



REPÚBLICA BOLIVARIANA DE VENEZUELA  
UNIVERSIDAD DEL ZULIA  
FACULTAD EXPERIMENTAL DE CIENCIAS  
DIVISIÓN DE PROGRAMAS ESPECIALES  
LICENCIATURA EN COMPUTACIÓN



## **Herramienta Web para la Clasificación y Recuperación de Información Digital Basada en Etiquetas**

**Proyecto de Trabajo Especial de Grado presentado como  
requisito para optar al título de Licenciado en Computación**

Autor: Br. Edinson Padrón Urdaneta

Tutor: MSc. Gerardo Pirela

Maracaibo, abril de 2014

**Herramienta Web para la Clasificación y Recuperación de Información Digital  
Basada en Etiquetas**

---

*Br. Edinson Padrón Urdaneta*

*Cl. No.: 19.216.488*

*Teléfono: 0414-6574149*

*Venezuela, Zulia, Maracaibo, Urb. La Paz, Calle 96D, Avenida 56, Casa #56-06*

*Correo electrónico: edinson.padron.urdaneta@gmail.com*

---

*MSc. Gerardo Pirela*

*Cl.: 12.404.565*

*Teléfono: 0412-1734718*

*Correo electrónico: gepirela@fec.luz.edu.ve*

Padrón Urdaneta, Edinson. “**Herramienta Web para la Clasificación y Recuperación de Información Digital Basada en Etiquetas**”. Trabajo Especial de Grado. Universidad del Zulia. Facultad Experimental de Ciencias. División de Programas Especiales. Maracaibo, Venezuela. 2014. 13p.

## **RESUMEN**

A diario, el ser humano se ve en la necesidad de manipular una gran cantidad de información digital, distribuida entre una variedad de archivos y dispositivos de hardware. Con este trabajo de investigación se pretende elaborar una herramienta web que permita clasificar y recuperar la información digital mediante el uso de etiquetas, para así ofrecer a los potenciales usuarios una alternativa más intuitiva y práctica de manipular sus datos, sin importar la localización física o lógica de los mismos. Se implementará la metodología Scrum de desarrollo ágil, a lo largo de cuatro lanzamientos conformados por nueve *sprints*.

**Palabras claves:** Aplicación web, Etiquetado, Clasificación, Recuperación, Información digital

**Correo electrónico:** edinson.padron.urdaneta@gmail.com

Padrón Urdaneta, Edinson. “**Herramienta Web para la Clasificación y Recuperación de Información Digital Basada en Etiquetas**”. Trabajo Especial de Grado. Universidad del Zulia. Facultad Experimental de Ciencias. División de Programas Especiales. Maracaibo, Venezuela. 2014. 13p.

## **ABSTRACT**

Nowadays, people have the need to handle a large amount of digital information in files which are distributed along a variety of hardware devices. The goal of this research is to develop a web application to classify and recover digital information using tags, and thus offer potential users a more intuitive and pragmatic alternative to manage their data, regardless of its physical or logical location. This research will follow a Scrum methodology for agile development along four releases composed of nine sprints.

**KeyWords:** Web application, Tagging, Classification, Retrieval, Digital information

**Email:** edinson.padron.urdaneta@gmail.com

## ÍNDICE DE CONTENIDO

	Pág.
Resumen .....	3
Abstract .....	4
Índice de Tablas .....	6
Planteamiento y justificación del problema .....	7
Formulación del problema .....	9
Objetivos .....	9
Objetivo General .....	9
Objetivos Específicos .....	9
Delimitaciones de la Investigación .....	9
Metodología de la Investigación .....	10
Estudio de Factibilidad .....	10
Recursos Humanos .....	10
Recursos Materiales .....	11
Recursos Financieros .....	11
Recursos Legales .....	11
Viabilidad de la Investigación .....	11
Resultados Esperados .....	11
Cronograma de Actividades .....	12
Referencias bibliográficas .....	13

