

ARTÍCULO

PROPUESTA DE CONSTRUCCIÓN Y OPTIMIZACIÓN DE UN BUSCADOR SEMÁNTICO

Gerson Villa González, Luis Manuel Hernández Simón, Agustín Gutiérrez
Tornes, Uwe Villa González

Propuesta de Construcción y optimización de un buscador semántico

Resumen

Muchas personas, entidades públicas y en especial entidades privadas, están tratando de obtener el mayor provecho en la utilización en la infraestructura tecnológica de la información y comunicaciones, adquirida no sólo como una herramienta para el desarrollo de procesos y actividades en el desempeño cotidiano de sus labores, sino como una oportunidad para la construcción de conocimiento a través del aprendizaje colaborativo.

El presente artículo enfatiza el uso de las tecnologías de la Web Semántica, como lo es la construcción de un buscador semántico para el descubrimiento del conocimiento, a fin de contribuir en la toma de decisiones tácticas y estratégicas en una organización, proporcionando un sentido automatizado para la generación de conocimiento.

Además, se describe brevemente cómo temas relacionados con la taxonomía, ontología y folksonomía contribuyen de forma acertada a la generación de nuevo conocimiento. Se consideran varios elementos, como: qué son, quién las utiliza, qué beneficios tienen y qué oportunidades brindan a los usuarios de la web.

Palabras clave: semántica, ontología, arquitectura, conocimiento, taxonomía e hipótesis.

Construction Proposal and semantic search engine optimization

Abstract

Many people, especially public and private entities are trying to get the most benefit in using the information technology infrastructure and communications, acquired not only as a tool for the development of processes and activities in their daily performance work, but as an opportunity to construct knowledge through collaborative learning.

This article emphasizes the use of Semantic Web technologies such as building a semantic search engine for the discovery of knowledge, to contribute in making tactical and strategic decisions in an organization, providing automated way for knowledge generation.

You can briefly describe as issues related to taxonomy, ontology and folksonomy rightly contribute to the generation of new knowledge. Several elements were considered as they are, who uses them, what benefits and opportunities are provided to web users.

Keywords: semantic, ontology, architecture, knowledge, taxonomy, hypothesis

Introducción

En los últimos años ha existido un gran crecimiento en nuestras capacidades de generar y coleccionar datos, debido básicamente al gran poder de procesamiento de las máquinas y por su bajo costo de almacenamiento.

Sin embargo, dentro de estas enormes masas de datos existe una gran cantidad de información "oculta", de gran importancia estratégica, a la que no se puede acceder por las técnicas clásicas de recuperación de la información. El descubrimiento de esta información oculta es posible gracias a la minería de datos, la teoría de agentes y los algoritmos genéticos, que entre otras sofisticadas técnicas aplicamos a la inteligencia artificial para encontrar patrones y relaciones dentro de los datos, permitiendo la creación de modelos, los cuales dan un significado a los patrones encontrados.

Así, el valor real de los datos reside en la información que se puede extraer de ellos, información que ayude a tomar decisiones o mejorar nuestra comprensión de los fenómenos que nos rodean. Hoy, más que nunca, los métodos analíticos avanzados son el arma secreta de muchos negocios exitosos hoy en día, como sucede con EBay y Amazon.

Empleando métodos analíticos avanzados para la explotación de datos, los negocios incrementan sus ganancias, maximizan la eficiencia operativa, reducen costos y mejoran la satisfacción del cliente.

Además, la generación de conocimiento en las organizaciones ha cambiado drásticamente, como resultado del surgimiento de nuevos paradigmas asociados a la denominada sociedad de la información y a una nueva economía, basada en el conocimiento. [1]

La Web actual trabaja por medio del léxico, mientras que la Web 2.0, la Web 3.0 y la Web 4.0, que son las propuestas del futuro, se les ha dotado de más significado y semántica.

El desarrollo de las Tecnologías de la Información y la Comunicación (TIC) modifican a la sociedad, y si el medio se transforma, también cambia el acceso de la información, que hoy es mucho más rápido, amplio y preciso.

Los conceptos de taxonomía, ontología y folksonomía se toman como base para la construcción de conocimiento, tratando de indicar, en un principio, cómo se definen, quién los usa y qué beneficios y oportunidades representan para los usuarios y colaboradores de las organizaciones, además de cómo, desde la clasificación y organización de la información, se empiezan a construir los mecanismos generales para la gestión integral de los procesos de negocio. [2]

Taxonomía

Una taxonomía es un tipo de vocabulario controlado en que todos los términos están conectados mediante algún modelo estructural (jerárquico, arbóreo, facetado, etcétera) y especialmente orientado a los sistemas de navegación, organización y búsqueda de contenidos de los sitios Web. [3]

La base para el desarrollo de una buena taxonomía son las características taxonómicas, que son las propiedades o atributos de los objetos que se categorizan, y deben satisfacer ciertos requisitos. Estos son:

- **Objetividad:** cuando la característica se identifica sobre la base de un conocimiento objetivo exclusivamente.
- **Determinación:** cuando existe un proceso claro que pueda seguirse para identificar la característica.
- **Reproducibilidad:** cuando varias personas describen de forma independiente las características de un mismo objeto y coinciden con el valor observado.
- **Mutuamente excluyente:** cuando la inclusión de un grupo en una categoría lo excluye de la categorización en cualquier otra.
- **Exhaustivos:** cuando los grupos incluyen todas las posibilidades.
- **Aceptable:** cuando la característica es lógica e intuitiva, de forma que la comunidad acepte las categorías.
- **Útil:** cuando la característica puede utilizarse para obtener conocimiento.

El cumplimiento de los atributos en una taxonomía aplicada a un sitio Web, favorece la navegabilidad y facilita además acceder al capital intelectual de las organizaciones.

En el ambiente tecnológico, para la gestión eficaz de información y contenidos, la taxonomía es elemento

esencial, y dentro de ella, el XML se considera como tecnología fundamental. Hoy por hoy es clave en cualquier proyecto con grandes volúmenes de información establecer una política adecuada de taxonomías.

¿Quién las utiliza? Hoy en día los principales usuarios de las taxonomías son las grandes agencias internacionales de inteligencia (CIA, FBI, M16, etcétera), que utilizan desde hace muchos años tecnologías de rastreo de información, clasificación y recuperación muy sofisticadas.

Entre las ventajas que producen, se encuentra el mejoramiento de la organización de los contenidos mediante la representación de la información. Permiten definir la estructura de navegación y, por tanto, la organización de los contenidos de un sitio Web, garantizando el éxito en el uso del sitio.

Dentro del entorno digital las taxonomías mejoran la navegación y el desarrollo de sistemas de búsqueda basados en la exploración (browsing) y en la recuperación (searching). Un sitio que contiene información bien estructurada y clasificada puede ser fácilmente entendido por otros sitios Web y por los buscadores, así como cumplir mejores objetivos de diseminación de la información lo más ampliamente posible, además de permitir al usuario encontrar lo que busca en forma ágil y oportuna.

