Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

Tipo de artículo: Artículo original
Temática: Tecnologías de la información y las telecomunicaciones
Recibido: 11/12/2017 | Aceptado: 22/01/2018

Procesamiento Semántico de información en Sistemas de Recuperación de Información

Semantic processing of information in Information Retrieval Systems

Hubert Viltres Sala [1*], Paúl Rodríguez Leyva [2], Juan Pedro Febles[3], Vivian Estrada Sentí[3]

[1]Departamento de Práctica Profesional, Universidad de las Ciencias Informáticas, La Habana, Cuba. hviltres@uci.cu

[2]Departamento de Soluciones Informáticas para Internet, Universidad de las Ciencias Informáticas, La Habana, Cuba. pleyva@uci.cu

[3]Departamento Metodológico de Postgrado, Universidad de las Ciencias Informáticas, La Habana, Cuba. {febles, vivian}@uci.cu

[*] **Autor para correspondencia**: hviltres@uci.cu

Resumen

El procesamiento de información con anotación semántica permite identificar la intención de búsqueda del usuario y ajustar el resultado según el contexto de la información. La presente investigación propone realizar el procesamiento semántico de información para mejorar la relevante de los resultados brindados a los usuarios cuando acceden a un Sistema de Recuperación de Información. La propuesta propone desarrollar tres componentes (Rastreo-Indexación, Procesamiento y Presentación) para identificar la necesidad de información del usuario mediante el procesamiento, selección y posterior publicación de la información recuperada. El componente de rastreo e indexación permite identificar los sitios web disponibles para extraer la información y realizar la anotación semántica aplicando diferentes técnicas de procesamiento de información. El componente de procesamiento se analiza las preferencias del usuario y se procesa la consulta realizada para calcular la similitud de la información indexada. Posteriormente se ordenan los resultados según la relevancia para mostrar en el componente de Presentación una cantidad de información que pueda ser asimilada por los usuarios. Para la validación de la propuesta se emplearon las métricas de precisión y exhaustividad que permitieron demostrar la calidad, pertinencia y relevancia de la recuperación de información con anotación semántica.

Palabras clave: anotación semántica, recuperación de información, relevancia, similitud, web semántica.

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

Abstract

The processing of information with semantic annotation allows identifying the user's search intention and adjusting the result according to the context of the information. The present investigation proposes to carry out the semantic processing of information to improve the relevant of the results provided to the users when they access an Information Retrieval System. The proposal proposes to develop three components (Crawling-Indexing, Processing and Presentation) to identify the need for user information through the processing, selection and subsequent publication of retrieved information. The tracking and indexing component allows identifying the available websites to extract the information and make the semantic annotation applying different information processing techniques. The processing component analyzes the user's preferences and the query made is processed to calculate the similarity of the indexed information. Subsequently, the results are ordered according to the relevance to show in the Presentation component a quantity of information that can be assimilated by the users. For the validation of the proposal, precision and completeness metrics were used to demonstrate the quality, relevance and relevance of information retrieval with semantic annotation.

Keywords: *information retrieval, relevance, semantic annotation, semantic web, similarity*

Introducción

El desarrollo de la sociedad, el surgimiento de tecnologías y herramientas para mejorar el acceso a la información y el rápido crecimiento de Internet en los últimos años, ha posibilitado que se genere un gran volumen de contenido web. La información disponible en la web se encuentra dispersa, está poco estructurada o es invisible al usuario común, dificultando el proceso de acceso a información de alta calidad y valor para el usuario. En este contexto los usuarios cuando acceden a Internet se sienten abrumados por la sobrecarga de información y no obtienen de forma fácil y rápida la información que más se ajusta a sus necesidades, limitan su experiencia en la utilización de un sistema de recuperación de información. Existen más de un billón de sitios web en Internet y cada día se incrementa exponencialmente la cantidad de información disponible. Generando nuevas oportunidades y disímiles retos para los usuarios cuando intentan obtener información relevante. Debido a la gran cantidad de información disponible en Internet y la dificultad de asimilarlas, los usuarios se apoyan en los sistemas de recuperación de información (SRI) para encontrar lo que buscan.

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

Los sistemas de recuperación de información mediante la utilización de diferentes herramientas, métodos y técnicas recuperan la información pública de la web para su posterior análisis, seleccionando y ordenan la información más relevante para la necesidad del usuario. Entre las principales fuentes para obtener información se encuentran los repositorios de componentes, base de datos y los buscadores que permiten simplificar y agrupar información relevante, utilizando determinados conceptos de organización de la información. El objetivo principal de un SRI según se plantean en Deco, Reyes y Bender (2012) es satisfacer la necesidad de información planteada por un usuario en una consulta en lenguaje natural especificada a través de un conjunto de palabras claves (ver figura 1), que ayudan a identificar la información más relevante para el usuario.

