
12

MODELO DE WEB SEMÁNTICA PARA UNIVERSIDADES.

*Karla Abad Sacoto, Walter Orozco Iguasnia,
Washington Torres, Alfredo González Tomalá*

*Fecha de recepción: 24 de septiembre del 2015
Fecha de aceptación: 15 de diciembre del 2015*

MODELO DE WEB SEMÁNTICA PARA UNIVERSIDADES

Karla Abad Sacoto¹, Walter Orozco Iguasnia¹, Washington Torres¹, Alfredo González Tomalá¹

¹ Escuela de Informática, Universidad Estatal Península de Santa Elena (UPSE).
Campus La Libertad, vía principal Santa Elena – La Libertad. La Libertad-Ecuador
E-mail: karla.abad.13@gmail.com

Resumen

A raíz del estudio de estado actual de micrositos y repositorios en la Universidad Estatal Península de Santa Elena se encontró que su información carecía de semántica óptima y adecuada. Bajo estas circunstancias, se plantea entonces la necesidad de crear un modelo de estructura de web semántica para Universidades, el cual posteriormente fue aplicado a micrositos y repositorio digital de la UPSE, como caso de prueba. Parte de este proyecto incluye la instalación de módulos de software con sus respectivas configuraciones y la utilización de estándares de metadatos como DUBLIN CORE, para la mejora del SEO (optimización en motores de búsqueda); con ello se ha logrado la generación de metadatos estandarizados y la creación de políticas para la subida de información. El uso de metadatos transforma datos simples en estructuras bien organizadas que aportan información y conocimiento para generar resultados en buscadores web. Al culminar la implementación del modelo de web semántica es posible decir que la universidad ha mejorado su presencia y visibilidad en la web a través del indexamiento de información en diferentes motores de búsqueda y posicionamiento en la categorización de universidades y de repositorios de Webometrics (ranking que proporciona clasificación de universidades de todo el mundo).

Palabras Claves: *Dublin Core, indexamiento, metadatos, motores de búsqueda, modelo, SEO, web semántica.*

Abstract

After examining the current microsites and repositories situation in University, Peninsula of Santa Elena's, it was found that information lacked optimal and appropriate semantic. Under these circumstances, there is a need to create a semantic web structure model for Universities, which was subsequently applied to UPSE's microsites and digital repositories, as a test study case. Part of this project includes the installation of software modules with their respective configurations and the use of metadata standards such as DUBLIN CORE, to improve the SEO (Search Engine Optimization); with these applications, it was achieved the creation of standardized metadata and the creation of uploading information policies. The use of metadata transforms raw data into well-organized structures that provide information and knowledge to generate web engine search results. Upon completion of the implementation of semantic web model, it is possible to say that the university had improved its presence and visibility on the web through the indexing of information in different search engines and the categorization positioning of universities and Repositories in the Webometrics ranking (ranking Web of Higher Education Institutions Worldwide).

Keywords: *Dublin Core, indexing, metadata, model, search engine, semantic web, SEO.*

1. Introducción

En el presente, el internet es un medio de uso diario para muchas instituciones académicas públicas o privadas, ya que por medio de sitios web se publica y difunde información ya sea de índole investigativo, científicos, culturales, etc. facilitando la comunicación entre instituciones, estudiantes y la comunidad en general.

Estos sitios Web contienen información representada mayoritariamente en forma de texto, ésta es fácilmente comprensible por personas que la leen, sin embargo no lo es para los motores de búsqueda que necesitan algún tipo pista o alguna palabra que indique que existe relación entre la búsqueda y la información que posee el sitio.

Al no existir estas palabras que guíen a encontrar la información deseada, los buscadores arrojarán resultados de otros sitios web que contienen información similar pero no del sitio de nuestro interés. Entonces es necesario tener en cuenta que ciertos motores no podrán encontrarlo así este tenga la información que se está solicitando, este hecho se da por la falta de semántica en los datos.

