



Ciclo de Extensión Centro Regional Chivilcoy

Entendiendo la Web: Motores de Búsqueda y Aplicaciones en la Universidad y la Empresa

Mg. Gabriel H.Tolosa
tolosoft@unlu.edu.ar

Programa MySISC

Modelos y Servicios de Información sobre Sistemas Complejos

- Noviembre de 2007 -

» AGENDA

» Conceptos sobre la web

» La Web y la recuperación de información

» Aplicaciones en la universidad y las organizaciones

» El impacto de los buscadores y posicionamiento

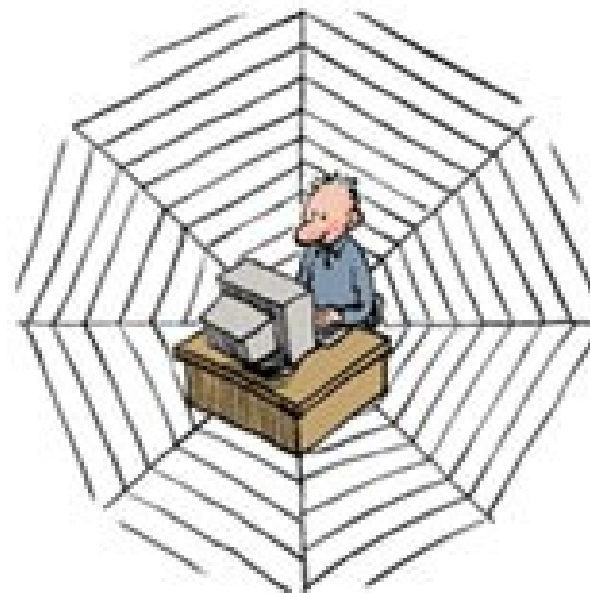
» Minería de la Web

» Algunas tendencias de evolución

» ¿Qué es la web?

» Repositorio de información

- Distribuido (millones de servidores)
- Basado en hipertexto
- Vinculado
- Publicación abierta
- Interface: Navegador
- Hoy es multimedia

