

El Impacto de los Buscadores en el Futuro de la Web

Ricardo Baeza Yates
ricardo@baeza.cl

Centro de Investigación de la Web
www.ciw.cl
Depto. de Ciencias de la Computación
Universidad de Chile

El Impacto de los Buscadores en el Futuro de Internet

Ricardo Baeza Yates
ricardo@baeza.cl

Centro de Investigación de la Web
www.ciw.cl
Depto. de Ciencias de la Computación
Universidad de Chile

El Impacto de los Buscadores en el Futuro de los Medios de Comunicación

Ricardo Baeza Yates

ricardo@baeza.cl

Centro de Investigación de la Web

www.ciw.cl

**Depto. de Ciencias de la Computación
Universidad de Chile**

“Al borde del mar,
entre la tierra de los átomos
y el mar de los bits,
estamos enfrentando el desafío de
reconciliar nuestra ciudadanía dual
en el mundo físico y en el mundo digital”

Hiroshi Ishii, MIT Media Lab

Y la marea está subiendo

“Hoy en día, la gente espera ser capaz de encontrar la información en un instante, una suposición que hubiera sido impensable antes de que la Web se convirtiera en un medio masivo”.

Mike Yamamoto, CNET.com, Abril 2003

“Lo que encontramos cambia lo que seremos”

Peter Morville

Ambient Findability, O'Reilly, 2005

Agenda

- La Web Mundial y Chilena
- Impacto en el Diseño Web
- Impacto Social
- Historia de los Buscadores
- La Búsqueda como Negocio:
Publicidad Dirigida
- Oportunidades Futuras

La Web

- Creada en 1989, emerge en 1993 y se consolida en pocos años
- Libertad de publicación
- Más datos publicados en los últimos 5 años que todos los escritos realizados por la humanidad hasta el siglo XX (50Tb/día)
- Exceso de datos e información
- La solución actual: búsqueda

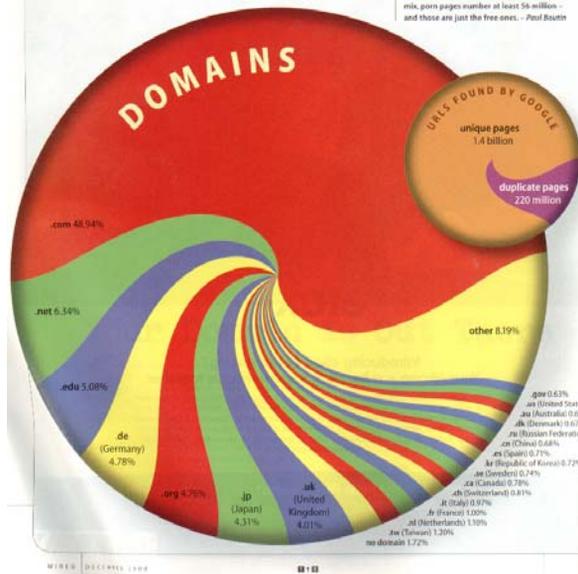
La Web Mundial

- Más de 60 millones de servidores Web
 - 1 de cada 4 computadores en Internet
 - Apache (68%) y Microsoft (21%)
- ¿Más de 6 mil millones de páginas estáticas?
- Muchas más páginas dinámicas (URL/contenido)
- Más de 600 mil millones de archivos
- Inglés es el idioma de más de la mitad de ellas
- Muy heterogénea, redundante y desorganizada

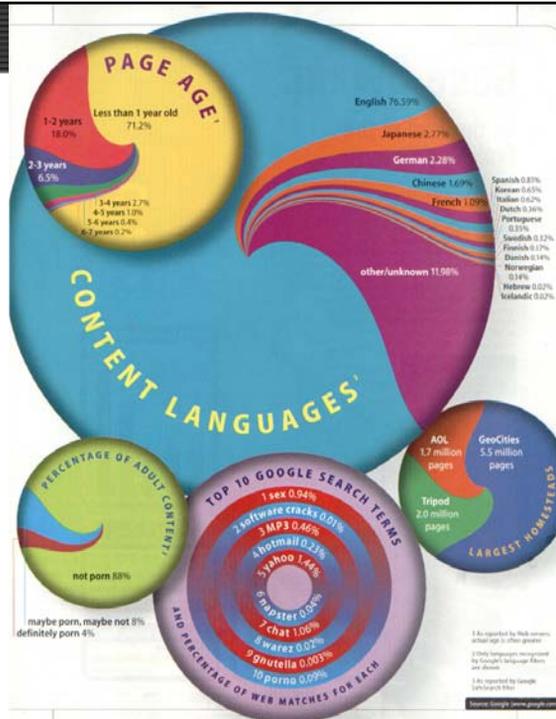
1.6 Billion Served

The Web According to Google

How big is the World Wide Web? Google's recent gigapixel of 1.6 billion URLs suggests popular estimates have been too conservative. By contrast, the happy talk of globalization ignores the ongoing dominance of American content: Germany's .de domain, number two after the US-centric .com and company, is home to less than 5 percent of Google's page population; and at 3 percent, Japanese-language pages run a distant second to sites in English. Not so surprising is the spread of adult content in all available tongues: With a 4 percent share of the mix, porn pages number at least 50 million — and those are just the free ones. —Paul Boutin