El diseñador y comunicólogo Shedroff N. [1] considera que el Diseño de Información trata con la organización y presentación de datos: su transformación a información valiosa y significativa. Así mismo Jiménez M. [2] indica que el Instituto Internacional de Diseño de Información, mejor conocido como IIID, considera al Diseño de Información como la definición de requerimientos tales como la selección, transformación y transmisión de la información con el propósito de transmitir conocimiento, así como de optimizar la información en relación a dichos requerimientos.

La semántica que se maneja a nivel de sitios web, dependerá mucho del administrador de contenido que se esté utilizando, ya que es necesaria la instalación de módulos y configuración de los mismos, así como de llevar un estándar de subida de información. Uno de los estándares más recomendados por medidores de rankings es el formato Dublin Core [3], el cual permite la generación de metadatos para poder construir una web semántica. Esto se complementa con la ayuda de taxonomías e indexación de sitios que además permitirán el monitoreo del tráfico que se genera en base a resultados producidos por los principales buscadores a nivel mundial como lo son: Google, Bing y Yahoo.

Este artículo pretende presentar y explicar el modelo de estructura de web semántica creado para la UPSE luego de la realización de un análisis de situación actual, de herramientas y estándares adecuados para ello. Además busca demostrar las

ventajas de la tecnología de la Web Semántica mediante la aplicación de dicho modelo.

2. Metodología de Investigación

Para el estudio realizado se utilizó la investigación documental, el cual es un procedimiento científico, un proceso sistemático de indagación, recolección, organización, análisis e interpretación de información o datos en torno a un determinado tema. Al igual que otros tipos de investigación, éste es conducente a la construcción de conocimientos. Según [4] la investigación documental tiene la particularidad de utilizar como una fuente primaria de insumos, más no la única y exclusiva, el documento escrito en sus diferentes formas: documentos impresos, electrónicos y audiovisuales.

Aplicando este tipo de investigación fue posible recopilar y analizar los documentos de facultades, centros y departamentos de la Universidad, los cuales serán subidos en el sitio web correspondiente y en cada comunidad del repositorio digital.

Se utilizó además la investigación exploratoria. Los estudios exploratorios sirven para familiarizarnos con fenómenos en estudio, obtener información sobre la posibilidad de llevar a cabo una investigación más completa respecto de un contexto particular, investigar nuevos problemas, identificar conceptos o variables promisorias, establecer prioridades para investigaciones futuras, o sugerir afirmaciones y postulados. Esta clase de estudios son comunes en la investigación, sobre todo en situaciones donde existe poca información [5].

En este caso se usará la investigación exploratoria para realizar el seguimiento de los movimientos del tráfico en los Micrositios y el Repositorio Digital y verificar el posicionamiento en los rankings universitarios de Webometrics [6].

3. Situación inicial

Al inicio del proyecto de implementación de web semántico aplicado a los Micrositios y repositorio digital de la UPSE, se encontraron falencias en la estructura semántica de los sitios web de las diversas áreas y en el repositorio no se estaba utilizando un estándar óptimo de generación de metadatos además de que algunas colecciones creadas estaban vacías. Poseer colecciones vacías provoca que los evaluadores de repositorios consideren que en esas áreas no se está trabajando y este hecho incide finalmente en las calificaciones.

Durante el proceso de investigación se aplicaron fichas de observación las cuales ayudaron a tener un

panorama de la situación en que se encuentra la web de la universidad en base a ciertos indicadores que se describen a continuación:

En la figura 1 se muestra el estado en el cual se encontraba el Repositorio Digital de la Universidad Estatal Península de Santa Elena. Se puede notar que se estaban generando metadatos pero estos no poseían un formato óptimo. También se encontró un problema que repercute en la presencia web, el cual es tener pocos archivos almacenados; además la información no es actualizada periódicamente y por ende la visibilidad del repositorio no llegaría a mejorar.