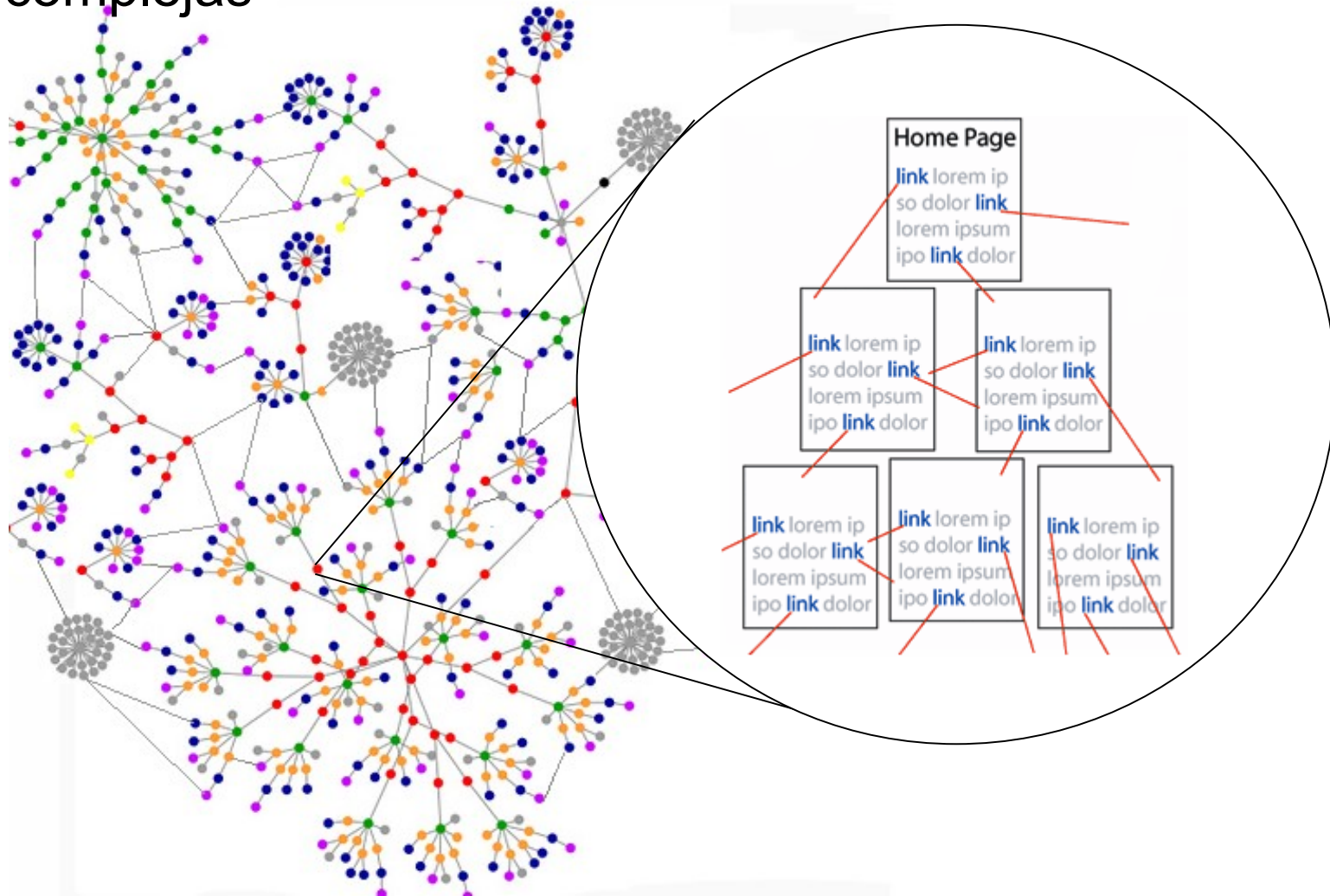


» Elemento “fundamental”: Link URLs

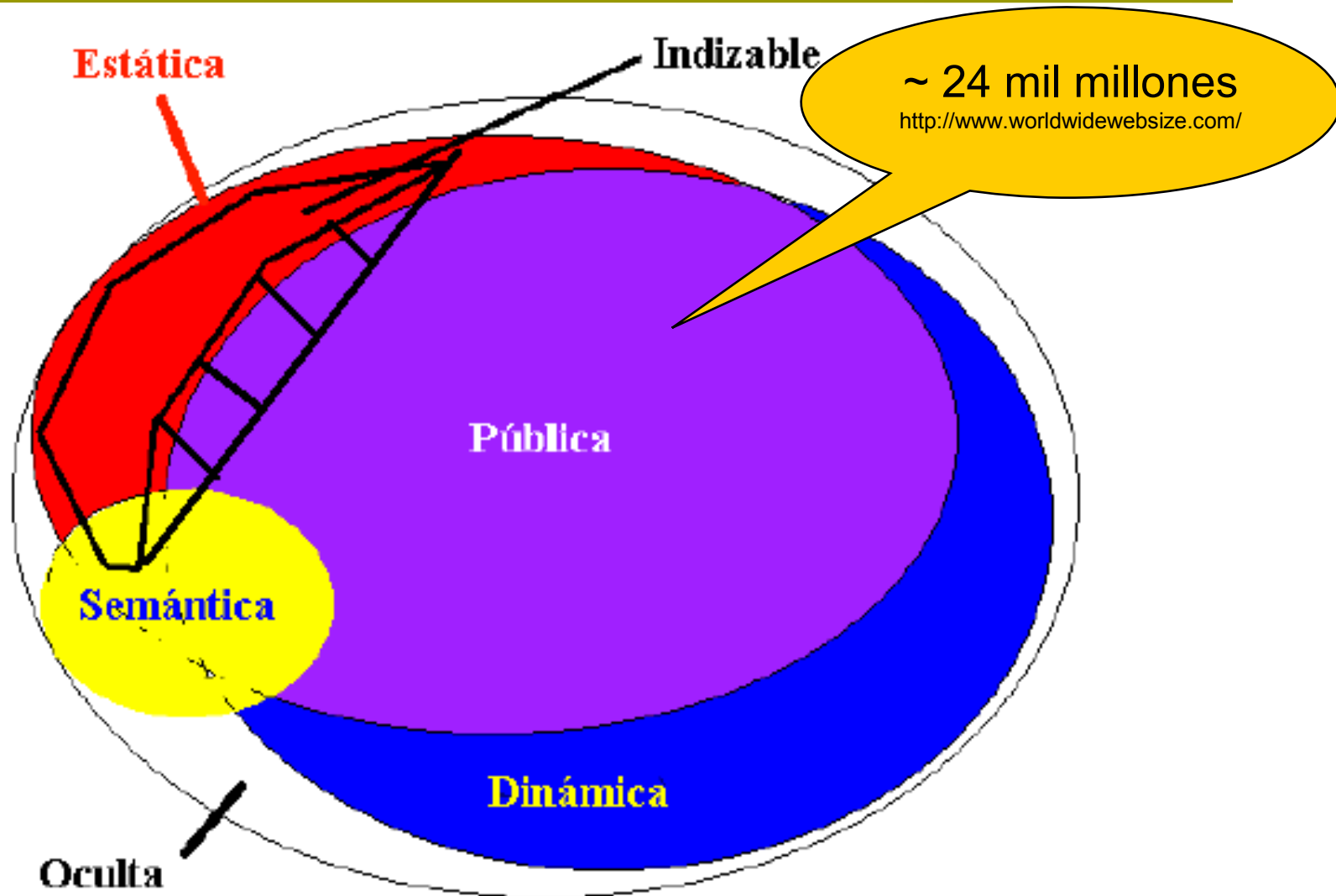
`Universidad Nacional de Luján`

» Estructuras en la web

» Son complejas

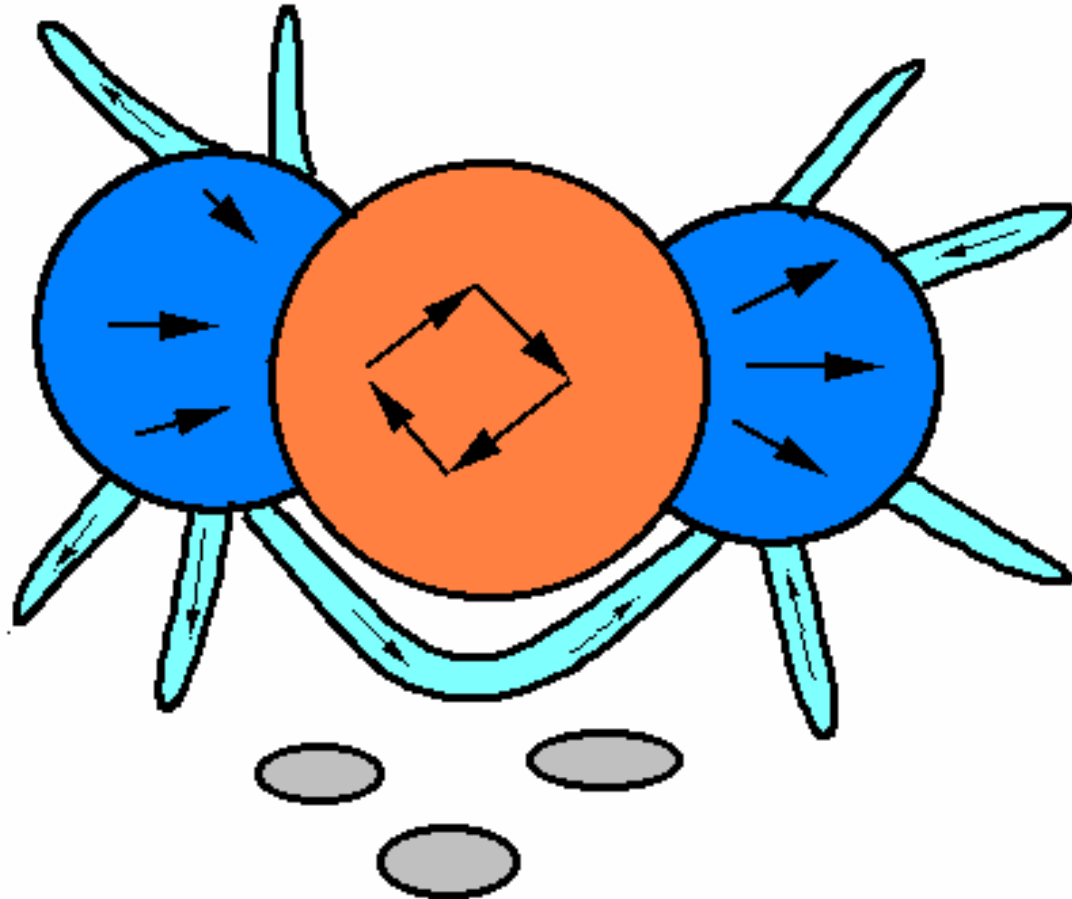


» Otra forma de visualizarla [Baeza-Yates, 2003]



» De acuerdo a los enlaces

- » Modelo de Bow-tie [Broder, 2000]
- » Altamente dinámica!



