

Marco de trabajo para búsqueda efectiva de imágenes

Clenyr Rivas¹, Francisco Luengo^{1} y Gerardo Pirela-Morillo²*

¹Laboratorio de Simulación Gráfica. ²Laboratorio de Lenguajes y Modelos Computacionales.
Universidad del Zulia. Maracaibo, Venezuela

Recibido: 10-06-2011 Aceptado: 10-10-2012

Resumen

Internet se ha convertido en un inmenso repositorio de información sobre diversos temas. Dicha información es almacenada en un formato difícil de analizar por los computadores, lo cual genera muchos problemas en lo que se refiere a la búsqueda y recuperación de un contenido específico. Esto ha motivado el surgimiento de la Web Semántica, la cual pretende mejorar la funcionalidad de la Internet; convirtiéndose en el centro de muchas investigaciones en el mundo. En tal sentido, el presente artículo se enfoca en el desarrollo de un marco de trabajo para mejorar la efectividad de las búsquedas y recuperación de imágenes a través de consultas hechas en lenguaje natural; para lo cual se introducen dos herramientas, un editor de metadatos para imágenes y un buscador de imágenes que procesa consultas hechas en lenguaje natural. El primero permite agregar metadatos descriptivos del contenido a las imágenes, para que el segundo pueda ser capaz de procesar, analizar y realizar deducciones lógicas a partir de lo expresado por el usuario, y al final mostrar las imágenes que mejor se ajusten a la consulta hecha; todo basado en una ontología especialmente desarrollada para este fin.

Palabras clave: búsquedas, imágenes, metadatos, ontología, web semántica.

Framework for effective retrieval of images

Abstract

Internet has become a huge repository of information about many diverse themes. Such information is stored in heterogeneous formats which makes it difficult for computers to analyze and in turn creates problems with regards to searching and retrieving specific contents information. This has motivated the emergence of the Semantic Web which aims to improve and enhance Internet's functionality; turning into the subject of many a research effort around the world. In that respect, we have developed a framework to improve effectiveness in image web searching and retrieval via natural language queries. The framework is made of two independent tools, an image metadata editor and an image search engine with natural language queries processing capabilities. The former allows edition of descriptive metadata unto image files based upon an ontology especially created for this retrieval task so that the latter is able to process, analyze, and carry out logical deductions from the user's query to finally retrieve and show the images that best match the user's search intention.

Keywords: search, images, metadata, ontology, semantic web.

* Autor para la correspondencia: fluengo@fec.luz.edu.ve

Introducción

Desde sus inicios, Internet ha revolucionado al mundo al ofrecerle a los usuarios diversos servicios que facilitan la comunicación, el entretenimiento, el acceso a la gran cantidad de información y muchas otras bondades propias de este medio. Los motores de búsqueda están entre las herramientas más populares y usadas de la Internet. Ellos usualmente se basan en palabras claves proporcionadas por el usuario, que intentan ser identificadas dentro de los documentos objeto de las búsquedas con el propósito de ofrecer como resultado una lista de enlaces a los sitios que albergan los referidos documentos.

Los buscadores han facilitado la localización y recuperación de información en la Web, pero no son del todo efectivos, hecho que se manifiesta por la enorme cantidad de resultados obtenidos no relacionados con el objeto de la búsqueda. Esta realidad se hace más evidente aun cuando se trata de encontrar imágenes en internet.

Los archivos de imagen son un recurso muy popular y de gran demanda en la web; tanto que buscadores como Google, Yahoo o Bing, incorporan un enlace especial en su portal principal para la búsqueda específica de dicho recurso.

Con el propósito de obtener resultados más efectivos en cuanto a la localización de imágenes, esta investigación propone una estrategia basada en la incorporación de metadatos ontológicos a los archivos de las imágenes, sin alterar su estructura, junto con el análisis semántico de las consultas hechas por el usuario en lenguaje natural; logrando con ello una correspondencia más exacta entre la intención de búsqueda y la imagen o imágenes requeridas.

