



**Snoop**  
CONSULTING

| Visión innovadora



## **Facts & Authorities: Text Mining en Web 2.0**

ENJOY YOUR COFFEE,  
WE TAKE CARE OF THE PROJECT

[www.snoopconsulting.com](http://www.snoopconsulting.com)

[info@snoopconsulting.com](mailto:info@snoopconsulting.com)

# Administración de Conocimiento es más que Administración de Documentos

El **conocimiento** en una empresa...

- Tiene forma imprecisa:
  - ➔ Quiero poder encontrar aún lo que no estoy buscando.
- Es más asequible a través de expertos:
  - ➔ Quiero poder saber a quién preguntarle.

**¿Cómo?** Mediante...

- Asignación automática de tags
- Browsing sensible al contexto.
- Reconocimiento automático de autoridades en diferentes temas = poder encontrarlos + darles incentivo del reconocimiento.

En Snoop Consulting, lo llamamos **Facts & Authorities**.

**¿Por qué?**

- Analizando discusiones podemos mejorar significativamente como detectar los expertos, y como empleados nuevos y experimentados exploran la base de conocimiento de su empresa.

# Problemas del Tagging

- Los tags permiten una vista general de "acerca de qué" es una la colección de documentos.
- **¿Por qué entonces, la mayoría de las veces, los tags son sólo una curiosidad?**
  - ➔ Tagging es muy personal: la gente acuerda en pocos tags y desacuerda en la mayoría.
  - ➔ La interfaz de usuario normal para motrstrarlos (nube de tags) es deficiente porque:
    - Solo se puede mostrar un grupo reducido de tags a la vez: La selección de los tags determina su utilidad.
    - Los tags seleccionados no son sensibles al contexto.

# Solución: Retrotagger

- **El Retrotagger permite incorporar tags a texto ya existente**, mediante la identificación automática de conceptos que son útiles al momento de resumir "sobre qué" es una conversación o texto.
- **Problema:**
  - ➔ Errores de escritura son parte integral del contenido, no un problema pasajero. P.ej. en un sitio web, "Microsoft Windows" está escrito como Windwos, windwows, windos, windoz, window\$...
- **Solución:**
  - ➔ Facts & Authorities entiende lenguaje coloquial, aún con verbos rioplatentes.
  - ➔ Reglas de corrección sensibles al contexto gramatical.

# Retrotagger sobre Yahoo! Answers

## Autotagger de preguntas acerca de Linux en Yahoo! Answers

Texto de la consulta:

Cantidad de Respuestas:

Buscar




conceptos importantes

### ¿como manejar dinamicamente memoria en C (linux) ?

tengo que hacer una funcion en la cual un arreglo A[ ] valla cambiando de tamaño pero no puedo establecer exactamente el tamaño porque no quiero desperdiciar memoria con los espacios vacios que no ocupo

por: marro\_8 el: 2007-10-25 16:35:17

 **Tags sugeridos: linux**

[Mostrar / Ocultar Respuestas](#)

### ¿Puedo hacer un servidor Multi Terminal Linux que de recursos a terminales Windows? ¿Como?

Lo que quiero es hacer un servidor Multiterminal; pero quiero que los usuarios de las terminales, computadoras de baja capacidad, sigan usando el Windows que siempre utilizan, el servidor Linux les daría recursos como memoria y espacio en disco duro; pero las terminales seguirían siendo Windows, he estado buscando información en el internet, pero siempre es Servidor y Terminales en un mismo sistema operativo. ¿Existe alguna manera de hacer un hibrido de eso? ¿Servidor en un sistema, y terminales en otro? Si alguien pudiera ayudarme, se lo agradecería muchísimo.

por: Artemis el: 2007-02-08 17:18:22

 **Tags sugeridos: servidor multiterminal , windows , servidor linux , disco duro , internet , hibrido , linux**

[Mostrar / Ocultar Respuestas](#)

Tags que describen la pregunta y sus respuestas, identificados por el **Retrotagger** automáticamente.

## Solución: Navegación por Tags sensible al Contexto

- Una nube de tags estática es de poca utilidad. No me sirve como siguiente punto de exploración.
- Seleccionar tags por frecuencia es poco útil: la navegación tiene poco efecto sobre la selección.
- En Facts & Authorities la nube de tags es siempre relevante al criterio de selección:
  - ➔ **Cuando casi todos los documentos son visibles,** la Nube de Tags muestra aquellos tags que representen grupos de documentos relevantes con mas o menos el mismo número de items.
  - ➔ **Cuando el usuario selecciona qué le interesa, y hay varios documentos excluidos,** la Nube de Tags muestra sólo los tags relevantes, aquellos que sólo describen al conjunto de documentos interesantes.

# Detección y Ranking de Autoridades

- **Autoridad:** Una persona que contesta mucho y contesta bien.

