

Búsqueda semántica de recursos educativos en repositorios libres

Semantic Search of Educational Resources Repository Free

Recibido: 29 de abril de 2014 • Aceptado: 15 de diciembre de 2014

Para citar este artículo: L. Bonilla y J. Otálora, « Búsqueda semántica de recursos educativos en repositorios libres », Ingenium, vol. 16, n.º 32, pp. 42-56, junio, 2015.



Luis Ernesto Bonilla Orduz*
Jorge Enrique Otálora Luna**

Resumen

En la actualidad existen mecanismos de búsqueda de información en la web que no dan suficiente ilustración, además al realizar una búsqueda de información utilizando los motores de búsqueda conocidos, cargan información siguiendo muchos vínculos, que dan un porcentaje muy bajo de resultados esperados.

El propósito es presentar un modelo que permita la búsqueda y recuperación de la información de recursos educativos de una manera precisa, basado en un proceso de investigación de tipo cuasi experimental, fundamentalmente por las características que la señalan y haciendo uso de la web semántica con aplicaciones y técnicas que ayudan a los usuarios a encontrar información, toda vez que se ha publicado mucha información de recursos educativos, como objetos virtuales de aprendizaje, libros, trabajos de grado, monografías, entre otros de mucha utilidad, para todos los estudiantes y docentes que desea tener información puntual de un tema específico.

Igualmente se plasma una serie de análisis de los resultados que se obtienen con los buscadores actuales e información de la situación actual en cuanto al tema de las tecnologías web, recursos educativos para la búsqueda y recuperación de información en la web.

* M.Sc. Tecnología Informática, Grupo de Investigación de Software GIS, Universidad Pedagógica y Tecnológica de Colombia. E-mail: luisernesto.bonilla@uptc.edu.co

** M.Sc Ingeniería, Universidad EAFIT, Director Escuela de Ingeniería de Sistemas y Computación, Grupo de Investigación de Software GIS, Universidad Pedagógica y Tecnológica de Colombia. E-mail: Jorge.otalora@uptc.edu.co

Palabras clave

Tecnologías web, recursos educativos, inteligencia artificial

Abstract

Currently there are mechanisms for finding information on the web that do not give enough illustration, ie there is no precise information, since a lot of scattered information exists and it takes too long searching and organization. When a search is performed using information known search engines, upload information following many links, which give us a very low percentage of expected results.

The purpose is to present a model that allows search and retrieval of information of educational resources in a precise manner, based on a process type quasi experimental research, mainly by the characteristics that point and using semantic web applications and techniques that help users find information, since it is an extensive literature of educational resources, such as virtual learning objects, books, theses or dissertations, monographs, among other very useful for all students and teachers want to have timely information on a specific topic.

Also a series of analyzes of the results obtained with the current search engines and information about the current situation is reflected in the subject of web technologies, educational resources for search and retrieval of information on the web. The model proposed provides tools for developers to undertake a system that meets the task of finding free educational resource repositories.

Keywords

Web technologies, educational resources, artificial intelligence

1. Introducción

En la actualidad existen mecanismos de búsqueda de información en la web que no dan suficiente ilustración, es decir no hay precisión en las consultas, toda vez que existe mucho material educativo disperso y toma demasiado tiempo su recuperación y organización.

Cuando se realiza una búsqueda de información utilizando los motores de búsqueda conocidos como: Google, Yahoo, Bing, Lycos y Ozú, entre otros; se recupera información siguiendo muchos vínculos, que en ocasiones presentan el resultado esperado y en otras guían hacia sitios donde hay información "basura". Según Netcraft "a mayo de 2012 existían 662,959,946 sitios Web accesibles por los usuarios y continúa creciendo día a día, por lo que constituye una fuente de información demasiado engorrosa para su consulta"[1]. Además, "En la actualidad, la red mundial cuenta con unos 5 millones de sitios, y cerca de 1.500 millones de páginas de datos, se estima que diariamente la cantidad de información en Internet se expande a 2 millones de páginas y los motores de búsqueda no alcanzan a abarcar más del 16 por ciento" [2], donde muchos recursos educativos se encuentran en la web invisible.

Los modelos actuales de búsqueda están orientados a consultas de manera general y no puntualmente a recursos educativos almacenados en repositorios libres; por lo tanto las búsquedas no son precisas, al realizar exploraciones con los algunos motores de búsquedas existentes el resultado no es el esperado, toda vez que los procesos utilizados son elementales razón por la cual se ha convertido en tareas engorrosas, complejas y poco eficientes[3]. Por lo tanto, se requiere realizar búsquedas utilizando tecnologías de la web semántica para aprovechar sus propiedades como el significado de las palabras, ya que utilizando la sintaxis no se pueden tener búsquedas optimas porque ésta se encarga de la parte gramatical.

2. Marco de referencia

2.1. Tecnologías De La Web Semántica

Las tecnologías web son el avance de las ciencias aplicadas a nivel de software en la web, donde implica un conjunto de aplicaciones y técnicas que ayudan a los usuarios a encontrar información en un mundo extenso de datos, “Puede definirse la Web 2.0 como la evolución natural de la Web, que tiene su fundamento en el desarrollo de servicios que centran la atención en el usuario y en la participación activa de éstos. No se trata de una Web paralela o alternativa sino más bien de nuevas funcionalidades que permiten una mayor interactividad y conexión entre los usuarios”, tomado de [4] .