Web Semántica

Desde sus inicios, la World Wide Web fue desarrollada para el intercambio de información a través de Internet, y para ello se

creó el HTML (HyperText Markup Language), el cual permite adecuar el aspecto visual e incluir objetos multimedia (imágenes, videos, etc.) en los documentos mostrados en la Web. Sin embargo, dicho lenguaje ofrece pocas posibilidades para dotar el material que se publica en internet con información que ayude a describir su contenido.

Según el padre de la Web, Berners-Lee (1), la Web Semántica surge como una propuesta para superar las limitaciones de la Web actual mediante la incorporación de conocimiento semántico explícito que describa y/o estructure la información y servicios disponibles en internet, de forma susceptible a ser procesada automáticamente por un programa de computadora. Para tal fin, se emplean lenguajes de marcado como XML (Extensible Markup Language), XML Schema, RDF (Resource Description Framework), RDF Schema y OWL (Web Ontology Language); los cuales intentan dotar a la información existente en Internet de lógica y significado.

Recuperación de la información

La información en la Web se caracteriza por el dinamismo; es decir, cambios continuos en su contenido. Al contrario que en los medios tradicionales, la redundancia o publicación repetida de información es una constante en internet, por lo que el problema de la recuperación de información se agrava.

La Recuperación de Información aborda la búsqueda de documentos, imágenes, videos, u otros recursos, ya sea de internet o de cualquier otro medio. Actualmente, los métodos más rápidos y eficaces de recuperación de información en la Web son a través de los buscadores. Los principales tipos de buscadores son los de índices temáticos (los cuales manejan catálogos o directorios como Yahoo) y los buscadores por contenido (o motores de búsqueda, los cuales consultan bases de datos, a partir de palabras claves, con referencias a los sitios web y sus contenidos, como Google). Sin embargo, la cantidad de

resultados irrelevantes que los buscadores arrojan puede resultar frustrante.

La Web Semántica permitiría que cualquier usuario pudiera encontrar en Internet respuestas de forma más rápida y sencilla, gracias a una información mejor definida.

Recuperación de imágenes

Para el caso particular de la recuperación de imágenes (2), los motores de búsqueda convencionales suelen emplear el sistema de recuperación basado en el contexto; es decir, se apoyan en las coincidencias con el texto del entorno del documento en el que se encuentran. Una vez seleccionadas, éstas son ordenadas según su relevancia. Entre los factores que permiten determinar la pertinencia de una imagen para una consulta están: cantidad de sitios que contienen alguna otra idéntica o similar, su tamaño, relación entre sus vínculos, número de veces que aparece dentro de un sitio, ponderación del texto por su distancia a ella, el nombre de su archivo, entre otros.

Por otro lado, la recuperación basada en el contenido puede considerar las características visuales de la imagen (color, forma, etc.) o las características no visuales (descriptores libres almacenados como etiquetas dentro de su formato). Algunos trabajos relacionados son (3-6).

Metodología

El propósito principal de este trabajo es definir un marco para la búsqueda efectiva de imágenes, el cual entra en la categoría de las técnicas basadas en contenido. Considerando todo lo que implica el proceso de recuperación de información (7); esta sección aborda los pasos seguidos para la definición de dicho marco.

Definición de una Ontología

Uno de los ejes principales de la Web Semántica es la noción de ontología (8) como herramienta clave para alcanzar un entendimiento entre las partes (usuarios, desa-

rolladores, programas) que participan de este conocimiento común.

Para la definición de una ontología que permitiera describir el contenido de una imagen, se siguió la metodología Methontology (9); y para validar que no hubiesen inconsistencias en su implementación se empleó Protégé (10) bajo el lenguaje OWL-DL y con el uso del razonador lógico Pellet.