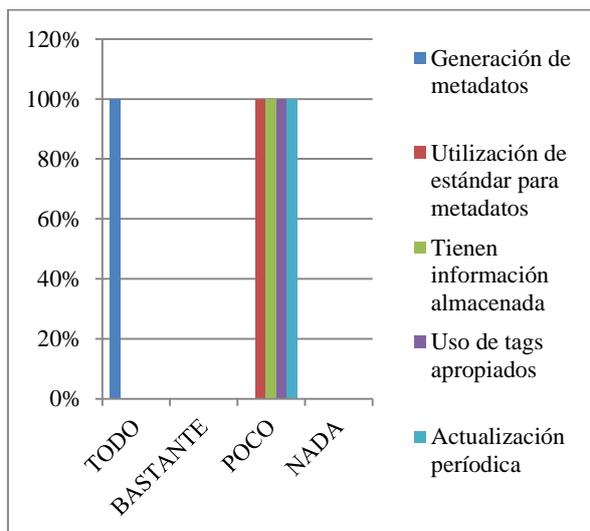


Figura 1. Situación encontrada Repositorio digital.

Se analizó la situación actual del repositorio de la Universidad, en el cual se pudo visualizar que no se estaba utilizando el estándar Dublin Core, establecido al inicio del proyecto para la generación de semántica de cada artículo que se subía al repositorio digital. También se halló que existían comunidades creadas en las cuales no se tiene información y la falta de palabras claves relevantes en algunos de los archivos del repositorio digital.

En la figura 2 se muestra el estado en el cual se encontraban los Micrositios de la Universidad Estatal Península de Santa Elena. Se puede notar varias falencias entre las cuales se listan: poca taxonomía, inexistencia de metadatos, falta de uso de un estándar, información no actualizada, existencia de enlaces perdidos y demora en carga en algunos de los sitios.

Durante la revisión de los micrositios de la Universidad, se observó que en su mayoría solo poseían módulos y configuraciones básicas lo cual conduce a la no generación de resultados que

conlleven al aumento del tráfico o resultados en buscadores que redireccionen a los micrositios que están bajo el dominio upse.edu.ec.

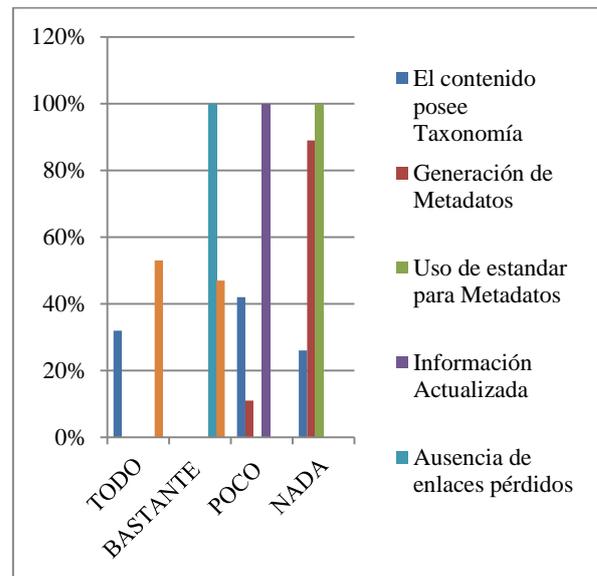


Figura 2. Situación encontrada Micrositios web.

Los contenidos no poseían el estándar de generación de metadatos adecuado y por consecuencia no se estaba construyendo la taxonomía correspondiente por cada archivo subido a los micrositios. Esto colabora en el aumento de resultados en los buscadores web a nivel mundial.

La información de todos los sitios no se encontraba actualizada y en cuanto al diseño existían sitios con plantillas que ocasionan demoras en carga de las páginas y en su mayoría no son responsivas (es decir que posean la capacidad de acoplarse a diversos dispositivos).

