser interpretado por un ordenador. De este aspecto se ocupan los lenguajes de ontologías web (OWL por sus siglas en inglés).

La Web de los Datos no está basada solamente en estos tres medios, sino que está formada por auténticas y diversas normas, protocolos y lenguajes, encargados cada uno de funciones específicas para que el proyecto cumpla con todas las metas trazadas.

Berners-Lee, basándose en su gran capacidad, creó una especie de diagrama que pretende abarcar toda la estructura de la “Web Semántica” (Figura1), la que describe en 7 niveles o capas (ANTONIOU y VAN HARMELEN, 2008).

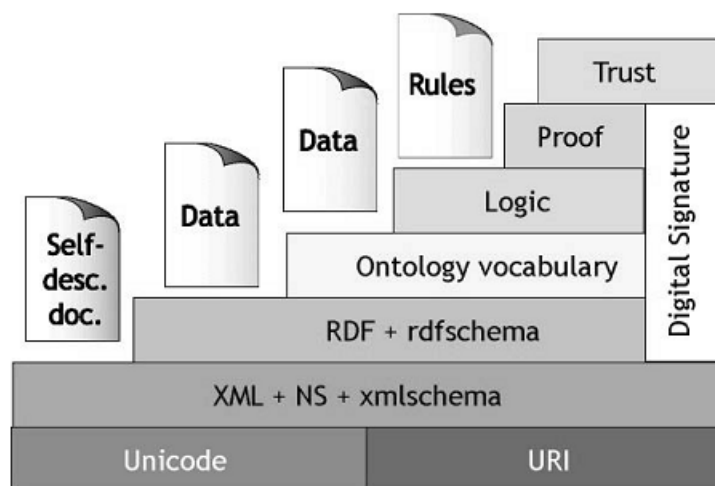


Figura 1: La Web Semántica vista como un sistema de capas (ANTONIOU y VAN HARMELEN, 2008).

La distribución de capas o niveles (Figura 1) en la actualidad ya no cumple con los requisitos para ser la estructura básica de la “Web Semántica”, debido a las diferentes transformaciones, nuevas tecnologías y nuevos estándares desarrollados, que han brindado numerosos avances y nuevos alcances en la temática.

Se ha construido una alternativa del diagrama anterior, que está siendo debatida en las diferentes instituciones especializadas en la materia (Figura 2). La siguiente estructura por capas se ajusta más a los recientes desarrollos y cambios tecnológicos asociados (ANTONIOU y VAN HARMELEN, 2008).

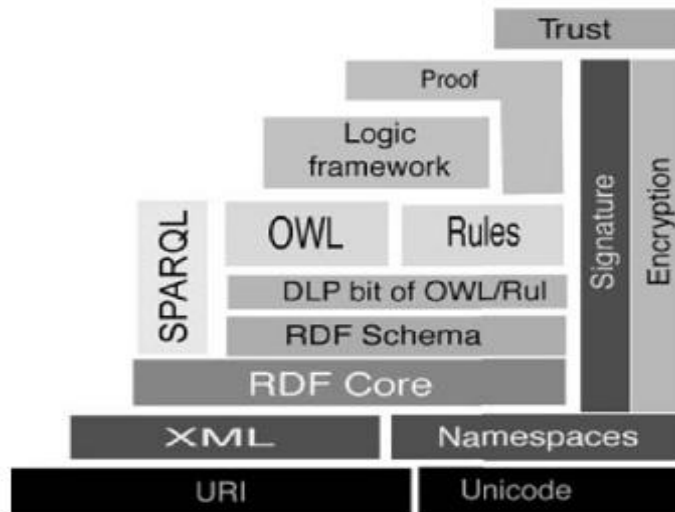


Figura 2: Alternativa del sistema de capas de la Web Semántica (ANTONIOU y VAN HARMELEN, 2008).

La capa de la ontología está identificada con dos alternativas: el OWL normal actual y un lenguaje basado en reglas. El DLP es como la intersección de la nueva capa OWL/Rul y el RDF *Schema*.

En la Tabla 1 se expone un significado más detallado de las capas que contiene el nuevo diagrama (figura 2).

Tabla 1: Significado de las capas estructurales de la Web Semántica.

<p>1. Unicode + URI</p>	<p><i>Unicode</i> es un sistema internacional, estándar que proporciona un número único para cada carácter, sin importar la plataforma ni el programa. Esto permite representar caracteres de cualquier idioma con una codificación unificada. <i>Uniform Resource Identifier (URI)</i> es un sistema de direccionamiento e identificación de recursos. El sistema que usamos actualmente para acceder a los recursos de la Web (URL) es una parte de URI.</p>
<p>2. XML + NameSpaces</p>	<p>XML es un sistema que permite definir lenguajes de marcas para usos específicos. <i>NameSpaces (NS)</i> permite combinar diversos lenguajes de marcado creados con XML en un mismo documento.</p>
<p>3. RDF</p>	<p><i>Resource Description Framework (RDF)</i> es un modelo de representación de metadatos que, entre otras cosas, permite representar recursos digitales tales como sitios o páginas web. RDF está concebido para representar cualquier clase de recursos (no solamente páginas publicadas en la Web).</p>
<p>4. RDF Schema + OWL+Rules + SPARQL</p>	<p><i>RDF Schema</i>, es una extensión de RDF que aporta un lenguaje con mayor capacidad para representar relaciones semánticas complejas. <i>OWL</i>: añade más vocabulario para describir propiedades y clases: tales como relaciones entre clases, cardinalidad, igualdad, tipologías de propiedades más complejas, caracterización de propiedades o clases enumeradas. <i>Rules</i>, Son un conjunto de reglas específicas de algunas relaciones y términos que van más allá de las capacidades expresivas del OWL. <i>SPARQL</i>, (<i>SPARQL Protocol and RDF Query Language</i>) es un lenguaje que debe entender los RDF, es decir, no sólo debe entender la sintaxis, sino que también debe</p>

	entender, los modelos de datos de RDF y la semántica del vocabulario. Es un lenguaje de recuperación para RDF/RDFS. Esta tecnología de consulta SPARQL permite que las personas puedan centrarse en la información que quieren, sin tener en cuenta la tecnología de la base de datos o el formato utilizado para almacenarlos.
5. Logic framework	<i>Logic</i> , se refiere al estudio de las reglas formales que permiten determinar si un razonamiento se sigue de sus premisas. La lógica estudia, por tanto, la estructura de los razonamientos válidos. Se espera que los ordenadores del futuro puedan efectuar razonamientos sobre los recursos y servicios de la Web combinando los conocimientos expresados en las ontologías, los hechos declarados en los metadatos y la aplicación de reglas lógicas.
6. Proof	<i>Proof (prueba)</i> , en este contexto significa demostración [matemática]. Se considera que un ordenador alcanza la máxima fiabilidad en sus razonamientos cuando es capaz de realizar demostraciones o, lo que es lo mismo a efectos prácticos, cuando es capaz de justificar el motivo por el cual tomó (o aconsejó tomar) una decisión.
7. Trust (+ Digital Signature + Encryption)	<i>Trust (confianza)</i> , la última capa, debe servir para otorgar confianza a las transacciones en la Web, que se llevarán a cabo no solamente entre usuarios y sitios web sino también entre programas de software y todo ello tanto en el plano C2B (<i>consumer to business</i>) como en el B2B (<i>business to business</i>). <i>Digital Signature (firma digital)</i> y <i>Encryption (Encriptación)</i> , proporcionarán soporte específico a esta capa, tal como muestra el diagrama.

Hasta ahora, solamente las cuatro primeras capas disponen de algún nivel de desarrollo, y ni siquiera lo están en su totalidad, aunque se han desarrollado algunas tecnologías nuevas y se han logrado numerosos avances en el trabajo con las ontologías y se han puesto en funcionamiento en algunos sistemas web y en redes pequeñas (CODINA y ROVIRA, 2006).

1.3.2- Tecnologías y lenguajes recientes en desarrollo

Dentro de la arquitectura de la Web Semántica un pilar fundamental son las tecnologías ontológicas y los lenguajes ontológicos, que ayudan a representar el conocimiento y fueron desarrolladas en el área de inteligencia artificial para facilitar el intercambio y reúso del conocimiento.

El grupo de desarrollo de ontologías del W3C ha dividido los OWL, para una mejor estructuración y organización, debido a algunos requisitos en tres sub-lenguajes, OWL-Lite, OWL-DL y OWL-Full, cada uno incremental, respecto al anterior en capacidad expresiva (ANTONIOU y VAN HARMELEN, 2008). Para mayor comprensión se desglosan a continuación:

- *OWL-Lite*: Permite establecer clasificaciones jerárquicas de conceptos de forma similar a RDFS además de proveer de mecanismos restrictivos básicos con respecto a la naturaleza de relaciones como el hecho de asignar cardinalidades por ejemplo enunciar el hecho de que una persona sólo puede tener un padre o una madre biológica (ANTONIOU y VAN HARMELEN, 2008).
- *OWL DL*: Proporciona una mayor riqueza expresiva con respecto a OWL-Lite garantizando la computabilidad de cualquier conclusión en un tiempo finito. Debe su nombre a la Lógica Descriptiva (*Description Logics*), un campo de investigación reciente en la actualidad (ANTONIOU y VAN HARMELEN, 2008).
- *OWL Full*: Pensado para usuarios que requieran la máxima expresividad dentro de la libertad sintáctica de RDF prescindiendo de cualquier garantía computacional en lo que respecta a la extracción de conclusiones (ANTONIOU y VAN HARMELEN, 2008).

Según Valentina Janev y Sanja Vranes (2009), las investigaciones recientes se han enfocado en regularizar las tecnologías de la Web Semántica, para los servicios y las herramientas reforzando la interoperabilidad. El último esfuerzo involucra el desarrollo de las reglas del lenguaje, lenguajes de intercambio de reglas, artefactos que refuercen el razonamiento y la mejora de los lenguajes ontológicos (Tabla 2).

Tabla 2: Tecnologías emergentes de la Web Semántica (AYERS, 2009).

Tecnologías y Lenguajes	Definición
OWL-S	Es un servicio ontológico que permite a los software agentes descubrir, invocar, componer y monitorear recursos web.
OWL 2	Extiende el lenguaje ontológico de la Web con un conjunto pequeño pero útil de características (EL, QL, RL) que habiliten el razonamiento efectivo.
WSMO	El servicio modelador de ontologías Web (WSMO por sus siglas en inglés) proporciona un marco conceptual y un lenguaje formal para ir describiendo semánticamente todos los aspectos relevantes de los servicios Web, para facilitar la automatización de hallar, combinar, e invocar servicios electrónicos sobre la Web.
WSML	Los <i>Web Services Markup Language</i> (WSML) proporcionan una sintaxis y una semántica formal para el WSMO y constan de varias variantes, tales como: WSML-Core, WSML-DL, WSML-Flight, WSML-Rule, and WSML-Full.
SWRL	Las reglas del lenguaje de la Web Semántica (SWRL por sus siglas en inglés) apuntan a ser en la Web Semántica los estándares de las reglas del lenguaje y están basadas en una combinación del OWL DL, OWL lite y RuleML.
SKOS	El Sistema de Organización de Conocimiento Simple (SKOS por sus siglas en inglés) es, en el sentido estricto de la palabra, un vocabulario construido en RDFS y OWL, con las especificaciones todavía en desarrollo. Sin embargo, comparte similares áreas de aplicación como los otros lenguajes. El SKOS deja expresar los sistemas de clasificación, tales como

	<p>taxonomías en el modelo de RDF cuando RDFS y las contracciones lógicas de OWL puedan ser demasiado fuertes. Ofrece un camino de migración directo de sistemas de organización de conocimiento en vigor para tecnologías de la Web Semántica.</p>
GRDDL	<p>Los Descriptores de Recursos de Dialectos del Lenguajes (GRDDL por sus siglas en inglés) en especificaciones, proporcionan un juego de procesos que interpretarán los documentos automáticamente como RDF.</p>
N3	<p>La Notación 3 (N3 por sus siglas en inglés) es básicamente diseñar un sencillo y ligero formato RDF, sobre el que se puedan construir fácilmente aplicaciones, las que también puedan ser sencillamente transformadas en caso de ser necesario.</p>
POWDER	<p>El Protocolo para los Recursos de Descripción de la Web (POWDER por sus siglas en inglés) según sus características técnicas de muchas maneras devuelve a las raíces de la Plataforma de Selección de Contenidos en Internet (PICS por sus siglas en inglés), El POWDER permiten, por ejemplo, etiquetar los recursos a través de expresiones regulares igualando las cadenas de caracteres de sitios direccionados con URIS.</p>
RuleML	<p>Constituye una familia del módulo de sublenguajes de la Web incluyendo la derivación de reglas, preguntas y limitaciones de integridad, así como la producción y reacción de reglamentos.</p>
RDFa	<p>El RDFa es una versión más simple del lenguaje básico de la Web Semántica RDF. La finalidad principal del RDFa es añadir metadatos a páginas Web HTML o XHTML ya existentes, por lo que resulta más fácil de incorporar que el RDF.</p>
RIF	<p>Las Reglas de Intercambio de Formatos (RIF por sus siglas en inglés) apuntan por ser la norma que gobierne el lenguaje y las reglas de intercambio de la Web Semántica.</p>

1.3.3- Herramientas en desarrollo

Dentro de la línea de la Web Semántica, actualmente se han desarrollado variedad de herramientas y plataformas que proporcionan servicios para crear y procesar elementos semánticos. Existen diferentes herramientas en el desarrollo de ontologías que se clasifican, según PÉREZ (2002) en:

- *Desarrollo de Ontologías:* Este grupo contiene herramientas y paquetes integrados que pueden ser utilizados para construir una nueva ontología a partir de cero. Aparte de las funcionalidades de edición y navegación, estas herramientas incluyen documentación importación y exportación de ontologías, a, o desde, diferentes formatos, librerías y motores de inferencia.
- *Evaluación de Ontologías:* Son utilizadas para evaluar aspectos de consistencia, corrección redundancia en el contenido de las ontologías; es decir para asegurar la calidad de las mismas y evitar problemas posteriores.
- *Combinación e Integración de Ontologías:* Son utilizadas para resolver el problema de fusión y alineamiento o combinación e integración de diferentes ontologías que pertenecen al mismo dominio.
- *Herramienta de Anotación basadas en Ontologías:* Herramientas en las cuales el usuario puede insertar instancias de conceptos y de relaciones en ontologías y mantener semi-automáticamente la anotación o marcas en páginas Web basadas en ontologías. La mayor parte de estas herramientas están siendo integradas en un entorno de desarrollo de ontologías.
- *Almacenamiento y Consulta de Ontologías (motores de inferencia):* Estas permiten la consulta de ontologías de manera sencilla y llevan a cabo inferencias con ellas. Para ello surgieron lenguajes para consulta de ontologías, como: RDQL, SPARQL, SeRQL.
- *Herramientas de aprendizaje de ontologías:* Estas herramientas son usadas para derivar ontologías a partir de textos en lenguaje natural de forma semi-automática; utilizando para ello técnicas en el área de aprendizaje automático y de procesamiento de lenguaje natural. La parte más complicada en la Web Semántica es el diseño de ontologías que equipare

todos los requerimientos. Elegir una correcta ontología es un elemento crítico de implementación satisfactoria de proyectos de Web Semántica.

En la literatura consultada ha sido posible identificar algunos proyectos que implementan soluciones para la Web Semántica, incluso ya existen en mercado algunas aplicaciones versión beta, Melo y Botia (2008) mencionan como los de mayor connotación a:

- *SWoogle2006*: proyecto para la recuperación de documentos de la Web Semántica. Además es para archivos escritos en RDF y OWL y utiliza una interfaz muy similar a la de de Google.
- *Spock*: primer motor de búsqueda para encontrar información de las personas en la Web como los sitios web, fotos, videos y blogs.
- *ZOMM*: motor de búsqueda semántica para la indexación y búsquedas por medio de lenguaje natural.
- *AKT (Advanced Knowledge Technologies)*: consorcio especializado en la generación de soluciones tecnológicas de Web Semántica para la gestión del conocimiento a partir de los activos informacionales de las organizaciones
- *DBin*: aplicación de la Web Semántica que permite a usuarios crear grupos de discusión para expresar y recuperar conocimiento compartiendo archivos e información estructurada semánticamente usando RDF.
- *DIP (Data, Information, and Process Integration with Semantic Web Services)*: desarrollo de soluciones en Web Semántica a partir de tecnologías de Web Services para la interoperabilidad automática de diferentes servicios web.
- *KW (Knowledge Web)*: enfocada al fortalecimiento de proveedores de servicios industriales, el teletrabajo y el comercio electrónico en el contexto europeo a partir de tecnologías de la Web Semántica.
- *NeOn*: proyecto cuya finalidad es permitir la integración y gestión de diferentes ontologías
- *Powerset*: proyecto enfocado al procesamiento del lenguaje natural por ordenadores para servicios de búsqueda y recuperación de información.