Las tecnologías web implican lenguajes , sistemas y herramientas de programación como: Ajax (Asynchronous JavaScript and XML), API (Application Programming Interface), P2P (Peer-to-peer), RSS (Really Simple Syndication) y XML (Extended Markup Language), para mostrar resultados finales conocidas como las redes sociales, motores de búsqueda, podemos dar ejemplos conocidos entre los que están: Amazon (<http://www.amazon.com/>), Facebook(<http://www.facebook.com/>), YouTube.(<http://www.youtube.com/>) y Wikipedia (<http://www.wikipedia.org/>)

2.2. Web semántica

Según Tim Berners-Lee, James Hendler, and Ora Lassila, La Web Semántica no es una web independiente, sino una extensión de la actual, facilitando que computadores y personas trabajen en conjunto. “Es una fuente para recuperar la información de la web (utilizando la arañas de los archivos RDF (Resource Description Framework, Marco de Descripción de Recursos, Es un formato que puede ser usado para almacenar recursos tales como marcadores o bookmarks y correos, tomado de la especificación RDF [5])) y acceder a los datos a través de agentes de la Web Semántica como se puede ver en [6].

La organización actual de la Web presenta algunos inconvenientes, siendo el principal la falta de precisión al momento de realizar búsquedas con los motores disponibles como Google, Yahoo, etc., este problema surge debido a que: “el significado del contenido de la web no es accesible por las máquinas” lo dice [7]; es decir, la información es presentada en un formato que es accesible y entendible por los usuarios, mas no puede ser interpretada por las máquinas. Debido a éste y otros inconvenientes que tiene consigo la Web

actual, Tim Berners Lee [8] (Creador de la Web) plantea el proyecto de Web Semántica, denominado así por la característica de dotar de significado o sentido, a la información y recursos disponibles en torno a la Web. En la actualidad se han realizado varios esfuerzos para alcanzar el objetivo de la Web Semántica tales como: descripción de los datos -metadatos-, lenguaje base que sea interoperable -XML-, método o técnica que permita representar el conocimiento -ontologías-, y software que permita comunicar e interpretar la información disponible en la Web -agentes- Todos estos elementos juegan un papel importante dentro de la Web Semántica como lo menciona [9].

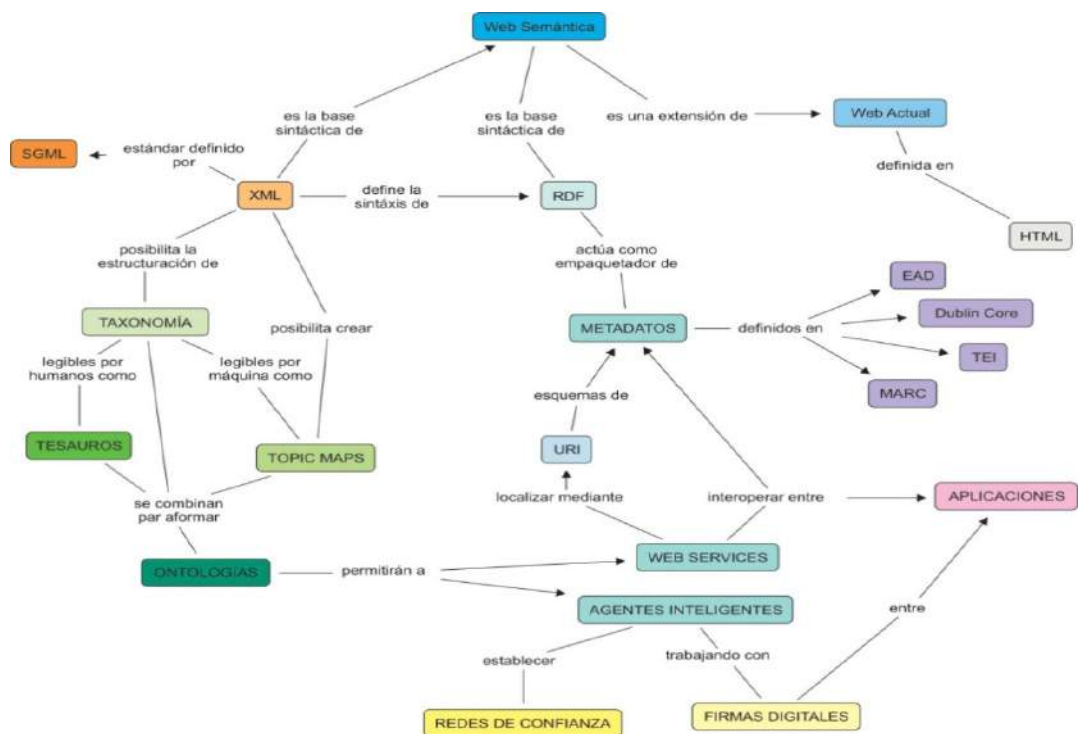


Figura 1. Mapa conceptual Web Semántica [6]

2.3. Ontologías

Uno de los mecanismos mediante el cual la Web semántica pretende alcanzar sus objetivos es con el uso de ontologías, las que permiten a desarrolladores o cualquier persona especificar: conceptos, relaciones, propiedades, taxonomías, reglas de inferencias, en torno a un dominio de interés; es decir, a través de dichos elementos es posible inferir conocimiento y obtener razonamiento automático. Sin embargo, las ontologías no se convierten en el único elemento que interviene en la obtención de conocimiento dentro de toda la Web, pues aquí también se ponen de manifiesto elementos como la inteligencia artificial (agentes inteligentes), buscadores, WebMining; lo podemos asegurar por intermedio de [10], donde manifiesta claramente sobre este concepto.