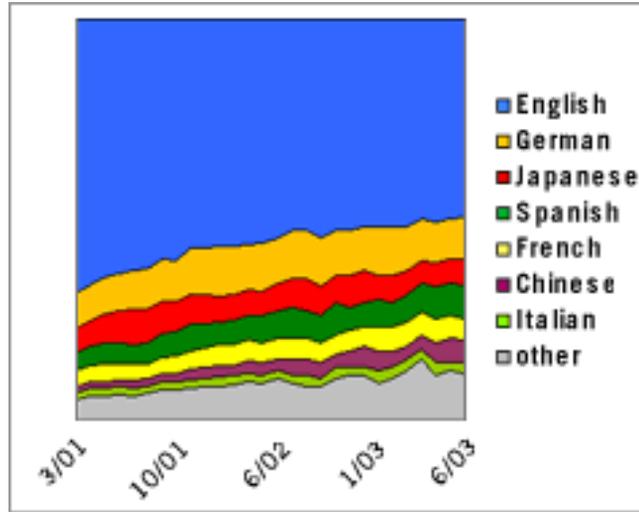
Web según
gle (1)
2000)



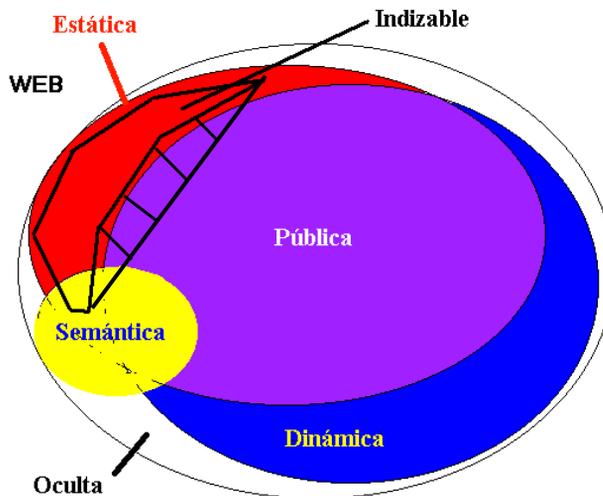
Web según
gle (2)

Idiomas de la Web

(según Google en 2003)

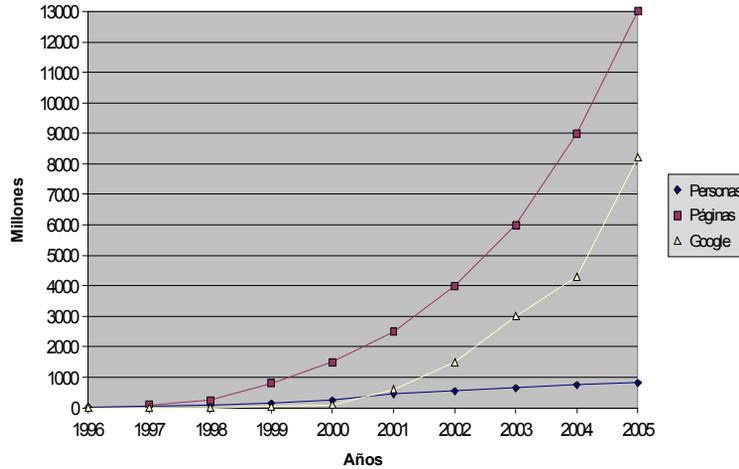


Las Múltiples Caras de la Web



Evolución

Crecimiento



La Web Chilena

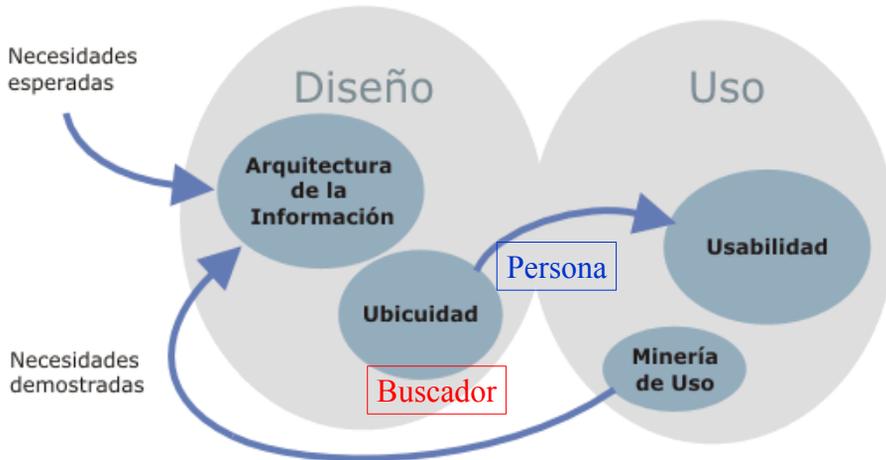


Estudio del Centro de Investigación de la Web y TodoCL.cl

Año	2000	2001	2002	2003	2004
Páginas	630.415	794.218	2.214.253	3.135.089	3.254.137
Sitios	7.483	21.207	39.965	56.699	53.528
Dominios	6.288	19.389	35.390	33.981	47.469

¡El texto cabe en un disco duro!