» ¿Cómo es la web?

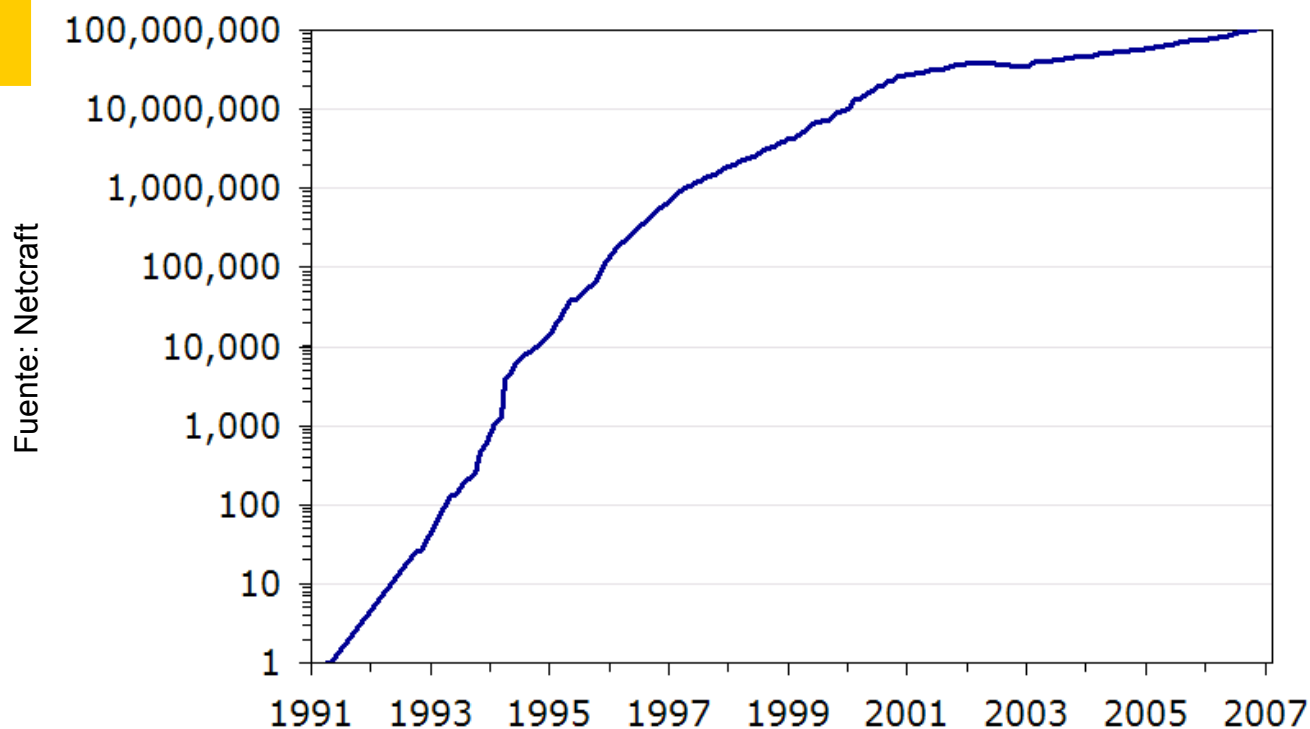
» Algunas características

- Gigantesco!
- En general, de baja calidad
- Débilmente organizado
- Alta duplicación (20-30%)
- Heterogénea
 - Datos, formatos, idiomas, usuarios
- Los sitios no son uniformes
 - ~ 1% de los sitios poseen el 50% de las páginas
- Ocurre lo mismo con los enlaces



»Crecimiento

» Cantidad de sitios desde 1991 al 2006



Tres etapas en el crecimiento:

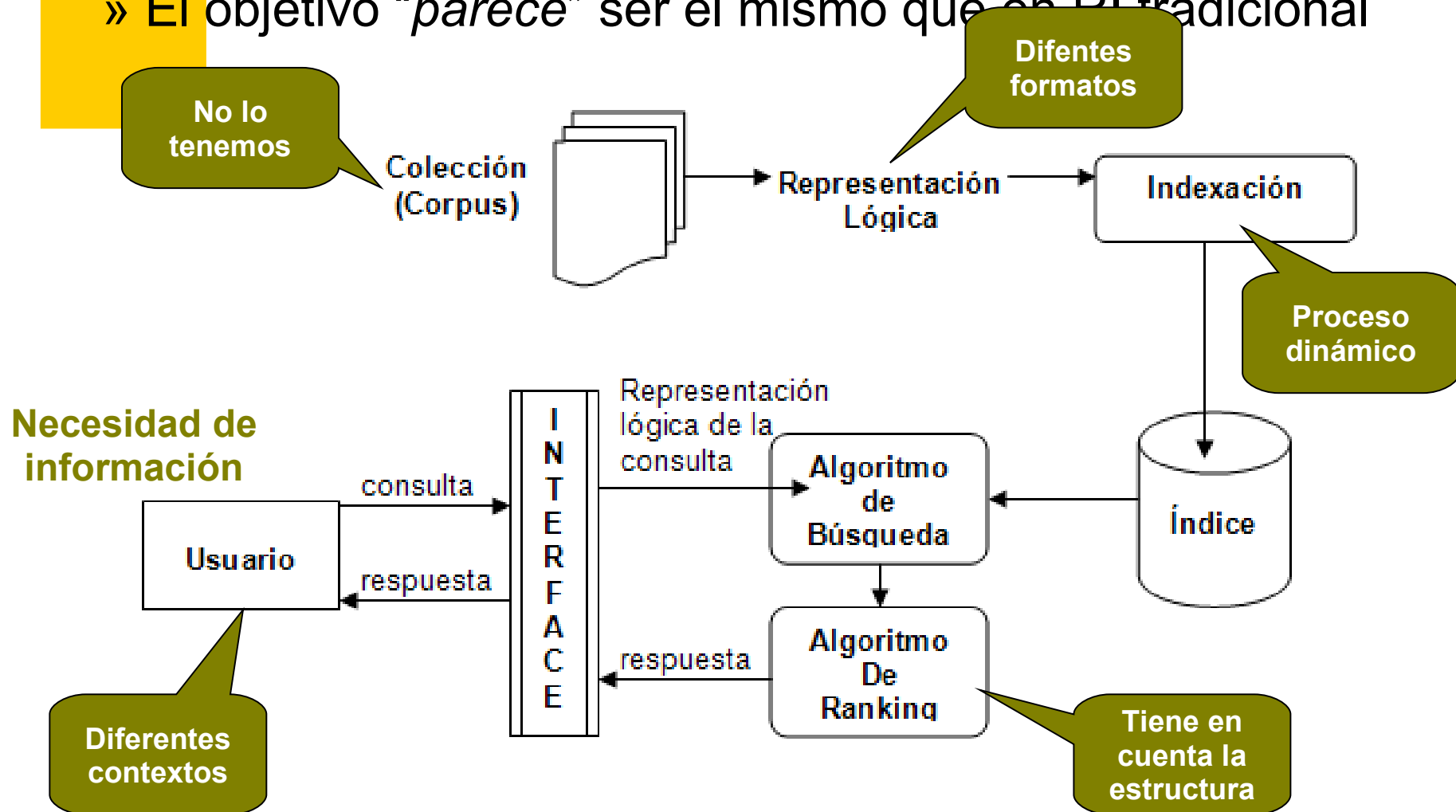
1991-1997
Explosivo
(850% / año)