2.4. Agentes inteligentes

En Ciencias de la Computación el término agente inteligente puede ser usado para referirse a un agente de software que tiene algo de inteligencia, independientemente

de si no es un agente racional por definición de Russell [11] y Norvig [12]. Por ejemplo, programas autónomos utilizados para asistencia de un operador o de minería de datos (a veces denominado robots) son también llamados “agentes inteligentes”.

Según Nikola Kasabov [13] dice que los sistemas de Agentes Inteligentes deben exhibir las siguientes características:

- Aprender nuevos problemas e incrementar normas de solución.
- Capacidad de adaptación en línea y en tiempo real.
- Ser capaz de analizar condiciones en términos de comportamiento, el error y el éxito.
- Aprender y mejorar a través de la interacción con el medio ambiente (realización).
- Aprender rápidamente de grandes cantidades de datos.
- Deben estar basados en memoria de almacenamiento masivo y la recuperación de dicha capacidad.

2.5. Objetos virtuales de aprendizaje

Un objeto de aprendizaje es un conjunto de recursos digitales que puede ser utilizado en diversos contextos, con un propósito educativo y constituido por al menos tres componentes internos: contenidos, actividades de aprendizaje y elementos de contextualización. Además, el objeto de aprendizaje debe tener una estructura de información externa (metadato) para facilitar su almacenamiento, identificación y recuperación[14]

Los contenidos se manejan como objetos digitales, ya que tendrán un tratamiento y una organización que los transforma en algo más que contenidos y se convierten en componentes de un repositorio que los administrará como objetos[15].

2.6. WEB INVISIBLE

La Web Invisible es un fenómeno que influye como una de las principales causas de la dificultad para encontrar información de calidad en La Web. Como se refiere en [16], la Web Invisible cuenta con muchos más recursos de calidad que en la Web Superficial, siendo esta última aquella que puede ser indexada por los buscadores más populares; al contrario de la Web Invisible cuyos recursos no son indexados o son mal indexados por los motores de búsqueda.

Por tal motivo y según [17] debería denominarse, en realidad, la web “no indizable”, lo cual es un término mucho más adecuado, pero claramente alejado de la capacidad sugeridora del término “invisible”. Dado que, sin embargo, es el término más habitual incluso en la bibliografía técnica, usaremos en este trabajo el término Web o Internet invisible para referirnos a la información publicada en servidores Web que por diversos motivos no puede ser indizada y, por tanto, no puede ser encontrada por los motores de búsqueda convencionales.

3. Trabajos previos

Por mencionar algunos trabajos realizados que ayudan a este proceso investigativo están los siguientes:

- Preservación de objetos de aprendizaje en repositorios digitales: [18] presenta en su artículo la preservación de objetos de aprendizaje en repositorios digitales.
- Software basado en agentes inteligentes y servicios web para búsqueda de productos en la web: [19] menciona en su artículo, una aproximación que pretende obtener resultados mucho más precisos, basándose en el uso de agentes inteligentes que rastrean la red según las necesidades informativas del usuario.
- Ontología para la recomendación de recursos educativos almacenados en el repositorio de objetos de aprendizaje (ROA) DSpace: [10] en su proyecto menciona el uso de un Repositorio de Objetos de Aprendizaje (ROA) denominado DSpace, que fue adaptado para que funcione bajo el estándar LOM (estándar de marcaje de materiales destinados a la educación). Con el fin de que los Objetos de Aprendizaje (OA) almacenados en el ROA, ayuden en el proceso de aprendizaje, por ello realiza una ontología de tal manera que los profesionales en formación de la UTPL, cuentan con OA relacionados con el contenido de cada una de las asignaturas en las que está matriculado.
- Construcción de un motor de búsqueda de contenidos en repositorios confiables, basado en crawlers, enmarcado en una arquitectura orientada a servicio: [16] en su trabajo de grado, desarrolló un motor de búsqueda en repositorios confiables, usando arquitectura orientada a servicios (SOA), acoplado un bus de integración (ESB) como middleware y tecnología NoSQL como almacén de datos; de esta forma poder dar respuesta a las necesidades de usuarios, en encontrar información de calidad en La Web y poder abarcar la información de La Web Invisible.
- Arquitectura basada en agentes inteligentes y servicios web para la adaptación de contenidos educativos en plataformas de e-learning: [20], presenta una arquitectura genérica basada en agentes inteligentes y servicios web para la publicación de servicios, que puedan ser usados en plataformas de e-learning para lograr un proceso de adaptación en la entrega de contenidos educativos a los usuarios.
- Software basado en agentes inteligentes y servicios web para búsqueda de productos en la web: [21], muestra un software que se aproxima a mostrar los resultados más precisos haciendo uso de agentes inteligentes que rastrean la red según las necesidades del usuario.
- Método de Extracción de Información Semántica en ontologías: [22], en este artículo nos propone la inferencia de ontologías, mediante un conjunto de transformaciones mostrando un grafo como resultado de dicho proceso.

4. Metodología

Antes de mencionar la relación de la investigación cuasi experimental con el trabajo, es de aclarar que existen dos tipos de investigación: la investigación experimental e in-

vestigación no experimental. A su vez, la investigación experimental puede dividirse de acuerdo con las categorías en: preexperimentos, experimentos, “puros” verdaderos y cuasiexperimentos[23].

La investigación cuasi experimental proviene del ámbito educativo, donde la investigación de ciertos fenómenos no podía llevarse a cabo siguiendo los procedimientos experimentales. No hay manipulación de variables, éstas se observan y se describen tal como se presentan en su ambiente natural. Su metodología es fundamentalmente descriptiva, aunque puede valerse de algunos elementos cuantitativos y cualitativos[24].