4. Modelo general para la construcción de Web Semántica de Universidades.

Los hallazgos del análisis inicial evidenciaron la necesidad de construir un modelo general de web semántica que permita aprovechar las ventajas de las tecnologías y herramientas que este proyecto contempla. En la figura 3 se visualiza el esquema que comprende el modelo de web semántica.

El modelo desarrollado consta de cuatro grandes grupos:

- Micrositios
- Repositorio Digital
- Herramientas de Indexación
- Herramientas de Monitoreo

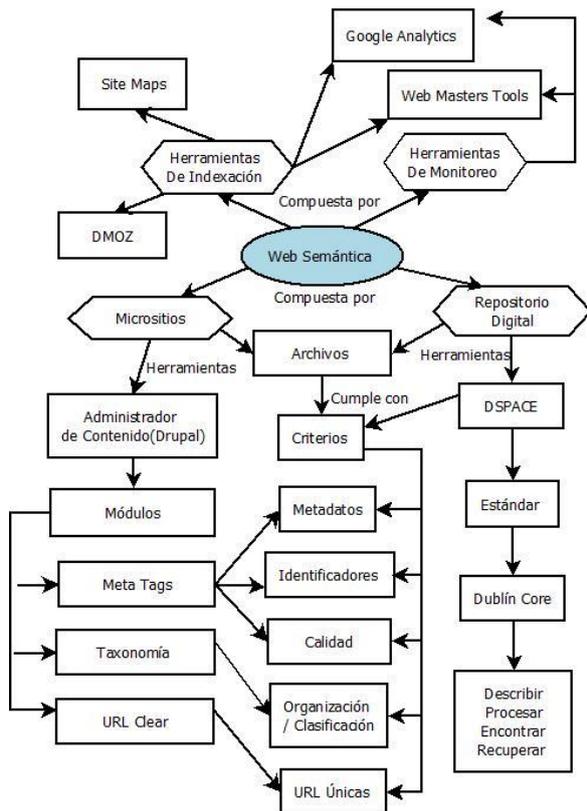


Figura 3. Modelo General – Web Semántica

4.1 Micrositios

Esta sección se refiere a la información que se subirá a los sitios web de centros, departamentos, facultades y áreas administrativas de la universidad, estos sitios están creados con el Administrador de Contenido Drupal, el cual nos ofrece módulos orientados definir la semántica en sitios web.

Entre los módulos que se utilizan para la generación de semántica están: Meta Tags, Taxonomía y URL Clear, los cuales colaboran con la generación de los siguientes criterios de semántica web:

- **Metadatos:** Son las palabras relevantes sobre el archivo a subir, para esto hacemos uso del módulo Meta-Tags.
- **Identificadores:** Estos ayudarán a redireccionar a los buscadores hacia la información que posee el sitio, para la creación de estos identificadores se usará el módulo Meta-Tags.
- **Calidad:** Para personificar los archivos en la web se necesita tener parámetros como el creador, tipo de archivo entre otros, para esto se utilizará el módulo Meta-Tags.
- **Organización/ Clasificación:** Permitirá que la información de cada sitio sea clasificada según

su procedencia, para esto se hace uso de la Taxonomía.

- **URL únicas:** Esto permitirá generar automáticamente URL para evitar conflictos. Se hará uso del módulo URL-Clear

Los criterios anteriormente mencionados se tomarán en cuenta en la subida de información en cada sitio web además de considerar el estándar Dublin Core.

4.2 Repositorio Digital

El repositorio digital fue creado con la herramienta Dspace la cual posee comunidades y sub comunidades que están creadas para el alojamiento de documentos generados por las diversas áreas de la universidad. Al momento de crear sub comunidades se debe tener en cuenta el estándar Dublin Core para la generación de metadatos. Este estándar posee los siguientes elementos: Describir, Procesar, Encontrar y Recuperar.