Como resultado, y comprendiendo la complejidad implícita en la elaboración de una ontología para la descripción del contenido de cualquier imagen, se definió una que intenta categorizar de manera general diversidad de contenido. La ontología definida consta de alrededor de 700 clases, 320 atributos y 320 relaciones, además de un diccionario de conceptos; definidos en una Base de Conocimiento desarrollada en lenguaje Prolog (11). Una parte de dicha Base de Conocimiento se muestra a continuación.

```

%%% Hechos %%%
Clase(imagen) .
Subclase(arte,imagen) .
Subclase(ciencia,imagen) .
Subclase(entretencion,imagen) .
Subclase(farandula,imagen) .
Subclase(imagenes_3d,imagen) .
Subclase(imagenes_abstractas,imagen) .
...
conjugacion(comer,como) .
conjugacion(comer,comes) .
conjugacion(comer,come) .
conjugacion(comer,comemos) .
conjugacion(comer,comen) .

%%%Relaciones %%%
rel(provenir,carnes,animales) .
rel(comer,herbivoros,plantas) .
rel(comer,omnivoros,carnes) .
rel(comer,omnivoros,plantas) .
...

```

```

% Reglas %
comen_animales(X):-rel(V,X,Y),rel
(provenir,Y,Z).

camino(Ini,Fin,Camino:
-camino(Ini,Fin,Camino,[Ini]).
camino(I, F, [I,F], V):-subclase(F,I).
camino(I, F, [I|T], V):-subclase(Z,I),
camino(Z,F,T,[Z|V]).

```

Con esta ontología se consigue una taxonomía general para la clasificación de imágenes que se puede ir particularizando según la especificidad del contenido que se desee describir. Un extracto de la misma en formato OWL es mostrado a continuación.

```

<owl:ObjectProperty rdf:about="#son_comidos_por">
<owl:inverseOf>
  <owl:ObjectProperty rdf:about="#comen"/>
</owl:inverseOf>
<rdfs:range>
  <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
      <owl:Class rdf:about="#Carnivoros"/>
      <owl:Class rdf:about="#Herbivoros"/>
    </owl:unionOf>
  </owl:Class>
</rdfs:range>
<rdfs:domain>
  <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
      <owl:Class rdf:about="#Animales"/>
      <owl:Class rdf:about="#Planta"/>
    </owl:unionOf>
  </owl:Class>
</rdfs:domain>
</owl:ObjectProperty>

```

Editor de imágenes

El siguiente paso consiste en añadirle a las imágenes la descripción ontológica de su

contenido; para lo cual se desarrolló un editor (figura 1).

Este editor permite leer archivos en formato de imágenes, a los cuales se les puede añadir o modificar sus metadatos. La información agregada no afecta la visualización de las mismas por otras aplicaciones. Los metadatos a añadir son seleccionados de una estructura tipo árbol que representa la taxonomía de clases definida en la ontología (panel izquierdo del editor, figura 1), y la cual es previamente cargada por el editor desde un archivo OWL. Estos metadatos son mostrados en el panel inferior del editor, y actualizados a medida que se agregan, eliminan o modifican.

Adicionalmente, es posible agregar un nivel adicional de información a la clases seleccionadas, de manera de describirlas aun mejor, en caso de no contemplarse en la ontología (caja de texto bajo lo imagen cargada; figura 1).

Estrategia de búsqueda

Para la efectiva identificación de una imagen se consideran dos aspectos fundamentales: la correcta identificación de la intención de búsqueda del usuario (saber con certeza lo que se quiere encontrar), y que el objeto de la búsqueda (en este caso, la imagen) proporcione la suficiente información para identificarlo inequívocamente.

El segundo aspecto lo solventamos con la ontología discutida previamente; almacenando dentro del formato de la imagen contenido mas descriptivo sobre ella, en forma de metadatos.

Con respecto al primer aspecto, más allá de consultas empleando palabras claves, se parte de consultas en lenguaje natural que aporten mayor información sobre la intención de búsqueda del usuario.