## ÍNDICE DE TABLAS

	Pág.
Tabla 1. Cronograma .....	12

## 1. Planteamiento y justificación del problema

La época actual es llamada la “era de la información” (Giuliano, 1983) debido a la relevancia de ésta en la sociedad. Furth (1994) escribió: “La información es actualmente tan vital, y tan intangible, como el aire que respiramos, el cual está lleno de ondas de radio”. Hoy en día es el bien más importante y valioso para toda compañía, país y grupo social, por lo que su refinamiento a partir de la abundante cantidad de datos que son creados cada segundo es un proceso primordial y que ofrece una ventaja significativa.

Desde la creación de la Internet, el volumen de información disponible ha crecido exponencialmente. En el 2010 se produjo y almacenó un estimado de dos exabytes. Esto incluye todos los medios: libros, revistas, documentos, Internet, fotografías, televisión, radio, música, entre otros; sin embargo, la mayoría era digital (93% para ser más preciso) (Burgin, 2010). Para el año 2012 la cantidad de información digital creada se elevó a 2.5 exabytes al día (IBM, 2012), evidenciando una tendencia que se consolida cada vez más.

A la Internet se le suma el creciente uso de dispositivos móviles (teléfonos inteligentes, tabletas, lectores electrónicos, entre otros) que permiten la accesible y cómoda creación y consumo de información gracias a la web, lo cual provoca una cuantiosa producción de datos, además de contribuir con la heterogeneidad y descentralización de los mismos, ocasionando así que su tratamiento sea una tarea penosamente difícil de realizar.

En la actualidad, los distintos repositorios presentan al usuario la información digital en ellos almacenados mediante una estructura jerárquica de directorios en forma de árbol invertido. Esta representación, nacida en 1969 con el lanzamiento del sistema operativo Multics en respuesta a la necesidad de un sistema de almacenamiento secundario en un ambiente de multiprogramación (Multics, 1968), rígida por causa de su inherente propiedad taxonómica, limita la manera en la que puede ser organizada la información, mediante una relación padre-hijo, y hace de la búsqueda de la misma un proceso arduo debido al uso de engorrosas rutas de directorios. Además, esta aproximación dista en gran medida del método natural usado por el hombre para almacenar y recuperar información de su memoria a largo plazo, el cual consiste en codificarla semánticamente para su almacenamiento (Baddeley, 1966), asociando los datos que percibe (Atkinson y Shiffrin, 1968). Por lo tanto, el ser humano no es una máquina que trabaje en base a datos aislados, sino interrelacionados, describiéndolos de manera

inherentemente subjetiva, breve y dinámica; haciendo necesario un método semejante para clasificar y recuperar la información digital, siendo éste más flexible, intuitivo y natural.

Afortunadamente, la web ha dado origen a un nuevo sistema de clasificación. Originalmente diseñado e implantado por el servicio del.icio.us (Mathes, 2004), el sistema ofrece la posibilidad de describir recursos mediante un conjunto de palabras clave llamadas etiquetas (tags, por su término en inglés), las cuales son consideradas por los usuarios como relevantes para caracterizar dichos recursos de acuerdo a sus necesidades sin depender de un vocabulario controlado o de una estructura previamente definida, estableciendo así una relación entre el recurso y un concepto en su mente, con el objeto de organizar el contenido para uso futuro de una manera fácil y flexible (Specia y Motta, 2007). Este esquema de catalogación se ha extendido en uso por una gran variedad de servicios web debido a su popularidad entre los usuarios, que ven en éste una manera más natural, sencilla, rápida y personal de clasificar la información digital que le es de interés, hallándola posteriormente con mínimo esfuerzo, siendo ésta una mejor alternativa al problema de darle tratamiento a una gran cantidad de datos, contrastada con la estructura jerárquica que ofrecen los sistemas de archivos modernos.