Partiendo de este tipo de investigación se procede a enunciar los siguientes pasos que se desarrollan en todo el trabajo, como son:

- Plantear el problema de Investigación: En esta actividad tiene como fin, realizar el planteamiento del problema, donde se menciona la situación actual, los síntomas y las posibles causas que lo generan, dichos elementos son necesarios para la construcción de los objetivos que serán desarrollados a lo largo de la investigación, así como la justificación e importancia de la investigación
- Elaborar el marco de Referencia: Esta etapa define la recopilación de información relevante acerca de trabajos anteriores. Además de documentar la base teórica de la investigación consolidando el estado del arte de la misma y generando un soporte de consulta necesaria para afrontar el desarrollo de la investigación, se realiza una recolección de información para ser reflejada en el marco de referencia donde incluye el marco teórico, marco conceptual, marco legal y marco educativo
- Establecer la hipótesis: De acuerdo a todo el planteamiento del problema de investigación y objetivos se determina la hipótesis a la cual se quiere comprobar
- Recolección y Análisis los datos: Durante todo el proceso de investigación se realiza una recolección de información utilizando mecanismos como muestras de información, luego se realiza un proceso de organización y selección, que sirven de insumo en la determinación y análisis de los componentes que se deben incluir en el modelo y que complementan desarrollo y presentación de resultados
- Desarrollo y Presentación de los resultados: Para construir el modelo de búsqueda de información contemplado dentro de la investigación, es necesario recurrir a una serie de pasos secuenciales, que conlleven al desarrollo de cada uno de los objetivos específicos y por ende al cumplimiento del objetivo general de la investigación. Por ello a continuación se menciona cada uno de los pasos así:
 - Para el desarrollo en cuanto a la selección de un modelo de búsqueda y recuperación mediante un estudio comparativo que facilite la gestión semántica de información en recursos educativos, se debe realizar una serie de pruebas plasmadas en una matriz de resultados que se refleje el modelo óptimo a utilizar en la búsqueda y recuperación de la información.

- En cuanto a la Caracterización de los recursos educativos teniendo en cuenta su información semántica con el fin de determinar los criterios de búsqueda en repositorios libres, se realiza la investigación de cada una de la características de los recursos educativos para determinar la mejor alternativa en la búsqueda y recuperación de la información; para luego a partir de la caracterización anterior seleccionar los componentes conceptuales para el diseño del modelo.
- Diseñar un esquema del modelo de representación y búsqueda de recursos educativos haciendo uso de la web semántica, donde se utilicen los esquemas de representación como casos de uso, diagrama de actividades, diagrama de secuencia y diagrama de componentes, para mostrar el modelo resultado de la investigación.
- Mediante una metodología de validación aprobar el modelo definido haciendo uso de una simulación teórica de búsqueda de información, mostrando el resultado en una matriz, indicando la viabilidad del modelo.

5. Análisis y resultados

5.1. Análisis de los Modelos de Búsqueda

De acuerdo con Martínez, «Las técnicas de recuperación de información empleadas en Internet, proceden de las empleadas en los sistemas de recuperación de Información tradicionales, y por ello surgen problemas cuando se realizan operaciones de recuperación de información, en tanto que el entorno de trabajo no es el mismo y las características intrínsecas de los datos almacenados difieren considerablemente. Al mismo tiempo, en la web surgen nuevos problemas, tales como los populares fenómenos denominados cloaking , links farms y guest spamming , o los vinculados con el enorme tamaño del índice de estos sistemas de recuperación de información, que poco a poco llega a alcanzar magnitudes impresionantes, muy difíciles de gestionar adecuadamente con los modelos tradicionales»[25], por lo anterior se hace referencia a técnicas las cuales optimizan los resultados de la recuperación.

A continuación se presenta una relación de técnicas de búsqueda, como insumo para determinar la mejor.

Técnica/ característica	Comparación de estados	Conocimiento del dominio	Examina todos los nodos	Alta Complejidad	Problemas pequeños
Exhaustiva	SI	SI	SI	SI	SI
Heurística	NO	NO	SI	SI	NO
Búsqueda dirigida por los datos	SI	SI	SI	NO	NO
Búsqueda dirigida por metas	SI	NO	NO	SI	SI

Tabla 1. Comparativo técnicas de búsqueda
Fuente: Autor

Por tal motivo realizando esta comparación se determina que la mejor técnica es la llamada «búsqueda dirigida por datos» también llamada búsqueda hacia adelante.

A continuación se presenta los resultados de las pruebas aplicadas, realizando un muestreo simple (todos los elementos pueden ser escogidos o tomados), donde lo separamos por áreas importantes como la tecnología, educación, deportes y entretenimiento; además como se observa en la siguiente figura, los motores de búsqueda más utilizados son Google, Bing y Yahoo, por este motivo se realizan las pruebas con dichos motores.