Para el análisis de las consultas se empleó un analizador de gramática de cláusulas definidas (12, 13), DCG por sus siglas en ingles, el cual verifica que las consultas hechas por el usuario, en lenguaje natural, co-

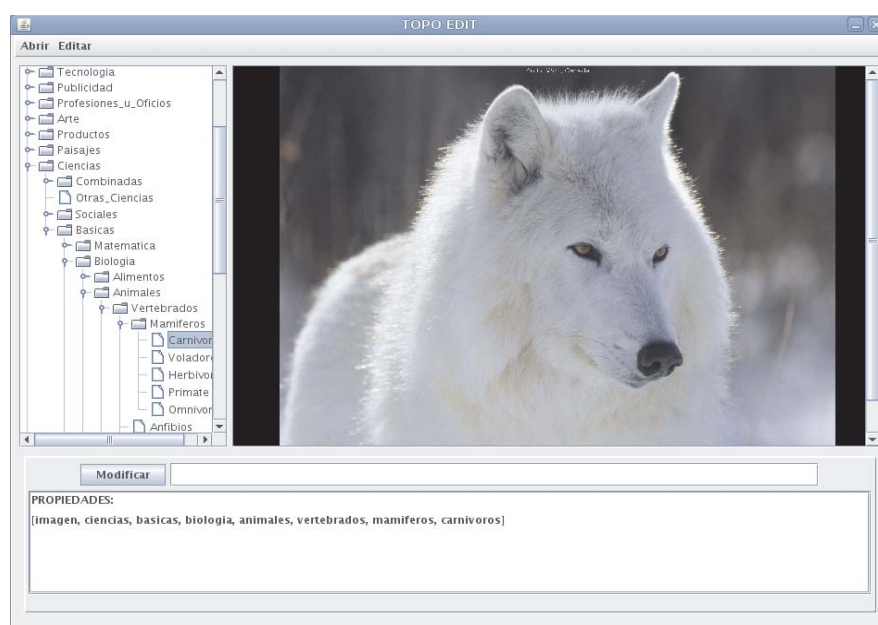


Figura 1. Editor de Imágenes.

respondan con alguna de las reglas gramaticales definidas y con el vocabulario de la ontología. Las reglas gramaticales determinan la estructura de las oraciones, y fueron previamente estructuradas y almacenadas en la Base de Conocimiento. Estas reglas permiten un mejor entendimiento de lo que el usuario intenta transmitir.

Luego de que la consulta es procesada y reconocida, el analizador DCG devuelve una lista de las palabras que conforman la frase con sus respectivas características. Con esta lista, se forman nuevas consultas hechas a la Base de Conocimiento de la Ontología, y los resultados obtenidos son las mejores aproximaciones a las categorías definidas en la taxonomía, las cuales son comparadas con la información contenida en las imágenes; según el porcentaje de coincidencia, las imágenes son devueltas para ser mostradas. Este proceso es ilustrado en la figura 2.

Resultados

Para la implementación de la estrategia de búsqueda descrita en la sección anterior se desarrolló una aplicación, hecha en Java,

la cual se llamó TOPO, con la misma apariencia del buscador Google, por simplicidad (figura 3). Java, como lenguaje de programación de alto nivel, ofrece una rica gama de librerías entre las que se encuentran aquellas para trabajar con ontologías (14) y conexión con Prolog (15).

Para el diseño de las pruebas, se agregaron metadatos a un conjunto de imágenes de diversa naturaleza (animales, personas, plantas, vehículos, dibujos animados, objetos, etc.), creando de esta forma un repositorio inicial de alrededor de 150 imágenes. Adicionalmente, se estructuraron tres niveles de consultas. Dichas consultas están definidas dentro del rango abarcado por la ontología y el análisis semántico definido para esta investigación. El primer nivel de consulta está asociada a palabras específicas (claves), y los otros dos a frases cada vez más elaboradas que describen la intención de búsqueda de manera más detallada. Un ejemplo de estos niveles de consultas son:

Consulta Nivel 1: “animales”

Consulta Nivel 2: “animales carnívoros”