Hoy en día, existe una serie de herramientas que extienden las capacidades de los sistemas de archivos al permitir el uso de etiquetas; sin embargo, estas herramientas presentan algunas limitaciones, tales como: el tratamiento exclusivo de una fracción de la información digital existente, su operación bajo un número reducido de plataformas, la prestación de un conjunto restringido de funcionalidades, entre otros.

Con este trabajo de investigación se busca desarrollar una herramienta web integral que garantice la interoperabilidad entre los distintos sistemas operativos y dispositivos al permitir un acceso convenientemente centralizado a la información digital sin importar la distribución física de ésta, brindando así transparencia de localización mediante el uso de un servidor central que contenga la herramienta y aplicaciones nativas que enlacen los repositorios de los usuarios con dicho servidor. Además, se implantará una manera sencilla, versátil y personal de clasificar y recuperar la información digital contenida en los distintos repositorios al permitir su descripción mediante el uso de etiquetas definidas a conveniencia e implementando criterios de búsqueda basados en expresiones lógicas. Adaptando así la tecnología nacida en la web para el tratamiento conveniente, efectivo y eficiente de los datos gestionados por los sistemas



de almacenamiento secundario contenidos en los distintos dispositivos usados todos los días.

## 2. Formulación del problema

¿Será la clasificación y recuperación de información digital, basada en el uso de etiquetas a través de una herramienta web, más intuitiva y práctica que los actuales sistemas de archivos?

## 3. Objetivos

### 3.1. Objetivo General

- Desarrollar una herramienta web para la clasificación y recuperación de información digital basada en etiquetas

### 3.2. Objetivos Específicos

- Realizar una revisión documental sobre herramientas de clasificación y recuperación de información digital
- Implementar la aplicación servidor para el almacenamiento y gestión de etiquetas de recursos digitales
- Desarrollar el módulo de la aplicación cliente encargada del rastreo personalizable de la información a etiquetar
- Desarrollar el módulo de la aplicación cliente correspondiente a la interfaz web de usuario
- Realizar las pruebas de caja gris de la herramienta

## 4. Delimitaciones de la Investigación

La investigación se llevará a cabo en la Unidad Académica de Lenguajes y Modelos Computacionales de la Facultad Experimental de Ciencias de la Universidad del Zulia, en el contexto de la línea de investigación: “Modelos Computacionales”. La misma se desarrollará durante un año, en el periodo comprendido desde mayo de 2014 hasta mayo de 2015, en los

espacios del laboratorio de la mencionada unidad académica.

## 5. Metodología de la Investigación

En esta investigación será empleada la metodología Scrum de desarrollo ágil de software, tal y como la describen Schwaber y Sutherland (2011). En Scrum se recolectan todas las funcionalidades que debe poseer el sistema final, descritas desde el punto de vista del cliente y llamadas *user stories*; éstas se priorizan para planificar qué subconjunto será diseñado, desarrollado, integrado y testeado para el próximo lanzamiento, el cual es a su vez subdividido en un conjunto de iteraciones, llamadas *sprints*. El objetivo de un *sprint* es el de presentar al cliente un subconjunto del sistema en total y perfecto funcionamiento en un periodo no mayor a un mes, logrando de esta manera la retroalimentación necesaria para el desarrollo exitoso y adaptable del proyecto. En la etapa de planificación de un lanzamiento se estima el tiempo que tomará diseñar, desarrollar, integrar y testear cada *user story*, estimación usada como base para la estructuración de los *sprints*.

Se estima realizar un total de cuatro lanzamientos. En el primero, se entregará la interfaz web de usuario; en el segundo, el módulo de etiquetado; en el tercero, el módulo de recuperación y en el cuarto, la aplicación cliente. Se prevé que el primer lanzamiento estará constituido por tres sprints, mientras que el resto de tan sólo dos; se estima además que cada sprint durará un mes.