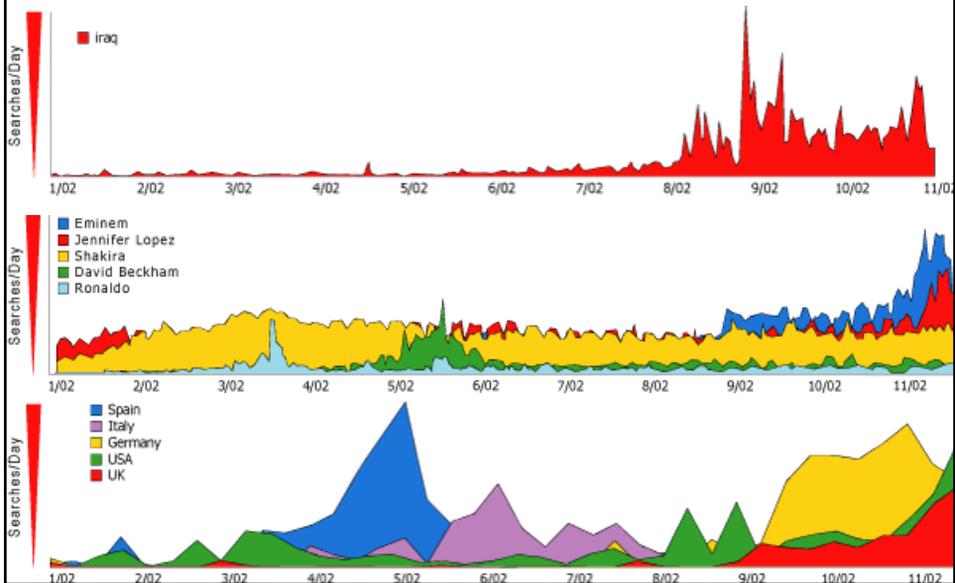
Impacto en Sitios Web



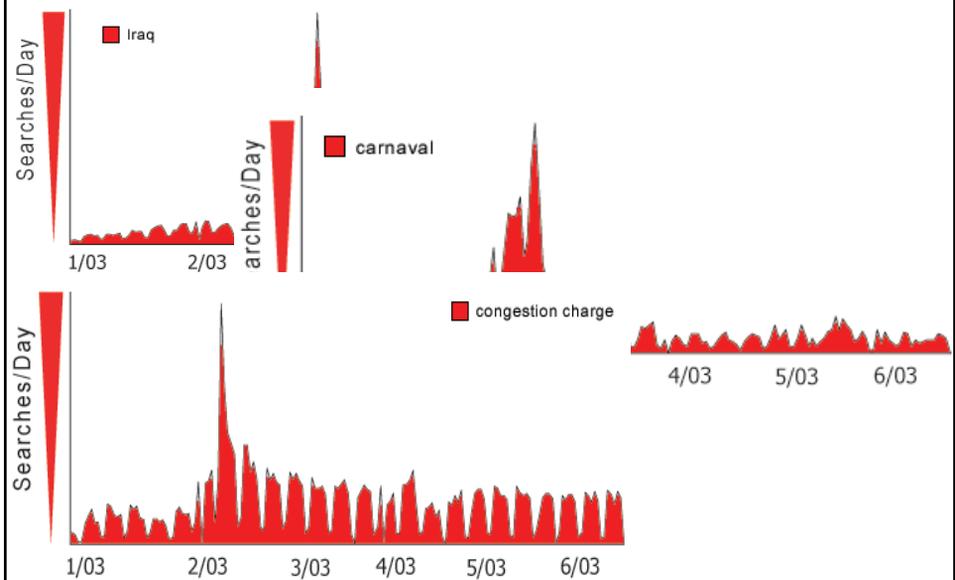
Impacto Social

- Globalización de datos, opiniones,
- **Pérdida de privacidad:**
cuidado con lo que se escribe
- **Acto de fé:** resultados relevantes pero no necesariamente los mejores
- **Navegación:**
Google por defecto+sitio+“tengo suerte”
- **Lenguaje:** “I will google you”, *spell-checker*

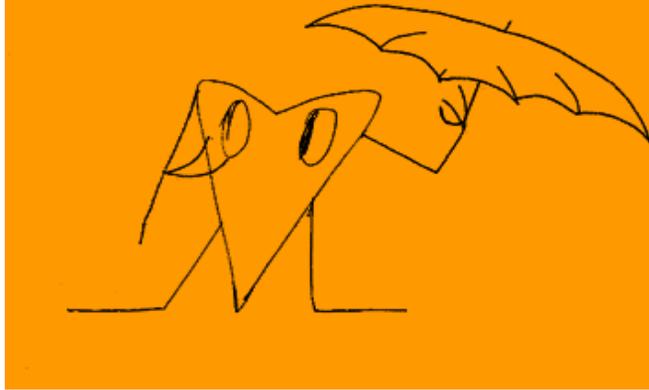
Comportamiento Social (2002)