Igualmente, Jiménez (2010) menciona la existencia de las siguientes plataformas y

herramientas:

- *SÉSAME*: API para Java que presenta un marco de desarrollo para almacenamiento, consulta y razonamiento con RDF y RDF Schema. Puede ser usado como base de datos RDF y RDF Schema o como una librería de Java para aplicaciones que necesitan trabajar internamente con RDF. Es un entorno para desarrollo de aplicaciones en el lenguaje de programación Java para Web Semántica.
- *JENA*: framework desarrollado por HP Lab para manipular metadatos desde una aplicación java, existen dos versiones:
 - *Jena 1*: con soporte para RDF y capacidades limitadas de razonamiento, permite crear modelos persistentes que son mantenidos de forma transparente al usuario en una base de datos relacional.
 - *Jena 2*: Incluye una API para el manejo de ontologías y soporta el lenguaje *OWL*, *MySQL*, *Oracle* y *PostgreSQL*.
- *TWINE*: primera red social de la Web Semántica, permite organizar, compartir y descubrir información de interés para los usuarios, utiliza RDF y OWL, lo que le permite encontrar información relevante y significativa para el usuario final.
- *CORESE*: herramienta de desarrollo Web, consiste en un motor RDF basado en Grafos Conceptuales, su principal funcionalidad es recuperar recursos Web anotados en RDF Schema mediante el uso de un lenguaje de consulta basado en SPARQL y un motor de reglas de inferencias, permite extraer la información externa de documentos XML.
- *KOWARI*: herramienta de Web Semántica para almacenamiento y recuperación de ontologías, consiste en una base de datos escalable y transaccional construida con el objetivo de almacenar, recuperar y analizar metadatos, está escrita en Java y soporta RDF y OWL y se desarrolló en código abierto.

1.4. Investigaciones realizadas sobre Web Semántica

Existen algunas investigaciones relacionadas con la “Web Semántica” de valor científico-técnico, realizadas en diferentes instituciones del mundo, por ejemplo:

Los autores Berners-Lee, Hendler y Lassila (2001) brindan explicaciones de lo que significa la Web Semántica, sus principales funciones, los gigantescos alcances del proyecto, su representación y hacen alusión a las ontologías y agentes inteligentes como pilares fundamentales de la Web Semántica.

También Ding, Fensel y Klein (2002) hacen mención al interés que le han puesto las empresas informáticas al desarrollo de ontologías y exponen los diferentes componentes de la Web Semántica, brindando una explicación del funcionamiento y papel que desempeñan en el proyecto. Muestran además herramientas y repositorios semánticos, los que describen de forma sencilla mediante diagramas.

El artículo de Castells (2003) recalca la situación actual de la web y sus limitaciones, motivando y explicando las propuestas de la Web Semántica. Describe además, las tecnologías más importantes desarrolladas, así como el punto en el que se encuentra el proyecto promovido por el propio inventor de la web.

Otros autores como Fernández y García (2005) exponen también la problemática de la web actual y plantean como posible solución una evolución, en la que los contenidos disponibles en la Web tuviesen asociadas descripciones formales procesables por aplicaciones. Analizan los componentes fundamentales de la Web Semántica y su estado del arte: anotación semántica, ontologías y razonamiento lógico.

La investigación de Castro y Larraud (2007) plantea ideas y soluciones acerca de la recuperación de información bilingüe en la Web Semántica. La búsqueda de un esquema de representación adecuado, como una ontología multilingüe y de lo que es básicamente la Web Semántica, su estructura y características así como componentes presentes.

El libro de Antoniou y Van Harmelen (2008) evidencia el veloz desarrollo de la Web Semántica en los últimos años, mostrando los nuevos elementos que han aparecido en la Web Semántica, las nuevas áreas de aplicación y las nuevas herramientas que están siendo producidas. Dedicar un capítulo para cada una de las tecnologías que componen la Web de los Datos.

Los autores Melo y Botia (2008) presentan a la comunidad académica e industrial, un breve estado del arte de la Web Semántica, en qué consiste y sus beneficios con respecto a la Web actual, mostrando su proyección hasta el año 2030. Además explican conceptos básicos, Ontologías Web, utilidades de las Ontologías Web, lenguajes de las Ontologías Web y trabajos que se han realizado con la Web Semántica.

Por su parte, Danny Ayers (2009) enfatiza en las nuevas tendencias, especificaciones y tecnologías con las que se está trabajando la Web Semántica, describiéndolas en cuanto a sus usos, estructura y funcionamiento. Enfoca con especial atención los vocabularios y las ontologías mostrando algunos cambios y avances en estos términos.

El artículo de David Shotton (2009) brinda un enfoque sobre la próxima revolución en la publicación de revistas científicas atendiendo al reciente desarrollo en las tecnologías de las publicaciones semánticas y haciendo énfasis en los beneficios sustanciales que esto representa para la comunicación científica.

En su trabajo, Jiménez (2010) plasma una amplia explicación de los componentes y la arquitectura de la Web Semántica. Además, muestra una importante representación de Servicios Web Semánticos, que son una tecnología nueva, que resulta al unirse la Web Semántica y los Servicios Web.

1.4.2- A nivel nacional

En Cuba se han llevado a cabo investigaciones relacionadas con la Web Semántica también de mucho valor científico y con un gran aporte al desarrollo del país, a continuación ejemplos de los trabajos realizados.

Perojo (2006) examina los componentes y tecnologías que integran el modelo teórico de la Web Semántica. Explica el modelo constructivo de capas que garantizará el funcionamiento de toda su infraestructura. Ofrece algunas consideraciones en torno a las posibilidades que brinda la Web Semántica en la concepción y desarrollo de sistemas de gestión de contenidos más dinámicos.

Céspedes (2004) expone algunos criterios referentes al desarrollo del WWW como uno de los servicios fundamentales de Internet. Se analizan algunos de sus principales problemas en la representación, organización, búsqueda y recuperación de la información, y se valora la posibilidad de la Web Semántica como alternativa para poner orden al caos.

Labrada (2003) ilustra cuales serían las nuevas perspectivas de desarrollo del profesional con la Web Semántica. Se identifican nuevos horizontes de realización intelectual y de habilidad práctica para estos. Señala un nuevo entorno de trabajo para el profesional de la información con las nuevas tecnologías y se abordan conceptos que en la sociedad de la información de hoy, no se puede permitir desconocer.

1.4.3- Investigaciones en la UCI

En la UCI existen pocos trabajos relacionados con la temática tratada.

Suárez y Paz (2008) formulan una propuesta teórica para la creación de manera automática y semiautomática de las ontologías a partir de texto, para aplicarla en un futuro, en entornos e-learning, específicamente a los repositorios de objetos de aprendizaje, de manera tal que se conviertan en repositorios semánticos.

Bravo y Reytor (2009) abordan la creación de ontologías para modelar la organización de la meta-información de los objetos de aprendizaje y un módulo de inferencia para la gestión de información en las mismas, formando con esto la base para hacer una extensión semántica del repositorio de objetos de aprendizaje, extensión que contará con otros módulos.

1.5. Papel de la producción científica

La práctica de la comunicación y transferencia de los resultados de investigación por medio de revistas científicas surge a finales del siglo XVII con la aparición del *Journal des Savans*, en Francia y la *Philosophical Transactions of the Royal Society*, en Inglaterra (PATALANO, 2005). Esa experiencia se extendería y adaptaría según las necesidades de las comunidades científicas, incrementándose el número de publicaciones existentes, la especialización de sus temáticas y la cantidad de artículos publicados (KUHN, 2004). La gran explosión en cuanto al incremento del número de revistas académicas y su protagonismo en el reconocimiento de la actividad científica se concretaría en el período post bélico de la Segunda Guerra Mundial a raíz de la revolución científica en las esferas de la química y la biología (BRIAN, 2009).

Paralelamente al desarrollo de las revistas científicas surgirían fuentes secundarias dedicadas a la sistematización, evaluación, promoción y distribución de todo el volumen de información generado en los canales de publicaciones científicas, estas se pueden clasificar en índices de fuentes e índices de citas, el principal exponente de este tipo de fuentes lo constituyó el *Science Citation Index* publicado por el *Information Science Institute* en 1963 (PORTAL, 2009), y cuyas principales aplicaciones han sido: la evaluación cuantitativa de las publicaciones científicas, el análisis del desarrollo histórico de la ciencias y la recuperación bibliográfica (ARCHAMBAULT, ÉRIC y LARIVIÈRE, 2009).

A partir de los años 90 con el advenimiento de Internet como fenómeno global ocurriría otro cambio en el paradigma de los canales de comunicación científica con la aparición de las revistas electrónicas y las bases de datos bibliográficas en línea, que transformaron completamente el escenario de la industria editorial de ciencia y técnica y la forma en que los investigadores consumían y compartirían la información (GOODRUM *et al.*, 2001).

La Ciencia de la Computación no fue ajena a estas prácticas, durante mucho tiempo predominó como canal de comunicación principal el de memorias de eventos, tendencia que se ha revertido en los últimos años, consolidándose el de revistas científicas como vía principal según estudios realizados por el *Research Information Network and Joint Committee on Quantitative Assessment of Research* (2009).

Los primeros estudios y análisis de la producción científica se llevaron a cabo en la primera mitad del siglo XX, tomándose como pioneros los realizados por Coles y Eales, seguidos por los trabajos estadísticos de historia de la ciencia realizados por Hulme y Gross a partir de las referencias utilizadas en *The Journal of the American Chemistry Society* (YANG, YANG, 2009), muy notable para el desarrollo de esta disciplina serían los aportes realizados por Bradford (CHAO-CHIH, 2009) más tarde conocidos como Ley de Bradford para la distribución de la producción científica, estas investigaciones sentarían las bases de la Bibliometría a partir de 1969. Tagguy y Stuclyff definen que la Bibliometría “estudia los factores cuantitativos de la producción, uso y difusión de la información registrada, mientras permite tomar decisiones, valiéndose de los resultados provenientes de la construcción de modelaciones y estimaciones matemáticas” (ROCHE RODRÍGUEZ y MARTÍ LAHERA, 2008).

Los antecedentes de estudios bibliométricos enfocados a la producción científica de la Ciencia de la Computación aparecen en las década de los setenta, donde se verá reflejada como subdisciplina de las ramas del conocimiento que le dieron origen, fundamentalmente desde la Matemática, como parte de la Matemática Aplicada y de la Ingeniería (AZDEN, 1976; GARFIELD, 1973). En 1976 se publica el primer intento por definir la zona núcleo de publicaciones para Ciencia de la Computación a través de la aplicación de la Ley de Bradford (SUBRAMANYAM) y Hirts (1977) divulga en una publicación de la IEEE el primer análisis de citas para la temática, y Culnan (CULNAN, 1978) estudia los patrones de citas de los científicos a partir de la producción científica en eventos especializados.

En 1979 se encuentra el primer antecedente de publicación de la Ley de Lotka para determinar los autores más productivos (SUBRAMANYAN, 1979) y Salton realiza un estudio de citas a partir de la literatura científica en Ciencia de la Computación. Estos trabajos sirven de evidencia para demostrar la existencia de un volumen de publicaciones científicas considerable que ya alcanzan un nivel de especialización hacia la Ciencia de la Computación, de la misma forma que demuestran la presencia de una comunidad de investigadores con patrones de comportamientos diferentes a las comunidades donde se gestó la Ciencia de la Computación como disciplina. El principal el valor de estas contribuciones con respecto a trabajos anteriores será la consideración de la Ciencia de la Computación como una rama del conocimiento independiente.

En la década de los ochenta Garfield estudia la influencia de la multidisciplinariedad en la Ciencia de la Computación a partir de un análisis de la producción científica disponible en *CompuMath Citation Index* para el período de 1976-1980 (GARFIELD, 1984b) y este mismo autor determina en un trabajo a partir de la misma fuente cuales son las monografías de mayor impacto en el dominio (GARFIELD, 1984a). A estas aportaciones se les suman las realizadas por Subramanyam (SUBRAMANYAM, 1984) y utiliza como fuente los artículos indexados por *Computer and Control Abstracts* donde identifica a partir de un núcleo de 419 la dispersión de la productividad en la disciplina.

En los años noventa Barnhill y Linton (BARNHILL y LINTON, 1992) analizan la productividad de los departamentos de computación en los Estados Unidos, este ha sido el primer trabajo donde se realiza un análisis en específico de la productividad a nivel institucional. Por su parte, Wagner (WAGNER-DOBLER, 1997) analiza el apareo bibliográfico en los patrones de citación entre la Matemática y la Ciencia de la Computación.

Hasta el año 2000 las investigaciones recuperadas en la búsqueda bibliográfica estudian el estado de la investigación en áreas geográficas muy específicas donde se incluían a los países tradicionalmente reconocidos como potencias científicas: Estados Unidos y algunos países europeos con reconocimiento dentro de la comunidad científica. Y ya en el 2000 comienzan a realizarse investigaciones en países en vías de desarrollo o emergentes como son los estudios realizados por Yinian sobre la productividad de investigadores malayos presentes en el *Science Citation Index* (YINIAN, 2002; YINIAN y ZAINAB, 2001) y las propuestas de Liang, Kumar y Yang sobre la productividad de investigadores chinos, su estructura de colaboración con investigadores homólogos en países del tercer mundo y el caso específico de colaboración con investigadores indios (KUMAR y GARG, 2005; LIANG *et al.*, 2001; YANG, ZHONGKAI, 2010); la productividad de los departamentos docentes de Ciencia de la Computación en universidades griegas (KATSAROS *et al.*, 2008) y resulta interesante por lo cercano al entorno cubano la investigación de Arruda y Bezera sobre la productividad de los investigadores brasileños (ARRUDA *et al.*, 2006).

Al mismo tiempo comienzan a realizarse estudios que analizan el reconocimiento dentro

de la comunidad académica de las fuentes bibliográficas utilizadas (PETRICEK *et al.*, 2005) y el impacto que ha tenido Internet como herramienta de trabajos (GOODRUM *et al.*, 2001) y de los recursos disponibles en la web como fuentes de información (ZHAO y LOGAN, 2002).

Como un indicador del nivel de especialización alcanzado por la Ciencia de Computación y de la importancia de los estudios métricos para determinar su estado pueden tomarse las propuestas de Melo y Rogers a partir del estado de la investigación sobre interacción hombre/computadora (MEHO, LI y ROGERS, 2008) y las aportaciones realizadas por Marion y McCain a partir de estudios de la producción científica en Ingeniería de Software (MARION y MCCAIN, 2001).

Por muchos años el *Science Citation Index* (SCI) del *Web of Science* constituyó el único servicio disponible para la realización de estudios sobre producción científica a partir de indicadores bibliométricos (TORRES-SALINAS y JIMÉNEZ-CONTRERAS, 2010), a pesar de diversas investigaciones críticas sobre sus métodos de selección de publicaciones, volúmenes de sus listas de títulos y sesgos presentes en el análisis de citas.

En los últimos años han surgido diversas propuestas como alternativas, entre las que se destaca la base de datos Scopus, desarrollada por Elsevier, y servicios de búsqueda especializada de información como *Google Scholar*. Este fenómeno ha provocado una serie de trabajos enfocados a distinguir particularidades, ventajas y desventajas de sus procedimientos de selección de fuentes, indicadores utilizados y métodos de organización y representación de los contenidos, tanto de forma general (ARCHAMBAULT, É *et al.*, 2009; FRANCESHET, 2009; JACSÓ, 2005; LEYDESDORFF *et al.*, 2009; MOYA-ANEGÓN *et al.*, 2007; PISLYAKOV, 2009) como en disciplinas específicas (FALAGAS *et al.*, 2008; MEHO, L. I. y YANG, 2007; VARGAS-QUESADA *et al.*, 2007) precisamente enfocadas en Ciencia de la Computación. Llamen la atención las propuestas de Franceschet (2009) a partir de la comparación de los indicadores bibliométricos utilizados por *Web of Science* y *Google Scholar*, y las realizadas por Meho y Roger (2008) sobre el estado de la producción en la temática de *Human-Computer Interaction* a partir de la producción presente en el SCI y Scopus.

1.6. Aportes Investigativos

El resultado de la investigación dará una visión del estado de la producción científica en la temática Web Semántica, en el período comprendido del 2005 al 2009, lo que ayudará a definir los objetivos y alcances del proyecto perteneciente al departamento de Universidad Digital del centro CENIA de la Facultad 1 en la UCI. Los resultados obtenidos permitirán la identificación de líneas de investigación trabajadas por la comunidad científica, para trazar la proyección del proyecto.

El conocimiento de las instituciones con mayor producción científica, permitirá la identificación de alianzas de cooperación, definir y conocer quiénes son los actores que pautan la investigación en Web Semántica en el mundo, así como su proyección futura.

Se conocerán los autores más productivos y sus temas de especialización a partir de la producción científica y relaciones de colaboración con otros investigadores.

Se tendrá un control y dominio de las relaciones de colaboración a niveles nacionales e institucionales, además se podrá determinar el núcleo de revistas con mayor especialización en el tema de investigación

Capítulo II: Materiales, métodos y herramientas

2.1 Métodos y procedimientos

Para la realización del trabajo se utilizaron métodos de estudios métricos de la información, que posibilitaron la medición e identificación de tendencias y patrones a partir de la muestra de documentos seleccionada en el estudio.

Los estudios métricos de la información tienen sus antecedentes en los procesos de penetración de modelos y métodos matemáticos en otras ramas del conocimiento que se han llevado a cabo con el objetivo de garantizar la cientificidad del discurso intelectual de una rama del conocimiento determinada o en la formación de leyes que identifiquen el desarrollo de fenómenos o procesos sociales (GORBEA PORTAL, 2004). Bajo este término se representan un conjunto de especialidades métricas como: la bibliometría, informetría y cienciometría (PORTAL, 2009).

En el caso específico de esta investigación se utilizaron técnicas de análisis bibliométricos, definidas por primera vez en 1969 aunque ya existían antecedentes de su utilización desde 1917 por Cole y Eales (SPINAK, 2004).