Si nos centramos en el ámbito de la arquitectura de la información, una misma taxonomía puede constituir una herramienta básica o auxiliar para los diferentes sistemas de navegación, de organización, búsqueda de contenidos, de etiquetado y de personalización, la cual ofrece diferentes beneficios:

- Permite la rentabilización del esfuerzo inicial de creación de la taxonomía y de los esfuerzos subsiguientes de mantenimiento.
- Facilita la gestión de las funcionalidades que aplica la taxonomía, es decir, una modificación en las categorías o en las relaciones entre categorías de la taxonomía puede trasladarse uniforme y consistentemente a todas las funcionalidades.
- Mejora el uso del sistema Web en su conjunto, ya que reduce considerablemente las exigencias de carga cognitiva, de memoria y de aprendizaje.
- Por último, facilita la interacción con el sitio Web y la creación de una imagen consistente de la organización que crea y aplica la taxonomía.

Ontología

Las ontologías incluyen las definiciones de conceptos básicos en un campo determinado y las relaciones entre ellos, es decir, las ontologías se encargan de definir los términos utilizados para describir y representar un área de conocimiento. Sirven para definir vocabularios que facilitan la búsqueda mediante una herramienta (buscador) que pueda entender, ya que son especificados con la suficiente precisión como para diferenciar términos y referenciarlos de manera precisa.

Una ontología es una descripción formal de los conceptos y las relaciones entre ellos. Las ontologías se caracterizan por tener componentes que sirven para representar el conocimiento de algún dominio. Estos son:

- Conceptos. Son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etcétera.
- Relaciones. Representan la interacción y el enlace entre los conceptos del dominio.
- Funciones. Son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología.
- Instancias. Se utilizan para representar objetos determinados de un concepto.
- Axiomas. Son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología.

Los axiomas permiten, junto con la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos. Las ontologías toman un papel clave en la resolución de interoperabilidad semántica entre sistemas de información y su uso dentro del contexto, ya que son una herramienta para el intercambio y el uso del conocimiento, porque proveen una comprensión compartida y consensuada de un dominio del conocimiento, que puede ser comunicada entre personas y sistemas heterogéneos. [4]

¿Quién las utiliza? Son utilizadas por los usuarios finales, las bases de datos y las aplicaciones que necesitan compartir información específica, por ejemplo, el sector bancario, el de salud, gestión de proyectos, etcétera.

De acuerdo con el grupo de trabajo W3C, las ontologías se utilizan generalmente en las siguientes áreas:

- Portales Web
- Colecciones multimedia
- Administración de sitios Web corporativos.
- Documentación de diseño.
- Agentes inteligentes
- Servicios Web y computación ubicua.

La principal función que tiene el lenguaje de ontologías es adicionar lógica descriptiva al lenguaje de esquemas de descripción RDFS. Además, por medio de estas tecnologías (lenguajes y herramientas) generan marcado y procesamiento semántico, considerando que sirven de referencia para la adquisición de información, lo cual permite compartir el conocimiento y posibilitan el trabajo colaborativo al funcionar como soporte común de conocimiento entre organizaciones y comunidades científicas, etcétera.

Folksonomía

El término folksonomía se emplea para designar a un sistema de etiquetado o clasificación de objetos Web no jerárquico, que nace de forma natural y democrática de los propios internautas, que son quienes asignan las etiquetas espontáneamente y de cuya gestión se encarga un sistema automático.

Los mejores exponentes son los sitios compartidos, como: Del.icio.us (enlaces favoritos), Flickr (fotos), Tagzania (lugares), o 43 Things (deseos).

Las folksonomías se apartan de las estructuras jerarquizadas para aproximarse a una organización basada en la colaboración. Son un método de clasificación explotado por varios servicios Web, donde los usuarios añaden información o fotografías o clasifican páginas Web.

Las decisiones de etiquetado las toman los propios usuarios, permitiéndose el uso de más de una etiqueta para clasificar cada elemento relacionado con un mismo tema. Las etiquetas aportan metadatos sobre lo que el usuario piensa que es el tema solicitado.

La folksonomía permite generar datos producidos por la participación de miles de usuarios. Este sistema colaborativo usado por muchas aplicaciones en red, materializa la arquitectura de la participación, así como las ideas de la inteligencia colectiva y la intercreatividad.

La forma de descripción de las folksonomías es característica de lo que se conoce como la Web 2.0, en la que la participación y la colaboración de los internautas es crucial.

Este uso colectivo de etiquetas genera un sistema de categorización no jerárquico. Construir una jerarquía no es algo esencial, mientras que sí lo es un instrumento que ofrezca la posibilidad de emplear etiquetas semánticas para efectuar la indización.

De acuerdo con lo anterior, no sólo son importantes los términos más populares, sino también aquellas etiquetas no dominantes (meta-noise) pero que añaden comprensión semántica. Por lo tanto, deben cumplir las siguientes condiciones para que se cumpla lo anterior:

- Están hechos por cualquier persona
- Para crearlos se requiere aprender muy pocas reglas.
- Se producen sin interés propio.
- Las ventajas de los usuarios crecen con la agregación.
- No se rompen cuando hay datos incompletos.
- Se presentan en forma de conjuntos, y no de jerarquías.
- No están diseñados a priori, lo que les hace más flexibles.
- No pertenecen a nadie, pues nadie la centraliza o controla.
- Son relevantes para los propósitos e intereses de un sitio Web.

¿Quién las utiliza? Este sistema de tags o etiquetas es muy común entre los weblogs. Permite incluso la búsqueda por tags o etiquetas en las entradas de los blogs y sitios sociales.

¿Qué ventajas producen? Entre las principales ventajas con las que contribuyen las folksonomías, se encuentran: la formación de los usuarios en temas específicos, el desarrollo de aprendizaje colaborativo y la habilidad de construir nuevo conocimiento que redundará en un mayor y mejor aprovechamiento para el crecimiento del capital intelectual, “recordando que el aprendizaje auténtico se sitúa en la experiencia real y está basado en la actividad colaborativa”.

Las folksonomías rompen con teorías clásicas sobre la construcción de conocimiento, las cuales afirmaban que éste se veía enfrentando una serie de dificultades que provenían del entorno, especialmente relacionadas con factores culturales, los individualismos, la falta de una cultura basada en el conocimiento, como con la capacidad en la utilización de herramientas informáticas como apoyo.

Son todo lo contrario: propician una cultura de colaboración y los individualismos desaparecen, ya que los usuarios comparten y se apropian de la información. Muchas son las comunidades que han salido favorecidas, ya que han logrado clasificar y organizar temas específicos que contribuyen a la generación de nuevo conocimiento.

Las folksonomías no se rompen si hay datos incorrectos o incompletos. Propician economías; reducen los costos mediante la optimización e intercambio de recursos y por la retroalimentación que existe entre ellas; contribuyen a la cooperación entre redes de información; animan la construcción de nuevas bases, por la sencillez con que son utilizadas por el usuario y por lo atractivo de su representación; son flexibles porque no están diseñadas con antelación, y crecen con la adición de nueva información.

La popularidad de las folksonomías crece día a día, y son cada vez más empleadas por todo tipo de usuarios (no es necesario poseer conocimientos especializados sobre sistemas de clasificación), puesto que responden al propósito por el que surgieron: la simple organización de los recursos Web.