Comportamiento Social (2003)



Webiar Humanum Est



Nicanor Parra

Historia de los Buscadores

- Primera Generación
 - 1993 Excite (Stanford/Kleiner-Perkins)
 - 1994 WebCrawler
 - 1994 Yahoo! (Primer directorio – semi-manual)
 - 1994 Lycos
 - 1994 Infoseek (Servicio de suscripción)
 - 1995 AltaVista (DEC Labs)
 - 1996 Inktomi

Historia de los Buscadores

- Segunda Generación
 - 1997 Ask Jeeves (indexar preguntas)
 - 1997 Goto/Overture (pagar por ranking)
 - 1998 DirectHit (ranking por clicks)
 - 1998 Google (ranking por links - PageRank)
 - 1999 Fast/AlltheWeb (escalabilidad)
 - 2000 Yahoo! cambia Inktomi por Google
 - 2000 Overture externaliza publicidad
 - 2002 Google externaliza publicidad (AOL)

Historia de los Buscadores

- Consolidación
 - 2002 Yahoo! compra Inktomi (235 MUS\$)
 - 2003 Overture compra AltaVista y Fast/AlltheWeb (500 MUS\$)
 - 2003 Microsoft anuncia su propio buscador
 - 2003 Yahoo! compra Overture (1630 MUS\$)
 - 2004 Aparece A9 (Amazon: buscar en libros)
 - 2004 Yahoo! cambia Google por Inktomi
 - 2004 Los laboratorios de investigación fuera de EEUU se consolidan (China, India, Europa)

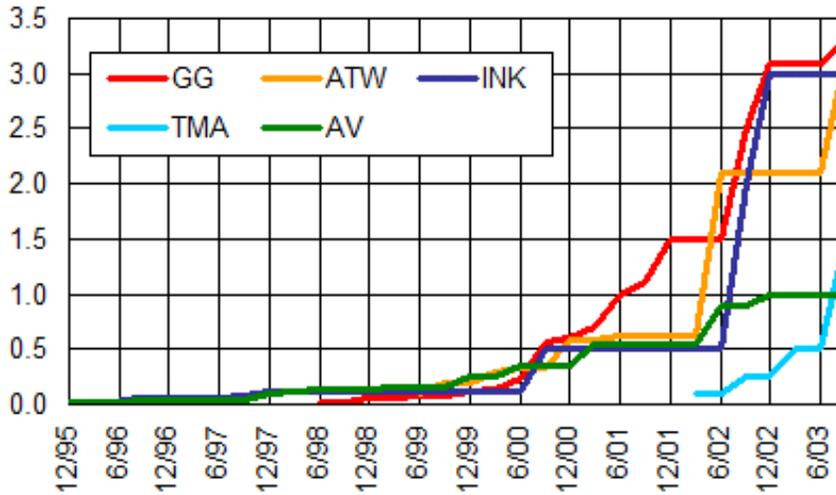
Historia de los Buscadores

- La Batalla Comienza
 - Abr 2004: Google comienza Gmail
 - Aug 2004: Google sale a bolsa: 1670 MUS\$ (85)
 - Feb 2005: Microsoft lanza su propio buscador
 - Jul 2005: Yahoo! Research se refuerza
 - Jul 2005: Google contrata vicepresidente de Microsoft para montar Google China
 - Sep 2005: E-Bay compra Skype for US\$ 2100 mill.
 - Sep 2005: Google consigue 4200 MUS\$ en bolsa (285)

Historia de los Buscadores

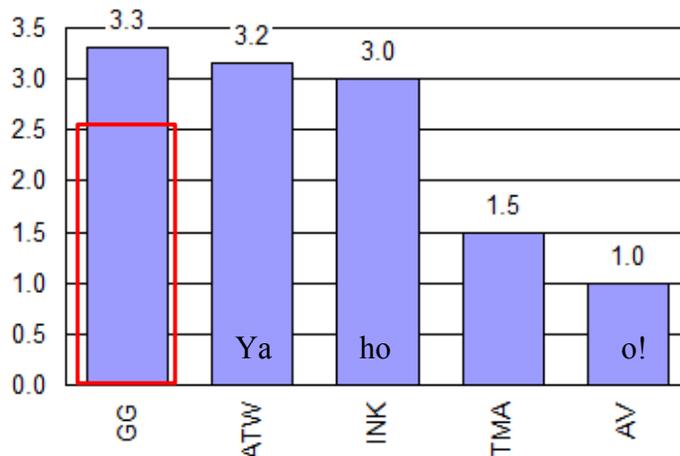
- ¡También en Latinoamérica!
 - Abr 2005: Yahoo! compra TeRespondo (Colombia)
 - Ago 2005: Google compra Akwan (Brasil) e inicia desarrollo de tecnología en la región
 - Ago 2005: UOL (Brasil) queda entre dos fuegos
- ¿Estamos fuera de la cancha?