Los metadatos Dublin Core tratan de ubicar, en el entramado de Internet, los datos necesarios para describir, identificar, procesar, encontrar y recuperar un documento introducido en la Red. Si el conjunto de elementos Dublin Core se logrará aceptar internamente supondría que todos los motores de búsqueda que indexan documentos en Internet, encontrarían en la cabecera de los mismos, todos los datos necesarios para su indexación [6].

4.3 Herramientas de Indexación

Para que la semántica aplicada tanto a los micrositios y al repositorio tenga un mayor impacto en los principales buscadores web tenemos que indexarlos con ayuda de las siguientes herramientas:

DMOZ: Ayuda en la presentación formal de un sitio en el internet y además lo indexará a los principales buscadores web a nivel mundial.

XML Sitemap: Este módulo de Drupal ayuda a la indexación específicamente a indexar sitios a dos buscadores como Bing y Google.

Web Masters Tool: Esta herramienta permite optimizar la visibilidad de los sitios creados, además permite que el principal buscador a nivel mundial Google pueda ver el contenido de los sitios con facilidad.

Google Analytics: Esta herramienta permite la indexación de los sitios web a Google por medio de un módulo que se agrega al administrador de contenido. La indexación se dará a través de un código que genera esta herramienta, el cual se

insertará en la cabecera de cada sitio pues ahí se encuentran las palabras claves del mismo.

4.4 Herramientas de Monitoreo

Google ofrece dos grandes herramientas de monitoreo: Web Masters Tool y Google Analytics, las cuales nos proporcionan reportes del tráfico que se genera en cada uno de los sitios agregados a cada una de estas dos herramientas.

Web Masters Tool: entre una de sus características especiales está proporcionar un reporte de fallas que posee el sitio como enlaces dañados o amenazas y ataques.

Google Analytics: proporciona reportes gráficos de las visitas a nivel mundial, estadísticas de tráfico y estado del sitio.

Luego de haber diseñado el modelo anteriormente expuesto, éste fue aplicado para la construcción de la web semántica de la UPSE, con la finalidad de validar el modelo propuesto y aumentar la presencia de los microsítios y el repositorio digital en los buscadores web, mejorar la posición de la universidad en el ranking web de universidades de webometrics e incrementar el tráfico hacia los sitios que estén bajo el dominio upse.edu.ec.

5. Diseño de políticas de subida de Información.

El modelo de web semántica se complementa con el diseño de un conjunto de políticas que deberán ser aplicadas para subir información a la web de la Universidad. El seguimiento de estas normas asegurará que la información publicada cuente con criterios semánticos que permitan el indexamiento de la web UPSE.

En la creación de contenido es necesario considerar que tipo de información se subirá a los microsítios; se tendrá muy en cuenta puntos específicos que permitirán que éste tenga una mayor visibilidad a nivel de buscadores web:

- Título del contenido
- Cuerpo del contenido (Body)
- Formato del texto
- Tags
- Opciones de ruta URL
- Opciones de comentarios
- Meta tags(dublin core)
- Imágenes alt
- Nombres de archivos

Título del contenido: En esta opción se colocara el título del documento. Aunque el editor no limita

ingresar caracteres especiales ni el largo del texto, se recomienda colocar solo caracteres, con un máximo de 70 caracteres. El nombre será referente al documento.

Cuerpo del contenido (Body): El administrador de contenido permite agregar información en un ambiente similar a un procesador de texto, en el cual podemos colocar tablas, imágenes, numeración entre otras opciones que nos proporciona el administrador de contenido.

Formato del texto: En esta opción se deberá elegir el formato de texto "FULL HTML" el cual nos permitirá presentar la información con un formato elegante.

Tags: En esta sección se colocará las palabras relevantes del documento es decir palabras claves, se recomienda 8 palabras claves separadas por comas, en minúsculas y sin tildes. Usar las palabras claves: NOMBRE DE INSTITUCIÓN, LUGAR, permitirá aumentar los resultados de búsqueda y referenciar el lugar de origen del archivo.