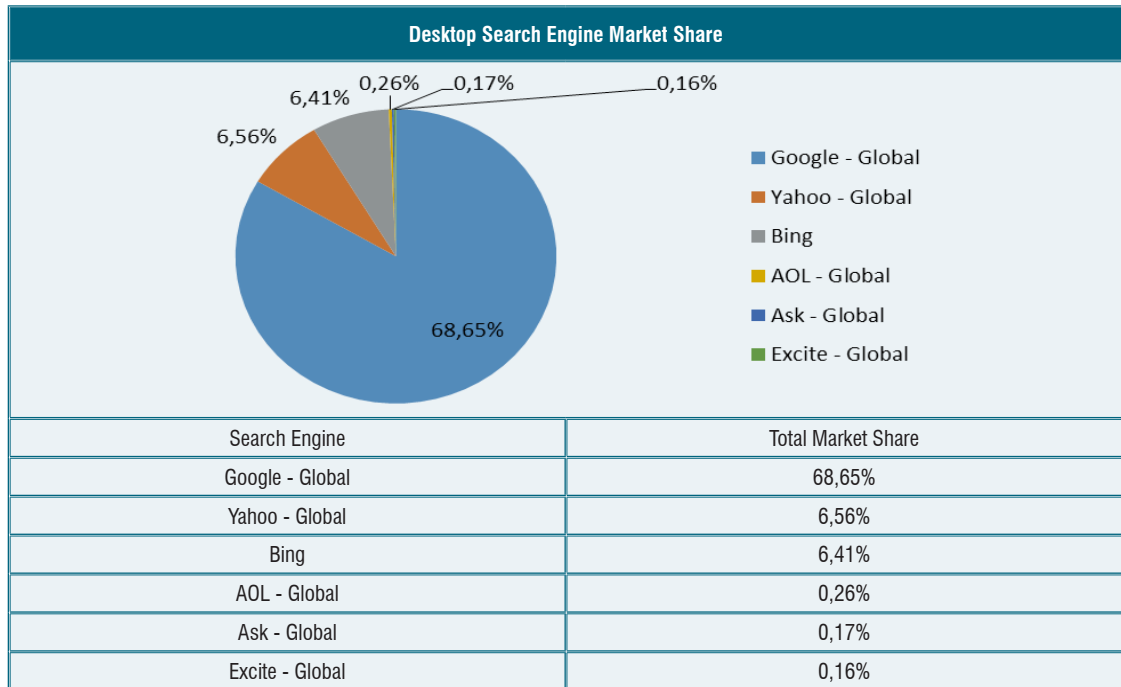
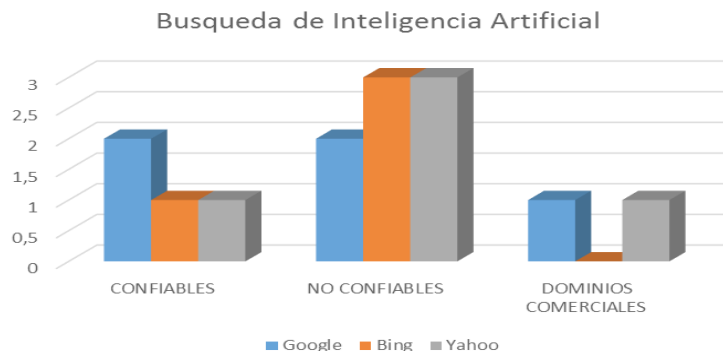


Figura 2: Analítica web en tiempo real
Fuente: Market Share Statistics for Internet Technologies [26]

Luego se realiza una clasificación del sitio Web recuperado en:

- C: Repositorios confiables
- NC: Repositorios no confiables
- DC: Dominios Comerciales

Una vez realizada la consulta, se toman los primeros cinco sitios web recuperados por cada motor de búsqueda, por lo que maneja una escala es de 1 a 5 para cada una de las clasificaciones, como se observa en la siguiente gráfica:



Para realizar el diseño de un modelo de representación y búsqueda de recursos educativos haciendo uso de la web semántica, se debe realizar una fundamentación en cuanto a que el modelo permita hacer búsqueda, extracción e indexación de contenidos en repositorios libres (semillas), partiendo de un conjunto de semillas suministradas por un administrador, La información del metadato de cada página debe ser obtenida individualmente, para ser identificadas, además el contenido tendrá que ser indexado para ayudar en el proceso de búsqueda.

Luego se debe permitir la extracción de varios contenidos al mismo tiempo, cumpliendo con las políticas de exclusión de robots y de no sobrecargar los distintos servidores; por esta razón debe ordenar el tiempo de acceso a cada servidor y página.

Partiendo que los motores de búsqueda tienen tres componentes fundamentales que son: Robot, indexador y mecanismo de búsqueda, en la siguiente imagen se relaciona algunos procesos así:

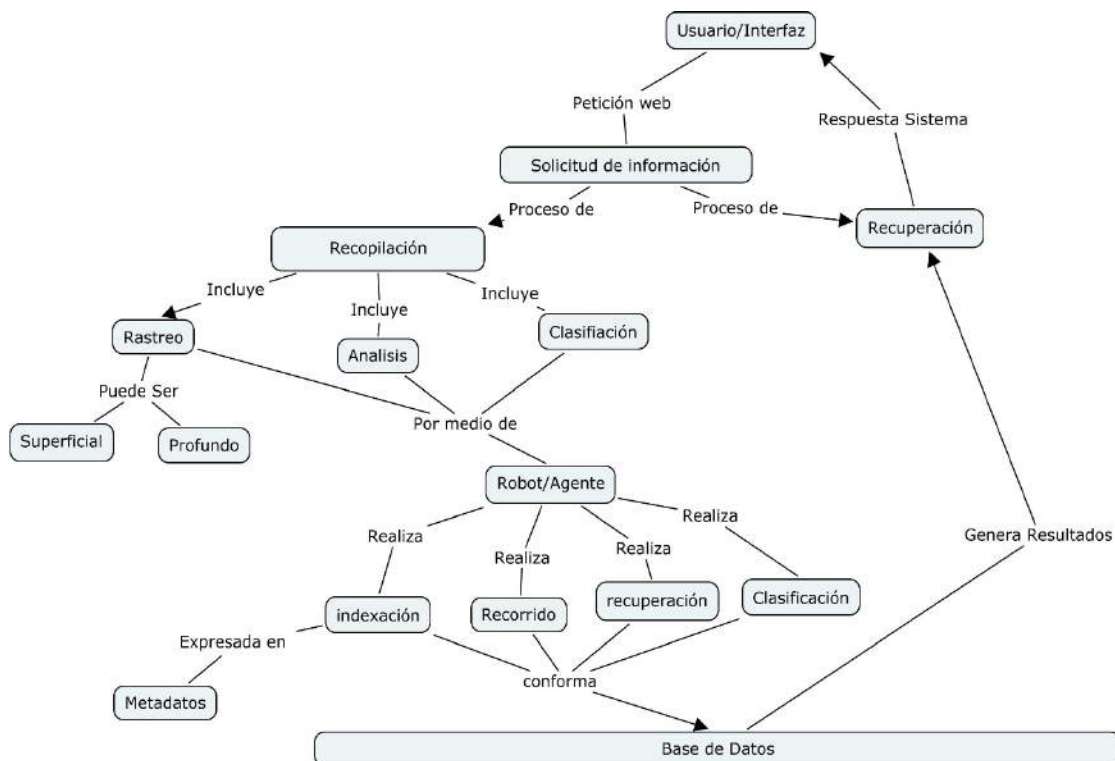


Figura 3. Funcionamiento del buscador.

Para describir el funcionamiento del buscador se puede mencionar que hay dos procesos: uno que es la recopilación de información donde se hace el rastreo, análisis, indexación y clasificación de las semillas o páginas, el otro proceso es la recuperación de la información. La recopilación se realiza por medio de robots y agentes.

Los robots o agentes “son programas que rastrean la estructura hipertextual de la Web, recogen información sobre las páginas, indexan la información, la clasifican y conforman

una base de datos que es la que posteriormente se muestra en la interfaz como resultado de dicha búsqueda”[27].

La indización también llamada granularidad se realiza extrayendo los datos que se encuentran en los metadatos de cada semilla, permitiendo la precisión en la localización de una palabra o descriptor con exactitud.

El rastreo, búsqueda o exploración puede ser superficial o profundo con capacidad para rastrear imágenes y texto.

El usuario por medio de la interfaz realiza una solicitud de información o una petición web, luego el sistema realiza un proceso de recopilación o recuperación, dicho proceso rastrea en la base de datos por medio de un agente o robot informático la coincidencia de la petición con lo almacenado, una vez realizado el recorrido y capturados los datos del sitio de localización de la información, se realiza un análisis y clasificación dando lugar a la generación de resultados.

A continuación se describe detalladamente los componentes fundamentales:

Componentes del Robot

El robot es una aplicación que se encarga de procesar información, basándose primero en la consulta a realizarse y segundo en la base de datos de repositorios libres previamente catalogados, por lo que su trabajo consiste en capturar una dirección URL o identificador del repositorio a partir de una Base de datos, descargar su contenido, catalogar y aprovechar los enlaces de dichas páginas para hacer una nueva búsqueda con cada documento vinculado. A su vez, cada nuevo documento vinculado se clasifica nuevamente. En la siguiente figura se muestra cómo es realizado dicho proceso.

Los componentes que lo conforman son los siguientes: un componente llamado gestor de descargas el cual explora el contenido de un sitio web, crea un documento con sus metadatos y almacena el contenido en una base de datos de resultados. A su vez, busca en dicho sitio más enlaces o URLs, los cuales son enviados a una cola de espera para su procesamiento posterior. El componente llamado programador, se encarga de tomar los enlaces de la cola de espera para enviarlos al programador y realizar de nuevo el proceso.

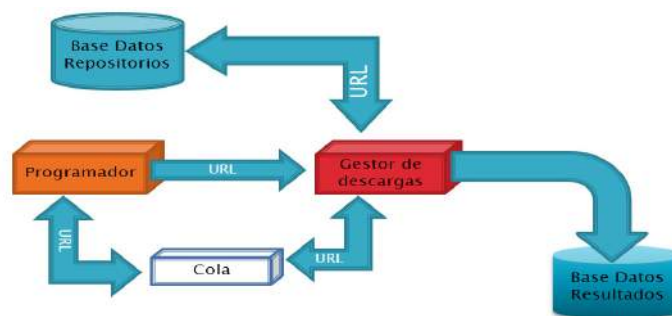


Figura 4. Componentes de un Robot. Mejorado de [28]

Componentes del Indexador

Una vez tomada la información de la página rastreada, se somete a la indexación. Es un proceso complejo en el cual se analiza el sitio y se ordena en el índice según su importancia respecto a una palabra clave que se consultó. En la siguiente figura se muestran dos componentes fundamentales como son: indexador y analizador, los cuales permiten realizar dicha labor.