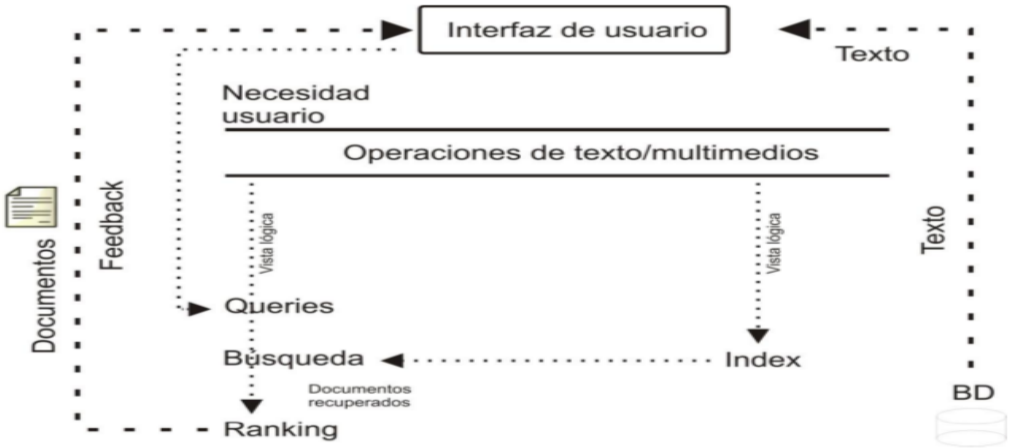

Figura 1: Proceso de búsqueda de información- fuente Vuotto, Bogetti y Fernández, 2015

Autores como Salton y Mcgill (1983), Gonzalo, et al. (2017) y Martínez (2004) plantean que la Búsqueda y Recuperación de Información tiene como principal objetivo proporcionar información relevante al usuario para satisfacer su necesidad de información. Dentro de la BRI se definen cinco actividades principales (localizar, seleccionar, interpretar, sintetizar y comunicar la información) para guiar el proceso de obtener información ajustada a la necesidad del usuario. Estas cinco actividades están contempladas en los tres principales componentes de un buscador en la actualidad (rastreador, indexador y procesador).

Durante el proceso de recuperación de información los motores de búsqueda tradicionales generalmente utilizan técnicas que determinan la relevancia por la coincidencia de las palabras claves en los documentos y no analizan las relaciones que existen entre el significado implícito de las palabras claves y el documento. Por ellos se hace necesario

realizar un proceso de identificación de la intención del usuario detrás de la pregunta realizada y ajustar el resultado al contexto de la pregunta. Varios autores plantean que la recuperación semántica de información mejora la calidad y relevancia de la información mostrada a los usuarios, ya que emplea técnicas de procesamiento en lenguaje natural, utiliza ontologías para identificar el contexto y la relevancia se establece por la similitud semántica de la consulta y los documentos indexados.

Recuperación semántica de información

La Web Semántica está cambiando la forma de obtener información en internet, es una de las tecnologías que más impacto ha generado para los usuarios de internet por la calidad de la información que obtiene. Berners-Lee, et al. (2001) define la Web Semántica como "…una extensión de la Web actual, en la cual la información tiene un significado bien definido, facilitando a las computadoras trabajar mejor en cooperación con los humanos" y su objetivo principal ha sido permitir que los datos almacenados en la Web puedan ser procesados por las máquinas de manera inteligente, facilitando a las personas la búsqueda, integración y análisis de la información disponible. La web semántica tiene como principio el procesamiento de información de forma automática mediante la utilización de inteligencia artificial utilizando una gran variedad de algoritmos. Pretende además comprender la necesidad expresada por el usuario en una consulta realizada y dotar la búsqueda de un significado, identificando y brindando información confiable. Para realizar la búsqueda semántica se emplean buscadores semánticos que son "sistemas de recuperación de información que entienden la necesidad el usuario y analizan la información disponible en la Web mediante el uso de algoritmos que simulan comprensión o entendimiento".

El funcionamiento general de un buscador semántico en Martínez, et al. (2010) está asociado a las siguientes características:

- Permite realizar búsquedas por campos.
- Tiene capacidad para extender los términos de la consulta mediante sinónimos o palabras relacionadas.
- Identifica entidades nombradas, como nombres de empresas, organizaciones o personas, que se utilizan con ese significado en el proceso de búsqueda.
- Emplea técnicas de agrupamiento para construir categorizaciones de contenidos sobre los que buscar o para agrupar términos clave. Es el caso de las nubes de etiquetas que muestran los términos clave de un sitio web según su importancia.

- Detecta relaciones entre términos de búsqueda y palabras que aparecen en contenidos basándose en modelos de conocimiento representados a través de ontologías.
- Ofrece la posibilidad de emplear lenguaje natural para expresar consultas e incluso preguntas factuales, para las que se obtienen respuestas concretas (Martínez, et al., 2010).

Las características antes expuestas evidencian las posibilidades de la web semántica en la recuperación de información donde un usuario expresa en lenguaje natural su intención de búsqueda y el buscador analizar y seleccionar la información ajustada a esa necesidad. En el contexto de la web cubana donde las limitaciones tecnológicas dificultad el proceso de recuperación de información para resolver este problema se necesario emplear la recuperación de información semántica.

Recuperación de información en la web cubana

En Cuba existen más de 6 mil sitios web alojados bajo el dominio .cu con información variada. Para acceder a la información almacenada en la web cubana los usuarios utilizan diferentes sistemas de recuperación de información, pero no siempre obtienen información relevante, debido principalmente a la:

- Heterogeneidad de fuentes de información.
- Calidad de la información.
- Visibilidad de la información.
- Accesibilidad de la información.

Además de los elementos anteriormente mencionados otro factor que afecta la recuperación de la información es la utilización de sistemas que emplean algoritmos de cálculo de la similitud por palabras, donde no se analiza la semántica de la información. Un análisis realizado sobre los sistemas que determinan la similitud por palabras claves evidenció las siguientes deficiencias:

- Dificultad para entender la necesidad del usuario expresada en lenguaje natural.
- Poca precisión de los resultados porque se potencia la similitud de las palabras clave.
- Sensibilidad de los resultados frente a los términos exactos introducidos.
- Selección de la información por la relevancia del posicionamiento del sitio web.

