



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**

**Informe 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 Morillo

Maracaibo, marzo 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: +58 414 6574149*

*Venezuela, Zulia, Maracaibo, Urb. La Paz, Calle 96D, Avenida 56*

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

---

*MSc. Gerardo Pirela Morillo*

*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. 14p.

## **RESUMEN**

**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. 14p.

## **ABSTRACT**

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

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

## ÍNDICE DE CONTENIDO

	Pág.
Resumen .....	3
Abstract .....	4
Introducción .....	6
Capítulo I. El Problema	
Planteamiento del Problema y Justificación de la Investigación .....	7
Alcance del problema .....	9
Objetivos .....	9
Objetivo General .....	9
Objetivos Específicos .....	9
Capítulo II. Marco Teórico	
Antecedentes de la Investigación .....	10
Bases Teóricas .....	12
Referencias bibliográficas .....	14

## INTRODUCCIÓN

# **CAPÍTULO I**

## **El Problema**

### **1. Planteamiento del Problema y Justificación de la Investigación**

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.

En el año 2010 se produjo y almacenó un estimado de dos exabytes de información (esto incluye todos los medios: libros, revistas, documentos, Internet, fotografías, televisión, radio, música, entre otros) de los cuales, el 93% era digital (Burgin, 2010). Dos años después, la cantidad de información digital que se generó, se elevó a 2.5 exabytes al día (IBM, 2012), evidenciando un crecimiento exponencial con la Internet como principal catalizador de este fenómeno.

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, quienes ven en éste una manera más natural, sencilla, rápida y personal de clasificar la información digital que les 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 buscó desarrollar una herramienta web integral que garantizara 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 que enlacen los repositorios de los usuarios con dicho servidor. Además, se implantó una manera sencilla, versátil y personal de clasificar



y recuperar la información digital contenida en los mencionados repositorios al permitir su descripción y recuperación mediante el uso de etiquetas definidas a conveniencia. Adaptando así la tecnología nacida en la web al ámbito de los actuales sistemas de archivo.

## 2. Alcance del problema

La investigación estuvo delimitada al desarrollo de un sistema cliente-servidor donde el cliente, desarrollado para trabajar solo bajo los entornos Windows®, OSX® y Linux®, se comunica con una aplicación web, manteniendo sincronizadas las etiquetas asociadas a los archivos ubicados en los repositorios del usuario, permitiéndole así clasificarlos y recuperarlos mediante el uso de las mencionadas etiquetas.

## 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

## CAPÍTULO II

### Marco Teórico

#### 1. Antecedentes de la Investigación

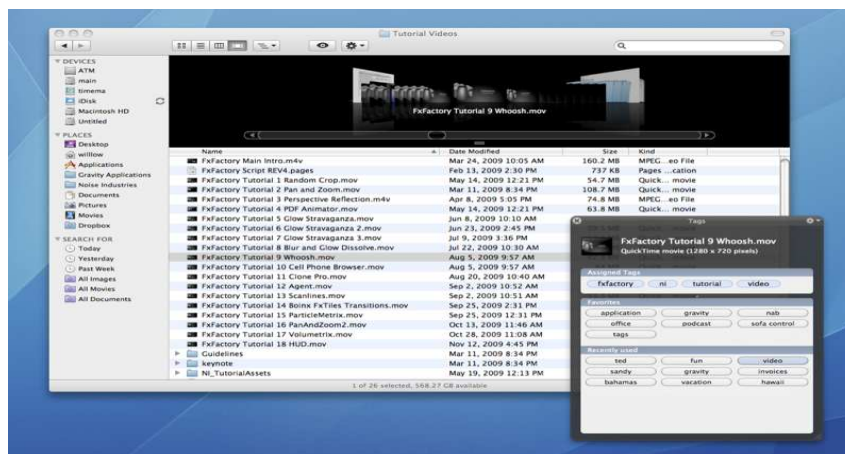
En la actualidad, los sistemas de clasificación basados en etiquetas son ampliamente usados por innumerables aplicaciones web, sin embargo, su uso como complemento a los sistemas de archivo modernos presenta una reducida gama de opciones, dentro de la cual se pueden mencionar:

##### 1.1. Tag

Tags es una aplicación que permite etiquetar archivos y directorios en *finder* (explorador), correos electrónicos en *Mail* (gestor de emails), fotos en *iPhoto* (gestor de imágenes) y enlaces en *Safari* (navegador web); con el objetivo de mantener organizados dichos recursos y poder hallarlos de manera rápida y sencilla.