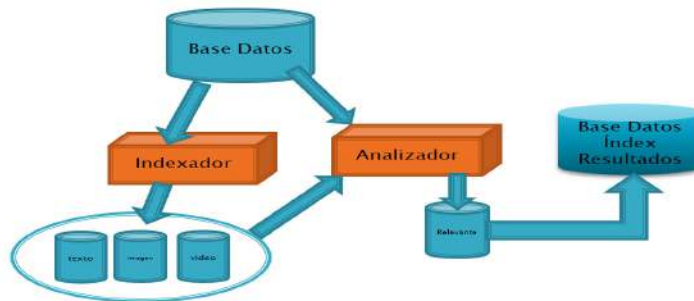


Figura 5. Componentes del Indexador. Autor

Componentes del Mecanismo de Búsqueda

La función de este componente es encontrar páginas del interés para los usuarios, por lo que se realiza todo el proceso de búsqueda, se indexan y se devuelven los resultados. Por lo que una vez se tenga la base de datos indexada, se envía los resultados obtenidos al mecanismo de ranqueo de búsqueda, como se observa en la siguiente figura y por último se envía a la interfaz del usuario donde

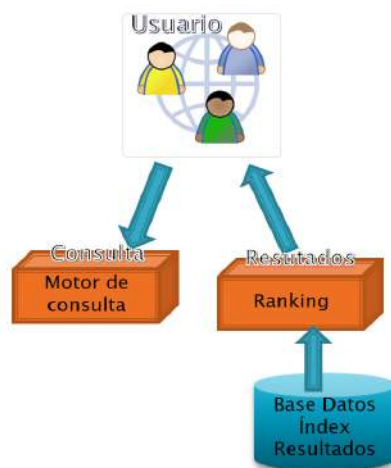


Figura 6. Componentes del mecanismo de búsqueda. Autor

Para concluir se realiza una validación del modelo por lo que se utiliza la lógica funcional, la cual se refiere al funcionamiento y tareas soportadas por el modelo. Su valoración se determina por la precisión de los resultados obtenidos al momento de realizar búsquedas. Se define matemáticamente como:

$$Precisión(x) = \frac{1}{N} + \sum_{n=1}^N \left(\frac{i_o}{n_i} \right)$$

En donde:

N= número total del búsquedas

i_o =instancias correctas que satisfacen el criterio de búsqueda

n_i =número total de instancias retornadas en la búsqueda

Se realiza una prueba con un total de cinco búsquedas así:

- Primera búsqueda: ontologías
- Segunda búsqueda: inteligencia artificial
- Tercera búsqueda: célula
- Cuarta búsqueda: historia de Colombia
- Quinta búsqueda: Futbol

Criterio	Descripción	Valor
Lógica Funcional	Número total de búsquedas	5
	i_o Primera búsqueda	6
	n_i Primera búsqueda	4
	i_o Segunda búsqueda	4
	n_i Segunda búsqueda	5
	i_o Tercera búsqueda	5
	n_i Tercera búsqueda	7
	i_o Cuarta búsqueda	7
	n_i Cuarta búsqueda	6
	i_o Quinta búsqueda	7
n_i Quinta búsqueda	5	
Precisión		0,94

Tabla 2. Validación a nivel funcional
Fuente: Autor

Como se puede observar una vez realizado el recorrido del modelo, se puede determinar que tiene un porcentaje muy alto en cuanto a la precisión de las búsquedas.

6. Conclusiones

La mejor técnica de búsqueda y recuperación semántica de información en recursos educativos es la llamada “búsqueda dirigida por datos” también llamada búsqueda hacia adelante, la cual se determinó realizando esta comparación con las demás técnicas de búsqueda y posteriormente evaluada con pruebas realizadas en algunos motores de búsqueda como google, Yahoo y Bing.

La caracterización de los recursos educativos teniendo en cuenta su información semántica se determinan y se conocen de acuerdo a la estructura de los documentos almacenados en los repositorios, donde poseen quince propiedades para su uso, con el fin de determinar los criterios de búsqueda en repositorios libres, de acuerdo al concepto de Dublin Core Metadata Initiative (DCMI).

Los componentes conceptuales del modelo basados en la caracterización de los recursos educativos permiten tener una integridad entre las técnicas de búsqueda frente al modelo desarrollado, mejorando las búsquedas y permitiendo tener una administración de las semillas o sitios web validados para que la indexación y recuperación sea efectiva.

El modelo desarrollado permitirá hacer búsqueda, extracción e indexación de contenidos en repositorios libres (semillas), partiendo de un conjunto de semillas suministradas por un administrador basado en tres componentes fundamentales (Robot, indexador y mecanismo de búsqueda) y ayudados por medio de robots y agentes que no son más que programas que rastrean la estructura hipertextual de la Web, recogen información sobre las páginas, indizan la información, la clasifican y conforman una base de datos que es la que posteriormente se muestra en la interfaz como resultado de dicha búsqueda

Una vez desarrollado el modelo de búsqueda, mediante un modelo de validación llamado algoritmo de la colonia de hormigas la cual es una técnica probabilística para solucionar problemas computacionales que pueden reducirse a buscar los mejores caminos, se pudo comprobar la efectividad del modelo donde los resultados calculados en una simulación teórica, como el promedio de nodos recorridos en las rutas conseguidas por las hormigas en el proceso construcción de caminos es de 9 nodos dentro del modelo, lo cual determina que siempre hay un camino eficiente en conseguir los resultados.

Al realizar una validación funcional se pudo determinar que tiene una precisión del 94%, lo cual indica que es un excelente resultado superando la satisfacción de los usuarios, donde arroja información que pertenece a repositorios confiables.

Con el modelo que se presenta y la recopilación de información plasmada en este trabajo, se dan herramientas para que desarrolladores realicen un sistema que cumpla con la tarea de encontrar recursos educativos en repositorios libres, con efectividad

y eficiencia, coadyuvando a la formación académica y la reducción de tiempo en las búsquedas en la web.

Existe mucho material educativo digital, objetos virtuales de aprendizaje y portales interactivos que ayudan a las labores académicas de todos los usuarios y que es desconocido por falta de mecanismos de búsqueda efectivos.