De lo anterior se desprende que una buena organización de contenidos expuestos en la Web facilitará el uso y la apropiación de información al usuario final o colaborador de la organización, mejorará la eficacia en la recuperación de la información y, por ende, logrará una eficiente gestión para la construcción del

conocimiento. Mientras mayor sea la estructuración, clasificación y organización de la información, será más fácil para el usuario o colaborador de la organización hallar lo que realmente busca.

Recuperación inteligente de información

La recuperación digital de información es un campo muy amplio que incluye todas las formas de almacenamiento y envío digital de datos de cualquier índole. En el caso de la Lingüística Computacional, se trata principalmente de técnicas para la extracción de contenidos en textos y su transmisión a los usuarios. Para ello se usan actualmente métodos de procesamiento estadísticos y simbólicos diversos. Los buscadores de Internet se basan en uno o más de estos métodos de recuperación de información.

Todo aspecto del lenguaje humano es de interés para el desarrollo de buscadores semánticos para sitios Web, los cuales deben abordar áreas como fonética, morfología, sintaxis, semántica y la pragmática.

Para las búsquedas más sofisticadas se trata de desambiguar el sentido de las palabras mediante el uso combinado de estadísticas, redes semánticas, sistemas ontológicos, etcétera. [5]

Propuesta de una arquitectura de adquisición

La contribución a la automatización de la tarea de adquisición del conocimiento y de su relleno en modelos definidos, es un problema ya identificado en la construcción de sistemas basados en el conocimiento. [6] Dentro del contexto de la Web Semántica, esta tarea se redefine y adapta para abordar los nuevos problemas que surgen en procesos de adquisición para un repositorio de gran volumen de información no estructurada, como puede ser la World Wide Web actual. En este caso son las ontologías el formalismo elegido para el modelado del conocimiento de las aplicaciones (o agentes software). La tarea de adquisición de conocimiento, se puede dividir en dos partes:

- Adquisición del esquema de la ontología (ontology learning): en un primer momento es necesario diseñar un modelo semántico del dominio que represente los hechos de manera satisfactoria para los propósitos de las aplicaciones que lo exploten.
- Relleno del esquema definido (ontology population): la segunda parte consiste en instanciar el esquema de la ontología de dominio definido.

Para ambas tareas se necesitan ver sistemas automáticos y semi-automáticos, cuyo objetivo es localizar información en fuentes online, normalmente textuales. El fin de este módulo es ejercer un control sobre las distintas posibilidades de cada aproximación mediante estrategias de extracción. [7]

La llamada información de adquisición hace explícita la información sobre las distintas restricciones y relaciones que rigen sobre los datos, y le permite automatizar su identificación, extracción y posterior inserción en una ontología. [8]

Pre-proceso: Abastecimiento de Interpretaciones de Documentos

La arquitectura permite incluir interpretaciones de documentos como resultado de la fase pre-proceso, [9]

siendo identificadas seis interpretaciones de alto nivel:

- Interpretaciones de texto plano: la fuente es procesada como una cadena de caracteres sin tener en cuenta su significado, ubicación o relación con otras cadenas dentro del documento.
- Interpretaciones de HTML: permite modelar las relaciones estructurales entre los distintos

datos dentro de un documento hipertexto o entre varios documentos considerados para su análisis.

- Interpretaciones de Aspecto: modela aspectos visuales de los documentos.
- Interpretación de aspecto en coordenadas lógicas. La ventaja que ofrece ésta es una mayor eficiencia y menor costo computacional, que podría ser determinante en algunos dominios con fuertes requisitos sobre el tiempo de proceso.
- Interpretación de aspecto en coordenadas físicas. En este proceso se tienen en cuenta atributos de las etiquetas, especialmente aquellos referentes a las dimensiones de los objetos HTML (width, height, size, etcétera). El objetivo de esta interpretación es que cada pieza de información disponga de unas coordenadas asignadas con precisión de un pixel. Esto permite invocar operadores que comprueben posicionamientos mucho más precisos: IN ROW e IN COLUMN.
- Interpretaciones de Lenguaje: permiten identificar y recuperar datos de acuerdo a criterios sobre sus estructuras lingüísticas.

La potencia de la presente arquitectura estriba en la combinación de las interpretaciones en las tareas de identificación de la información. Existen varias maneras de enlazar las interpretaciones existentes entre sí, para poder referenciar una misma parte de la fuente original en más de una interpretación. La inclusión de interpretaciones nuevas en el sistema debe tener en cuenta qué granularidad tiene el sistema de enlace entre interpretaciones, como se muestra en la figura 1.

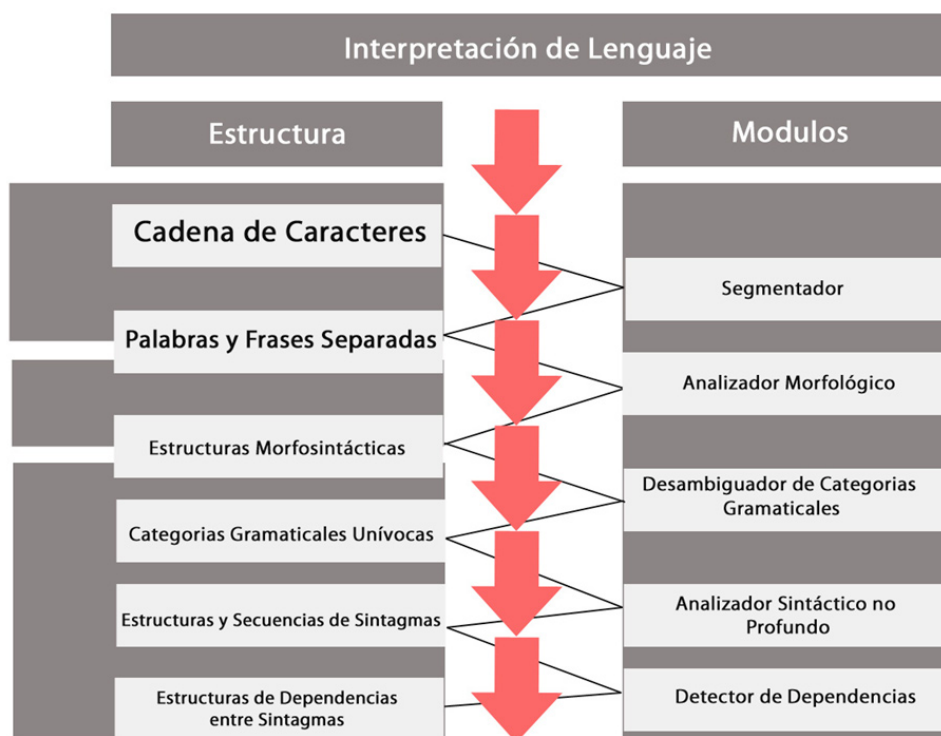


Figura 1. Estructura interna del proceso de interpretación de lenguaje

Hay dos posibles soluciones para enlazar las distintas interpretaciones entre sí:

- Enlace de Interpretaciones por posición Absoluta. Éstas deben disponer de un sistema de referencias común.
- Enlace de Interpretaciones Usando un Modelo Común. En algunos casos no es posible garantizar la estabilidad de las fuentes durante el proceso, por lo que se propone la utilización de un modelo central que sirva de referencia común para las interpretaciones en cuanto a las piezas de información, y es necesario prestar mucha atención a la granularidad mínima exigida entre todas las interpretaciones.