1998-2001
Rápido
(150% / año)

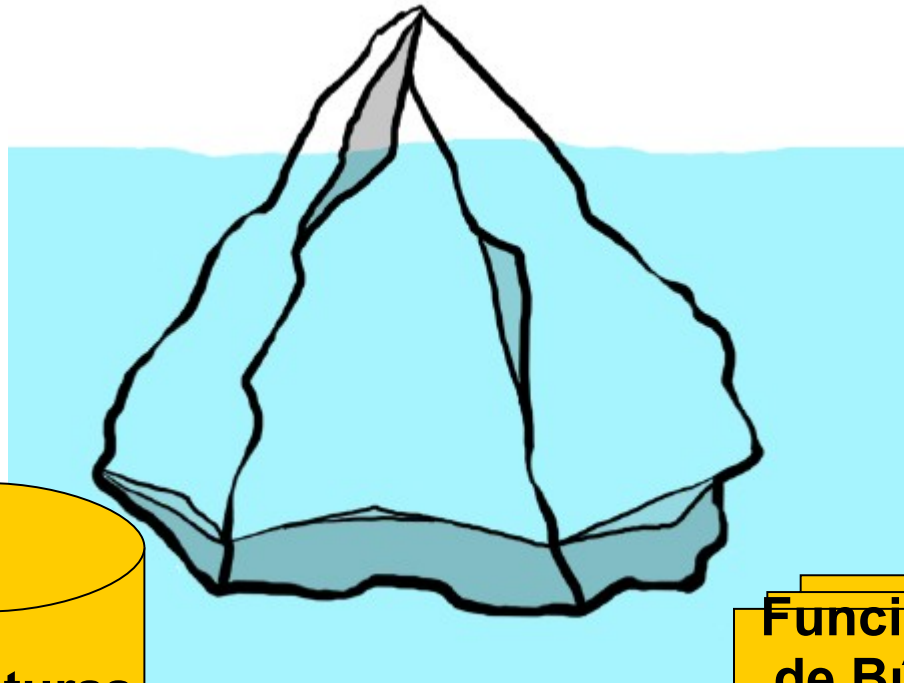
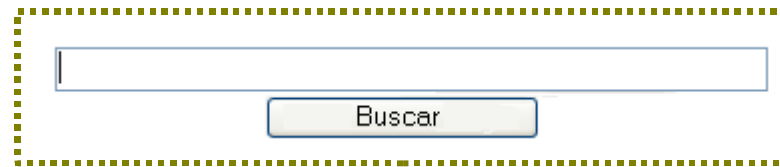
2002-2006
Maduro
(25% / año)

» ¿Y respecto de la RI?

» El objetivo “*parece*” ser el mismo que en RI tradicional



» La Búsqueda es como un Iceberg



**Estructuras
Contenido**



**Funcionalidad
de Búsqueda
y Ranking**

» Herramientas de Búsqueda Web

» Motores de Búsqueda

- Herramientas automáticas de recolección de contenido
 - Extracción del texto y de la estructura
 - Manejo de diferentes formatos complejos
 - Armado de estructuras de datos
-
- Procesamiento de la consulta
 - Recuperación
 - Ranking por relevancia
 - Presentación de resultados



» Bueno, pero no es para tanto...



El significado de la vida? Espere un minuto que lo busco en Google.

» Ejemplo: AskX.com

The screenshot shows the Ask.com search results for the query "chevrolet". The search bar on the left contains "chevrolet" and has a search button. Below the search bar are sections for "Narrow Your Search" (listing categories like Chevrolet History, Parts, Camaro, Corvette, Cavalier, Impala, Truck Parts) and "Expand Your Search" (listing related brands like Ford, General Motors). A "Related Names" section lists "Louis Chevrolet".

The main search results area shows "Showing 1-10 of 18,640,000" results. The top result is "Latest News: Chevrolet" with a sub-link "Truex wins pole for NASCAR Dickies 500 (PA SportsTicker) Yahoo! 48 minutes ago" and another link "NASCAR-Nextel Cup-Dickies 500 Lineup KVIA.com 24 minutes ago". Below this is a sponsored result for "Chevrolet en DeMotores" with the text "Comprá acá el Chevrolet que buscás más fácil y rápido" and the URL "www.DeMotores.com.ar".

Other results include "Chevrolet" (Official GM site), "Corvette: Cars: Chevrolet" (Key features and pricing for Corvette), and "Chevrolet Europe" (Official site of Chevrolet's European operations).