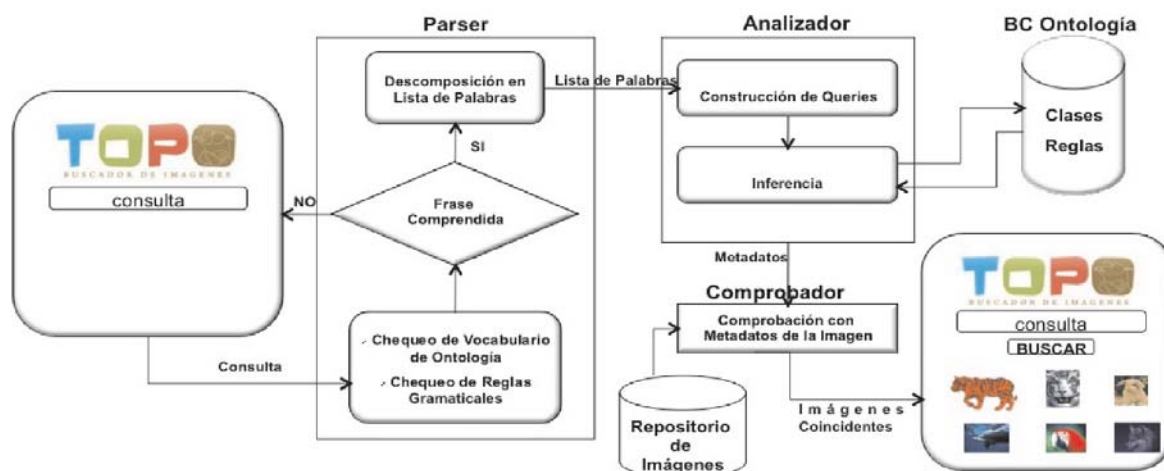


Figura 2. Proceso de Búsqueda de Imágenes.

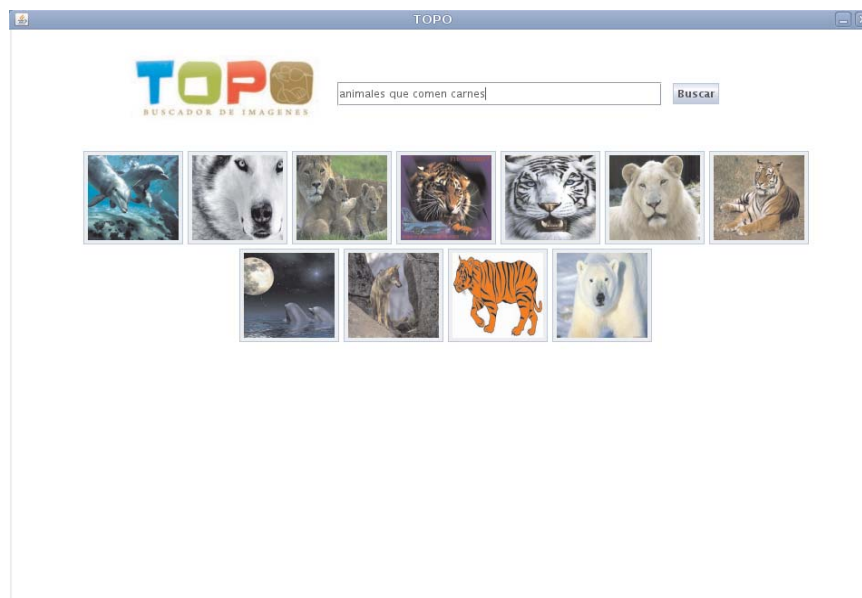


Figura 3. Buscador Semántico TOPO.

Consulta Nivel 3: “animales que comen carnes”

El propósito de estos niveles es ir reduciendo el espacio de resultados al ofrecer mayor información en la consulta sobre lo que se pretende encontrar.