Las dificultades antes expuestas evidencian poca precisión y exactitud en el proceso de recuperación de información y disminuyen la experiencia del usuario al realizar una búsqueda de información. Estas deficiencias aparejadas a la necesidad de brindarle a los usuarios información de alta calidad plantean la necesidad de desarrollar un sistema de

recuperación de información con anotación semántica que permita seleccionar la información más ajustada a las necesidades de los usuarios y con ello mejorar su experiencia en la web cubana.

Búsqueda semántica de información

La web semántica es una extensión de la web actual, autores como Vuotto, Bogetti y Fernández (2015) Berners-Lee, et al. (2001), Martínez, et al. (2010), García (2015) y Redondo (2017) plantean que permite obtener información de forma eficiente mediante la integración, automatización y reutilización de los datos empleando diversas técnicas para mejorar la relevancia de la información recolectada. Las búsquedas semánticas permiten obtener resultados relevantes al entender la necesidad de información del usuario expresada en lenguaje natural.

Según Redondo (2017) el objetivo de la búsqueda semántica es mejorar la precisión de la búsqueda mediante la comprensión de la intención del usuario cuando realiza una consulta y el significado contextual de los datos en la fuente de conocimiento. La búsqueda semántica predice lo que él usuario expresa explícitamente (intención de búsqueda) y ajusta su necesidad (contexto) a la información disponible seleccionando la más relevante para el usuario. Los sistemas de recuperación de información centran su implementación en comprender la búsqueda empleando procesamiento de consulta, extracción de conocimiento de las fuentes de datos, ajuste de las preferencias del usuario y cálculo de la relevancia. El modelo que se propone en la investigación se sustenta sobre la base de recuperar información relevante para el usuario utilizando la tecnología semántica.

Materiales y métodos

Con el objetivo de obtener información relevante para los usuarios se implementa un modelo computacional que permite procesar la información disponible semánticamente. En el modelo se plantean los tres componentes (Rastreo-Indexación, Procesamiento y Presentación) principales; que permitirán identificar la necesidad de información del usuario mediante el procesamiento, selección y posterior publicación de la información recuperada. En la figura 2 se presentan los componentes que sustentan el proceso de búsqueda y recuperación de información en la web. A continuación, se describen cada uno de los tres componentes.

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

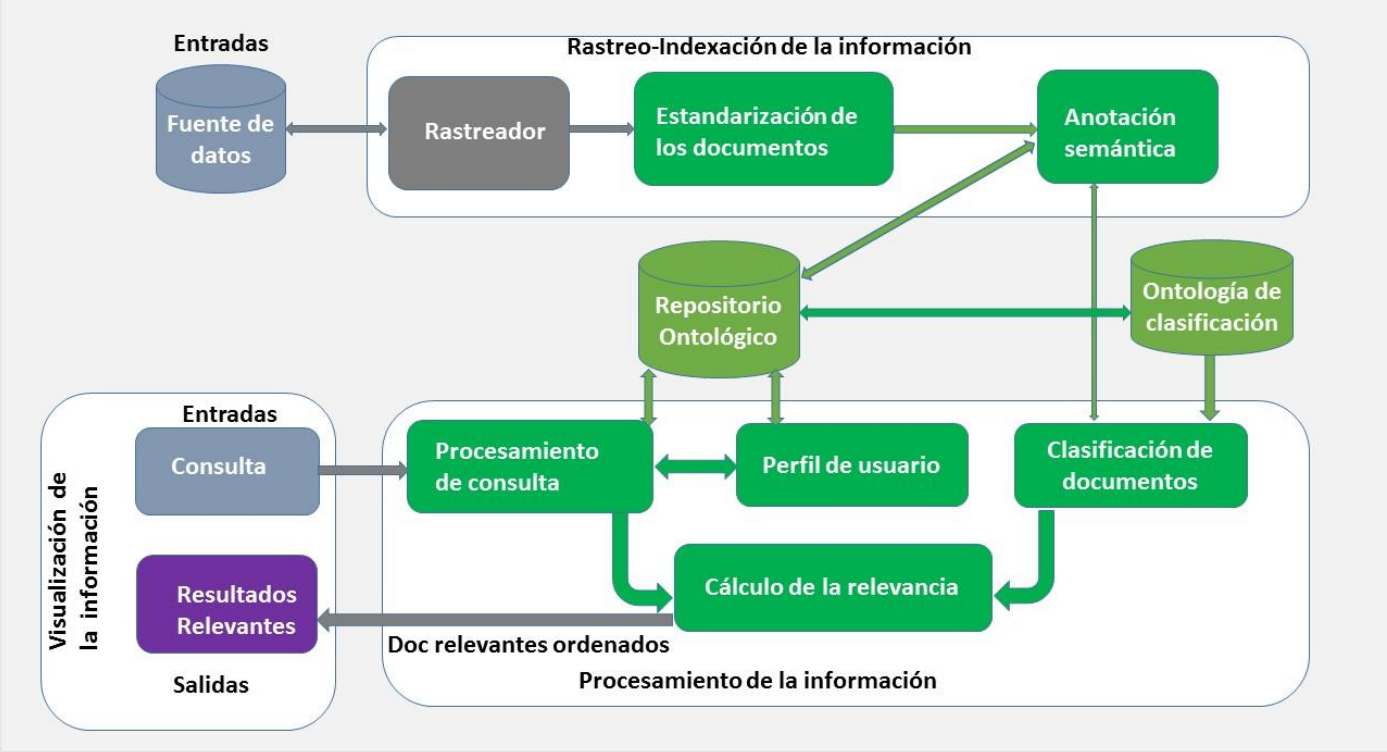

Figura 2: Modelo computacional para el procesamiento semántico de información (elaboración propia)