Crecimiento de los Buscadores



Tamaño de los Buscadores

- SearchEngineWatch (2003)



¿Es Google Dios?

- Artículo en NY Times 29/6/2003
- Wi-fi + Google es como dios:
es inalámbrico, está en todas partes y conoce todo
- Busque cualquier cosa, donde y cuando quiera
- 40 días después: 1400 páginas sobre el tema
- **Sesgar el contenido:** search+cut+merge+publish
e.g.: “Bush fumando con Aznar en la Casablanca”
(40 páginas de Mayo a Diciembre del 2003)

Google no es Dios

- Cualquier tecnología avanzada es magia
- Google no es omnipotente ni eterno, y además es ciego y estúpido
... pero siempre responde
aunque no siempre lo que uno quiere
- Google recuerda hasta cuando la página ya no existe (*backup vs. copyright*)

Google es casi Omnisciente

- Conoce más de 8000 millones de páginas
- Busca en texto, imágenes, noticias, precios, catálogos,
- Sabe mostrar mapas, hacer cálculos, reconocer conjuntos, traducir texto,
- Tecnología generada por 400 ingenieros, instalada en decenas de miles de PCs con Linux y vendida por un equipo de 600 personas
- Tecnología similar existe

Google no es perfecto

- Sesgo negativo para páginas nuevas
- Sitios populares se hacen más populares
 - Los ricos se hacen más ricos
- No sabe de idiomas
- Nosotros no somos perfectos
 - Malas descripciones (títulos, enlaces, metadatos, etc.)
 - Abuso de páginas dinámicas
 - Abuso de duplicación

Problemas

- Tamaño, crecimiento y cambio permanente
- Diversidad y duplicidad
- Ignorancia de los usuarios
- Spam de metadatos, contenido, enlaces, etc:
 - “weapons of mass destruction”
 - “french military victories”
 - “miserable failure”
- Recorrer la Web y encontrar las páginas más relevantes

Desafíos

- ¿Otro Medio de Comunicación guiado por Popularidad?
- Nuevos canales: blogs, portales verticales, etc.
- Búsqueda Integrada (a9.com) y completa
- Servicios donde el buscador es la herramienta
- Búsqueda Colaborativa (e.g. eurekster.com)
- La Web Semántica
- Búsqueda Multimedial (WebFaces)

Septiembre 22, 2004



Departamento de Ciencias de la Computación
 Universidad de Chile

[Presentación WebFaces](#) |
 [Descripción](#) |
 [Reconocimiento Facial](#) |
 [Detección de Caras](#) |
 [Detección de Piel](#) |
 [Publicaciones](#) |
 [Preguntas](#)

[WebFaces](#) |
 [Imágenes de Referencia](#) |
 [Buscar](#) |
 [Resultados Búsqueda](#)



Busqueda por aproximación...



0.184652	0.320111	0.322731	0.346323	0.352621	0.357363
----------	----------	----------	----------	----------	----------

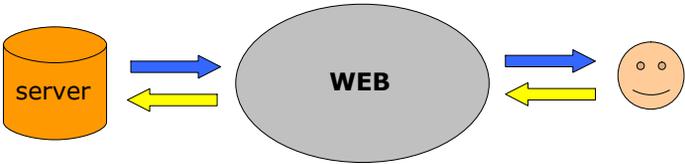
El numero correspondiente a la imagen es el indice de parecido -- Base de Datos Feret: 762 componentes

Tiempo de busqueda: 0.100 segundos



Departamento de Ciencias de la Computación
 Universidad de Chile

El Mercado



- ▶ **Interactividad** → detectar el interés del usuario
- ▶ Asociar la publicidad al interés del usuario
- ▶ ¿Cuándo en la Web una persona demuestra interés específico en algo?
- ▶ Al buscar
 - ¡La búsqueda es la clave!

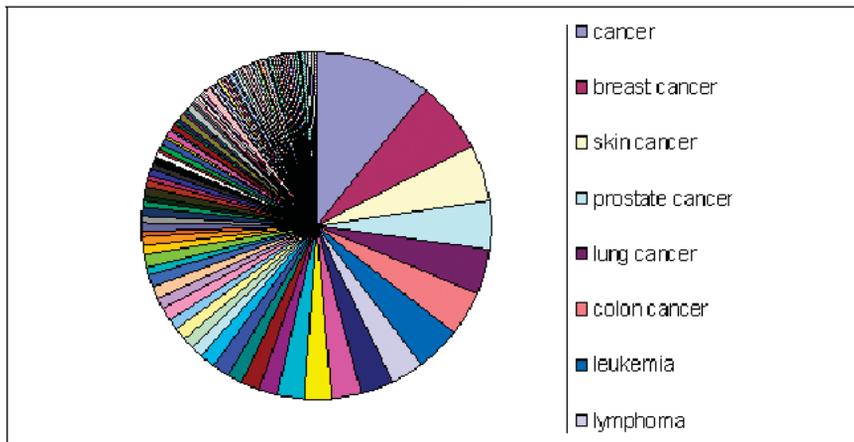
Sesión Típica

Google España (2003)