La importancia de la correcta interpretación de la intensidad de búsqueda del

usuario es mejor entendida con la consulta nivel 3 del ejemplo anterior. Si dicha consulta es hecha en un buscador convencional el cual interpreta la consulta como un conjunto de palabras claves a las que debe buscar coincidencia, es casi seguro que entre las imágenes resultantes de la búsqueda, además de diversos animales, se encuentren imágenes de carnes. Por el contrario, ha-

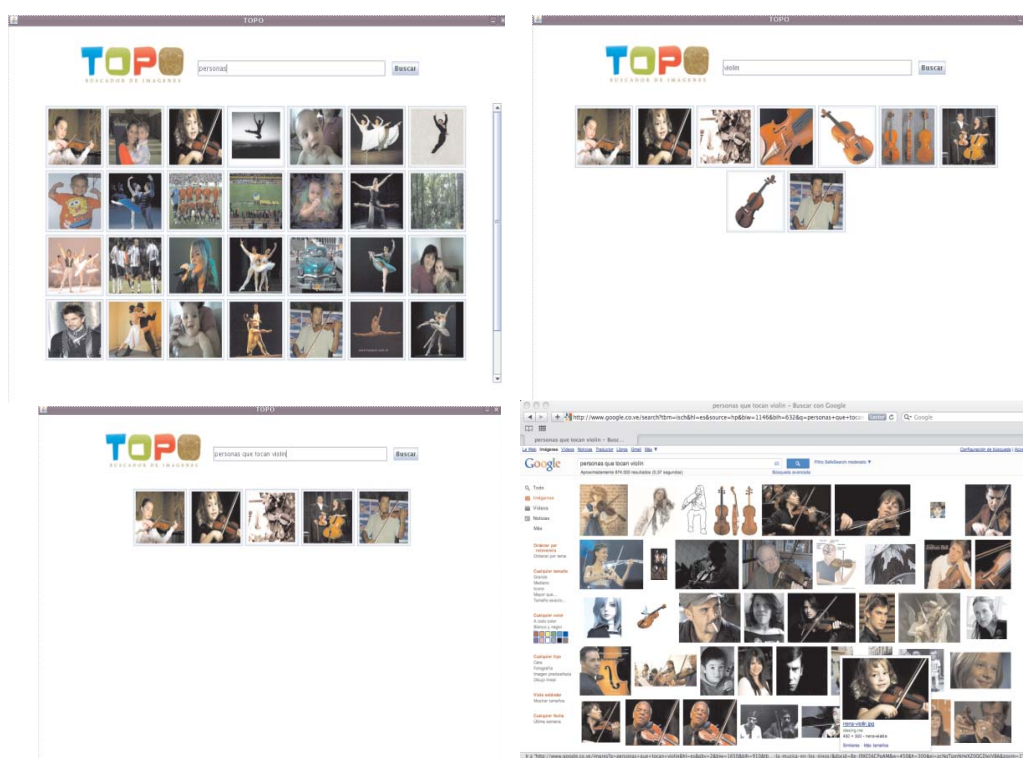


Figura 4. Resultado para las consultas: “personas”, “violin” y “personas que tocan violin”, en TOPO, y “personas que tocan violin” en Google; respectivamente.

ciendo un análisis adecuado de la frase ingresada por el usuario, el buscador TOPO solo devuelve imágenes de animales carnívoros (figura 3), a pesar de poseer en su repositorio, imágenes con descripciones específicas de carnes.

Otros resultados similares se obtuvieron con consultas como: “personas que tocan violin”, en la que esquemas de búsqueda convencionales devuelven no sólo personas tocando el violín, sino también imágenes de violines o simplemente personas; mientras que TOPO retornó sólo específicamente lo que se pedía. La figura 4 ilustra lo comentado.