Componente rastreo e indexación

El componente de rastreo e indexación permite identificar los sitios web disponibles, además de recupera y almacena la información de cada página web para su posterior procesamiento y presentación a los usuarios cuando realice una consulta. Los rastreadores se encargan de explorar la web identificando las páginas que han sido creadas o actualizadas para continuar actualizando su índice de información. Después de rastreada se almacenan diferentes metadatos (url, resumen del contenido, enlaces, palabras claves, idioma) que son utilizados para extraer conocimiento empleando técnicas de la web semántica.

Rastreo de la web

El proceso de rastreo comienza con una lista de enlaces a sitios web proporcionados por rastreos anteriores o por sitemap; entre mayor sea el número de enlaces mejor será el proceso de rastreo. Durante este proceso se brinda especial atención a los sitios web nuevos, a los cambios en los sitios web actuales y a los enlaces rotos. El rastreador analiza cada página, descarga su contenido e identifica nuevos enlaces para continuar el proceso de manera recurrente. Se utiliza para realizar el rastreo Nutch de forma distribuida empleando las políticas de selección, re-

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

visita, cortesía y paralelización que permiten realizar un rastreo exhaustivo. En la configuración del rastreador se determina qué sitios rastrear, con qué frecuencia y cuál es el número de páginas que se deben explorar en cada sitio (Google, 2017).

Indexación de la web

Después de realizar el proceso de rastreo se analiza cada página web para identificar los principales elementos para luego almacenar la información y crear un índice de contenidos que permita mejorar el proceso de recuperación de información. En el proceso de indexación se estandariza la información rastreada definiendo los metadatos necesarios para el procesamiento de la información ver figura 3. Posteriormente se procede a generar el gráfico de conocimiento extrayendo de cada página el contenido según el contexto mediante la utilización de una ontología general y otra específica de acuerdo a la categoría de la página web. Como herramientas para realizar el procesamiento de información se emplean Solr y Apache Jena que utilizan diferentes técnicas y algoritmos para extraer el conocimiento implícito de las páginas web.

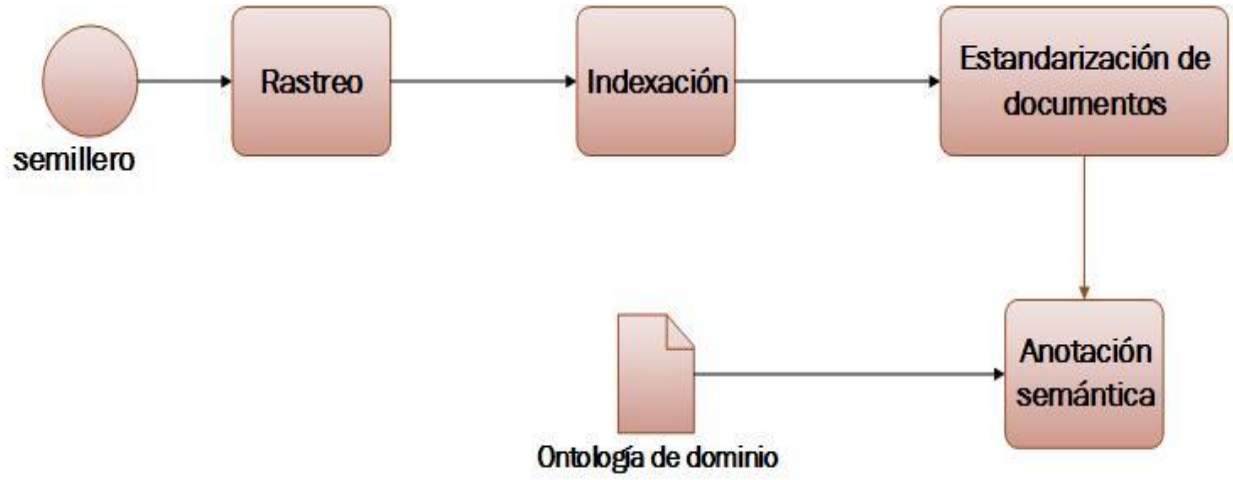

Figura 3: Componente Rastreo-indexación. Fuente: elaboración propia

Solr implementa el modelo de espacio vectorial y utiliza un sistema de fichero invertido para la creación del índice; además para realizar el proceso de normalización dispone de múltiples analizadores y se pueden definir analizadores propios (Montero y Placencia, 2016). Para el razonamiento semántico de la información se utiliza Apache Jena que proporciona una API para leer, escribir, extraer y procesar grafos RDF. También dispone de un motor de inferencia para razonar sobre las ontologías y realizar consultas con especificación SPARQL. Adicionalmente se emplea para la creación del índice el algoritmo CF-IDF (Frecuencia del concepto – frecuencia inversa de documento) en base a las

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

anotaciones realizadas, que según Goossen, et al., (2011) y García (2015) mejora el proceso de recuperación de información.