Módulo de identificación de información

Es el módulo de control del sistema propuesto que dirige todo el proceso de extracción y relleno de datos. Este proceso se ejecuta siguiendo estrategias de adquisición, que consisten en llamadas a diferentes operadores capaces de trabajar sobre las interpretaciones de documentos proporcionados en el pre-proceso de la fuente, y cuyo objetivo es construir un conjunto de hipótesis sobre la asignación de distintas partes de las fuentes a las piezas de información buscadas.

Ontología de Adquisición

La información contenida en la ontología de adquisición viene principalmente de dos fuentes:

- Esquema del modelo de dominio.
- Descripción del dominio.

La ontología de adquisición es una extensión de la ontología de dominio donde se añade información necesaria para el proceso de extracción y relleno. De esta manera no sufre modificaciones y es reutilizable para otros propósitos. Asimismo se puede aplicar sobre otras ontologías del mismo dominio con modificaciones mínimas.

El sistema presentado como proposición permite aprovechar tanto las instancias existentes previas al comienzo de la tarea de extracción, como las instancias adquiridas durante la misma, para reforzar o rechazar las hipótesis que se generan en el módulo central. Esto permite que a medida que el sistema vaya extrayendo información, vaya mejorando la eficiencia del mismo proceso gracias a la información adquirida. La arquitectura de la ontología se muestra en la figura 2.

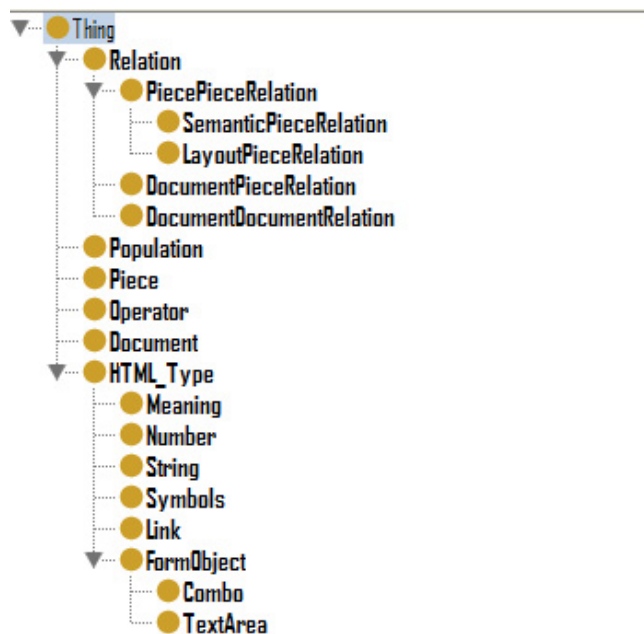


Figura 2. Ontología de Información

Relaciones

En la ontología de adquisición existen conceptos específicos para modelar relaciones entre las piezas y los documentos existentes.

Relaciones entre piezas:

- Relaciones de aspecto: una relación de aspecto entre dos piezas define su posición relativa en la visualización del documento. Especialmente en documentos de la WWW, el archivo fuente no determina la aparición final del documento directamente y es necesario procesar las marcas HTML para ubicar las distintas partes de la página.
- Relaciones semánticas: las relaciones semánticas permiten relacionar piezas de información a través del significado del contenido. Las relaciones semánticas se verifican usando técnicas de procesamiento de lenguaje natural para determinar qué relación existe entre dos piezas. Todos los operadores de comprobación tienen como pre-requisito la disponibilidad de interpretación de documento lenguaje, que contiene el resultado de un análisis sintáctico no profundo de la fuente (chunk parsing), identificando y relacionando sintagmas presentes.

Operadores

Las distintas estrategias propuestas para la implementación en el módulo de identificación de información, modificarán el orden de ejecución de los operadores sobre los modelos de documentos. Los operadores son piezas de software que realizan las tareas de recuperación, navegación o comprobación sobre las interpretaciones de documentos disponibles y son llamadas por el módulo de identificación de información de acuerdo a la estrategia de información.

La ejecución de los operadores de recuperación tiene por objeto localizar una cierta información esperada en el documento fuente. Es por ello que el presente sistema que se pretende proponer construya una hipótesis que forme las distintas posibilidades de asignación, de partes de la fuente a piezas de información esperadas.

Estrategias

Son las estrategias que construyen dinámicamente las secuencias de los operadores, con el objetivo de elaborar las hipótesis para el relleno. El algoritmo principal de ejecución de estrategias se realiza en un bucle de tres pasos:

Algoritmo principal del módulo de identificación

MIENTRAS queden decisiones por tomar REPETIR:

1. Tomar una decisión estratégica.
2. Ejecutar la decisión tomada.
3. Aplicar los resultados de la ejecución

FIN de Bucle

El objetivo de este algoritmo es construir un conjunto de hipótesis que relacionen piezas de información con partes del texto fuente, para que éstas puedan pasar a la fase final de relleno.

Dependiendo de la estrategia ejecutada, el sistema generará una sola o varias hipótesis. Las estrategias actúan sobre el primer paso de toma de decisiones. Las decisiones que se pueden tomar, son:

- Recuperar una pieza.
- Recuperar un documento.
- Recuperar una relación.
- Insertar hipótesis.

La manera de integrar una estrategia en módulo de identificación de información, consiste en generar partes disjuntas de la ontología de adquisición y usar la estrategia para generar un conjunto de hipótesis para cada una de estas partes. El conjunto de hipótesis, correspondiente a cada parte, se evalúa y ordena de acuerdo a su plausibilidad y aportación de información y se entrega al siguiente módulo que rellena la ontología de dominio.

Estrategia de Fuerza Bruta

La estrategia de fuerza bruta (Greedy) tiene por objetivo tomar decisiones encaminadas hacia una instanciación de datos lo más rápida posible. La restricción impuesta en esta estrategia es el mantenimiento de una única hipótesis para cada parte del espacio de búsqueda. La toma de decisiones sigue la siguiente secuencia:

- Recuperación de todos los documentos con URL conocida.
- Para cada documento se recuperan todas las piezas que le pertenecen y para cada pieza se recuperan todos los posibles candidatos.
- Para el documento se recuperan todas las posibles relaciones entre las piezas.
- Se comprueban las relaciones recuperadas y se eliminan los candidatos que lo incumplen.
- Si la hipótesis es válida (las cardinalidades se cumplen) se pasa a la inserción. En caso contrario, se procede con el siguiente documento recuperado.

Existen tres posibilidades en la toma de decisiones para la recuperación de una pieza P con una cardinalidad (N, M) para sus posibles valores:

- $(K < N)$ El número de candidatos recuperados K es menor que la cardinalidad mínima $(0)^1$ de la pieza: se descarta la hipótesis y el sistema termina anunciando la imposibilidad de cumplir con los requisitos de la ontología de adquisición.
- $(N \leq K \leq M)$ El número de candidatos recuperados K encaja en la cardinalidad de la pieza permitida: el sistema asigna los candidatos a la pieza.
- $(M < K)$ El número de candidatos recuperados K es mayor que la cardinalidad máxima de la pieza: se toman M piezas según una heurística que basa la selección únicamente en la lista de candidatos.