On the right side, there are three highlighted sections: "News Images" featuring a "1963 Chevrolet Corvette Coupe 'Rondine' Concept Car...", "Video" featuring a "Chevrolet Commercial" and a "Police Search For Kidnapped Lo..." video, and "Business Listings" with a search box for "chevrolet dealer" and a "Go" button.

» Y los usuarios?

» Algunos datos

- Las mayoría de las búsquedas son de términos sencillos (el 25% tiene 2)
- Sólo ~10% de las búsquedas son de términos complejos (el 25% tiene 2)
- La mayoría sigue siendo de términos sencillos (el 25% tiene 2)
- Poco refinamiento (el 25% tiene 2)
- Sólo examinar los primeros resultados (el 25% tiene 2)
- La interfaz de usuario es importante (el 25% tiene 2)

La Única Solución

(al menos por ahora)

=

Motores de Búsqueda WEB

*“O cómo adivinar, y rápido,
lo que la mitad del planeta
quiere decir y lo que la otra
mitad quiere saber”*

Ricardo Baeza-Yates

» ¿Para qué?

- **Información** (el 25% tiene 2)
- **Navegación** (el 25% tiene 2)
- **Transaccional** (el 25% tiene 2)

» La Web en la Universidad (estudiantes)

» Acceso a recursos educativos "casi" ilimitados

15 millones de libros en los próximos 10 años

- Ver un libro sin tenerlo:
- Live Search Académicos y
- Conferencias, congresos, seminarios...
- Wikis/foros/Bibliotecas digitales
- Software Libre
- Y para los informáticos:



» La Web en la Universidad (estudiantes)

» Existen algunas cuestiones a tener en cuenta:

- Confianza/reputación de las fuentes
El concepto de reputación != PageRank
- La “tentación” del “*copy&paste*”
- Documentos con *copyright*
- Sitios que introducen *ruido*: multiventas
La clave – por ahora – está en la consulta

» La Web y las organizaciones

» Hoy es una desventaja que una organización no posea un sitio web

Estar:

- Cada vez

Igualmente, no alcanza con tener presencia...

Hacer:

- Permite mostrar características, precios)
- Facilita la

Hay que lograr que los usuarios nos encuentren!

Ser:

- Una empresa con una imagen no más seria, formal, innovadora (mejor im

» ¿Por qué es importante que su sitio esté bien posicionado?

» Los Motores de Búsqueda:

- Son la primera interface entre los usuarios y la web
- Son usados por más del 90% de los usuarios para “navegar”
Generan ~ 90% del tráfico a un sitio
- Atraen mayor cantidad y diversidad de usuarios que cualquier otro sitio/servicio

» Acerca de la posición en la lista de resultados:

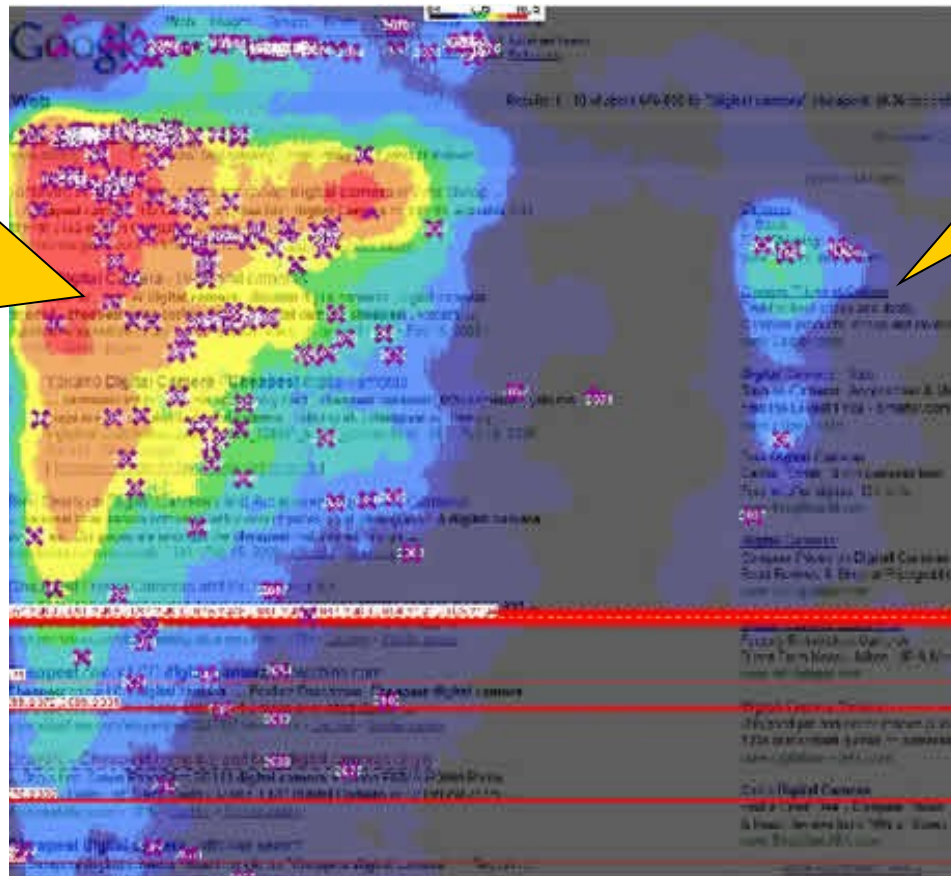
- Los 30 primeros sitios atraen el 90% del tráfico
- Su sitio es útil SOLO si la gente lo utiliza. Pero para ello deben primero encontrarlo
- Si su posición está +30 el sitio es funcionalmente “invisible”
- Un usuario que encuentra un sitio es un cliente potencial

» ¿Por qué es importante que su sitio esté bien posicionado?

» A dónde miran los usuarios?