Pritchard, define el término de Bibliometría como: “la aplicación de las matemáticas y los métodos estadísticos para analizar el curso de la comunicación escrita y el curso de una disciplina” o “como la aplicación de tratamientos cuantitativos a las propiedades del discurso escrito y los comportamientos típicos de este” (PRITCHARD, 1969).

Otros autores han definido como razón principal de los estudios bibliométricos la identificación de los actores que intervienen en los procesos de comunicación científica: autores, instituciones, países; sus relaciones y tendencias a partir del estudio de la organización de los sectores científicos y tecnológicos reflejados en las fuentes bibliográficas (SETIÉN, 2004).

El Diccionario Enciclopédico de Bibliometría, Cienciometría en Informetría (SPINAK, 2004) atendiendo a las fuentes de datos utilizadas por los estudios bibliométricos, los divide en:

- Basados en bibliografías, servicios de indexación y resúmenes.
- Basados en referencias y citaciones.
- Basados en directorios o catálogos colectivos de títulos de revista.

Y clasifica las aplicaciones de los estudios basados en esta especialidad en:

- Selección de libros y publicaciones periódicas.
- Caracterización temática de la literatura.
- Historia de la ciencia.
- Estudios de sociología de la ciencia.

2.2 Indicadores bibliométricos

El Diccionario de la Real Academia de la Lengua Española (RAE, 2001), define el término indicador como “aquello que indica o sirve para indicar”. En el ámbito de los estudios bibliométricos el uso de los indicadores permitirá describir y evaluar un fenómeno, su naturaleza, estado y evolución. Spinak (SPINAK, 2004) reconoce que será una medida que proveerá información sobre resultados de la actividad científica, institución, país, autores individuales o colectivos. Chaviano (CHAVIANO, 2008) menciona que facilitan la identificación de información cuantitativa y objetiva relacionada al estado de la actividad investigativa según su volumen, evolución, visibilidad y estructura. Macías Chapula resalta que su utilización se debe realizar desde un enfoque comparativo (MACIAS-CHAPULA, 1998).

Existen diferentes propuestas de clasificación para los indicadores bibliométricos, en la investigación se utilizará la propuesta realizada por Spinak (SPINAK, 1998), donde se organizan dos grandes grupos: indicadores de actividad o producción e indicadores de impacto, seleccionándose la categoría de indicadores de producción por ser la que más se ajusta a los objetivos de la investigación definiéndose las siguientes variables:

- Productividad por año: número de artículos publicados por las revistas referenciadas en CASC en los años comprendidos en el estudio.
- Productividad por autores: número de artículos publicados por autor.
- Productividad por publicación: número de artículos publicados en una publicación seriada y referenciada por CASC de EBSCO.
- Superposición de cobertura: número absoluto de publicaciones indexadas simultáneamente por dos servicios de indexación y resumen.
- Productividad por institución: número de artículos publicados según la adscripción institucional de los autores a partir de los datos declarados en el registro bibliográfico en el momento de publicación del artículo.
- Productividad por país: número de artículos publicados por país según la adscripción de los autores a partir de los datos declarados en el registro bibliográfico en el momento de la publicación del artículo.
- Productividad por palabras clave: número de artículos que presentan una determinada palabra clave en común asignada por el sistema de indexación de CASC en el momento de agregación del artículo a la base de datos.
- Colaboración autoral: números de relaciones establecidas entre diferentes autores para la publicación de artículos.
- Colaboración institucional: número de relaciones establecidas entre diferentes instituciones a partir de la publicación de artículos.
- Colaboración entre países: número de relaciones establecidas entre diferentes países a partir de la publicación de artículos.

Para una mejor interpretación de los datos obtenidos se utilizó como técnica de análisis la representación de los resultados por medio de técnicas de redes sociales, permitiéndose así la visualización de las diferentes relaciones establecidas en las variables de colaboración autoral, colaboración institucional, colaboración entre países y análisis de co-ocurrencia de palabras clave. El uso de redes sociales ha sido ampliamente utilizado y documentado en los estudios métricos de la información como lo demuestran un número importante de investigaciones consultadas (ALLEN *et al.*, 2006; BALLARD y HENRY, 2006; BJORNEBORN, 2006; COLINA, 2005; CHAO-CHIH, 2009; HANNEMAN, 2000; HJORLAND y ALBRECHTSEN, 1995; KLINK *et al.*, 2006; MINGERS y LIPITAKIS, 2010; NASCIMENTO y MARTELETO, 2008; ZINS y GUTTMANN, 2003).

2.3 Fuentes

Como fuente de información principal se utilizó la base de datos *Computers & Applied Sciences Complete* (CASC) distribuida por la empresa de EBSCO y suscrita por la Universidad de las Ciencias Informáticas desde el 2004 por medio de compra cooperada a través de las instituciones de información del Ministerio de la Informática y las Comunicaciones de Cuba (figura 3).

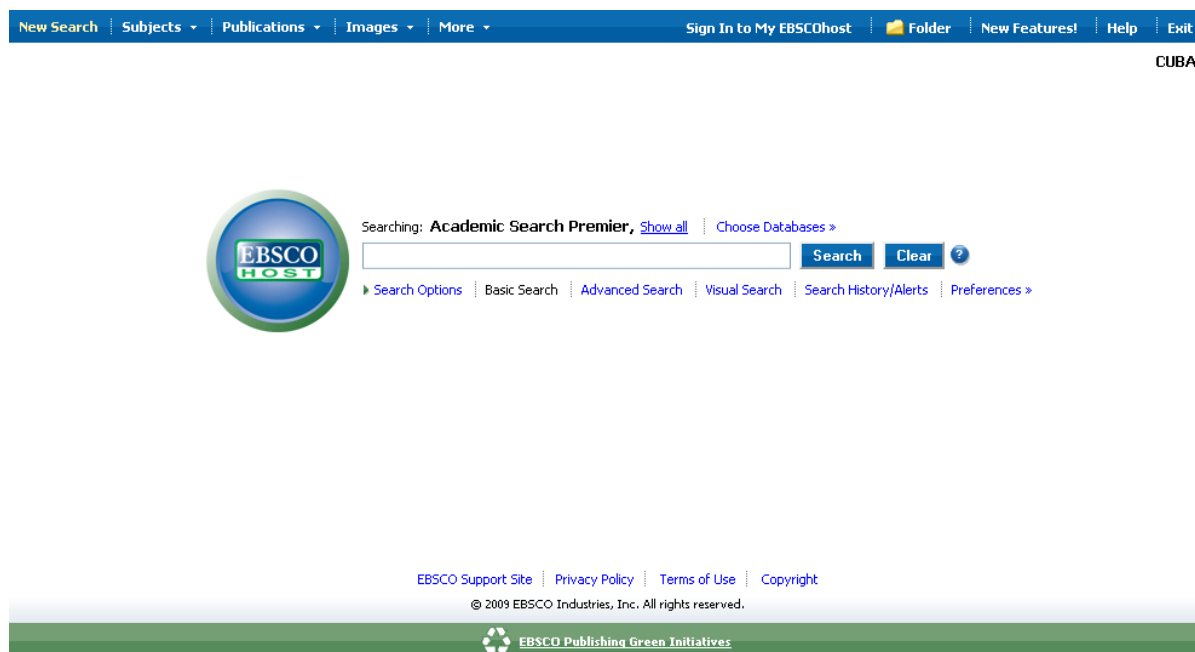


Figura 3: Interfaz de trabajo de las bases de datos bibliográficas disponibles en EBSCO.

CASC está conformada por un acervo de 2152 publicaciones de diferentes categorías (figura 4) de las cuales 1329 (62%) corresponde a la categoría de revistas científicas, 284 (13%) a reportes técnicos de tecnologías o herramientas informáticas y 265 (12%) a libros o partes de libros. Entre estos tres tipos de publicaciones suman más del 87% del total de publicaciones, además se incluyen otras categorías de documentos como pueden ser reportes genéricos, información promocional y enciclopedias pero que entre todos no son representativos de la base de datos al totalizar menos del 1% en la muestra.

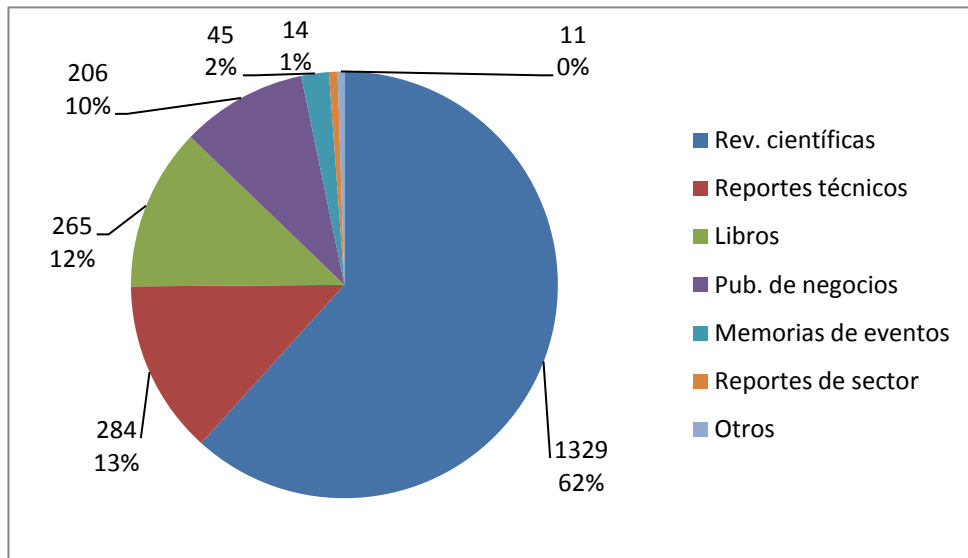


Figura 4. Distribución según tipo de publicaciones en CASC.

Atendiendo al nivel de acceso a los documentos referenciados por la base de datos (figura 5), el servicio permite consultar el 100% a nivel descriptivo más resumen, y 1357 publicaciones dan acceso al texto completo de sus contenidos (para un 62% del total) de los cuales solo 314 están limitadas por alguna forma de embargo en un período de tiempo que va desde los 15 días hasta los 24 meses, como una forma de los editores para proteger los ingresos por suscripción a las revistas. Como un valor agregado de la base de datos aunque no representativo, el servicio permite la búsqueda y disponibilidad de las referencias bibliográficas en 502 (23%) títulos, mayoritariamente en el subgrupo de las revistas académicas.

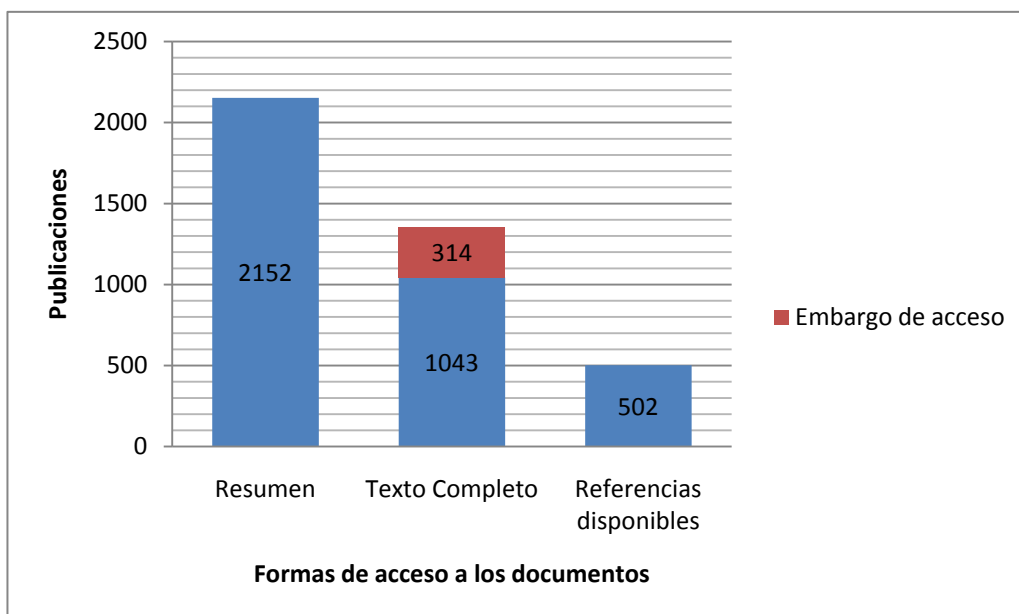


Figura 5. Distribución según niveles de acceso a los documentos en CASC.

Si se analiza la presencia de los grupos editoriales en la base de datos, como se observa en la figura 6, las diez casas editoriales con mayor número de publicaciones cubren un total de 835 títulos alcanzando el 39%, se observará que predominan los más grandes grupos editoriales de ciencia y técnica: *Wiley*, *Springer*, *Elsevier* y *Taylor & Francis*, y también están presentes las editoriales de la IEEE, la ACM e IET como sellos especializados en Ciencia de la Computación.

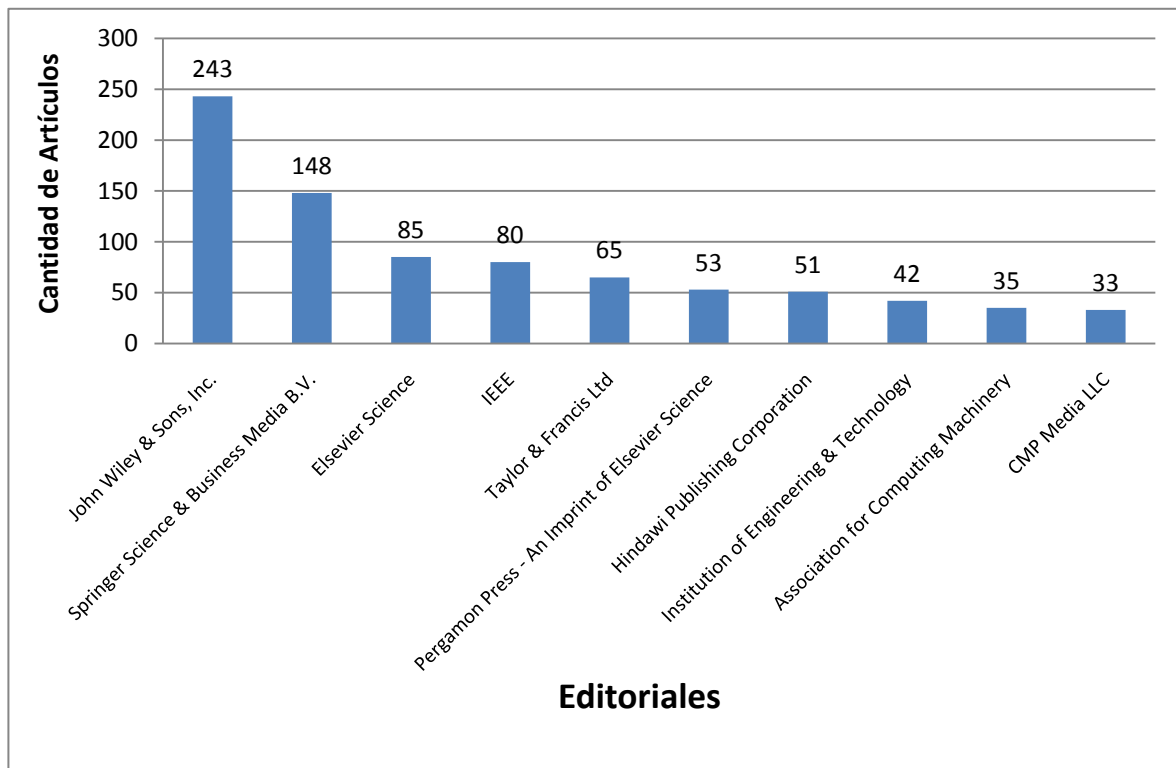


Figura 6. Cobertura editorial presente en CASC.

Para contrastar el impacto de las revistas referenciadas por CASC se utilizó la lista de títulos del servicio de indexación Scopus, disponible a través del sitio del Grupo Scimago de la Universidad de Granada España (SCIMAGO GRUPO, 2007), aquí se identificaron los títulos pertenecientes a las doce categorías de Ciencia de la Computación totalizando 916 publicaciones. De la misma forma se utilizaron los títulos disponibles en la página informativa de cobertura de títulos del grupo Thomson Reuters para el servicio *Science Citation Index Expanded*, (SCIE) identificándose siete categorías relacionadas directamente con la Ciencia de la Computación para un total de 459 publicaciones (THOMSON REUTERS, 2010).

2.4 Estrategia de búsqueda

Para recuperar la producción pertinente a la temática de Web Semántica en la base de datos se utilizó el artículo de reconocido impacto en la temática de Berners-Lee publicado en 2001 y se realizó una búsqueda por el campo de autor, en el registro recuperado se identificaron las palabras clave utilizadas por el sistema de indexación, encontrándose la presencia del descriptor “*semantic web*” como parte del tesoro del sistema de búsqueda y recuperación.

Al realizar la primera interacción con la base de datos se obtuvieron un total de 886 artículos, sin limitar el marco temporal de la búsqueda. En una segunda interacción con el sistema se limitó el marco temporal al período 2005-2009, no se consideran los resultados correspondientes al 2010 pues la carga de los registros en las bases de datos no se realiza a tiempo real y la consideración de los mismos pudieran generar malas interpretaciones de los datos. Como resultado final se obtuvieron un total de 622 artículos, los cuales se descargaron en formato RIS y se llevaron posteriormente a un gestor bibliográfico para su normalización (figura 7).