## 6. Estudio de Factibilidad

### 6.1. Recursos Humanos

El equipo de trabajo estará constituido por el MSc. Gerardo Pirela, quien desempeñará funciones como tutor y asesor; y el Br. Edinson Padrón, autor de esta investigación, además de diseñador y desarrollador de la herramienta.

## 6.2. Recursos Materiales

Para el desarrollo de esta investigación se requiere de un computador con las siguientes especificaciones mínimas: procesador de 2GHz, 1GB de memoria RAM, 10GB de espacio libre en disco duro y conexión a Internet.

## 6.3. Recursos Financieros

La mayor parte de los egresos serán asumidos por el laboratorio de la Unidad Académica de Lenguajes y Modelos Computacionales. El investigador se hará cargo de los costos que el laboratorio no pueda cubrir.

## 6.4. Recursos Legales

Se empleará software de código abierto de manera exclusiva; esto incluye el sistema operativo y ambientes de desarrollo, acatando las cláusulas dispuestas en sus respectivas licencias.

## 7. Viabilidad de la Investigación

Debido a que se dispone de todos los recursos humanos, materiales, legales y financieros, necesarios para el desarrollo de esta investigación, se concluye que la misma es viable.

## 8. Resultados Esperados

Se espera desarrollar una herramienta web que, a través de un navegador, muestre y dé acceso a los usuarios la información digital (documentos, fotos, música, enlaces, entre otros) contenida en los distintos repositorios de su propiedad, y que les permita clasificarla y hallarla mediante el uso de etiquetas. Todo esto a través de la implementación de un servicio web ubicado en un servidor y aplicaciones clientes que se encargarán de comunicar los repositorios con el servidor.

## 9. Cronograma de Actividades

Tabla 1. Cronograma

Año	2013							2014					
Mes	Junio	Julio	agosto	septiembre	octubre	noviembre	diciembre	enero	febrero	marzo	abril	mayo	junio
Lanzamiento				1º			2º		3º		4º		
Sprint				1º	2º	3º	4º	5º	6º	7º	8º	9º	
Asesorías con el tutor													
Revisión de literatura													
Diseño e implementación de la base de datos													
Diseño e implementación de la interfaz web de usuario													
Codificación e implementación de la aplicación servidor													
Codificación e implementación de la aplicación cliente													
Pruebas de caja gris de la herramienta													
Elaboración del informe final del TEG													

Fuente: Padrón (2014)

## REFERENCIAS BIBLIOGRÁFICAS

- Atkinson, R.C.; Shiffrin, R.M. (1968). Chapter: Human memory: A proposed system and its control processes. The psychology of learning and motivation. pp. 89-195.
- Baddeley, A.D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. The Quarterly Journal of Experimental Psychology. pp. 302-309.
- Burgin, Mark (2010). Theory of Information: Fundamentality, Diversity and Unification. World Scientific. p. VI.
- Furth, J. (1994). The Information Age in Charts, Fortune International.
- Giuliano, V.E. (1983). The United States of America in the Information Age, en Information Policy and Scientific Research, Elsevier, Amsterdam, pp. 59-76.
- IBM (2012) (Página consultada el 25 de junio de 2013) [On-line]. Dirección: [www-01.ibm.com/software/data/bigdata/](http://www-01.ibm.com/software/data/bigdata/)
- Mathes, Adam (2004) (Página consultada el 25 de junio de 2013) [On-line]. Dirección: [www.adammathes.com/academic/computer-mediated-communication/folksonomies.html](http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html)
- Multics (1968) (Página consultada el 25 de junio de 2013) [On-line]. Dirección: [www.multicians.org/fjcc4.html](http://www.multicians.org/fjcc4.html)
- Schwaber, Ken; Sutherland, Jeff (2011). The Definitive Guide to Scrum: The Rules of the Game.
- Specia, Lucia; Motta, Enrico (2007). Integrating Folksonomies with the Semantic Web. Knowledge Media Institute - The Open University.