Lado izquierdo

- 1 - 100%
- 2 - 100%
- 3 - 100%
- 4 - 85%
- 5 - 60%
- 6 - 50%
- 7 - 50%
- 8 - 30%
- 9 - 30%
- 10 - 20%

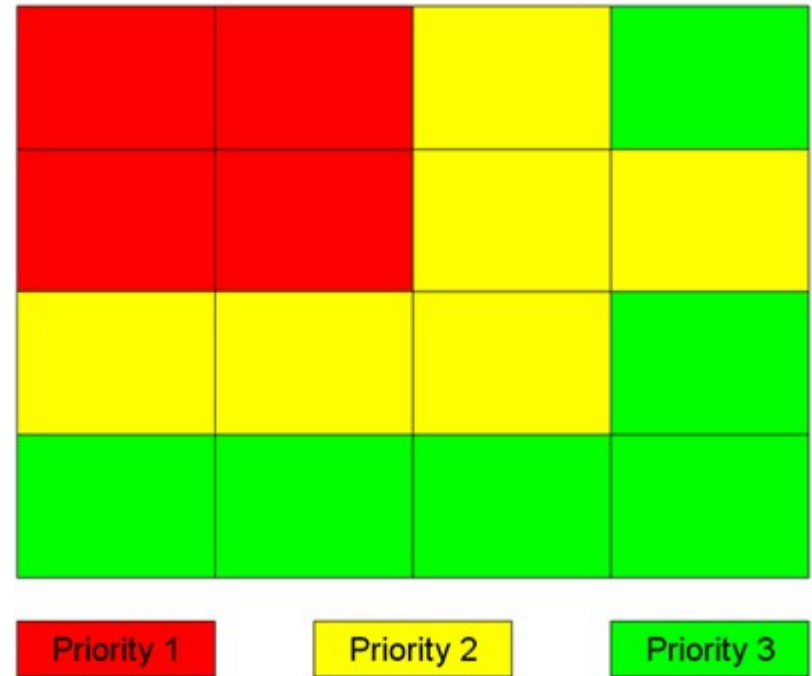
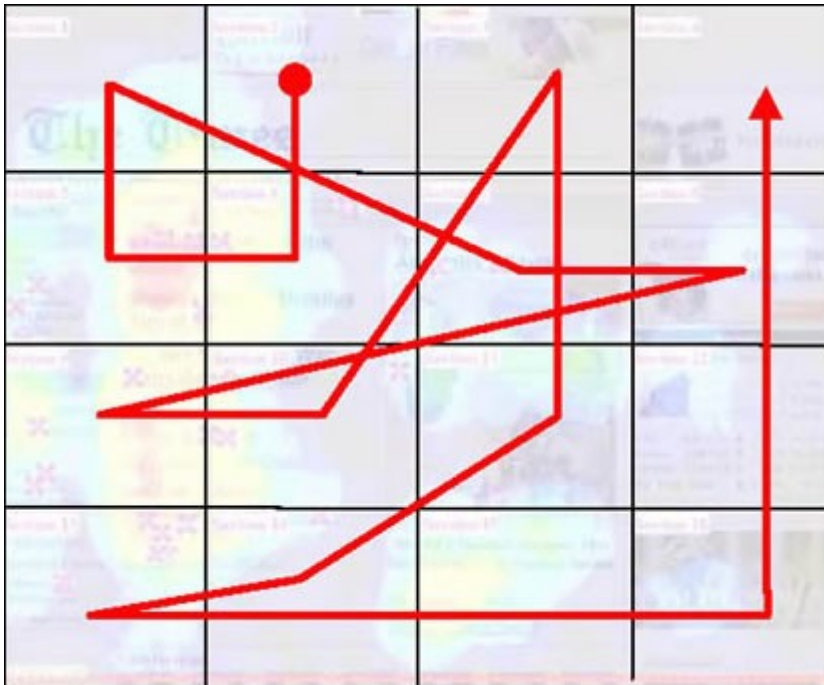


Lado derecho

- 1 - 50%
- 2 - 40%
- 3 - 30%
- 4 - 20%
- 5 - 10%

» ¿Por qué es importante que su sitio esté bien posicionado?

» A dónde miran los usuarios?

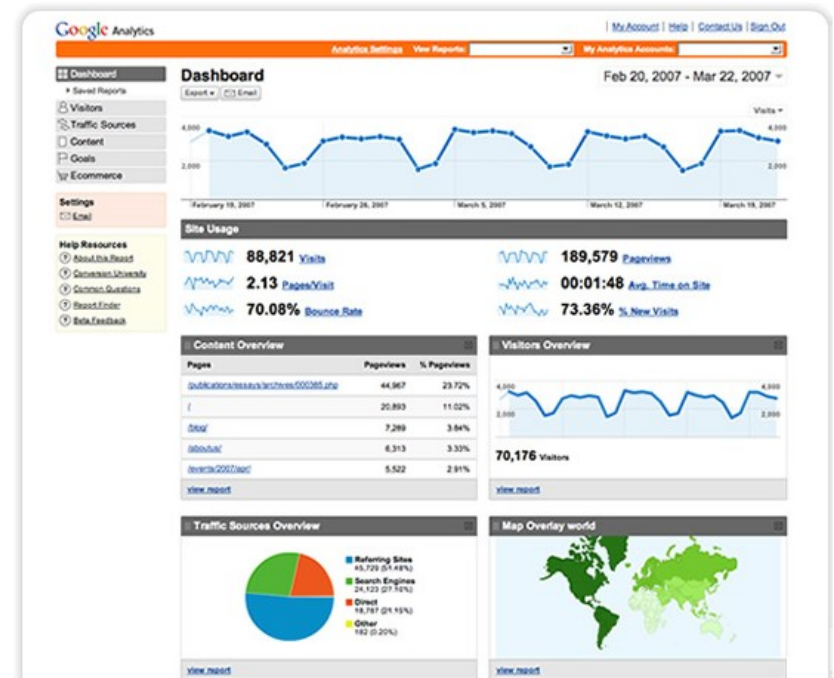


» Bien, ¿Cómo se hace?

» **Fácil (pero poco)**
Estadísticas de Uso
Análisis de tráfico

» **Mediana Complejidad**
SEM

» **Mayor Complejidad**
Minería de la Web
Search Analytics



» Search Engine Marketing

» **Idea fundamental:**

“Que su sitio web sea visible a TODOS”

SEO – Search Engine Optimization

SEA – Search Engine Advertising

» Search Engine Optimization

» Optimización para Motores de Búsqueda

Objetivo: Situar una página en los primeros lugares

Cómo se hace? Hay diversas técnicas (combinadas)

Alta en directorios importantes

Conseguir enlaces entrantes

Mejorar la accesibilidad (js y flash)

Títulos únicos y pertinentes

Visibilidad: Sitemap

Actualizaciones frecuentes

» Search Engine Optimization

» Algunos ejemplos:

	Suma	Resta
Términos	En la tag <title>, en la URL	Relleno
Enlaces	Desde sitios .edu o .gov	Patrón de Cross-linking
Meta Tags	Meta Tag “Description”	Meta Tag “Refresh “
Contenido	Actualizaciones frecuentes	Texto invisible
Otros	Accesibilidad	Flash en exceso

Cuidado con el spamdexing!!!