TI- A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags.
 AU- Tsui, Eric^1
 AU- Wang, W.M.^1
 AU- Cheung, C.F.^1
 AU- Lau, Adela S.M. Email Address: Adela.Lau@inet.polyu.edu.hk
 JN- Information Processing & Management
 PD- Jan2010, Vol. 46 Issue 1, p44-57
 PG- 14p
 DT- 20100101
 PT- Article
 AB- Abstract: Taxonomy construction is a resource-demanding, top-down, and time consuming effort. It does not always cater for the prevailing context of the captured information. This paper proposes a novel approach to automatically convert tags into a hierarchical taxonomy. Folksonomy describes the process by which many users add metadata in the form of keywords or tags to shared content. Using folksonomy as a knowledge source for nominating tags, the proposed method first converts the tags into a hierarchy. This serves to harness a core set of taxonomy terms; the generated hierarchical structure facilitates users' information navigation behavior and permits personalizations. Newly acquired tags are then progressively integrated into a taxonomy in a largely automated way to complete the taxonomy creation process. Common taxonomy construction techniques are based on 3 main approaches: clustering, lexico-syntactic pattern matching, and automatic acquisition from machine-readable dictionaries. In contrast to these prevailing approaches, this paper proposes a taxonomy construction analysis based on heuristic rules and deep syntactic analysis. The proposed method
 SU- SEMANTIC Web
 SU- TAGS (Metadata)
 SU- FOLKSONOMIES
 SU- NATURAL language processing (Computer science)
 SU- WEB personalization
 SU- TAXONOMY
 SU- AUTOMATION
 AD- ^1 Knowledge Management Research Centre, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong
 IS- 03064573
 DI- 10.1016/j.ipm.2009.05.009
 AN- 44699595
 UR-
<http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=44699595&site=ehost-live>
 UR-
<http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=44699595&site=ehost-live>

Figura 7. Ejemplo de registro bibliográfico en formato RIS descargado de la base de datos bibliográfica.

2.5 Herramientas

Para la realización del trabajo se utilizaron diversas tipologías de herramientas especializadas en estudios métricos de información atendiendo a los diferentes procedimientos que se llevan a cabo: gestión y organización de la información analizada, procesamiento, análisis y generación de metáforas visuales para la interpretación de los resultados

2.5.1 Gestión y manejo de los registros bibliográficos

Para la gestión y manejo de los registros bibliográficos recuperados en CASC se utilizó el EndNote X4. Es un software comercial para la gestión de colecciones bibliográficas, se utiliza para administrar bibliografías y referencias al escribir ensayos y artículos. Es creado y comercializado por Thomson Reuters, y dentro de su categoría es la herramienta de gestión bibliográfica más utilizada dentro de la comunidad científica. Permite la importación de registros de forma directa desde las bases bibliográficas en línea al igual que a partir de un fichero txt en formato RIS, al mismo tiempo posee herramientas para la edición global de los registros, la generación de conteos y da al investigador la posibilidad de modificar los formatos de exportación de datos y estructuras lo que facilita su exportación a un gran número de herramientas y servicios de análisis bibliográficos.

2.5.2 Bibexcel (v.2011)

Es un software desarrollado por Olle Persson, de la Universidad de Umea, Suecia, diseñado para ayudar al usuario en el análisis de los datos bibliográficos o cualquier otro dato de carácter textual (PILKINGTON, 2006). Genera archivos de datos que se pueden importar a Excel u otro programa que tome los datos de los registros bibliográficos y permite la integración con las bases de datos generadas en EndNote (ALÍ HERRERA, 2009). Como herramienta goza de gran prestigio y respaldo dentro de la comunidad de investigadores de estudios métricos de la información (ÅSTRÖM y SCHLEMMER, 2009). Aunque la herramienta ha sido desarrollada para entornos Windows sus desarrolladores afirman que puede utilizarse en otras plataformas por medio de emuladores como Wine para Linux o Mac.

2.5.3 Ucinet 6

Es una suite de herramientas para la generación de redes sociales desarrollado por Borgatti, y colaboradores en la universidad de Harvard, integra un conjunto de programas (*NetDraw*, *Pajek* y *Mage*) que permiten la gestión de los datos y algoritmos de visualización y al mismo tiempo herramientas para la visualización de las redes tanto en gráficos planos como en tercera dimensión. Es una herramienta desarrollada para entornos Windows, pero según sus desarrolladores puede utilizarse en plataformas Mac y Linux a través del uso de emuladores recomendando la utilización de Wine (MOLINA, 2005).

Capítulo III: Análisis de los resultados

3.1 Producción de artículos por año

El análisis de la producción de artículos por años (figura 8) mostró un crecimiento constante en los cinco años comprendidos en el estudio, con un crecimiento de 148 ítems, aumentando la producción un 616% con relación al primer año del estudio, para un promedio de 29 artículos por año. Los años de mayor crecimiento de la producción corresponden al 2006, con 44 artículos con respecto al 2005 y al 2009 con 47 documentos en relación al 2008. El crecimiento más considerable en la etapa estudiada lo constituyó el bienio 2005 -2006, donde la presencia de ítems relacionados con la temática en estudio creció un 183% en relación al 2005. La línea de tendencia confirmó el patrón de incremento anual dentro de la producción en Web Semántica.

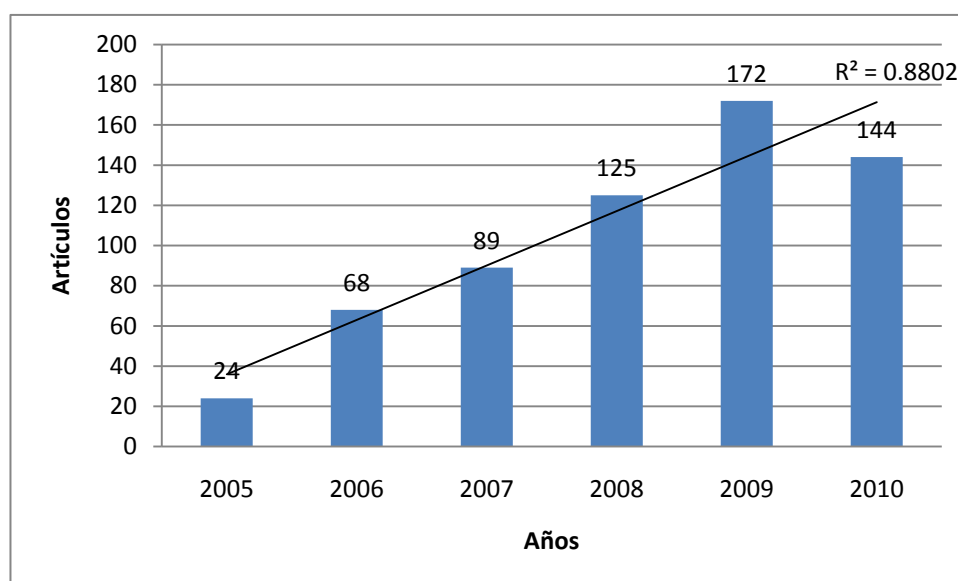


Figura 8: Distribución de la producción de artículos en Web Semántica en CASC de 2005 a 2009.

3.2 Solapamiento y superposición de la *Computers & Applied Sciences Complete* con Scopus y el *Science Citation Index Expandex (SCIE)*

El análisis de la superposición de cobertura para la base de datos CASC al compararla con las listas de títulos de Scopus y el SCIE, se observa en la tabla 3. Entre CASC y Scopus se totalizan un total de 4969 títulos, de los cuales se solapan 389, por lo que CASC da cobertura al 42% de las revistas de Ciencia de la Computación presentes en Scopus. En el caso de SCIE se totalizaron 2294, siendo coincidentes 317 publicaciones para un 69%. Se puede afirmar que CASC garantiza una mayor cobertura de títulos indexados en el *Web of Science*.

Tabla 3: Solapamiento de CASC con Scopus y SCle.

Servicio\ Cant. de publicación	Exclusivas en Scopus y JCR	Solapadas	Exclusivas CASC	% de solapamiento
Scopus + CASC	527	389	1763	42%
JCR + CASC	142	317	1835	69%

Debe tenerse en cuenta que el número de publicaciones SCle es menor en comparación con la lista de títulos presentes en Scopus, de forma tal que a pesar de estar solapadas un mayor número de revistas de Scopus, el por ciento de cobertura CASC para SCle sea mayor. Si se aplica la fórmula de cálculo para el índice superposición (IS) donde:

$$IS = \frac{\text{número de revistas indexadas en ambos sistemas}}{\text{revistas distintas en sistema A + sistema B}}$$

Se obtienen los valores de 0,160344 para SCle y CASC y de 0.168559 para Scopus y CASC, por lo que se confirma que la base de datos Scopus presenta una mayor superposición.

3.3 Producción y distribución por publicaciones

A partir de los 622 registros recuperados en la base de datos se contabilizaron un total de 198 publicaciones seriadas académicas, lo que representa que se recuperaron artículos en el 9% de las revistas referenciadas en CASC, de estas 198 publicaciones, 101, el 51%, son referenciadas al mismo tiempo por SCle y 116 de ellas aparecen simultáneamente en Scopus.

Para determinar la zona núcleo de las revistas más productivas en la muestra identificada, se utilizó la Ley de Bradford o ley de la dispersión bibliográfica. Se decidió organizar la dispersión de las publicaciones en tres zonas: zona núcleo, revistas más productivas; zona núcleo II, revistas medianamente productivas; zona III, revistas poco productivas. En la tabla 4 se observa la composición de cada una de las zonas.

Tabla 4. Dispersión bibliográfica para la temática de Web Semántica en las revistas de CASC.

Zonas	No. de revistas	No. de artículos
Zona núcleo	9	175
Zona II	41	219
Zona III	148	228
Total	198	622

La zona núcleo o principal conformada por las revistas más productivas representa solo el 4% de las revistas, entre ellas agrupan el 28% de todos los artículos publicados. La zona II de las revistas medianamente productivas acumula el 20% de los títulos y cubren el 35% de los trabajos y la zona III de las revistas de baja productividad conforman el 75% por lo que constituyen y acumulan el 37% de los ítems.

En la figura 9 se observa la gráfica de distribución para la zona núcleo de las revistas más productivas, donde resalta como principal productor la *IEEE Intelligent Systems*, y que es referenciada por los servicios de Scopus y SCle, de estos títulos solamente, como mismo sucede con *Expert Systems wiht Applications*, *IEEE Internet Computing*, *IEEE Transactions on Knowledge and Data Engineering*, *Theory and Practice of Logic Programming*, *Knowledge-Based Systems*, *Information Processing and Management*; no sucede con *Journal of Software* referenciada solamente por Scopus, ni con *Concurrency & Computation: Practice & Experience* incluida dentro de SCle. De forma general todas las revistas de la zona principal se encuentran dentro de la corriente principal de revistas académicas para Ciencia de la Computación.

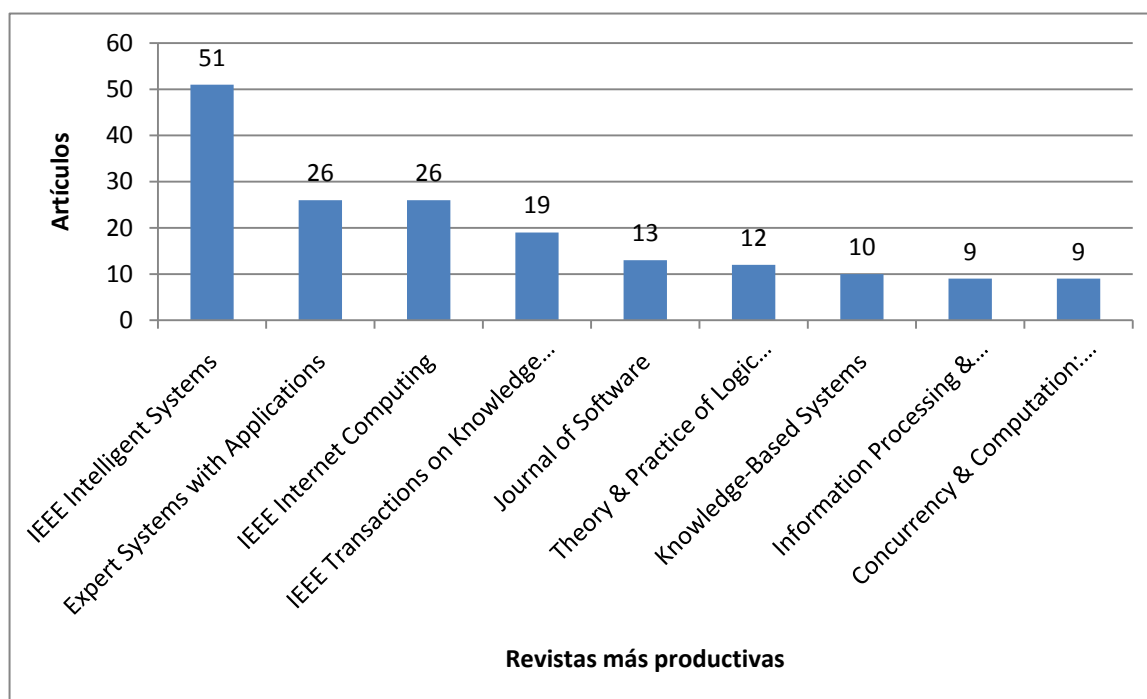


Figura 9: Distribución de artículos en Web Semántica en la zona de revistas más productivas dentro de CSAC en el período 2005-2009.

3.4 Producción y distribución por autores

En la muestra fue posible realizar la identificación de 1486 autores, sin discriminar el orden de aparición en el campo de autor de los registros bibliográficos para un promedio de 2.38 autores por artículo.

En la figura 10 se aprecia la distribución de autores por número de contribuciones realizadas en el período estudiado, se observa que 1317 de ellos, el 89% de los identificados, solo participaron en la publicación de un artículo en la temática de Web Semántica en los cinco años comprendidos en el estudio. Los autores con una productividad de entre dos y cuatro artículos conforman un conjunto de 164 personas que representan el 11%. Se consideraron como autores más productivos aquellos que lograron mantener una producción igual o mayor de cinco artículos, contabilizándose solamente 5 investigadores del total para una representatividad de 0.33%.

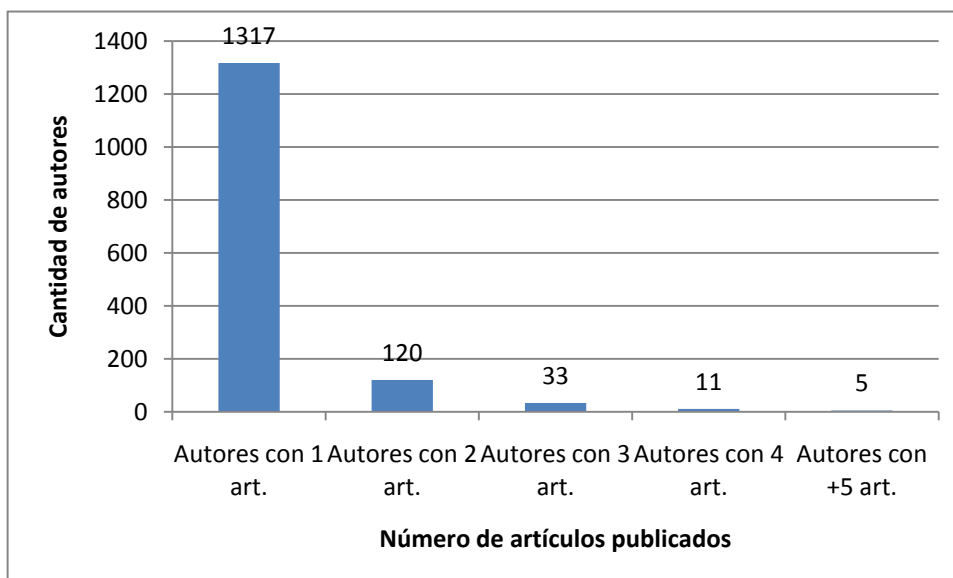


Figura 10: Distribución de autores por cantidad de contribuciones realizadas para la temática de Web Semántica en CSAC en el período 2005-2009.

El patrón de los autores más productivos se observa en la figura 11. El autor con mayor productividad en la disciplina es Sheth, A. con un acumulado de nueve trabajos, de los cuales solo aparece como primer autor en cuatro de ellos, en el resto no se indica como autor principal; es seguido por Hendler, J. con un total de siete contribuciones en las que aparece reflejado como autor principal en cuatro de ellas; le sigue Wilks, Y. con seis, en todos como autor principal; Petrie, C. y Decker ocupan los dos últimos lugares con cinco contribuciones cada uno.

Estos resultados concuerdan con las afirmaciones de Price sobre la colaboración entre investigadores, donde destaca que en las comunidades científicas existirá un núcleo pequeño de investigadores muy productivos y una “población flotante” de otros investigadores que colaborarán con los más productivos.