Opciones de ruta URL: Permite generar una URL automática para el contenido que se está creando, por lo general toma el texto que se puso en el título del contenido para generar la URL, también evita tener URL duplicadas en los microsítios. Si se desea crear URL de forma manual se debe tener en cuenta la categoría de la información y el título del archivo, evitar tildes o caracteres especiales.

Opciones de comentarios: Esta opción estará a consideración del administrador de cada sitio, ya que dependerá de él que coloque si desea que comenten sobre este artículo o no.

Meta Tags (Dublin Core): Esta opción va sumada con la de opción de Tags ya vista anteriormente, la diferencia es que aquí se tendrá que colocar información más específica del contenido que se está creando como se demuestra a continuación:

- Título de la página
- Descripción
- Abstracto
- Palabras clave

Dentro del estándar internacional para generación de metadatos Dublin Core existen muchos parámetros a considerar, sin embargo los buscadores web toman como prioridad los siguientes:

- Dublin Core Título: Máximo de 70 caracteres.
- Dublin Core Creador: Universidad o Área específica.

- **Dublin Core Tipo:** Dependerá de que información se suba al sitio, por ejemplo tesis, examen, artículo.

Imágenes Alt: Las imágenes que se coloquen en cada página básica deberán tener un nombre referente a la misma, ya que así estaremos generando un valor a la imagen la cual podrá ser encontrada por los buscadores web.

Nombres de archivos: Los nombres de los archivos que se suban adjuntos deberán tener el nombre completo o en caso de ser algo extenso un título referente al documento.

Entre las principales políticas de subida de un archivo al repositorio digital se considerarán:

- Tener presente que antes de subir información se debe revisar que los nombres de los archivos no contengan caracteres especiales y el mismo sea referente al título del documento.
- Los archivos adjuntos deberán estar en formato Adobe Acrobat (.pdf), MS Word (.doc, .docx), MS Powerpoint (.ppt, .pptx) [7].
- Además, se valorarán los siguientes puntos:

- **Autor:** El nombre del autor en mayúscula.
- **Abstract:** El resumen igual al que está en el archivo.
- **Título:** Título del archivo en mayúscula.
- **Palabras Claves:** Utilizar un mínimo de 10 palabras claves (Mabel-Joanna, 2008), de estas palabras todas en mayúsculas. También el uso de estas palabras base: [NOMBRE DE INSTITUCIÓN], [LUGAR] y el nombre del área correspondiente, estas permitirán aumentar los resultados de búsqueda y referenciar el lugar donde origen del archivo.
- **Descripción:** se deberá colocar un comentario o una pequeña descripción del documento.

6. Resultados

En la figura 4 es posible verificar los cambios en el repositorio digital en donde se encuentra mayor generación de metadatos, uso de un estándar, visualización de información en todas las colecciones de las comunidades y subcomunidades. Tras esta revisión también se consiguió una actualización de la información, sin embargo el que ésta sea periódica depende de cada una de los centros y departamentos de la Universidad.

En la figura 5 se muestra el estado de los Micrositios de la UPSE después de la implementación de web semántica. Se puede notar que actualmente se está generando taxonomía y metadatos, se está aplicando un estándar, la

información está actualizada, no existen enlaces perdidos y todos los sitios tienen buen tiempo de carga.