Tags también ofrece una vista previa de la información, atajos de teclado y una excelente integración con el sistema operativo. Desafortunadamente, el sistema solo está disponible para el sistema operativo OS X®, está sujeto a una licencia privativa y no es capaz de sincronizar recursos localizados en dispositivos diferentes.

Gráfica 1. Tags



Fuente: Tags (2014)

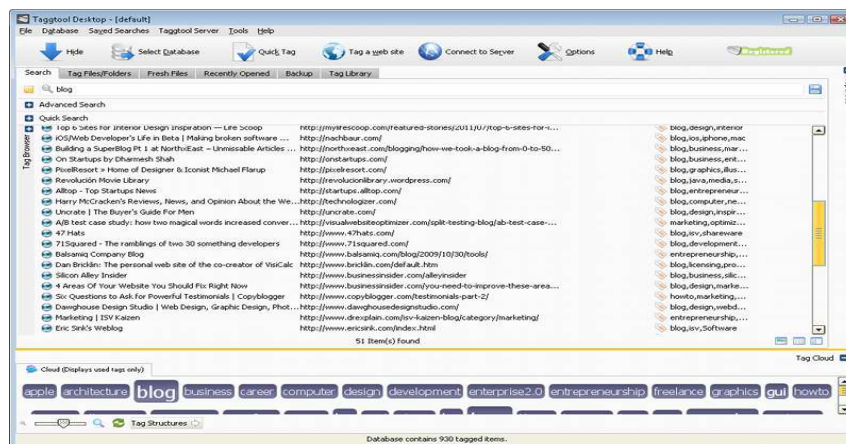
## 1.2. TagTool

TaggTool permite añadir etiquetas a archivos, directorios y «bookmarks». Adicionalmente, el usuario puede proporcionar una descripción del recurso, además de una calificación cuantitativa (*rating*). Esta herramienta ofrece la posibilidad de importar metadatos pre-existentes en los recursos y realizar búsquedas en base a estos.

TaggTool también posee una alta integración con el sistema operativo y brinda una vista previa de los recursos gestionados. Además, facilita un sistema de etiquetado automático basado en la naturaleza de los recursos, y muestra una nube de etiquetas para la visualización de las mismas.

Al igual que Tags, TaggTool está sujeto a una licencia privativa y no es capaz de sincronizar recursos localizados en dispositivos diferentes. Otra desventaja que presenta este sistema es que está disponible solo para el sistema operativo *Microsoft Windows®*.

**Gráfica 2. TaggTool**



**Fuente: TaggTool (2014)**

## 1.3. Tables

Tabbles ofrece la posibilidad de clasificar, buscar, organizar y compartir (a través de una LAN o la Internet) archivos, carpetas y «bookmarks», mediante el uso de etiquetas. Adicionalmente, la herramienta ofrece un sistema de auto-etiquetado, además de una excelente integración con el sistema operativo.

A pesar de ser capaz de gestionar, compartir y sincronizar recursos localizados en dispositivos distintos, Tabbles está disponible únicamente para el sistema operativo Microsoft Windows® bajo una licencia privativa.

**Gráfica 3. Tabbles**



**Fuente: Tabbles (2014)**

## 2. Bases Teóricas

### 2.1. Aplicación web

Una aplicación web es un programa desarrollado mediante el uso de herramientas y lenguajes (como por ejemplo: CSS, HTML, JavaScript) soportados por navegadores web y que dependen de éstos para su uso por parte de los usuarios (PC Magazine, 2014).

### 2.2. Sistema de archivo

TODO (linfo.org, 2004).

### 2.3. Metadatos

TODO (W. R. Durrell, 1985).

## 2.4. Etiqueta

Las etiquetas son un conjunto de palabras clave empleadas en la organización de contenido. El uso de etiquetas le permite a los usuarios clasificar sus recursos de manera flexible y personal, caracterizándolos de acuerdo a sus propias necesidades sin tener que depender en un vocabulario controlado o una estructura previamente definida (Specia y Motta, 2007).

## 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/)
- linfo.org (2004) (Página consultada el 23 de enero de 2014) [On-line]. Dirección: [goo.gl/pJqjHc](http://goo.gl/pJqjHc).
- 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)
- PC Magazine (2014) (Página consultada el 23 de enero de 2014) [On-line]. Dirección: <http://www.pcmag.com/encyclopedia/term/54272/web-application>
- 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.
- W. R. Durrell (1985). Data Administration. A Practical Guide to Data Administration. McGraw-Hill.