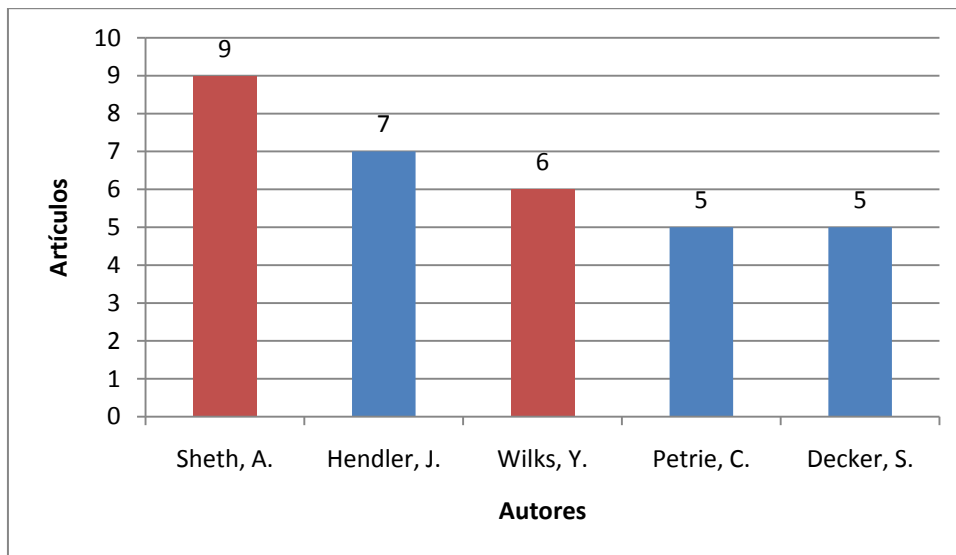


Figura 11: Autores con mayor productividad en la temática de Web Semántica a partir de la producción disponible en CASC en el período 2005-2009.

Es interesante igualmente el análisis de los primeros autores donde solo se presentan 168, cubriendo el 11% del total de autores recuperados, siendo Sheth, A. y Wilks, Y. señalados en rojo en la figura 11, los más productivos a un primer nivel de autoría con más de cinco artículos como autores principales.

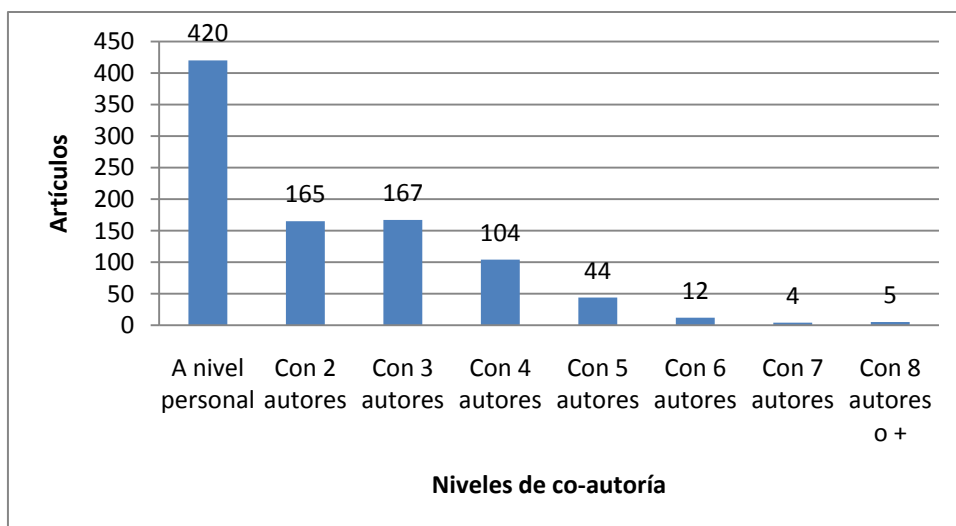


Figura 12: Distribución de los niveles de colaboración en la publicación de artículos en Web Semántica disponible en CASC en el período 2005-2009.

Al analizar los niveles de co-autoría se encontró que de la muestra total, 502 trabajos se publican en forma de co-autoría, para un 54%. Los niveles de coautoría van desde colaboraciones de dos investigadores hasta trabajos conformados por diez individuos. En la figura 12 se observan más detalles de los niveles de colaboración. La forma de

colaboración predominante es entre tres autores, seguida por la de dos autores, entre estas dos formas de colaboración alcanzan el 62% de los trabajos publicados de forma colaborativa.

La red de la figura 13 muestra las relaciones de colaboración entre los autores con más de tres publicaciones entre el 2005 y el 2009. La intensidad de la red la conforman 49 nodos que representan a los autores, el volumen de los mismos está dado a partir de la centralidad nodal según el número de relaciones establecidas con otros actores en el grafo; el grosor de las líneas de conexión establecidas denota el volumen de la colaboración existente.

Se identificaron catorce redes de colaboración, y once nodos se encontraron aislados. Las redes de colaboración más intensas la conformaron Petrie, Bussler, Finin, Sheth, Lytras y Pablos con grado 6; la establecida por Shadbolt, Berners-Lee, Hendler y Goldbeck; y Chen, Studer, Hitzler y Qi, ambas de grado 4.

Las relaciones de colaboración más fuertes no ocurren dentro de las redes de mayor tamaño si no entre redes menores conformadas en su mayoría por solo dos actores como las existentes entre Brewster y Wilks y Goble y Stevens.

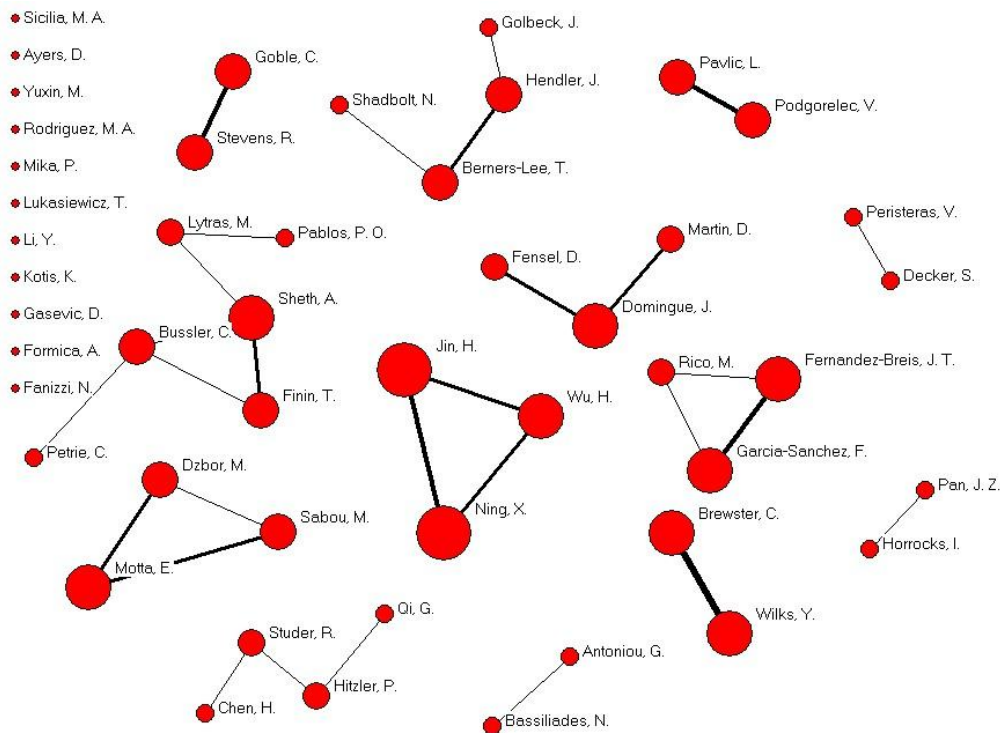


Figura 13: Redes de colaboración entre autores, productividad ≥ 3 .

3.5 Producción y distribución por institución

A partir de las adscripciones declaradas por los autores a la hora de realizar la publicación se pudieron recuperar 624 instituciones, solamente tres de los registros no poseían información que permitiera la identificación de la institución.

El criterio de consideración para las instituciones más productivas fue la presencia de al menos un acumulado de 5 artículos en la muestra. A partir de este elemento clasificaron como instituciones más productivas 29 entidades que representan el 4% de las identificadas (figura 14), como tipología de institución prima la presencia de universidades, destacándose únicamente IBM (en rojo) como entidad corporativa en la muestra, los centro de investigación no presentaron una producción suficiente para considerarse.

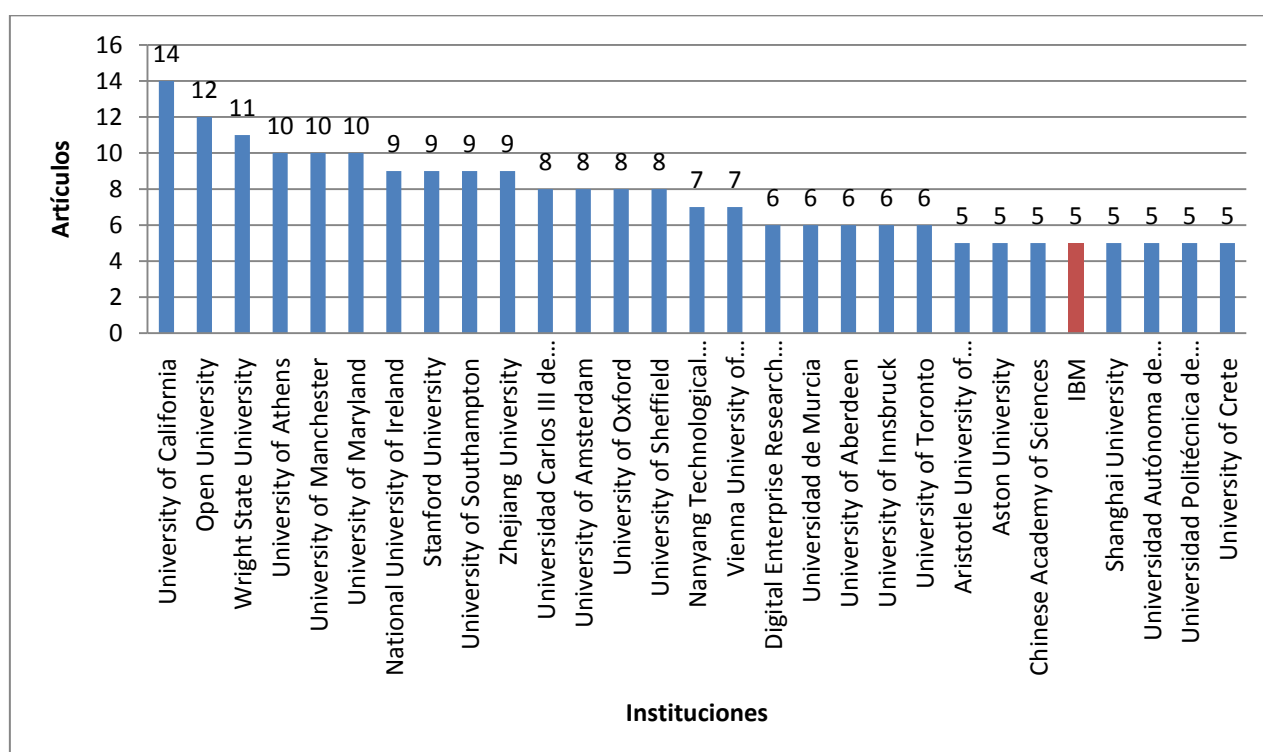


Figura 14: Niveles de producción entre las instituciones más productivas en la publicación de artículos en Web Semántica disponible en CASC en el período 2005-2009.

Se utilizaron tres criterios para la clasificación de las entidades presentes en la muestra: universidades, centros de investigación y empresas (figura 15) siendo las universidades los centros más activos en la producción con 525 entidades identificadas, para un 84%; las empresas presentes superan por muy poco (cinco unidades) el número de los centros de investigación.

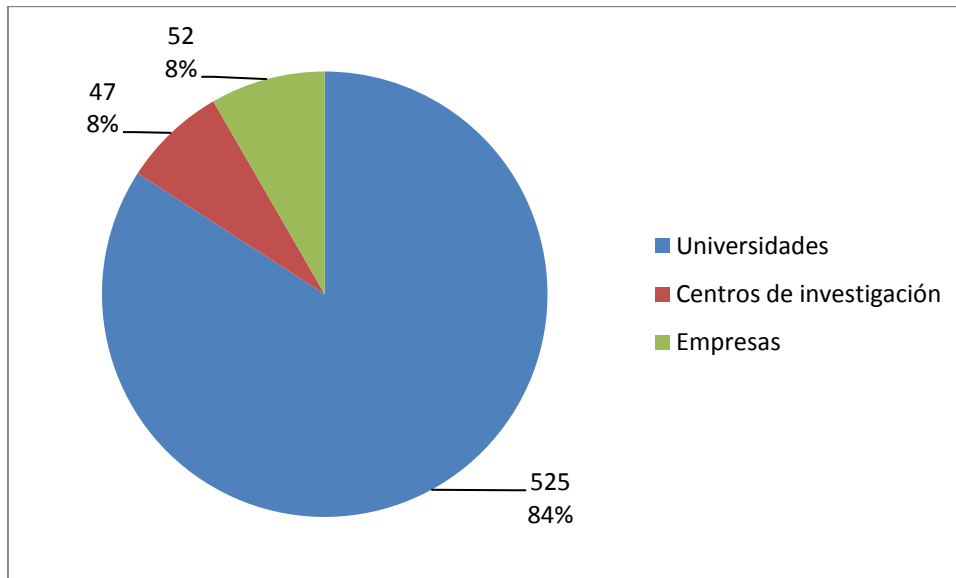


Figura 15: Distribución de las instituciones presentes en la muestra según tipo de entidad.

La red de colaboración institucional (figura 16) está conformada por un total de 151 instituciones, tomándose en cuenta solamente aquellas que presentaron una productividad mayor que tres contribuciones.

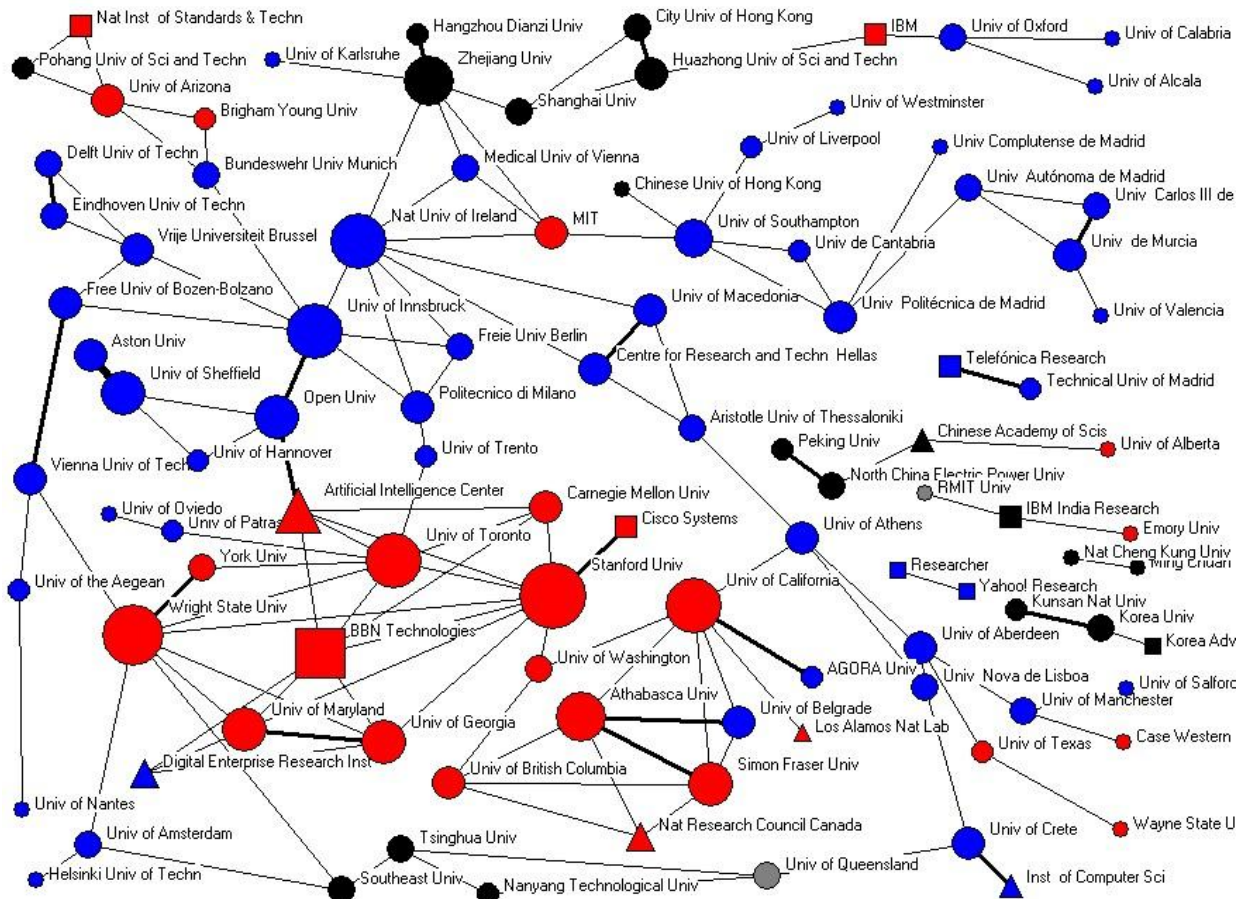


Figura 16: Red de colaboración entre autores, productividad ≥ 3 .