La estrategia de fuerza bruta es adecuada para sistemas con altos requisitos en tiempo de respuesta. Su complejidad es lineal dependiendo del número de documentos y piezas descritas y halladas en las fuentes.

Estrategias de búsqueda con retroceso (backtraking)

Esta estrategia despliega todo el posible árbol de búsqueda que consiste en la toma de decisión de las hipótesis, explotando todas las combinaciones. Cuando se encuentran candidatos para una pieza, más de lo que su cardinalidad permite o si la cardinalidad es mayor de uno, es preciso bifurcar la construcción de las hipótesis, ofreciendo todas las posibles combinaciones.

El proceso genérico de un algoritmo de búsqueda con retroceso sigue el siguiente pseudo código, que se muestra en la figura 3.

¹ Si cada instancia de la entidad no está obligada a participar en la relación

```
Vector Backtraking (H)
{
  IF esFinal (Hi) THEN
  {
    RETURN new Vector (Hi);
  }
  ELSE
  {
    Vector nuevas_hipotesis = ampliar (Hi);
    Vector buenas_hipotesis = filtrar_imposibles (nuevas_hipotesis);
    Vector resultado = new Vector ( );
    FOR (int = 0; i<buenas_hipotesis; i++)
    {
      Hipotesis h = (Hipotesis) buenas_hipotesis.elementAt (i);
      Vector resultado_parcial = Backtraking (H);
      resultado = union (resultado, resultado_parcial);
    } // END FOR
  } // END IF
} // END BCKTR
```

Figura 3. Proceso genérico de un algoritmo de búsqueda de retroceso

Es responsabilidad del proceso principal del módulo de identificación generar las distintas partes del espacio total de búsqueda, delimitado por la ontología de adquisición, para luego servir los conjuntos de hipótesis de las distintas partes a módulo de relleno de la ontología de dominio.

Estrategias de búsqueda con retroceso optimizada

Existen muchas variedades del algoritmo original para paliar este consumo computacional. En el sistema propuesto se ha optado por combinar dos de ellas:

- Poda. Poda del árbol de búsqueda de acuerdo a una función de costo
- Heurística. Introducción de heurísticas que no reducen el número de ramas en el árbol de búsqueda, pero que persiguen retrasar las ampliaciones de la hipótesis que mayor ambigüedad generan, lo máximo posible.

El proceso genérico de un algoritmo de búsqueda, con retroceso optimizado, sigue el siguiente pseudo código, como se muestra en la figura 4.

```
Vector Backtraking (Hi, funcion_costo)
{
  IF esFinal (Hi) THEN
    {
      costo_minimo = MIN (costo_minimo, funcion_costo(Hi));
      RETURN new Vector(Hi);
    }
  ELSE
    {
      Vector nuevas_hipótesis = ampliar (H);
      Vector buenas_hipotesis = filtrar_imposibles (nuevas_hipótesis);
      Vector hipotesis_podadas = podar (buenas_hipótesis, funcion_costo);
      Vector resultado = new Vector ( );
      FOR (int = 0; i<hipótesis_ordenadas; i++)
        {
          Hipotesis h = (Hipotesis) hipotesis_ordenadas.elementAt(i);
          Vector resultado_parcial = Backtraking (h, funcion_costo);
          resultado = union(resultado, resultado_parcial);
        } // END FOR
    } // END IF
} // END BCKTR_Opt
```

Figura 4. Pseudocódigo de un algoritmo de búsqueda con retroceso optimizado

Donde las funciones que dirigen el proceso, son una variante de las funciones de la estrategia genérica de búsqueda:

- Ampliar
- Podar
- ordenar.

El orden computacional de este algoritmo no es distinto, que el orden del algoritmo genérico de búsqueda con retroceso, pero el uso de podas y heurísticas permite disminuir ambigüedades tratadas y de esta manera reducir el tiempo y el consumo de memoria del mismo.

Hipótesis

Las hipótesis son estructuras de datos resultantes de la fase de identificación que albergan la información sobre la asignación de distintas partes de los documentos fuentes, con piezas de información esperadas y descritas en la ontología de adquisición. Como efecto de la existencia de restricciones sobre las piezas de información esperadas (su tipo de datos, relaciones entre ellas, cardinalidad, etcétera), se puedan considerar varias configuraciones de asignación de partes de las fuentes a piezas de información. Estas configuraciones se reflejan en las hipótesis resultantes del módulo de identificación.

Evaluación de las hipótesis

Con el fin de obtener un relleno correcto de los datos en la ontología de dominio, es necesario seleccionar una única hipótesis. Esto es posible solamente si el mecanismo de evaluación incluye una función de estimación que permita calcular el posible futuro valor de la hipótesis, aunque ésta no esté completamente construida. La función de evaluación está concebida conceptualmente como una función de costo: cuanto más alto valor devuelve, menos plausible es la hipótesis que evalúa.

$$Evaluar(H) = Evaluar(H_1, \psi, \phi) + Estimar(H_2, \psi, \phi) \setminus * MERGEFORMAT (1)$$

donde:

- H_1 : Parte construida de la hipótesis:
- H_2 : Parte que queda por construir de la hipótesis
- ψ : Función que expresa en grado de incumplimiento de la hipótesis con las restricciones opcionales en la ontología de adquisición.
- ϕ : Función que expresa el beneficio de la hipótesis.

La función de evaluación tiene un doble objetivo: permitir a las estrategias implementar heurísticas basadas en la función de costo y estimación (podas, priorización, etcétera) y, por otra parte, permite ordenar las hipótesis para ser rellenas en la ontología de dominio. La mejor hipótesis será aquella que mejor evaluada salga del módulo de identificación y menos inconsistencias genere en el modelo.

Relleno de Ontologías

El relleno de ontologías es la última fase de ejecución del sistema propuesto. Comprende la inserción de valores recuperados en lugares de un modelo de dominio. La aproximación más frecuente es dotar a un sistema tradicional de extracción de información, con un módulo de inserción de los datos en lugares preestablecidos en la ontología de dominio. Cada pieza identificada en las fuentes tiene asociada una información sobre su lugar en el modelo de dominio: a qué concepto corresponde y qué atributo se ha de rellenar.

Operaciones en la instanciación de ontología de dominio

El relleno de una ontología con datos extraídos de fuentes no estructuradas, puede implicar alguna de las siguientes operaciones:

- Creación de nuevas instancias para el relleno de un atributo.
- Relleno de atributos de instancias existentes con valores extraídos del texto.

- Modificación de valores de instancias existentes.
- Relleno de atributos para relacionar instancias existentes o de nueva creación.
- Modificación de relaciones entre instancias existentes.

Simulación

El objetivo de las simulaciones es calcular lo costoso que resulta insertar instancias en la ontología de dominio, en términos de creación de nuevas instancias o modificación de las existentes. Partimos del supuesto que la información encontrada no tiende a contradecir la información ya almacenada, y se premia a aquellas hipótesis que no introducen inconsistencias. Esta medida permite enriquecer la información de evaluación proporcionada por el módulo anterior y reordenar las hipótesis.