Referencias bibliograficas

- [1] Netcraft, "Internet Research, Anti-Phishing and PCI Security Services," 2012. [Online]. Available: <http://news.netcraft.com/>. [Accessed: 02-Jun-2012].
- [2] El tiempo, "MOTORES DE BÚSQUEDA EN REPARACIÓN," 04/02/2012, p. 1, 2014.
- [3] P. L. Guzmán, "Motores de búsqueda; soluciones con aplicaciones de Google.," *eFormadores*, pp. 1–12, 2011.
- [4] M. V. (universitat P. F. Mari-Carmen Marcos (Universitat Pompeu Fabra), "Las bibliotecas en un entorno Web 2.0 - Hipertext - (UPF)." [Online]. Available: <http://www.upf.edu/hipertextnet/numero-7/bibliotecas-2.0.html>. [Accessed: 02-Jun-2012].
- [5] W3C, "RDF Primer." Feb-2004.
- [6] "semanticweb.org." [Online]. Available: http://semanticweb.org/wiki/Main_Page. [Accessed: 02-Jun-2012].
- [7] G. A. F. V. H. Grigoris Antoniou, *Semantic Web Primer*. <http://books.google.com.co/books?op=lookup&id=cYAiS1wbRsC&contnue=http://books.google.com.co/books%3Fid%3DclYAiS1wbRsC%26printsec%3Dfrontcover%26hl%3Des&hl=es,2004>, p. 234.
- [8] "Tim Berners-Lee," 2012. [Online]. Available: <http://www.w3.org/People/Berners-Lee/>. [Accessed: 05-Nov-2012].
- [9] R. Keilyn, Rodríguez Perojo y Rodrigo, "Web semántica: un nuevo enfoque para la organización y recuperación de información en el web," 2005. [Online]. Available: http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm. [Accessed: 21-Oct-2012].
- [10] C. Flores and M. Alberto, "Ontología para la recomendación de recursos educativos almacenados en el repositorio de objetos de aprendizaje (ROA) DSpace," Jan. 2012.
- [11] D. Marcovecchio, L. Molas, E. Montenegro, F. Sisul, M. Torres, A. Garc, D. Mart, G. Simari, and L. G. Simari, "Modelo DBI con argumentación y planificación aplicado a juegos con agentes inteligentes autónomos," pp. 182–186, 2012.
- [12] N. Peter, "Artificial intelligence: The coming superintelligence," *New Sci.*, vol. 216, no. 0262–4079, 2012.
- [13] N. KASABOV, *Springer handbook of bio-/ neuroinformatics*. 2012, p. 1200.
- [14] M. de E. Nacional, "Docentes y Directivos de Superior -." [Online]. Available: <http://www.colombiaprende.edu.co/html/directivos/1598/article-99543.html>. [Accessed: 02-Jun-2012].
- [15] C. López Guzmán and F. J. García Peñalvo, "Los repositorios digitales en el ámbito universitario," 2007. [Online]. Available: <http://gredos.usal.es/jspui/handle/10366/55713>. [Accessed: 02-Jun-2012].
- [16] Edward J. Valera G, *CONSTRUCCIÓN DE UN MOTOR DE BÚSQUEDA DE CONTENIDOS EN REPOSITORIOS CONFIABLES, BASADO EN CRAWLERS, ENMARCADO EN UNA ARQUITECTURA ORIENTADA A SERVICIO*. 2012, p. 246.
- [17] Lluís, Codina, "Internet invisible y web semántica," *Nov*, 2003, vol. 2, Nov. 2003.
- [18] J. Boté Vericad and J. Minguillón, "Preservación de objetos de aprendizaje en repositorios digitales," *RUSC. Rev. Univ. y Soc. del Conoc.*, vol. 9, no. 1, pp. 22–35, Jan. 2012.
- [19] ESPEJO BOHORQUEZ, Yeismer; TELLEZ, Magaly and RODRIGUEZ, Jorge Enrique, "Intelligent-agents-based software and web services for products searchover the internet," *Mar*. 2012, vol. 16, no. 31, Mar. 2012.
- [20] D. Huerva, R. Fabregat, C. Mejía, and S. Gómez, "ARQUITECTURA BASADA EN AGENTES INTELIGENTES Y SERVICIOS WEB PARA LA ADAPTACIÓN DE CONTENIDOS EDUCATIVOS EN PLATAFORMAS DE E-LEARNING."
- [21] J. E. Rodríguez, "Software basado en agentes inteligentes y servicios web para búsqueda de productos en la web."
- [22] H. D. Díaz, A. F. Orquín, U. De Matanzas, R. M. Guillena, U. De Alicante, and S. V. Pérez, "Método de Extracción de Información Semántica en ontologías Semantic Information Extraction method on ontologies."
- [23] P. B. Lucio, "Metodología de la Investigación," pp. 1–25, 2003.
- [24] D. t. Campbell and J. C. Stanley, *Experimental And Quasi-Experimental Designs For Research*. Boston: Houghton Mifflin, 1963, pp. 171–246.
- [25] F. Martínez Méndez, *RECUPERACIÓN DE INFORMACIÓN: MODELOS, SISTEMAS Y EVALUACIÓN, EL KIOSKO*. 2004, p. 106.
- [26] NetMarketshare, "NetMarketshare," *Abril*, 2014, 2014. [Online]. Available: www.netmarketshare.com.
- [27] J. O. Gómez, "Elementos teórico-prácticos útiles para comprender el uso de los motores de búsqueda en Internet," *Scielo*, 2003.
- [28] I. F. Camargo Sarmiento and S. Ordóñez Salinas, "EVOLUCIÓN Y TENDENCIAS ACTUALES DE LOS WEB CRAWLERS," *Ing. Univ. Dist. Fr. Jose Caldas*, vol. 18, 2013.