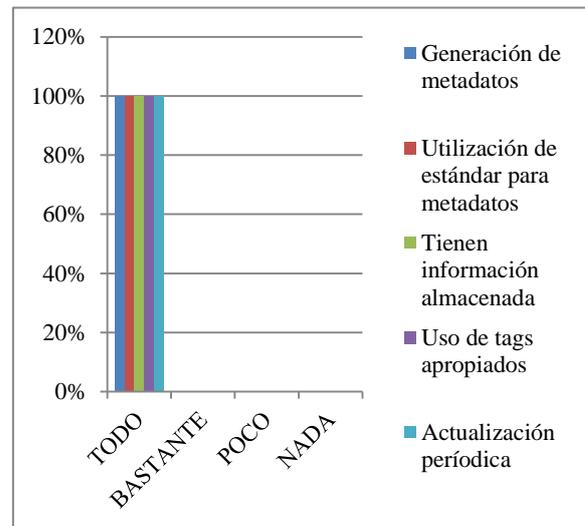


Figura 4. Situación actual Repositorio Digital.

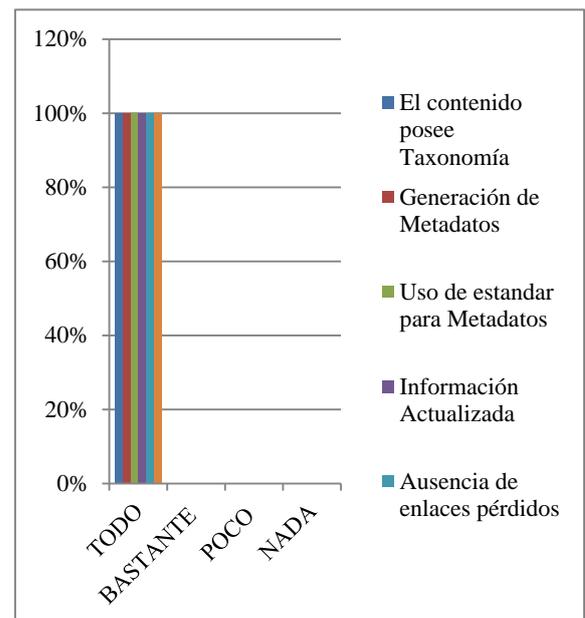


Figura 5. Situación actual Micrositios web.

Por otro lado en las tablas 1 y 2 se demuestra que los archivos indexados en google, principal motor de búsqueda web, han aumentado considerablemente. Esto se da tanto con el repositorio digital y los micrositios. Se demuestra también que en el medidor de rankings Universitario Webometrics [8] [9] se ha escalado puestos considerables, cabe destacar que anteriormente ya existieron escalas en las posiciones del ranking dadas por el desarrollo del proyecto de implementación de metadatos y creación de micrositios y repositorio digital.

Tabla 1. Indicadores de resultados Repositorio Digital.

Indicadores	REPOSITORIO DIGITAL	
	Antes	Después
# Resultados Indexamiento en Google	27.900	56.100
# Resultados Indexamiento en Google Scholar	1.090	1.200
Ubicación en el ranking – webometrics	23	19

Tabla 2. Indicadores de resultados Micrositios Web.

Indicadores	MICROSITIOS	
	Antes	Después
# Resultados Indexamiento en Google en promedio por sitio	50	65
Ubicación en el ranking – webometrics	34	31

7. Conclusiones

En base al proceso de implementación del modelo de web semántica y considerando la experiencia obtenida se puede concluir que:

- Se ha logrado difundir la información de investigaciones para que sean de fácil acceso para la comunidad local, nacional y mundial a través del repositorio digital universitario.
- El usos de Tags y taxonomía establecidos y estandarizados representan un gran soporte en la generación de tráfico hacia el domino upse.edu.ec, mejorando la accesibilidad en la nube semántica.
- El uso de todo los metadatos ayudó a transformar datos simples en estructuras bien organizadas que aportan conocimiento a la información y generar más resultados en los buscadores web tanto como para micrositios y el repositorio digital.
- Se logró en una gran mayoría hacer conciencia de las ventajas de tener un Micrositio y mostrar que la difusión de la información vía web, es la mejor vía para llegar a la comunidad.
- La mayor dificultad a resolver es crear y mantener la cultura de presentar y difundir la

información de cada entidad a través de los medios web con los que gracias a este proyecto cuenta la universidad. Esto con el fin de que la presencia y visibilidad de la Institución de Educación Superior siga creciendo en la web como todas las instituciones en esta era, era de la información lo están llevando a cabo.