- Dos consultas
- Dos palabras
- Dos páginas de respuestas
- Dos clicks
- Tres tipos:
 - Información (50%)
 - Transacción (30%)
 - Navegación (20%)

1. **letizia ortiz**
2. **terra**
3. **gran hermano**
4. **juegos**
5. **coches**
6. **britney spears**
7. **dibujos**
8. **renfe**
9. **iberia**
10. **real madrid**

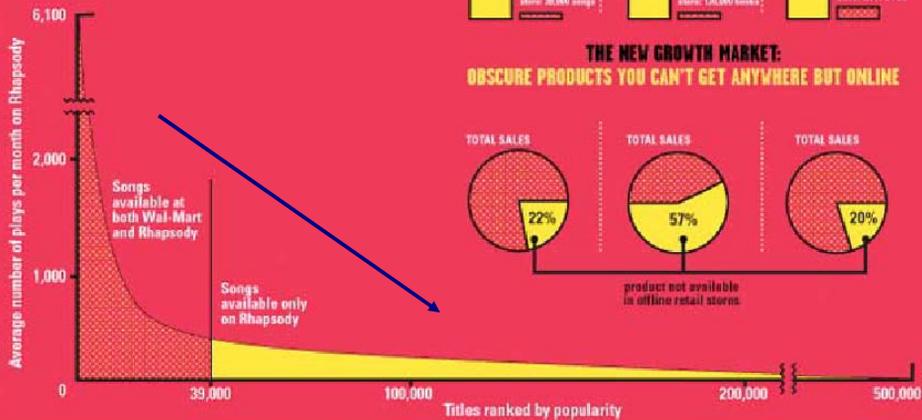
Diversidad de Búsqueda



– Como casi todo en la Web: Ley de Zipf,
Pocos tienen mucho y muchos tienen

ANATOMY OF THE LONG TAIL

Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-Mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in yellow) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.

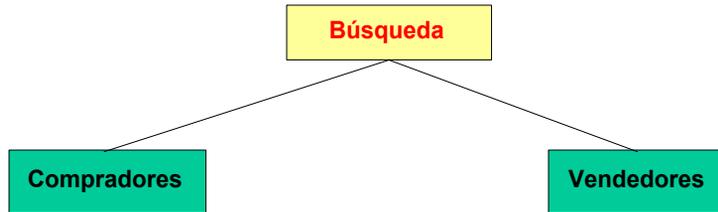


Source: Erik Brynjolfsson and Jeffrey He, MIT, and Michael S. Smith, Carnegie Mellon; Barnes & Noble; Netflix; Paul Networks

Modelos de Negocios

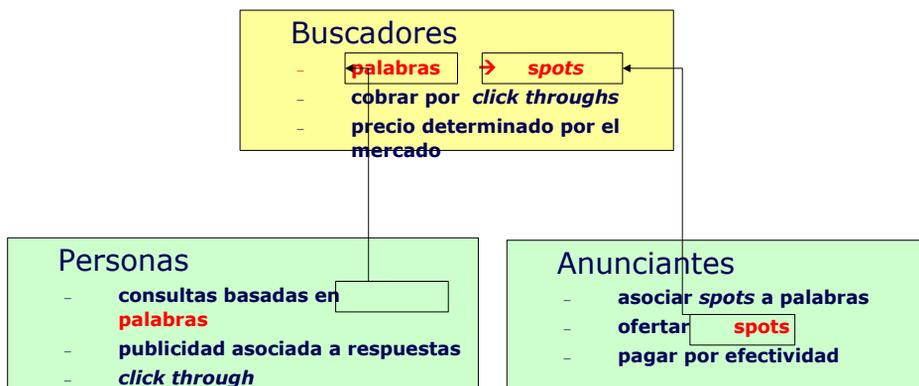
- Venta de publicidad dirigida y focalizada
 - medio interactivo, la publicidad tradicional no funciona
 - *adwords, adsense*
 - *pay-per-ranking, pay-per-click*
- Servicio de búsqueda externalizado (*outsourcing*)
- Búsqueda en Intranets (e.g. *Google-in-a-box*)
- Licenciamiento de tecnología
- API de búsqueda como Web Service
- Búsquedas expertas

El Concepto



- ▶ Esencialmente, **buscar** es un método para asociar compradores con vendedores
- ▶ **Compradores** son personas que adquieren bienes y servicios
- ▶ **Vendedores** son compañías que venden bienes y servicios

El Modelo de Negocios



ciwi.cl
centro de investigación del web

Departamento de Ciencias de la Computación
Universidad de Chile

La Web [Imágenes](#) [Grupos](#) [Directorio](#) [News](#)

Google [Búsqueda Avanzada](#) [Preferencias](#)

Búsqueda: la Web páginas en español páginas de España

La Web Resultados 1 - 10 de aproximadamente 24.000.000 de hotel. (0,17 segundos)

Pay-per-rankink Enlace Patrocinado

Resultados de noticias que contienen **hotel** - [Vea las últimas noticias de hoy](#)