» Search Engine Optimization

» Diseñar basado en estándares



» Comprender primero cómo operan los MB

Webmaster Guidelines (Google)

Search Engine Indexing (Yahoo!)

Site Owner (MSN/Live)

» Search Engine Advertising

» Publicidad en Motores de Búsqueda

Web | Images | Video | Local | Shopping | more ▾

marketing Search Options ▾

Search In: the Web pages from Argentina | 1-10 of 988,000,000 for marketing (About) - 0.02 sec

Also try: [yahoo search marketing](#), [marketing plan](#), [marketing strategy](#), [More...](#)

PPR

SPONSOR RESULTS

- [Marketing - Factura Electrónica](#)
www.gire.com/corporativo/c_fe.htm - Adquiera mayor productividad, ahorro de costos, reducción de plazos...
- [Licenciatura en Marketing - UCES](#)
www.uces.edu.ar - Universidad de Ciencias Empresariales y Sociales, carreras,...

1. [Marketing - Wikipedia, the free encyclopedia](#)
Direct **marketing** / Personal Sales. Product placement / Public relations ... **Marketing** is a social process which satisfies consumers' wants. ...
Quick Links: [Introduction](#) - [Two Levels of Marketing](#) - [Four Ps](#)
en.wikipedia.org/wiki/Marketing - 94k - [Cached](#)

2. [Marketing - Marketing Strategy - Marketing Plan - Marketing Careers and ...](#)
Articles and links for **marketing** strategy, plans, careers, and jobs.
marketing.about.com - 29k - [Cached](#)

3. [American Marketing Association](#)
Dedicated to serving the educational and professional needs of **marketing** executives.
www.marketingpower.com - 48k - [Cached](#)

4. [KnowThis: For Marketing, Market Research, Internet Marketing ...](#)

PPC

SPONSOR RESULTS

Marketing en UP
Estudiá en la Universidad de Palermo, Facultad de Ciencias...
www.palermo.edu/

Libros Sobre Marketing - BuscaPé.com
Más de 450 tiendas en un solo sitio compará precios en BuscaPé.
www.BuscaPe.com.ar/

Marketing Multinivel
Gane dinero extra, máxima ganancia con mínima inversión, su propio...
www.trabaje.net/avision

Marketing Directo
Aviones publicitarios, remolque de carteles gigantes, Mar del Plata, ...
publicidadaaerea.com

Licenciatura en Marketing
Estudiá en la UdeMM, título intermedio al finalizar tercer año.
www.udemm.edu.ar

» Search Engine Advertising

» Algunas cuestiones a tener en cuenta

- Elección del sistema publicitario (Adwords, Overture)
- Sitios donde aparecer: Redes de afiliados
- ¿Qué palabras “comprar”?
- ¿Con qué presupuesto por período (precio a pagar por click)
- Segmentación a utilizar. Geográfica es posible!!!

» Se puede ir por más...

» **No siempre alcanza con el SEM**

» **Hay información que no está en la “superficie”**

» **La ciencia aporta técnicas más sofisticadas para “descubrir” esta información:**

Minería de la Web

Permite descubrir tendencias y relaciones (por ejemplo, en el comportamiento de los usuarios)

Aporta información para mejorar diferentes aspectos del objeto de estudio

» Minería de la Web

» Ejemplos: Información en la “superficie”

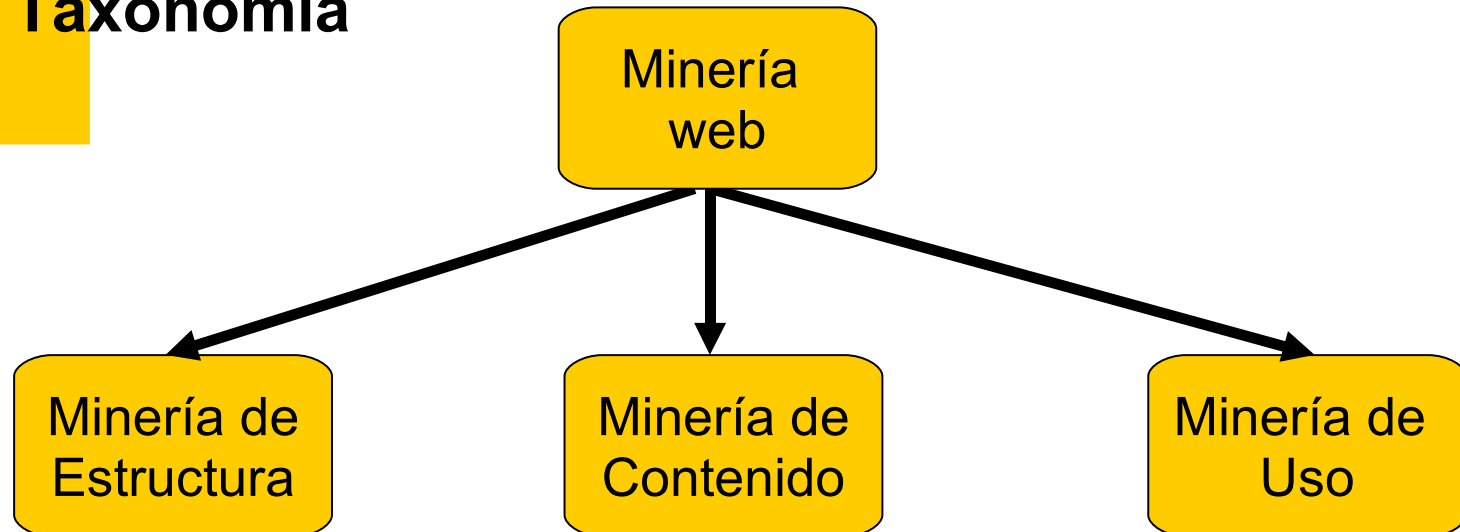
- Día y hora que se solicita una página
- Tiempo de duración de una visita (total y por página)
- País, navegador y sistema operativo utilizado por el usuario
- Desde donde se llega a una página
- Si se accedió a través de un buscador, palabras clave usadas