Propuesta de implementación de arquitectura

Se presenta un posible uso del contenido semántico, que se divide en dos partes: la primera, titulada Gestión y Búsqueda de Información Semántica, que describirá un sistema genérico de publicación de datos semánticos a través de un portal Web, que permite visualizar los datos adquiridos por el sistema propuesto. En la segunda parte, titulada Portal Semántico, se aplica el sistema de publicación a cualquier dominio.

Arquitectura Lógica de un Portal Semántico

Un portal semántico, como se propone en este trabajo, permite usar ontologías para modelar la información, y permite navegar por las instancias de los conceptos. Estas instancias constituyen la pieza básica del portal, siendo la información publicada, ligada a los documentos relacionados, y siendo también la información encontrada en los buscadores. Los buscadores evolucionan de esta manera y en vez de devolver documentos relevantes, como se hace en la mayoría de los portales actualmente, los buscadores de los portales semánticos devuelven instancias de conceptos como respuestas a las preguntas.

El portal semántico consta de tres módulos diferenciados, como se muestra en la figura 5:

- Módulo de interpretación de búsquedas: encargado de interpretar las búsquedas hechas por el usuario dentro del modelo semántico definido. El interfaz sencillo, consiste en una serie de formularios que se corresponden con conceptos definidos en la ontología, donde el navegante puede definir criterios de búsqueda, rellenando parcialmente los valores de los atributos.
- Módulo de adquisición: recupera información de fuentes online y los inserta en la ontología de dominio para que forme parte del portal.
- Módulo de presentación: encargado de presentar las instancias de la ontología de dominio y permitir una navegación entre ellos.

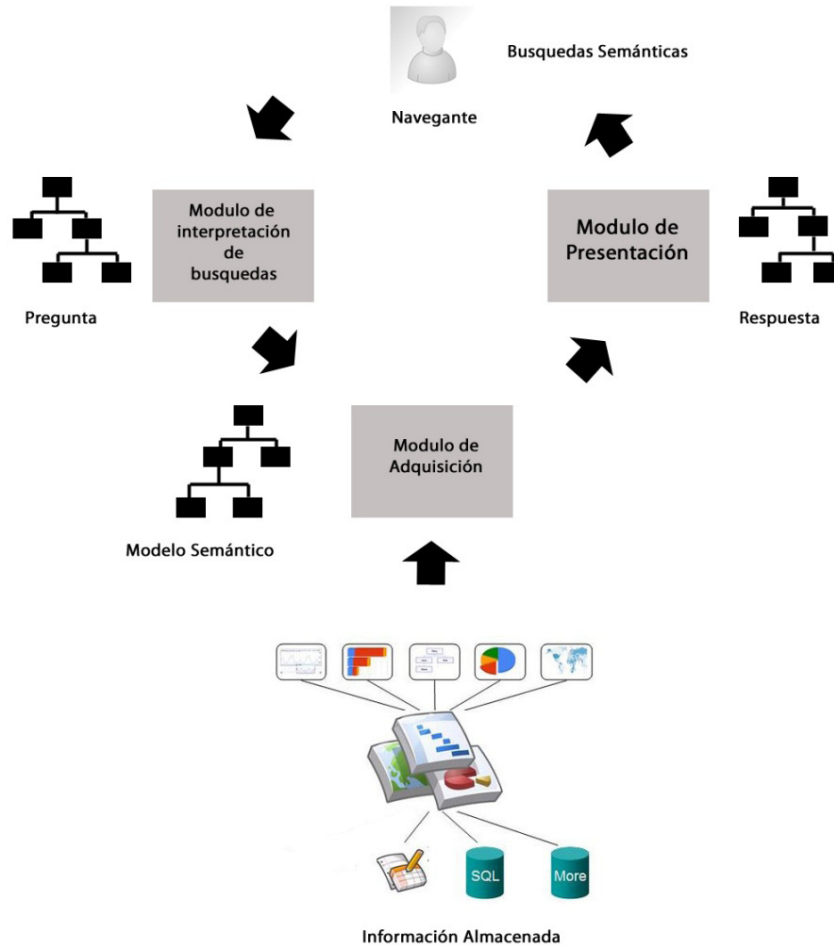


Figura 5. Esquema de un Portal Semántico

Modelo de Conocimiento Publicable

El proceso de publicación propuesto en este trabajo se basa en la existencia de una ontología auxiliar, llamada ontología de visualización o modelo de publicación, que permite definir vistas sobre la ontología de dominio. Estas vistas se definen de acuerdo con criterios de usabilidad y estéticos, con el fin de exponer una presentación legible del modelo semántico.

El modelo de publicación sigue el formalismo de ontología y actúa como un contenedor de entidades y atributos publicables, que extraen los valores mediante lenguajes de consultas sobre la ontología de dominio. Estas entidades publicables contienen aquellos atributos de la ontología de dominio que se van a presentar al usuario final. El contenido de una instancia publicable puede agrupar varios conceptos del modelo semántico original o, al contrario, pueden dividir un concepto de dominio complejo en varias instancias publicables.

Más concretamente, la ontología de visualización incluye dos conceptos predefinidos, que realizan la función de meta-modelo:

- Entidad de publicación: concepto que encapsula objetos, tal como se verán publicados. Todo concepto definido en la ontología de publicación heredará de él, y deberá definir los siguientes atributos:

- Hoja de estilo asociada al concepto que traduce sus instancias publicables.
 - Consulta que obtiene todos los valores de los atributos de la instancia correspondiente en la ontología de dominio.
- Atributo de publicación: todos los atributos que se muestren en la aplicación final deben heredar de este concepto. La forma que en el atributo se muestre en la página Web se define mediante las siguientes propiedades:
 - Etiqueta: la posible etiqueta que aparecerá con el valor del atributo.
 - Consulta: se ejecuta para obtener el valor del atributo.
 - Enlace: si el valor publicado debe realizar alguna acción al pulsar sobre él (enlace Web, mail, botón, etcétera).

Las componentes de una página del portal para visualizar instancias de la ontología, se describen como subclases de Entidad publicación, y sus instancias se definen de acuerdo con el canal de publicación (HTML, WAP, Voice-VML, XML) a generar, a través de transformaciones XSL.

La separación entre la ontología para la representación del conocimiento del dominio y la ontología de visualización, facilita la independencia de estas dos labores. Otra ventaja de la aproximación seguida, es que para ambas se pueden utilizar las mismas herramientas de manejo de lenguajes de ontologías, como Protégé.