En la generación de la red se utilizaron los atributos de: grado de centralidad nodal para el volumen de los nodos, grosor de las líneas de conexión para representar la intensidad de las relaciones entre los actores, forma del nodo a partir de la tipología de institución que represente: círculo, universidad; triángulo, centro de investigación y un cuadrado para las empresas; para representar el área geográfica en la que se encuentra enclavada la institución se utilizaron los colores: rojo, América del Norte; azul, Europa; negro, Asia y gris para Australia y Oceanía. Se generaron varios grafos de la red a partir de diferentes podas que permitirán un mejor análisis de las visualizaciones. Además se eliminaron los nodos que no presentaban conexiones.

En la figura 16, se muestran todos los actores de la red, identificándose un total de siete subredes o clúster, seis de ellos muy pequeños atendiendo al número de nodos presentes que oscilan entre redes de tamaño dos a cuatro con predominio de entidades asiáticas.

La red más importante la conforman un total de 81 actores, donde predominan las instituciones europeas en primer lugar y norteamericanas en segundo, en esta red se establece relaciones con nodos de áreas poco representadas en la red como son las instituciones de Australia y Oceanía y Asia, que aunque si bien son poco representadas y de baja centralidad, sirven para conectar a universidades localizadas en una misma región.

En el grafo de la figura 17 se visualizan únicamente entidades de Europa y Norteamérica, al sustraer las entidades asiáticas dejan de estar conectados en el grafo principal, por no tener relaciones directas con sus homólogos regionales.

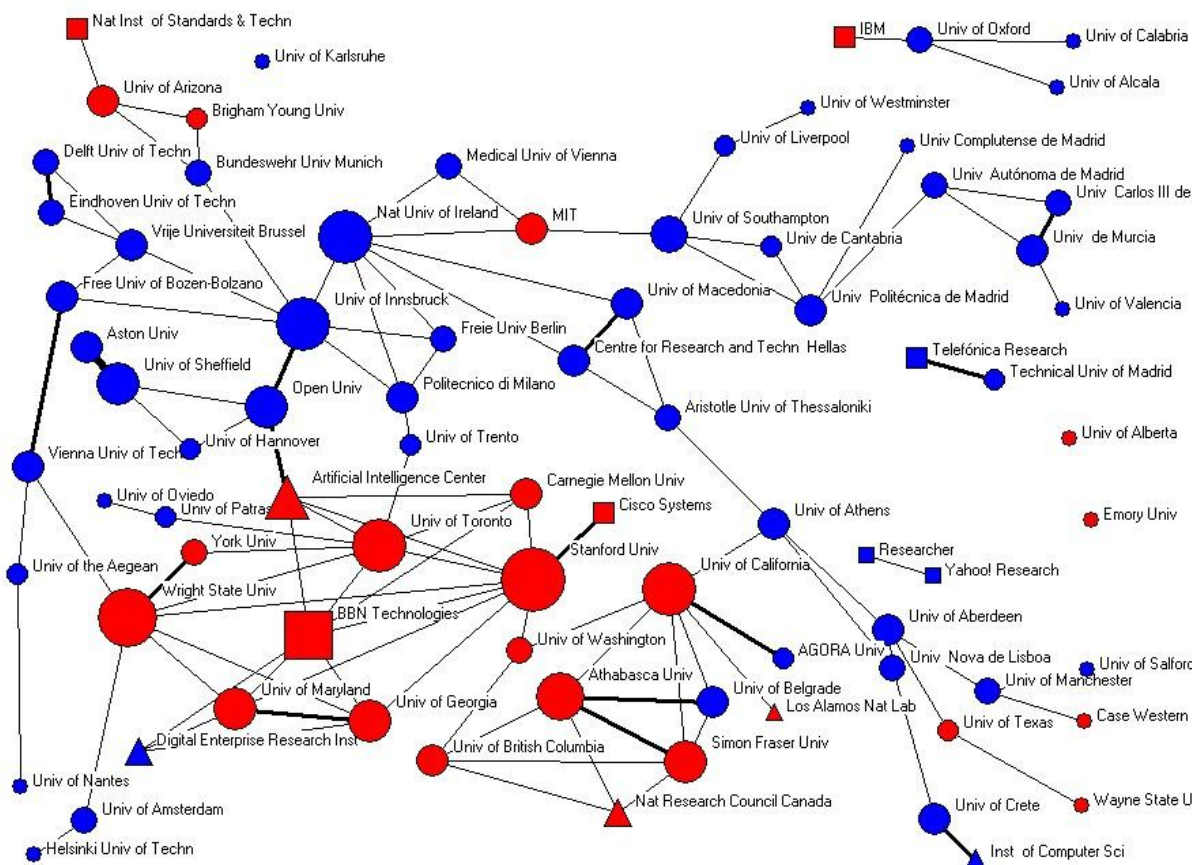


Figura 17: Red de colaboración entre autores norteamericanos y europeos.

El grafo muestra que aunque existen relaciones de colaboración entre entidades de ambas regiones geográficas estas son débiles y a través de nodos de baja centralidad y no de forma homogénea, el enlace más importante ocurre entre la *Open University* de Bélgica y el *Artificial Intelligence Center* de Estados Unidos y se pueden mencionar también los de *Agora University*, de Grecia con la estadounidense *University of California* y el de *Athabasca University*, estadounidense, con la yugoslava *University of Belgrade*.

El grueso de las relaciones se establece entre instituciones de una misma región, destacándose por las europeas los enlaces entre la *Aston University* y la *Sheffield University*, ambas del Reino Unido, las establecidas entre la *Vienna University of Technology*, de Austria, y la italiana *Free University of Bozen-Bolzano*; por las norteamericanas la *Stanford University* y la empresa *Cisco Systems*, de Estados Unidos; la *Athabasca University* y la *Simon Fraser University*, ambas norteamericanas así como la *University of Maryland* y *Universidad of Georgia*, igual de Estados Unidos.

Se resalta además que a pesar de Europa contener el mayor número de entidades productoras en la materia con respecto a Norteamérica, cuenta con solo dos instituciones con alto grado de centralidad, siendo superadas por las *Wright State University*, *University of Toronto*, *BBN Technologies*, *Stanford University* y *University of California*.

Un dato importante del análisis de la producción institucional de Web Semántica en el período estudiado es la ausencia de instituciones latinoamericanas tanto en los grafos como en las tablas de instituciones más productivas. Según los resultados la presencia de las instituciones latinoamericanas es baja identificándose solo doce entidades, conformando el 12% del total de las entidades, todas pertenecientes a la categoría de universidades y que no realizaron más de una publicación en el período estudiado (tabla 5).

Tabla 5. Presencia de instituciones latinoamericanas en la producción de Web Semántica dentro de CASC en el período 2005-2009.

Universidad	Artículos	País
Tecnológico de Monterrey	1	México
Universidad Federal Juiz de Fora	1	Brasil
Universidad de Sao Paulo	1	Brasil
Universidad de Fortaleza	1	Brasil
Universidad Nacional del Sur	1	Argentina
Consejo Nacional de Investigaciones Científicas y Técnicas	1	Argentina
Universidad de la Sierra del Sur	1	México
Universidad Tecnológica de Panamá	1	Panamá
Universidad de Chile	1	Chile
Pontificia Universidad Católica de Chile	1	Chile
Universidad Autónoma de Bucaramanga	1	Colombia
Universidad Simón Bolívar	1	Venezuela

3.6 Producción y distribución por países

En el análisis de los registros recuperados se detectaron un total de 61 países productores, en la figura 18 se observa la distribución de artículos publicados por países, donde 24 de ellos, para un 64%, poseen más de cinco artículos en el período analizado y 18, 29% del total de países, contabilizan una sola contribución.

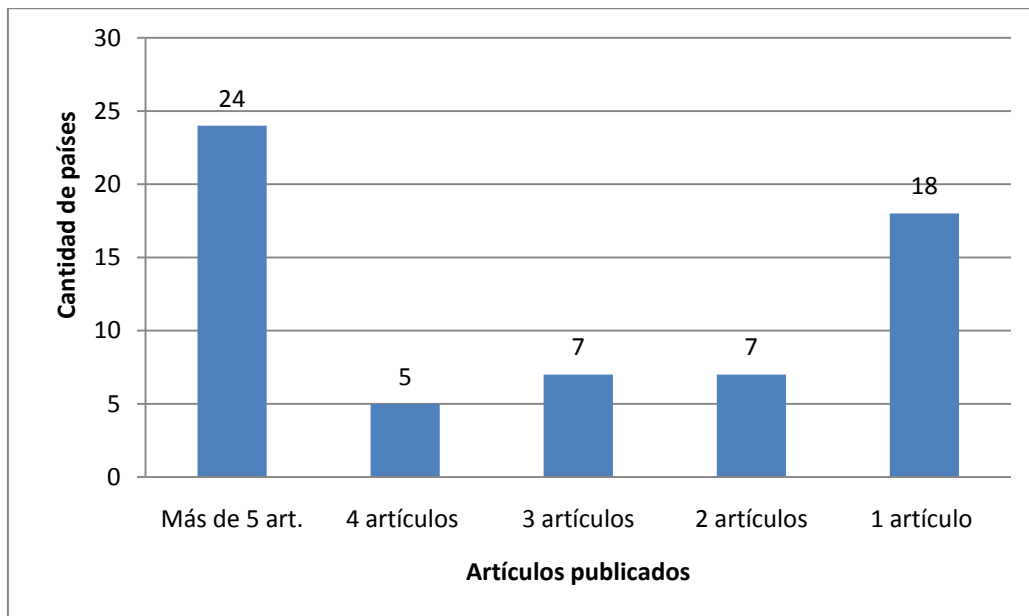


Figura 18: Distribución de la producción de artículos por países en la temática de Web Semántica disponible en CASC en el período 2005-2009.

Se consideraron como países más productivos todos aquellos que presentaron una producción de más de 5 artículos en el período analizado, contabilizando 24 naciones, las producciones nacionales de estos países (figura 19) confirman el papel preponderante de Estados Unidos como líder en las temáticas de Ciencia de la Computación; el tercer cupo de China confirma el aumento de la producción científica china, fenómeno que ha sido estudiado por varios autores. Entre los países más productivos llama la atención la presencia de naciones con baja presencia en los canales tradicionales de comunicación científica en Ciencia de la Computación como Rumania y Turquía, y resulta interesante el cuarto lugar ocupado por España, que podría servir como elemento de retroalimentación para el establecimiento de proyectos de colaboración a partir de las similitudes culturales e idiomáticas que presenta con Cuba.

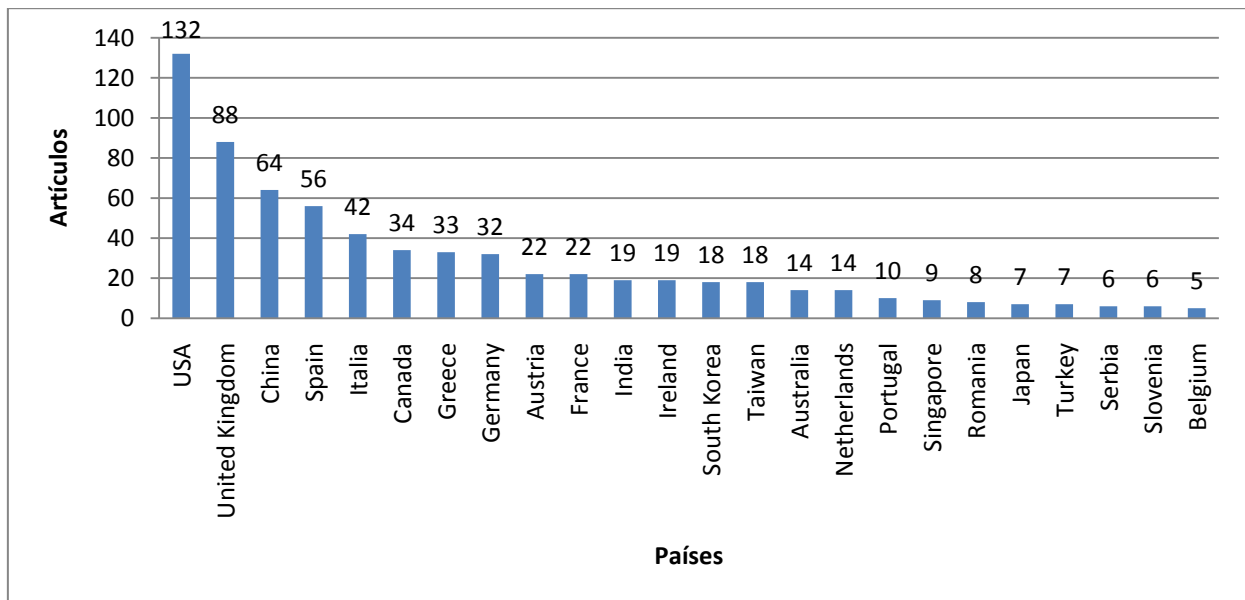


Figura 19: Niveles de producción nacional, países más productivos en la publicación de artículos en Web Semántica disponibles en CASC en el período 2005-2009.

Al igual que en el análisis de la producción institucional una de las conclusiones a las que se puede llegar es la ausencia de países latinoamericanos en la muestra. Se pudo comprobar la existencia de producciones latinoamericanas (figura 20), en su mayoría en el grupo de países con bajos niveles de producción: Argentina, Chile, Colombia, Panamá y Venezuela y los de mayor producción en el grupo de los de productividad media: Brasil y México con tres y dos comunicaciones respectivamente.

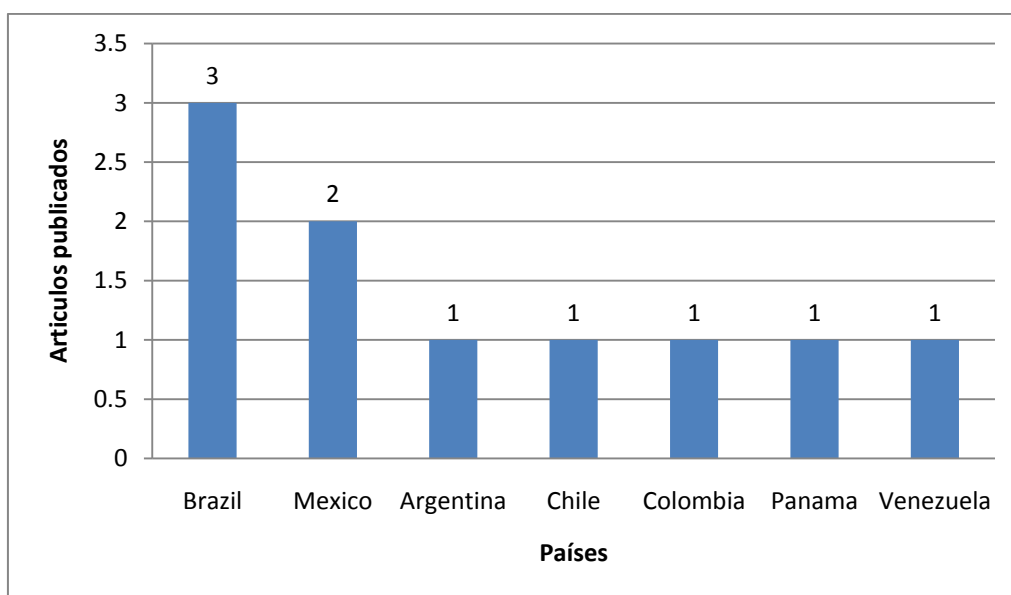


Figura 20: Producción nacional, países latinoamericanos presentes en la muestra.

La colaboración entre los países presentes en la muestra se observa a través del grafo de la figura 21, al igual que en las redes anteriores se ha asignado la centralidad al volumen de los nodos, y el grosor de las líneas de enlace representa la intensidad de las relaciones entre los actores, según el área regional donde se encuentre el nodo se ha asignado un color: azul, Europa; rojo, América del Norte, negro, Asia; gris, Australia y Oceanía; magenta, África y verde para América Latina.