Conclusiones

En este artículo se presentó un marco de trabajo para la búsqueda efectiva de imágenes, el cual se centra en la correcta inter-

pretación de la intención de búsqueda del usuario, y en dotar a las imágenes con información descriptiva sobre su contenido. Para ello se definió una ontología, a partir de la cual se crean y almacenan en la propia imagen metadatos sobre su contenido. Adicionalmente, se emplea un analizador DCG para la revisión gramatical de la consulta, la que luego es procesada para generar nuevas consultas a la base de conocimiento de la ontología, y de esta forma obtener los datos de la taxonomía que el usuario espera encontrar en las imágenes buscadas.

Este proceso demuestra ser más efectivo que los métodos convencionales en cuanto a la relevancia de los resultados obtenidos. Con consultas que expresen más claramente lo que se está buscando, si estas son “entendidas”, es de esperar respuestas acordes; mientras que si se descompone la consulta en palabras claves, es de esperar re-

sultados más diversos. Por otro lado, la cantidad de procesamiento del esquema propuesto es mayor. Sin embargo, bajo este enfoque, el tipo de consultas abarca desde palabras claves hasta frases más elaboradas, pudiendo proporcionar más información que restrinja el espacio de resultados.

El principal reto para esta estrategia es el desarrollo de una base de conocimientos ontológica que describa completamente el contenido de cualquier imagen. Aunque para dominios más restringidos, su construcción es más cómoda de realizar, con la consecuente mejora en los resultados.

Referencias bibliográficas

1. BERNERS-LEE T., HENDLER J., LASSILA O. "The Semantic Web". *Scientific American Magazine* May 2001.
2. HOI S., LUO J., BOLL S., XU D., JIN R., KING I. *Social Media Modeling and Computing* Springer. 2011.
3. MEZARIS V., KOMPATSIARIS I., STRINTZIS M.G. "An ontology approach to object-based image retrieval". *IEEE International Conference on Image Processing ICIP 2003* pp II 511-514, vol 3. Sep 2003.
4. HANBURY A. "A survey of methods for image annotation". *J Visual Lang Comput* Vol. 19, issue 5, pp. 617-627. 2008.
5. ENSER P. "The evolution of visual information retrieval". *J Inf Sci* Vol 34, no. 4, pp. 531-546. 2008.
6. LEE H.J. "A new model for semantic photograph description combining basic levels and user-assigned descriptors". *J Inf Sci* Vol. 36, No. 5, pp. 547-565. 2010.
7. BELKIN N.J., CROFT W.B. "Information filtering and information retrieval: Two sides of the same coin". *Communications of the ACM* Vol. 35, No. 12: pp. 29-38. 1992.
8. USCHOLD M., GRUNINGER M. "Ontologies: principles, methods and applications". *Knowl Eng Rev* 11: 93-136. 1996.
9. CORCHO O., FERNANDEZ-LOPEZ M., GOMEZ-PEREZ A. "Methodologies, tools and languages for building ontologies. Where is their meeting point?". *Data Knowl Eng* Vol. 46, Issue 1, pp. 41-64. 2003.
10. KNUBLAUCH H., HORRIDGE M., MUSEN M., RECTOR A., STEVENS R., DRUMMOND N., LORD P., NOY N. F., SEIDENBERG J., WANG H. "The Protege OWL experience". *OWL Experiences and Directions (OWLED 2005)* Galway, Ireland. 2005.
11. KOWALSKI, R. "Prolog as a Logic Programming Language". *AICA Congress*. Pavia, Italy. pp. 1029-1034. 1981.
12. PEREIRA F., WARREN D. "Definite clause grammars for language analysis—A survey of the formalism and a comparison with augmented transition networks". *Artif Intell*. Vol. 13, Issue 3. pp. 231-278. 1980.
13. VAN ROY P. "A useful extension to Prolog's definite clause grammar notation". *ACM SIGPLAN Notices* Vol. 24. Issue 11. 1989.
14. Matthew Horridge, Sean Bechhofer. "The OWL API: A Java API for Working with OWL 2 Ontologies" OWL Experiences and Directions (OWLED 2009), Washington D.C., USA.
15. JIProlog. <http://www.ugosweb.com/jiprolog/index.aspx>. 09-2011.