En discusiones se pueden distinguir:

- Answer people:
  - ➡ Le contesta a muchos que no se contestan mucho entre sí.
  - ➡ Pregunta relativamente poco.
- Reply Magnets (¿flamers?):
  - ➡ Atrae muchísimas respuestas, no participa mucho en conversaciones excepto al principio.

...esto no alcanza, no diferencia bocones de autoridades, porque..

- **¿Qué es una Autoridad?** Es una answer-person que es confirmado por sus pares.
- El concepto de autoridad se puede extender a otras fuentes, como Wikis.

# Solución: Detectando verdaderos expertos

- Luego de identificando potenciales expertos, las opiniones de otras personas sirven para respaldarlos o contradecirlos como expertos, así y diferenciarlos de aquellos que opinan sin saber.
- **Ejemplos reales de opiniones:**
  - ➡ *“Muy groso lo de exchange... lo estoy viendo!”*
  - ➡ *“genial man, te agradezco, lo voy a intentar y despues lo comento”*
  - ➡ *"Gracias pero paso. No me gusta para nada el Opera."*
- Los métodos usuales de text mining necesitan la asignación manual de asignación de categorías (a favor/en contra/neutro) a comentarios, lo que es manualmente intensivo.
- **Facts & Authorities** es capaz de detectar la polaridad de una opinión, sin necesidad de asignar categorías manualmente, lo que es económicamente efectivo.

# Facts & Authorities

Interfaz de Usuario permite exploración dinámica

Nube de Tags Sensible al Contexto

Encontrar documentos así:

buscar por términos  con estos tags:  tag  tag   internet  modem  Tag aquí?  Tag aquí?  Tag aquí?

### Listado de threads

Asunto	Autor	Fecha	Respuestas
<a href="#">[INFO] Como instalar ubuntu</a> Foro: Psicofxp Linux	elnegro1973	16/Mayo/2007	8
Hola amigos quiero saber si cuesta mucho instalar y configurar ubuntu, si como i...			
<a href="#">[INFO] Como instalar ubuntu</a> Foro: Psicofxp Linux	elnegro1973	16/Mayo/2007	9
Hola amigos quiero saber si cuesta mucho instalar y configurar ubuntu, si como i...			
<a href="#">[PEDIDO] Internet con Kubuntu</a> Foro: Psicofxp Linux	nkop...	15/Mayo/2007	2
Hola estoy tratando de configurar mi Kubuntu para usar internet.Tengo cablemodem...			
<a href="#">[CONSULTA] Programa escondido me usa la red</a> Foro: Psicofxp Linux	sanity...	15/Mayo/2007	23
Tengo Ubuntu 7.04 Desde que inicié la PC tengo una taza de descarga de 15Kb/s. ...			
<a href="#">[CONSULTA] Flash Cablemodem en SuSe 7.1</a>	kseba	11/Mayo/2007	10

Página 1 de 7 Vista Previa Mostrando threads: 1 - 20 of 122

### Tags Asociados más Relevantes

arnet conexión dhcp driver drivers ethernet fibertel internet linux modem ethernet modem usb modem zyxel modems usb placa pppoe speedy usb zyxel zyxel prestige 630

### Tags Asociados por Frecuencia

cd conexión configuración consola debian distro distros driver drivers foro instalación internet kernel linux maquina pc placa ubuntu usb windows

### Autores Influyentes

caleb	Dr. Falken	lechuga	dEiCiE	storax
★★★★★	★★★★★	★★★★★	★★★★☆	★★★★☆
Temas: <b>modem</b> , <b>linux</b> , <b>internet</b>	Temas: <b>modem</b> , <b>linux</b> , <b>internet</b>	Temas: <b>modem</b> , <b>linux</b> , <b>internet</b>	Temas: <b>modem</b> , <b>linux</b> , <b>internet</b>	Temas: <b>modem</b> , <b>linux</b> , <b>ubuntu</b>
Ultimo mensaje en: <a href="#">[CONSULTA] LILO: Timestamp mismatch...</a>	Ultimo mensaje en: <a href="#">[PEDIDO] me pueden pasar una buena...</a>	Ultimo mensaje en: <a href="#">[INFO] Indiana, la nueva distro li...</a>	Ultimo mensaje en: <a href="#">[CONSULTA] actualizar ubuntu...</a>	Ultimo mensaje en: <a href="#">sitema operativo en un disco usb...</a>

Discusiones, opiniones, emails, etc...

Expertos Identificados Automáticamente

Tags identificados por el Retrotagger

# Puntos Destacados

- **Facts & Authorities** permite agregar valor a la información sin trabajo extra, llevando su sitio web a web 2.0.
- Facts & Authorities está disponible tanto como una serie de componentes como en una versión integrada con una interfaz de usuario RIA.
- Puede incorporar información disponible de foros y blogs en formato de texto, feeds Atom 1.0 y RSS sin necesidad de pasos extra.
- Exportadores disponibles y validados para PhpBB, mvnForum, POP3 (email) y MediaWiki.
- Control de Acceso por usuario y por grupo de documentos.
- Integrable con CMS y cualquier Base de Datos.