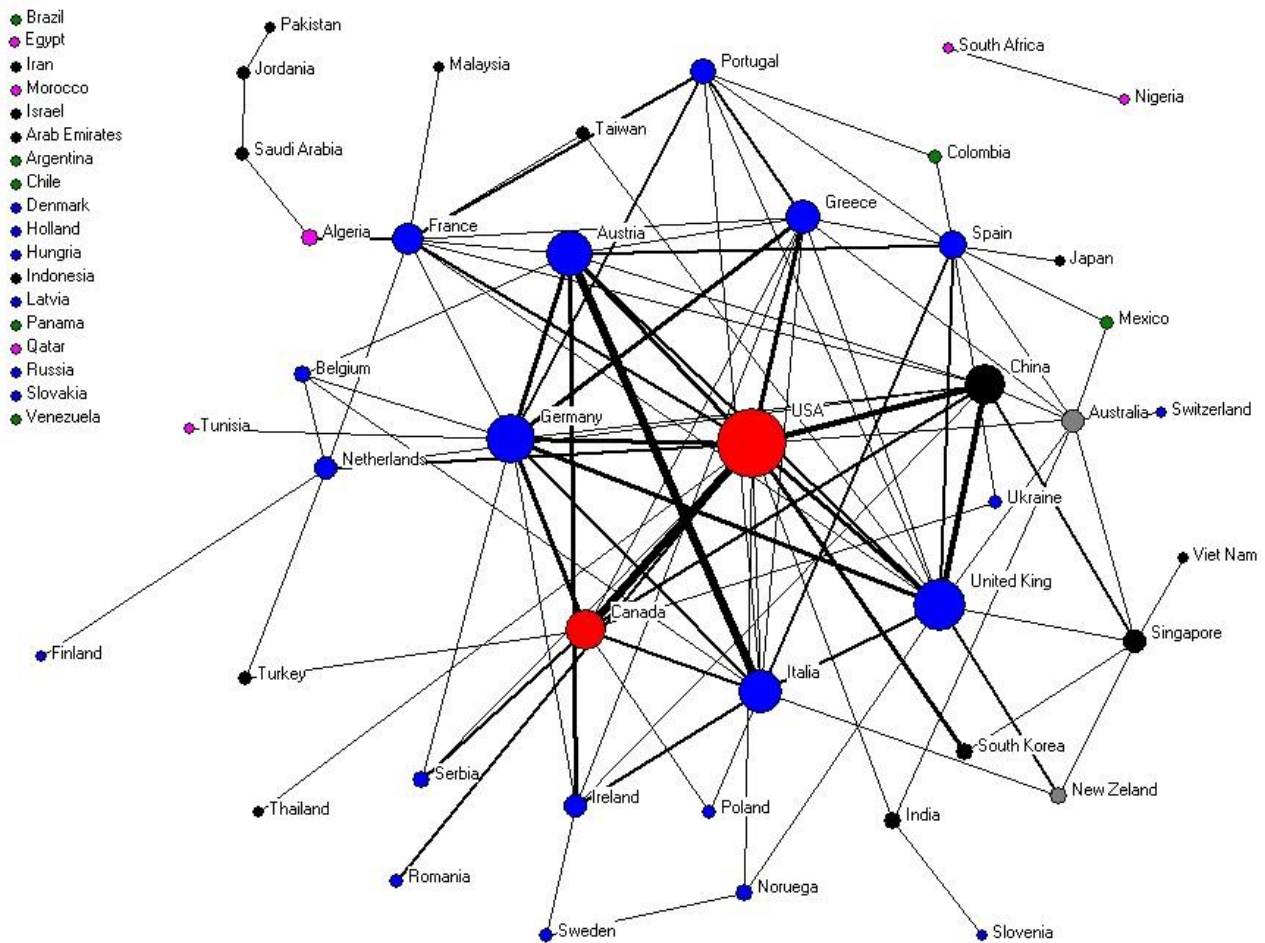


Figura 21: Red de colaboración entre países presentes en la muestra de producción en Web Semántica disponible en CASC en el período 2005-2009.

El grafo muestra la presencia de dos redes y un conjunto de nodos desconectados, 18 en total, que no establecen relaciones de colaboración en el área de Web Semántica. Una de las redes independientes está conformada por dos nodos de países africanos, con bajos niveles de colaboración entre sí, y la segunda red, con un nivel de densidad de 41 nodos es la que muestra con mayor énfasis las relaciones de colaboración internacional en la temática. Las relaciones de colaboración más fuertes se establecen entre Estados Unidos y Canadá y entre Austria e Italia, en ambos casos llama la atención que ocurre entre países ubicados en las mismas áreas geográficas.

El nodo con mayor nivel de centralidad representa a los Estados Unidos, país desde el cual se establecen un mayor número de colaboraciones, al mismo tiempo es el nodo que mayor heterogeneidad de relaciones presenta atendiendo a la diversidad de regiones geográficas con las cuales establece relaciones de colaboración, los enlaces más fuertes los establece con Canadá, ubicado en la misma región geográfica y con China, aunque no presenta relaciones con todos los nodos de países europeos, sí las mantiene con los de mayor centralidad dentro de esa área geográfica, llama la atención que no presente relaciones con España o Irlanda identificados en la relación de países con mayor producción.

Se observa que los países latinoamericanos se encuentran desconectados, menos Colombia y México con relaciones muy débiles con España, Portugal y Austria. Ninguno de los países latinoamericanos estableció relaciones de colaboración entre sí.

3.7 Producción temática a través del análisis de palabras clave

Por medio del conjunto de descriptores y palabras clave utilizadas en el procesamiento y representación de los contenidos indexados en CASC se procedió a determinar la producción temática a partir de la producción recuperada.

El análisis arrojó la utilización de 1705 términos en la descripción de los contenidos recuperados de ello, de estos 1257 se utilizaron solo una vez para la descripción de un artículo. Los diez términos más utilizados (figura 22) representan solo el 0.5% del total de los descriptores recuperados.

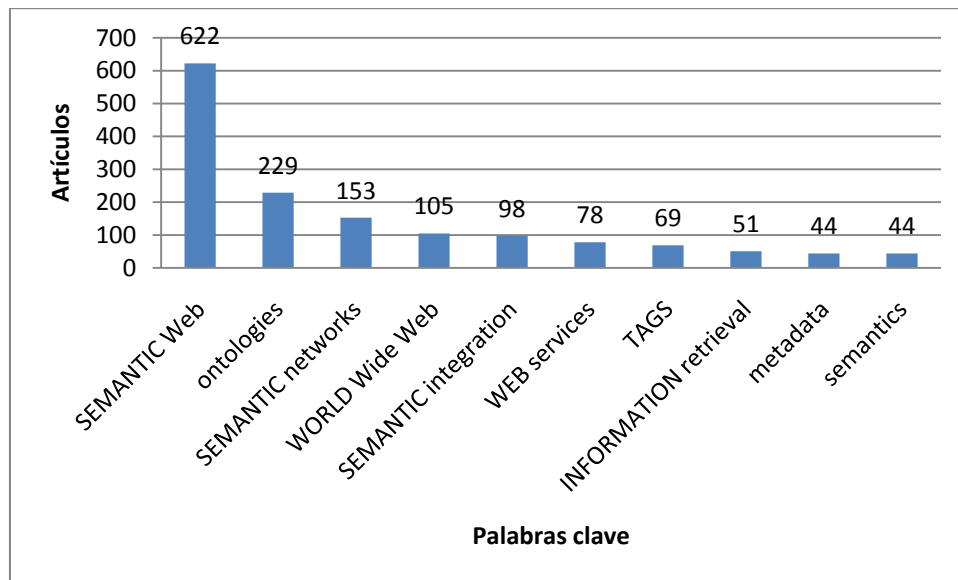


Figura 22: Palabras clave más utilizadas en la muestra de artículos sobre Web Semántica disponibles en CASC en el período 2005-2009.

El término más representado es “Semantic web” debido a que se utilizó como criterio de búsqueda en el proceso de recuperación de información en la base de datos, el resto de los términos guarda relación estrecha con la temática que analiza el artículo, como se pudo comprobar en la realización del estado del arte en Web Semántica en el capítulo 1.

Coincidentemente en el grafo generado para el análisis de co-ocurrencia de las palabras clave se puede confirmar la presencia de estos términos atendiendo al nivel de centralidad de los nodos que los representan (figura 23).

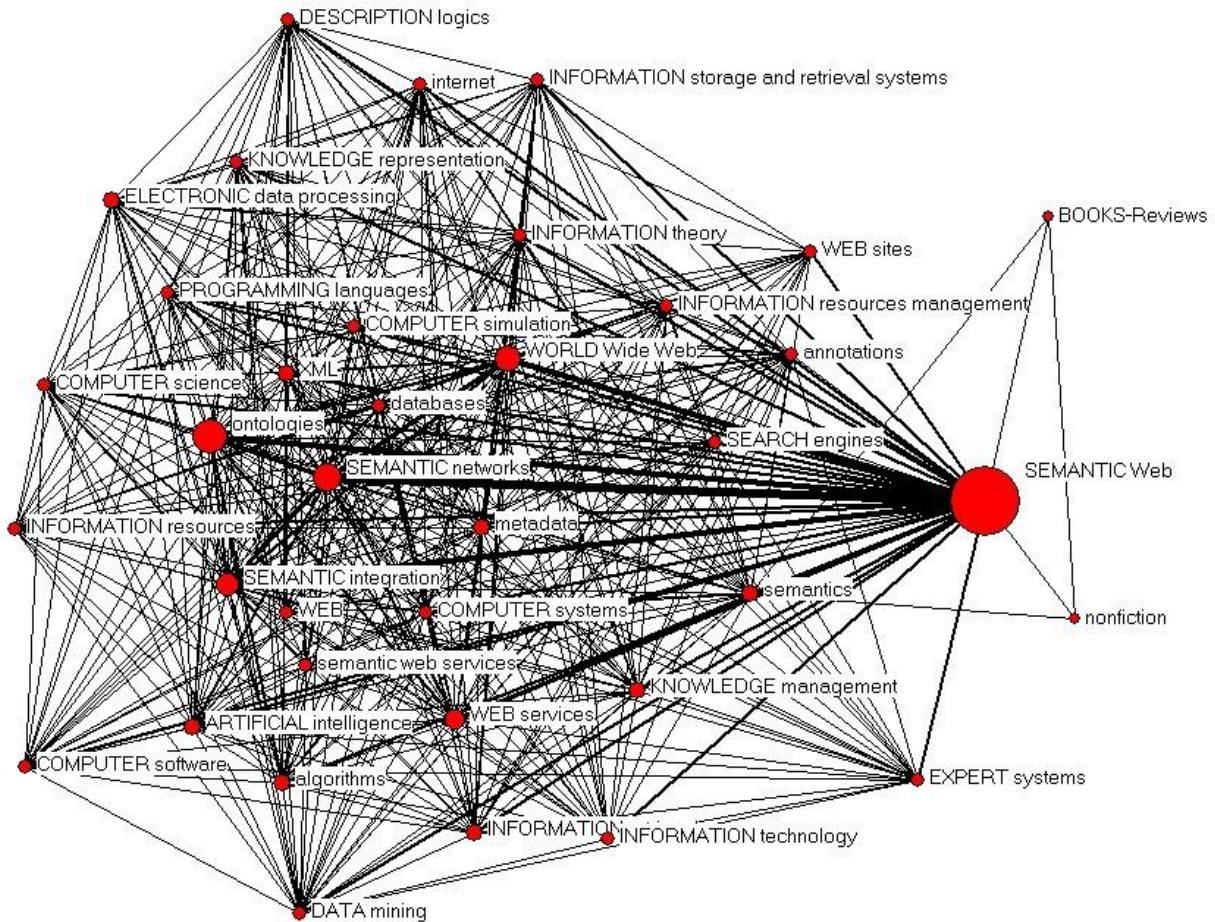


Figura 23: Red de co-ocurrencia de palabras clave ≥ 50 .

La presencia del descriptor “*Ontologies*” con 229 ocurrencias, es el segundo tema más abordado dentro de las investigaciones en Web Semántica como se visualiza claramente en el grafo, debido a que constituyen el pilar fundamental para la implementación de sistemas semánticos sobre los que se construye la Web Semántica. Presenta fuertes relaciones de co-ocurrencia con los principales nodos (“*Semantic Web*” y *Semantic networks*).

“*Semantic network*” aparece en la descripción de 153 documentos, agrupa el conjunto de investigaciones relacionadas con el desarrollo de las redes en Web Semántica, presentando también una amplia co-ocurrencia con la gran mayoría de los nodos mostrados en el grafo (Figura 14). Se comprueba así su importancia como término de referencia dentro de la temática.

El término “*World Wide Web*” presente en 105 documentos, hace referencia a los antecedentes y campo de acción de las técnicas en Web Semántica, según las relaciones que establece en la red se destacan los enlaces hacia “*ontologies*”, “*information resources management*”, “*information theory*”, “*Semantic networks*” y “*Semantic web*”.

“*Semantic integration*” con 98 ocurrencias, menciona las aplicaciones y soluciones para la integración de tecnologías semánticas, se interrelaciona más fuertemente con “*Semantic networks*”, “*ontologies*”, “*World Wide Web*” y “*Semantic web*”.

“*Web services*”, presente en 78 artículos, destaca el abordaje, diseño y conceptualización de servicios web, aparece más frecuentemente acompañado de los términos “*semantics*”, “*algorithms*”, “*computer systems*” y “*semantic web*”.

“*Tags*” co-ocurrente en 69 documentos, agrupa al conjunto de contribuciones que utiliza el conocimiento social por medio de las palabras clave y descriptores para la representación social de los contenidos y su interrelación con el desarrollo de ontologías, no es visible en la red, pues su grado de centralidad es menor de 50 relaciones.

“*Information retrieval*”, presente en 51 trabajos, describe la aplicación de soluciones de Web Semántica para solucionar u optimizar procesos de búsqueda y recuperación de información, no es visible en la red por presentar un grado de centralidad menor a 50.

“*Metadata*”, presente en 44 documentos, se refiere a la utilización de los metadatos como elementos para la elaboración de soluciones de Web Semántica, las relaciones más fuertes las establece con “*World Wide Web*” y “*Web services*”.

Conclusiones

- La producción científica correspondiente a la temática de Web Semántica en la base de datos bibliográfica *Computers & Applied Sciences Complete* (CASC) en el período 2005-2009 la conforman un total de 622 artículos.
- La producción de artículos científicos en Web Semántica presentes en CASC, mostró un crecimiento anual constante en los años comprendidos por el estudio. Las mayores tasas de crecimiento con respecto a años anteriores le correspondieron a los bienios 2005/2006 y 2008/2009. El crecimiento total con respecto al primer año de estudio fue de 148 artículos para un 606%.
- A partir de las 192 publicaciones seriadas identificadas como pertinentes para la temática de Web Semántica se logró definir la zona núcleo de revistas más productivas, conformada por nueve títulos: *IEEE Intelligent Systems*, *Expert Systems with Applications*, *IEEE Internet Computing*, *IEEE Transactions on Knowledge & Data Engineering*, *Journal of Software, Theory & Practice of Logic Programming*, *Knowledge-Based Systems*, *Information Processing & Management*, *Concurrency & Computation: Practice & Experience*.
- Fueron identificados un total de 1486 autores, de los cuales el 89%, 1317, solo publicaron un artículo en los cinco años de análisis. Los autores considerados como más productivos en la muestra fueron: Sheth, con 9 artículos; Hendler, con 7; Wilks, 6; Petrie, 5 y Decker con 5 trabajos. La forma de publicación predominante es en colaboración, aunque la publicación en formato individual no es despreciable. Aunque se identificaron catorce nodos de colaboración autoral los mismos están conformados por pocos actores. Las relaciones de colaboración autoral más fuertes se establecieron entre Brewster y Wilks; Goble y Stevenson.
- Se identificaron un total de 624 instituciones, donde predominó la universidad como tipo de entidad, aunque también estuvieron presentes las formas de centros de investigación y empresas. El núcleo de instituciones más productivas lo conforman una relación de 29 entidades entre las que se encuentra IBM como

única institución no universitaria, la *University of California*, de Estados Unidos, con 14 artículos, resultó ser la institución más productiva, seguida por la *Open University* de Bélgica, con 12 trabajos y la *Wright State University*, estadounidense, con 11 contribuciones.

- La colaboración institucional en la temática de Web Semántica está dominada por entidades europeas y norteamericanas. Son más frecuentes las relaciones de colaboración entre organizaciones de la misma región geográfica. La presencia de entidades latinoamericanas es pequeña, y se localizan en la zona de baja productividad institucional. Las relaciones de las instituciones latinoamericanas son débiles y de muy baja centralidad.
- En la muestra estudiada fue posible constatar la presencia de 61 países con producción de artículos en Web Semántica. El núcleo de países más productivos lo conformaron 24 naciones, donde ocupó el primer lugar Estados Unidos con 132 artículos; Reino Unido, con 88, y China en tercero con 64 trabajos. Dentro del núcleo de mayores productores nacionales no se ubicó ningún país latinoamericano. Se logró identificar la presencia de siete países latinoamericanos, resultando los de mayor productividad Brasil, con tres contribuciones y México con dos, el resto: Argentina, Chile, Colombia, Panamá y Venezuela, solo lograron una publicación en los cinco años de estudio.
- Los países con mayores relaciones de colaboración en la investigación de Web Semántica por medio de la publicación de artículos son: Estados Unidos, Reino Unido, Alemania y Austria. De los países latinoamericanos presentes en la red de colaboración, solo Panamá y México establecieron relaciones con otros países, no obstante sus niveles de centralidad nodal y la intensidad de sus relaciones es baja.
- El análisis de los descriptores y palabras clave permitió identificar los diez términos más utilizados, indicando de este modo la especialización temática en la producción científica en Web Semántica, estas son: ontologías, redes semánticas, *world wide web*, *web services*, *tags*, *information retrieval*, *metadata* y *semantics*.

Recomendaciones

- Divulgar los resultados de investigación dentro de la comunidad de investigadores que trabajan los temas relacionados con Web Semántica.
- Actualizar los resultados de la investigación por medio del monitoreo constante de las revistas identificadas dentro de la zona núcleo de publicaciones más productivas.
- Desarrollar estudios similares en otras líneas de investigación en la facultad.
- Confrontar los resultados del estudio con otros servicios de indización y resúmenes especializados en Ciencia de la Computación.
- Buscar alternativas dentro del movimiento de Software libre para la sustitución de las herramientas de análisis utilizadas.