» Ejemplos: Información “profunda”

- Patrones de comportamiento de los usuarios
 - Relaciones entre la “visibilidad” de una página y su ubicación
 - Preferencias de los usuarios
 - Relaciones entre búsquedas, resultados y compras
-
- Caso de estudio: **Amazon**

» Minería de la Web

» Taxonomía



Técnicas

- Clasificación, agrupamiento, asociación, discriminación, predicción

Aplicaciones

- Patrones de navegación => Rediseñar sitios
- Conjuntos de "objetos" que ocurren juntos => Recomendar productos
- Patrones de comportamiento => Ofrecer productos acordes a perfiles

» Search Analytics

» Pero también debemos preocuparnos por las búsquedas sobre nuestro sitio

- Es cada vez más relevante frente al browsing
- Frente a situaciones que generan compras masivas (por ejemplo, navidad), muchos usuarios prefieren realizar una búsqueda cuando llega a una tienda online (compra rápida).
- Los usuarios que no encuentran rápidamente lo que necesitan tienen 3 veces más probabilidades de abandonar el sitio



The screenshot shows the header of the Clarín.com website. The search bar is highlighted with a red oval. The search bar contains the text "En Clarín" and a dropdown arrow. To the right of the search bar is a red button labeled "Buscar". Below the search bar, the website displays the date "lunes 12 noviembre 2007" and the time "10:40". The weather for Buenos Aires is shown as "T 16.4° ST 16.4° H 81%". The navigation menu includes "Deportes", "El Mundo", "El País", "La Ciudad", "Sociedad", and "Ediciones Anteriores".

Clarín.com  lunes 12 noviembre 2007 10:40 Archivo

286.410 Usuarios únicos

En Clarín  **Buscar**

 Buenos Aires T 16.4° ST 16.4° H 81%  Av. San Juan 

Deportes El Mundo El País La Ciudad Sociedad Ediciones Anteriores

» Search Analytics

»Es una forma de *”conversación con nuestros clientes”* (Escuchar y comprender sus necesidades)

- ¿Qué buscan?
- ¿Con qué término?
- ¿Cuántos?
- ¿Cómo se comportan?
- ¿Están satisfechos?

Recordemos:

**No podemos mejorar
si no sabemos
qué y cómo
hacerlo (intuición?)**

» Para ir finalizando...



» Los motores de búsqueda no son 100% correctos!

» Dado el contexto, ofrecen resultados satisfactorios
Evaluar cuál es la herramienta adecuada!

» La optimización es un proceso dinámico, un ciclo:

diseñar, implementar, optimizar, analizar



» Debemos considerar el entorno web de mi sitio, de mis visitantes (clientes?) o ambos (*Por ej., webs nacionales*)

“No hay garantías en la vida, solo oportunidades...”

Michael Michalko

» Para ir finalizando...



- » Estamos en un momento de evolución en las estrategias de búsqueda
- » El tamaño del MB ya no es lo + importante, sino resultados mejores
- » Google no “*parece*” ir en esa dirección (Maps, Suggest, Books, etc)

¿Cuál es el rumbo de las búsquedas en la web? [Baeza-Yates]

Modos de búsqueda más avanzados (Q&A, investigar, aprender)

Incorporar Semántica

Redes Sociales (PeopleRank)

Personalización e información contextual

Multimedia y múltiples lenguajes

ROLLYO
BETA



¿Y las aplicaciones emergentes?

Alternativas de búsquedas

Para móviles

Locales/escriptorio

Nichos verticales

Agrupamiento de Resultados

Selección de Fuentes



Lexxe[™]alpha

Feedster 

 **Technorati**[™]

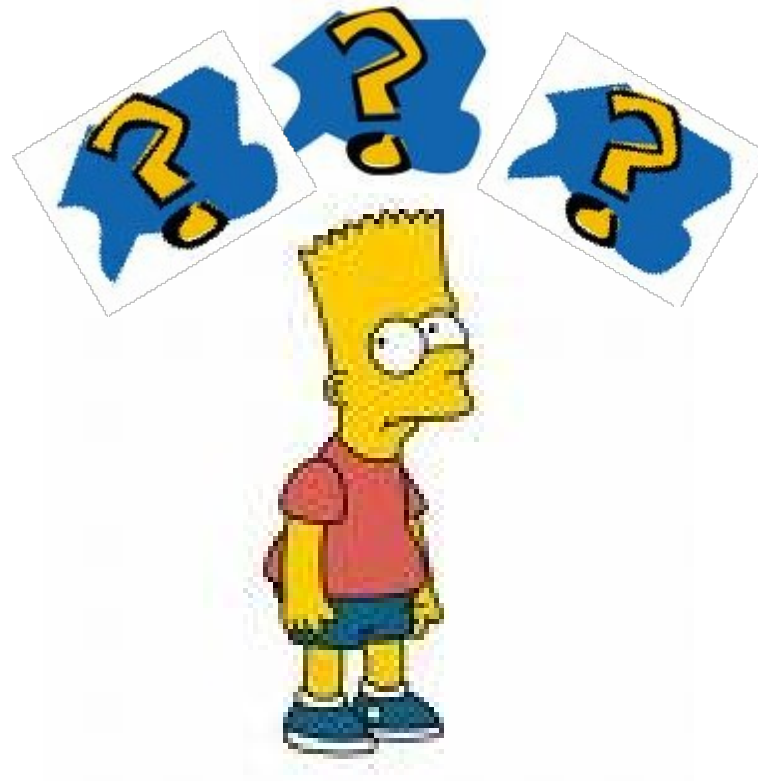
» Para ir finalizando...



» Y las búsquedas en la Web 3.0?

- » Se basa en la idea de agregar una “capa de sentido” (significado) a la web
- » Es pensar mas allá y tratar de construir un espacio web “mejor” conectado que permita “gestionar” conceptos semánticamente vinculados.
- » ¿Qué le gustaría/necesitaría que haga un servicio web?

» ¿PREGUNTAS?





Ciclo de Extensión Centro Regional Chivilcoy

Muchas gracias!

Gabriel H.Tolosa
tolosoft@unlu.edu.ar