Proceso de extracción

Las distintas estrategias tienen por objetivo secuenciar la ejecución de los operadores, con el objetivo de generar una hipótesis sobre posibles rellenos en la ontología de dominio. En la propuesta de implementación del portal semántico para cualquier dominio que se desee en particular, se ha optado por la utilización de una estrategia de búsqueda con retroceso aumentada con alguna heurística para la optimización del proceso de relleno, disminuyendo el número de hipótesis creadas, y con ello perdiendo algunas soluciones posibles, como se ilustra a continuación:

El proceso comienza con el procesamiento del contenido de la Web de Prueba, donde se localiza la página inicial (home page). En esta página se localiza el nombre del objeto buscado, descrito como una pieza contenida en una lista desplegable (combo box), que sirve de enlace, a través de una relación pieza-documento a las páginas descriptivas de la palabra computacional. La ontología de adquisición especifica que la pieza de nombre del objeto en la página principal, tiene una cardinalidad de (0: 300). El sistema encuentra 3 candidatos para la pieza de nombre computacional, todos ellos incluidos en la lista desplegable. Al ser el número de candidatos menos que la cardinalidad máxima permitida, el sistema tiene una única hipótesis sobre el documento, donde se incluyen todos los candidatos hallados como posible asignación a la pieza de nombre computacional, como se muestra en las figuras 6 y 7.

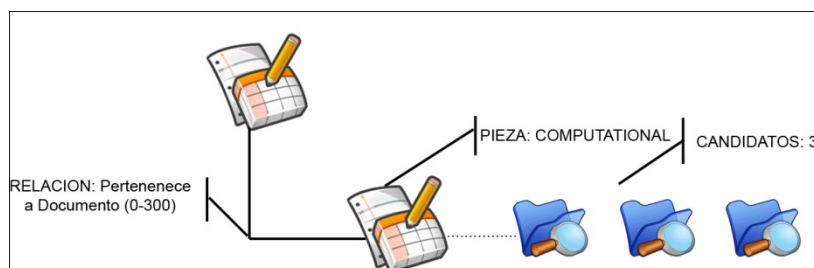


Figura 6. Búsqueda dentro del portal de prueba de la palabra computacional

Basic Configuration	
Search every tag name:	Yes ▾
Search custom taxonomies:	Yes ▾
Search every category name and description:	Yes ▾
Search every comment:	Yes ▾
Search comment authors:	Yes ▾
Search approved comments only:	Yes ▾
Search every excerpt:	Yes ▾
Search every draft:	Yes ▾
Search every attachment: (post type = attachment)	Yes ▾
Search every custom field: (metadata)	Yes ▾

Figura 7. Especificación de la pieza con nombre computacional

En este paso existe una posible pérdida de información en el proceso de construcción de la hipótesis. En el caso de encontrar un número de candidatos menor que la cardinalidad máxima permitida, de acuerdo a la estrategia de búsqueda con retroceso pura, deberían generarse tantas hipótesis como posibles combinaciones de asignación de los candidatos a la pieza, es decir: N factorial, siendo N el número de candidatos encontrados.

La estrategia propuesta para este dominio asume la posible pérdida de información, mejorando así la eficiencia de la recuperación tanto en tiempo como en recursos consumidos. La estrategia pura de búsqueda en anchura sería inviable en este caso, con 2861 hipótesis creadas. Esta modificación en el algoritmo original es posible al alto grado de estructura presente en las fuentes.

Para la palabra computacional localizada en el portal, se procede a navegar hacia la página de descripción detallada. En esta página se localizan las piezas descritas en la ontología de adquisición. La descripción consta de un conjunto de piezas formadas por una etiqueta que está en línea visual con un dato buscado. Esta descripción permite una eficiente extracción a partir del documento, gracias a su estructura bien definida.

Finalmente, tras procesar la Web de prueba, el sistema obtiene 3 resultados posibles, de los cuales se elige el que se desea con todos sus atributos.

Esta pérdida de información se trata de compensar con la búsqueda por taxonomía e hyphenation, a través de una indexación y métodos de búsqueda.

Para tratar de optimizar tanto la búsqueda y extracción de información, se utiliza el algoritmo de Huffman, para optimizar el funcionamiento del proceso mostrado anteriormente.

Algoritmo de Huffman

El Algoritmo de Huffman es una técnica ampliamente usada y muy efectiva para la comprensión de datos. Esta técnica reduce en gran porcentaje el espacio en memoria de un archivo. Tal reducción depende de las características del mismo, esto es con el fin de optimizar el motor de búsqueda y no se sature el cache.

El problema de los códigos de Huffman consiste en determinar el código binario para representar cada uno de los caracteres, de tal manera que el número de bits requerido para representar el texto sea mínimo. Este

20 -xx

problema se define formalmente de la siguiente manera:

Instancia. Los caracteres $c_1, c_2, c_3, \dots, c_n$ y sus frecuencias f_1, f_2, \dots, f_n

Un conjunto S de códigos binarios (para nuestros datos) $cod(c_1), cod(c_2), \dots, cod(c_n)$

Medidas. $m(S) = \sum_{i=1}^n f_i * cod(c_i)$, donde $cod(c_i)$ es la longitud del código binario $cod(c_i)$.

Objetivo. Minimizar $m(S)$

La solución que se encuentra representada por un árbol binario, de la siguiente manera:

- Las hojas del árbol son los caracteres
- Al recorrer el camino de la raíz a una hoja determinada se obtiene el código de dicha hoja, con la interpretación cero si el siguiente nodo del camino es hijo izquierdo y 1 si es hijo derecho.
- $cod(c_i)$ corresponde a la profundidad del carácter c

La siguiente función extrae el menor elemento del montículo C , conservando esta estructura en C

def extraermin (C):

i, menor: cardinal

1 extraermin = C. vec [1]

2 intercambiar (C.vec1, C, n)

3 dec (C. n)

4 hundir (C.vec, C, n, 1)

La función extermin tiene una complejidad $O(\log(n))$.

El procedimiento insertar introduce el elemento x al final de C , y posteriormente lo hace flotar para mantener las propiedades del montículo.

def insertar (C,x):

1 inic (C.n)

2 C.vec[C.n] = x

3 Flotar (C.vec, c, n)

El procedimiento tiene una complejidad de $\log(n) + 2$

La función Huffman inicialmente crea un montículo con el vector de caracteres C; paso seguido extrae los dos elementos de menor peso del montículo; luego crea un nuevo caracter que tenga éstos como hijos y cuyo peso sea la suma de los pesos de sus hijos.

Por el ultimo, el nuevo carácter es insertado en el montículo. Este proceso es realizado $n - 1$ veces, hasta haber conformado el primer caracter de C un árbol binario que representa la codificación de los caracteres iniciales.

Por lo tanto $f(n) = O(n \log(n))$ es la función de complejidad de este algoritmo.

Resultados

Se ha logrado la construcción de una plantilla semántica gratuita (en Wordpress y Joomla). La puede usar cualquier persona no experta en el campo, para poder optimizar la búsqueda y estructuración de cualquier sitio Web. La herramienta es escalable en cualquier lenguaje de programación. Se pueden indexar en los motores de búsqueda principales, como Yahoo!, Bing, Ask y Google. Además se destacan los siguientes elementos que contribuyen a la generación de conocimiento, los cuales se presentan a continuación:

The screenshot displays the WordPress Taxonomy management interface, organized into three main sections:

- Taxonomies create with More Taxonomies:** This section shows a table with columns for 'Taxonomy' and 'Actions'. Below the table, it indicates 'No Taxonomies defined' and provides a blue button labeled 'Add new Taxonomy'.
- Taxonomies elsewhere:** This section shows taxonomies created in function.php or elsewhere. It includes a table with 'Captions' listed under 'Taxonomy' and 'Override | Export' under 'Actions'.
- Default WordPress taxonomies:** This section lists default taxonomies built into WordPress. The table includes entries for 'Category', 'Post Tag', 'Navigation Menu', and 'Link Category', each with 'Override | Export' actions. A 'Format' entry is also present with 'Override | Export' actions.