Premia el público a "Hotel Rwanda" en Toronto - La Hora (Ecuador) - 20 Sep 2004

hotels.com - the best prices at the best places. Guaranteed. - [Traduzca esta página]
Find the Lowest **Hotel** Rates Guaranteed! From Luxury Hotels to Budget Accommodations. Hotels.Com has the Best Deals and Discounts for **Hotel** Rooms Anywhere. ...
[www.hotels.com/ - 57k](#) - [En caché](#) - [Páginas similares](#)

The official site for a Marriott hotel, resort, or suite worldwide ... - [Traduzca esta página]
Official site for a Marriott **hotel**, resort, or suite worldwide at Marriott.com. ... Marriott's Look No Further SM best rate guarantee >> Find a **Hotel**. City. ...
[marriott.com/default.mi - 45k](#) - 21 Sep 2004 - [En caché](#) - [Páginas similares](#)

Habbo Hotel - Home Page - [Traduzca esta página]
Habbo **Hotel**. ... check in 2. promo1.gif promo2.gif, Habbo **Hotel** USA. Habbo **Hotel** is a virtual **hotel** where you meet up with your friends. ...
[www.habbohotel.com/ - 27k](#) - 21 Sep 2004 - [En caché](#) - [Páginas similares](#)

Get low rates on Best Western hotels, motels and suites worldwide. - [Traduzca esta página]
Book a **hotel** room on bestwestern.com to save on international Best Western reservations in over 80 countries. ... Home/Find a **Hotel**, ...
[www.bestwestern.com/ - 49k](#) - [En caché](#) - [Páginas similares](#)

Ofertas en Promovaje
Viajes de todos los estilos:
paquetes, cruceros y vuelos
[www.promovaje.com](#)

Pay-per-click
[www.laterooms.com](#)

Hotel
Precio mínimo garantizado!
Reserve online o en el 900953320
[www.hotels.com](#)

Hotels - Reserva on line
Oferta en hoteles 2, 3, 4 y 5* en España, Europa y resto del mundo
[www.travelprice.es](#)

Hotel Reservations
Book your cheap **hotel** reservations right here! Affiliate.
[www.orbitz.com](#)

Internet

ciwi.cl
centro de investigación del web

Departamento de Ciencias de la Computación
Universidad de Chile

Ad-sense

→

→

→

La Publicidad en Internet Madura

- Más de 2000 MUS\$ en 2004, 4100 MUS\$ en 2005 (6%) y estimado en 5700 MUS\$ el 2006
- Capital a invertir limita a nuevos contendores
- Publicidad por palabras y por efectividad
- 20% de la industria de publicidad en Web (900 MUS\$ en 2006)
- Concentración del tráfico:
Yahoo!, MSN, Google, AOL
- Concentración del negocio:
Tecnología + Tráfico + Canal de Publicidad

Ad Spending by Media: Full Year 2004 vs. 2003¹

Publ

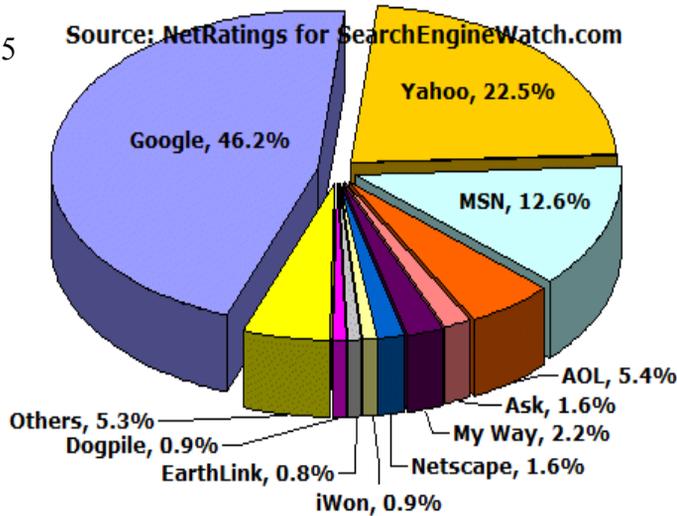
MEDIA	Full Year 2004 (MM)	Full Year 2003 (MM)	% CHANGE
NEWSPAPERS (LOCAL)	\$24,555.50	\$23,018.50	6.70%
NETWORK TV	\$22,522.40	\$20,340.60	10.70%
CONSUMER MAGAZINES	\$21,292.20	\$19,145.90	11.20%
SPOT TV ²	\$17,305.40	\$15,499.20	11.70%
CABLE TV ³	\$14,248.80	\$12,521.60	13.80%
INTERNET	\$7,441.50	\$6,131.90	21.40%
LOCAL RADIO ⁴	\$7,330.50	\$7,250.40	1.10%
B-TO-B MAGAZINES	\$5,214.10	\$5,125.60	1.70%
SYNDICATION - NATIONAL	\$3,930.90	\$3,395.80	15.80%
HISPANIC MEDIA ⁵	\$3,888.60	\$3,713.60	4.70%
NATIONAL NEWSPAPERS	\$3,255.20	\$3,053.10	6.60%
OUTDOOR	\$3,213.00	\$2,674.40	20.10%
NATIONAL SPOT RADIO	\$2,616.50	\$2,635.20	-0.70%
SUNDAY MAGAZINES	\$1,497.40	\$1,383.80	8.20%
FSI's ⁶	\$1,391.60	\$1,317.20	5.60%
NETWORK RADIO	\$1,027.80	\$1,000.70	2.70%
LOCAL MAGAZINES	\$360.20	\$325.40	10.70%
TOTAL⁷	\$141,091.70	\$128,533.00	9.80%