- Se ha conseguido mejorar e incrementar la presencia y visibilidad de la UPSE en la web a través de micrositios para cada entidad y el repositorio digital. Paulatinamente además se ha logrado subir en el ranking web de universidades y repositorios de Webometrics.
- Se espera que mediante el constante ingreso de contenidos al repositorio se puedan mejorar los valores SIZE, RICH FILES, SCHOLAR y de IMPACT del Webometrics y así mantener y de ser posible incrementar la ubicación de la UPSE en este ranking web de Universidades.

8. Recomendaciones

Finalmente luego de la implementación del modelo de web semántica, se recomienda:

- La actualización de la información debe ser periódica para lograr así mayores resultados de tráfico web en el futuro para mejorar el posicionamiento en los medidores de ranking de web universitaria.
- Mejorar la información que se presenta en el repositorio digital ya que hay facultades que carecen de información y de documentos de tesis. Esto genera un impacto desfavorable ante el resto de repositorios a nivel nacional.
- No crear comunidades o colecciones en el repositorio digital si no se tiene la información para subir en ellas, ya que estaría demostrando que en esas áreas no se está trabajando por parte de la Universidad.
- Actualizar la versión de la herramienta DSPACE para el repositorio digital de la universidad.
- Además de las políticas de información presentadas en este trabajo, considerar incluir el uso de otros estándares utilizados a nivel mundial, como por ejemplo la iniciativa europea DRIVER 2.0.

9. Referencias

[1] Shedroff, N., 1999. Information interaction design: A unified field theory of design. Information design, 267-292. The MIT Press, San Francisco, CA.

[2] Jimenez, M., 2005. El ensayo fotográfico como Diseño de Información. El uso de la fotografía en la investigación exploratoria de un fenómeno social. Colección de Tesis Digitales Universidad de las Américas Puebla, pp. 13 a 23.

[3] Harper, C., 2010. Dublin Core Metadata Initiative: Beyond the Element Set. Information Standards Quarterly Winter 2010, vol. 22, pp. 20 a 28.

[4] Morales, O., 2003. Fundamentos de la Investigación Documental y la Monografía. Manual para la elaboración y presentación de la monografía. Mérida, Facultad de Odontología de la Universidad de Los Andes.

[5] Hernández, R., Fernández, C., Baptista, M., 2010. Metodología de la Investigación. Mc Graw Hill.

[6] Arango, M., Cruz H., Ordoñez J., 2003. Los Metadatos y el proyecto Biblioteca Digital desde Colombia. Biblioteca Digital Andina.

[7] Laboratorio de Cibermetría – CSIC, s.f. Metodología. Extraído el 01 de Septiembre del 2013 desde <http://repositories.webometrics.info/es/metodologia>

[8] CybermetricsLab, CCHS – CSIC, (s.f.). Ranking web de repositorios. Extraído el 20 de Febrero de 2015 desde http://repositories.webometrics.info/es/Latin_America_es/Ecuador

[9] CybermetricsLab, CCHS – CSIC, (s.f.). Ranking web de universidades. Extraído el 20 de Febrero de 2015 desde http://www.webometrics.info/es/Latin_America_es/Ecuador

10. Agradecimientos

Este trabajo ha sido financiado parcialmente por el Proyecto de Investigación “Implementación de metadatos en la Web de la UPSE, por medio de un repositorio Digital y desarrollo de micro-sitios semánticos” y comprobado a través de la tesis “Implementación de una web semántica sobre los micrositiros y repositorio digital de la UPSE”. Los

autores desean expresar su agradecimiento al Grupo de Investigación en Tecnologías Web y Móviles por su colaboración y prestación de instalaciones.