2.2 Componente de procesamiento

Se encarga de procesar y analizar textos en lenguaje natural asociando cada sentencia de un texto a una representación semántica empleando como base una ontología con miles de palabras, donde las palabras se categorizan según los distintos significados que tienen y donde se definen las relaciones entre ellas. Gruber (1983) define una ontología como "una especificación explícita de una conceptualización" que permiten añadir un sentido a la información que se necesita procesar. Consta de 5 componentes (conceptos, relaciones, funciones, instancias y axiomas) que describen las relaciones de las palabras y le añaden un sentido natural. La utilización de Ontologías posibilita mejorar el procesamiento en lenguaje natural de la consulta realizada por el usuario y la información recopilada por los rastreadores en la web. En la figura 4 se presenta el componente de procesamiento de información y se describe la relación entre cada elemento para mejorar la recuperación de información.

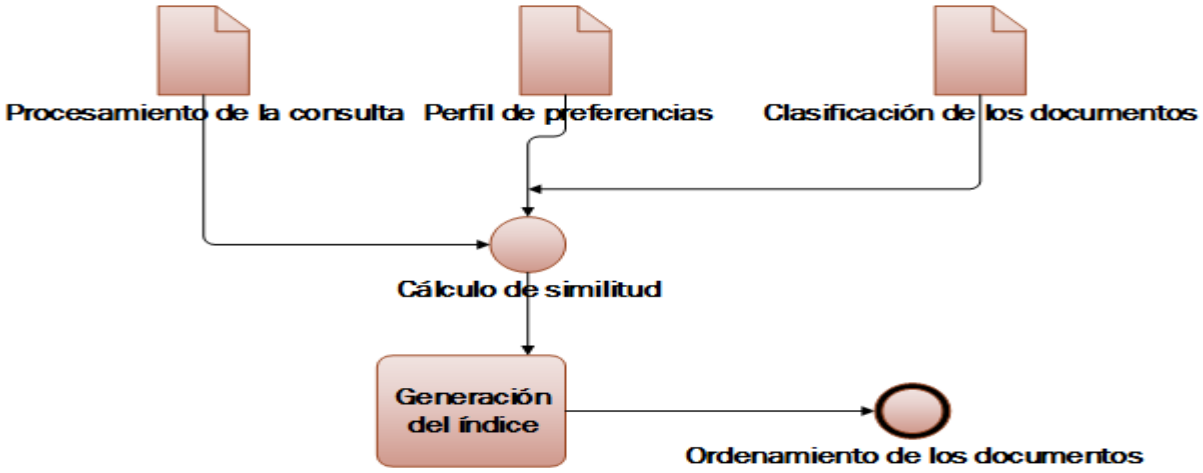

Figura 4: Componente procesamiento de la información. Fuente: elaboración propia.

Procesamiento de la Consulta

Los usuarios cuando acceden a un SRI formulan la consulta en lenguaje natural, mediante la introducción textos cortos de uso popular dificultando identificar el contexto de la pregunta y la necesidad del usuario. Los SRI comparan los términos introducidos en la consulta con la información almacenadas, mediante diferentes técnicas que determinan la frecuencia de aparición de los términos permitiendo seleccionar los documentos más relevantes según el índice de búsqueda creado. Los métodos tradicionales de procesamiento de la consulta dificultan entender la intención detrás de la pregunta realizada por el usuario y limita la capacidad del SRI de recuperar documentos relevantes.

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

Para mejorar el procesamiento de las consultas autores como Segura (2010), Kuna, et al. (2014) proponen realizar la expansión de la consulta. En la expansión de consulta se pueden utilizar tesauros, diccionarios, sistemas expertos, ontologías; que permiten identificar términos similares a los introducidos por el usuario y realizar la desambiguación de los términos. En el proceso de expansión de la consulta se aplican principalmente dos métodos relacionados con técnicas probabilísticas sobre el corpus de búsqueda y el de análisis de colecciones o estructuras de conocimiento.