Figura 8. Creación de Taxonomías dentro de la Plantilla Semántica

Hide verbose explanations of settings

OpenGraph (Facebook)

Add OpenGraph meta data

Add OpenGraph meta data to your site's <head> section. (very experimental)

Indexation Rules

Below you'll find checkboxes for each of the sections of your site that you might want to disallow the search engines from indexing. Be aware that this is a powerful tool, blocking category archives, for instance, really blocks all category archives from showing up in the index.

This site's search result pages

Prevents the search engines from indexing your search result pages, by a `noindex, follow` robots tag to them. The `follow` part means that search engine crawlers *will* spider the pages listed in the search results.

The login and register pages

(warning: don't enable this if you have the [minimeta widget](#) installed!)

All admin pages

The above two options prevent the search engines from indexing your login, register and admin pages.

Subpages of the homepage

Prevent the search engines from indexing your subpages, if you want them to only index your category and / or tag archives.

Author archives

Figura 9. Creación de un Grafo en Facebook (Red Social) y reglas de indexación

Creación de Sinónimos dentro del portal, considerando que una palabra puede ser polisémica.

Synonyms

Add synonyms here in 'key = value' format. When searching with the OR operator, any search of 'key' will be expanded to include 'value' as well. Using phrases is possible. The key-value pairs work in one direction only, but you can of course repeat the same pair reversed.

Figura 10. Creación de Sinónimos dentro del portal semántico.

Además, optimización de palabras dentro de la plantilla semántica.

Keyword Optimizer Options

The following fields except comma separated values example: seo, blogging

 Keywords	<input type="text"/>	Limit: <input type="text" value="1"/>	<input type="checkbox"/> Enable?
 Keywords	<input type="text"/>	Limit: <input type="text" value="1"/>	<input type="checkbox"/> Enable?
<u> Keywords	<input type="text"/>	Limit: <input type="text" value="1"/>	<input type="checkbox"/> Enable?

Figura 11. Optimización de palabras

Por último, se presenta una estadística de las personas que han visitado el portal semántico.



Figura 12. Gráficos de visitantes del portal semántico

Conclusiones y propuestas futuras

La capacidad para almacenar datos ha crecido en los últimos años a velocidades exponenciales. En contrapartida, la capacidad para procesarlos y utilizarlos no ha ido a la par.

Por este motivo un buscador semántico se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento obtenido, usando grandes volúmenes de datos. Sin embargo, en su aplicación no sólo se obtienen patrones que no sirven, mientras no se les encuentre significado, y su valor real reside en la información que ayude a tomar decisiones o mejorar la comprensión de los fenómenos que nos rodean.

Las técnicas estadísticas son fundamentales a la hora de validar hipótesis y analizar datos, por lo cual la estadística desempeña un papel muy importante, pero debemos considerar que esta no es la única herramienta para analizar datos, ya que los resultados a veces carecen de significado al utilizar este tipo de técnicas. Sólo se basan en el patrón de búsqueda de un usuario o conjunto de usuarios.

La intención de este trabajo es contribuir a la superación de uno de los retos más importantes, identificados en la consecución del éxito de la Web Semántica. Cabe destacar que es una iniciativa sin precedentes

sobre el propósito de la creación de una base de conocimientos global, formal y distribuida, equivalente a la WWW, pero a diferencia de ésta, la Web Semántica está definida para el proceso y consumo por parte de aplicaciones de software.

Se propone una arquitectura que permite implementar sistemas en los cuales las distintas tecnologías se unen en la consecución de la tarea de obtención de contenido semántico. El proceso de cooperación entre ellas va dirigido por una estrategia, adecuando el proceso al tipo de fuente y su dominio. Es una arquitectura extensible que modela el proceso de extracción en tres fases: pre-proceso de las fuentes según su interpretación, extracción de la información y formación de hipótesis sobre su semántica y finalmente la fase de inserción. El diseño de la arquitectura, de manera abierta, permite incorporar nuevas interpretaciones de las fuentes, así como nuevas estrategias de control en el módulo de extracción, adecuando el sistema completo a las necesidades de cada dominio o cada aplicación. La flexibilidad y apertura de la arquitectura ha sido un requisito esencial en su concepción, para servir de plataforma de desarrollo y extensión del alcance de los sistemas finales en nuevas fuentes y nuevas estrategias de procesamiento y extracción de información.

La propuesta de arquitectura hecha, ha tenido, desde sus orígenes, en cuenta el requisito de apertura hacia su expansión, inclusión de nuevas aproximaciones tecnológicas y aplicación en distintos dominios. Es allí donde se centran las posibles futuras condiciones y ampliaciones que pueden englobar interpretaciones digitales: de fuentes PDF, de Bases de Datos, de Sistemas de gestión documental y Multimedia.

Asimismo, la arquitectura soporta la inclusión de nuevas estrategias de extracción: Heurísticas empíricas o estadísticas, Información adicional externa al sistema e Interacción con el usuario. También se ha estudiado la posibilidad de construcción de sistemas de extracción sobre nuevos dominios: Dominio financiero y Cultural.

Referencias

1. Artz, D.; Gil, Y. A survey of trust in computer science and the Semantic Web, en: *Web Semantics: Science, Services and Agents on the World Wide*. 2007; 2:58-71.
2. Christoph M. A survey and classification of semantic search approaches, en: *International Journal of Metadata, Semantics and Ontologies*. 2007; 1:23 – 34.
3. Corcho, O.; Alper, P.; Kotsiopoulou, I.; Missier, P.; Bechhofer, S.; Goble, C. An overview of S-OGSA: A Reference Semantic Grid Architecture, en *Web Semantics: Science, Services and Agents on the World Wide*. 2006; 2:102-115.
4. Fabian M. Suchanek; Gjergji Kasneci; Gerhard Weikum. YAGO: A Large Ontology from Wikipedia and WordNet, en: *Journal of Web Semantics*. 2008; 3:203-217.
5. Qi wang; Shu Wang; Haitao Gong. Synthesis of relational web services, en: *International Journal of Semantic Computing*. 2010; 3: 385-417.
6. Mika, P. Ontologies are us: A unified model of social networks and semantics, en: *Web Semantics: Science, Services and Agents on the World Wide Web*. 2007; 1:5-15.
7. Ronald D.; Catherine D.; Glen H.; Vania D.; Anthony G. Supporting Domain Experts to Construct Conceptual Ontologies: A Holistic Approach, en: *Web Semantics: Science, Services and Agents on the World Wide Web*. 2011; 1:54-78.
8. Schraefel; Lloyd R. User interaction in semantic web research, en: *Web Semantics: Science, Services and Agents on the World Wide Web*. 2010; 4: 375-376.
9. Yingxu W. On concept algebra for computing with words, en: *International Journal of Semantic Computing*. 2010; 3: 331-35.