Bibliografía

- ALBARELLO, F. Historia de Internet. 2005. [Consultado el: 19/02/2011]. Disponible en: http://www.newsmatic.e-pol.com.ar/index.php?Accion=VerArticulo&NombreSeccion=UNIDAD%202&aid=12972&pub_id=147.
- ALÍ HERRERA, A. Usando Bibexcel con datos en EndNote *ACIMED*, 2009, vol. 20, nº 5, p. 75-79.
- ALLEN, M.; JACOBS, S. K., *et al.* Mapping the literature of nursing: 1996-2000. *Journal of the Medical Library Association (JMLA)*, April 2006, vol. 94, nº 2, p. np-np.
- ANTONIOU, G. y VAN HARMELEN, F. *A Semantic Web Primer, (Cooperative Information Systems)*. 2008. ISBN 0262012421.
- ARCHAMBAULT, É.; CAMPBELL, D., *et al.* Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus. *Arxiv preprint arXiv:0903.5254*, 2009, nº
- ARCHAMBAULT, É. y LARIVIÈRE, V. History of the journal impact factor: Contingencies and consequences. *Scientometrics*, 2009, nº Disponible en: <http://dx.doi.org/10.1007/s11192-007-2036-x>.
- ARRUDA, D.; BEZERRA, F., *et al.* Brazilian computer science research: Gender and regional distributions. *Scientometrics*, 2006, vol. 67, nº 1, p. 67-86. Disponible en: <http://dx.doi.org/10.1007/s11192-007-1944-0>.
- ÅSTRÖM, F. y SCHLEMMER, B. *Celebrating Scholarly Communication Studies. A Festschrift for Olle Persson at his 60th Birthday*. International Society for Scientometrics and Informetrics. ISSI, 2009, vol. 5 (Special volumen),
- AYERS, D. Delivered Deliverables. *IEEE Computer Society*, 2009, nº p. 4.
- AZDEN, B. W. The Computer Science and Engineering Research Study. *ACM Communications*, 1976, vol. 19, nº 12, p. 670-673.
- BALLARD, S. y HENRY, M. Citation searching: new players, new tools. *Searcher*, October 2006, vol. 14, nº 9, p. 24-33. Disponible en: <http://www.infotoday.com>
- BARNHILL, R. y LINTON, M. *Promoting research in science and engineering departments: A chair's perspective*. En Haden, C. R. y Brink, J. R. (editor). *Innovative models for university research*. New York: North Holland: 1992,
- BERGMAN, M. The deep web: Surfacing hidden value. *Journal of Electronic Publishing*, 2001, vol. 7, nº 1, p. 07-01.
- BERNERS-LEE, T.; HENDLER, J., *et al.* The semantic web. *Scientific American*, 2001, vol. 284, nº 5, p. 34-43.

- BJORNEBORN, L. 'Mini small worlds' of shortest link paths crossing domain boundaries in an academic Web space. *Scientometrics*, 2006, vol. 68, nº 3, p. 395-414. ISSN 0138-9130.
- BRAVO, S. y REYTOR, M. *Módulo de Inferencia para un Repositorio Semántico* FACULTAD 10 Universidad de las Ciencias Informáticas 2009.
- BRIAN, M. A Protocol for Exchanging Scientific Citations. En 2009. p. 171-177.
- CARVIN, A. Tim Berners-Lee: Weaving a Semantic Web 2005, nº Disponible en: http://faculty.cbpp.uaa.alaska.edu/afef/weaving%20the%20web-tim_bernierslee.htm.
- CASTELLS, P. La Web Semántica *Universidad Autónoma de Madrid*, 2003, nº
- CASTRO, S. y LARRAUD, M. *Recuperación de Información Bilingüe en la Web Semántica*. Facultad de Ingeniería. Instituto de Computación, 2007.
- CAVENDER, S. *Legends* [Consultado el: 22/04/2011]. Disponible en: <http://www.forbes.com/asap/1998/1005/126.html>.
- CODINA, L. y ROVIRA, C. La web semántica. *Tendencias en documentación digital*, Gijón, Trea, 2006, nº p. 9-54.
- COLINA, C. Bases socio-metodológicas para el Análisis de Redes Sociales, ARS. *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 2005, nº p. 9-35. Disponible en: http://dialnet.unirioja.es/servlet/dfichero_articulo?codigo=1374915&orden=72939.
- CUEVA, S. P. WEB SEMANTICA: Principios y Estándares. 2008, nº
- CULNAN, M. J. Analysis of the information usage patterns of academics and practitioner in the computer field: a citation analysis of national conference proceedings. *Information Processing & Management*, 1978, vol. 14, nº 6, p. 395-404. Disponible en: [http://dx.doi.org/10.1016/0306-4573\(78\)90004-3](http://dx.doi.org/10.1016/0306-4573(78)90004-3)
- CHAO-CHIH, H. The Use of Social Network Analysis in Knowledge Diffusion Research from Patent Data. En 2009. p. 393-398.
- CHAVIANO, O. G. Aplicaciones y perspectivas de los estudios métricos de la información (emi) en la gestión de información y el conocimiento en las organizaciones. *Revista AIBDA*, 2008, vol. 29, nº p. 1-2.
- DING, Y.; FENSEL, D., *et al.* The semantic web: yet another hip? *Data & Knowledge Engineering*, 2002, nº p. 1-23.
- FALAGAS, M. E.; PITSOUNI, E. I., *et al.* Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, 2008, vol. 22, nº 2, p. 338-342.
- FERNÁNDEZ, L. S. y GARCÍA, N. F. La Web Semántica: fundamentos y breve "estado del arte". *Asociación de Técnicos de Informática (NOVÁTICA)*, 2005, nº

- FRANCESCHET, M. A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics (on line firsts)*, 2009, n^o June,
- GARFIELD, E. The 100 most-cited books in the CompuMath Citation Index, 1976-1980. *Essays of an information scientist*, 1984a, vol. 7 n^o p. 264-269.
- GARFIELD, E. Highly cited works in mathematics. Part 2. *Essays of an information scientist*, 1973, vol. 1, n^o p. 509-513. Disponible en: www.garfield.library.upenn.edu/essays/V1p509y1962-73.pdf.
- GARFIELD, E. The multidisciplinary impact of math and computer science in the most-cited articles in CompuMath Citation Index, 1976-1980. *Essays of an information scientist*, 1984b, vol. 7, n^o p. 232-239.
- GOODRUM, A. A.; MCCAIN, K. W., et al. Scholarly publishing in the Internet age: a citation analysis of computer science literature. *Information Processing and Management*, Sep 2001, vol. 37, n^o 5, p. 661-675. Disponible en: <http://clgiles.ist.psu.edu/papers/IPM-2001-scholarly-publishing.pdf>.
- GORBEA PORTAL, S. *Modelación matemática de la actividad bibliotecaria: una revisión*. En Martínez Rodríguez, A. (editor). *Selección de lecturas de estudios metricos de la información*. La Habana: Editorial Félix Varela, 2004, p. 3-21.
- HANNEMAN, R. A. *Introducción a los métodos del análisis de redes sociales*. Revista REDES, 2000.
- HIRST, G. y N., T. Computers science journals - an iterated citation analysis. *IEEE Transactions on Professional Communications*, 1977, vol. 20, n^o 4, p. 233-238.
- HJORLAND, B. y ALBRECHTSEN, H. Towards a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, Jul 1995, vol. 46, n^o 6, p. 400-425.
- JACSÓ, P. *As we may search: Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citationenhanced databases*. Current Science. publicado el: 28 de abril de 2009 de 2005, última actualización: 28 de abril de 2009. vol. 89, 1537-1547 p. Disponible en: <http://www.ias.ac.in/currsci/nov102005/1537.pdf>.
- JANEV, V. y VRANES, S. *Semantic Web Technologies: Ready for Adoption?* The IEEE Computer Society 2009,
- JIMÉNEZ, G. M. C. *“Tecnologías de WEB Semántica orientada al desarrollo de servicios WEB”* Ciencias de la Computación Universidad Tecnica Particular de LOJA, 2010.
- KATSAROS, D.; MATSOUKAS, V., et al. *Evaluating Greek Departments of Computer Science/Engineering using Bibliometric Indices*. En *Πρακτικά PCI2008*. 2008.
- KLINK, S.; REUTHER, P., et al. Analysing social networks within bibliographical data. *Database and Expert Systems Applications 17th International Conference, DEXA 2006 Proceedings Lecture Notes in Computer Science Vol 4080*, 2006, vol. 4080,

nº p. 234-243. Disponible en: http://dbis.uni-trier.de/Mitarbeiter/reuther_files/private/paper/DEXA06.pdf. ISSN 3540378715.

KUHN, T. S. *La estructura de las revoluciones científicas*. Traducido por: Contín, A. 8^{va} ed. México D.F.: Fondo de Cultura Económica, 2004. Breviarios.

KUMAR, S. y GARG, K. C. Scientometrics of computer science research in India and China. *Scientometrics*, 2005, vol. 64, nº 2, p. 121-132. Disponible en: <http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=20136838&site=ehost-live>

KUNDER, M. D. *The size of the World Wide Web* [Consultado el: 03/12/2010]. Disponible en: <http://www.worldwidewebsite.com/>.

LABRADA, O. L. Web Semántica y Ontologías, desafíos para el Profesional de la Información en el siglo XXI. 2003, nº

LEYDESDORFF, L.; MOYA-ANEGÓN, F. D., *et al.* Journal Maps on the Basis of Scopus Data: A Comparison With the Journal Citation Reports of the ISI. *ASIS&T*, 2009, nº

LIANG, L.; KRETSCHMER, H., *et al.* Age structures of scientific collaboration in Chinese computer science. *Scientometrics*, 2001, vol. 52, nº p. 471-486. Disponible en: <http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=ISTA3703758&site=ehost-live> ISSN 01389130.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. *Ci. Inf*, 1998, vol. 27, nº 2, p. 134-140.

MARION, L. S. y MCCAIN, K. W. Contrasting views of software engineering journals: Authors cocitation choices and indexer vocabulary assignments. *Journal of the American Society for Information Science and Technology*, 2001, vol. 52, nº 4, p. 297–308.

MARTÍNEZ, I. *Que es Internet y como surgió* [Consultado el: 02/12/2010]. Disponible en: <http://indira-informatica.blogspot.com/>.

MEHO, L. y ROGERS, Y. Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 2008, vol. 59, nº 11, p. 1711-1726.

MEHO, L. I. y YANG, K. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 2007, vol. 58, nº 13, p. 2105–2125.

MELO, M. A. y BOTIA, I. A. De la web actual a la web semántica. 2008, nº

MINGERS, J. y LIPITAKIS, E. Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics*, 2010, nº 85, p. 613–625.

- MOLINA, J. L. *Operaciones básicas con Ucinet 6*. 2005. 12 p.
- MOYA-ANEGÓN, F. D.; CHINCHILLA-RODRÍGUEZ, Z., *et al.* Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 2007, vol. 73, nº 1, p. 53-78.
- NASCIMENTO, D. y MARTELETO, R. Social field, domains of knowledge and informational practice. *Journal of Documentation*, 2008, vol. 64, nº 3, p. 397-412.
- PATALANO, M. Las publicaciones del campo científico: las revistas académicas de América Latina. *ANALES DE DOCUMENTACION*, 2005, nº 8, p. 217-235. Disponible en: <http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=17783081&site=ehost-live>
- PÉREZ, A. G. *A Survey on Ontology Tools, OntoWebOntology-based information exchange for knowledge management and electronic commerce*. 2002, Disponible en: http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/OntoWeb_Del_1-3.pdf
- PEROJO, K. R. Web Semántica: un nuevo enfoque hacia la Organización de Información en los Sistemas de Gestión de Contenidos. 2006, nº
- PETRICEK, V.; COX, I. J., *et al.* A comparison of on-line computer science citation databases. En *Vienna, Austria. 2005*. p. 438-449.
- PILKINGTON, A. *Bibexcel – Quick Start Guide to Bibliometrics and Citation Analysis*. 2006,
- PISLYAKOV, V. Comparing two “thermometers”: Impact factors of 20 leading economic journals according to Journal Citation Reports and Scopus. *Scientometrics*, 2009, vol. 79, nº 3, Disponible en: <http://dx.doi.org/10.1007/s11192-007-2016-1>.
- PORTAL, G. Principios teóricos y metodológicos de los estudios métricos de la información. *INVESTIGACIÓN BIBLIOTECOLÓGICA*, 2009, vol. 9, nº 017, ISSN 0187-358X.
- PRITCHARD, A. Statistical bibliography or bibliometrics. *Journal of Documentation*, 1969, vol. 25, nº p. 348.
- RAE. *Diccionario de la Real Academia de la Lengua Española*. Diccionario de la Real Academia de la Lengua Española. 22ª ed. Madrid: 2001,
- RAMÍREZ, Z. Web semántica: una alternativa para poner orden al caos. . 2004, nº
- RESEARCH INFORMATION NETWORK y
JOINT COMMITTEE ON QUANTITATIVE ASSESSMENT OF RESEARCH.
Communicating knowledge: How and why researchers publish and disseminate their findings. Bibliometric analysis Publication, dissemination & citation behaviour of researchers Londres: Research Information Network, 2009.
- ROCHE RODRÍGUEZ, D. y MARTÍ LAHERA, Y. *La cultura organizacional en el campo de la Bibliotecología y la Ciencia de la Información Una aproximación a su estudio*

desde un enfoque métrico *Web of Science*, 1991–2006. Trabajo de Diploma en opción al título de Licenciado en Bibliotecología y Ciencia de la Información, Departamento de Bibliotecología y Ciencia de la Información de la Facultad de Comunicación. Universidad de La Habana, 2008.

SCIMAGO GRUPO. *SJR — SCImago Journal & Country Rank*. [Página web]. Granada: 2007, vol. 2011, Disponible en: <http://www.scimagojr.com>.

SEGAL, B. A Short History of Internet Protocols at CERN 1995, nº

SETIÉN, E. *Conceptos métricos en las disciplinas bibliotecario-informativas*. En Martínez, A. (editor). *Selección de lecturas de estudios métricos de la información*. La Habana: Editorial Félix Varela, 2004, p. 22-32.

SHOTTON, D. Semantic publishing: The coming revolution in scientific journal publishing. *LEARNED PUBLISHING*, 2009, nº p. 1-10.

SIGNORINI, A. *The Indexable Web is More than 11.5 billion pages*. 2005,

SPINAK, E. *Diccionario Enciclopédico de Bibliometría, Cienciometría e Informetría*. La Habana: Editorial Félix Varela, 2004.

SPINAK, E. Indicadores cientiométricos. *Ciência da Informação*, 1998, vol. 27, nº p. nd.

SUÁREZ, Y. y PAZ, C. M. *Mecanismos Semiautomáticos y Automáticos para la Generación de Ontologías de Dominio*. . FACULTAD 10 Universidad de las Ciencias Informáticas, 2008.

SUBRAMANYAM, K. Core journals in computer science. *IEEE Transaction on Professional Communication*, 1976, vol. 19, nº 2, p. 22-25.

SUBRAMANYAM, K. Research productivity and breadth of interest of computer scientists. *Journal of the American Society for Information Science*, 1984, vol. 35, nº 6, p. 369-371. ISSN 0002-8231.

SUBRAMANYAN, K. Lotka's law and literature of computer science. *IEEE Transactions on Professional Communications*, 1979, vol. 22, nº 4, p. 187-189.

THOMSON REUTERS. *Master Journal List : Science Citation Index Expanded*. [Página web]. Pensilvania: 2010, vol. 2011, Disponible en: <http://science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=D>.

TORRES-SALINAS, D. y JIMÉNEZ-CONTRERAS, E. Introducción y estudio comparativo de los nuevos indicadores de citación sobre revistas científicas en Journal Citation Reports y Scopus. *El Profesional de la Información*, 2010, vol. 19, nº 2, p. 201-208.

VARGAS-QUESADA, B.; DE MOYA ANEGÓN, F., et al. Evolución de la estructura científica española: ISI Web of Science 1990-2005. *El Profesional de la Información*, 2007, vol. 17, nº 1, p. 22-37.

W3C, W. W. W. C. *W3C Semantic Web Activity* [página web]. [Consultado el: 10 de noviembre de 2009]. Disponible en: <http://www.w3.org/2001/sw/>.

- W3C, W. W. W. C. *XHTML2 Working Group Home Page* [Consultado el: 10/02/2011]. Disponible en: <http://www.w3.org/MarkUp/>.
- WAGNER-DOBLER, R. Science-technology coupling: the case of mathematical logic and computer science. *Journal of the American Society for Information Science*, 1997, vol. 48, nº 2, p. 171-183. ISSN 0002-8231.
- WIKIPEDIA. *Historia de Internet* [página web]. [Consultado el: 20 de Diciembre de 2010]. Disponible en: http://es.wikipedia.org/wiki/Historia_de_Internet.
- YANG, Y. The Effect of Open Access Journals on Citation Impact: A Citation Analysis of Open Access Journals Using Google Scholar. En 2009. p. 278-280.
- YANG, Z. Exploration and Analysis of the Relations between China and Foreign Countries Based on Knowledge Flows of Patent Citation. En 2010. p. 603-607.
- YINIAN, G. An exploratory study of Malaysian publication productivity in computer science and information technology. *Journal of the American Society for Information Science and Technology*, 2002, vol. 53, nº 12, p. 974-986. Disponible en: file:///E:/Documentos/Doctorado/Análisis/yinian_1.pdf. ISSN 1532-2882.
- YINIAN, G. y ZAINAB, A. N. Publication productivity of Malaysian researchers in the field of computer science and information technology. *Malaysian Journal of Library and Information Science*, 2001, vol. 6, nº 1, p. 1-23. ISSN 1394-6234.
- ZHAO, D. y LOGAN, E. Citation analysis using scientific publications on the Web as data source: A case study in the XML research area. *Scientometrics*, 2002, vol. 53, nº 2, p. 449-472.
- ZINS, C. y GUTTMANN, D. Domain analysis of social work: An example of an integrated methodological approach. *Knowledge Organization*, 2003, vol. 30, nº 3-4, p. 196-212.