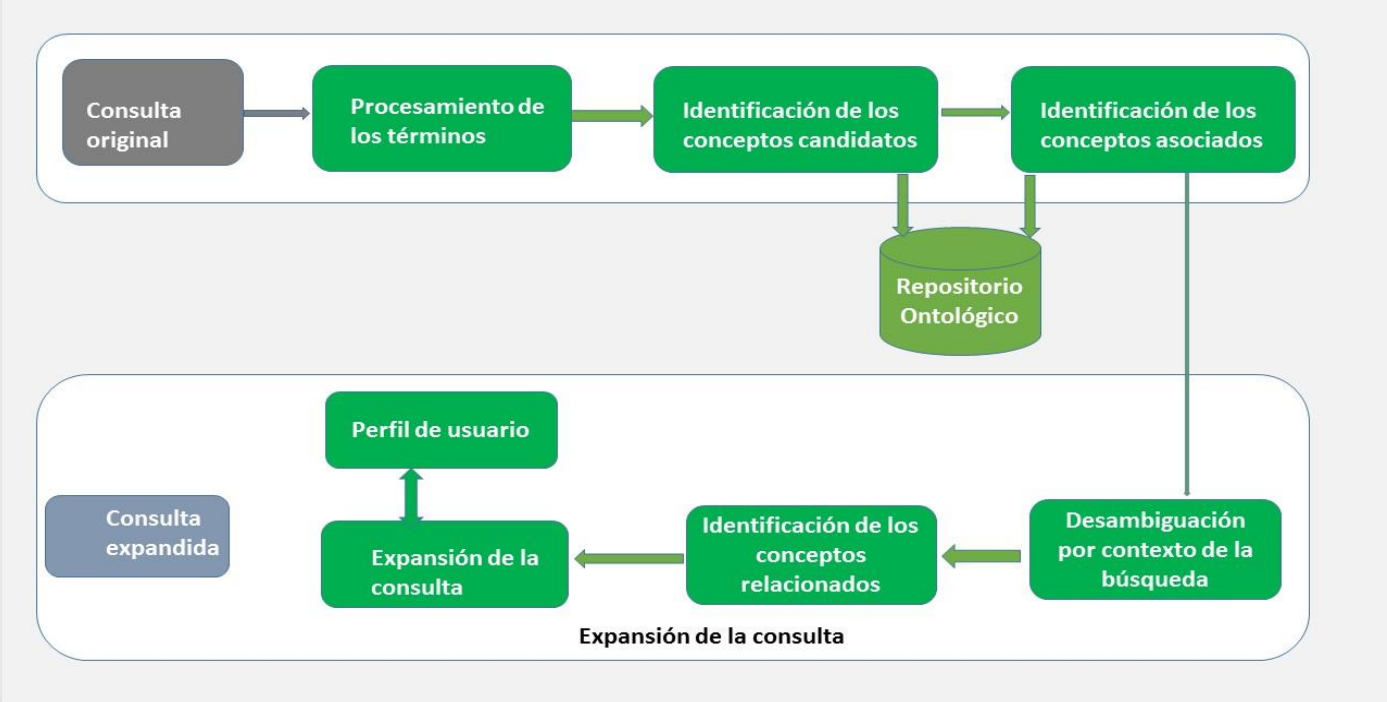

Figura 5: Expansión de consulta. Fuente: elaboración propia.

El procesamiento de la consulta tiene como principal objetivo la desambiguación de los términos introducidos por el usuario generando como salida una tripleta en formato RDF. El procesamiento de la consulta, analiza los términos introducidos, el perfil de preferencia del usuario y determina el contexto y la intención del usuario para que el SRI seleccione los documentos más relevantes.

Procesamiento del perfil de usuario

Permite generar y actualizar el perfil del usuario según sus preferencias implícitas y explícitas utilizando varios elementos (categorías seleccionadas en su perfil, historial de búsqueda y ubicación del usuario) para obtener mejor resultado cuando un usuario realiza una búsqueda. En la figura 6 se describe el procesamiento del perfil de usuario para identificar sus preferencias y mejorar la personalización de los resultados de búsqueda brindados a los usuarios.

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

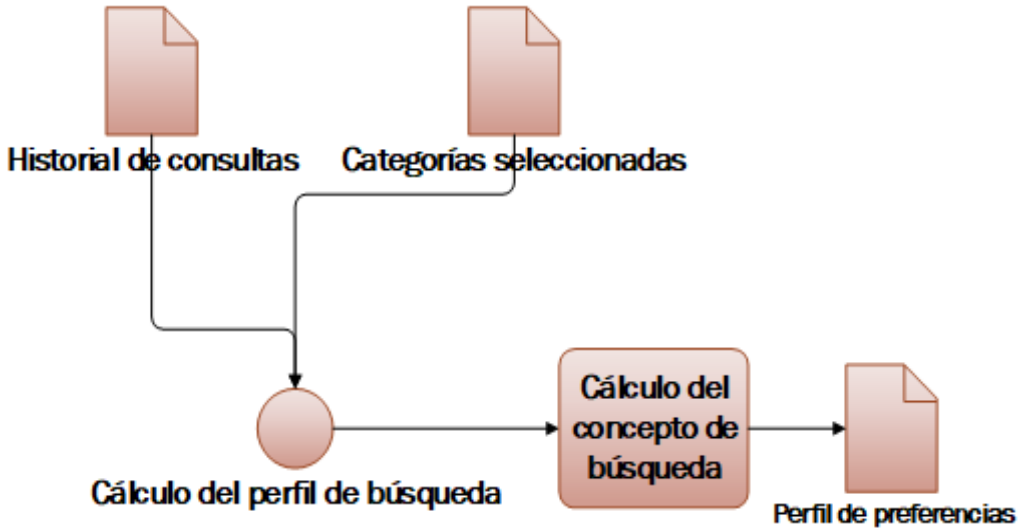

Figura 6: Perfil de preferencias. Fuente: elaboración propia.

Clasificación de los documentos

El mecanismo de clasificación (figura 7) obtiene un documento con las anotaciones realizadas en el componente de rastreo-indexación. Cada documento tiene asociado todos los conceptos con su índice semántico. A continuación, se identifica a que categoría pertenece cada concepto, para determinar el por ciento de relevancia y categorizar el documento.