Source: TNSMedia Intelligence

Los Competidores

Julio 2005

Source: NetRatings for SearchEngineWatch.com



Las Ganancias Aumentan

- ▶ Efectividad de la inversión es más alta que en medios alternativos (TV, radio, diarios)

Key-Word Search Is Working Well...and It's Just Beginning...

		CQ3:02	CQ4:02	CQ1:03	CQ2:03
Overture					
Revenue (MM)	↑	\$173	\$200	\$225	\$265
Y/Y Growth		138%	97%	57%	74%
Advertisers (K)	↑	73	80	88	95
Y/Y Growth		49%	51%	47%	42%
Average Price per Click	↑	\$0.34	\$0.35	\$0.37	\$0.40
Y/Y Growth		62%	52%	54%	33%

Source: Morgan Stanley, Internet Trends, Dec 2003.

Limitaciones

- ▶ **Número de anunciantes.** Sólo 250,000 anunciantes en EEUU
- ▶ **Número de spots.** Lugar reducido a 5-10 spots por consultas que son usados por las ofertas más altas
- ▶ **Calidad.** Spots mejores y más apropiados (por que se paga sólo si hay un click)

Búsqueda Local

- Aprovechar la localidad física
 - **Más anunciantes:** empresas locales de todo tamaño
 - (a) Kelsey Group: 25% de las búsquedas son locales
 - (b) millones de medianas y pequeñas empresas
 - **Más spots:** una consulta como ‘gasfiter’ se puede localizar de muchas maneras, como ‘gasfiter Santiago’, ‘gasfiter Providencia’, etc.
 - **Mayor calidad:** tanto anunciantes como compradores obtienen lo que quieren (más *click throughs* = más dinero)

¿Qué Se Necesita?

- Tráfico Web
- Canal de Publicidad
 - En Chile la publicidad en Internet aún no madura
 - Oportunidad para nuevos actores
- Tecnología
 - Adaptarla a necesidades locales
 - Localizar geográficamente sitios Web y personas

Experiencia Chilena

- Centro de Investigación de la Web (2002-2007)
 - Núcleo Programa Científico Milenio
 - Nivel internacional, líder regional
 - Generador de Tecnología
- TodoCL (2000-2005)
 - www.todo.cl, www.todo.cl.com
 - Páginas del dominio .cl u hospedadas en Chile
 - Más de dos millones de consultas al mes



Tecnología y Consultas Diarias (est.)

Portal	Buscador	Directorio	Publicidad	Mill.
Google	Google	ODP	Google	250
Yahoo!	Yahoo!	Yahoo!	Yahoo!	150
Yahoo!	Yahoo!		Yahoo!	100
MSN Search	Microsoft	Looksmart	Yahoo!	100
AOL	Google		Google	50
AskJeeves	Teoma		Google	30

Entendiendo el Mercado

Compañía	Datos de Compradores	Datos de Vendedores	Tecnología de Búsqueda	Costo del Cambio
Google	Parcialmente (datos de Gmail, Orkut y de direcciones IP)	Parcialmente (anunciantes)	Si	No
Yahoo!	Si (datos de Yahoo Mail, agenda, otros servicios, etc.)	Parcialmente (anunciantes)	Si	Si (inversión en direcciones de email, agenda, sitios de citas, etc.)
Yellow Pages	No (¿dispositivos móviles?)	Si	No	No

Entendiendo el Mercado

Compañía	Datos de Compradores	Datos de Vendedores	Tecnología de Búsqueda	Costo del Cambio
AOL	Si (datos de abonados)	No (tratando pero es difícil)	No	Si
Microsoft	Si (datos de registros de productos, Hotmail, sitios MSN, etc.)	No	Si	Si (inversión en direcciones de email, agenda, sitios de citas, etc.)
Amazon	Si (datos de clientes)	Parcialmente	Avanzando (A9)	No

E-Bay

Consecuencias

- Las compañías que tienen datos de compradores deben obtener los datos de los vendedores
- Las compañías que tienen datos de vendedores deben obtener los datos de los compradores
- Tanto a nivel global como local
- Los que tienen ambos pueden entrar al juego obteniendo tecnología de búsqueda
- Pero lo importante es que todos necesitan **búsqueda**

¿Preguntas?

El Impacto de los Buscadores en el Futuro de Internet

Ricardo Baeza Yates
ricardo@baeza.cl

Centro de Investigación de la Web
www.ciw.cl
Depto. de Ciencias de la Computación
Universidad de Chile