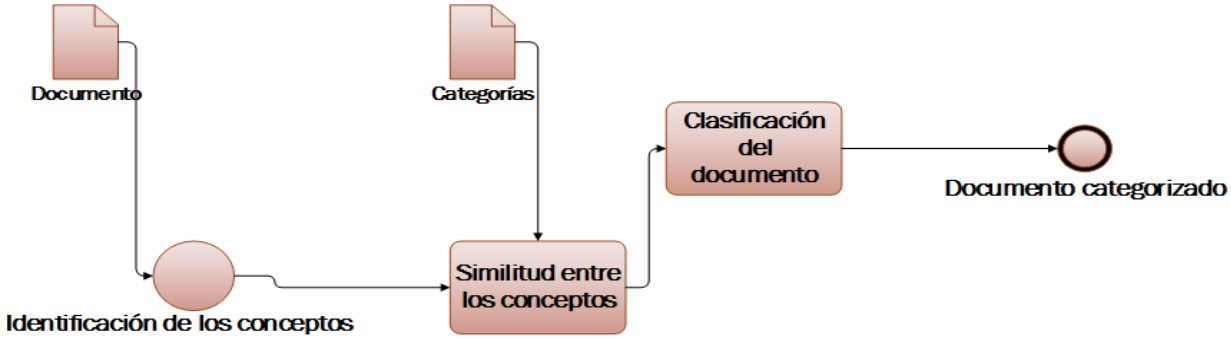

Figura 7: Clasificación de documentos. Fuente: elaboración propia.

Cálculo de la similitud

Para determinar la similitud entre la consulta realizada por el usuario y la información indexada en el buscador se utilizan los resultados del procesamiento de la consulta, del procesamiento del perfil de usuario y el índice de relevancia de la anotación semántica realizada durante el proceso de almacenamiento de la información. La similitud se determina utilizando el algoritmo de Levenshtein para textos cortos y la función del coseno.

Revista Cubana de Ciencias Informáticas
Vol. 12, No. 1, Enero-Marzo, 2018
ISSN: 2227-1899 | RNPS: 2301
Pág. 102-116
http://rcci.uci.cu

Cálculo de la relevancia

Después de obtener la similitud semántica se procede a calcular la relevancia para mostrar la información más relevante para el usuario. En este proceso se utiliza el algoritmo propuesto en Baquerizo (2017) que determina el coeficiente de relevancia según el perfil de usuario, la consulta y el índice de similitud semántico. El coeficiente de relevancia obtenido se utiliza para ordenar los resultados y mostrar un número de elementos que puedan ser asimilados por el usuario.

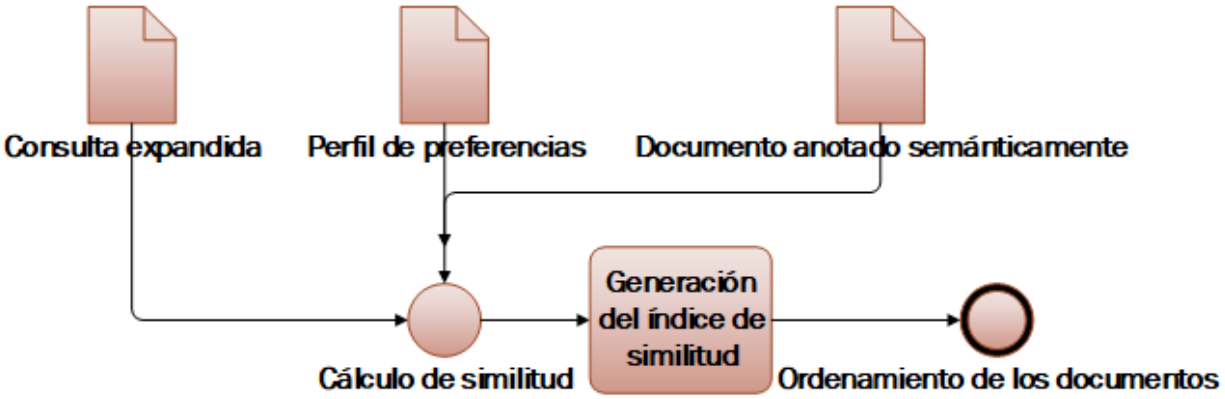

Figura 8: Cálculo de la relevancia. Fuente: elaboración propia.

Componente de presentación

Empleando las técnicas de experiencia de usuario se diseña la interfaz del sistema donde el usuario puede realizar la consulta y obtener los resultados. El sistema de recuperación de información dispone de una búsqueda simple y una avanzada que cumplen con los principios de diseño centrado en el usuario.

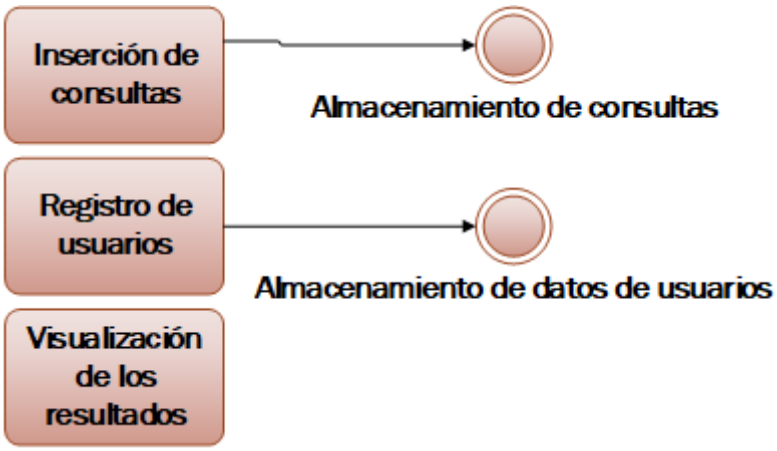

Figura 9: Presentación de la información. Fuente: elaboración propia.

En la búsqueda simple el usuario introduce la pregunta y se le muestran los resultados más relevantes. La búsqueda avanzada le permite al usuario un nivel mayor de personalización de los resultados utilizando alguno de los siguientes filtros:

- Con alguna de las palabras: devuelve resultados que contenga una o algunas de las palabras del criterio de búsqueda.
- Con todas las palabras: devuelva resultados que contengan específicamente todas las palabras del criterio.
- Con la frase exacta: devuelve resultados que contengan específicamente la frase exacta introducida en el criterio de búsqueda.
- Sitio: permite buscar resultados definiendo el sitio web o el dominio.

Resultados y discusión

En la evaluación del modelo propuesto se utilizaron las métricas de Precisión (P) y Exhaustividad (E) que permiten comprobar la calidad de los resultados obtenidos. Para la validación se diseñó un experimento sobre la información publicada en la web cubana. En el experimento se analizó el resultado brindados a las preguntas formulados por los usuarios utilizando un SRI sin procesamiento semántico y el modelo propuesto.

Tabla 1: Resultado de aplicar P y E antes y después de implementar el modelo. Fuente: elaboración propia.

P sin modelo	P con el modelo	E sin modelo	E con el modelo
0,29	0.94	0,24	0.90

Los valores de precisión obtenidos fueron aceptables, corroborando que la recuperación de información con anotación semántica mejora la recuperación de información. Adicionalmente se realizó una consulta de expertos donde la concordancia demostró un nivel alto de satisfacción con la aplicación del modelo propuesto.

La evaluación utilizando las métricas y la consulta a los expertos demuestra la calidad, pertinencia y relevancia de la recuperación de información con anotación semántica. Permitiendo ajustar los resultados más relevantes a las necesidades del usuario, incrementando su experiencia en la utilización de sistemas de recuperación de información semántico.

Conclusiones

- El análisis sobre el proceso de recuperación de información permitió identificar como principales deficiencias la sobre carga de información, la heterogeneidad de fuentes de información y la interoperabilidad que dificultan en gran medida el procesamiento adecuado de la información disponible.

- La utilización de un componente para el rastreo-indexación, procesamiento y presentación de la información permitió recuperar información relevante para los usuarios.
- El cálculo de la relevancia utilizando la similitud semántica permite mejorar el proceso de recuperación de información.
- La validación del modelo utilizando las métricas de Precisión y Exhaustividad y la consulta a expertos permite comprobar la calidad de los resultados obtenidos.

Referencias

Baquerizo, R. P., et al. Algorithm for calculating relevance of documents in information retrieval systems. International Research Journal of Engineering and Technology. 2017, 4(3). pp. 1-5.

Berners-Lee, T. et al. "The semantic web," Scientific american, vol. 284, no. 5, pp. 28-37, 2001

Deco, C.; Reyes, N. y Bender, C: Recuperación de Información en Bases de datos no estructuradas, XIV Workshop de Investigadores en Ciencias de la Computación, 2012

García Moreno, C. "Desarrollo de un modelo para la gestión de la I+D+i soportado por tecnologías de la Web Semántica" ,2015.

Gonzalo, C.; Codina, L., *et al*. Recuperación de información centrada en el usuario y SEO: Categorización y determinación de las intenciones de búsqueda en la Web. [Consultado el: 15 de enero de 2017] Disponible en: http://journals.sfu.ca/indexcomunicacion/index.php/indexcomunicacion/article/download/197/175

Goossen, Frank, et al. News personalization using the CF-IDF semantic recommender. En Proceedings of the International Conference on Web Intelligence, Mining and Semantics. ACM, 2011. p. 10.

Gruber, .T. R. "A Translation Approach to Portable Ontology Specifications". Knowledge Acquisition, 5(2), 1993. pp.199-220.

Kuna, H.; Rey, M.; Podkowa, L.; Martini, E. y Solonezen, L.: "Expansión de Consultas basada en Ontologías para un Sistema de Recuperación de Información", XVI Workshop de Investigadores en Ciencias de la Computación, pp. 500 - 505, 2014.

Martínez Méndez, F. J. Recuperación de información: modelos, sistemas y evaluación. Murcia, KIOSKO JMC, 2004. 106 p.

Martínez-Fernández,J. L. et al. Búsqueda semántica a través del Procesamiento de Lenguaje Natural, 2010 p. 2-3.

Montero Puñales, E. M. y Placencia Salgueiro, A. Sistema de recuperación y análisis de información para investigadores del Instituto Investigativo ICIMAF. INFO 2016, 2016, p 2-15.

Redondo, S. ¿Qué es la búsqueda semántica y por qué me debe importar? [Consultado el: 15 de marzo de 2017] Disponible en: http://www.senormunoz.es/SEO-MARBELLA/que-es-la-busqueda-semantica-y-por-que-me-debe-importar

Rodríguez García, M. A., et al. Creating a semantically-enhanced cloud services environment through ontology evolution. Future Generations in Computer Systems, 32, 2014, p 295–306.

Salton, G. y Mcgill, M. Introduction to Modern Information Retrieval. McGraw-Hill, Inc., 1983.

Segura Navarrete A., A.: "Aplicaciones de la expansión de consultas basada en Ontologias de Dominio a la Búsqueda de Objetos de Aprendizaje en Repositorios", Tesis de doctorado, Universidad de Álcala de Henares, 2010.

Vuotto, A.; Bogetti, C. y Fernández, G. Application of TF-IDF factor in the semantic analysis of a documentary collection, biblios, 2015, vol 60, p. 1-13.